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 multiplicativity 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.

Manuscript received August 15, 2011; revised March 6, 2012; Accepted May 4, 2012. This work was supported in part by National Natural Science Foundation of China under Grant No. 61101130, Sino-Swedish IMT-Advanced and Beyond Cooperative Program under Grant No.2008DFA11780, Grant GRF HKU 7191/11E, the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20110002120059) and by the open research fund of National Mobile Communications Research Laboratory, Southeast University (No. 2011ID02).

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 and is with National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, and is also with the School of Engineering and Science, Jacobs University, Bremen, Germany (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).

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 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 perfect channel state information (CSI) is discussed in [12]. Furthermore, robust design, which takes channel estimation errors into account, is recently investigated in [2], [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
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 $k$th relay has $N_{R,k}$ receive antennas and $N_{R,k+1}$ 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 $\mathbf{a} = [a_1, a_2, \ldots, a_N]^T$ 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 $\mathcal{A} = \{ \pm 1, \pm 3, \ldots, \pm (\sqrt{M} - 1) \}$.

As shown by Fig. 1, at the transmitter, the data vector $\mathbf{a}$ is fed into the a precoding unit which consists of a $N \times N$ feedback matrix $\mathbf{B}$ and a nonlinear modulo operator $\text{MOD}_{\mathcal{M}}(\cdot)$. The square matrix $\mathbf{B}$ is a strictly lower triangular matrix which allows data precoding in a recursive fashion and the $\text{MOD}_{\mathcal{M}}(\cdot)$ is defined as

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

where the symbol $[z]$ denotes the largest integer not exceeding $z$. The nonlinear modulo operator reduces the output signals into a square region $[-\sqrt{M}, \sqrt{M}] \times [-\sqrt{M}, \sqrt{M}]$. In the equation, $\text{Re}(x)$ and $\text{Im}(x)$ 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

$$b_k = a_k - \sum_{l=1}^{k-1} \left\lfloor |\mathbf{B}|_{k,l} b_l + d_k \right\rfloor$$

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

$$\mathbf{b} = (\mathbf{B} + i\mathbf{C})^{-1} (\mathbf{a} + \mathbf{d}) \quad \text{mod} \quad \mathbb{C}$$

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

After the nonlinear operation, the vector $\mathbf{b}$ is multiplied with a precoder matrix $\mathbf{P}_1$ under a transmit power constraint $\text{Tr}(\mathbf{P}_1 \mathbf{R}_b \mathbf{P}_1^H) \leq P_1$ where $P_1$ is the maximum transmit power.
power at the source. When the elements of $a$ are independent and identically distributed (i.i.d.) over the constellation and the dimension of modulation constellation $\mathcal{M}$ is large, $b$ can be considered as i.i.d. \cite{19}, i.e.,

$$R_b = 2(\mathcal{M} - 1)/3I_N \triangleq \sigma_b^2 I_N.$$  \hfill (4)

The received signal $x_1$ at the first relay is formulated as

$$x_1 = H_1 P_1 b + n_1$$  \hfill (5)

where $H_1$ is the channel between the source and the first relay and $n_1$ is additive Gaussian noise with mean zero and covariance matrix $R_{n_1} = \sigma_n^2 I_{N_1}$. At the first relay, the received signal $x_1$ is multiplied by a forwarding matrix $P_2$ and then the resultant signal is transmitted to the second relay. The received signal at the second relay can be written as

$$x_2 = H_2 P_2 H_1 b + H_2 P_2 n_1 + n_2$$  \hfill (6)

where $H_2$ is the MIMO channel matrix between the first and second relay, and $n_2$ is the additive Gaussian noise vector at the second hop with zero mean and covariance matrix $R_{n_2} = \sigma_n^2 I_{N_2}$. Similarly, at the $k^{th}$ relay the received signal is

$$x_k = H_k P_k x_{k-1} + n_k$$  \hfill (7)

with $H_k$ and $n_k$ are the channel and additive noise at the $k^{th}$ hop, respectively. In this paper, we considered slow fading channels with $H_k$ being fixed in each transmission.

The covariance matrix of $n_k$ is denoted as $R_{n_k} = \sigma_n^2 I_{N_k}$. Finally, for a $K$-hop AF MIMO relaying system, the received signal at the destination is

$$y = \left[ \prod_{k=1}^{K} (H_k P_k) \right] b + \sum_{k=1}^{K-1} \left\{ \prod_{l=k+1}^{K} (H_l P_l) \right\} n_k + n_K.$$  \hfill (8)

where $\prod_{k=1}^{K} Z_k$ denotes $Z_K \times \cdots \times Z_1$. In order to guarantee the transmitted data $s$ can be recovered at the destination, it is assumed that $N_{T,k}$ and $N_{R,k}$ are greater than or equal to $N$. \hfill \textit{4}

In practice, the channels $H_k$ are estimated and channel estimation errors are inevitable. Therefore, the channel $H_k$ can be expressed as

$$H_k = \bar{H}_k + \Delta H_k,$$  \hfill (9)

where $\bar{H}_k$ is the estimated channels, and $\Delta H_k$ is the corresponding channel estimation error, whose elements are zero mean Gaussian random variables. Furthermore, the $N_{R,k} \times N_{T,k}$ matrix $\Delta H_k$ can be decomposed by the widely used Kronecker model \cite{7,8,20} as $\Delta H_k = \Sigma_k^{1/2} W_k \Psi_k^{1/2}$, where the elements of the $N_{R,k} \times N_{T,k}$ matrix $W_{k}$ are i.i.d. Gaussian random variables with zero mean and unit variance. The specific formulas of $\Sigma_k$ and $\Psi_k$ are determined by the training sequences and channel estimators \cite{7,8,11,21}.

\begin{itemize}
  \item In this paper, only channel estimation errors are taken into account.
\end{itemize}

\subsection*{B. Problem Formulation}

As shown by Fig.1 at the destination, a linear equalizer $G$ is adopted and is followed by a modulo operator. As the real and imaginary parts of $d$ are both integer multiples of $2\sqrt{\mathcal{M}}$, the effect of $d$ will be perfectly removed by modulo operator at the destination. As a result, estimating $s$ is equivalent to estimating $a$ \cite{14,15}. Thus at the destination, a linear equalizer $G$ is used to detect the data vector $s$. The MSE matrix of the data vector is defined as $E\{(Gy - s)(Gy - s)^H\}$ \cite{15,19}, where the expectation is taken with respect to random data, channel estimation errors, and noise. Following a similar derivation to that in \cite{8}, it can be shown that

$$\Phi(G, \{P_k\}_{k=1}^K, C) = E\{(Gy - Cb)(Gy - Ch)^H\}$$

$$= G[\bar{H}_K P_K R_{x_{K-1}} P_K^H \bar{H}_K^H + \text{Tr}(P_K R_{x_{K-1}} P_K^H \Psi_K)] \Sigma_K$$

$$+ R_{n_K} G^H - \sigma_b^2 G \prod_{k=1}^{K} (\bar{H}_k P_k) C^H$$

$$- \sigma_b^2 \left[ G \prod_{k=1}^{K} (\bar{H}_k P_k) C \right]^H + \sigma_b^2 CC^H.$$  \hfill (10)

where matrices $R_{x_k}$ is defined as

$$R_{x_k} \triangleq E\{x_k x_k^H\} = \bar{H}_k P_k R_{x_{k-1}} P_k^H \bar{H}_k + \text{Tr}(P_k R_{x_{k-1}} P_k^H \Psi_k) \Sigma_k + R_{n_k}.$$  \hfill (11)

It is obvious that $R_{x_k}$ is the covariance matrix of the received signal at the relay. Notice that $R_{x_0} = R_0 = \sigma_b^2 I_N$.

