

A General Robust Linear Transceiver Design for Multi-Hop Amplify-and-Forward MIMO Relaying Systems

Chengwen Xing, Shaodan Ma, Zesong Fei, Yik-Chung Wu and H. Vincent Poor

Abstract

In this paper, linear transceiver design for multi-hop amplify-and-forward (AF) multiple-input multiple-out (MIMO) relaying systems with Gaussian distributed channel estimation errors is investigated. Commonly used transceiver design criteria including weighted mean-square-error (MSE) minimization, capacity maximization, worst-MSE/MAX-MSE minimization and weighted sum-rate maximization, are considered and unified into a single matrix-variate optimization problem. A general robust design algorithm is proposed to solve the unified problem. Specifically, by exploiting majorization theory and properties of matrix-variate functions, the optimal structure of the robust transceiver is derived when either the covariance matrix of channel estimation errors seen from the transmitter side or the corresponding covariance matrix seen from the receiver side is proportional to an identity matrix. Based on the optimal structure, the original transceiver design problems are reduced to much simpler problems with only scalar variables whose solutions are readily obtained by iterative water-filling algorithm. A

This work was supported in part by National Natural Science Foundation of China under Grant No. 61101130. The material in this paper was partially presented at WCSP 2011, Nanjing, China, Nov. 2011.

Chengwen Xing and Zesong Fei are with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. (e-mails : chengwenxing@ieee.org, feizesong@bit.edu.cn).

Shaodan Ma is with the Department of Electrical and Computer Engineering, University of Macau, Macao. (e-mail: shaodanma@umac.mo).

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

H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ , USA. (e-mail: poor@princeton.edu).

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

number of existing transceiver design algorithms are found to be special cases of the proposed solution. The differences between our work and the existing related work are also discussed in detail. The performance advantages of the proposed robust designs are demonstrated by simulation results.

Index Terms

Amplify-and-forward (AF), MIMO relaying, matrix-variate optimization, robust transceiver design.

I. INTRODUCTION

With significant potential to enable the emerging requirements for high speed ubiquitous wireless communications, cooperative communications has been adopted as one of the key components in future wireless communication standards such as long term evolution (LTE), international mobile telecommunications-advanced (IMT-Advanced), the Winner project, etc. Specifically, these developments involve the deployment of relays to enhance the coverage of base stations and to improve the communication quality of wireless links [1]. In general, relays can adopt different relaying strategies, e.g., amplify-and-forward (AF), decode-and-forward (DF) and compress-and-forward (CF). Among these relaying strategies, the AF scheme is the most attractive for practical implementation due to its low complexity and independence of the underlying modulation. On the other hand, it is well-established that employing multiple antennas provides spatial diversity and multiplexing gain in a wireless communication system. It is straightforward to combine AF transmission with multi-input multi-output (MIMO) systems so that the virtues of both techniques can be obtained. The resulting system (termed an AF MIMO relaying system) has attracted considerable interest [2] in recent years.

Transceiver design for AF MIMO relaying systems, which refers to the design of source precoder, relay amplifier and receiver equalizer, has been widely discussed in the literature [3]–[16]. Generally speaking, transceiver design varies from system to system and depends heavily on the design criteria and objectives. The most commonly used criteria are capacity maximization [3], [4], [8] and data mean-square-error (MSE) minimization [5]–[8]. Usually these two criteria are contradictory to each other and call for different algorithms to solve the optimization problems. Interestingly, in [8] a unified framework which is applicable to both capacity maximization and MSE minimization is proposed for transceiver design in dual-hop AF MIMO relay systems. Since multi-hop AF transmission is a promising technique to increase the

coverage of a transmitter, transceiver design for a multi-hop system is further investigated in [10]. It reveals that optimal solutions for both capacity and MSE criteria in a multi-hop system should have diagonal structures. However, in most of the previous works on transceiver design including [8] and [10], channel state information (CSI) is assumed to be perfectly known/estimated. This is difficult to achieve in practice and channel estimation errors are inevitable due to limited training and quantization operation, resulting in significant performance degradation. In order to mitigate the performance degradation, such channel estimation errors are necessary to be taken into account in the transceiver design process. This kind of transceiver is called robust transceiver. It has been shown in [17] and [18] that robust transceiver design is essentially different from the transceiver design with perfect CSI. It is more challenging and different algorithms are required to solve the challenging robust design problem.

In general, channel estimation errors can be modeled in two different ways: norm-bounded errors with known error bound and random errors with certain distribution. Correspondingly, robust transceiver designs can also be classified into two main categories: worst-case robust design for norm-bounded errors [19] and Bayesian robust design for randomly distributed errors [20]. For linear channel estimators, the estimation errors can be accurately modeled as being random with a Gaussian distribution [11]. Under this kind of Gaussian estimation errors, Bayesian robust transceiver design for dual-hop AF relaying systems has been investigated in [14] and solutions for capacity maximization and MSE minimization respectively are proposed by implicitly approximating a design-variable dependent covariance matrix (the matrix \mathbf{A} in [14]) as being constant. Since the approximation is tight only when the covariance matrix of channel estimation errors seen from the receiver side is proportional to an identity matrix, the proposed solutions are sub-optimal for general cases. In [11] and [12], Bayesian robust transceiver design targeting at weighted MSE minimization is discussed for dual-hop AF relaying systems and an optimal solution is found without considering the source precoder. The optimality of the proposed solution is proved to hold under a wide range of cases, i.e., when either the covariance matrix of channel estimation errors from the transmitter side or the corresponding covariance matrix seen from the receiver side is proportional to an identity matrix. These works have been extended to systems with source precoder design and an iterative algorithm has been proposed to find a good solution without guaranteed optimality [15]. Similarly, the robust transceiver design for maximizing mutual information rate for dual-hop AF relaying systems under Gaussian

channel estimation errors at all nodes has been investigated in [16] and a solution without global optimality is proposed with an iterative algorithm. Unfortunately, the aforementioned algorithms are applicable only to dual-hop AF systems and their extension to multi-hop AF systems is by no means straightforward as shown in [10].

In this paper, we investigate robust transceiver design for a general **multi-hop** AF relaying system with Gaussian distributed channel uncertainties. The robust design problem is significantly different from that in the literature and is challenging due to the existence of random channel uncertainties and the complexity of the multihop system. A number of widely used design criteria including weighted MSE minimization, capacity maximization, worst case MSE minimization, and weighted sum-rate maximization are considered and their corresponding robust design problems are unified into one matrix-variate optimization problem. A general robust design algorithm is proposed to solve the unified problem, i.e., to jointly design the precoder at the source, multiple forwarding matrices at the relays, and the equalizer at the destination. Specifically, the structure of the optimal solution for the unified problem is derived based on majorization theory [21], [22] and properties of matrix-monotone functions [22]. It is demonstrated that the derived optimal structure is significantly different from its counterpart with perfect CSI [10] and its optimality holds under a wide range of cases, i.e., when either the covariance matrix of channel estimation errors seen from the transmitter side or the corresponding covariance matrix seen from the receiver side is proportional to an identity matrix. With the optimal structure, the robust design problem is simplified into a design problem with only scalar variables. An iterative water-filling algorithm is then proposed to obtain the remaining unknown parameters in the transceiver. The performance of the proposed robust designs is finally corroborated by simulation results. In addition, it is shown that the proposed solutions cover some existing transceiver design solutions as special cases.

The rest of the paper is organized as follows. In Section II, the signal model for a multi-hop AF system is introduced. Then a unified robust transceiver design problem applicable to weighted MSE minimization, capacity maximization, MAX-MSE minimization and weighted sum-rate maximization, is formulated in Section III. In Section IV, the optimal structure for the robust transceiver is derived and the unified transceiver design problem is reduced to a problem of finding a set of diagonal matrices, which can be solved by an iterative water-filling algorithm. The performance of the proposed robust designs is demonstrated by simulation results

in Section V. Finally, this paper is concluded in Section VI.

The following notation is used throughout this paper. Boldface lowercase letters denote vectors, while boldface uppercase letters denote matrices. The notation \mathbf{Z}^H denotes the Hermitian of the matrix \mathbf{Z} , and $\text{Tr}(\mathbf{Z})$ is the trace of the matrix \mathbf{Z} . The symbol \mathbf{I}_M denotes the $M \times M$ identity matrix, while $\mathbf{0}_{M,N}$ denotes the $M \times N$ all zero matrix. The notation $\mathbf{Z}^{1/2}$ is the Hermitian square root of the positive semidefinite matrix \mathbf{Z} , such that $\mathbf{Z}^{1/2}\mathbf{Z}^{1/2} = \mathbf{Z}$ and $\mathbf{Z}^{1/2}$ is also a Hermitian matrix. The symbol $\lambda_i(\mathbf{Z})$ represents the i^{th} largest eigenvalue of \mathbf{Z} . The symbol \otimes denotes the Kronecker product. For two Hermitian matrices, $\mathbf{C} \succeq \mathbf{D}$ means that $\mathbf{C} - \mathbf{D}$ is a positive semi-definite matrix. The symbol $\mathbf{\Lambda} \searrow$ represents a rectangular diagonal matrix with decreasing diagonal elements.

II. SYSTEM MODEL

In this paper, a multi-hop AF MIMO relaying system is considered. As shown in Fig. 1, one source with N_1 antennas wants to communicate with a destination with M_K antennas through $K - 1$ relays. The k^{th} relay has M_k receive antennas and N_{k+1} transmit antennas. It is obvious that the dual-hop AF MIMO relaying system is a special case of this configuration when $K = 2$.

At the source, an $N \times 1$ data vector \mathbf{s} with covariance matrix $\mathbf{R}_s = \mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_N$ is transmitted after being precoded by a precoder matrix \mathbf{P}_1 . The received signal \mathbf{x}_1 at the first relay is $\mathbf{x}_1 = \mathbf{H}_1\mathbf{P}_1\mathbf{s} + \mathbf{n}_1$ where \mathbf{H}_1 is the MIMO channel matrix between the source and the first relay, and \mathbf{n}_1 is an additive Gaussian noise vector at the first relay with zero mean and covariance matrix $\mathbf{R}_{n_1} = \sigma_{n_1}^2 \mathbf{I}_{M_1}$.

At the first relay, the received signal \mathbf{x}_1 is multiplied by a forwarding matrix \mathbf{P}_2 and then the resulting signal is transmitted to the second relay. The received signal \mathbf{x}_2 at the second relay is $\mathbf{x}_2 = \mathbf{H}_2\mathbf{P}_2\mathbf{x}_1 + \mathbf{n}_2$, where \mathbf{H}_2 is the MIMO channel matrix between the first relay and the second relay, and \mathbf{n}_2 is an additive Gaussian noise vector at the second relay with zero mean and covariance matrix $\mathbf{R}_{n_2} = \sigma_{n_2}^2 \mathbf{I}_{M_2}$. Similarly, the received signal at the k^{th} relay can be written as

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

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

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

$$\mathbf{y} = \left[\prod_{k=1}^K \mathbf{H}_k \mathbf{P}_k \right] \mathbf{s} + \sum_{k=1}^{K-1} \left\{ \left[\prod_{l=k+1}^K \mathbf{H}_l \mathbf{P}_l \right] \mathbf{n}_k \right\} + \mathbf{n}_K, \quad (2)$$

where $\prod_{k=1}^K \mathbf{Z}_k$ denotes $\mathbf{Z}_K \times \cdots \times \mathbf{Z}_1$. It is generally assumed that N_k and M_k are greater than or equal to N in order to guarantee that the transmitted data \mathbf{s} can be recovered at the destination [5].

In practical systems, because of limited length of training sequences, channel estimation errors are inevitable. With channel estimation errors, the channel matrix can be written as

$$\mathbf{H}_k = \bar{\mathbf{H}}_k + \Delta \mathbf{H}_k, \quad (3)$$

where $\bar{\mathbf{H}}_k$ is the estimated channel matrix in the k^{th} hop and $\Delta \mathbf{H}_k$ is the corresponding channel estimation error matrix whose elements are zero mean Gaussian random variables. Moreover, the $M_k \times N_k$ matrix $\Delta \mathbf{H}_k$ can be decomposed using the widely used Kronecker model as $\Delta \mathbf{H}_k = \Sigma_k^{1/2} \mathbf{H}_{W,k} \Psi_k^{1/2}$ [11]–[13], [17], [18], [20]. The elements of the $M_k \times N_k$ matrix $\mathbf{H}_{W,k}$ are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance. The specific properties of the row correlation matrix Σ_k and the column correlation matrix Ψ_k are determined by the training sequences and channel estimators being used [11], [17]. Note that Σ_k and Ψ_k correspond to the covariance matrices of the channel estimation errors seen from the transmitter and receiver sides, respectively.

