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

TRANSCIEVER 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 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 [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 (DFC). 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
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 $Z^H$ denotes the Hermitian of the matrix $Z$, and $\text{Tr}(Z)$ is the trace of the matrix $Z$. The symbol $I_N$ denotes an $N \times N$ identity matrix. The notation $Z^{1/2}$ is the Hermitian square root of the positive semidefinite matrix $Z$, such that $Z^{1/2}Z^{1/2} = Z$ and $Z^{1/2}$ is also a Hermitian matrix. The symbol $E\{\bullet\}$ represents the statistical expectation. For two Hermitian matrices, $C \succeq D$ means that $C - D$ is a positive semi-definite matrix. The $(n, m)^{th}$ entry of a matrix $Z$ is denoted as $[Z]_{n, m}$ and $\lambda(Z)$ represents the vector consisting of the eigenvalues of $Z$.

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 $k^{th}$ 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 $a = [a_1, a_2, \cdots, a_N]^T$ 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 $\mathcal{A} = \{\pm 1, \pm 3, \cdots, \pm (\sqrt{M} - 1)\}$.

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

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

(1)

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

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

(2)

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

$$b = (B + I_N)^{-1}(a + d)$$

(3)

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

After the nonlinear operation, the vector $b$ is multiplied with a precoder matrix $P_1$ under a transmit power constraint $\text{Tr}(P_1R_bP_1^H) \leq P_1$ where $P_1$ is the maximum transmit 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 $M$ is large, $b$ can be considered as i.i.d. [19], i.e.,

$$R_b = 2(M - 1)/3N \triangleq \sigma_b^2 I_N.$$  

(4)
The received signal $x_1$ at the first relay is formulated as
\[ x_1 = H_1 p_1 b + n_1 \]  
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 $R_{n_1} = \sigma_n^2 I_{N_R,1}$.

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
\[ x_2 = H_2 p_2 H_1 p_1 b + H_2 p_2 n_1 + n_2 \]
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 $R_{n_2} = \sigma_n^2 I_{N_R,2}$. Similarly, at the $k^{th}$ relay the received signal is
\[ x_k = H_k p_k x_{k-1} + n_k \]
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 $R_{n_k} = \sigma_n^2 I_{N_R,k}$. Finally, for a $K$-hop AF MIMO relaying system, the received signal at the destination is
\[ y = \prod_{k=1}^{K} (H_k p_k) b + \sum_{k=1}^{K-1} \left\{ \prod_{l=k+1}^{K} (H_l p_l) \right\} n_k + n_K, \]
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 $N$.[4]

In practice, the channels $H_k$ are estimated and channel estimation errors are inevitable. Therefore, the channel $H_k$ can be expressed as
\[ H_k = \bar{H}_k + \Delta H_k, \]
where $\bar{H}_k$ is the estimated channels, and $\Delta H_k$ 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 $\Delta H_k$ can be decomposed by the widely used Kronecker model [7], [8], [20] as
\[ \Delta H_k = \Sigma_k / \sqrt{2} H_{W,k} \Psi_k / \sqrt{2}, \]
where the elements of the $N_{R,k} \times N_{T,k}$ matrix $H_{W,k}$ 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 [7], [8], [11], [21].

### 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 $2\sqrt{M}$, the effect of $d$ will be perfectly removed by modulo operator at the destination. As a result, estimating $s$ is equivalent to estimating $a$ [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 $E\{(G y - s)(G y - s)^H\}$ [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
\[
\Phi(G, \{P_k\}_{k=1}^{K}, C) = E\{(G y - Cb)(G y - Cb)^H\}
\]
\[
= G (\bar{H}_K p_K r_{x_{K-1}} p_K^H \bar{H}_K^H + tr(p_K r_{x_{K-1}} p_K^H \Psi_k) \Sigma_k + r_{n_K}) G^H - \sigma_n^2 G \sum_{k=1}^{K} (\bar{H}_k p_k C)^H
\]
\[ - \sigma_n^2 G \left[ N \prod_{k=1}^{K} (\bar{H}_k p_k C)^H \right]^H + \sigma_n^2 C C^H \]
where matrices $r_{x_k}$ is defined as
\[
r_{x_k} \triangleq E\{x_k x_k^H\} = \bar{H}_k p_k R_{x_{k-1}} p_K^H \bar{H}_k^H + tr(p_K r_{x_{k-1}} p_K^H \Psi_k) \Sigma_k + r_{n_K}. \]