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 $f(\text{MSE}_1, \cdots, \text{MSE}_N) = \sum_{n=1}^{N} \text{MSE}_n$, where $\text{MSE}_n = \Phi(G, \{P_k\}_{k=1}^K, C)|_{n,n}$. For product MSE minimization, the objective function is $f(\text{MSE}_1, \cdots, \text{MSE}_N) = \prod_{n=1}^{N} \text{MSE}_n$. Furthermore, worst-case MSE minimization corresponds to minimizing the objective function given as $f(\text{MSE}_1, \cdots, \text{MSE}_N) = \max_{n=1,2,\cdots,N} \{\text{MSE}_n\}$ \cite{9,13,15,22}. On the other hand, weighted geometric mean MSE minimization corresponds to minimizing the following objective function $f(\text{MSE}_1, \cdots, \text{MSE}_N) = \prod_{n=1}^{N} \text{MSE}_n^w$ with $w_1 \geq w_2 \cdots \geq w_N \geq 0$. Therefore, a unified transceiver design optimization problem can be formulated as

$$\min_{G, P_k, C} f(\text{MSE}_1, \cdots, \text{MSE}_N)$$

s.t. $\text{MSE}_n = \Phi(G, \{P_k\}_{k=1}^K, C)|_{n,n}$

$$\text{Tr}(P_k R_{x_{k-1}} P_k^H) \leq P_k, \quad k = 1, \cdots, K$$  \hfill (12)

where the matrix $C$ is a lower triangular matrix with unit diagonal elements and $P_k$ is the maximum transmit power at the $k^{th}$ node.

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

1. $f(\bullet)$ is an increasing real-valued vector function $\mathbb{C}^N \rightarrow \mathbb{R}$, i.e., for two vectors $u = [u_1, u_2, \cdots, u_N]^T$ and $v = [v_1, v_2, \cdots, v_N]^T$, when $u_n \geq v_n$, we have $f(u) \geq f(v)$.
This property is natural in transceiver design. This is because for two designs resulting in $[\text{MSE}_1, \cdots, \text{MSE}_N]^T$ and $[\text{MSE}_1, \cdots, \text{MSE}_N]^T$, suppose $\text{MSE}_n < \tilde{\text{MSE}}_n$ 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 $x \in \mathbb{R}^n$, let $z_{(k)}$ denotes the $k^{th}$ largest elements of $z$ and $z_{(k)}$ denotes the $k^{th}$ smallest elements of $z$, i.e., $z_{(1)} \geq \cdots \geq z_{(N)}$ and $z_{(1)} \leq \cdots \leq z_{(N)}$. For two vectors $v, u$ whose elements are nonnegative, $v \prec_X u$ is defined as

$$\prod_{i=1}^{k} u_{[i]} \leq \prod_{i=1}^{k} v_{[i]}, \quad k = 1, \cdots, N - 1$$

and

$$\prod_{i=1}^{N} u_{[i]} = \prod_{i=1}^{N} v_{[i]}.$$  

(13)

**Definition 2:** A function $\phi(\bullet)$ is multiplicatively Schur-convex if and only if $v \prec_X u$ implies $\phi(v) \leq \phi(u)$. 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 $D = \{z : z_1 \geq \cdots \geq z_N \geq 0\}$. Then $\phi(\bullet)$ is multiplicatively Schur-convex if and only if for all $z \in D$,

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

is decreasing in $e$ over the following regions

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

(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 $z_1, z_2, \cdots, z_N > 0$. In contrast, **Lemma 1** provides a stronger result and allows elements of $z$ 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 $G$ AND $C$

The linear minimum mean-square-error (LMMSE) equalizer is obtained by setting the differentiation of the trace of (10) with respect to $G^*$ (the conjugate of $G$) to be zero, and we have

$$G_{\text{LMMSE}} = \sigma_0^2 \left[ \prod_{k=1}^{K} (\bar{H}_k P_k) \right]^H \bar{H}_K P_K R_{\bar{x}_K} \bar{P}_K^H R_K^H$$

+ $\text{Tr}(P_K R_{\bar{x}_{K-1}} P_K^H \Psi \Sigma_K + R_{\bar{n}_K})^{-1}$.  

(15)

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

$$\Phi(G_{\text{LMMSE}}, \{P_k\}_k)^K_{k=1}, C) \leq \Phi(G, \{P_k\}_k)^K_{k=1}, C)$$  

(16)

which implies

$$\Phi(G_{\text{LMMSE}}, \{P_k\}_k)^K_{k=1}, C)] \leq \Phi(G, \{P_k\}_k)^K_{k=1}, C]$$  

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

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

$$\Phi_{\text{MSE}}(\{P_k\}_k)^K_{k=1}, C)$$

$$= \sigma_0^2 C (I_N - \sigma_0^2 \left[ \prod_{k=1}^{K} (\bar{H}_k P_k) \right]^H \bar{H}_K P_K R_{\bar{x}_K-1}$$

$$\times P_K^H \bar{P}_K^H + \text{Tr}(P_K R_{\bar{x}_{K-1}} P_K^H \Psi \Sigma_K + R_{\bar{n}_K})^{-1}$$

$$\times \left[ \prod_{k=1}^{K} (\bar{H}_k P_k) \right]^H)^C^H$$  

(17)

based on which the optimization problem (12) becomes

$$\min_{P_k, C} \quad f(\text{MSE}_1, \cdots, \text{MSE}_N)^T$$

s.t. $\quad \text{MSE}_n = [\Phi_{\text{MSE}}(\{P_k\}_k)^K_{k=1}, C)]_{n,n}$

$$\text{Tr}(P_K R_{\bar{x}_{K-1}} P_K^H) \leq P_k.$$  

(18)

From the definition of $R_{\bar{x}_K}$ in (11), it is noticed that $R_{\bar{x}_K}$ is a function of $P$ 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

$$F_1 = P_1 P_{\bar{b}}^{1/2} Q_0^H$$

and

$$F_k = P_k K_{F,k-1}^{1/2} (K_{F,k-1}^{-1/2} \bar{H}_k - \bar{F}_{k-1} F_k^{-1} \bar{F}_{k-1}^H \bar{H}_{k-1} K_{F,k-1}^{-1/2} + I_{N_{R,k-1}})^{1/2} Q_{k-1}^H$$  

(20)

where $K_{F,k}$ is defined as

$$K_{F,k} \triangleq \text{Tr}(F_k F_k^H \Psi) \Sigma_k + \sigma_{n_k}^2 I_{N_{R,k}},$$  

(21)

and the matrix $Q_k$ is an additional unknown unitary matrix. Based on the definition of $F_k$ in (19) and (20), it is easy to show that $F_k^H F_k = P_k R_{\bar{x}_{K-1}} P_k^H$ and thus the power constraints becomes

$$\text{Tr}(P_K R_{\bar{x}_{K-1}} P_K^H) \leq P_k.$$  

(22)

Therefore, in terms of the new variables $F_k$, the power constraints become independent of each other, which facilitates further manipulations.