At the destination, a linear equalizer \mathbf{G} is employed to detect the desired data vector \mathbf{s} . The resulting data MSE matrix equals to $\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K) = \mathbb{E}\{(\mathbf{G}\mathbf{y} - \mathbf{s})(\mathbf{G}\mathbf{y} - \mathbf{s})^H\}$, where the expectation is taken with respect to random data, channel estimation errors, and noise¹. Following a similar derivation in dual-hop systems [12], the MSE matrix is derived to be

$$\begin{aligned} \Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K) &= \mathbb{E}\{(\mathbf{G}\mathbf{y} - \mathbf{s})(\mathbf{G}\mathbf{y} - \mathbf{s})^H\} \\ &= \mathbf{G} [\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^H \bar{\mathbf{H}}_K^H + \text{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^H \Psi_K) \Sigma_K \\ &\quad + \mathbf{R}_{\mathbf{n}_K}] \mathbf{G}^H + \mathbf{I}_N - \left[\prod_{k=1}^K \bar{\mathbf{H}}_k \mathbf{P}_k \right]^H \mathbf{G}^H - \mathbf{G} \left[\prod_{k=1}^K \bar{\mathbf{H}}_k \mathbf{P}_k \right], \end{aligned} \quad (4)$$

¹Here the channel estimation errors are assumed unknown at all the nodes. The data MSE matrix at the receiver should thus be computed by taking expectation against all the unknown random variables including data, noise and channel estimation errors.

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

$$\mathbf{R}_{\mathbf{x}_k} = \bar{\mathbf{H}}_k \mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^H \bar{\mathbf{H}}_k^H + \text{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^H \boldsymbol{\Psi}_k) \boldsymbol{\Sigma}_k + \mathbf{R}_{n_k}, \quad (5)$$

and $\mathbf{R}_{\mathbf{x}_0} = \mathbf{R}_s = \mathbf{I}_N$ represents the signal covariance matrix at the source.

III. TRANSCIVER DESIGN PROBLEMS

A. Objective Functions

There are various performance metrics for transceiver design. In the following, four widely used metrics are discussed.

(1) *Weighted MSE*: With the data MSE defined in (4), weighted MSE can be directly written as

$$\text{Obj 1: } \text{Tr}[\mathbf{W} \Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)] \quad (6)$$

where the weighting matrix \mathbf{W} is a positive semi-definite matrix [23]. Here \mathbf{W} is not restricted to be a diagonal matrix. Given any two matrices \mathbf{N} and \mathbf{M} satisfying $\mathbf{N} \succeq \mathbf{M} \succeq \mathbf{0}$, we have $\text{Tr}[\mathbf{W}\mathbf{N}] \geq \text{Tr}[\mathbf{W}\mathbf{M}]$. The weighted MSE is thus a matrix-monotonically increasing function of $\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)$ [32]. Clearly, transceiver design with weighted MSE minimization aims at minimizing the distortion between the recovered and the transmitted signal [5], [24], [25].

(2) *Capacity*: Capacity maximization is another important and widely used performance metric for transceiver design. Denoting the received pilot for channel estimation as \mathbf{r} , the channel capacity between the source and destination is $I(\mathbf{s}; \mathbf{y}|\mathbf{r})$ [26]. To the best of our knowledge, the exact capacity of MIMO channels with channel estimation errors is still open even for point-to-point MIMO systems [18], [26]. However, a lower bound of the capacity can be found as

$$-\log|\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)| \leq I(\mathbf{s}; \mathbf{y}|\mathbf{r}). \quad (7)$$

The equality in (7) holds when perfect CSI is known [4], [24]. For imperfect CSI, the tightness of this bound is extensively investigated in [26], [29]. This lower bound $-\log|\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)|$ can be interpreted as the sum-rate of multiple transmitted data streams when linear equalizer \mathbf{G} is employed. It has been widely used to replace the unknown exact capacity as a performance

metric. Based on this lower bound, the robust transceiver design maximizing capacity can be replaced by minimizing the following objective function [18], [29]:

$$\text{Obj 2: } \log|\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)|. \quad (8)$$

(3) *Worst MSE*: Notice that capacity maximization criterion (**Obj 2**) does not impose any fairness on the simultaneously transmitted multiple data streams, while the weighted MSE minimization criterion (**Obj 1**) imposes only a limited degree of fairness on the data as it involves only a linear operation on the MSE. When fairness is required to balance the performance across different data streams, worst MSE minimization is a good alternative for such transceiver design. In general, the worst MSE can be represented as [24]

$$\text{Obj 3: } \psi_1[\text{d}(\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K))] \quad (9)$$

where $\psi_1(\bullet)$ is an increasing Schur-convex function and $\text{d}(\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K))$ denotes a vector consisting of the diagonal elements of $\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)$, i.e.,

$$\text{d}(\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)) = [[\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)]_{1,1} \cdots [\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)]_{N,N}]^T, \quad (10)$$

with the symbol $[\mathbf{Z}]_{i,j}$ representing the $(i, j)^{\text{th}}$ entry of \mathbf{Z} . It follows that $\psi_1[\text{d}(\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K))]$ is also a matrix-monotonically increasing function with respect to $\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)$. Note that the objective function in (9) is applicable to other design criteria involving fairness considerations.

(4) *Weighted sum rate*: When a preference is required to be given to a certain data stream (e.g., loading more resources to the data streams with better channel state information so that the weighted sum rate is maximized), the objective function can be written as [24]

$$\text{Obj 4: } \psi_2[\text{d}(\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K))] \quad (11)$$

where $\psi_2(\bullet)$ is an increasing Schur-concave function. Similarly to (9), this function $\psi_2[\text{d}(\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K))]$ is a matrix-monotonically increasing function with respect to $\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)$.

Remark 1: Some objective functions on signal to inference plus noise ratio (SINR) and bit error rate (BER) can also be formulated as (9) or (11) and thus can be incorporated into our framework. For example, when the objective is to maximize a sum of weighted SINRs, the objective function can be formulated as the form of (11) as an increasing Schur-concave function of the diagonal elements of the MSE matrix. Similarly, when the objective is to maximize the harmonic mean of SINRs or to maximize the minimal SINR, the objective function can be formulated as the form

of (9) as an increasing Schur-convex function of the diagonal elements of the MSE matrix. On the other hand, when BER minimization is concerned, when all the data streams are modulated using the same scheme, the average BER can be approximated as an increasing Schur-convex function of the diagonal elements of the MSE matrix [21] and can be incorporated into the category of Objective 3 in (9).

B. Problem Formulation

Although the above four criteria aim at different objectives, they have one common feature, that is, the objective functions are matrix-monotonically increasing functions with respect to the data MSE. The corresponding transceiver design problems can therefore be unified into a single form:

$$\begin{aligned} \min_{\mathbf{P}_k, \mathbf{G}} \quad & \mathbf{f}(\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)) \\ \text{s.t.} \quad & \text{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^H) \leq P_k, \quad k = 1, \dots, K \end{aligned} \quad (12)$$

where $\mathbf{f}(\bullet)$ is a real-value matrix monotonically increasing function with $\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)$ as its argument. Notice that the constraints here are imposed on the powers averaged over the channel estimation errors.

With the definition of the data MSE (4) and by differentiating the trace of the MSE with respect to \mathbf{G} and setting the result to zero, we can easily obtain a linear minimum MSE (LMMSE) equalizer as [27]

$$\mathbf{G}_{\text{LMMSE}} = \left[\prod_{k=1}^K \bar{\mathbf{H}}_k \mathbf{P}_k \right]^H \left[\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^H \bar{\mathbf{H}}_K^H + \text{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^H \Psi_K) \Sigma_K + \mathbf{R}_{n_K} \right]^{-1}, \quad (13)$$

with the following property [23], [24]:

$$\Phi(\mathbf{G}_{\text{LMMSE}}, \{\mathbf{P}_k\}_{k=1}^K) \preceq \Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K). \quad (14)$$

The above equality holds when $\mathbf{G} = \mathbf{G}_{\text{LMMSE}}$. Because $\mathbf{f}(\bullet)$ is matrix-monotonically increasing, it follows easily from (14) that $\mathbf{f}(\Phi(\mathbf{G}_{\text{LMMSE}}, \{\mathbf{P}_k\}_{k=1}^K)) \leq \mathbf{f}(\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K))$. It means that $\mathbf{f}(\Phi(\mathbf{G}_{\text{LMMSE}}, \{\mathbf{P}_k\}_{k=1}^K))$ is a tight lower bound of the objective function in (12). Together with

the fact that the equalizer \mathbf{G} is not involved in the constraints in (12), the optimization problem in (12) is equivalent to

$$\begin{aligned} \min_{\mathbf{P}_k} \quad & f(\Phi(\mathbf{G}_{\text{LMMSE}}, \{\mathbf{P}_k\}_{k=1}^K)) \\ \text{s.t.} \quad & \text{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^H) \leq P_k, \quad k = 1, \dots, K. \end{aligned} \quad (15)$$

It implies that the optimal equalizer of (12) is $\mathbf{G}_{\text{LMMSE}}$ in (13). Substituting the optimal equalizer into $\Phi(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K)$ in (4) and denoting $\Phi(\mathbf{G}_{\text{LMMSE}}, \{\mathbf{P}_k\}_{k=1}^K) = \Phi_{\text{MMSE}}(\{\mathbf{P}_k\}_{k=1}^K)$ for simplicity, we have

$$\begin{aligned} \Phi_{\text{MMSE}}(\{\mathbf{P}_k\}_{k=1}^K) = \mathbf{I}_N - & \left[\prod_{k=1}^K \bar{\mathbf{H}}_k \mathbf{P}_k \right]^H [\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^H \bar{\mathbf{H}}_K^H \\ & + \text{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^H \boldsymbol{\Psi}_K) \boldsymbol{\Sigma}_K + \mathbf{R}_{n_K}]^{-1} \left[\prod_{k=1}^K \bar{\mathbf{H}}_k \mathbf{P}_k \right]. \end{aligned} \quad (16)$$

For multi-hop AF MIMO relaying systems, the received signal at the k^{th} relay depends on the forwarding matrices at all preceding relays, causing the power allocations at different relays to be coupled to each other (as seen in the constraints of (15)), and thus making the problem (15) difficult to solve. To proceed, we define the following new variables in terms of \mathbf{P}_k :

$$\begin{aligned} \mathbf{F}_1 & \triangleq \mathbf{P}_1 \mathbf{Q}_0^H, \\ \mathbf{F}_k & \triangleq \mathbf{P}_k \mathbf{K}_{\mathbf{F}_{k-1}}^{1/2} \underbrace{(\mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2} \bar{\mathbf{H}}_{k-1} \mathbf{F}_{k-1} \mathbf{F}_{k-1}^H \bar{\mathbf{H}}_{k-1}^H \mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2} + \mathbf{I}_{M_{k-1}})}_{\triangleq \boldsymbol{\Pi}_{k-1}}^{1/2} \mathbf{Q}_{k-1}^H \end{aligned} \quad (17)$$

where $\mathbf{K}_{\mathbf{F}_k} \triangleq \text{Tr}(\mathbf{F}_k \mathbf{F}_k^H \boldsymbol{\Psi}_k) \boldsymbol{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I}_{M_k}$ and \mathbf{Q}_k is an unknown unitary matrix. The introduction of \mathbf{Q}_k is due to the fact that for a positive semi-definite matrix \mathbf{M} , its square root has the form $\mathbf{M}^{1/2} \mathbf{Q}$ where \mathbf{Q} is a unitary matrix. With the new variables, the MMSE matrix $\Phi_{\text{MMSE}}(\{\mathbf{P}_k\}_{k=1}^K)$ (16) is reformulated as

$$\begin{aligned} \Phi_{\text{MMSE}}(\{\mathbf{P}_k\}_{k=1}^K) = \mathbf{I}_N - & \mathbf{Q}_0^H \left[\prod_{k=1}^K \mathbf{Q}_k \boldsymbol{\Pi}_k^{-1/2} \mathbf{K}_{\mathbf{F}_k}^{-1/2} \bar{\mathbf{H}}_k \mathbf{F}_k \right]^H \left[\prod_{k=1}^K \mathbf{Q}_k \underbrace{\boldsymbol{\Pi}_k^{-1/2} \mathbf{K}_{\mathbf{F}_k}^{-1/2} \bar{\mathbf{H}}_k \mathbf{F}_k}_{\triangleq \mathbf{A}_k} \right] \mathbf{Q}_0 \\ = \mathbf{I}_N - & \mathbf{Q}_0^H \mathbf{A}_1^H \mathbf{Q}_1^H \cdots \mathbf{A}_K^H \mathbf{Q}_K^H \mathbf{Q}_K \mathbf{A}_K \cdots \mathbf{Q}_1 \mathbf{A}_1 \mathbf{Q}_0. \end{aligned} \quad (18)$$

