Optimal Linear Non-Regenerative Multi-Hop MIMO Relays with MMSE-DFE Receiver at the Destination

Yue Rong, *Member, IEEE*

***Abstract*—In this paper, we study multi-hop non-regenerative multiple-input multiple-output (MIMO) relay communications with any number of hops. We design the optimal source pre-coding matrix and the optimal relay amplifying matrices for such relay network where a nonlinear minimal mean-squared error (MMSE)-decision feedback equalizer (DFE) is used at the destination node. We first derive the structure of the optimal source and relay matrices. Then based on the link between most commonly used MIMO system design objectives and the diagonal elements of the MSE matrix, we classify the objective functions into two categories: Schur-convex and Schur-concave composite objective functions. We show that when the composite objective function is Schur-convex, the MMSE-DFE receiver together with the optimal source and relay matrices enable an arbitrary number of source symbols to be transmitted at one time, and yield a significantly improved BER performance compared with non-regenerative MIMO relay systems using linear receivers at the destination. We also show that for Schur-concave composite objective functions, the optimal source and relay matrices, and the optimal feed-forward matrix at the destination node jointly diagonalize the multi-hop MIMO relay channel, and thus in such case, the nonlinear MMSE-DFE receiver is essentially equivalent to a linear MMSE receiver.**

***Index Terms*—MIMO relay network, multi-hop relay, MMSE, DFE, non-regenerative relay, majorization.**

I. INTRODUCTION

MULTI-HOP relay communication is well known for being a cost-effective approach in improving the energy-efficiency of communication system in the case of long source-destination distance [1]. When nodes in the relay network are equipped with multiple antennas, we call such system a multiple-input multiple-output (MIMO) relay system. For a basic three-node two-hop MIMO relay system, optimal algorithms are developed in [2]-[5] to maximize the mutual information (MI) between source and destination. In [6], [7], the optimal relay amplifying matrix is developed to minimize the mean-squared error (MSE) of the signal waveform estimation at the destination.

Recently, serially configured multi-hop MIMO relay communications with any number of hops attracted much research interest [8]-[12]. The asymptotic capacity of such system is studied in [8]-[10]. In [11], the authors investigated the diversity gain of multi-hop MIMO relay channel when the relays use diagonal amplifying matrices. Assuming that a linear minimal mean-squared error (MMSE) receiver is used at the destination, the MIMO relay design issue for most commonly used objective functions is investigated in [12]. It is shown in [12] that the optimal source precoding matrix, the optimal relay amplifying matrices, and the optimal linear MMSE receiving matrix jointly diagonalize the multi-hop MIMO relay channel when the objective function is Schur-concave. And such joint diagonalization along with a rotation of the source precoding matrix is also optimal for Schur-convex objective functions.

As we know from point-to-point MIMO systems, a nonlinear decision feedback equalizer (DFE) recovers the source signals successively by exploiting the finite alphabet property of the source signals. Therefore, a nonlinear DFE receiver has a much better bit-error-rate (BER) and MI performance than linear receivers [13], [14]. The DFE technique is also well-known as the successive interference cancellation (SIC) technique [13] or the V-BLAST technique [14]. In this paper, we focus on the nonlinear DFE receiver where the MMSE technique is applied to recover the source signal at each layer. We refer to such receiver as the MMSE-DFE receiver. It is well-known that the MMSE-DFE receiver is information lossless [13, Ch.8]. We assume that the multi-hop MIMO relay link being considered has already been established by protocols at higher layers (link and/or network layers) [15]. We also assume that when a link failure is detected, the failed node can be bypassed by increasing the transmission power of the previous relay node, see [16] and the references therein. Given the established MIMO relay link, we optimize the source precoding matrix and the relay amplifying matrices by applying the majorization theory [17] and the recently developed matrix generalized triangular decomposition (GTD) tool [18]. We show that given the structure of the optimal source and relay matrices, the complicated matrix-variable optimization problem is simplified to an equivalent optimization problem with scalar variables.

It has been shown in [19] that most commonly used objective functions in MIMO system design can be classified into Schur-concave and Schur-convex composite functions. In this paper we show that when the composite objective function is Schur-convex, the MMSE-DFE receiver together with the optimal source and relay matrices enable an arbitrary number

Manuscript received May 26, 2009; revised October 15, 2009 and February 16, 2010; accepted April 30, 2010. The associate editor coordinating the review of this paper and approving it for publication was R. Nabar.

The author is with the Department of Electrical and Computer Engineering, Curtin University of Technology, Bentley, WA 6102, Australia (e-mail: y.rong@curtin.edu.au).

Digital Object Identifier 10.1109/TWC.2010.07.090769

1536-1276/10$25.00 © 2010 IEEE

of source symbols to be transmitted at one time, and the multi-hop MIMO relay channel is uniformly decomposed into identical subchannels. Numerical examples demonstrate that in such case, the non-regenerative MIMO relay system with a nonlinear MMSE-DFE receiver yields a significantly lower BER compared with a non-regenerative MIMO relay system using a linear MMSE receiver at the destination. We also show that for Schur-concave composite objective functions, the optimal source and relay matrices, and the optimal feed-forward matrix at the destination node jointly diagonalize the multi-hop MIMO relay channel, and thus in such case, the nonlinear MMSE-DFE receiver is essentially equivalent to the linear MMSE receiver discussed in [12]. Therefore, [12] can be treated as a special case of this paper where the composite objective function is Schur-concave. The relay scheme using the optimal source, relay matrices and the nonlinear MMSE-DFE receiver developed in this paper is more general than the relay scheme in [12]. In this paper, for notational convenience, we consider a narrow band single-carrier system. However, our results can be straightforwardly generalized to wide band multi-carrier multi-hop MIMO relay systems.

We would like to mention that the optimal source matrix design for a single-hop (point-to-point) MIMO system using the nonlinear MMSE-DFE receiver is addressed in [19]-[21]. In particular, the optimal source precoding matrices for Schur-convex and Schur-concave composite objective functions are summarized in [19]. Our paper extends the results in [19], [21] from a single-hop MIMO channel to multi-hop non-regenerative MIMO relay communication systems with any number of hops. Note that the proof of the theorems for multi-hop MIMO relay system is much more involved than that for the single-hop MIMO channel. The extension from single-hop system to multi-hop system is significant, as the results are important for multi-hop wireless backhaul networks [1].

The rest of this paper is organized as follows. In Section II, we introduce the model of a multi-hop linear non-regenerative MIMO relay communication system with a nonlinear MMSE-DFE receiver at the destination. The structure of the optimal source and relay matrices are shown in Section III. In Section IV, we show some numerical examples. Conclusions are drawn in Section V. Notations used throughout the paper are summarized in Appendix A.

II. System Model

We consider a wireless communication system with one source node, one destination node, and $L-1$ relay nodes ($L \geq 2$). We assume that due to the propagation path-loss, the signal transmitted by the ith node can only be received by its direct neighboring nodes, i.e., the $(i+1)$-th and $(i-1)$-th nodes. Thus, the source signals travel through L hops until they are received by the destination node. Moreover, in order to avoid any interference, the neighboring nodes transmit signals at orthogonal channels (time and/or frequency). We also assume that the number of antennas at each node is N_i, $i = 1, \cdots, L+1$, and the number of source symbols in each transmission is N_b. Like [3]-[12], a linear non-regenerative relay matrix is used at each relay. However, in contrast to [3]-[12], a nonlinear MMSE-DFE receiver is deployed at the destination node.

The $N_1 \times 1$ signal vector transmitted by the source node is

$$\mathbf{x}_1 = \mathbf{F}_1 \mathbf{s} \tag{1}$$

where $\mathbf{s} = [s_1, s_2, \cdots, s_{N_b}]^T$ is the $N_b \times 1$ source symbol vector, $(\cdot)^T$ is the matrix (vector) transpose, and $\mathbf{F}_1$ is the $N_1 \times N_b$ source precoding matrix. We assume that $\mathrm{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_{N_b}$, where $\mathrm{E}[\cdot]$ stands for the statistical expectation, $(\cdot)^H$ denotes the Hermitian transpose, and $\mathbf{I}_n$ is an $n \times n$ identity matrix. The $N_i \times 1$ signal vector received at the ith node is written as

$$\mathbf{y}_i = \mathbf{H}_{i-1}\mathbf{x}_{i-1} + \mathbf{v}_i, \qquad i = 2, \cdots, L+1 \tag{2}$$

where $\mathbf{H}_{i-1}$ is the $N_i \times N_{i-1}$ MIMO channel matrix of the $(i-1)$-th hop (between the ith and the $(i-1)$-th nodes), $\mathbf{v}_i$ is the $N_i \times 1$ independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) vector at the ith node, and $\mathbf{x}_{i-1}$ is the $N_{i-1} \times 1$ signal vector transmitted by the $(i-1)$-th node. We assume that the noises are complex circularly symmetric with zero mean and unit variance.

Using the linear non-regenerative strategy, the input-output relationship at node i is given by

$$\mathbf{x}_i = \mathbf{F}_i \mathbf{y}_i, \qquad i = 2, \cdots, L \tag{3}$$

where $\mathbf{F}_i$ is the $N_i \times N_i$ relay amplifying matrix at node i. Combining (1)-(3), the received signal vector at the destination node (the $(L+1)$-th node) can be written as [12]

$$\mathbf{y}_{L+1} = \bar{\mathbf{H}}\mathbf{s} + \bar{\mathbf{v}}$$

where $\bar{\mathbf{H}}$ and $\bar{\mathbf{v}}$ are the equivalent MIMO channel matrix and the equivalent noise vector, and given respectively by [12]

$$\bar{\mathbf{H}} = \bigotimes_{i=L}^{1}(\mathbf{H}_i\mathbf{F}_i) \tag{4}$$

$$\bar{\mathbf{v}} = \sum_{l=2}^{L}\Big(\bigotimes_{i=L}^{l}(\mathbf{H}_i\mathbf{F}_i)\mathbf{v}_l\Big) + \mathbf{v}_{L+1}. \tag{5}$$

Here for matrices $\mathbf{A}_i$, $\bigotimes_{i=l}^{k}(\mathbf{A}_i) \triangleq \mathbf{A}_l \cdots \mathbf{A}_k$. From (5) we see that $\bar{\mathbf{v}}$ is nonwhite with the following covariance matrix [12]

$$\mathbf{C}_{\bar{v}} = \sum_{l=2}^{L}\Big(\bigotimes_{i=L}^{l}(\mathbf{H}_i\mathbf{F}_i)\bigotimes_{i=l}^{L}\big(\mathbf{F}_i^H\mathbf{H}_i^H\big)\Big) + \mathbf{I}_{N_{L+1}}. \tag{6}$$

We assume that all MIMO channels are quasi-static block fading channel, i.e., $\mathbf{H}_i$, $i = 1, \cdots, L$ are constant over some time T_c before they change to another realization. It is assumed that T_c is much longer than the time required for a symbol to be transmitted from the source to the destination. We also assume that the source node has the channel state information (CSI) knowledge of $\mathbf{H}_1$, the destination node knows $\bar{\mathbf{H}}$, and the ith node, $i = 2, \cdots, L$, knows the CSI of its backward channel $\mathbf{H}_{i-1}$ and its forward channel $\mathbf{H}_i$. In practice, the backward CSI can be obtained through standard training methods. The forward CSI required at the ith node ($\mathbf{H}_i$) is exactly the backward CSI at the $(i+1)$-th node, and thus can be obtained by a feedback from the $(i+1)$-th node. The assumptions on the channels made above are justifiable, since for wireless relays the fading is often relatively slow whenever the mobility of the relays is relatively low, and for static relays, the channel state information can be almost constant. Therefore, for quasi-static channels, the necessary CSI can be obtained at each node with a reasonably high precision.

At the destination node, a nonlinear DFE receiver is used to detect the source symbols successively with the N_bth symbol detected first and the first symbol detected last. At each step, a linear feed-forward filter is applied to the received signal vector and the previously detected symbols are fed back and subtracted from the filtered signal. Assuming that there is no error propagation in the DFE receiver, the kth source symbol is estimated as [13, Ch.8]

$$\hat{s}_k = \mathbf{w}_k^H \mathbf{y}_{L+1} - \sum_{l=k+1}^{N_b} b_{k,l} s_l, \qquad k = 1, \cdots, N_b \quad (7)$$

where $\mathbf{w}_k$ is the feed-forward vector for the kth symbol, and $b_{k,l}$, $l = k+1, \cdots N_b$ are the feedback coefficients for the kth symbol. The error-propagation-free assumption is justified by information theory: Powerful coding can be applied to each layer such that an arbitrarily small error probability is achieved if the data rate at each layer is less than the sub-channel capacity. In Section IV it will be seen that even for uncoded system, the error propagation effect only slightly increases the system BER. By introducing $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_{N_b}]$, $\hat{\mathbf{s}} = [\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_{N_b}]^T$, and an $N_b \times N_b$ strictly upper-triangle matrix $\mathbf{B}$ with nonzero elements $b_{k,l}$, we can represent (7) in matrix form as

$$\hat{\mathbf{s}} = \mathbf{W}^H \mathbf{y}_{L+1} - \mathbf{B}\mathbf{s} = \big(\mathbf{W}^H \bar{\mathbf{H}} - \mathbf{B}\big)\mathbf{s} + \mathbf{W}^H \bar{\mathbf{v}}. \quad (8)$$

Here $\mathbf{W}$ and $\mathbf{B}$ are the feed-forward and feedback matrix of the DFE receiver, respectively.