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

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 $f(MSE_1, \cdots, MSE_N) = \sum_{n=1}^{N} MSE_n$, where $MSE_n = [\Phi(G, \{P_k\}_{k=1}^{K}, C)]_{n,n}$. For product MSE minimization, the objective function is $f(MSE_1, \cdots, MSE_N) = \prod_{n=1}^{N} MSE_n$. Furthermore, worst-case MSE minimization corresponds to minimizing the objective function given as $f(MSE_1, \cdots, MSE_N) = \max_{n=1,\cdots,N} MSE_n$ [9], [13], [15], [23]. On the other hand, weighted geometric mean MSE minimization corresponds to minimizing the following objective function $f(MSE_1, \cdots, MSE_N) = \prod_{n=1}^{N} MSE_n^{w_n}$ with $w_1 \geq w_2 \cdots \geq w_N \geq 0$. Therefore, a unified transceiver design optimization problem can be formulated as
\[
\begin{align*}
\min_{G,p_k,c} & \quad f(MSE_1, \cdots, MSE_N) \\
\text{s.t.} & \quad MSE_n = [\Phi(G, \{P_k\}_{k=1}^{K}, C)]_{n,n} \\
& \quad tr(p_k r_{x_{k-1}} p_K^H) \leq P_k, \quad k = 1, \cdots, K
\end{align*}
\]
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 $\mathbb{R}^{N} \rightarrow \mathbb{R}$, 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 $[MSE_1, \cdots, MSE_N]^T$ and $[MSE_1, \cdots, MSE_N]^T$, suppose $MSE_n < MSE_n$ 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 $z \in \mathbb{R}^n$, let $z_{(k)}$ denotes the $k^{th}$ largest elements of $z$ and $z_{(k)}$ denotes the $k^{th}$ smallest elements of $z$.\(^2\)
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, \( \mathbf{v} \preceq \mathbf{u} \) is defined as
\[
\prod_{i=1}^{k} v[i] \leq \prod_{i=1}^{k} u[i], \quad k = 1, \ldots, N-1 \quad \text{and} \quad \prod_{i=1}^{N} v[i] = \prod_{i=1}^{N} u[i].
\] (13)

**Definition 2:** A function \( \phi(\bullet) \) is multiplicatively Schur-convex if and only if \( \mathbf{v} \preceq \mathbf{u} \) 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 \( D = \{ \mathbf{z} : z_1 \geq \cdots \geq z_N \geq 0 \} \). Then \( \phi(\bullet) \) is multiplicatively Schur-convex if and only if for all \( \mathbf{z} \in D \),
\[
\phi(z_1, \ldots, z_{k-1}, z_k/x, z_{k+1} \times e, z_{k+2}, \ldots, z_N)
\]
is decreasing in \( e \) over the following regions
\[
1 \leq e \quad \text{and} \quad z_k/e \geq z_{k+1} \times e \quad \text{for} \quad k = 1, \ldots, N-1.
\] (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 notation, 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 \( z_1, z_2, \ldots, z_N > 0 \). 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 system 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) estimator is obtained by setting the differentiation of the trace of (11) with respect to \( \mathbf{G}^* \) (the conjugate of \( \mathbf{G} \)) to be zero, and we have
\[
\begin{align*}
\Phi(\mathbf{G}_{\text{LMMSE}}, \{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C}) &= \Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C}) \\
&\preceq \Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C}).
\end{align*}
\] (16)

which implies
\[
\Phi(\mathbf{G}_{\text{LMMSE}}, \{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C}) \leq \Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C}),
\]
for each \( \Phi \) in (11) and (16), the matrix \( \mathbf{K}_{\mathbf{F}_k} \) can be interpreted as the equivalent noise covariance matrix at the \( k \)-th hop.
The following multiplicative majorization relationship can be established [24]

\[ \sigma^2_{\theta} \left[ \prod_{n=1}^{N} (1 - \lambda_n(\Theta)) \right] \succeq \mathbf{1}_N \succeq \mathbf{1} \succeq \Theta \mathbf{1}_{1,\ldots,1}^T, \]