Meanwhile, the power constraint in the k^{th} hop (i.e., $\text{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^H) \leq P_k$) can now be rewritten as

$$\text{Tr}(\mathbf{F}_k \mathbf{F}_k^H) \leq P_k. \quad (19)$$

It is clear that with the new variables \mathbf{F}_k , the constraints become independent of each other. Moreover, the latter transformation of the objective function in the unified problem will not affect the constraints, thus improving the tractability of the problem. Putting (18) and (19) into (15), the unified transceiver design problem can be reformulated as

$$\begin{aligned}
\mathbf{P1}: \quad & \min_{\mathbf{F}_k, \mathbf{Q}_k} \quad f(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0) \\
\text{s.t.} \quad & \text{Tr}(\mathbf{F}_k \mathbf{F}_k^H) \leq P_k, \quad k = 1, \dots, K \\
& \Theta = \mathbf{A}_1^H \mathbf{Q}_1^H \cdots \mathbf{A}_K^H \mathbf{Q}_K^H \mathbf{Q}_K \mathbf{A}_K \cdots \mathbf{Q}_1 \mathbf{A}_1 \\
& \mathbf{Q}_k^H \mathbf{Q}_k = \mathbf{I}_{M_k}. \tag{20}
\end{aligned}$$

From the definition of \mathbf{A}_k in (18) and noticing that $\mathbf{K}_{\mathbf{F}_k} = \text{Tr}(\mathbf{F}_k \mathbf{F}_k^H \Psi_k) \Sigma_k + \sigma_{n_k}^2 \mathbf{I}_{M_k}$, it can be seen that the design variable \mathbf{F}_k appears at multiple positions in the objective function and is involved in matrix inversion and square root operations through $\mathbf{K}_{\mathbf{F}_k}$. This is significantly different from transceiver design for multi-hop MIMO relaying systems with perfect CSI in [10]. Therefore, the optimization problem is much more complicated than its counterpart with perfect CSI. Indeed, as demonstrated by, e.g., [11] and [17], [18], [20], robust transceiver design is much more complicated and challenging than its counterpart with perfect CSI even for point-to-point or dual-hop relaying MIMO systems.

IV. OPTIMAL SOLUTION FOR THE ROBUST TRANSCEIVER

Clearly from the formulation of **P1** in (20), two sets of matrix variables (i.e., $\mathbf{F}_k, \mathbf{Q}_k$) need to be determined. In this section, their optimal structures will be derived first, which enables the simplification of the optimization problem in (20) into a problem with only scalar variables. An iterative water-filling algorithm is then applied to solve the simplified problem. The relationship between our proposed solution and a number of existing solutions will also be discussed in detail.

A. Optimal \mathbf{Q}_k

Based on the formulations of the objectives given in (6), (8), (9) and (11), we have the following property of the optimization problem **P1**.

Property 1: At the optimal value of $\mathbf{P1}$, $\mathbf{Q}_0^H \Theta \mathbf{Q}_0$ and the objective function $f(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0)$ can be written respectively as

$$\mathbf{Q}_0^H \Theta \mathbf{Q}_0 = \mathbf{U}_\Omega \text{diag}[\boldsymbol{\lambda}(\Theta)] \mathbf{U}_\Omega^H, \quad (21)$$

$$f(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0) = f(\mathbf{I}_N - \mathbf{U}_\Omega \text{diag}[\boldsymbol{\lambda}(\Theta)] \mathbf{U}_\Omega^H) \triangleq g[\boldsymbol{\lambda}(\Theta)], \quad (22)$$

where $g(\bullet)$ is a monotonically decreasing and Schur-concave function with respect to $\boldsymbol{\lambda}(\Theta)$ ²; the vector $\boldsymbol{\lambda}(\Theta) = [\lambda_1(\Theta), \dots, \lambda_N(\Theta)]^T$ with $\lambda_i(\Theta)$ being the i^{th} largest eigenvalue of Θ ; and

$$\mathbf{U}_\Omega = \begin{cases} \mathbf{U}_W & \text{for Obj 1} \\ \mathbf{U}_{\text{Arb}} & \text{for Obj 2} \\ \mathbf{Q}_{\text{DFT}} & \text{for Obj 3} \\ \mathbf{I}_N & \text{for Obj 4} \end{cases}. \quad (23)$$

In (23), the matrix \mathbf{U}_W is unitary and defined from the eigen-decomposition of the weighting matrix \mathbf{W} , i.e., $\mathbf{W} = \mathbf{U}_W \boldsymbol{\Lambda}_W \mathbf{U}_W^H$ with $\boldsymbol{\Lambda}_W \searrow$; the matrix \mathbf{U}_{Arb} is an arbitrary unitary matrix; and \mathbf{Q}_{DFT} is the discrete Fourier transform (DFT) matrix making $\mathbf{Q}_{\text{DFT}} \text{diag}[\boldsymbol{\lambda}(\Theta)] \mathbf{Q}_{\text{DFT}}^H$ have identical diagonal elements.

Proof: See Appendix A. ■

We notice that the equality in (21) will hold directly, when

$$\mathbf{Q}_0 = \mathbf{U}_\Theta \mathbf{U}_\Omega^H \quad (24)$$

where \mathbf{U}_Θ is the unitary matrix corresponding to the eigen-decomposition of Θ with eigenvalues in decreasing order. Since \mathbf{Q}_0 is not involved in the constraints in (20), it follows from **Property 1** that $\mathbf{Q}_0 = \mathbf{U}_\Theta \mathbf{U}_\Omega^H$ is the optimal solution of \mathbf{Q}_0 for **P1**.

Using **Property 1**, the objective function of (20) can be directly replaced by $g[\boldsymbol{\lambda}(\Theta)]$ and thus the optimization problem is equivalent to

$$\begin{aligned} \mathbf{P2:} \quad & \min_{\mathbf{F}_k, \mathbf{Q}_k} g[\boldsymbol{\lambda}(\Theta)] \\ \text{s.t.} \quad & \Theta = \mathbf{A}_1^H \mathbf{Q}_1^H \cdots \mathbf{A}_K^H \mathbf{Q}_K^H \mathbf{Q}_K \mathbf{A}_K \cdots \mathbf{Q}_1 \mathbf{A}_1 \\ & \text{Tr}(\mathbf{F}_k \mathbf{F}_k^H) \leq P_k, \quad \mathbf{Q}_k^H \mathbf{Q}_k = \mathbf{I}_{M_k}, \quad k = 1, \dots, K. \end{aligned} \quad (25)$$

²The specific expressions for $g(\bullet)$ are given in Appendix A, but they are not important for the derivation of the optimal structures.

For this optimization problem, we have another property as follows.

Property 2: As $g(\bullet)$ is a decreasing and Schur-concave function, the objective function in **P2** satisfies

$$g(\boldsymbol{\lambda}(\boldsymbol{\Theta})) \geq g(\underbrace{[\gamma_1(\{\mathbf{F}_k\}_{k=1}^K) \cdots \gamma_N(\{\mathbf{F}_k\}_{k=1}^K)]^T}_{\triangleq \gamma(\{\mathbf{F}_k\}_{k=1}^K)}) \quad (26)$$

$$\text{with } \gamma_i(\{\mathbf{F}_k\}_{k=1}^K) \triangleq \prod_{k=1}^K \frac{\lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}{1 + \lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}, \quad (27)$$

with the equality in (26) holds when

$$\mathbf{Q}_k = \mathbf{V}_{\mathbf{A}_{k+1}} \mathbf{U}_{\mathbf{A}_k}^H, \quad k = 1, \dots, K-1, \quad (28)$$

and \mathbf{Q}_K is an arbitrary unitary matrix. In (28), unitary matrices $\mathbf{U}_{\mathbf{A}_k}$ and $\mathbf{V}_{\mathbf{A}_k}$ are defined based on the singular value decomposition (SVD) $\mathbf{A}_k = \mathbf{U}_{\mathbf{A}_k} \boldsymbol{\Lambda}_{\mathbf{A}_k} \mathbf{V}_{\mathbf{A}_k}^H$ with $\boldsymbol{\Lambda}_{\mathbf{A}_k} \searrow$.

Proof: See Appendix B. ■

It is clear from (28) that $\mathbf{Q}_k, k = 1, \dots, K-1$ can be uniquely computed from \mathbf{A}_k which is determined only by \mathbf{F}_k as shown in (18). Similarly, according to (24) and the definition of $\boldsymbol{\Theta}$, it can be concluded that \mathbf{Q}_0 is determined by $\mathbf{Q}_k, k = 1, \dots, K$ and \mathbf{A}_k , and therefore it is eventually determined only by \mathbf{F}_k . With this fact and **Property 2**, the optimization problem with two set of variables of \mathbf{F}_k and \mathbf{Q}_k in **P2** (25) can be reduced to the optimization problem with only one set of variables of \mathbf{F}_k as follows:

$$\begin{aligned} \mathbf{P3:} \quad & \min_{\mathbf{F}_k} \quad g[\gamma(\{\mathbf{F}_k\}_{k=1}^K)] \\ \text{s.t.} \quad & \gamma_i(\{\mathbf{F}_k\}_{k=1}^K) = \prod_{k=1}^K \frac{\lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}{1 + \lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)} \\ & \text{Tr}(\mathbf{F}_k \mathbf{F}_k^H) \leq P_k, \quad k = 1, \dots, K \end{aligned} \quad (29)$$

B. Optimal Structure of \mathbf{F}_k

Since $g(\bullet)$ is a monotonically decreasing function of its vector argument, we have the following additional property of the optimal solution of \mathbf{F}_k in **P3**.

Property 3: The optimal solutions of the optimization problem **P3** in (29) always occur on the boundary, i.e., $\text{Tr}(\mathbf{F}_k \mathbf{F}_k^H) = P_k$ and the power constraint is equivalent to

$$\text{Tr}[\mathbf{F}_k \mathbf{F}_k^H (\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})] / \eta_{f_k} = P_k, \quad (30)$$

where α_k is a constant as $\alpha_k = \text{Tr}(\Sigma_k)/M_k$ and

$$\eta_{f_k} \triangleq \text{Tr}(\mathbf{F}_k \mathbf{F}_k^H \Psi_k) \alpha_k + \sigma_{n_k}^2. \quad (31)$$

Proof: See Appendix C. ■

With **Property 3**, the optimal solution of the optimization problem (29) is exactly the optimal solution of the following optimization problem with different constraints:

$$\begin{aligned} \mathbf{P4:} \quad & \min_{\mathbf{F}_k} \quad g[\gamma(\{\mathbf{F}_k\}_{k=1}^K)] \\ \text{s.t.} \quad & \gamma_i(\{\mathbf{F}_k\}_{k=1}^K) = \prod_{k=1}^K \frac{\lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}{1 + \lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)} \\ & \text{Tr}[\mathbf{F}_k \mathbf{F}_k^H (\alpha_k P_k \Psi_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})] / \eta_{f_k} = P_k. \end{aligned} \quad (32)$$

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

$$(\mathbf{K}_{\mathbf{F}_k} / \eta_{f_k})^{-1/2} \bar{\mathbf{H}}_k (\alpha_k P_k \Psi_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} = \mathbf{U}_{\mathcal{H}_k} \Lambda_{\mathcal{H}_k} \mathbf{V}_{\mathcal{H}_k}^H \quad (33)$$

with singular values in decreasing order, we have the key result about the optimal structure of \mathbf{F}_k as follows.

Property 4: When $\Psi_k \propto \mathbf{I}$ or $\Sigma_k \propto \mathbf{I}$, the matrix $\mathbf{K}_{\mathbf{F}_k} / \eta_{f_k}$ is constant and independent of \mathbf{F}_k . Meanwhile, the optimal solution of the optimization problem (32) has the following structure:

$$\mathbf{F}_{k,\text{opt}} = \sqrt{\xi_k(\Lambda_{\mathcal{F}_k})} (\alpha_k P_k \Psi_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} \mathbf{V}_{\mathcal{H}_k, N} \Lambda_{\mathcal{F}_k} \mathbf{U}_{\mathcal{F}_k, N}^H, \quad (34)$$

where $\mathbf{V}_{\mathcal{H}_k, N}$ and $\mathbf{U}_{\mathcal{F}_k, N}$ are the matrices consisting of the first N columns of $\mathbf{V}_{\mathcal{H}_k}$ and $\mathbf{U}_{\mathcal{F}_k}$, respectively; $\mathbf{U}_{\mathcal{F}_k}$ is an arbitrary unitary matrix; $\Lambda_{\mathcal{F}_k}$ is a $N \times N$ unknown diagonal matrix; and the scalar $\xi_k(\Lambda_{\mathcal{F}_k})$ is a function of $\Lambda_{\mathcal{F}_k}$ and equals

$$\begin{aligned} \xi_k(\Lambda_{\mathcal{F}_k}) &= \sigma_{n_k}^2 / \{1 - \alpha_k \text{Tr}[\mathbf{V}_{\mathcal{H}_k, N}^H (\alpha_k P_k \Psi_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} \Psi_k (\alpha_k P_k \Psi_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} \mathbf{V}_{\mathcal{H}_k, N} \Lambda_{\mathcal{F}_k}^2]\} \\ &= \eta_{f_k}. \end{aligned} \quad (35)$$

Proof: See Appendix D. ■

Remark 2: When reversing the direction of data transmission in the multi-hop system, we can get a dual multi-hop system where the estimated channel matrix in its $(K - k + 1)^{\text{th}}$ hop becomes $\bar{\mathbf{H}}_k^H$ and the roles of row correlation matrices and column correlation matrices are interchanged. Using (17) and **Property 4** and after some tedious manipulation, the optimal precoder matrices

$\mathbf{P}'_{k,\text{opt}}$ for the dual multi-hop system can be found to be $\beta_k \mathbf{P}_{k,\text{opt}}^{\text{H}}$ where β_k is a scalar. This means that there exists an uplink-downlink duality in the multi-hop AF MIMO relaying systems with channel estimation errors.