The performance of the kth data stream in (7), $k = 1, \cdots, N_b$, measured in terms of the MSE and the signal-to-interference-noise ratio (SINR), is given respectively by

$$\begin{aligned}\mathrm{MSE}_k &= \mathrm{E}\left[|s_k - \hat{s}_k|^2\right] \\ &= \left|\mathbf{w}_k^H \left[\bar{\mathbf{H}}\right]_k - 1\right|^2 + \sum_{l=1}^{k-1} \left|\mathbf{w}_k^H \left[\bar{\mathbf{H}}\right]_l\right|^2 \\ &\quad + \mathbf{w}_k^H \mathbf{C}_{\bar{v}} \mathbf{w}_k + \sum_{l=k+1}^{N_b} \left|\mathbf{w}_k^H \left[\bar{\mathbf{H}}\right]_l - b_{k,l}\right|^2 \end{aligned} \quad (9)$$

$$\mathrm{SINR}_k = \frac{\left|\mathbf{w}_k^H \left[\bar{\mathbf{H}}\right]_k\right|^2}{\sum_{l=1}^{k-1}\left|\mathbf{w}_k^H[\bar{\mathbf{H}}]_l\right|^2 + \mathbf{w}_k^H \mathbf{C}_{\bar{v}} \mathbf{w}_k + \sum_{l=k+1}^{N_b}\left|\mathbf{w}_k^H[\bar{\mathbf{H}}]_l - b_{k,l}\right|^2} \quad (10)$$

where $\left[\bar{\mathbf{H}}\right]_k$ is the kth column vector of $\bar{\mathbf{H}}$. From (9) and (10) we find that the $b_{k,l}$ which minimizes MSE_k and maximizes SINR_k (and thus optimizes most commonly used objective functions in MIMO system design) should satisfy $\left|\mathbf{w}_k^H \left[\bar{\mathbf{H}}\right]_l - b_{k,l}\right| = 0$. Therefore, the optimal $b_{k,l}$ should be $b_{k,l} = \mathbf{w}_k^H \left[\bar{\mathbf{H}}\right]_l$, or equivalently in matrix form as

$$\mathbf{B} = \mathcal{U}\big[\mathbf{W}^H \bar{\mathbf{H}}\big] \quad (11)$$

where $\mathcal{U}[\mathbf{A}]$ denotes the strictly upper-triangular part of $\mathbf{A}$. Substituting (11) back into (7), we obtain

$$\hat{s}_k = \mathbf{w}_k^H \left(\left[\bar{\mathbf{H}}\right]_{1:k} [\mathbf{s}]_{1:k} + \bar{\mathbf{v}}\right), \qquad k = 1, \cdots, N_b \quad (12)$$

where $[\mathbf{a}]_{1:k}$ denotes a vector containing the first k elements of vector $\mathbf{a}$, $[\mathbf{A}]_{1:k}$ stands for a matrix containing the first k columns of $\mathbf{A}$.

When the MMSE criterion is used to estimate each symbol, from (12) the feed-forward matrix $\mathbf{W}$ is given by [13, Ch.8]

$$\mathbf{w}_k = \left(\left[\bar{\mathbf{H}}\right]_{1:k} \left[\bar{\mathbf{H}}\right]_{1:k}^H + \mathbf{C}_{\bar{v}}\right)^{-1} \left[\bar{\mathbf{H}}\right]_k, \quad k = 1, \cdots, N_b \quad (13)$$

where $(\cdot)^{-1}$ denotes the matrix inversion. Let us introduce the following QR decomposition [22]

$$\mathbf{G} \triangleq \begin{bmatrix} \mathbf{C}_{\bar{v}}^{-\frac{1}{2}} \bar{\mathbf{H}} \\ \mathbf{I}_{N_b} \end{bmatrix} = \mathbf{Q}\mathbf{R} = \begin{bmatrix} \bar{\mathbf{Q}} \\ \underline{\mathbf{Q}} \end{bmatrix} \mathbf{R} \quad (14)$$

where $\mathbf{R}$ is an $N_b \times N_b$ upper-triangular matrix with all positive diagonal elements, $\mathbf{Q}$ is an $(N_b + N_{L+1}) \times N_b$ semi-unitary matrix with $\mathbf{Q}^H \mathbf{Q} = \mathbf{I}_{N_b}$, $\bar{\mathbf{Q}}$ is a matrix containing the first N_{L+1} rows of $\mathbf{Q}$, and $\underline{\mathbf{Q}}$ contains the last N_b rows of $\mathbf{Q}$. Obviously, we have from (14) that

$$\mathbf{C}_{\bar{v}}^{-\frac{1}{2}} \bar{\mathbf{H}} = \bar{\mathbf{Q}}\mathbf{R}, \qquad \underline{\mathbf{Q}} = \mathbf{R}^{-1}. \quad (15)$$

Using (14) and (15), we can express $\mathbf{W}$, $\mathbf{B}$, and the MSE matrix defined by $\mathbf{E} = \mathrm{E}\big[(\hat{\mathbf{s}} - \mathbf{s})(\hat{\mathbf{s}} - \mathbf{s})^H\big]$ as follows.

THEOREM 1: Using the QR decomposition (14), the feed-forward weight matrix $\mathbf{W}$, the feedback matrix $\mathbf{B}$, and the MSE matrix $\mathbf{E}$ can be written as

$$\mathbf{W} = \mathbf{C}_{\bar{v}}^{-\frac{1}{2}} \bar{\mathbf{Q}}\, \mathbf{D}_R^{-1}, \quad \mathbf{B} = \mathbf{D}_R^{-1}\mathbf{R} - \mathbf{I}_{N_b}, \quad \mathbf{E} = \mathbf{D}_R^{-2} \quad (16)$$

where $\mathbf{D}_R$ is a matrix taking the diagonal elements of $\mathbf{R}$ as the main diagonal and zero elsewhere.

PROOF: See Appendix B. □

Interestingly, from (16) we find that for non-regenerative MIMO relay systems using the nonlinear MMSE-DFE receiver at the destination, the MSE matrix $\mathbf{E}$ is diagonal for any channel realization. Note that a similar theorem is proved in [19] for the case of white $\bar{\mathbf{v}}$ (using our notation). Obviously, Theorem 1 extends the result obtained in [19] to nonwhite $\bar{\mathbf{v}}$. In the following, we use (16) to derive the optimal source precoding matrix and the optimal relay amplifying matrices.

III. OPTIMAL SOURCE AND RELAY MATRICES

It has been shown in [23] that most commonly used objective functions for MIMO systems can be represented as functions of the main diagonal elements of the MSE matrix. Denoting $q(\mathbf{x})$ as a unified objective function which is increasing with respect to each element of $\mathbf{x}$ and using Theorem 1, the objective function can be represented as $q\big(\mathbf{d}\big[\mathbf{D}_R^{-2}\big]\big)$. Here, for a matrix $\mathbf{A}$, $\mathbf{d}[\mathbf{A}]$ is a column vector containing all main diagonal elements of $\mathbf{A}$. It can be seen from (4), (6), and (14) that $\mathbf{D}_R$ depends on $\mathbf{F}_i$, $i = 1, \cdots, L$. Thus, the source precoding matrix and all relay amplifying matrices need to be appropriately designed in order to optimize $q\big(\mathbf{d}\big[\mathbf{D}_R^{-2}\big]\big)$. Note that for a multi-hop single-input single-output (SISO) relay network [24] and serially configured sensor network [25], the generalized SNR objective function depends only on the last hop relay function, when the estimate-and-forward relay protocol is used. Therefore, the multi-hop non-regenerative MIMO relay optimization problem is much more challenging than the SISO relay design problem.

The non-regenerative MIMO relay design problem can be formulated as

$$\min_{\mathbf{D}_R, \{\mathbf{F}_i\}} \quad q\big(\mathbf{d}\big[\mathbf{D}_R^{-2}\big]\big) \tag{17}$$

$$\text{s.t.} \quad \begin{bmatrix} \mathbf{C}_{\bar{v}}^{-\frac{1}{2}} \bar{\mathbf{H}} \\ \mathbf{I}_{N_b} \end{bmatrix} = \mathbf{QR} \tag{18}$$

$$\operatorname{tr}\big(\mathbf{F}_1 \mathbf{F}_1^H\big) \le p_1 \tag{19}$$

$$\operatorname{tr}\Bigg(\mathbf{F}_i \Bigg(\sum_{l=1}^{i-1} \Big(\bigotimes_{k=i-1}^{l} (\mathbf{H}_k \mathbf{F}_k) \bigotimes_{k=l}^{i-1} \big(\mathbf{F}_k^H \mathbf{H}_k^H\big)\Big) + \mathbf{I}_{N_i}\Bigg) \mathbf{F}_i^H\Bigg) \le p_i, \quad i = 2, \cdots, L \tag{20}$$

where $\{\mathbf{F}_i\} \triangleq \{\mathbf{F}_i, i = 1, \cdots, L\}$, $p_i > 0$ is the transmission power available at node i, and $\operatorname{tr}(\cdot)$ denotes the trace of a matrix. Here (19) and (20) are the transmission power constraints at the source node and all relay nodes, respectively. The following definitions are required to state the theorem on the solution of problem (17)-(20).

DEFINITION 1: Consider any two nonnegative $N \times 1$ vectors $\mathbf{x}$ and $\mathbf{y}$, let $x_{[1]} \ge x_{[2]} \ge \cdots \ge x_{[N]}$, $y_{[1]} \ge y_{[2]} \ge \cdots \ge y_{[N]}$ denote the elements of $\mathbf{x}$ and $\mathbf{y}$ sorted in decreasing order, respectively. We say that $\mathbf{x}$ is multiplicatively majorized by $\mathbf{y}$, denoted as $\mathbf{x} \prec_{\times} \mathbf{y}$, if $\prod_{i=1}^{n} x_{[i]} \le \prod_{i=1}^{n} y_{[i]}$, for $n = 1, \cdots, N-1$, and $\prod_{i=1}^{N} x_{[i]} = \prod_{i=1}^{N} y_{[i]}$. We say that $\mathbf{x}$ is weakly multiplicatively submajorized by $\mathbf{y}$, denoted as $\mathbf{x} \prec_{\times(\mathrm{w})} \mathbf{y}$, if $\prod_{i=1}^{n} x_{[i]} \le \prod_{i=1}^{n} y_{[i]}$, for $n = 1, \cdots, N$.

DEFINITION 2 [17, 1.A.1, 1.A.2]: For any two real-valued $N \times 1$ vectors $\mathbf{x}$ and $\mathbf{y}$, we say that $\mathbf{x}$ is additively majorized by $\mathbf{y}$, denoted as $\mathbf{x} \prec_{+} \mathbf{y}$, if $\sum_{i=1}^{n} x_{[i]} \le \sum_{i=1}^{n} y_{[i]}$, for $n = 1, \cdots, N-1$, and $\sum_{i=1}^{N} x_{[i]} = \sum_{i=1}^{N} y_{[i]}$. We say that $\mathbf{x}$ is weakly additively submajorized by $\mathbf{y}$, denoted as $\mathbf{x} \prec_{+(\mathrm{w})} \mathbf{y}$, if $\sum_{i=1}^{n} x_{[i]} \le \sum_{i=1}^{n} y_{[i]}$, for $n = 1, \cdots, N$.

DEFINITION 3 [17, 3.A.1]: A real-valued function f is called Schur-convex if $f(\mathbf{x}) \le f(\mathbf{y})$ for $\mathbf{x} \prec_{+} \mathbf{y}$, or called Schur-concave if $f(\mathbf{x}) \ge f(\mathbf{y})$ for $\mathbf{x} \prec_{+} \mathbf{y}$.

Let us write the singular value decomposition (SVD) of $\mathbf{H}_i$ as

$$\mathbf{H}_i = \mathbf{U}_i \boldsymbol{\Sigma}_i \mathbf{V}_i^H, \qquad i = 1, \cdots, L \tag{21}$$

where the dimensions of $\mathbf{U}_i$, $\boldsymbol{\Sigma}_i$, $\mathbf{V}_i$ are $N_{i+1} \times N_{i+1}$, $N_{i+1} \times N_i$, $N_i \times N_i$, respectively. We assume that the main diagonal elements of $\boldsymbol{\Sigma}_i$, $i = 1, \cdots, L$, are arranged in the decreasing order. We also introduce $M = \min(R_h, N_b)$, where $R_h \triangleq \min(\operatorname{rank}(\mathbf{H}_1), \operatorname{rank}(\mathbf{H}_2), \cdots, \operatorname{rank}(\mathbf{H}_L))$ and $\operatorname{rank}(\cdot)$ denotes the rank of a matrix. The following theorem establishes the optimal structure of the source and relay matrices.