3 Putting the definition of $F_k$ into (11) and comparing (11) with (23), the matrix $K_{F,k}$ can be interpreted as the equivalent noise covariance matrix at the $k^{th}$ hop.
Meanwhile, using (19) and (20), the MSE matrix is further reformulated as (23) on the top of the page. Based on (22) and (23), the optimization problem (18) is simplified as

\[
\min_{F_k, Q_k} f([\text{MSE}_1, \ldots, \text{MSE}_N]^T)
\]

\[
\text{s.t. } \quad \text{MSE}_n = \sigma^2_b [C(I_N - Q^H_0Q_0)C^H]^n, n, n
\]

\[
\Theta = M_1^HQ_1^H \cdots M_K^HQ_K^HM_K \cdots Q_1 M_1
\]

\[
\text{Tr}(F_k F_k^H) \leq P_k, \quad Q_k^H Q_k = I_{N_{R,k}}.
\]

(24)

Notice that the largest singular value of \(M_k\) is smaller than one. Therefore, the largest eigenvalue of \(\Theta\) is smaller than one (see Appendix A) and then \(I_N - Q_0^H\Theta Q_0\) is a positive definite matrix. With the Cholesky factorization

\[
(I_N - Q_0^H\Theta Q_0)\sigma^2_b = LL^H
\]

(25)

where \(L\) is a lower triangular matrix, and the definition of MSE\(_n\) in the second line of (24), we have

\[
\text{MSE}_n = \sigma^2_b [C(I_N - Q^H_0\Theta Q_0)C^H]^n, n, n
\]

\[
= \left(\left[C^H\right]_n, n\right)HLL^H\left[C^H\right]_n, n
\]

\[
= \sum_{i=1}^{n-1} |L_{i,i}|^2 + |L_{i,n}^2| + |L_{n,n}^2|
\]

\[
\geq |L_{n,n}^2|
\]

(26)

where \(D\) is a diagonal matrix defined as

\[
D = \text{diag}\{[L_{1,1}, \ldots, L_{N,N}]^T\}
\]

(27)

In order to make the equality in the final line of (26) to hold, we need \(\sum_{i=1}^{n-1} |L_{i,i}|^2 |C_{\text{LD}}^{-1}|^2_{i, i, n} = 0\), whose solution is

\[
C_{\text{opt}} = DL^{-1}.
\]

(28)

As a result MSE\(_n\) = \(L_{n,n}^2\), and the optimization problem for robust transceiver design is formulated as

\[
\min_{F_k, Q_k} f([L_{1,1}^2, \ldots, L_{N,N}^2]^T)
\]

\[
\text{s.t. } \quad \sigma^2_b [I_N - Q_0^H\Theta Q_0] = LL^H
\]

\[
\Theta = M_1^HQ_1^H \cdots M_K^HQ_K^HM_K \cdots Q_1 M_1
\]

\[
\text{Tr}(F_k F_k^H) \leq P_k, \quad Q_k^H Q_k = I_{N_{R,k}}.
\]

(29)

IV. OPTIMIZATION PROBLEM REFORMULATION FOR \(F_k\)

A. Optimal Solution of \(Q_0\)

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

\[
|\sigma^2_b (I_N - Q_0^H\Theta Q_0)| = \prod_{n=1}^{N} [L_{n,n}^2] \prod_{n=1}^{N} (1 - \lambda_n(\Theta))
\]

(30)

where \(\lambda_n(\Theta)\) is the \(n\)th largest eigenvalue of \(\Theta\). Based on (30), the following multiplicative majorization relationship can be established (23)

\[
\sigma^2_b \left[\prod_{n=1}^{N} (1 - \lambda_n(\Theta))\right] \sim \left[|L_{1,1}^2, \ldots, |L_{N,N}^2| \right]^T
\]

(31)

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

\[
f\left([L_{1,1}^2, \ldots, |L_{N,N}^2|]^T\right) \geq f\left(\sigma^2_b \left[\prod_{n=1}^{N} (1 - \lambda_n(\Theta))\right] \otimes 1_N\right)
\]

\[
\triangleq g(\lambda(\Theta))
\]

(32)

where \(\lambda(\Theta) = [\lambda_1(\Theta), \ldots, \lambda_N(\Theta)]^T\). The equality in (32) holds when \(\sim\) in (31) is replaced by equality, which means that \(|L_{n,n}^2\) are identical for all \(n\). Notice that from (25), we can write \(LL^H = \sigma^2_b Q_0^H(I - \Theta)Q_0\). Since \(I - \Theta\) is positive definite, there always exists an unitary matrix \(Q_0\) which makes the Cholesky factorization matrix of \(Q_0^H(I - \Theta)Q_0\) have identical diagonal elements (15). An explicit algorithm for constructing such \(Q_0\) is given in Appendix A.

M-Schur-concave:

From definition of \(L\) in (25) and based Weyl’ theorem (23), we have

\[
[L_{1,1}^2, \ldots, |L_{N,N}^2|]^T \sim \sigma^2_b [I_N - \lambda(\Theta)]
\]

(33)

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

\[
f\left([L_{1,1}^2, \ldots, |L_{N,N}^2|]^T\right) \geq f\left(\sigma^2_b [I_N - \lambda(\Theta)]\right)
\]

\[
\triangleq g(\lambda(\Theta))
\]

(34)
The equality in (34) holds when \( \prec_x \) in (33) is replaced by equality, which means that \( [L]_{i,n}^2 \) equals to \( \sigma^2_1[1 - \lambda_n(\Theta)] \). On the other hand, taking eigenvalues on both sides of (33), we can obtain \( \sigma^2_1[1 - \lambda_n(\Theta)] = [\lambda_N(\bar{L}L), \cdots, \lambda_1(\bar{L}L)]^T \). Therefore, \( [\bar{L}]_{1,1}^2, \cdots, [\bar{L}]_{N,N}^2 \) = \([\lambda_N(\bar{L}L), \cdots, \lambda_1(\bar{L}L)]^T \), which implies \( L \) is a diagonal matrix. With \( L \) being a diagonal matrix, \( Q_k^H \Theta Q_0 \) is also a diagonal matrix. This can be satisfied if we take \( Q_0 = U \Theta \), where the unitary matrix \( U \Theta \) is defined based on the eigendecomposition \( \Theta = U \Theta \Lambda \Theta U^H \) with the elements of \( \Lambda \Theta \) arranged in decreasing order.

Notice that since \( L \) is a diagonal matrix, \( C_{\text{opt}} \) in (28) is also a diagonal matrix. Based on the definition of \( C \) in (5) and with the fact that \( C \) is a lower triangular matrix with unit diagonal elements, it can be seen that the feedback matrix \( B \) 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 [13, 15, 17].