In the optimal structure given by (34), the scalar variable $\xi_k(\Lambda_{\mathcal{F}_k})$ can be uniquely determined by the matrix $\Lambda_{\mathcal{F}_k}$ and therefore the only unknown variable in (34) is $\Lambda_{\mathcal{F}_k}$. The computation of $\Lambda_{\mathcal{F}_k}$ will be addressed in detail in the following subsection.

C. Computation of $\Lambda_{\mathcal{F}_k}$

Substituting the optimal structures given by **Property 4** into **P4** and defining $[\Lambda_{\mathcal{H}_k}]_{i,i} = h_{k,i}$ and $[\Lambda_{\mathcal{F}_k}]_{i,i} = f_{k,i}$ for $i = 1, \dots, N$, the optimization problem for computing $\Lambda_{\mathcal{F}_k}$ becomes

$$\begin{aligned} \min_{f_{k,i}} \quad & \mathbf{g}[\gamma(\{\mathbf{F}_k\}_{k=1}^K)] \\ \text{s.t.} \quad & \gamma_i(\{\mathbf{F}_k\}_{k=1}^K) = \frac{\prod_{k=1}^K f_{k,i}^2 h_{k,i}^2}{\prod_{k=1}^K (f_{k,i}^2 h_{k,i}^2 + 1)} \\ & \sum_{i=1}^N f_{k,i}^2 = P_k. \end{aligned} \quad (36)$$

The exact expression for $\mathbf{g}(\bullet)$ depends on the specific design criterion used for transceiver design. For all four criteria discussed in Section III-A, a widely used and computationally efficient iterative algorithm can be applied to solve for $f_{k,i}$ from (36) [31], even though the optimization problem (36) is non-convex in nature. For completeness, the optimal solution for $f_{k,i}$ will be given case by case in the following.

1) *Weighted MSE Minimization*: For weighted MSE minimization, it is proved in Appendix A that $\mathbf{g}[\gamma(\{\mathbf{F}_k\}_{k=1}^K)] = \sum_{i=1}^N (w_i - w_i \gamma_i(\{\mathbf{F}_k\}_{k=1}^K))$ where $w_i = [\Lambda_{\mathbf{w}}]_{i,i}$. Therefore, the optimization problem (36) can be rewritten as

$$\begin{aligned} \min_{f_{k,i}} \quad & \sum_{i=1}^N \left(w_i - \frac{w_i \prod_{k=1}^K f_{k,i}^2 h_{k,i}^2}{\prod_{k=1}^K (f_{k,i}^2 h_{k,i}^2 + 1)} \right) \\ \text{s.t.} \quad & \sum_{i=1}^N f_{k,i}^2 = P_k. \end{aligned} \quad (37)$$

Using the iterative water-filling algorithm, $f_{k,i}$ can be directly computed with given $f_{l,i}$'s where $l \neq k$ as

$$f_{k,i}^2 = \left(\sqrt{\frac{w_i}{\mu_k h_{k,i}^2}} \sqrt{\prod_{l \neq k} \left\{ \frac{f_{l,i}^2 h_{l,i}^2}{1 + f_{l,i}^2 h_{l,i}^2} \right\}} - \frac{1}{h_{k,i}^2} \right)^+, \quad i = 1, \dots, N, \quad (38)$$

where μ_k is the Lagrange multiplier that makes $\sum_{i=1}^N f_{k,i}^2 = P_k$. Notice that this iterative water-filling algorithm is guaranteed to converge, as discussed in [31].

2) *Capacity Maximization*: As proved in Appendix A, the objective function for capacity maximization is given by $g[\gamma(\{\mathbf{F}_k\}_{k=1}^K)] = \sum_{i=1}^N \log(1 - \gamma_i(\{\mathbf{F}_k\}_{k=1}^K))$, based on which the optimization problem (36) can be written as

$$\begin{aligned} \min_{f_{k,i}} \quad & \sum_{i=1}^N \log \left(1 - \frac{\prod_{k=1}^K f_{k,i}^2 h_{k,i}^2}{\prod_{k=1}^K (f_{k,i}^2 h_{k,i}^2 + 1)} \right) \\ \text{s.t.} \quad & \sum_{i=1}^N f_{k,i}^2 = P_k. \end{aligned} \quad (39)$$

Similarly, the iterative water-filling algorithm can be used to solve for $f_{k,i}$ with guaranteed convergence. More specifically, when the $f_{l,i}$'s are given with $l \neq k$, the solution for $f_{k,i}$ can be derived as

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

with $a_{k,i} = \prod_{l \neq k} f_{l,i}^2 h_{l,i}^2 / (f_{l,i}^2 h_{l,i}^2 + 1)$ (40)

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

3) *MAX-MSE Minimization*: MAX-MSE is in fact a special case of **Obj 3** and in this case, $\psi_1(d(\Phi_{\text{MSE}}(\{\mathbf{P}_k\}_{k=1}^K))) = \max[\Phi_{\text{MSE}}(\{\mathbf{P}_k\}_{k=1}^K)]_{i,i}$. As shown in Appendix A, $g(\lambda(\{\mathbf{F}_k\}_{k=1}^K)) = \psi_1[\mathbf{1}_N - (\sum_{i=1}^N \lambda_i(\{\mathbf{F}_k\}_{k=1}^K)/N) \otimes \mathbf{1}_N]$. It follows that $g[\gamma(\{\mathbf{F}_k\}_{k=1}^K)]$ equals

$$\begin{aligned} g[\gamma(\{\mathbf{F}_k\}_{k=1}^K)] &= \max \left\{ \mathbf{1}_N - \left(\sum_{i=1}^N \gamma_i(\{\mathbf{F}_k\}_{k=1}^K)/N \right) \otimes \mathbf{1}_N \right\} \\ &= 1 - \frac{1}{N} \sum_{i=1}^N \gamma_i(\{\mathbf{F}_k\}_{k=1}^K). \end{aligned} \quad (41)$$

Clearly this expression for $g(\bullet)$ is similar to that for weighted MSE minimization. The optimal solution for $f_{k,i}$ can then be easily found as (38) with $w_i = 1$.

4) *Weighted Sum-Rate Maximization*: Under weighted sum-rate maximization, the objective function **Obj 4** can be further specified as

$$\psi_2(d(\Phi_{\text{MMSE}}(\{\mathbf{P}_k\}_{k=1}^K))) = \sum_{i=1}^N v_i \log(d_{[N-i+1]}(\Phi_{\text{MMSE}}(\{\mathbf{P}_k\}_{k=1}^K)))$$

where v_i is the i^{th} largest positive weighting factor and $d_{[i]}(\Phi_{\text{MMSE}}(\{\mathbf{P}_k\}_{k=1}^K))$ is the i^{th} largest diagonal element. Roughly speaking, this design scheme exhibits preference for data streams with better channel state information. It is proved in Appendix A that for this objective, $\mathbf{g}(\boldsymbol{\lambda}(\{\mathbf{F}_k\}_{k=1}^K)) = \psi_2[\mathbf{1}_N - \boldsymbol{\lambda}(\{\mathbf{F}_k\}_{k=1}^K)]$. It follows that

$$\mathbf{g}[\boldsymbol{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)] = \sum_{i=1}^N v_i \log(1 - \gamma_i(\{\mathbf{F}_k\}_{k=1}^K)) \quad (42)$$

and the optimization problem is formulated as

$$\begin{aligned} \min_{f_{k,i}} \quad & \sum_{i=1}^N v_i \log \left(1 - \frac{\prod_{k=1}^K f_{k,i}^2 h_{k,i}^2}{\prod_{k=1}^K (f_{k,i}^2 h_{k,i}^2 + 1)} \right) \\ \text{s.t.} \quad & \sum_{i=1}^N f_{k,i}^2 = P_k. \end{aligned} \quad (43)$$

The optimization problem in (43) has a similar form to that in (39), except that there are a number of weighting factors v_i in the objective function of (43). Therefore, the iterative water-filling solution of $f_{k,i}$ can be obtained similarly to that in (40) but with μ_k replaced by μ_k/v_k .

D. Relationship with Existing Solutions

By comparing our proposed optimal solution given by **Property 4** with existing solutions for various systems in the literature, we find that our proposed solution reduces to the following existing solutions by setting some system parameters accordingly:

- the robust design with weighted MSE minimization for a dual-hop AF MIMO relaying system without source precoder in [11], by setting $K = 2$, $\boldsymbol{\Sigma}_2 \propto \mathbf{I}_{M_2}$, and $\mathbf{P}_1 = \mathbf{I}_N$;
- the robust design for a dual-hop AF MIMO relaying system in [12], by setting $K = 2$, $\boldsymbol{\Psi}_2 \propto \mathbf{I}_{N_2}$, and $\mathbf{P}_1 = \mathbf{I}_N$;
- the transceiver design with weighted MSE minimization for a dual hop system with perfect CSI in [5], by setting $K = 2$, $\boldsymbol{\Psi}_k = \boldsymbol{\Sigma}_k = \mathbf{0}$, $\mathbf{W} = \mathbf{I}_N$ and $\mathbf{P}_1 = \mathbf{I}_N$;
- the transceiver design for a dual hop system with perfect CSI in [8], by setting $K = 2$, $\boldsymbol{\Psi}_k = \boldsymbol{\Sigma}_k = \mathbf{0}$ and $\mathbf{W} = \mathbf{I}_N$;

- the transceiver design with capacity maximization for a dual hop system with perfect CSI in [4], by setting $K = 2$, $\Psi_k = \Sigma_k = \mathbf{0}$ and $\mathbf{P}_1 = \mathbf{I}_N$;
- the robust design with weighted MSE minimization for a point-to-point MIMO system in [17], by setting $K = 1$; and
- the robust design with capacity maximization for a point-to-point MIMO system [18], by setting $K = 1$.

In other words, our proposed solution covers the above designs as special cases. It further verifies the correctness and optimality of our proposed solution.

E. Discussions

The optimal structure of \mathbf{F}_k in (34) is derived under the condition $\Psi_k \propto \mathbf{I}$ or $\Sigma_k \propto \mathbf{I}$. This condition can be easily satisfied in practice. We notice that the expressions for Ψ_k and Σ_k generally depend on specific channel estimation algorithms. Denote the transmit and receive antenna correlation matrices and the channel estimation error variance in the k^{th} hop as $\mathbf{R}_{T,k}$, $\mathbf{R}_{R,k}$ and $\sigma_{e,k}^2$, respectively. Applying the widely used channel estimation algorithms in [17] and [18], the covariance matrices for channel estimation errors can be written as $\Psi_k = \mathbf{R}_{T,k}$ and $\Sigma_k = \sigma_{e,k}^2(\mathbf{I}_{M_k} + \sigma_{e,k}^2 \mathbf{R}_{R,k}^{-1})^{-1}$. Clearly, when the receive antennas are spaced widely, i.e., $\mathbf{R}_{R,k} \propto \mathbf{I}_{M_k}$, we directly have $\Sigma_k \propto \mathbf{I}_{M_k}$. Moreover, when the length of training is long, the value of $\sigma_{e,k}^2$ will be small and $\mathbf{I}_{M_k} + \sigma_{e,k}^2 \mathbf{R}_{R,k}^{-1} \approx \mathbf{I}_{M_k}$. As a result, Σ_k can be approximated as an identity matrix even when $\mathbf{R}_{R,k} \not\propto \mathbf{I}_{M_k}$. Furthermore, when the channel statistics are unknown and the least-squares channel estimator is applied, it can be derived that $\Sigma_k \propto \mathbf{I}_{M_k}$ [11]. On the other hand, when the transmit antennas are spaced widely, i.e., $\mathbf{R}_{T,k} \propto \mathbf{I}_{N_k}$, we can obtain $\Psi_k \propto \mathbf{I}_{N_k}$.