THEOREM 2: For multi-hop non-regenerative MIMO relay systems using nonlinear MMSE-DFE receiver at the destination, assuming that $\operatorname{rank}(\mathbf{F}_i) = M$, $i = 1, \cdots, L$, the optimal source precoding matrix and the relay amplifying matrices are given by

$$\mathbf{F}_1 = \mathbf{V}_{1,1} \boldsymbol{\Lambda}_1 \mathbf{V}_{F_1}^H, \quad \mathbf{F}_i = \mathbf{V}_{i,1} \boldsymbol{\Lambda}_i \mathbf{U}_{i-1,1}^H, \quad i = 2, \cdots, L \tag{22}$$

where $\boldsymbol{\Lambda}_i = \operatorname{diag}(\lambda_{i,1}, \lambda_{i,2}, \cdots, \lambda_{i,M})$ are $M \times M$ diagonal matrices, $\mathbf{U}_{i,1}$ and $\mathbf{V}_{i,1}$ contain the leftmost M vectors of $\mathbf{U}_i$ and $\mathbf{V}_i$, respectively, and $\mathbf{V}_{F_1}$ is an $N_b \times M$ semi-unitary matrix ($\mathbf{V}_{F_1}^H \mathbf{V}_{F_1} = \mathbf{I}_M$) such that the QR decomposition in (18) holds with $\mathbf{D}_R$ being the solution to the problem (23)-(26) given below. In particular, $\langle \lambda_{i,k} \rangle \triangleq \{\lambda_{i,k}, 1 \le i \le L, 1 \le k \le M\}$ are obtained from the following optimization problem

$$\min_{\mathbf{d}[\mathbf{D}_R], \langle \lambda_{i,k} \rangle} q\big(\mathbf{d}\big[\mathbf{D}_R^{-2}\big]\big) \tag{23}$$

$$\text{s.t.} \;\; \mathbf{d}\big[\mathbf{D}_R^2\big] \prec_{\times(\mathrm{w})} \Bigg[\Bigg\{1 + \frac{\prod_{i=1}^{L} \sigma_{i,k}^2 \lambda_{i,k}^2}{1 + \sum_{i=2}^{L} \prod_{l=i}^{L} \sigma_{l,k}^2 \lambda_{l,k}^2}\Bigg\}_M^T, \mathbf{1}_{N_b - M}\Bigg]^T \tag{24}$$

$$\sum_{k=1}^{M} \lambda_{1,k}^2 \le p_1 \tag{25}$$

$$\sum_{k=1}^{M} \lambda_{i,k}^2 \Bigg(\sum_{j=1}^{i-1} \prod_{l=j}^{i-1} \lambda_{l,k}^2 \sigma_{l,k}^2 + 1\Bigg) \le p_i, \quad i = 2, \cdots, L \tag{26}$$

where $\sigma_{i,k}$, $i = 1, \cdots, L, k = 1, \cdots, M$, are the kth main diagonal element of $\boldsymbol{\Sigma}_i$, for a scalar x, $\{x\}_n \triangleq [x_1, x_2, \cdots, x_n]^T$, and $\mathbf{1}_n$ denotes a $1 \times n$ vector with all 1 elements.

PROOF: See Appendix C. □

Interestingly, from Theorem 2 we see that the structure of the optimal $\mathbf{F}_i$, $i = 1, \cdots, L$, is similar to that of the optimal source and relay matrices when the linear MMSE receiver is used at the destination node [12]. The motivation of the assumption $\operatorname{rank}(\mathbf{F}_i) = M$ is to avoid any transmission power loss at each node. It is worth noting that Theorem 2 simplifies the matrix-variable optimization problem (17)-(20) to the problem (23)-(26) with scalar variables. Moreover, (22) presents the structure of the optimal $\mathbf{F}_i$ in the form of its singular value decomposition. For $i = 2, \cdots, L$, the left and right singular vector matrices of $\mathbf{F}_i$, (i.e., $\mathbf{V}_{i,1}$ and $\mathbf{U}_{i-1,1}$) are known, while the singular value matrix $\boldsymbol{\Lambda}_i$ is obtained by solving the problem (23)-(26). For the optimal $\mathbf{F}_1$, $\mathbf{V}_{1,1}$ is known, $\boldsymbol{\Lambda}_1$ is obtained from (23)-(26), while $\mathbf{V}_{F,1}$ is chosen to guarantee that the QR decomposition (14) yields the $\mathbf{D}_R$ in (23). In particular, we carry out the GTD of the following matrix

$$\boldsymbol{\Psi} \triangleq \begin{bmatrix} \mathbf{C}_{\bar{v}}^{-\frac{1}{2}} \bigotimes_{i=L}^{2} (\mathbf{H}_i \mathbf{F}_i) \mathbf{H}_1 \mathbf{V}_{1,1} \big[\boldsymbol{\Lambda}_1, \mathbf{0}_{M \times (N_b - M)}\big] \\ \mathbf{I}_{N_b} \end{bmatrix} = \mathbf{Q}_{\Psi} \mathbf{R} \mathbf{P}_{\Psi}^H$$

where $\mathbf{R}$ has the diagonal elements $\mathbf{D}_R$ obtained from the first step, and $\mathbf{0}_{p \times m}$ denotes a $p \times m$ matrix with all zero entries. Then we obtain $\mathbf{V}_{F_1} = \big[\mathbf{P}_{\Psi}^H\big]_{1:M}$ and consequently $\mathbf{F}_1$ in (22). The justification of this step can be found in Appendix C.

It can be seen from (22) that the optimal relay amplifying matrices $\mathbf{F}_i$, $i = 2, \cdots, L$, have the classical SVD beamforming structure. In fact, the ith node first performs receive beamforming using the Hermitian transpose of the left singular matrix of its direct backward channel $\mathbf{H}_{i-1}$. Then it conducts a power loading operation. Finally, a transmit beamforming is performed by the ith node using the right singular matrix of its direct forward channel $\mathbf{H}_i$. Substituting (22) back into (4) and (6), we have

$$\bar{\mathbf{H}} = \mathbf{U}_{L,1} \mathbf{D}_h \mathbf{V}_{F_1}^H \tag{27}$$

$$\mathbf{C}_{\bar{v}} = \mathbf{U}_{L,1} \mathbf{D}_c \mathbf{U}_{L,1}^H + \mathbf{I}_{N_{L+1}} \tag{28}$$

where $\mathbf{D}_h$ and $\mathbf{D}_c$ are $M \times M$ diagonal matrices with the kth diagonal elements $k = 1, \cdots, M$, given respectively by

$$d_{h,k} = \prod_{l=1}^{L} \lambda_{l,k}\sigma_{l,k} \qquad d_{c,k} = \sum_{l=2}^{L}\prod_{i=l}^{L} \lambda_{i,k}^2\sigma_{i,k}^2 .$$

Computing the optimal $\langle\lambda_{i,k}\rangle$ by solving the problem (23)-(26) requires a centralized processing. The processor which performs the optimization can reside at any node depending on the capability of all nodes. This processor first collects the information on $\sigma_{i,k}$, $i = 1, \cdots, L$, $k = 1, \cdots, M$. Then it performs the optimization and computes $\mathbf{V}_{F_1}$. Finally, it sends the optimal $\lambda_{i,k}$, $k = 1, \cdots, M$ to the ith node, $i = 1, \cdots, L$, and $\mathbf{V}_{F_1}$ to the source node. At the ith node, after the optimal $\lambda_{i,k}$, $k = 1, \cdots, M$ are received (also after the reception of $\mathbf{V}_{F_1}$ for the source node), the optimal matrix $\mathbf{F}_i$ is assembled using (22).

For general objective functions, the problem (23)-(26) is still complicated, especially the constraint (24). Let us define the composite objective function of $q \diamond \exp(\mathbf{x}) \triangleq q\left(e^{x_1}, e^{x_2}, \cdots, e^{x_{N_b}}\right)$. In the following, we show that for Schur-concave $q \diamond \exp$ and Schur-convex $q \diamond \exp$, the problem (23)-(26) can be further simplified. Examples of Schur-convex $q \diamond \exp$ can be found in [19] such as the sum of MSEs of all data streams (SMSE), and the negative of the product of the SINRs of all data streams (NPSINR). Using $\mathbf{E} = \mathbf{D}_R^{-2}$ in (16), the MSE of each date stream is given by $[\mathbf{E}]_{k,k} = [\mathbf{D}_R]_{k,k}^{-2}$, and the SINR of each stream can be written as $[\mathbf{E}]_{k,k}^{-1} - 1 = [\mathbf{D}_R]_{k,k}^2 - 1$. Therefore, these two objectives can be expressed respectively as

$$\text{SMSE} \triangleq q_1\big(\mathbf{d}\big[\mathbf{D}_R^{-2}\big]\big) = \sum_{k=1}^{N_b} [\mathbf{D}_R]_{k,k}^{-2}$$

$$\text{NPSINR} \triangleq q_2\big(\mathbf{d}\big[\mathbf{D}_R^{-2}\big]\big) = -\prod_{k=1}^{N_b}\left([\mathbf{D}_R]_{k,k}^2 - 1\right) .$$

Some Schur-concave $q \diamond \exp$ are also listed in [19] such as the exponentially weighted product of MSEs (WPMSE) and the negative of the weighted sum of SINRs (NWSSINR). Using the relation between $\mathbf{D}_R$ and the MSE and SINR of each data stream, they are written respectively as

$$\text{WPMSE} \triangleq q_3\big(\mathbf{d}\big[\mathbf{D}_R^{-2}\big]\big) = \prod_{k=1}^{N_b} [\mathbf{D}_R]_{k,k}^{-2\alpha_k}$$

$$\text{NWSSINR} \triangleq q_4\big(\mathbf{d}\big[\mathbf{D}_R^{-2}\big]\big) = -\sum_{k=1}^{N_b}\alpha_k\left([\mathbf{D}_R]_{k,k}^2 - 1\right)$$

where $0 < \alpha_1 \le \cdots \le \alpha_{N_b}$ are the weighting coefficients.

There are two interesting links between the Schur-convexity of q and $q \diamond \exp$. First, if $q(\mathbf{x})$ is Schur-convex with respect to $\mathbf{x}$, $q \diamond \exp(\mathbf{x})$ is also Schur-convex with respect to $\mathbf{x}$. This can be easily shown by using [17, 3.B.2] and the fact that e^x is convex and q is increasing and Schur-convex. Second, if $q \diamond \exp$ is Schur-concave, then q is also Schur-concave [19, Lemma 2.12]. However, for Schur-concave q, $q \diamond \exp$ is not necessarily Schur-concave. It is shown in [19] that for some Schur-concave q, $q \diamond \exp$ is Schur-convex. In the following we introduce two theorems that establish the optimal $\mathbf{V}_{F_1}$ and $\langle\lambda_{i,k}\rangle$ for Schur-concave and Schur-convex $q \diamond \exp$, respectively.

THEOREM 3: For the relay design problem (23)-(26), if $q \diamond \exp$ is increasing and Schur-concave, then the optimal $\mathbf{V}_{F_1}$ is equal to $\left[\mathbf{I}_M, \mathbf{0}_{M\times(N_b-M)}\right]^T$, and $\langle\lambda_{i,k}\rangle$ is the solution to the following problem

$$\min_{\langle\lambda_{i,k}\rangle} \; q\left(\left[\left\{\left(1 + \frac{\prod_{i=1}^{L}\sigma_{i,k}^2\lambda_{i,k}^2}{1 + \sum_{i=2}^{L}\prod_{l=i}^{L}\sigma_{l,k}^2\lambda_{l,k}^2}\right)^{-1}\right\}_M^T, \mathbf{1}_{N_b-M}\right]^T\right) \quad (29)$$

$$\text{s.t.} \; \sum_{k=1}^{M}\lambda_{1,k}^2 \le p_1 \quad (30)$$

$$\sum_{k=1}^{M}\lambda_{i,k}^2\left(\sum_{j=1}^{i-1}\prod_{l=j}^{i-1}\lambda_{l,k}^2\sigma_{l,k}^2 + 1\right) \le p_i, \quad i = 2, \cdots, L \quad (31)$$

PROOF: See Appendix D. □

Interestingly, the problem (29)-(31) is identical to the multi-hop non-regenerative MIMO relay design problem with Schur-concave q, when a linear MMSE receiver is adopted at the destination node [12]. We find that for Schur-concave $q \diamond \exp$, without wasting the transmission power at the source node, the number of symbols N_b should be no greater than R_h.