**Remark 3:** The equal bit rate assumption at the beginning of Section [11] 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 in (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

\[
\min_{F_k, Q_k} \quad g[\lambda(\Theta)] \\
\text{s.t.} \quad \Theta = M^H_1 Q^H_1 \cdots M^H_k Q^H_k \Lambda M_k \cdots Q_1 M_1 \\
\text{Tr}(F_k F_k^H) \leq P_k, \quad Q^H_k Q_k = I_{N \times K}.
\]

where \( g[\lambda(\Theta)] \) equals to

\[
g[\lambda(\Theta)] = \begin{cases} 
\frac{\sigma^2_1}{(1 - \lambda_n(\Theta))} \Theta \otimes 1_N, & \text{if } f(\bullet) \text{ is M-Schur-convex}, \\
\sigma^2_1[1 - \lambda_n(\Theta)] \Lambda \Theta, & \text{if } f(\bullet) \text{ is M-Schur-concave}.
\end{cases}
\]

It is difficult to directly solve the optimization problem (35), because \( \Theta \) is a product consists of matrices \( M_k \)'s which in turn are complicated functions of the variables \( F_k \)'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 \( F_k \). 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 \( v, u \in D \) with \( D = \{ z : z_1 \geq \cdots \geq z_N \geq 0 \} \), \( v \prec_{x,w} u \) is defined as

\[
\prod_{i=1}^{k} v[i] \leq \prod_{i=1}^{k} u[i], \quad k = 1, \cdots, N.
\]

Notice that there is a subtle difference between Definition 2 in (13) and Definition 3. In Definition 3, when \( k = N \), \( \prod_{i=1}^{N} v[i] \leq \prod_{i=1}^{N} u[i] \) rather than \( \prod_{i=1}^{N} v[i] = \prod_{i=1}^{N} u[i] \) in Definition 2.

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

\[
v \prec_{x,w} u \Rightarrow \phi(v) \geq \phi(u).
\]

**Proof:** See Appendix C.

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

\[
v \prec_{x,w} u \Rightarrow \phi(1_N - v) \geq \phi(1_N - u).
\]

**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, the two useful properties of the objective function \( g(\bullet) \) are first derived based on the multiplicative majorization theory.

**Property 1:** The vector \( \lambda(\Theta) \) has the following relationship

\[
\lambda(\Theta) \prec_{x,w} [\gamma_1(F_k^M)^{K_1}, \gamma_2(F_k^M)^{K_2}, \cdots, \gamma_N(F_k^M)^{K_N}]^T
\]

with \( \gamma_n(F_k^M)^{K_n} = \prod_{k=1}^{K_n} \lambda_n(F_k^M)^{K_n} = \prod_{k=1}^{K_n} \left( 1 + \lambda_n(F_k^M)^{K_n} - \frac{1}{\lambda_n(F_k^M)^{K_n}} \right) \),

\[
\prod_{i=1}^{k} \left( 1 - \lambda_n(\Theta[i]) \right) \Theta[i] \otimes I_N
\]

where the equality holds when

\[
Q_k = V_{M_k \times K} U_{M_k}^H, \quad k = 1, \cdots, K - 1
\]

where \( V_{M_k} \) and \( V_{M_k} \) are defined based on the singular value decomposition \( M_k = U_{M_k} \Lambda_{M_k} V_{M_k}^H \) with the diagonal elements of \( \Lambda_{M_k} \) arranged in decreasing order. Notice that (21) does not cover the design of \( Q_k \), but it can be any unitary matrix because it always appears in the form \( Q_k^H Q_k \) and equals to an identity matrix in the objective function.

**Proof:** See Appendix E.

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

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

When \( f(\bullet) \) is M-Schur-convex, \( g[\lambda(\Theta)] = f(\sigma^2_1[\prod_{n=1}^{N}(1 - \lambda_n(\Theta))]^2 \Theta \otimes I_N) \). Using Lemma 1, \( \prod_{n=1}^{N}(1 - \lambda_n(\Theta)) \) can be proved to be a M-Schur-concave function of \( \lambda(\Theta) \). Furthermore, it can be easily seen that if \( \prod_{n=1}^{N}(1 - \lambda_n(\Theta)) \) is...
that the relays transmit at the maximum power. Together with the fact that \( f(\bullet) \) is an increasing function, it is concluded that

\[
\begin{align*}
&\left(1 - \lambda_n(\Theta)\right) \geq \lambda_n(\Theta) - \lambda(\Theta) \implies \\
&\gamma(\min_{P_k} g[\{F_k\}_{k=1}^K]) \\
\end{align*}
\]

On the other hand, when \( f(\bullet) \) is increasing and M-Schur-concave, \( g[\lambda(\Theta)] = \gamma(\{(F_k)_{k=1}^K]) \) with equality achieved when \( \Theta = \Theta_k \) is satisfied. When the lower bound is achieved, we have the following additional observations:

(a) The constraints \( Q_k^H Q_k = I_{N_{R,k}} \) are automatically satisfied.
(b) The objective function \( g[\gamma(\{(F_k)_{k=1}^K]) \) is independent of \( Q_k \).
(c) When \( F_k \)'s are known, \( Q_k \)'s can be directly computed using (41).

Applying these three observations into (45), we have the reformulated optimization problem

\[
\begin{align*}
&\min_{P_k} \quad g[\gamma(\{(F_k)_{k=1}^K]) \\
&\text{s.t.} \quad \gamma(\{(F_k)_{k=1}^K]) = \prod_{k=1}^K \frac{\lambda_n(F_k^H H_k^H F_k^{-1} H_k F_k)}{1 + \lambda_n(F_k^H H_k^H F_k^{-1} H_k F_k)} \\
&\quad \text{Tr}(F_k^H F_k) \leq P_k.
\end{align*}
\]

V. Solution of \( F_k \)

In the following, we first derive the optimal structure of \( F_k \) and then present an algorithm to solve for the remaining unknown variables.

A. Optimal Structure of \( F_k \)

Notice that \( g(\bullet) \) is a decreasing function, and \( \gamma(\{(F_k)_{k=1}^K]) \) is an increasing function of \( \lambda_n(F_k^H H_k^H F_k^{-1} H_k F_k) \). Therefore, \( g[\gamma(\{(F_k)_{k=1}^K]) \) is a decreasing matrix-monotone function of \( F_k^H H_k^H F_k^{-1} H_k F_k \). Following the derivation in [11], it can be proved that at the optimal solution, the power constraints hold at the equality, i.e., \( \text{Tr}(F_k^H F_k) = P_k \), meaning that the relays transmit at the maximum power.