For the general case when $\Psi_k \not\propto \mathbf{I}_{N_k}$ and $\Sigma_k \not\propto \mathbf{I}_{M_k}$, to the best of our knowledge, finding a closed-form optimal solution of the robust design problem is still open, even for point-to-point MIMO systems [17], [20]. The main difficulty comes from the fact that when $\Psi_k \not\propto \mathbf{I}_{N_k}$ and $\Sigma_k \not\propto \mathbf{I}_{M_k}$, $\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k}$ varies with \mathbf{F}_k , and so is not a constant. However, for this general case, $\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k}$ in (33) can be replaced by

$$\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k} \approx P_k \lambda_1(\Psi_k) / (P_k \lambda_1(\Psi_k) \alpha_k + \sigma_{n_k}^2) \Sigma_k + \sigma_{n_k}^2 / (P_k \lambda_1(\Psi_k) \alpha_k + \sigma_{n_k}^2) \mathbf{I}_{M_k}, \quad (44)$$

such that it is not a function of \mathbf{F}_k . Notice that the above inequality becomes an equality when $\Psi_k \propto \mathbf{I}_{N_k}$ or $\Sigma_k \propto \mathbf{I}_{M_k}$. Then the proposed solution can still be applied for this general case.

When there are two hops ($K = 2$), our proposed optimal structure is different from that derived in [14] (comparing (34) with Equation (16) in [14]). In [14], the solution structure is obtained by implicitly approximating a design-variable-dependent covariance matrix (the matrix \mathbf{A} in [14]) as constant. Since the approximation is tight only when the covariance matrix of channel estimation errors seen from the receiver side is proportional to an identity matrix, i.e., $\mathbf{\Psi}_k \propto \mathbf{I}_{N_k}$, $k = 1, 2$, the proposed solution in [14] is sub-optimal when $\mathbf{\Psi}_k \not\propto \mathbf{I}_{N_k}$, $k = 1, 2$. Using the duality between the uplink and downlink discovered in our paper, the optimal solution for the case of $\mathbf{\Sigma}_k \propto \mathbf{I}_{N_k}$, $k = 1, 2$, can also be obtained based on the approach of [14]. Unfortunately, this result has not been discovered in [14]. Generally speaking, with the duality, the work of [14] can lead to the optimal solution when $\mathbf{\Sigma}_k \propto \mathbf{I}_{N_k}$, $k = 1, 2$ or $\mathbf{\Psi}_k \propto \mathbf{I}_{N_k}$, $k = 1, 2$. However, even for dual-hop systems, our proposed solution is optimal under a wider range of cases than that in [14], since it is optimal when either $\mathbf{\Sigma}_k$ or $\mathbf{\Psi}_k$, $\forall k$, is proportional to an identity matrix³. Notice that the condition of ($\mathbf{\Sigma}_k \propto \mathbf{I}_{N_k}$ or $\mathbf{\Psi}_k \propto \mathbf{I}_{N_k}$), $\forall k$ is more relaxed than that of ($\mathbf{\Sigma}_k \propto \mathbf{I}_{N_k}$, $\forall k$ or $\mathbf{\Psi}_k \propto \mathbf{I}_{N_k}$, $\forall k$). In conclusion, our solution is more general than that in [14].

With respect to the complexity, it is clear from (34) that the complexity of our algorithm is due to two kinds of operations, i.e., the iterative water-filling computation for the inner diagonal matrix in (34) and the decomposition/multiplication for the matrices on the lefthand and righthand sides of the diagonal matrix in (34). Comparing the structures of the solution in (34) and that in [14], similar operations are needed to obtain the solution in [14]. So we can expect that the complexity of our approach is comparable to that in [14].

V. SIMULATION RESULTS

In this section, the performance of the proposed robust designs is evaluated by simulations. In the simulations, the number of antennas at each node is set to four. At the source node, four independent data streams are transmitted and in each data stream, $N_{Data} = 10^4$ independent quadrature phase shifting keying (QPSK) symbols are transmitted. The correlation matrices corresponding to the channel estimation errors are chosen according to the widely used exponential model, i.e., $[\mathbf{\Psi}_k]_{i,j} = \sigma_e^2 \alpha^{|i-j|}$ and $[\mathbf{\Sigma}_k]_{i,j} = \beta^{|i-j|}$, where α and β are

³It is possible to extend the work of [14] to multi-hop systems. With the duality found in our paper, the extension would lead to the optimal solution for the cases of ($\mathbf{\Sigma}_k \propto \mathbf{I}_{N_k}$, $\forall k$ or $\mathbf{\Psi}_k \propto \mathbf{I}_{N_k}$, $\forall k$). However, the extension is by no means straightforward as shown in [10].

the correlation coefficients, and σ_e^2 denotes the variance of the channel estimation error [12], [20]. The estimated channel matrices $\bar{\mathbf{H}}_k$'s, are generated following the widely used complex Gaussian distributions, $\bar{\mathbf{H}}_k \sim \mathcal{CN}_{M_k, N_k}(\mathbf{0}_{M_k, N_k}, (1 - \sigma_e^2)/\sigma_e^2 \boldsymbol{\Sigma}_k \otimes \boldsymbol{\Psi}_k^T)$ [12], [28], such that channel realizations $\mathbf{H}_k = \bar{\mathbf{H}}_k + \Delta \mathbf{H}_k$ have unit variance. The signal-to-noise ratio (SNR) for the k^{th} link is defined as $P_k/\sigma_{n_k}^2$ and each point in the following figures shows an average result of 10^4 trials.

A dual hop system ($K = 2$) with error correlation coefficients of $\alpha = 0.6$ and $\beta = 0$ (i.e., $\boldsymbol{\Psi}_k \not\propto \mathbf{I}, \boldsymbol{\Sigma}_k \propto \mathbf{I}, k = 1, 2$) is considered first. Fig. 2 shows the weighted MSE at the destination when the weighting matrix is arbitrarily chosen as $\mathbf{W} = \text{diag}\{[0.3 \ 0.3 \ 0.26 \ 0.26]\}$ and $P_k/\sigma_{n_k}^2 = 30\text{dB}$. For comparison, the performance of the algorithm based on the estimated channel only (labeled as non-robust design) [8], the robust algorithm proposed by Rong in [14] and the robust algorithm without source precoding in [12] is also shown. It is clear from the figure that our proposed robust design offers the best performance, while the non-robust design is the worst. Fig. 3 shows the sum-rates of various algorithms for the considered two-hop AF MIMO relaying system. It can be seen that the robust algorithms generally have better performance than the algorithm based on estimated CSI only. Furthermore, the performance of the proposed robust design is much better than that of the robust algorithm in [14].

Next a three-hop AF MIMO relaying system, i.e., $K = 3$, is considered to further investigate the effectiveness of the proposed robust design. Since there are few (if any) robust transceiver design algorithms proposed for multi-hop AF MIMO systems in the literature, our proposed robust design is mainly compared with the non-robust design in [10] in the following. With the weighting matrix being arbitrarily selected as $\mathbf{W} = \text{diag}\{[0.26 \ 0.25 \ 0.25 \ 0.24]\}$, Fig. 4 shows the weighted MSE at the destination when $P_k/\sigma_{n_k}^2 = 30\text{dB}$. Here two sets of error correlation coefficients, $(\alpha = 0.6, \beta = 0)$, and $(\alpha = 0, \beta = 0.6)$, are taken as examples. They correspond to the cases of $(\boldsymbol{\Psi}_k \not\propto \mathbf{I}, \boldsymbol{\Sigma}_k \propto \mathbf{I})$ and $(\boldsymbol{\Psi}_k \propto \mathbf{I}, \boldsymbol{\Sigma}_k \not\propto \mathbf{I})$, respectively. It can be seen that the proposed algorithm shows similar performance for the two cases and always outperforms the non-robust design based on the estimated CSI only. When there is no channel estimation error, i.e., $\sigma_e^2 = 0$, the performance of the two algorithms is the same as expected.

Fig. 5 shows the sum-rates at different SNRs ($\text{SNR} = P_k/\sigma_{n_k}^2$) for the three-hop system. The SNRs at various hops are set as the same for simplicity. The correlation coefficients for the channel estimation errors are taken as $\alpha = 0.6$ and $\beta = 0$. It is further demonstrated that the

proposed algorithm shows better performance than the non-robust algorithm based on estimated CSI only. Furthermore, as the estimation errors increase, the performance gap between the two algorithms enlarges. This result coincides with that for the weighted-MSE-based robust design shown in Fig. 4. The performance of the maximum MSE across four data streams with $\alpha = 0$ and $\beta = 0.6$ is then shown in Fig. 6. Similarly, it is observed that the performance gain of the proposed robust design over the non-robust design with estimated CSI only becomes larger as SNR increases. The performance gap is also more apparent when σ_e^2 increases.

Finally, Fig. 7 shows the bit-error-rate (BER) performance for the three-hop systems with different design criteria: capacity maximization, sum MSE minimization (i.e., weighted MSE minimization with $\mathbf{W} = \mathbf{I}$) and MAX-MSE minimization. The parameters are chosen as $\alpha = 0.6$, $\beta = 0$ and $\sigma_e^2 = 0.004$. It can be seen that in terms of BER performance, the former two criteria perform worse than the latter one since the latter criterion targets the BER performance more. Moreover, the non-robust design with capacity maximization based on estimated CSI only is also given and the results further verify the performance advantage of the proposed robust designs over the non-robust design with estimated CSI only.

VI. CONCLUSIONS

Bayesian robust transceiver design for multi-hop AF MIMO relaying systems with channel estimation errors has been considered. Various transceiver design criteria including weighted MSE minimization, capacity maximization, worst MSE minimization and weighted sum-rate maximization have been discussed and formulated into a unified optimization problem. Using majorization theory and properties of matrix-variate functions, the optimal structure of the robust transceivers has been derived. Then the transceiver design problems have been greatly simplified and solved by iterative water-filling algorithm. The performance of the proposed transceiver designs has been demonstrated via simulation results.

APPENDIX A

PROOF OF PROPERTY 1

The proof of Property 1 depends on the specific objective function in (20). In the following, we will discuss the optimization problem (20) with different objective functions case by case.

Obj 1: With the objective function of (6) and the MMSE matrix in (18), we have

$$f(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0) = \text{Tr}(\mathbf{W}) - \text{Tr}(\mathbf{W} \mathbf{Q}_0^H \Theta \mathbf{Q}_0) \geq \underbrace{\text{Tr}(\mathbf{W}) - \sum_{i=1}^N \lambda_i(\mathbf{W}) \lambda_i(\Theta)}_{\triangleq g(\lambda(\Theta))} \quad (45)$$

where the inequality follows from the fact that for two positive semi-definite matrices \mathbf{W} and Θ , $\text{Tr}(\mathbf{W} \mathbf{Q}_0^H \Theta \mathbf{Q}_0) \leq \sum_i \lambda_i(\mathbf{W}) \lambda_i(\mathbf{Q}_0^H \Theta \mathbf{Q}_0)$ with $\lambda_i(\mathbf{Z})$ denoting the i^{th} largest eigenvalue of \mathbf{Z} . Furthermore, the second equality in (45) holds when $\mathbf{Q}_0^H \Theta \mathbf{Q}_0 = \mathbf{U}_W \text{diag}(\lambda(\Theta)) \mathbf{U}_W^H$ where $\lambda(\Theta) = [\lambda_1(\Theta), \dots, \lambda_N(\Theta)]^T$ and \mathbf{U}_W is the unitary matrix containing the eigenvectors of \mathbf{W} as columns [32]. It implies that the optimal value of $f(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0)$ is $g(\lambda(\Theta))$ and is achieved when $\mathbf{Q}_0^H \Theta \mathbf{Q}_0 = \mathbf{U}_W \text{diag}(\lambda(\Theta)) \mathbf{U}_W^H$.

Using the Lemma 2.A.2 [32] and the definition of $g(\lambda(\Theta))$ in (45), it can be easily found that $g(\lambda(\Theta))$ is a Schur-concave function with respect to $\lambda(\Theta)$. Furthermore, for two vectors $\mathbf{v} \leq \mathbf{u}$ (i.e., $v_i \leq u_i$), from the definition of $g(\lambda(\Theta))$ in (45), it can be concluded that $g(\mathbf{v}) \geq g(\mathbf{u})$. It means that $g(\bullet)$ is a decreasing function.