(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

\[ f([\mathbf{L}]^2_{1,\ldots,1}^T) \geq f\left( \sigma^2_{\theta} \left[ \prod_{n=1}^{N} (1 - \lambda_n(\Theta)) \right] \right) \]

\[ \triangleq \mathbf{e}_\theta(\lambda(\Theta)) \]

(32)

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

**M-Schur-convex:**

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

\[ [\mathbf{L}]^2_{1,\ldots,1}^T \succeq \mathbf{1} \succeq [\mathbf{L}]^2_{N,N}^T \succeq [\mathbf{L}]^2_{N,1}^T \]

(33)

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

\[ f([\mathbf{L}]^2_{1,\ldots,1}^T) \geq f\left( \mathbf{e}_\theta(\lambda(\Theta)) \right). \]

(34)
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 [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 (32) is equivalent to

$$\min_{F_k,Q_k} g[\lambda(\Theta)]$$

subject to

$$\Theta = M^H_1Q^H_1\cdots M^H_KQ^H_KM_K\cdots Q_1M_1,$$

$$\text{Tr}(F_k^HH_k^H) \leq P_k, \quad Q^H_kQ_k = I_{N,k},$$

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

$$g[\lambda(\Theta)] = \begin{cases} f(\sigma^2_k[\prod_{n=1}^N(1-\lambda_n(\Theta))]^{1/N}_n) \otimes 1_N, & \text{if } f(\bullet) \text{ is M-Schur-convex}, \\ f(\sigma^2_k[1_N - \lambda(\Theta)]) \otimes 1_N, & \text{if } f(\bullet) \text{ is M-Schur-concave}. \end{cases}$$

(36)

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 derivations.

**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

$$\prod_{n=1}^N v_n \leq \prod_{n=1}^N u_n, \quad k = 1, \cdots, N.$$  

(37)

Notice that there is a subtle difference between **Definition 1** in (13) and **Definition 3**. In **Definition 3**, when $k = N$, $\prod_{n=1}^N v_n \leq \prod_{n=1}^N u_n$ rather than $\prod_{n=1}^N v_n = \prod_{n=1}^N u_n$ in **Definition 1**.

**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

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

(38)

**Proof:** See Appendix C.

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

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

(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 $\lambda(\Theta)$ has the following relationship

$$\lambda(\Theta) \prec_{x,w} \frac{\gamma_1([F_k^k]_{k=1}^K), \gamma_2([F_k^k]_{k=1}^K), \cdots, \gamma_N([F_k^k]_{k=1}^K)^T}{\gamma_k([F_k^k]_{k=1}^K)},$$

(40)

where $\gamma_n([F_k]_{k=1}^K) = \prod_{k=1}^K \lambda_n(F^H_kH_kF_k + 1) + \lambda_n(F^H_kH_kF_k^H + 1)$, $k = 1, \cdots, K$. Notice that (41) does not cover the design of $Q_K$, but it can be any unitary matrix because it always appears in the form $Q^H_KQ_K$ 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 $\lambda(\Theta) \prec_{x,w} \lambda(\Theta)$ implies $g[\lambda(\Theta)] \geq g[\lambda(\Theta)]$. In the following, we will prove the latter.

When $f(\bullet)$ is M-Schur-convex, $g[\lambda(\Theta)] = f(\sigma^2_k[\prod_{n=1}^N(1-\lambda_n(\Theta))]^{1/N}_n) \otimes 1_N$. 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 $\prod_{n=1}^N(1 - \lambda_n(\Theta))$ is a decreasing function. If $\lambda(\Theta) \prec_{x,w} \lambda(\Theta)$ is true, based on **Lemma 2**, we have

$$\prod_{n=1}^N(1 - \lambda_n(\Theta)) \geq \prod_{n=1}^N(1 - \lambda_n(\Theta)).$$

(42)