Assuming that $N_b = R_h = M$ and substituting $\mathbf{V}_{F_1} = \mathbf{I}_{N_b}$ back into (27) we obtain $\bar{\mathbf{H}} = \mathbf{U}_{L,1}\mathbf{D}_h$, and

$$[\bar{\mathbf{H}}]_k = d_{h,k}[\mathbf{U}_{L,1}]_k \quad (32)$$

$$\left[\bar{\mathbf{H}}\right]_{1:k} = [\mathbf{U}_{L,1}]_{1:k}[\mathbf{D}_h]_{1:k,1:k}, \qquad k = 1, \cdots, N_b \quad (33)$$

where for any matrix $\mathbf{A}$, $[\mathbf{A}]_{1:k,1:k}$ is a principal sub-matrix of $\mathbf{A}$ lying in the first k rows and the first k columns of $\mathbf{A}$. Substituting (28), (32), and (33) back into (13), we have

$$\begin{aligned}\mathbf{w}_k &= d_{h,k}\big([\mathbf{U}_{L,1}]_{1:k}[\mathbf{D}_h^2]_{1:k,1:k}[\mathbf{U}_{L,1}]_{1:k}^H \\ &\quad + \mathbf{U}_{L,1}\mathbf{D}_c\mathbf{U}_{L,1}^H + \mathbf{I}_{N_{L+1}}\big)^{-1}[\mathbf{U}_{L,1}]_k \\ &= \frac{d_{h,k}}{d_{h,k}^2 + d_{c,k} + 1}[\mathbf{U}_{L,1}]_k . \end{aligned} \quad (34)$$

Thus, the feed-forward matrix $\mathbf{W}$ can be represented as $\mathbf{W} = \mathbf{U}_{L,1}\mathbf{D}_w$, where $\mathbf{D}_w$ is an $N_b \times N_b$ diagonal matrix with the kth diagonal element given by $d_{w,k} = d_{h,k}/(d_{h,k}^2 + d_{c,k} + 1)$. Now we have $\mathbf{W}^H\bar{\mathbf{H}} = \mathbf{D}_w\mathbf{U}_{L,1}^H\mathbf{U}_{L,1}\mathbf{D}_h = \mathbf{D}_w\mathbf{D}_h$ which is an $N_b \times N_b$ diagonal matrix with the kth diagonal element as $d_{h,k}^2/(d_{h,k}^2 + d_{c,k} + 1)$. This indicates that when $q \diamond \exp$ is Schur-concave, the optimal source precoding matrix (22) along with $\mathbf{V}_{F_1} = \mathbf{I}_{N_b}$, the optimal relay amplifying matrices (22), and the optimal feed-forward matrix at the destination (34) jointly diagonalize the multi-hop MIMO relay channel. From (11), we find that the feedback matrix $\mathbf{B} = \mathbf{0}_{N_b\times N_b}$. Thus, the nonlinear MMSE-DFE receiver is identical to a linear MMSE receiver.

We would like to mention that the capacity of a MIMO relay channel without any constraints on the relay strategy at each node is a challenging open problem. However, under the constraint of using the *linear non-regenerative* strategy at all relay nodes, the mutual information between source and destination is given by $I(\mathbf{y}_{L+1}, \mathbf{s}) = \log_2|\mathbf{I}_{N_b} + \bar{\mathbf{H}}^H\mathbf{C}_{\bar{v}}^{-1}\mathbf{H}| =$

$-\log_2\left(\prod_{k=1}^{N_b}[\mathbf{D}_R]_{k,k}^{-2}\right)$. Thus, under such constraint, the capacity of the MIMO relay channel is given by maximizing $-\log_2\left(\prod_{k=1}^{N_b}[\mathbf{D}_R]_{k,k}^{-2}\right)$, which is equivalent to minimizing $\prod_{k=1}^{N_b}[\mathbf{D}_R]_{k,k}^{-2}$. Since the latter function is a Schur-concave composite objective function, the optimal source and relay matrices are given by Theorem 2 and Theorem 3. Interestingly, for Schur-concave composite objective functions, the nonlinear MMSE-DFE receiver is identical to a linear MMSE receiver. Thus, the capacity can be achieved by both receivers.

As we mentioned before, for all Schur-concave $q \diamond \exp$, q is also Schur-concave. Thus, the objective functions encompassed in Theorem 3 are in a subset of the Schur-concave objective functions discussed in [12]. For Schur-convex $q\diamond\exp$, the following theorem is in order.

THEOREM 4: For the relay design problem (23)-(26), if $q \diamond \exp$ is increasing and Schur-convex, then the optimal $\mathbf{V}_{F_1}$ is chosen such that $\mathbf{d}[\mathbf{D}_R]$ has equal entries, and $\langle\lambda_{i,k}\rangle$ can be obtained from

$$\min_{\langle\lambda_{i,k}\rangle} -\sum_{k=1}^{M}\log\left(1+\frac{\prod_{i=1}^{L}\sigma_{i,k}^2\lambda_{i,k}^2}{1+\sum_{i=2}^{L}\prod_{l=i}^{L}\sigma_{l,k}^2\lambda_{l,k}^2}\right) \quad (35)$$

$$\text{s.t.} \sum_{k=1}^{M}\lambda_{1,k}^2 \le p_1 \quad (36)$$

$$\sum_{k=1}^{M}\lambda_{i,k}^2\left(\sum_{j=1}^{i-1}\prod_{l=j}^{i-1}\lambda_{l,k}^2\sigma_{l,k}^2+1\right)\le p_i, \quad i=2,\cdots,L. \quad (37)$$

PROOF: See Appendix E. □

From (35)-(37) we find that with Schur-convex $q\diamond\exp$, the problem of optimizing $\langle\lambda_{i,k}\rangle$ is identical to the optimization of the power loading in a multi-hop MIMO relay system using the linear MMSE receiver at the destination and the maximal MI criterion [12]. It can be seen from Appendix E that for Schur-convex $q\diamond\exp$, $\mathbf{d}[\mathbf{D}_R]$ has identical elements. Thus, the multi-hop MIMO relay channel is uniformly decomposed into N_b identical subchannels. In contrast to the case of Schur-concave $q \diamond \exp$, there is no constraint on N_b. In fact, N_b can be greater than R_h for Schur-convex $q\diamond\exp$ as explained later. Note that for single-hop MIMO channel, such equal-diagonal QR decomposition has been addressed in [20], [21]. An important observation is that the multi-hop MIMO relay channel is *not* diagonalized by the source precoding matrix, the relay amplifying matrices (22) and the feed-forward matrix as in the case of Schur-concave $q \diamond \exp$. We explain it in the following.

Substituting (27) and (28) back into (13) we have

$$\begin{aligned}\mathbf{w}_k &= \Big(\mathbf{U}_{L,1}\mathbf{D}_h[\mathbf{V}_{F_1}^H]_{1:k}[\mathbf{V}_{F_1}^H]_{1:k}^H\mathbf{D}_h\mathbf{U}_{L,1}^H \\ &\quad +\mathbf{U}_{L,1}\mathbf{D}_c\mathbf{U}_{L,1}^H+\mathbf{I}_{N_{L+1}}\Big)^{-1}\mathbf{U}_{L,1}\mathbf{D}_h[\mathbf{V}_{F_1}^H]_k \\ &= \mathbf{U}_{L,1}\Big(\mathbf{D}_h[\mathbf{V}_{F_1}^H]_{1:k}[\mathbf{V}_{F_1}^H]_{1:k}^H\mathbf{D}_h+\mathbf{D}_c+\mathbf{I}_M\Big)^{-1} \\ &\quad \times\mathbf{D}_h[\mathbf{V}_{F_1}^H]_k \\ \mathbf{w}_k^H[\bar{\mathbf{H}}]_l &= [\mathbf{V}_{F_1}^H]_k^H\mathbf{D}_h\Big(\mathbf{D}_h[\mathbf{V}_{F_1}^H]_{1:k}[\mathbf{V}_{F_1}^H]_{1:k}^H\mathbf{D}_h \\ &\quad +\mathbf{D}_c+\mathbf{I}_M\Big)^{-1}\mathbf{D}_h[\mathbf{V}_{F_1}^H]_l. \end{aligned} \quad (38)$$

We find from (38) that due to $\mathbf{V}_{F_1} \ne [\mathbf{I}_M, \mathbf{0}_{M\times(N_b-M)}]^T$, $\mathbf{w}_k^H[\bar{\mathbf{H}}]_l \ne 0$. Consequently, $\mathbf{W}^H\bar{\mathbf{H}}$ is not diagonal and from (11) $\mathbf{B} \ne \mathbf{0}_{N_b\times N_b}$. This is the reason that more data streams N_b can be supported than the number of subchannels R_h. Thus, in this case, the MIMO relay system with nonlinear MMSE-DFE receiver at the destination has a different performance compared with the MIMO relay system using the linear MMSE receiver at the destination. In the next section, we will show through numerical simulations that the former system has a much better performance than the latter one. Note that since for all Schur-convex q, $q\diamond\exp$ is also Schur-convex, the Schur-convex objective functions discussed in [12] are in a subset of the objective functions included in Theorem 4.

After the optimal structure of the source and relay matrices is determined, the remaining problem is to optimize the power loading vector $\langle\lambda_{i,k}\rangle$ by solving the problem (29)-(31) and the problem (35)-(37) for Schur-concave and Schur-convex $q \diamond \exp$, respectively. Unfortunately, for $L \ge 2$, both the problem (29)-(31) and the problem (35)-(37) are nonconvex. Hence, a globally optimal solution is difficult to obtain with affordable computational complexity, especially when L is large. However, a locally optimal solution can be obtained by iteratively updating the power allocation vector of one node by fixing the power allocation vectors of all other nodes [12]. This iterative algorithm provides an excellent performance-complexity tradeoff. Finally, for the MIMO relay design problem with Schur-convex $q \diamond \exp$ (35)-(37), matrix $\mathbf{V}_{F_1}$ can be computed using the numerical method developed in [18].

IV. NUMERICAL EXAMPLES

In this section, we study the performance of the proposed non-regenerative MIMO relay technique with the nonlinear MMSE-DFE receiver at the destination through numerical simulations. All channel matrices have Gaussian entries with zero-mean and variances σ_i^2/N_i for $\mathbf{H}_i$, $i=1,\cdots,L$. Consequently, we define $\mathrm{SNR}_i \triangleq \sigma_i^2 p_i N_{i+1}/N_i$ as the signal-to-noise ratio (SNR) of the ith hop. All simulation results are averaged over 1000 independent channel realizations.

We compare the relay algorithm for Schur-convex $q \diamond \exp$ (denoted as the NL algorithm) with the maximal MI (MMI) algorithm and the MMSE algorithm developed in [12]. The MMI algorithm refers to a multi-hop MIMO relay system with a linear MMSE receiver at the destination node, and the source and relay matrices designed to maximize the source-destination mutual information. Thus, both the MMI and the MMSE algorithms apply the linear MMSE receiver at the destination. The comparison is fair since the MMI and MMSE relay algorithms in [12] and the proposed NL algorithm require the same amount of channel information at each node. Moreover, all algorithms use the iterative approach to obtain the optimal power allocation vectors, and thus, have the same computational complexity order. As a benchmark, we also show the performance of the genie-aided NL algorithm, where at each layer the error propagation is eliminated by a genie.

In the first example, we simulate a relay system with $L = 2$ hops and choose $N_1 = N_2 = N_3 = 3$ and $N_b = 2$. Fig. 1 shows BERs of all algorithms versus SNR_1 for SNR_2=20dB. It can be seen that the proposed NL algorithm

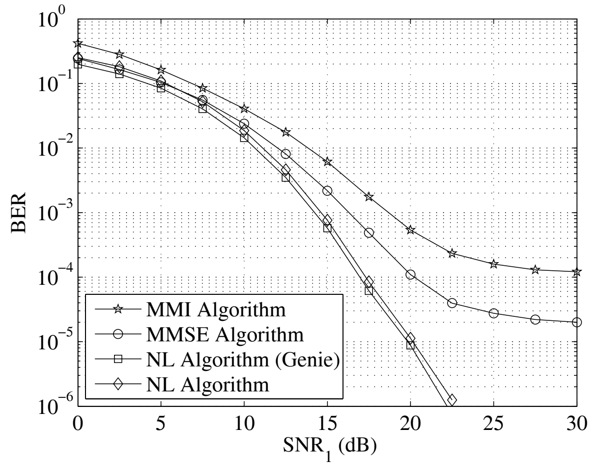


Fig. 1. Example 1: Two hops. $N_1 = N_2 = N_3 = 3$, $N_b = 2$, $\mathrm{SNR}_2 =$ 20dB.