Obj 2: For the second objective function given by (8), it is directly obtained that

$$f(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0) = \log|\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0| = \underbrace{\sum_{i=1}^N \log[1 - \lambda_i(\Theta)]}_{\triangleq g(\lambda(\Theta))}. \quad (46)$$

Obviously, the above equality holds unconditionally and thus the objective function $f(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0)$ is independent of \mathbf{Q}_0 . It follows from the optimization problem (20) that \mathbf{Q}_0 can take any arbitrary unitary matrix since it is only involved in the constraint of $\mathbf{Q}_0^H \mathbf{Q}_0 = \mathbf{I}$. Therefore, $\mathbf{Q}_0^H \Theta \mathbf{Q}_0 = \mathbf{U}_{\text{Arb}} \text{diag}(\lambda(\Theta)) \mathbf{U}_{\text{Arb}}^H$ with \mathbf{U}_{Arb} being an arbitrary unitary matrix always holds. Meanwhile, the optimal value of $f(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0)$ can always be written as $g(\lambda(\Theta))$. Based on the Lemma 2.A.2 [32] and the definition of $g(\lambda(\Theta))$ in (46), it can also be proved that $g(\lambda(\Theta))$ is a decreasing Schur-concave function with respect to $\lambda(\Theta)$.

Obj 3: For the diagonal elements of the positive semi-definite matrix $\Phi_{\text{MMSE}}(\{\mathbf{P}_k\}_{k=1}^K) = \mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0$, we have the following majorization relationship [32]:

$$d(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0) \succ \mathbf{1}_N - \left(\sum_{i=1}^N \lambda_i(\Theta) / N \right) \otimes \mathbf{1}_N \quad (47)$$

where the equality holds if and only if $[\mathbf{Q}_0^H \Theta \mathbf{Q}_0]_{i,i} = \sum_{i=1}^N \lambda_i(\Theta) / N$, and $\mathbf{1}_N$ is the $N \times 1$ all-one vector.

For the third objective function in (9), as $\psi_1(\bullet)$ is increasing and Schur-convex, the objective function in (20) satisfies [24]

$$\mathbf{f}(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0) = \psi_1(d(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0)) \geq \underbrace{\psi_1\left(\mathbf{1}_N - \left(\sum_{i=1}^N \lambda_i(\Theta)/N\right) \otimes \mathbf{1}_N\right)}_{\triangleq g[\lambda(\Theta)]}, \quad (48)$$

with equality if and only if $[\mathbf{Q}_0^H \Theta \mathbf{Q}_0]_{i,i} = \sum_{i=1}^N \lambda_i(\Theta)/N$. As shown in [24], when $\mathbf{Q}_0^H \Theta \mathbf{Q}_0 = \mathbf{Q}_{\text{DFT}} \text{diag}(\lambda(\Theta)) \mathbf{Q}_{\text{DFT}}^H$ where \mathbf{Q}_{DFT} is a DFT matrix, $\mathbf{Q}_0^H \Theta \mathbf{Q}_0$ has identical diagonal elements. It follows that when $\mathbf{Q}_0^H \Theta \mathbf{Q}_0 = \mathbf{Q}_{\text{DFT}} \text{diag}(\lambda(\Theta)) \mathbf{Q}_{\text{DFT}}^H$, the objective function $\mathbf{f}(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0)$ will take minimum/optimal value of $g(\lambda(\Theta))$.

Based on the fact that $\psi_1(\bullet)$ is an increasing and Schur-convex function, it can be directly concluded from (48) that $g(\lambda(\Theta))$ is a decreasing function of $\lambda(\Theta)$. Furthermore, based on the Lemma 2.A.2 [32], $\psi_1(\bullet)$ is also a Schur-concave function of $\lambda(\Theta)$.

Obj 4: Notice that for the positive semi-definite matrix $\Phi_{\text{MMSE}}(\{\mathbf{P}_k\}_{k=1}^K) = \mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0$, $d(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0) \prec \lambda(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0)$ [24]. With the Schur-concave function of $\psi_2(\bullet)$ in (11), we have

$$\mathbf{f}(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0) = \psi_2(d(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0)) \geq \underbrace{\psi_2([\mathbf{1}_N - \lambda(\Theta)])}_{\triangleq g[\lambda(\Theta)]}, \quad (49)$$

where the equality holds when $[\mathbf{Q}_0^H \Theta \mathbf{Q}_0]_{i,i} = \lambda_i(\Theta)$. It is easy to see that when $\mathbf{Q}_0^H \Theta \mathbf{Q}_0 = \mathbf{I}_N \text{diag}(\lambda(\Theta)) \mathbf{I}_N$, the preceding condition is satisfied and then the objective function $\mathbf{f}(\mathbf{I}_N - \mathbf{Q}_0^H \Theta \mathbf{Q}_0)$ achieves its minimum/optimal value of $g(\lambda(\Theta))$.

Since $\psi_2(\bullet)$ is increasing and Schur-concave, it is clear that $g(\lambda(\Theta))$ is decreasing with respect to $\lambda(\Theta)$. Moreover, using [32, 3.A.6.a], it can be proved that $\psi_2(\mathbf{1}_N - \lambda(\Theta))$ is also Schur-concave with respect to $\lambda(\Theta)$.

APPENDIX B

PROOF OF PROPERTY 2

First notice that for two matrices \mathbf{A} and \mathbf{B} with compatible dimensions, $\lambda_i(\mathbf{A}\mathbf{B}) = \lambda_i(\mathbf{B}\mathbf{A})$ [32, 9.A.1.a]. Together with the fact that for two positive semi-definite matrices \mathbf{A} and \mathbf{B} ,

$\prod_{i=1}^k \lambda_i(\mathbf{A}\mathbf{B}) \leq \prod_{i=1}^k \lambda_i(\mathbf{A})\lambda_i(\mathbf{B})$ [32, 9.H.1.a], we have⁴

$$\begin{aligned} & \prod_{i=1}^k \lambda_i(\mathbf{A}_1^H \mathbf{Q}_1^H \cdots \mathbf{A}_K^H \mathbf{Q}_K^H \mathbf{Q}_K \mathbf{A}_K \cdots \mathbf{Q}_1 \mathbf{A}_1) \\ & \leq \prod_{i=1}^k \lambda_i(\mathbf{A}_2^H \mathbf{Q}_2^H \cdots \mathbf{A}_K^H \mathbf{Q}_K^H \mathbf{Q}_K \mathbf{A}_K \cdots \mathbf{A}_2 \mathbf{Q}_2) \underbrace{\lambda_i(\mathbf{Q}_1 \mathbf{A}_1 \mathbf{A}_1^H \mathbf{Q}_1^H)}_{=\lambda_i(\mathbf{A}_1 \mathbf{A}_1^H)} \quad k = 1, \dots, N. \end{aligned} \quad (50)$$

Repeating this process, we have the following inequality:

$$\prod_{i=1}^k \lambda_i(\mathbf{A}_1^H \mathbf{Q}_1^H \cdots \mathbf{A}_K^H \mathbf{Q}_K^H \mathbf{Q}_K \mathbf{A}_K \cdots \mathbf{Q}_1 \mathbf{A}_1) \leq \prod_{i=1}^k \underbrace{\lambda_i(\mathbf{A}_K^H \mathbf{A}_K) \lambda_i(\mathbf{A}_{K-1}^H \mathbf{A}_{K-1}) \cdots \lambda_i(\mathbf{A}_1^H \mathbf{A}_1)}_{\triangleq \gamma_i(\{\mathbf{F}_k\}_{k=1}^K)}. \quad (51)$$

Based on (51) and **5.A.2.b** in [32], we directly have

$$\lambda(\mathbf{A}_1^H \mathbf{Q}_1^H \cdots \mathbf{A}_K^H \mathbf{Q}_K^H \mathbf{Q}_K \mathbf{A}_K \cdots \mathbf{Q}_1 \mathbf{A}_1) \prec_w [\gamma_1(\{\mathbf{F}_k\}_{k=1}^K) \cdots \gamma_N(\{\mathbf{F}_k\}_{k=1}^K)]^T \triangleq \gamma(\{\mathbf{F}_k\}_{k=1}^K) \quad (52)$$

where $\mathbf{a} \prec_w \mathbf{b}$ denotes that \mathbf{a} is weakly majorized by \mathbf{b} [32] and the equality holds if and only if the neighboring \mathbf{A}_k 's satisfy

$$\mathbf{Q}_k = \mathbf{V}_{\mathbf{A}_{k+1}} \mathbf{U}_{\mathbf{A}_k}^H, \quad k = 1, \dots, K-1 \quad (53)$$

where $\mathbf{U}_{\mathbf{A}_k}$ and $\mathbf{V}_{\mathbf{A}_k}$ are defined based on the following singular value decomposition: $\mathbf{A}_k = \mathbf{U}_{\mathbf{A}_k} \boldsymbol{\Lambda}_{\mathbf{A}_k} \mathbf{V}_{\mathbf{A}_k}^H$ with $\boldsymbol{\Lambda}_{\mathbf{A}_k} \searrow$. As $g(\bullet)$ is a decreasing and Schur-concave function, we have [32]

$$g[\lambda(\boldsymbol{\Theta})] \geq g[\gamma(\{\mathbf{F}_k\}_{k=1}^K)] \quad (54)$$

with equality if and only if (53) holds. Finally, based on the definition of \mathbf{A}_k in (18), using the matrix inversion lemma, the following equality holds:

$$\mathbf{A}_k^H \mathbf{A}_k = \mathbf{I}_{M_k} - (\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k + \mathbf{I}_{M_k})^{-1}. \quad (55)$$

It follows that $\lambda_i(\mathbf{A}_k^H \mathbf{A}_k) = \lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k) / [1 + \lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)]$. Based on this result, $\gamma_i(\{\mathbf{F}_k\}_{k=1}^K)$ in (51) equals

$$\gamma_i(\{\mathbf{F}_k\}_{k=1}^K) = \prod_{k=1}^K \frac{\lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}{1 + \lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}. \quad (56)$$

⁴Note that in general \mathbf{A}_k is not a square matrix.

APPENDIX C

PROOF OF PROPERTY 3

Suppose at the moment that using the optimal \mathbf{F}_k , denoted by $\mathbf{F}_{k,\text{opt}}$, transmission is not at the maximum power, i.e., $\text{Tr}(\mathbf{F}_{k,\text{opt}}\mathbf{F}_{k,\text{opt}}^H) < P_k$; then we have $a \triangleq \sqrt{P_k/\text{Tr}(\mathbf{F}_{k,\text{opt}}\mathbf{F}_{k,\text{opt}}^H)} > 1$. Defining $\hat{\mathbf{F}}_k \triangleq a\mathbf{F}_{k,\text{opt}}$, it follows that

$$\begin{aligned} & \hat{\mathbf{F}}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\hat{\mathbf{F}}_k}^{-1} \bar{\mathbf{H}}_k \hat{\mathbf{F}}_k \\ &= \mathbf{F}_{k,\text{opt}}^H \bar{\mathbf{H}}_k^H (\text{Tr}(\mathbf{F}_{k,\text{opt}}\mathbf{F}_{k,\text{opt}}^H \boldsymbol{\Psi}_k) \boldsymbol{\Sigma}_k + \sigma_{n_k}^2/a^2 \mathbf{I})^{-1} \bar{\mathbf{H}}_k \mathbf{F}_{k,\text{opt}} \\ &\succeq \mathbf{F}_{k,\text{opt}}^H \bar{\mathbf{H}}_k^H \underbrace{(\text{Tr}(\mathbf{F}_{k,\text{opt}}\mathbf{F}_{k,\text{opt}}^H \boldsymbol{\Psi}_k) \boldsymbol{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I})^{-1}}_{\mathbf{K}_{\mathbf{F}_{k,\text{opt}}}^{-1}} \bar{\mathbf{H}}_k \mathbf{F}_{k,\text{opt}}. \end{aligned} \quad (57)$$

Note that $\mathbf{A} \succeq \mathbf{B}$ means that $\lambda_i(\mathbf{A}) \geq \lambda_i(\mathbf{B})$ for all i , and therefore (57) implies

$$\lambda_i(\hat{\mathbf{F}}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\hat{\mathbf{F}}_k}^{-1} \bar{\mathbf{H}}_k \hat{\mathbf{F}}_k) \geq \lambda_i(\mathbf{F}_{k,\text{opt}}^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_{k,\text{opt}}}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_{k,\text{opt}}). \quad (58)$$

Moreover, it is clear from the definition of $\gamma_i(\{\mathbf{F}_k\}_{k=1}^K)$ in (27) that $\gamma_i(\{\mathbf{F}_k\}_{k=1}^K)$ is an increasing function of $\lambda_i(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)$. It then follows that $\gamma(\{\mathbf{F}_k = \hat{\mathbf{F}}_k\}_{k=1}^K) \geq \gamma(\{\mathbf{F}_k = \hat{\mathbf{F}}_{k,\text{opt}}\}_{k=1}^K)$. Together with the fact that $\mathbf{g}(\bullet)$ is a decreasing function, it is concluded that $\mathbf{g}[\gamma(\{\mathbf{F}_k = \hat{\mathbf{F}}_k\}_{k=1}^K)] \leq \mathbf{g}[\gamma(\{\mathbf{F}_k = \hat{\mathbf{F}}_{k,\text{opt}}\}_{k=1}^K)]$. It is obvious that this result contradicts the optimality of $\mathbf{F}_{k,\text{opt}}$, and therefore a necessary condition for the optimal \mathbf{F}_k is $\text{Tr}(\mathbf{F}_k\mathbf{F}_k^H) = P_k$. Furthermore, when $\text{Tr}(\mathbf{F}_k\mathbf{F}_k^H) = P_k$, the following equality holds:

$$\begin{aligned} \text{Tr}[\mathbf{F}_k\mathbf{F}_k^H(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I})] &= \alpha_k P_k \text{Tr}(\mathbf{F}_k\mathbf{F}_k^H \boldsymbol{\Psi}_k) + \sigma_{n_k}^2 \underbrace{\text{Tr}(\mathbf{F}_k\mathbf{F}_k^H)}_{=P_k} \\ &= \alpha_k P_k \text{Tr}(\mathbf{F}_k\mathbf{F}_k^H \boldsymbol{\Psi}_k) + \sigma_{n_k}^2 P_k. \end{aligned} \quad (59)$$