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

$$f(\sigma^2_k[\prod_{n=1}^N(1-\lambda_n(\Theta))]^{1/N}_n \otimes 1_N)$$

$$\geq f(\sigma^2_k[\prod_{n=1}^N(1-\lambda_n(\Theta))]^{1/N}_n \otimes 1_N).$$

(43)

$$g[\lambda(\Theta)].$$
On the other hand, when \( f(\bullet) \) is increasing and M-Schur-concave, \( g(\lambda(\Theta)) = f(\sigma^2_k[I_N - \lambda(\Theta)]) \). Using Lemma 3 we directly have \( \lambda(\Theta) \prec_{\lambda,\omega} \lambda(\hat{\Theta}) \) implies

\[
\frac{f(\sigma^2_k[I_N - \lambda(\Theta)])}{g(\lambda(\Theta))} \geq \frac{f(\sigma^2_k[I_N - \lambda(\hat{\Theta})])}{g(\lambda(\hat{\Theta}))}. \tag{44}
\]

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

(a) The constraints \( Q_k^H Q_k = I_{N_{R,k}} \) 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 (35), we have the reformulated optimization problem

\[
\begin{align*}
\min_{F_k} & \quad g(\gamma(\{F_k\}_{k=1}^K)) \\
\text{s.t.} & \quad \gamma_n(\{F_k\}_{k=1}^K) = \prod_{k=1}^K \frac{\lambda_n(F_k^H \hat{H}_k^H K^-1 \hat{H}_k F_k)}{1 + \lambda_n(F_k^H \hat{H}_k^H K^-1 \hat{H}_k F_k)} \\
& \quad \text{Tr}(F_k F_k^H) \leq P_k. \tag{45}
\end{align*}
\]

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

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

\[
\eta_{f_k} = \alpha_k \text{Tr}(F_k F_k^H \Psi_k) + \sigma_{n_k}^2 \quad \text{with} \quad \alpha_k = \text{Tr}(\Sigma_k)/N_{R,k}, \tag{46}
\]

\( \text{Tr}(F_k F_k^H) = P_k \) is exactly equivalent to \( \text{Tr}(F_k^H \hat{H}_k^H K^-1 \hat{H}_k F_k) = P_k \) as proved in [10], [11], [26]. Thus the robust transceiver design problem (45) is equivalent to

\[
\begin{align*}
\min_{F_k} & \quad g(\gamma(\{F_k\}_{k=1}^K)) \\
\text{s.t.} & \quad \gamma_n(\{F_k\}_{k=1}^K) = \prod_{k=1}^K \frac{\lambda_n(F_k^H \hat{H}_k^H \Sigma_k^{-1} \hat{H}_k F_k)}{1 + \lambda_n(F_k^H \hat{H}_k^H \Sigma_k^{-1} \hat{H}_k F_k)} \\
& \quad \text{Tr}(F_k F_k^H (\alpha_k P_k \Psi_k + \sigma_{n_k}^2 I_{N_{T,k}}))/\eta_{f_k} = P_k. \tag{47}
\end{align*}
\]

It is proved in Appendix F 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

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

with \( \xi_k(A_{\mathcal{F}_k}) = \sigma_{n_k}^2/(1 - \alpha_k \text{Tr}(V_{\mathcal{H}_k, N} (\alpha_k P_k \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} A_{\mathcal{F}_k}^H)), \tag{48}
\]

where \( A_{\mathcal{F}_k} \) is a \( N \times N \) unknown diagonal matrix, and \( V_{\mathcal{H}_k, N} \) and \( U_{\mathcal{A}_{R,b}}^{-N} \) are the matrices consisting of the first \( N \) columns of \( V_{\mathcal{H}_k} \) and \( U_{\mathcal{A}_{R,b}}^{-N} \), respectively. The unitary matrix \( U_{\mathcal{A}_{R,b}}^{-N} \) is an arbitrary \( N_{R,k-1} \times N_{R,k-1} \) unitary matrix, and the unitary matrix \( V_{\mathcal{H}_k} \) is defined based on the following singular value decomposition

\[
(K_{F_k}/\eta_{f_k})^{-1/2} \hat{H}_k (\alpha_k P_k \Psi_k + \sigma_{n_k}^2 I_{N_{T,k}})^{-1/2} = U_{\mathcal{H}_k} A_{\mathcal{H}_k} V_{\mathcal{H}_k}^H \tag{49}
\]

where the diagonal elements of \( A_{\mathcal{H}_k} \) 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 \( \sigma_{n,k}^2 \), respectively. When the channels are estimated based on the algorithm proposed in [21], [22], it can be shown that \( \Psi_k = R_{T,k} \) and \( \Sigma_k = \sigma_{n,k}^2 I_{N_{R,k}} + \sigma_{e,k}^2 R_{R,k}^{-1} \). 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 \( \sigma_{n,k}^2 \) will be small and \( \Sigma_k \propto I_{N_{R,k}} \). 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 \( A_{\mathcal{F}_k} \)