Defining a variable \( \eta_{f_k} \) as

\[
\eta_{f_k} = \alpha_k \text{Tr}(F_k^H \Psi_k) + \sigma_{n_k}^2 \quad \text{with} \quad \alpha_k = \text{Tr}(\Sigma_k)/N_{R,k},
\]

\( \text{Tr}(F_k^H F_k) = P_k \) is exactly equivalent to \( \text{Tr}[F_k^H \Psi_k + \sigma_{n_k}^2 I_{N_{T,k}}]/\eta_{f_k} = P_k \) as proved in [10], [11], [26]. Thus the robust transceiver design problem (45) is equivalent to

\[
\begin{align*}
\min_{P_k} \quad & g[\gamma(\{(F_k)_{k=1}^K]) \\
\text{s.t.} \quad & \gamma(\{(F_k)_{k=1}^K]) = \prod_{k=1}^K \frac{\lambda_n(F_k^H H_k^H F_k^{-1} H_k F_k)}{1 + \lambda_n(F_k^H H_k^H F_k^{-1} H_k F_k)} \\
&\quad \text{Tr}(F_k^H F_k) \leq P_k.
\end{align*}
\]

It is proved in Appendix [7] that when \( \Psi_k \propto I_{N_{T,k}} \) or \( \Sigma_k \propto I_{N_{R,k}} \), the optimal solutions of the optimization problem (47) have the following structure

\[
F_{k,\text{opt}} = \sqrt{\xi_k(A_{f_k})}(\alpha_k P_k \Psi_k + \sigma_{n_k}^2 I_{N_{T,k}})^{-1/2} \times V_{\mathcal{H}_k, N_{T,k}, N_{R,k}, \Psi_k} U_{\mathcal{H}_k, N_{T,k}, N_{R,k}, \Psi_k}^H
\]

with \( \xi_k(A_{f_k}) = \sigma_{n_k}^2/[1 - \alpha_k \text{Tr}(V_{\mathcal{H}_k, N_{T,k}, N_{R,k}, \Psi_k}^H \Psi_k + \sigma_{n_k}^2 I_{N_{T,k}})^{-1/2} \Psi_k(\alpha_k P_k \Psi_k + \sigma_{n_k}^2 I_{N_{T,k}})^{-1/2} \times V_{\mathcal{H}_k, N_{T,k}, N_{R,k}, \Psi_k}] \). (48)

where \( \Lambda_{f_k} \) is a \( N \times N \) unknown diagonal matrix, and \( V_{\mathcal{H}_k, N_{T,k}, N_{R,k}, \Psi_k} \) are the unitary matrix \( U_{\mathcal{H}_k, N_{T,k}, N_{R,k}, \Psi_k} \) is an arbitrary \( N_{R,k-1} \times N_{R,k-1} \) unitary matrix, and the unitary matrix \( V_{\mathcal{H}_k, N_{T,k}, N_{R,k}, \Psi_k} \) is defined based on the singular value decomposition

\[
(K_{f_k}/\eta_{f_k})^{-1/2} H_k(\alpha_k P_k \Psi_k + \sigma_{n_k}^2 I_{N_{T,k}})^{-1/2} \times V_{\mathcal{H}_k, N_{T,k}, N_{R,k}, \Psi_k}^H
\]

where the diagonal elements of \( \Sigma_{f_k} \) are arranged in decreasing order.

Remark 4: In general, the expressions of \( \Psi_k \) and \( \Sigma_k \) depend on specific channel estimation algorithms. Denote the transmit and receive antennas correlation matrices and the channel estimation error variance in the \( k \)-th hop as \( R_{T,k}, R_{R,k} \) and \( \sigma_{e,k}^2 \), respectively. When the channels are estimated based on the algorithm proposed in [21], [22], it can be shown that \( \Psi_k \propto R_{T,k} \) and \( \Sigma_k \propto \sigma_{e,k}^2 I_{N_{T,k}} \). If the transmit antennas or the receive antennas are spaced widely, we have \( R_{T,k} \propto I_{N_{T,k}} \) or \( R_{R,k} \propto I_{N_{R,k}} \). These imply \( \Psi_k \propto I_{N_{T,k}} \) or \( \Sigma_k \propto I_{N_{R,k}} \). Moreover, if the length of training is large, the value of \( \sigma_{e,k}^2 \) will be small and \( \sigma_{e,k}^2 R_{R,k} \approx I_{N_{R,k}} \). As a result, \( \Sigma_k \) will also approximate an identity matrix even when \( R_{R,k} \propto I_{N_{R,k}} \). On the other hand, if the channel statistics are unknown, and using least-square channel estimator, it can be derived that \( \Sigma_k \propto I_{N_{R,k}} \) always holds regardless of the antenna correlation or training length [8].

B. Computation of \( \Lambda_{f_k} \)

It is obvious that in (43), the only unknown variable is \( \Lambda_{f_k} \). In the following, we will discuss how to solve \( \Lambda_{f_k} \) in more detail. Denoting the following diagonal elements as

\[
[\Lambda_{h_k}]_{k,n} = h_{k,n}, \quad [\Lambda_{f_k}]_{k,n} = f_{k,n},
\]

(50)
substituting \( \xi_k(A_{\mathcal{F}_k}) = \eta_{f_k} \) (shown by (29) in Appendix F), after a straightforward derivation, the optimization for robust transceiver design is simplified as

\[
\begin{align*}
\min_{f_{k,n}} & \quad g(\gamma(\{F_k\}_k=1)) \\
\text{s.t.} & \quad \gamma_n(\{F_k\}_k=1) = \prod_{k=1}^{K} \frac{f_{2,n}h_{k,n}^2}{f_{k,n}h_{k,n}^2 + 1} \\
& \quad \sum_{n=1}^{N} f_{k,n}^2 = P_k.
\end{align*}
\tag{51}
\]

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

**M-Schur-convex functions:**
Notice that when \( f([\text{MSE}_1, \cdots, \text{MSE}_N]^T) \) is an M-Schur-convex function, regardless of the specific expression of \( f(\bullet) \), the optimization problem (51) is equivalent to minimize

\[
\prod_{n=1}^{N} (1 - \gamma_n(\{F_k\}_k=1)) \tag{52}
\]

Therefore, the transceiver design problem (51) equals to

\[
\begin{align*}
\min_{f_{k,i}} & \quad \sum_{n=1}^{N} \log \left( 1 - \frac{\prod_{k=1}^{K} f_{k,n}^2 h_{k,n}^2}{\prod_{k=1}^{K} (f_{k,n}^2 h_{k,n}^2 + 1)} \right) \\
\text{s.t.} & \quad \sum_{n=1}^{N} f_{k,n}^2 = P_k.
\end{align*}
\tag{52}
\]

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

\[
f_{k,n}^2 = \frac{1}{h_{k,n}^2} \left( -a_{k,n} + \sqrt{a_{k,n}^2 + 4(1 - a_{k,n} a_{k,n} h_{k,n}^2/\mu_k + 1)} \right) \tag{53}
\]

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

where \( \mu_k \) 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 \( f([\text{MSE}_1, \cdots, \text{MSE}_N]^T) \) is a M-Schur-concave function, there is no unified solution. In this case, \( \Lambda_{\mathcal{F}_k} \) should be solved case by case. In the following, we use the example \( f([\text{MSE}_1, \cdots, \text{MSE}_N]^T) = \prod_{n=1}^{N} \text{MSE}_n \) for \( w_1 \geq w_2 \cdots \geq w_N \geq 0 \) to illustrate how to compute \( \Lambda_{\mathcal{F}_k} \). For this objective function, using (36) it follows that

\[
g(\gamma(\{F_k\}_k=1)) = \sqrt{2} \sum_{n=1}^{N} w_n \left( \prod_{n=1}^{N} (1 - \gamma_n(\{F_k\}_k=1)) \right)^{2/3} 
\]

and the optimization (51) is equivalent to

\[
\begin{align*}
\min_{f_{k,i}} & \quad \sum_{n=1}^{N} w_n \log \left( 1 - \frac{\prod_{k=1}^{K} f_{2,n} h_{k,n}^2}{\prod_{k=1}^{K} (f_{k,n} h_{k,n}^2 + 1)} \right) \\
\text{s.t.} & \quad \sum_{n=1}^{N} f_{2,n} = P_k.
\end{align*}
\tag{54}
\]

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 \( \Lambda_{\mathcal{F}_k} \). In general, only local optimality of the solution can be guaranteed (29), which is a common problem for AF MIMO relaying design (6), (9), (29).