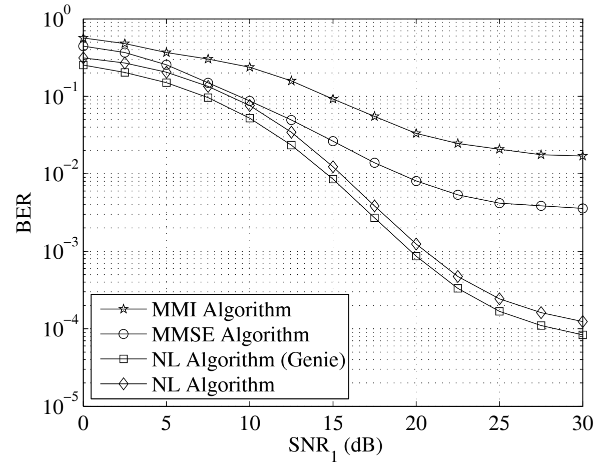


Fig. 3. Example 3: Two hops. $N_1 = 5$, $N_2 = 6$, $N_3 = 4$, $N_b = 4$, SNR_2 = 20dB.

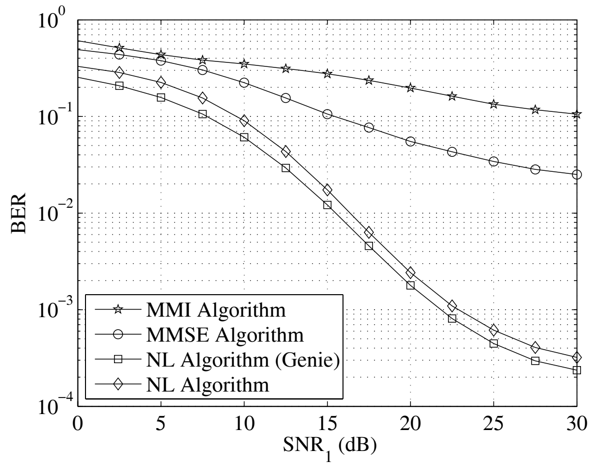


Fig. 2. Example 2: Two hops. $N_1 = N_2 = N_3 = 3$, $N_b = 3$, $\mathrm{SNR}_2 =$ 20dB.

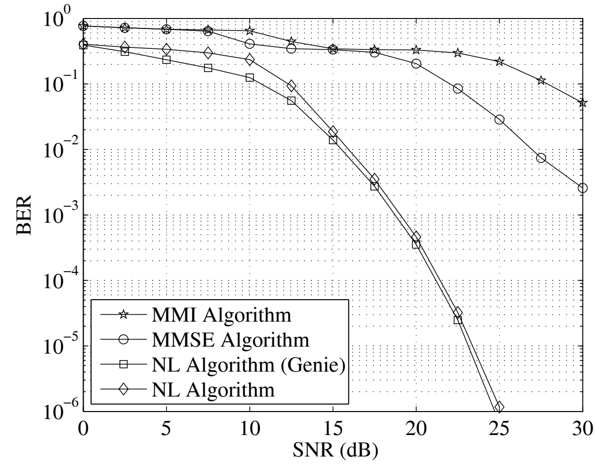


Fig. 4. Example 4: Five hops. $N_i = 3$, $i = 1, \cdots, 6$, SNR_i=SNR, $i = 1, \cdots, 5$.

performs consistently better than the competing algorithms over the whole SNR_1 range. In particular, at high SNR_1, the MMSE and MI algorithms display an error floor, while the NL algorithm does not.

In the second example, we simulate a fully loaded two-hop system with $N_1 = N_2 = N_3 = 3$, $N_b = 3$, and SNR_2=20dB. The performance of all algorithms is shown in Fig. 2. We see that for a fully loaded MIMO relay system, nonlinear MMSE-DFE receiver achieves a much higher diversity order than the linear receiver. The performance of all algorithms for a two-hop system with different number of antennas at each node is studied in our third example. We set $N_1 = 5$, $N_2 = 6$, $N_3 = 4$, $N_b = 4$, and SNR_2=20dB. It can be seen from Fig. 3 that similar to Figs. 1 and 2, the NL algorithm significantly outperforms the MMSE and MI algorithms.

In the last example, a multi-hop MIMO relay system with $L = 5$ and $N_i = 3$, $i = 1, \cdots, 6$ is simulated. Each hop has the same SNR, i.e, $\mathrm{SNR}_i = \mathrm{SNR}$, $i = 1, \cdots, 5$. Fig. 4 displays the BER performance of all three algorithms versus SNR. Obviously, for multi-hop systems, the NL algorithm still has the best performance. Therefore, for multi-hop MIMO relay communication, the nonlinear MMSE-DFE receiver together with the optimal source and relay matrices developed in this paper should be used in order to achieve a lower system BER. Comparing the practical NL algorithm with the genie-aided NL algorithm, we find from Figs. 1-4 that the propagation of the detection error at each layer only slightly increases the system BER.

V. Conclusions

We have developed the optimal source and relay matrices for non-regenerative multi-hop MIMO relay systems when the nonlinear MMSE-DFE receiver is used at the destination. For Schur-convex composite objective functions, the optimal scheme uniformly decomposes the MIMO relay channel into identical subchannels without any constraints on the number of subchannels. We have shown that the system BER performance can be significantly improved by using the MMSE-DFE receiver. We also demonstrated that for Schur-concave composite objective functions, the MMSE-DFE receiver is equivalent to a linear MMSE receiver.

APPENDIX A
SUMMARY OF NOTATIONS

$(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$	Matrix transpose, Hermitian transpose and inversion, respectively
$\mathrm{tr}(\cdot)$, $\mathrm{rank}(\cdot)$	Matrix trace and rank, respectively
$[\mathbf{A}]_k$, $[\mathbf{A}]_{k,k}$	kth column vector of $\mathbf{A}$ and (k,k)-th element of $\mathbf{A}$, respectively
$[\mathbf{A}]_{1:k}$	Matrix containing the first k columns of $\mathbf{A}$
$[\mathbf{A}]_{1:k,1:k}$	Principal sub-matrix of $\mathbf{A}$ lying in the first k rows and the first k columns of $\mathbf{A}$
$\mathbf{d}[\mathbf{A}]$	Column vector containing all main diagonal elements of $\mathbf{A}$
$\mathbf{D}_A$	Matrix taking the diagonal elements of $\mathbf{A}$ as the main diagonal and zero elsewhere
$\mathcal{U}[\mathbf{A}]$	Strictly upper-triangular part of $\mathbf{A}$
$\boldsymbol{\sigma}_A$	Column vector containing the singular values of $\mathbf{A}$
$[\mathbf{a}]_{1:k}$	Vector containing the first k elements of $\mathbf{a}$
$\underline{\mathbf{a}}$	Vector with identical elements of $\sum_{i=1}^N a_i/N$
$\{\mathbf{A}_i\}$	Stands for a set of L matrices $\{\mathbf{A}_i, i = 1,\cdots,L\}$
$\{a\}_n$	Stands for an $n\times 1$ vector $[a_1,a_2,\cdots,a_n]^T$
$\langle a_{i,k}\rangle$	Stands for a set of ML numbers $\{a_{i,k}, 1\le i\le L, 1\le k\le M\}$
$q(\mathbf{x})$	Unified objective function which is increasing with respect to each element of $\mathbf{x}$
$q\diamond\exp(\mathbf{x})$	Composite function defined as $q\left(e^{x_1},e^{x_2},\cdots,e^{x_{N_b}}\right)$
$\prec_+$, $\prec_{+(\mathrm{w})}$	Additive majorization and weakly additive submajorization, respectively
$\prec_\times$, $\prec_{\times(\mathrm{w})}$	Multiplicative majorization and weakly multiplicative submajorization, respectively
$\bigotimes_{i=l}^k(\mathbf{A}_i)$	Stands for the matrix multiplication of $\mathbf{A}_l\cdots\mathbf{A}_k$
$\mathrm{E}[\cdot]$	Statistical expectation
$\mathbf{I}_n$, $\mathbf{1}_n$	$n\times n$ identity matrix and $1\times n$ vector with all 1 elements, respectively
$\mathbf{0}_{p\times m}$	$p\times m$ matrix with all zero entries

APPENDIX B
PROOF OF THEOREM 1

Applying [26, Theorem 2.6.1], which uses the matrix inversion lemma, we have

$$\left([\bar{\mathbf{H}}]_{1:k}[\bar{\mathbf{H}}]_{1:k}^H+\mathbf{C}_{\bar{v}}\right)^{-1}[\bar{\mathbf{H}}]_{1:k} = \mathbf{C}_{\bar{v}}^{-1}[\bar{\mathbf{H}}]_{1:k}\left([\bar{\mathbf{H}}]_{1:k}^H\mathbf{C}_{\bar{v}}^{-1}[\bar{\mathbf{H}}]_{1:k}+\mathbf{I}_k\right)^{-1}.$$

Thus, (13) can be rewritten as

$$\mathbf{w}_k = \left[\mathbf{C}_{\bar{v}}^{-1}[\bar{\mathbf{H}}]_{1:k}\left([\bar{\mathbf{H}}]_{1:k}^H\mathbf{C}_{\bar{v}}^{-1}[\bar{\mathbf{H}}]_{1:k}+\mathbf{I}_k\right)^{-1}\right]_k \quad k=1,\cdots,N_b \qquad (39)$$

where $[\mathbf{A}]_k$ denotes the kth column of matrix $\mathbf{A}$. From (14) and (15), we have

$$[\mathbf{G}]_{1:k} = [\mathbf{Q}]_{1:k}[\mathbf{R}]_{1:k,1:k},\ \ \mathbf{C}_{\bar{v}}^{-\frac{1}{2}}[\bar{\mathbf{H}}]_{1:k} = [\bar{\mathbf{Q}}]_{1:k}[\mathbf{R}]_{1:k,1:k}. \qquad (40)$$

Substituting (40) back into (39) we have

$$\begin{aligned}\mathbf{w}_k &= \Big[\mathbf{C}_{\bar{v}}^{-\frac{1}{2}}[\bar{\mathbf{Q}}]_{1:k}[\mathbf{R}]_{1:k,1:k}\\ &\quad\times\left([\mathbf{R}]_{1:k,1:k}^H[\mathbf{Q}]_{1:k}^H[\mathbf{Q}]_{1:k}[\mathbf{R}]_{1:k,1:k}\right)^{-1}\Big]_k\\ &= \left[\mathbf{C}_{\bar{v}}^{-\frac{1}{2}}[\bar{\mathbf{Q}}]_{1:k}([\mathbf{R}]_{1:k,1:k})^{-H}\right]_k\\ &= [\mathbf{R}]_{k,k}^{-1}\mathbf{C}_{\bar{v}}^{-\frac{1}{2}}[\bar{\mathbf{Q}}]_k.\end{aligned}$$

Thus the feed-forward matrix can be written as $\mathbf{W} = \mathbf{C}_{\bar{v}}^{-\frac{1}{2}}\bar{\mathbf{Q}}\mathbf{D}_R^{-1}$. Using (15) and $\bar{\mathbf{Q}}^H\bar{\mathbf{Q}} = \mathbf{I}_{N_b} - \underline{\mathbf{Q}}^H\underline{\mathbf{Q}}$, we obtain

$$\begin{aligned}\mathbf{W}^H\bar{\mathbf{H}} &= \mathbf{D}_R^{-1}\bar{\mathbf{Q}}^H\mathbf{C}_{\bar{v}}^{-\frac{1}{2}}\bar{\mathbf{H}}\\ &= \mathbf{D}_R^{-1}\bar{\mathbf{Q}}^H\bar{\mathbf{Q}}\mathbf{R}\\ &= \mathbf{D}_R^{-1}(\mathbf{I}_{N_b}-\mathbf{R}^{-H}\mathbf{R}^{-1})\mathbf{R}\\ &= \mathbf{D}_R^{-1}\mathbf{R}-\mathbf{D}_R^{-1}\mathbf{R}^{-H}.\end{aligned} \qquad (41)$$

From (11), the feedback matrix $\mathbf{B}$ can be written as

$$\mathbf{B} = \mathbf{D}_R^{-1}\mathbf{R}-\mathbf{I}_{N_b}. \qquad (42)$$

Substituting (41) and (42) into (8), we have

$$\hat{\mathbf{s}}-\mathbf{s} = -\mathbf{D}_R^{-1}\mathbf{R}^{-H}\mathbf{s}+\mathbf{D}_R^{-1}\bar{\mathbf{Q}}^H\mathbf{C}_{\bar{v}}^{-\frac{1}{2}}\bar{\mathbf{v}}.$$

Therefore, the MSE matrix is

$$\begin{aligned}\mathbf{E} &= \mathrm{E}\left[(\hat{\mathbf{s}}-\mathbf{s})(\hat{\mathbf{s}}-\mathbf{s})^H\right]\\ &= \mathbf{D}_R^{-1}\mathbf{R}^{-H}\mathbf{R}^{-1}\mathbf{D}_R^{-1}+\mathbf{D}_R^{-1}\bar{\mathbf{Q}}^H\bar{\mathbf{Q}}\mathbf{D}_R^{-1}\\ &= \mathbf{D}_R^{-2}.\end{aligned}$$

APPENDIX C
PROOF OF THEOREM 2

The following lemmas are required to prove Theorem 2.