Defining $\eta_{f_k} = \alpha_k \text{Tr}(\mathbf{F}_k\mathbf{F}_k^H \boldsymbol{\Psi}_k) + \sigma_{n_k}^2$ with $\alpha_k = \text{Tr}(\boldsymbol{\Sigma}_k)/M_k$, (59) can be rewritten as $\text{Tr}[\mathbf{F}_k\mathbf{F}_k^H(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I})] = P_k \eta_{f_k}$. In other words, the power constraint $\text{Tr}(\mathbf{F}_k\mathbf{F}_k^H) = P_k$ is equivalent to

$$\text{Tr}[\mathbf{F}_k\mathbf{F}_k^H(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I})]/\eta_{f_k} = P_k. \quad (60)$$

APPENDIX D

PROOF OF PROPERTY 4

Problem reformulation: As shown in (27), $\gamma_i(\{\mathbf{F}_k\}_{k=1}^K)$ is a complicated function of $\boldsymbol{\lambda}(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)$. Clearly, \mathbf{F}_k appears in multiple positions. In particular, $\mathbf{K}_{\mathbf{F}_k}$ is a function of \mathbf{F}_k which complicates

the derivation of optimal solutions. In order to simplify the problem, $\lambda(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)$ is reformulated as

$$\begin{aligned} \lambda(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k) &= \lambda[\tilde{\mathbf{F}}_k^H (\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} \bar{\mathbf{H}}_k^H (\mathbf{K}_{\mathbf{F}_k} / \eta_{f_k})^{-1/2} \\ &\quad \times \underbrace{(\mathbf{K}_{\mathbf{F}_k} / \eta_{f_k})^{-1/2} \bar{\mathbf{H}}_k (\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} \tilde{\mathbf{F}}_k}_{\triangleq \mathcal{H}_k}], \end{aligned} \quad (61)$$

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

$$\tilde{\mathbf{F}}_k = 1/\sqrt{\eta_{f_k}} (\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{1/2} \mathbf{F}_k. \quad (62)$$

The right hand side of (61) is easier to handle than the left hand side. This is because when $\boldsymbol{\Psi}_k \propto \mathbf{I}_{N_k}$ or $\boldsymbol{\Sigma}_k \propto \mathbf{I}_{M_k}$, \mathcal{H}_k is independent of $\tilde{\mathbf{F}}_k$. In the following, we will prove this in detail.

It is obvious that \mathcal{H}_k being independent of $\tilde{\mathbf{F}}_k$ is equivalent to $\mathbf{K}_{\mathbf{F}_k} / \eta_{f_k}$ being independent of $\tilde{\mathbf{F}}_k$. First consider $\boldsymbol{\Psi}_k \propto \mathbf{I}_{N_k}$, i.e., $\boldsymbol{\Psi}_k = \beta_k \mathbf{I}_{N_k}$. With the definitions of $\mathbf{K}_{\mathbf{F}_k}$ in (17) and η_{f_k} , $\mathbf{K}_{\mathbf{F}_k} / \eta_{f_k}$ equals

$$\begin{aligned} \mathbf{K}_{\mathbf{F}_k} / \eta_{f_k} &= [\beta_k \text{Tr}(\mathbf{F}_k \mathbf{F}_k^H) \boldsymbol{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I}_{M_k}] / [\beta_k \alpha_k \text{Tr}(\mathbf{F}_k \mathbf{F}_k^H) + \sigma_{n_k}^2] \\ &= (\beta_k P_k \boldsymbol{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I}_{M_k}) / (\alpha_k \beta_k P_k + \sigma_{n_k}^2), \end{aligned} \quad (63)$$

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

$$\begin{aligned} \mathbf{K}_{\mathbf{F}_k} / \eta_{f_k} &= [\alpha_k \text{Tr}(\mathbf{F}_k \mathbf{F}_k^H \boldsymbol{\Psi}_k) \mathbf{I}_{N_k} + \sigma_{n_k}^2 \mathbf{I}_{N_k}] / [\alpha_k [\text{Tr}(\mathbf{F}_k \mathbf{F}_k^H \boldsymbol{\Psi}_k) + \sigma_{n_k}^2]] \\ &= \mathbf{I}_{N_k}. \end{aligned} \quad (64)$$

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

Using the substitution (62), the optimization problem (32) is reformulated as

$$\begin{aligned} \min_{\tilde{\mathbf{F}}_k} \quad & g[\gamma(\{\tilde{\mathbf{F}}_k\}_{k=1}^K)] \\ \text{s.t.} \quad & \gamma_i(\{\tilde{\mathbf{F}}_k\}_{k=1}^K) = \prod_{k=1}^K \frac{\lambda_i(\tilde{\mathbf{F}}_k^H \mathcal{H}_k^H \mathcal{H}_k \tilde{\mathbf{F}}_k)}{1 + \lambda_i(\tilde{\mathbf{F}}_k^H \mathcal{H}_k^H \mathcal{H}_k \tilde{\mathbf{F}}_k)} \\ & \text{Tr}(\tilde{\mathbf{F}}_k \tilde{\mathbf{F}}_k^H) = P_k. \end{aligned} \quad (65)$$

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

$$\mathcal{H}_k \tilde{\mathbf{F}}_{k,\text{opt}} = \mathbf{U}_{M_k} \Lambda_{M_k} \mathbf{V}_{M_k}^H \quad \text{with } \Lambda_{M_k} \searrow \quad \text{and } \mathcal{H}_k = \mathbf{U}_{\mathcal{H}_k} \Lambda_{\mathcal{H}_k} \mathbf{V}_{\mathcal{H}_k}^H \quad \text{with } \Lambda_{\mathcal{H}_k} \searrow, \quad (66)$$

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

$$\bar{\mathbf{F}}_k = \mathbf{V}_{\mathcal{H}_k} \Lambda_{\mathbf{X}_k} \mathbf{V}_{M_k}^H, \quad (67)$$

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

Because \mathcal{H}_k is independent of the unknown variable $\tilde{\mathbf{F}}_k$, using Lemma 12 in [24], we have

$$\bar{\mathbf{F}}_k^H \mathcal{H}_k^H \mathcal{H}_k \bar{\mathbf{F}}_k \succeq \tilde{\mathbf{F}}_{k,\text{opt}}^H \mathcal{H}_k^H \mathcal{H}_k \tilde{\mathbf{F}}_{k,\text{opt}}. \quad (68)$$

Taking eigenvalues of both sides, we have $\lambda_i(\bar{\mathbf{F}}_k^H \mathcal{H}_k^H \mathcal{H}_k \bar{\mathbf{F}}_k) \geq \lambda_i(\tilde{\mathbf{F}}_{k,\text{opt}}^H \mathcal{H}_k^H \mathcal{H}_k \tilde{\mathbf{F}}_{k,\text{opt}})$ [32]. Since $\gamma_i(\{\tilde{\mathbf{F}}_k\}_{k=1}^K)$ is an increasing function of $\lambda_i(\tilde{\mathbf{F}}_k^H \mathcal{H}_k^H \mathcal{H}_k \tilde{\mathbf{F}}_k)$, we directly have that $\gamma(\{\tilde{\mathbf{F}}_k = \bar{\mathbf{F}}_k\}_{k=1}^K) \geq \gamma(\{\tilde{\mathbf{F}}_k = \tilde{\mathbf{F}}_{k,\text{opt}}\}_{k=1}^K)$. Furthermore, as the objective function $\mathbf{g}(\bullet)$ of (65) is a decreasing function, we finally have $\mathbf{g}[\gamma(\{\tilde{\mathbf{F}}_k = \bar{\mathbf{F}}_k\}_{k=1}^K)] \leq \mathbf{g}[\gamma(\{\tilde{\mathbf{F}}_k = \tilde{\mathbf{F}}_{k,\text{opt}}\}_{k=1}^K)]$. Because $\tilde{\mathbf{F}}_{k,\text{opt}}$ is the optimal solution, $\tilde{\mathbf{F}}_{k,\text{opt}}$ must be in the form of $\bar{\mathbf{F}}_k$. Therefore, the structure of optimal $\tilde{\mathbf{F}}_k$ is given by (67), i.e., $\tilde{\mathbf{F}}_{k,\text{opt}} = \mathbf{V}_{\mathcal{H}_k} \Lambda_{\mathbf{X}_k} \mathbf{V}_{M_k}^H$.

As the minimum dimension of \mathbf{A}_k is N , on substituting (67) into $\gamma_i(\{\mathbf{F}_k\}_{k=1}^K)$ (51), it can be seen that for the optimal solution only the $N \times N$ principal submatrix of $\Lambda_{\mathbf{X}_k}$ can be nonzero, which is denoted as $\Lambda_{\mathcal{F}_k}$. As a result, $\tilde{\mathbf{F}}_{k,\text{opt}}$ has the following structure:

$$\tilde{\mathbf{F}}_{k,\text{opt}} = \mathbf{V}_{\mathcal{H}_k, N} \Lambda_{\mathcal{F}_k} \mathbf{V}_{M_k, N}^H. \quad (69)$$

It is clear that the values of \mathbf{V}_{M_k} 's do not affect the values of $\lambda_i(\tilde{\mathbf{F}}_k^H \mathcal{H}_k^H \mathcal{H}_k \tilde{\mathbf{F}}_k)$, the constraint $\text{Tr}(\tilde{\mathbf{F}}_k \tilde{\mathbf{F}}_k^H) = P_k$ and the objective function in the optimization problem (65). Therefore, \mathbf{V}_{M_k} can be an arbitrary unitary matrix.

Structure of optimal \mathbf{F}_k : Based on the relationship between \mathbf{F}_k and $\tilde{\mathbf{F}}_k$ given in (62),

$$\mathbf{F}_{k,\text{opt}} = \sqrt{\eta_{f_k}} (\alpha_k P_k \Psi_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} \tilde{\mathbf{F}}_{k,\text{opt}}. \quad (70)$$

Putting the structure of $\mathbf{F}_{k,\text{opt}}$ in (70) into η_{f_k} in (31), η_{f_k} can be solved to be

$$\eta_{f_k} = \sigma_{n_k}^2 / \{1 - \alpha_k \text{Tr}[\mathbf{V}_{\mathcal{H}_k, N}^H (\alpha_k P_k \Psi_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} \Psi_k (\alpha_k P_k \Psi_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} \mathbf{V}_{\mathcal{H}_k, N} \Lambda_{\mathcal{F}_k}^2]\}. \quad (71)$$

Clearly in (71), η_{f_k} is a function of $\Lambda_{\mathcal{F}_k}$ and it can be denoted as $\eta_{f_k} = \xi_k(\Lambda_{\mathcal{F}_k})$ for clarification.

REFERENCES

- [1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12 pp. 3062–3080, Dec. 2004.
- [2] S. Jin, M. R. McKay, C. Zhong, and K.-K. Wong, "Ergodic capacity analysis of amplify-and-forward MIMO dual-hop systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2204–2224, May 2010.
- [3] 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.
- [4] 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.
- [5] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 517–519, July 2008.
- [6] F.-S. Tseng, W.-R. Wu, and J.-Y. Wu "Joint source/relay precoder design in nonregenerative cooperative systems using an MMSE criterion," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 4928–4933, Oct. 2009.
- [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] 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.
- [9] C. Li, X. Wang, L. Yang, and W.-P. Zhu, "A joint source and relay power allocation scheme for a class of MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4852–4860, Dec. 2009.
- [10] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 6068–6077, Dec. 2009.
- [11] C. Xing, S. Ma, Y.-C. Wu, and T.-S. Ng, "Transceiver design for dual-hop non-regenerative MIMO-OFDM relay systems under channel uncertainties," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6325–6339, Dec. 2010.
- [12] 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.
- [13] B. K. Chalise and L. Vandendorpe, "Joint linear processing for an amplify-and-forward MIMO relay channel with imperfect channel state information," *EURASIP J. Adv. Signal Process.*, vol. 2010, Article ID 640186, doi:10.1155/2010/640186.
- [14] Y. Rong, "Robust design for linear non-regenerative MIMO relays with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2455–2460, May 2011.
- [15] 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 Global Commun. Conf.*, Houston, TX, USA, Dec. 2011.
- [16] C. Xing, Z. Fei, Y.-C. Wu, S. Ma, and J. Kuang, "Robust transceiver design for AF MIMO relaying systems with column correlations," in *Proc. IEEE International Conference on Signal Processing, Communications and Computing*, Xi'an, China, Sep. 2011.
- [17] 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.
- [18] 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.
- [19] J. Wang and D. P. Palomar, "Worst-case robust MIMO transmission with imperfect channel knowledge," *IEEE Trans. Signal Process.*, vol.57, no. 9, pp. 3575–3587, Sep. 2009.