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 \( F_{k,\text{opt}} \) in (48) and the optimal \( Q_k \) in (41), using the definition of \( F_k \) given by (19) and (20), we can derive that the optimal forwarding matrix at the \( k \)th node has the following structure

\[
P_k = (\alpha_k P_k \Psi_k + \sigma^2_k I_{N_{T,k}})^{-1/2} \times V_{H_k,N} \Lambda_{P_k} U_{H_{k-1,N} K_{F_{k-1}}^{-1/2}}.
\tag{55}
\]

where \( \Lambda_{P_k} \) is a diagonal matrix whose elements are functions of the diagonal elements of \( \Lambda_{\mathcal{F}_m} \) for all \( m \). It can be seen that except \( \Lambda_{P_k} \), all other matrices in (58) are only

| Table I | Summary of Robust Transceiver Design |
|---------|-----------------------------------|
| 1.     | Derive the data estimation MSE matrix (19). |
| 2.     | Formulate the optimization problem (24) with \( C \) as variables. |
| 3.     | Derive the optimal equalizer \( G \) as a function of \( P_k \) and \( C \) given by (15). Substitute the optimal \( G \) into the optimization problem (12) to reduce the number of variables and have a reformulated optimization problem (15). |
| 4.     | Simplify the constraints of the optimization problem (18) by replacing \( P_k \) with \( Q_k \) and \( F_k \) and obtain an equivalent optimization problem (21). |
| 5.     | Derive the optimal \( C \) as a function of \( Q_k \) and \( F_k \) given by (28). Substitute the optimal \( C \) into the optimization problem (25) to reduce the number of \( K \). |
| 6.     | Derive the optimal \( Q_k \) as a function of \( P_k \) based on majorization theory and substitute the optimal \( Q_k \) in (21) into the optimization problem (29) to reduce the number of variables. The optimization problem is then simplified to be (48). |
| 7.     | Derive the optimal structure of \( F_k \) given by (48). |
| 8.     | Solve for the unknown diagonal matrices \( \Lambda_{\mathcal{F}_k} \) in the optimal structure using (51). |

1. Derive the data estimation MSE matrix (19)
2. Formulate the optimization problem (24) with \( C \) and \( P_k \) as variables
3. Derive the optimal equalizer \( G \) as a function of \( P_k \) and \( C \) given by (15)
4. Simplify the constraints of the optimization problem (18) by replacing \( P_k \) with \( Q_k \) and \( F_k \)
5. Derive the optimal \( C \) as a function of \( Q_k \) and \( F_k \) given by (28)
6. Derive the optimal \( Q_k \) as a function of \( P_k \) based on majorization theory and substitute the optimal \( Q_k \) in (21) into the optimization problem (29) to reduce the number of variables. The optimization problem is then simplified to be (48)
7. Derive the optimal structure of \( F_k \) given by (48)
8. Solve for the unknown diagonal matrices \( \Lambda_{\mathcal{F}_k} \) in the optimal structure using (51)
the functions of the channels immediately preceding and succeeding the \( k \)th 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 \( \mathbf{A}_\mathbf{x}_n \) 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, 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 \( \Psi_k = \sigma_e^2 \rho_t^{i-j} \) and \( \Sigma_k = \rho_r^{i-j} \) \( \Psi_k \) where \( \rho_t \) and \( \rho_r \) are the correlation coefficients, and \( \sigma_e^2 \) denotes the estimation error variance. The estimated channels \( \mathbf{H}_k \)’s are randomly generated based on the following complex Gaussian distributions \( \mathcal{CN}(0, \mathbf{I}) \), \( \mathcal{CN}(0, \mathbf{I}) \), \( \mathcal{CN}(0, \mathbf{I}) \), \( \mathcal{CN}(0, \mathbf{I}) \), \( \mathcal{CN}(0, \mathbf{I}) \). such that channel realizations \( \mathbf{h}_k = \mathbf{H}_k + \Delta \mathbf{H}_k \) have unit variance. We define the signal-to-noise ratio (SNR) for the \( k \)th link as \( P_k / \sigma_n^2 \). At the source node, four independent data streams are transmitted and in each data stream, \( N_{\text{Data}} = 10000 \) 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 \( \mathcal{CN}(0, \mathbf{I}) \), \( \mathcal{CN}(0, \mathbf{I}) \), \( \mathcal{CN}(0, \mathbf{I}) \), \( \mathcal{CN}(0, \mathbf{I}) \), \( \mathcal{CN}(0, \mathbf{I}) \).

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 \( \rho_t = 0, \rho_r = 0.4, \sigma_e^2 = 0.001 \), \( \sigma_e^2 = 0.001, \sigma_e^2 = 0.005 \), \( P_1 / \sigma_n^2 = 30 \)dB, and \( P_2 / \sigma_n^2 \) 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 \( N(N-1)/2 \) 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 \( \rho_t = 0.5, \rho_r = 0, P_1 / \sigma_n^2 = 30 \)dB, and \( P_2 / \sigma_n^2 \) being varied from 10 to 35dB. The algorithm based on estimated CSI only is obtained by simply setting \( \Psi_k = 0 \) 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 \( \rho_r \), when \( \rho_t = 0, \sigma_e^2 = 0.002, P_1 / \sigma_n^2 = 30 \)dB, and \( P_2 / \sigma_n^2 \) 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 \( \rho_r \) 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, \( \Sigma_k \) 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 as the transmission directions, Figs. 4 shows that correlation becomes to prove that when correlation becomes stronger, \( \rho_r = 0 \), \( \sigma_r^2 = 0.002 \), \( P_2/\sigma^2_{\rho_1} = 30 \text{dB} \).

**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 that channel estimation errors exist in CSI in all hops. At MIMO relaying systems was investigated. It was assumed the non-robust design.

The proposed robust design has much better performance than iterative water-filling solution was proposed to solve for the original optimization problem was greatly simplified and an structure of the transceivers was first derived. Then, the majorization theory and matrix-monotone functions, the optimal Schur-concave. Using elegant properties of multiplicative ma-

**APPENDIX A**

**PROOF OF LEMMA 1**

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

For a vector \( z \in D \), define \( \tilde{z} = [\tilde{z}_1, \cdots, \tilde{z}_N]^T \) and \( \tilde{z}_k = \prod_{i=1}^{k} z_i \).

For \( v, u \in D \), it is obvious that \( v \prec_{\times} u \) is equivalent to \( \{ \tilde{v}_k \leq \tilde{u}_k \}_{k=1}^{N-1}, \text{ and } \tilde{v}_N = \tilde{u}_N \).