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

\[
[A_{\mathcal{H}_k}]_{n,n} = h_{k,n}, \quad [A_{\mathcal{F}_k}]_{n,n} = f_{k,n}, \tag{50}
\]

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

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

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

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

\[
\min_{f_{k,i}} \sum_{n=1}^{N} \log \left( 1 - \prod_{k=1}^{K} f_{k,n}^2 h_{k,n}^2 \right) \\
\text{s.t. } \sum_{n=1}^{N} f_{k,n}^2 = P_k. 
\]

(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

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

(53)

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

M-Schur-convex functions:

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

\[
g(\gamma(\{F_k\}_{k=1}^K)) = \sigma^2 \sum_{n=1}^{N} w_n \prod_{n=1}^{N} (1 - \gamma_n(\{F_k\}_{k=1}^K))^{w_n}
\]

and the optimization (51) is equivalent to

\[
\min_{f_{k,i}} \sum_{n=1}^{N} w_n \log \left( 1 - \prod_{k=1}^{K} f_{k,n}^2 h_{k,n}^2 \right) \\
\text{s.t. } \sum_{n=1}^{N} f_{k,n}^2 = P_k. 
\]

(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 \( f(\cdot) \) 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], [26].

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 design. 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 (49) 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

\[
P_k = (\alpha_k P_k \Psi_k + \sigma^2 m_n \mathbf{I}_{N, k})^{-1/2} \\
\times \mathbf{V}_{\mathcal{H}_k \cup \mathcal{N}} \Lambda_{\mathcal{P}_k} U_{\mathcal{H}_k \cup \mathcal{N}}^{-1} K_{F_{k-1}^{1/2}}^{-1/2}
\]

(55)

where \( \Lambda_{\mathcal{P}_k} \) 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 \( \Lambda_{\mathcal{P}_k} \), all other matrices in (55) are only 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 \( \Lambda_{\mathcal{F}_m} \) 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 \( \prod_{k=1}^{K} f_{k,n} h_{k,n}^2 / (\sum_{k=1}^{K} f_{k,n} h_{k,n}^2 + 1) \), 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 noise amplification. Furthermore, although the three-hop system performs not as well 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 \( \rho_t = 0.5, \rho_r = 0, P_1/\sigma_{n_1}^2 = 30\) dB. 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 when \( \rho_t = 0.5, \rho_r = 0, P_1/\sigma_{n_1}^2 = 30\) 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 \( \rho_r \) increases, the performance gain of the proposed robust design with respect to that based on CSI only becomes larger. It is most obvious when \( \rho_r = 0.9 \) 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. Fig. 4 shows that correlation of channel estimation error would affect the direction of all hops, and subsequently affect the final BER performance. Fig. 5 shows the corresponding BERs for different $\rho_r$, with $\rho_t = 0$, $\sigma_r^2 = 0.002$ and $P_1/\sigma_n^2 = 30$dB.

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

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

(58)

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

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

(59)

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

$$\phi(\hat{v}_1, \hat{v}_2/\hat{u}_1, \cdots, \hat{v}_{L_z}/\hat{v}_{L_z-1}, 0, \cdots)$$

$$\leq \phi(\hat{u}_1, \hat{u}_2/\hat{u}_1, \cdots, \hat{u}_{L_z}/\hat{u}_{L_z-1}, 0, \cdots),$$

(60)