LEMMA 1 [17, 5.A.1]: For two $N\times 1$ vectors $\mathbf{x}$ and $\mathbf{y}$, if $\mathbf{x}\prec_+\mathbf{y}$ and f is a convex function, then $\{f(x_i)\}_N \prec_{+(\mathrm{w})} \{f(y_i)\}_N$.

LEMMA 2 [17, 9.H.1.b]: For m $N\times N$ complex matrices $\mathbf{A}_1,\mathbf{A}_2,\cdots,\mathbf{A}_m$, let $\mathbf{B} = \bigotimes_{i=1}^m\mathbf{A}_i$, then $\boldsymbol{\sigma}_b \prec_\times (\boldsymbol{\sigma}_{a_1}\odot\boldsymbol{\sigma}_{a_2}\odot\cdots\odot\boldsymbol{\sigma}_{a_m})$, where $\boldsymbol{\sigma}_b$, and $\boldsymbol{\sigma}_{a_i}$, $i=1,\cdots,m$, denote $N\times 1$ vectors containing the singular values of $\mathbf{B}$ and $\mathbf{A}_i$ arranged in the same order, respectively, and $\odot$ denotes the Schur (element-wise) product of two vectors.

LEMMA 3 [17, 9.H.1.h]: For two $N\times N$ positive semidefinite matrices $\mathbf{A}$ and $\mathbf{B}$ with eigenvalues $\lambda_{a,i}$ and $\lambda_{b,i}$, $i=1,\cdots,N$, arranged in the same order, respectively, it follows that $\mathrm{tr}(\mathbf{AB})\ge\sum_{i=1}^N\lambda_{a,i}\lambda_{b,N+1-i}$.

LEMMA 4: For two $N\times 1$ vectors $\mathbf{x}$ and $\mathbf{y}$, if $\mathbf{x}\prec_\times\mathbf{y}$, then $\{(1-x_i)^{-1}\}_N \prec_{\times(\mathrm{w})} \{(1-y_i)^{-1}\}_N$.

PROOF: Let us define two vectors $\mathbf{a}$ and $\mathbf{b}$ with $a_i=\log x_i$ and $b_i=\log y_i$, $i=1,\cdots,N$. Obviously, $\mathbf{x}\prec_\times\mathbf{y}$ is equivalent to $\mathbf{a}\prec_+\mathbf{b}$. Since $f(x)=-\log(1-e^x)$ is a convex function, it follows from Lemma 1 that $\{-\log(1-e^{a_i})\}_N \prec_{+(\mathrm{w})} \{-\log(1-e^{b_i})\}_N$, which is equivalent to $\{(1-e^{a_i})^{-1}\}_N \prec_{\times(\mathrm{w})} \{(1-e^{b_i})^{-1}\}_N$. Thus we have $\{(1-x_i)^{-1}\}_N \prec_{\times(\mathrm{w})} \{(1-y_i)^{-1}\}_N$. □

Now we set out to prove Theorem 2. The proof is conducted in three steps: First, we show that the constraint (18) is equivalent to $\mathbf{d}[\mathbf{D}_R]\prec_\times\boldsymbol{\sigma}_G$ where $\boldsymbol{\sigma}_G$ is a column vector containing singular values of $\mathbf{G}$. Second, we prove $\boldsymbol{\sigma}_G \prec_{\times(\mathrm{w})}$

$\left[\left\{1+\frac{\prod_{i=1}^{L}\sigma_{i,k}^{2}\lambda_{i,k}^{2}}{1+\sum_{i=2}^{L}\prod_{l=i}^{L}\sigma_{l,k}^{2}\lambda_{l,k}^{2}}\right\}_{M}^{T},\mathbf{1}_{N_b-M}\right]^{T}$. At last, we show the structure of the optimal source and relay matrices.

Let us introduce the SVD $\mathbf{F}_1 = \tilde{\mathbf{U}}_{F_1}\tilde{\mathbf{\Lambda}}_{F_1}\tilde{\mathbf{V}}_{F_1}^H$, where the dimensions of $\tilde{\mathbf{U}}_{F_1}$, $\tilde{\mathbf{\Lambda}}_{F_1}$, $\tilde{\mathbf{V}}_{F_1}$ are $N_1 \times N_1$, $N_1 \times N_b$, $N_b \times N_b$, respectively. We assume that the main diagonal elements of $\tilde{\mathbf{\Lambda}}_{F_1}$ are arranged in the decreasing order. Since $\text{rank}(\mathbf{F}_1) = M$, we also have $\mathbf{F}_1 = \mathbf{U}_{F_1}\mathbf{\Lambda}_{F_1}\mathbf{V}_{F_1}^H$, where $\mathbf{U}_{F_1} = [\tilde{\mathbf{U}}_{F_1}]_{1:M}$, $\mathbf{V}_{F_1} = [\tilde{\mathbf{V}}_{F_1}]_{1:M}$, $\mathbf{\Lambda}_{F_1} = [\tilde{\mathbf{\Lambda}}_{F_1}]_{1:M,1:M}$. From (14) we have

$$\mathbf{G} = \begin{bmatrix} \mathbf{C}_{\bar{v}}^{-\frac{1}{2}}\tilde{\mathbf{H}}\tilde{\mathbf{U}}_{F_1}\tilde{\mathbf{\Lambda}}_{F_1}\tilde{\mathbf{V}}_{F_1}^H \\ \mathbf{I}_{N_b} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{N_{L+1}} & \mathbf{0}_{N_{L+1}\times N_b} \\ \mathbf{0}_{N_b\times N_{L+1}} & \tilde{\mathbf{V}}_{F_1} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{\bar{v}}^{-\frac{1}{2}}\tilde{\mathbf{H}}\tilde{\mathbf{U}}_{F_1}\tilde{\mathbf{\Lambda}}_{F_1} \\ \mathbf{I}_{N_b} \end{bmatrix} \tilde{\mathbf{V}}_{F_1}^H$$

where $\tilde{\mathbf{H}} = \bigotimes_{i=L}^{2}(\mathbf{H}_i\mathbf{F}_i)\mathbf{H}_1$. Let us write the GTD of $\mathbf{\Psi}$ as

$$\mathbf{\Psi} \triangleq \begin{bmatrix} \mathbf{C}_{\bar{v}}^{-\frac{1}{2}}\tilde{\mathbf{H}}\tilde{\mathbf{U}}_{F_1}\tilde{\mathbf{\Lambda}}_{F_1} \\ \mathbf{I}_{N_b} \end{bmatrix} = \mathbf{Q}_\Psi\mathbf{R}\mathbf{P}_\Psi^H \tag{43}$$

where $\mathbf{Q}_\Psi$ is an $(N_{L+1}+N_b)\times N_b$ semi-unitary matrix with $\mathbf{Q}_\Psi^H\mathbf{Q}_\Psi = \mathbf{I}_{N_b}$, and $\mathbf{P}_\Psi$ is an $N_b\times N_b$ unitary matrix. It can be shown from [18] that (43) holds *if and only if* $\mathbf{d}[\mathbf{D}_R] \prec_\times \boldsymbol{\sigma}_\Psi$, where $\boldsymbol{\sigma}_\Psi$ is a column vector containing singular values of $\mathbf{\Psi}$. Without affecting the power constraints, we take $\tilde{\mathbf{V}}_{F_1} = \mathbf{P}_\Psi^H$, or equivalently

$$\mathbf{V}_{F_1} = [\mathbf{P}_\Psi^H]_{1:M}. \tag{44}$$

Then we can write the QR decomposition of $\mathbf{G}$ as

$$\mathbf{G} = \mathbf{Q}\mathbf{R}, \qquad \mathbf{Q} = \begin{bmatrix} \mathbf{I}_{N_{L+1}} & \mathbf{0}_{N_{L+1}\times N_b} \\ \mathbf{0}_{N_b\times N_{L+1}} & \mathbf{P}_\Psi^H \end{bmatrix}\mathbf{Q}_\Psi. \tag{45}$$

Because $\mathbf{\Psi}$ and $\mathbf{G}$ have the same singular values, from (43) and (45) we know that the constraint (18) can be equivalently written as

$$\mathbf{d}[\mathbf{D}_R] \prec_\times \boldsymbol{\sigma}_G. \tag{46}$$

Let us define

$$\mathbf{X} \triangleq \bar{\mathbf{H}}^H\left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H + \mathbf{C}_{\bar{v}}\right)^{-1}\bar{\mathbf{H}} \tag{47}$$

$$\mathbf{A}_1 \triangleq \mathbf{H}_1\mathbf{F}_1\mathbf{F}_1^H\mathbf{H}_1^H \tag{48}$$

$$\mathbf{A}_i \triangleq \mathbf{H}_i\mathbf{F}_i(\mathbf{A}_{i-1}+\mathbf{I}_{N_i})\mathbf{F}_i^H\mathbf{H}_i^H, \quad i = 2,\cdots,L \tag{49}$$

and write $\mathbf{A}_i = \mathbf{U}_{A_i}\mathbf{\Lambda}_{A_i}\mathbf{U}_{A_i}^H, i = 1,\cdots,L$, as the eigen-decomposition of $\mathbf{A}_i$, where $\mathbf{\Lambda}_{A_i}$ is an $M\times M$ diagonal matrix containing M nonzero eigenvalues of $\mathbf{A}_i$ sorted in the decreasing order for all i, and $\mathbf{U}_{A_i}$ is the associated $N_{i+1}\times M$ matrix of eigenvectors. From (48) and (49) we have

$$\mathbf{H}_1\mathbf{F}_1 = \mathbf{U}_{A_1}\mathbf{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_0 \tag{50}$$

$$\mathbf{H}_i\mathbf{F}_i = \mathbf{U}_{A_i}\mathbf{\Lambda}_{A_i}^{\frac{1}{2}}\mathbf{S}_i(\mathbf{A}_{i-1}+\mathbf{I}_{N_i})^{-\frac{1}{2}}, \quad i = 2,\cdots,L \tag{51}$$

where $\mathbf{U}_0$ is an $M\times N_b$ semi-unitary matrices with $\mathbf{U}_0\mathbf{U}_0^H = \mathbf{I}_M$, and $\mathbf{S}_i$, $i = 2,\cdots,L$, are $M\times N_i$ semi-unitary matrices with $\mathbf{S}_i\mathbf{S}_i^H = \mathbf{I}_M$. It will be seen that the power constraints (19) and (20) are invariant to $\mathbf{U}_0$ and $\mathbf{S}_i$, $i = 2,\cdots,L$. Substituting (50) and (51) back into (47), we have

$$\begin{aligned}\mathbf{X} &= \mathbf{U}_0^H\mathbf{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_{A_1}^H\bigotimes_{i=2}^{L}\left((\mathbf{A}_{i-1}+\mathbf{I}_{N_i})^{-\frac{1}{2}}\mathbf{S}_i^H\mathbf{\Lambda}_{A_i}^{\frac{1}{2}}\mathbf{U}_{A_i}^H\right) \\ &\quad\times(\mathbf{A}_L+\mathbf{I}_{N_{L+1}})^{-1}\bigotimes_{i=L}^{2}\left(\mathbf{U}_{A_i}\mathbf{\Lambda}_{A_i}^{\frac{1}{2}}\mathbf{S}_i(\mathbf{A}_{i-1}+\mathbf{I}_{N_i})^{-\frac{1}{2}}\right) \\ &\quad\times\mathbf{U}_{A_1}\mathbf{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_0 \\ &= \mathbf{U}_0^H\mathbf{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_{A_1}^H\bigotimes_{i=2}^{L}\left(\mathbf{U}_{A_{i-1}}(\mathbf{\Lambda}_{A_{i-1}}+\mathbf{I}_M)^{-\frac{1}{2}}\mathbf{U}_{A_{i-1}}^H\mathbf{S}_i^H\right. \\ &\quad\left.\times\mathbf{\Lambda}_{A_i}^{\frac{1}{2}}\mathbf{U}_{A_i}^H\right)\mathbf{U}_{A_L}(\mathbf{\Lambda}_{A_L}+\mathbf{I}_M)^{-1}\mathbf{U}_{A_L}^H\bigotimes_{i=L}^{2}\left(\mathbf{U}_{A_i}\mathbf{\Lambda}_{A_i}^{\frac{1}{2}}\right. \\ &\quad\left.\times\mathbf{S}_i\mathbf{U}_{A_{i-1}}(\mathbf{\Lambda}_{A_{i-1}}+\mathbf{I}_M)^{-\frac{1}{2}}\mathbf{U}_{A_{i-1}}^H\right)\mathbf{U}_{A_1}\mathbf{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_0. \end{aligned}\tag{52}$$