On the other hand, based on (57), \( z_k \) equals to \( z_k = \tilde{z}_k/\tilde{z}_{k-1}, \text{ for } k \leq L_z \).

where \( L_z - 1 \) is the number of the nonzero elements of \( z \).

Therefore \( \phi(v) \leq \phi(u) \) can be written as

\[
\frac{\phi(v_1, v_2, \cdots, v_{L_z}, 0, \cdots)}{\phi(u_1, u_2, \cdots, u_{L_z}, 0, \cdots)} \\
\leq \frac{\psi(v_1, v_2, \cdots, v_{L_z}, 0, \cdots)}{\psi(u_1, u_2, \cdots, u_{L_z}, 0, \cdots)}.
\]

Based on (58) and (60), proving \( \phi(z) \) 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 \( \psi(\tilde{v}) \leq \psi(\tilde{u}) \). 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 \( \psi(\tilde{v}) \leq \psi(\tilde{u}) \).

As \( \tilde{v}_k \geq 0 \) and \( \tilde{u}_k \geq 0 \), \( \tilde{v}_k \leq \tilde{u}_k \) is equivalent to \( \tilde{v}_k - \tilde{u}_k = \tilde{u}_k/e \) 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 \( u_k = u_k/\tilde{u}_k \) for \( k \leq L_u - 1 \), proving \( \psi(\tilde{v}) \leq \psi(\tilde{u}) \) is equivalent to proving \( \phi(u_1, \cdots, u_k/e, u_k+1e, \cdots) \) is decreasing over \( e \geq 1 \) and \( u_k/e \geq u_{k+1}e \).
APPENDIX B
ALGORITHM FOR COMPUTING $Q_0$

Following the sufficient conditions given in [2], an explicit algorithm for constructing $Q_0$ 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 $A$ based on the following eigen-decomposition

$$ (I_N - M_1^H Q_1^H \cdot \cdot \cdot M_K^H Q_K^H Q_K M_K \cdot \cdot \cdot Q_1 M_1)^{1/2} \sigma_0 = U_M A_M U_M^H \equiv \Lambda $$

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

$$ [S]_{1,1} = \sqrt{[A^H A]_{1,1}/N - [\Lambda M]_{1,1}/N} $$

$$ [S]_{N,1} = \sqrt{[\Lambda_M]_{1,1}/N - [A^H A]_{1,1}/N} $$

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

$$ [S]_{:,1} = \begin{bmatrix} -[S]_{N,1} & 0 \\ [S]_{1,1} & 0 \end{bmatrix} $$

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

$$ (A[S]_{:,1:k}^H[I - A[S]_{:,1:k}(A[S]_{:,1:k}^H A[S]_{:,1:k})]^{-1} x \times [S]_{:,1:k}^H A^H) (A[S]_{:,1:k}^H V(k) A(k) V(k)^H) $$

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

$$ [S]_{:,k+1} = [S]_{:,1:k} V(k) y(k) $$

and

$$ y(k) = \begin{bmatrix} [A^H A]_{1,1} - [\Lambda(k)]_{N-k,N-k} \bigg/ [\Lambda(k)]_{1,1} - [\Lambda(k)]_{N-k,N-k} \bigg/ 0_{1,N-k-1}, \\
[A^H A]_{1,1} - [\Lambda(k)]_{N-k,N-k} \bigg/ [\Lambda(k)]_{1,1} - [\Lambda(k)]_{N-k,N-k} \bigg/ \end{bmatrix}^T $$

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

$$ [S]_{:,1:k+1} = [U_S]_{:,k+2:N} $$

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

$$ y(N-1) = \begin{bmatrix} [A(N-2)]_{1,1} - [A^H A]_{1,1} \bigg/ [A(N-2)]_{1,1} - [A^H A]_{2,2} \\
[A(N-2)]_{1,1} - [A^H A]_{1,1} \bigg/ [A(N-2)]_{2,2} \end{bmatrix}^T. $$

**Step 5:** Finally, $Q_0$ equals to $Q_0 = U_M S$.

APPENDIX C
PROOF OF Lemma 2

**Proof of “if” direction**

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

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

Furthermore, for $v, u \in D$ with $v_k \leq u_k$ and $v_i = u_i$ for all $i \neq k$, we have $v \prec_{x,w} u$. Then $v \prec_{x,u} u \Rightarrow \phi(v) \geq \phi(u)$ implies $\phi(\bullet)$ is a decreasing function. Therefore, when $v \prec_{x,u} u \Rightarrow \phi(v) \geq \phi(u)$, 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 $v \prec_{x,w} u \Rightarrow \phi(v) \geq \phi(u)$. For any two vectors $v, u \in D$ with $v \prec_{x,w} u$ we can construct a vector $\tau \in D$ with $\tau_i = u_i$ for $i < N$ and $\tau_N$ is chosen to makes $\prod_{i=1}^{N} \tau_i = \prod_{i=1}^{N} u_i$. It is obvious that $\tau_N \leq u_N$. Then if $v \prec_{x,u} u$, we have $v \prec_{x} \tau \prec_{x,u} u$.

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

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 $z \in C = \{ z : 1 > z_1 \geq \cdots \geq z_N \geq 0 \}$. Using Lemma 2, for $v, u \in C$ with $v \prec_{\Pi} u$, we have

$$ \prod_{i=1}^{k} (1 - v_i) \geq \prod_{i=1}^{k} (1 - u_i) > 0, \quad k = 1, \ldots, N. \quad (69) $$

We construct a vector $\hat{\tau} = [\hat{\tau}(1), \ldots, \hat{\tau}(N)]^T$ with $\hat{\tau}(i) = \hat{u}(i)$ for $i < N$ and $\hat{\tau}(N)$ is chosen to makes $\prod_{i=1}^{N} \hat{\tau}(i) = \prod_{i=1}^{N} \hat{u}(i)$ 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 $\phi(\hat{\tau}) \geq \phi(\hat{u})$ where $\hat{u} = [\hat{u}(1), \ldots, \hat{u}(N)]^T$.

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) $\prod_{i=1}^{k} \hat{v}(i) \geq \prod_{i=1}^{k} \hat{u}(i)$ for $1 \leq k < N$. Based on the definition of $\hat{\tau}_N$, it can also be concluded that (b) $\prod_{i=1}^{k} \hat{v}(i) = \prod_{i=1}^{k} \hat{u}(i)$ for $1 \leq k < N$. Results (a) and (b) implies $\hat{v} \prec_{\tau} \hat{u}$ where $\hat{v} = [\hat{v}(1), \ldots, \hat{v}(N)]^T$.

As $\phi(\bullet)$ is M-Schur-concave, using Lemma 1, we have $\phi(\hat{v}) \geq \phi(\hat{u})$. Together with the conclusion in the last paragraph, we can obtain $\phi(\hat{v}) \geq \phi(\hat{u})$. Finally, with $\hat{v} = 1_N - v$ and $\hat{u} = 1_N - u$, the proof is completed.