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

To prove $\psi(\mathbf{v})$, we only need to prove that when $\hat{v}_k \leq \hat{u}_k$ and $\hat{v}_l = \hat{u}_l$ for all $l \neq k$, we have $\psi(\mathbf{v}) \leq \psi(\mathbf{u})$ [24]. As $\hat{v}_k \geq 0$ and $\hat{u}_k \geq 0$, $\hat{v}_k \leq \hat{u}_k$ is equivalent to $\hat{v}_k = \hat{u}_k/e$ with $e \geq 1$. Substituting $\hat{v}_k = \hat{u}_k/e$ and $\hat{v}_l = \hat{u}_l$ for all $l \neq k$ into (60) and replacing $u_k = \hat{u}_k/\hat{u}_{k-1}$ for $k \leq L_u-1$, proving $\psi(\mathbf{v}) \leq \psi(\mathbf{u})$ 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 $Q_0$

Following the sufficient conditions given in [32], 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

\[
(I_N - M_1^HM_1^H \cdots M_K^HM_K^H Q_K R_K M_K \cdots Q_1 R_1 M_1)^{1/2} \sigma_b = U_M A_M U_M^H \quad \triangleq A
\]

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

\[
[S]_{i,1} = \sqrt{\frac{|A^H A|^{1/2} - [A_M]_{i,i}}{[A_M]_{i,i}^2 - [A]^2}}
\]

\[
[S]_{N,1} = \sqrt{\frac{|A_M|_{N,N} - |A^H A|^{1/2}}{[A]_{N,N}^2 - [A^H A]}}
\]

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

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

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

\[
(A[S]_{1,k+1})^H[I - A[S]_{1,k+1}]^H A[S]_{1,k+1} = V^{(k)} A^{(k)} V^{(k)^H}
\]

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

\[
[S]_{k+1} = [S]_{1,k+1} V^{(k)} y^{(k)}
\]

and

\[
y^{(k)} = \begin{bmatrix}
\sqrt{\frac{|A^H A|^{1/2} - [A^{(k)}]_{N-k,N-k,N-k}}{[A^{(k)}]_{N-k,N-k}^2 - [A^H A]}}
0_{N-k-1}, \ldots, 0_{N-k-1}, \\
\sqrt{\frac{|A^{(k)}|_{1,1} - |A^H A|^{1/2}}{[A^{(k)}]_{1,1}^2 - [A^H A]}}
\end{bmatrix}
\]

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

\[
y^{(N-1)} = \begin{bmatrix}
\sqrt{\frac{|A^{(N-2)}|_{1,1} - |A^{H} A|^{1/2}}{[A^{(N-2)}]_{1,1}^2 - [A^H A]}}
\sqrt{\frac{|A^{H} A|^{1/2} - [A^{(N-2)}]_{2,1,2}}{[A^{(N-2)}]_{2,1,2}^2 - [A^H A]}}
\end{bmatrix}
\]

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

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

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

\[
\prod_{i=1}^{k} (1 - v_i) \geq \prod_{i=1}^{k} (1 - u_i) > 0, \quad k = 1, \cdots, N.
\]

We construct a vector $\hat{\tau} = [\hat{\tau}_1, \cdots, \hat{\tau}_N]^T$ 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 $\tau_N$ and $\hat{u}_N$ is at $i = N$, when $\phi(\bullet)$ is increasing, we have $\phi(\hat{\tau}) \geq \phi(\hat{u})$ where $\hat{u} = [\hat{u}_1, \cdots, \hat{u}_N]^T$.