Applying Lemma 2 to (52) we obtain that

$$\boldsymbol{\lambda}_X \prec_\times \left[\left(\mathbf{d}[\tilde{\mathbf{X}}]\right)^T, \mathbf{0}_{1\times(N_b-M)}\right]^T \tag{53}$$

where $\boldsymbol{\lambda}_X$ is a column vector containing all eigenvalues of $\mathbf{X}$, and $\tilde{\mathbf{X}}$ is a diagonal matrix given by

$$\begin{aligned}\tilde{\mathbf{X}} &\triangleq \mathbf{\Lambda}_{A_1}^{\frac{1}{2}}\bigotimes_{i=2}^{L}\left((\mathbf{\Lambda}_{A_{i-1}}+\mathbf{I}_M)^{-\frac{1}{2}}\mathbf{\Lambda}_{A_i}^{\frac{1}{2}}\right)(\mathbf{\Lambda}_{A_L}+\mathbf{I}_M)^{-1} \\ &\quad\bigotimes_{i=L}^{2}\left(\mathbf{\Lambda}_{A_i}^{\frac{1}{2}}(\mathbf{\Lambda}_{A_{i-1}}+\mathbf{I}_M)^{-\frac{1}{2}}\right)\mathbf{\Lambda}_{A_1}^{\frac{1}{2}} \\ &= \bigotimes_{i=1}^{L}(\mathbf{\Lambda}_{A_i}(\mathbf{\Lambda}_{A_i}+\mathbf{I}_M)^{-1}). \end{aligned}\tag{54}$$

Applying Lemma 4 to (53), we have

$$\{(1-\lambda_{X,i})^{-1}\}_{N_b} \prec_{\times(\text{w})} \left[\left\{(1-\tilde{x}_{i,i})^{-1}\right\}_M^T, \mathbf{1}_{N_b-M}\right]^T \tag{55}$$

where $\lambda_{X,i}$ denotes the ith eigenvalue of $\mathbf{X}$ and $\tilde{x}_{i,i}$ denotes the (i,i)-th element of $\tilde{\mathbf{X}}$. From (14), we can write

$$\begin{aligned}\mathbf{G}^H\mathbf{G} &= \mathbf{I}_{N_b} + \bar{\mathbf{H}}^H\mathbf{C}_{\bar{v}}^{-1}\bar{\mathbf{H}} \\ &= \left[\mathbf{I}_{N_b} - \bar{\mathbf{H}}^H\left(\bar{\mathbf{H}}^H\bar{\mathbf{H}}+\mathbf{C}_{\bar{v}}\right)^{-1}\bar{\mathbf{H}}\right]^{-1} \\ &= (\mathbf{I}_{N_b}-\mathbf{X})^{-1} \end{aligned}\tag{56}$$

where the matrix inversion lemma is applied to obtain the second equation. From (56), we find that $\{\sigma_{G,i}^2\}_{N_b} = \{(1-\lambda_{X,i})^{-1}\}_{N_b}$, where $\sigma_{G,i}$ is the ith singular value of $\mathbf{G}$. Using (55) we obtain $\{\sigma_{G,i}^2\}_{N_b} \prec_{\times(\text{w})} \left[\left\{(1-\tilde{x}_{i,i})^{-1}\right\}_M^T, \mathbf{1}_{N_b-M}\right]^T$. Moreover, since (46) is equivalent to $\mathbf{d}[\mathbf{D}_R^2] \prec_\times \{\sigma_{G,i}^2\}_{N_b}$, we have

$$\mathbf{d}[\mathbf{D}_R^2] \prec_{\times(\text{w})} \left[\left\{(1-\tilde{x}_{i,i})^{-1}\right\}_M^T, \mathbf{1}_{N_b-M}\right]^T. \tag{57}$$

We would like to mention that for all $\mathbf{D}_R$ and $\{\mathbf{F}_i\}$ that satisfy (46), inequality (57) also holds. In other words, (57) has a relaxed feasible region than that of (46). Since (46) is equivalent to (18), we can replace the constraint (18) by (57) without increasing the value of the objective function (17). Moreover, from (52) we see that $\boldsymbol{\lambda}_X = \left[\left(\mathbf{d}[\tilde{\mathbf{X}}]\right)^T, \mathbf{0}_{1\times(N_b-M)}\right]^T$ holds at

$$\mathbf{S}_i = \boldsymbol{\Phi}\mathbf{U}_{A_{i-1}}^H, \qquad i = 2, \cdots, L$$

where $\boldsymbol{\Phi}$ stands for an arbitrary $M \times M$ diagonal matrix with unit-norm main diagonal elements, i.e., $|[\boldsymbol{\Phi}]_{i,i}| = 1, [\boldsymbol{\Phi}]_{i,j} = 0, i,j = 1,\cdots,M, i \neq j$. Without affecting the objective function (17), we choose $\mathbf{S}_i = \mathbf{U}_{A_{i-1}}^H$, $i = 2,\cdots,L$.

Now we set out to consider the power constraints (19) and (20). First, we introduce some notations: for $i = 1,\cdots,L$, $r_i \triangleq \text{rank}(\mathbf{H}_i)$, $\hat{\mathbf{F}}_i \triangleq \mathbf{V}_i^H\mathbf{F}_i$, $\mathbf{U}_i \triangleq [\mathbf{U}_{i,r_i}, \mathbf{U}_{i,\bar{r}_i}]$, where $\mathbf{U}_{i,r_i}$ and $\mathbf{U}_{i,\bar{r}_i}$ contain the left singular vectors of $\mathbf{H}_i$ associated with the nonzero and zero singular values of $\mathbf{H}_i$, respectively, $\boldsymbol{\Sigma}_{i,r_i}$ is a diagonal matrix containing the nonzero singular values of $\mathbf{H}_i$, $\boldsymbol{\Sigma}_{i,1}$ contains the largest M singular values of $\mathbf{H}_i$ sorted in the same order as the diagonal elements of $\boldsymbol{\Lambda}_{A_i}$. Substituting the SVD of $\mathbf{H}_1$ in (21) into (50) and left multiplying by $\mathbf{U}_1^H$ on both sides, we have

$$\begin{bmatrix} \boldsymbol{\Sigma}_{1,r_1} & \mathbf{0}_{r_1\times(N_1-r_1)} \\ \mathbf{0}_{(N_2-r_1)\times r_1} & \mathbf{0}_{(N_2-r_1)\times(N_1-r_1)} \end{bmatrix} \hat{\mathbf{F}}_1 = \mathbf{U}_1^H\mathbf{U}_{A_1}\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_0 . \tag{58}$$

If $N_1 = N_2 = r_1$, (58) holds if and only if

$$\hat{\mathbf{F}}_1 = \boldsymbol{\Sigma}_{1,r_1}^{-1}\mathbf{U}_{1,r_1}^H\mathbf{U}_{A_1}\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_0 . \tag{59}$$

If $N_1 > N_2 = r_1$, then (58) holds if and only if

$$\left[\ \boldsymbol{\Sigma}_{1,r_1} \quad \mathbf{0}_{r_1\times(N_1-r_1)}\ \right]\hat{\mathbf{F}}_1 = \mathbf{U}_{1,r_1}^H\mathbf{U}_{A_1}\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_0 . \tag{60}$$

Finally, if $N_1 > r_1, N_2 > r_1$, (58) is true if and only if $\mathbf{U}_{1,\bar{r}_1}^H\mathbf{U}_{A_1} = \mathbf{0}_{(N_2-r_1)\times M}$ and (60) holds. From (60), we see that in the latter two cases, there are many solutions for $\hat{\mathbf{F}}_1$. We should choose $\hat{\mathbf{F}}_1$ such that the transmission power at the source node is minimized. Since $\text{tr}(\mathbf{F}_1\mathbf{F}_1^H) = \text{tr}(\hat{\mathbf{F}}_1\hat{\mathbf{F}}_1^H)$, the transmission power minimization problem is written as

$$\min_{\hat{\mathbf{F}}_1} \ \text{tr}\left(\hat{\mathbf{F}}_1\hat{\mathbf{F}}_1^H\right)$$
$$\text{s.t.} \ \left[\ \boldsymbol{\Sigma}_{1,r_1} \quad \mathbf{0}_{r_1\times(N_1-r_1)}\ \right]\hat{\mathbf{F}}_1 = \mathbf{U}_{1,r_1}^H\mathbf{U}_{A_1}\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_0 . \tag{61}$$

The solution to problem (61) is given by

$$\hat{\mathbf{F}}_1 = \left[\ \boldsymbol{\Sigma}_{1,r_1}^{-1} \quad \mathbf{0}_{r_1\times(N_1-r_1)}\ \right]^T\mathbf{U}_{1,r_1}^H\mathbf{U}_{A_1}\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_0 . \tag{62}$$

To determine $\mathbf{U}_{A_1}$ in (59) and (62), we substitute (59) and (62) into the objective function of (61). Interestingly, both (59) and (62) lead to the same transmission power, given by

$$\text{tr}\left(\mathbf{F}_1\mathbf{F}_1^H\right) = \text{tr}\left(\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_{A_1}^H\mathbf{U}_{1,r_1}\boldsymbol{\Sigma}_{1,r_1}^{-2}\mathbf{U}_{1,r_1}^H\mathbf{U}_{A_1}\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}\right) . \tag{63}$$

We note that the transmission power (63) is invariant to $\mathbf{U}_0$ and $\mathbf{S}_i$, $i = 2,\cdots,L$. Using Lemma 3, we know that under $\text{rank}(\mathbf{F}_1) = M$, (63) is minimized if and only if $\mathbf{U}_{A_1}^H\mathbf{U}_{1,r_1} = [\boldsymbol{\Phi}, \mathbf{0}_{M\times(r_1-M)}]$. The minimum of (63) is $\text{tr}(\boldsymbol{\Lambda}_{A_1}\boldsymbol{\Sigma}_{1,1}^{-2})$. Without loss of generality, we choose $\boldsymbol{\Phi} = \mathbf{I}_M$. Therefore, we have $\mathbf{U}_{A_1} = \mathbf{U}_{1,1}$ and

$$\mathbf{F}_1 = \mathbf{V}_1\left[\boldsymbol{\Sigma}_{1,1}^{-1}\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}, \mathbf{0}_{M\times(N_1-M)}\right]^T\mathbf{U}_0 = \mathbf{V}_{1,1}\boldsymbol{\Sigma}_{1,1}^{-1}\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{U}_0 .$$

Note that $\mathbf{U}_0$ does not affect $\boldsymbol{\lambda}_X$ and the power constraints. In fact, $\mathbf{U}_0$ should be chosen as $\mathbf{V}_{F_1}^H$ in (44) such that the QR decomposition of $\mathbf{G}$ in (45) holds. Thus $\mathbf{F}_1 = \mathbf{V}_{1,1}\boldsymbol{\Sigma}_{1,1}^{-1}\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}\mathbf{V}_{F_1}^H$, and we have proved that the optimal structure of $\mathbf{F}_1$ is as in (22) with $\boldsymbol{\Lambda}_1 = \boldsymbol{\Sigma}_{1,1}^{-1}\boldsymbol{\Lambda}_{A_1}^{\frac{1}{2}}$.