APPENDIX E
PROOF OF Property 1

First notice two facts in matrix theory: (a) for two matrices $A$ and $B$ with compatible dimension $\lambda_i(AB) = \lambda_i(BA)$.
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 \( \lambda_i(\tilde{M}_k^H \tilde{M}_k) = \lambda_i(M_k^HM_k) \) we can obtain the following inequality

\[
\prod_{i=1}^{n} \lambda_i(\Theta) \leq \prod_{i=1}^{n} \left[ \lambda_i(\tilde{M}_k^HM_k) \lambda_i(M_{i-1}^H M_{i-1}) \cdots \lambda_i(M_1^H M_1) \right] = \Delta \gamma_i(\{F_k\}_{k=1}^{K}),
\]

(71)

where the equality holds when \( \Theta_k \)’s satisfy

\[
\Theta_k = V_{M_{k+1}} U_{N_k}^H, \quad k = 1, \ldots, K - 1,
\]

(72)

where \( U_{N_k} \) and \( V_{M_k} \) are defined based on the following singular value decomposition \( M_k = U_{M_k} \Lambda_{M_k} V_{M_k}^H \) with the diagonal elements of \( \Lambda_{M_k} \) arranged in decreasing order. Furthermore, based on the definition of \( M_k \) in (23), \( \gamma_i(\{F_k\}_{k=1}^{K}) \) in (71) equals to

\[
\gamma_i(\{F_k\}_{k=1}^{K}) = \left( \frac{1}{\lambda_i(M_{i-1}^H M_{i-1})} \right) \frac{1}{1 + \lambda_i(M_k^H J_k^H M_k)^{-1} M_k^H J_k^H M_k}
\]

(73)

\section{Appendix F: Optimal Structure of \( \tilde{F}_k \)}

Defining new variables

\[
\tilde{F}_k = 1/\sqrt{\eta f_k}(\alpha_k P_k \Psi_k + \sigma^2_{I_{N,T,k}})^{-1/2} F_k\]

and

\[
\tilde{H}_k = (K_{F_k}/\eta f_k)^{-1/2} H_k(\alpha_k P_k \Psi_k + \sigma^2_{I_{N,T,k}})^{-1/2},
\]

(74)

the optimization problem (47) is reformulated as

\[
\min_{\tilde{F}_k} g(\{\tilde{F}_k\}_{k=1}^{K})
\]

s.t. \( \gamma_i(\{\tilde{F}_k\}_{k=1}^{K}) = \left( \frac{1}{\lambda_i(M_k^H M_k)} \right) \frac{1}{1 + \lambda_i(M_k^H J_k^H M_k)^{-1} M_k^H J_k^H M_k}
\]

(75)

When \( \Psi_k \propto I_{N,T,k} \) or \( \Sigma_k \propto I_{N,M_k} \), for the optimal solution \( K_{F_k}/\eta f_k \) is constant \( 10, 11, 26 \) and thus \( \tilde{H}_k \) is constant. Let \( \tilde{F}_{k,\text{opt}} \) be the optimal solution of (75). With the following singular value decompositions,

\[
\tilde{H}_k \tilde{F}_{k,\text{opt}} = U_{\tilde{A}_k} \Lambda_{\tilde{A}_k} V_{\tilde{A}_k}^H, \quad \tilde{H}_k = U_{\tilde{H}_k} \Lambda_{\tilde{H}_k} V_{\tilde{H}_k}^H,
\]

(76)

where the diagonal elements of \( \Lambda_{\tilde{A}_k} \) and \( \Lambda_{\tilde{H}_k} \) are arranged in decreasing order, we can construct a matrix \( \tilde{F}_k \) equals to

\[
\tilde{F}_k = V_{\tilde{H}_k} \Lambda_{\tilde{H}_k} V_{\tilde{H}_k}^H,
\]

(77)

where \( \Lambda_{\tilde{H}_k} \) is a rectangular diagonal matrix with the same rank as \( \Lambda_{\tilde{A}_k} \) and \( 1/b \tilde{h}_{\tilde{A}_k} \Lambda_{\tilde{A}_k} = \Lambda_{\tilde{A}_k} \). The scalar \( b_k \) is chosen to make that \( \text{Tr}(\tilde{F}_k^H \tilde{F}_k) = P_k \) holds.

Using Lemma 12 in [23], we can show that \( \tilde{F}_k H_k^H H_k \tilde{F}_k \geq \tilde{F}_k H_k^H H_k \tilde{F}_k \). Together with the formulation of \( \gamma_i(\{F_k\}_{k=1}^{K}) \) in (73), it can be concluded that \( \gamma_i(\{F_k\}_{k=1}^{K}) \geq \gamma_i(\{\tilde{F}_k\}_{k=1}^{K}) \). Since \( g(\bullet) \) is an decreasing function, \( g(\{\gamma_i(\{F_k\}_{k=1}^{K})\}) \leq g(\{\gamma_i(\{\tilde{F}_k\}_{k=1}^{K})\}) \). Because \( F_{k,\text{opt}} \) is the optimal solution, it is impossible to have \( g(\{\gamma_i(\{F_k\}_{k=1}^{K})\}) \leq g(\{\gamma_i(\{\tilde{F}_k\}_{k=1}^{K})\}) \). Therefore, \( \tilde{F}_k \) must be the optimal solution. Furthermore, based on the relationship between \( \tilde{F}_k \) and \( F_k \), it follows that

\[
F_{k,\text{opt}} = \sqrt{n_{f_k} P_k \Psi_k + \sigma^2_{I_{N,T,k}}^{-1/2}} \tilde{F}_k \Lambda_{\tilde{A}_k} \tilde{H}_k^H.
\]

(78)

Notice that in general the unitary matrix \( V_{\tilde{A}_k} \) depends on the optimal solution \( \tilde{F}_{k,\text{opt}} \). However, from (73), it can be seen that the value of \( V_{\tilde{A}_k} \) does not affect the objective functions and therefore it can be an arbitrary unitary matrix. Meanwhile, as the minimum dimension of \( \tilde{F}_k H_k^H H_k \tilde{F}_k \) is \( N \), only \( N \times N \) principal submatrix of \( \Lambda_{\tilde{H}_k} \) can be nonzero. For notational convenience, we denote that \( [\Lambda_{\tilde{H}_k}]_{1:N,1:N} \equiv A_{\tilde{F}_k} \).

Substituting (78) into the definition of \( \eta f_k \) in (40), we obtain a simple linear function of \( \eta f_k \), and \( \eta f_k \) can be easily solved to be

\[
\eta f_k = \sigma^2_{I_{N,T,k}}^{-1} \frac{1 - \alpha_k \text{Tr}(\tilde{H}_k^H \tilde{H}_k N_{\tilde{A}_k} \tilde{H}_k^H \tilde{H}_k)}{\alpha_k P_k \Psi_k + \sigma^2_{I_{N,T,k}}^{-1/2}} \xi_k(A_{\tilde{F}_k}).
\]

(79)

\section{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 nonregenerative MIMO relay systems,” IEEE Commun. Lett., vol. 12, no. 7, pp. 517–519, July 2008.

[6] F.-S. Tseng, W.-R. Wu, and Y.-J. 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.

[7] 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.

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

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

[10] 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.

[11] 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. 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. Select. 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] Y. Rong, 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. Infor. 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. Kavčič, 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.