- [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] 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. 4–5, pp. 331–551, 2006.
- [22] E. Jorswieck and H. Boche, "Majorization and Matrix-Monotone Functions in Wireless Communications," *Foundations and Trends in Communications and Information Theory*, Now Publishers, vol. 3, no. 6, pp. 553–701, 2007.
- [23] A. Beck, A. Ben-Tal and Y. C. Eldar, "Robust mean-squared error estimation of multiple signals in linear systems affected by model and noise uncertainties," *Math. Programming*, vol.107, pp. 155-187, 2006.
- [24] 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.
- [25] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.
- [26] B. Hassibi and B. M. Hochwald, "How much training is needed in mutiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [27] S. Kay, *Fundamental of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [28] 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.
- [29] 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.
- [30] 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
- [31] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for Gaussian vector multiple access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp.145–151, Jan. 2004.
- [32] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.

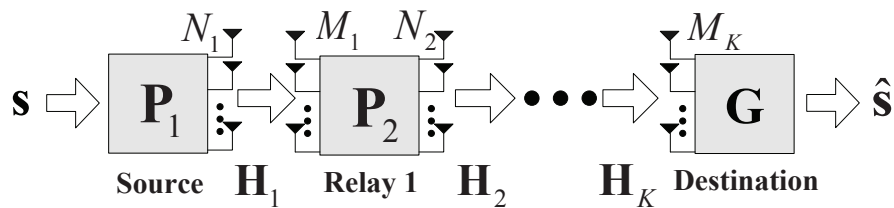


Fig. 1. A multi-hop amplify-and-forward MIMO relaying system.

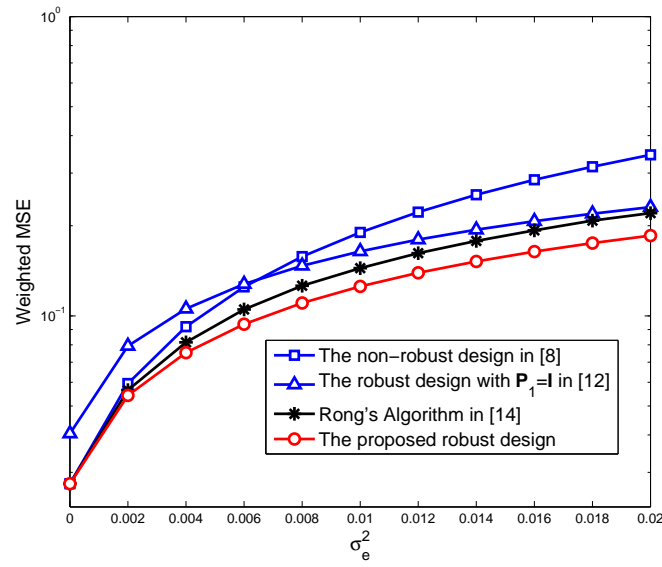


Fig. 2. Weighted MSE of the detected data in a dual-hop AF relaying system, when $\alpha = 0.6$, $\beta = 0$ and $P_k/\sigma_{n_k}^2 = 30\text{dB}$.

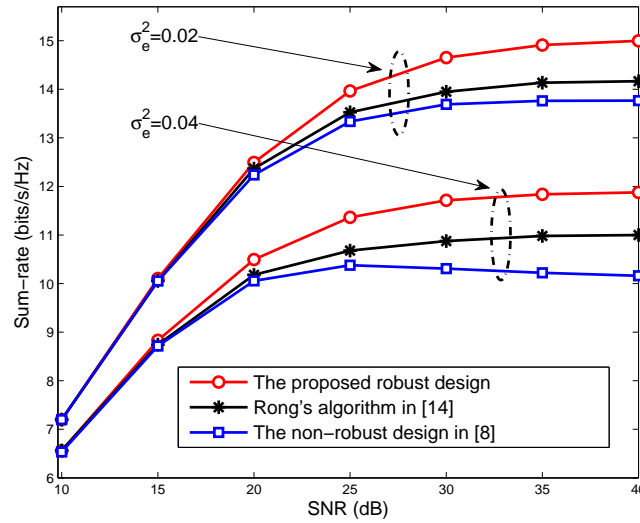


Fig. 3. Sum-rate of a dual-hop AF relaying system when $\alpha = 0.6$ and $\beta = 0$.



January 16, 2013

Chengwen Xing (S'08–M'10) received the B.Eng. degree from Xidian University, Xi'an, China, in 2005 and the Ph.D. degree in the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, in 2010. Since September 2010, he has been with the School of Information and Electronics, Beijing Institute of Technology (BIT), Beijing, China, where he is currently a Lecturer. He

DRAFT

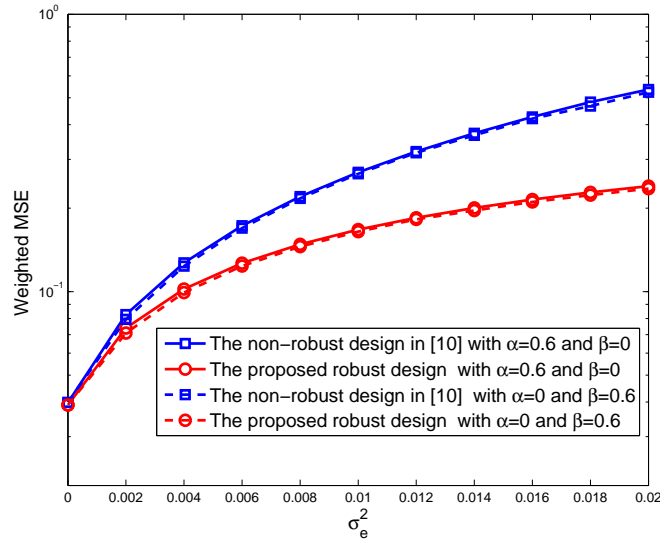


Fig. 4. Weighted MSE of the detected data in a three-hop AF relaying system when $P_k/\sigma_{n_k}^2 = 30\text{dB}$.

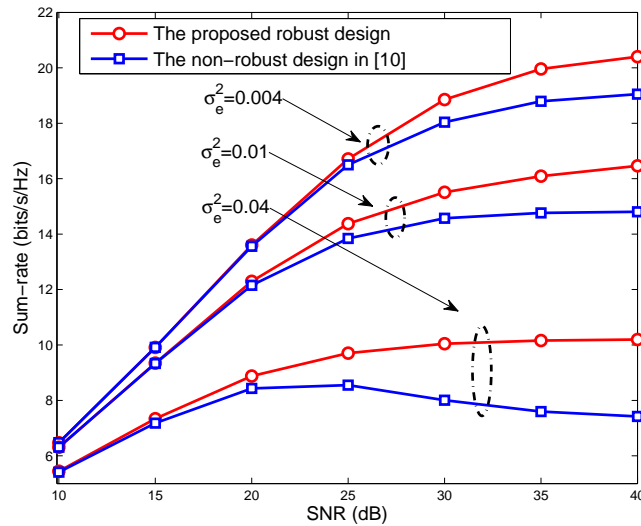


Fig. 5. Sum-rate of a three-hop AF relaying system when $\alpha = 0.6$ and $\beta = 0$.

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. He is currently serving as an editor for KSII Transactions on Internet and Information Systems.

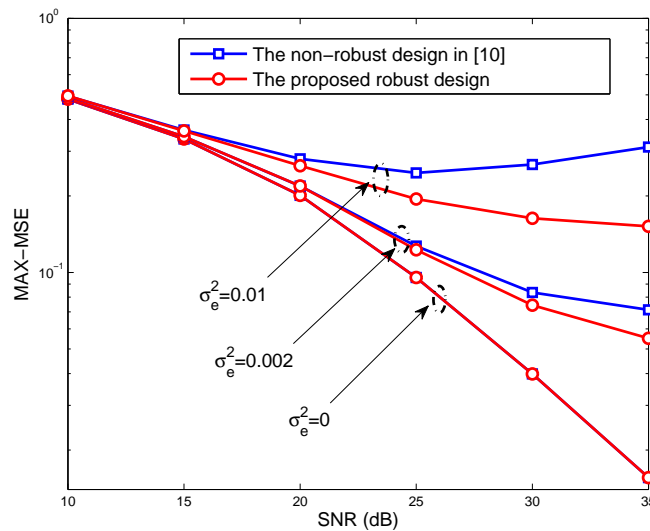


Fig. 6. Maximum MSE among different received data streams in a three-hop relaying system with $\alpha = 0$ and $\beta = 0.6$.

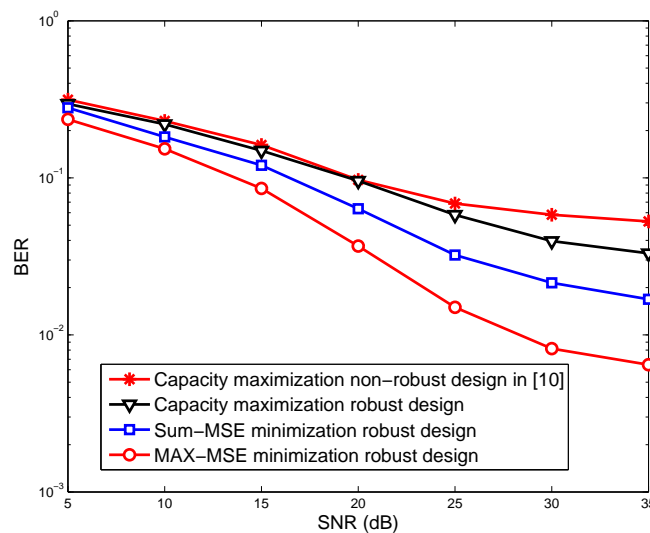


Fig. 7. BERs of the proposed robust design with different design objectives, when $\alpha = 0.6$, $\beta = 0$ and $\sigma_e^2 = 0.004$.



Shaodan Ma received her double Bachelor degrees in Science and Economics, and her Master degree of Engineering, from Nankai University, Tianjin, China. In 2006, she obtained her Ph. D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong. After graduation, she joined the University of Hong Kong as a Postdoctoral Fellow. Since August 2011, she has been with the University of Macau as an Assistant Professor. She was a visiting scholar in Princeton University in 2010 and is

January 16, 2013

DRAFT

currently an Honorary Assistant Professor in the University of Hong Kong. Her research interests are in the general areas of signal processing and communications, particularly, transceiver design, resource allocation and performance analysis, in OFDM, MIMO, and cooperative systems.



Zesong Fei (S03-M07) received the Ph.D. degree in Electronic Engineering in 2004 from Beijing Institute of Technology (BIT). He is now an Associate Professor in BIT and currently with the Research Institute of Communication Technology (RICT) of BIT, where he is involved in the design of the next generation high-speed wireless communication. His research interests include mobile communication, channel coding and modulation, cognitive radio and cooperative networking. He was chief investigator of China national Natural Science Fund project. He is the senior member of Chinese Institute of Electronics and China

Institute of Communications.



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, currently as an Associate 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 Transactions on Communications, IEEE Communications Letters, and Journal of Communications and Networks.



H. Vincent Poor (S'72, M'77, SM'82, F'87) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering and Dean of the School of Engineering and Applied Science. Dr. Poor's research interests are in the areas of stochastic analysis, statistical signal processing, and information theory, and their applications in wireless networks and related fields such as social networks and smart

grid. Among his publications in these areas the recent books *Smart Grid Communications and Networking* (Cambridge, 2012) and *Principles of Cognitive Radio* (Cambridge, 2013).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, a Fellow of the American Academy of Arts and Sciences, and an International Fellow of the Royal Academy of Engineering (U.K.). He is also a Fellow of the Institute of Mathematical Statistics, the Acoustical Society of America, and other organizations. He received the Technical Achievement and Society Awards of the SPS in 2007 and 2011, respectively. Recent recognition of his work includes the 2010 IET Ambrose Fleming Medal, the 2011 IEEE Eric E. Sumner Award, honorary doctorates from Aalborg University, the Hong Kong University of Science and Technology, and the University of Edinburgh.