Now we consider the power constraints (20). Similar to steps (58)-(62), for $i = 2,\cdots,L$, we have $\mathbf{U}_{i,\bar{r}_i}^H\mathbf{U}_{A_i} = \mathbf{0}_{(N_{i+1}-r_i)\times M}$ when $N_{i+1} > r_i$ and

$$\hat{\mathbf{F}}_i = \boldsymbol{\Sigma}_{i,r_i}^{-1}\mathbf{U}_{i,r_i}^H\mathbf{U}_{A_i}\boldsymbol{\Lambda}_{A_i}^{\frac{1}{2}}\mathbf{S}_i(\mathbf{A}_{i-1}+\mathbf{I}_{N_i})^{-\frac{1}{2}}, \qquad r_i = N_i$$
$$\hat{\mathbf{F}}_i = \left[\ \boldsymbol{\Sigma}_{i,r_i}^{-1} \quad \mathbf{0}_{r_i\times(N_i-r_i)}\ \right]^T\mathbf{U}_{i,r_i}^H\mathbf{U}_{A_i}\boldsymbol{\Lambda}_{A_i}^{\frac{1}{2}} \times\mathbf{S}_i(\mathbf{A}_{i-1}+\mathbf{I}_{N_i})^{-\frac{1}{2}}, \qquad r_i < N_i$$

where we solved the following problem

$$\min_{\hat{\mathbf{F}}_i} \ \text{tr}\left(\hat{\mathbf{F}}_i\left(\sum_{l=1}^{i-1}\Big(\bigotimes_{n=i-1}^{l}(\mathbf{H}_n\mathbf{F}_n)\bigotimes_{n=l}^{i-1}\left(\mathbf{F}_n^H\mathbf{H}_n^H\right)\Big)+\mathbf{I}_{N_i}\right)\hat{\mathbf{F}}_i^H\right)$$
$$\text{s.t.} \ \left[\ \boldsymbol{\Sigma}_{i,r_i} \quad \mathbf{0}_{r_i\times(N_i-r_i)}\ \right]\hat{\mathbf{F}}_i = \mathbf{U}_{i,r_i}^H\mathbf{U}_{A_i}\boldsymbol{\Lambda}_{A_i}^{\frac{1}{2}}\mathbf{S}_i(\mathbf{A}_{i-1}+\mathbf{I}_{N_i})^{-\frac{1}{2}} .$$

The transmission power at the ith node is

$$\text{tr}\left(\mathbf{F}_i(\mathbf{A}_{i-1}+\mathbf{I}_{N_i})\mathbf{F}_i^H\right) = \text{tr}\left(\boldsymbol{\Sigma}_{i,r_i}^{-1}\mathbf{U}_{i,r_i}^H\mathbf{U}_{A_i}\boldsymbol{\Lambda}_{A_i}\mathbf{U}_{A_i}^H\mathbf{U}_{i,r_i}\boldsymbol{\Sigma}_{i,r_i}^{-1}\right) . \tag{64}$$

Obviously, (64) is also invariant to $\mathbf{U}_0$ and $\mathbf{S}_i$, $i = 2,\cdots,L$. Similar to (63), (64) is minimized by $\mathbf{U}_{A_i} = \mathbf{U}_{i,1}$, and together with $\mathbf{S}_i = \mathbf{U}_{A_{i-1}}^H$, we obtain

$$\mathbf{F}_i = \mathbf{V}_{i,1}\boldsymbol{\Sigma}_{i,1}^{-1}\boldsymbol{\Lambda}_{A_i}^{\frac{1}{2}}(\boldsymbol{\Lambda}_{A_{i-1}}+\mathbf{I}_M)^{-\frac{1}{2}}\mathbf{U}_{i-1,1}^H .$$

Thus, the optimal structure of $\mathbf{F}_i$ is given by (22) with $\boldsymbol{\Lambda}_i = \boldsymbol{\Sigma}_{i,1}^{-1}\boldsymbol{\Lambda}_{A_i}^{\frac{1}{2}}(\boldsymbol{\Lambda}_{A_{i-1}}+\mathbf{I}_M)^{-\frac{1}{2}}$.

Substituting (22) into (48), (49) and (54), we obtain

$$\tilde{x}_{k,k} = \frac{\prod_{i=1}^{L}\sigma_{i,k}^2\lambda_{i,k}^2}{1+\sum_{i=1}^{L}\prod_{l=i}^{L}\sigma_{l,k}^2\lambda_{l,k}^2}, \qquad k = 1,\cdots,M . \tag{65}$$

Substituting (65) into (57) we have (24). Finally, by substituting (22) into (19) and (20), we obtain (25) and (26).

Appendix D
Proof of Theorem 3

The following lemma is needed to prove Theorem 3.

Lemma 5 [17, 3.A.8]: A real-valued function f satisfies $\mathbf{x} \prec_{+(\text{w})} \mathbf{y} \Rightarrow f(\mathbf{x}) \leq f(\mathbf{y})$ if and only if f is increasing and Schur-convex.

Let us define $d_k \triangleq \log([\mathbf{D}_R]_{k,k}^2)$, $k = 1,\cdots,N_b$, and rewrite the problem (23)-(26) as

$$\min_{\{d_k\}_{N_b},\langle\lambda_{i,k}\rangle} \ q\left(\left\{e^{-d_k}\right\}_{N_b}\right) \tag{66}$$
$$\text{s.t.} \ \{d_k\}_{N_b} \prec_{+(\text{w})} \boldsymbol{\eta} \tag{67}$$
$$\sum_{k=1}^{M}\lambda_{1,k}^2 \leq p_1 \tag{68}$$
$$\sum_{k=1}^{M}\lambda_{i,k}^2\left(\sum_{j=1}^{i-1}\prod_{l=j}^{i-1}\lambda_{l,k}^2\sigma_{l,k}^2+1\right) \leq p_i, \quad i = 2,\cdots,L \tag{69}$$

where

$$\boldsymbol{\eta} \triangleq \left[\left\{ \log\left(1 + \frac{\prod_{i=1}^{L} \sigma_{i,k}^2 \lambda_{i,k}^2}{1 + \sum_{i=2}^{L} \prod_{l=i}^{L} \sigma_{l,k}^2 \lambda_{l,k}^2}\right) \right\}_M^T, \mathbf{0}_{1\times(N_b-M)} \right]^T.$$

We know that for any function f, $f(-\mathbf{x})$ has the same Schur-convexity as $f(\mathbf{x})$ with respect to $\mathbf{x}$ [17]. Since $q \diamond \exp$ is Schur-concave, $q\left(\left\{e^{-d_k}\right\}_{N_b}\right)$ is Schur-concave with respect to $\{d_k\}_{N_b}$. Moreover, because $q \diamond \exp$ is increasing, $q\left(\left\{e^{-d_k}\right\}_{N_b}\right)$ is decreasing with respect to $\{d_k\}_{N_b}$. Obviously, $-q\left(\left\{e^{-d_k}\right\}_{N_b}\right)$ is increasing and Schur-convex with respect to $\{d_k\}_{N_b}$. Using Lemma 5 and (67), we have $-q\left(\left\{e^{-d_k}\right\}_{N_b}\right) \leq -q(\boldsymbol{\eta})$, or equivalently $q\left(\left\{e^{-d_k}\right\}_{N_b}\right) \geq q(\boldsymbol{\eta})$. Therefore, the solution to the problem (66)-(69) occurs at $\{d_k\}_{N_b} = \boldsymbol{\eta}$, which holds if and only if $\mathbf{V}_{F_1} = \left[\mathbf{I}_M, \mathbf{0}_{M\times(N_b-M)}\right]^T$. Thus, the problem (66)-(69) can be equivalently written as problem (29)-(31).

Appendix E
Proof of Theorem 4

To prove Theorem 4, we need the following lemma.

Lemma 6 [17, p.7]: For an $N \times 1$ real-valued vector $\mathbf{x}$, let us define an $N \times 1$ vector $\underline{\mathbf{x}}$ with identical elements of $\sum_{i=1}^{N} x_i/N$, there is $\underline{\mathbf{x}} \prec_+ \mathbf{x}$.

Let us define $d_k \triangleq \log([\mathbf{D}_R]_{k,k}^2)$, $k = 1, \cdots, N_b$. Applying Lemma 6 to (24), we obtain that

$$\underline{\mathbf{d}} \prec_+ \{d_k\}_{N_b} \tag{70}$$

where $\underline{\mathbf{d}}$ is an $N_b \times 1$ vector with identical elements of

$$\underline{d}_l = \frac{1}{N_b} \sum_{k=1}^{M} \log\left(1 + \frac{\prod_{i=1}^{L} \sigma_{i,k}^2 \lambda_{i,k}^2}{1 + \sum_{i=2}^{L} \prod_{j=i}^{L} \sigma_{j,k}^2 \lambda_{j,k}^2}\right), \quad l = 1, \cdots, N_b. \tag{71}$$

Since $q \diamond \exp$ is increasing and Schur-convex, applying Definition 3 to (70), we know that the solution to the problem (23)-(26) occurs at (71). It can be shown from [18] that there exists a GTD in (43) such that $\mathbf{R}$ has identical diagonal elements. Equivalently speaking, $\mathbf{V}_{F_1}$ should be chosen such that the QR decomposition of $\mathbf{G}$ in (14) yields a $\mathbf{d}[\mathbf{D}_R]$ with identical elements. Such $\mathbf{V}_{F_1}$ can be obtained by using the numerical method developed in [18].

Using (70) and (71), the objective function for Schur-convex $q \diamond \exp$ becomes

$$\max_{\langle \lambda_{i,k} \rangle} \quad \sum_{k=1}^{M} \log\left(1 + \frac{\prod_{i=1}^{L} \sigma_{i,k}^2 \lambda_{i,k}^2}{1 + \sum_{i=2}^{L} \prod_{l=i}^{L} \sigma_{l,k}^2 \lambda_{l,k}^2}\right)$$

which is equivalent to (35). Along with the power constraints, we obtain the problem (35)-(37).

References

[1] M. Cao, X. Wang, S.-J. Kim, and M. Madihian, "Multi-hop wireless backhaul networks: a cross-layer design paradigm," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 738-748, May 2007.

[2] B. Wang, J. Zhang, and A. Høst-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 29-43, Jan. 2005.

[3] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1398-1407, Apr. 2007.

[4] O. Muñoz-Medina, J. Vidal, and A. Agustín, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, pp. 2593-2604, June 2007.

[5] I. Hammerström and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2798-2802, Aug. 2007.

[6] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, pp. 517-519, July 2008.

[7] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, pp. 5062-5073, Oct. 2008.

[8] S. Yeh and O. Lévêque, "Asymptotic capacity of multi-level amplify-and-forward relay networks," in *Proc. IEEE ISIT*, Nice, France, June 2007, pp. 1436-1440.

[9] N. Fawaz, K. Zarifi, M. Debbah, and D. Gesbert, "Asymptotic capacity and optimal precoding strategy of multi-level precode & forward in correlated channels," in *Proc. IEEE ITW*, Porto, Portugal, May 2008, pp. 209-213.

[10] J. Wagner and A. Wittneben, "On capacity scaling of (long) MIMO amplify-and-forward multi-hop networks," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct. 2008, pp. 346-350.

[11] S. Yang and J.-C. Belfiore, "Diversity of MIMO multihop relay channels," submitted to *IEEE Trans. Inf. Theory*. [Online]. Avaliable: http://arxiv.org/abs/0708.0386v1.

[12] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 6068-6077, Dec. 2009.

[13] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[14] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multiple-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 1841-1852, Nov. 1999.

[15] H. Alwan and A. Agarwal, "A survey on fault tolerant routing techniques in wireless sensor networks," in *Proc. 3rd Int. Conf. Sensor Tech. Appl*, Athens, Greece, June 2009, pp. 366-371.

[16] N. A. Pantazis and D. D. Vergados, "A survey on power control issues in wireless sensor networks," *IEEE Commun. Surveys Tutorials*, vol. 9, pp. 86-107, 4th Quarter, 2007.

[17] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. Academic Press, 1979.

[18] Y. Jiang, W. Hager, and J. Li, "The generalized triangular decomposition," *Mathematics Computation*, vol. 77, pp. 1037-1056, Apr. 2008.

[19] D. P. Palomar and Y. Jiang, *MIMO Transceiver Design via Majorization Theory.* Now Publishers, 2007.

[20] J.-K. Zhang, A. Kavčić, and K. M. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection," *IEEE Trans. Inf. Theory*, vol. 51, pp. 154-172, Jan. 2005.

[21] Y. Jiang, W. Hager, and J. Li, "Tunable channel decomposition for MIMO communications using channel state information," *IEEE Trans. Signal Process.*, vol. 54, pp. 4405-4418, Nov. 2006.

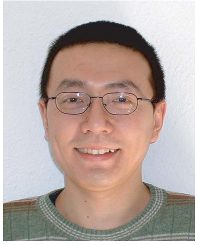
[22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[23] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, pp. 2381-2401, Sep. 2003.

[24] K. S. Gomadam and S. A. Jafar, "Optimal relay functionality for SNR maximization in memoryless relay networks," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 390-401, Feb. 2007.

[25] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors—part I: fundamentals," in *Proc. IEEE*, vol. 85, pp. 54-63, Jan. 1997.

[26] A. H. Sayed, *Fundamentals of Adaptive Filtering*. John Wiley & Sons, 2003.



Yue Rong (S'03-M'06) received the B.E. degree from Shanghai Jiao Tong University, China, the M.Sc. degree from the University of Duisburg-Essen, Duisburg, Germany, and the Ph.D. degree (summa cum laude) from Darmstadt University of Technology, Darmstadt, Germany, all in electrical engineering, in 1999, 2002, and 2005, respectively.

From April 2001 to October 2001, he was a research assistant at the Fraunhofer Institute of Microelectronic Circuits and Systems, Duisburg, Germany. From October 2001 to March 2002, he was with Nokia Ltd., Bochum, Germany. From November 2002 to March 2005, he was a Research Associate at the Department of Communication Systems in the University of Duisburg-Essen. From April 2005 to January 2006, he was with the Institute of Telecommunications at Darmstadt University of Technology, as a Research Associate. From February 2006 to November 2007, he was a Postdoctoral Researcher with the Department of Electrical Engineering, University of California, Riverside. Since December 2007, he has been with the Department of Electrical and Computer Engineering, Curtin University of Technology, Perth, Australia, as a Lecturer. His research interests include signal processing for communications, wireless communications, wireless networks, applications of linear algebra and optimization methods, and statistical and array signal processing.

Dr. Rong received the 2001-2002 Graduate Sponsoring Asia scholarship of DAAD/ABB (Germany) and the 2004 Chinese Government Award for Outstanding Self-Financed Students Abroad (China).