On the other hand, based on (69) and the fact that $\hat{\tau}_N = \hat{u}_N$ for $i < N$, it can be concluded that (a) $\prod_{i=1}^{k} \hat{\tau}_i \geq \prod_{i=1}^{k} \hat{u}_i$ for $1 \leq k < N$. Based on the definition of $\hat{\tau}_N$, it can also be concluded that (b) $\prod_{i=1}^{N} \hat{\tau}_i = \prod_{i=1}^{N} \hat{u}_i > 0$. Results (a) and (b) implies $v \prec_{x} \hat{\tau}$ where $\hat{\tau} = [\hat{v}_1, \cdots, \hat{v}_N]^T$ [24]. As $\phi(\bullet)$ is $M$-Schur-concave, using Lemma 1, we have $\phi(\hat{\tau}) \geq \phi(\hat{\tau})$. Together with the conclusion in the last paragraph, we can obtain $\phi(\hat{\tau}) \geq \phi(\hat{u})$. Finally, with $\hat{v} = \hat{1}_N - v$ and $\hat{u} = \hat{1}_N - u$, 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 $\lambda_i(AB) = \lambda_i(BA)$ [24, 9.A.1.a]; (b) for two positive semi-definite matrices $A$ and $B$, $\prod_{i=1}^{n} \lambda_i(AB) \leq \prod_{i=1}^{n} \lambda_i(A) \lambda_i(B)$ [24, 9.1.H.1.a], where the equality holds when $A$ and $B$ has the same unitary matrix in
eigendecomposition. With these two facts, we have
\[
\prod_{i=1}^{n} \lambda_i(M_i^HQ_i^H \cdots M_K^HQ_K^HQ_KM_K \cdots Q_1M_1) \\
= \prod_{i=1}^{n} \lambda_i(M_i^HQ_i^H \cdots M_K^HQ_K^HQ_KM_K \cdots Q_2M_2Q_1M_1M_1^HQ_1^H) \\
\leq \prod_{i=1}^{n} \lambda_i(M_i^HQ_i^H \cdots M_K^HQ_K^HQ_KM_K \cdots Q_2M_2Q_1M_1M_1^HQ_1^H) \\
\times \lambda_i(M_1M_1^H) \quad n = 1, \ldots, N, \quad (70)
\]
where the equality holds when \(Q_k\)s satisfy
\[
Q_k = V_{M_{k+1}}U_{M_k}^H, \quad k = 1, \ldots, K - 1, \quad (72)
\]
where \(U_{M_k}\) and \(V_{M_k}\) are defined based on the following singular value decomposition \(M_k = U_{M_k}A_{M_k}V_{M_k}^H\) with the diagonal elements of \(A_{M_k}\) arranged in decreasing order. Furthermore, based on the definition of \(M_k\) in (23), \(\gamma_i(\{F_k\}_k^{K=1})\) in (71) equals to
\[
\gamma_i(\{F_k\}_k^{K=1}) = \frac{\prod_{k=1}^{K} \lambda_i(F_kH_k^HkH_kF_k)}{\prod_{k=1}^{K} \lambda_i(F_kH_k^HkH_kF_k)} \quad (73).
\]

**APPENDIX F**

**Optimal Structure of \(F_k\)**

Defining new variables
\[
\tilde{F}_k = 1/\sqrt{\eta_{f_k}}(\alpha_kP_k\Psi_k + \sigma^2_{n_k}I_{N_{T,k}})^{-1/2}F_k, \quad \text{and} \quad \tilde{H}_k = (K_{F_k}/\eta_{f_k})^{-1/2}\tilde{H}_k(\alpha_kP_k\Psi_k + \sigma^2_{n_k}I_{N_{T,k}})^{-1/2}, \quad \text{for the optimization problem (47) is reformulated as}
\]
\[
\min_{F_k} \quad g(\gamma(\{F_k\}_k^{K=1})) \\
\text{s.t.} \quad \gamma_i(\{F_k\}_k^{K=1}) = \frac{\prod_{k=1}^{K} \lambda_i(F_kH_k^HkH_kF_k)}{\prod_{k=1}^{K} \lambda_i(F_kH_k^HkH_kF_k)} \quad \text{Tr}(\tilde{F}_kH_k^H) = P_k. \quad (75)
\]
When \(\Psi_k \propto I_{N_{T,k}}\) or \(\Sigma_k \propto I_{N_{R,k}}\), for the optimal solution \(K_{F_k}/\eta_{f_k}\) is constant [10], [11], [26] and thus \(\tilde{H}_k\) is constant. Let \(\tilde{F}_k, \tilde{H}_k\) be the optimal solution of (75). With the following singular value decompositions,
\[
\tilde{H}_k = U_{\tilde{\Lambda}_k}\Lambda_{\tilde{\Lambda}_k}V_{\tilde{\Lambda}_k}^H, \quad \tilde{F}_k = U_{\tilde{\Lambda}_k}\Lambda_{\tilde{\Lambda}_k}V_{\tilde{\Lambda}_k}^H, \quad (76)
\]
where the diagonal elements of \(\Lambda_{\tilde{\Lambda}_k}\) and \(\Lambda_{\tilde{\Lambda}_k}\) are arranged in decreasing order, we can construct a matrix \(\tilde{F}_k\) equals to
\[
\tilde{F}_k = V_{\tilde{\Lambda}_k}\Lambda_{\tilde{\Lambda}_k}V_{\tilde{\Lambda}_k}^H, \quad (77)
\]
where \(\Lambda_{\tilde{\Lambda}_k}\) is a rectangular diagonal matrix with the same rank as \(\Lambda_{\tilde{\Lambda}_k}\) and \(1/b_k\alpha_k\tilde{\Lambda}_k\Lambda_{\tilde{\Lambda}_k} = \Lambda_{\tilde{\Lambda}_k}\). The scalar \(b_k\) is chosen to make that \(\text{Tr}(\tilde{F}_kH_k^H) = P_k\) holds.

Using Lemma 12 in [23], we can show that \(\tilde{F}_kH_k^Hk\Lambda_{\tilde{\Lambda}_k} = \tilde{F}_kH_k^Hk\Lambda_{\tilde{\Lambda}_k}\). Together with the formulation of \(\gamma_i(\{F_k\}_k^{K=1})\) in (75), it can be concluded that \(\gamma_i(\{F_k\}_k^{K=1}) \geq \gamma_i(\{F_{k,\text{opt}}\}_k^{K=1})\). Since \(g(\bullet)\) is an decreasing function, \(g(\gamma_i(\{F_k\}_k^{K=1})) \leq g(\gamma_i(\{F_{k,\text{opt}}\}_k^{K=1}))\). Because \(F_{k,\text{opt}}\) is the optimal solution, it is impossible to have \(g(\gamma_i(\{F_k\}_k^{K=1})) < g(\gamma_i(\{F_{k,\text{opt}}\}_k^{K=1}))\). Therefore, \(\tilde{F}_k\) must be the optimal solution. Furthermore, based on the relationship between \(\tilde{F}_k\) and \(F_k\), it follows that
\[
F_{k,\text{opt}} = \sqrt{\eta_{f_k}}(\alpha_kP_k\Psi_k + \sigma^2_{n_k}I_{N_{T,k}})^{-1/2}V_{\tilde{\Lambda}_k}\Lambda_{\tilde{\Lambda}_k}V_{\tilde{\Lambda}_k}^H, \quad (78)
\]
Notice that in general the unitary matrix \(V_{\tilde{\Lambda}_k}\) depends on the optimal solution \(F_{k,\text{opt}}\). However, from (75), it can be seen that the value of \(V_{\tilde{\Lambda}_k}\) does not affect the objective functions and therefore it can be an arbitrary unitary matrix. Meanwhile, as the minimum dimension of \(F_{k,\text{opt}}H_{k,\text{opt}}\Lambda_{\tilde{\Lambda}_k}F_{k,\text{opt}}^H\) is \(N\), only \(N \times N\) principal submatrices of \(\Lambda_{\tilde{\Lambda}_k}\) can be nonzero. For notational convenience, we denote that \(\Lambda_{\tilde{\Lambda}_k}\) in (1) = 1, 2, N = \(\Lambda_{\tilde{\Lambda}_k}\).

Substituting (78) into the definition of \(\eta_{f_k}\) in (46), we obtain a simple linear function of \(\eta_{f_k}\), and \(\eta_{f_k}\) can be easily solved to be
\[
\eta_{f_k} = \sigma^2_{n_k}/\{1 - \alpha_k\text{Tr}[\tilde{V}_{\tilde{\Lambda}_k}N_{\Lambda_{\tilde{\Lambda}_k}}(\alpha_kP_k\Psi_k + \sigma^2_{n_k}I_{N_{T,k}})^{-1/2}] \times \Psi_k(\alpha_kP_k\Psi_k + \sigma^2_{n_k}I_{N_{T,k}})^{-1/2}/\tilde{V}_{\tilde{\Lambda}_k}N_{\Lambda_{\tilde{\Lambda}_k}}^2\} \quad \Rightarrow \quad \xi_k(\Lambda_{\tilde{\Lambda}_k}). \quad (79)
\]

**REFERENCES**


Chengwen Xing (S’08–M’10) received the B.Eng. degree from Xidian University, Xi’an, China, in 2006 and the Ph.D. degree from 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 multiuser 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.