IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 21, 2019, accepted March 13, 2019, date of publication March 25, 2019, date of current version April 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2907404

Multiuser Multi-Hop AF MIMO Relay System Design Based on MMSE-DFE Receiver

YANG LV[1], (Student Member, IEEE), ZHIQIANG HE[1,2], (Member, IEEE), AND YUE RONG[3], (Senior Member, IEEE)
[1]Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China
[3]School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Bentley, WA 6102, Australia

Corresponding author: Yue Rong (y.rong@curtin.edu.au)

This work was supported by the National Natural Science Foundation of China under Grant 61671080.

ABSTRACT To achieve a better long source-destination distance communication in uplink multiaccess scenarios, we propose a multiuser multi-hop amplify-and-forward (AF) multiple-input multiple-output (MIMO) relay technique with nonlinear minimal mean-squared error (MMSE)-decision feedback equalization (DFE) receiver. Under transmission power constraints, this paper focuses on the improvement of reliability, meanwhile, which doesn't lose the effectiveness or require higher complexity. We demonstrate that the optimal structures of relay amplifying matrices lead to a cascading construction for the mean-squared error (MSE) matrices of respective signal waveform estimations at the destination and each relay node. Hence, in (moderately) high signal-to-noise ratio (SNR) environment, the intractable nonconvex optimization problem can be decomposed into easier subproblems for separate optimizations of source precoding and relay amplifying matrices. The source precoding matrix, along with the decision feedback matrix, is obtained by an iterative process, which can converge to a Nash point within the reasonable time. As for the relay amplifying matrices, closed-form water-filling solutions are derived. The simulation and analysis results show that compared to other existing algorithms, which also utilize decomposition methods to simplify operations, the proposed algorithms have better MSE and bit-error-rate (BER) performance without increasing the computing time or signaling overhead, thus providing a new step forward in MIMO relay system design.

INDEX TERMS MIMO relay, AF, multi-hop relay, multiuser, MMSE, DFE, mutual information.

I. INTRODUCTION

During recent decades, multiple-input multiple-output (MIMO) wireless communication technique has undergone tremendous development in both academia and industry for its remarkable advantages obtained from multiplexing, diversity, coding or antenna gains [1], [2]. In order to further improve the system reliability, reduce the power consumption and expand the network coverage, extending conventional single-hop architecture to more complicated two- or multi-hop ones has always been regarded as a quite promising strategy, where one or more relay nodes are fixed to forward signals over long source-destination distance [3], [4]. The last dozen years has seen a flourishing progress in the field of MIMO relay system design with considerable achievements obtained.

The associate editor coordinating the review of this manuscript and approving it for publication was Min Li.

In general there are three major types of relay protocols [5], [6]: the simple and fast amplify-and-forward (AF) protocol retransmits the received signals in a linear non-regenerative manner [4]; the complicated yet high-performing decode-and-forward (DF) protocol, contrariwise, regenerates source signals via decoding and recoding operations [7]; and the compress-and-forward (CF) protocol is in between, offering a worthwhile complexity-performance tradeoff [8]. Thus, for multi-hop systems with constraints on implementation costs and signal processing latency, the AF protocol is a promising solution.

Among various AF MIMO relay models, either full-duplex (FD) or half-duplex (HD) strategy can be applied, depending on whether relay nodes transmit and receive signals simultaneously or not. The FD strategy, despite making more efficient use of communication resources, suffers from severe loop-back interference, and only under a high cancellation quality, may achieve an advantageous performance [9]. Therefore, we focus on the HD strategy in this paper.

In a classic single-user two-hop system, the direct link can be either neglected or considered. For both cases, the system capacity or its upper and lower bounds were discussed in [10]. Given that there may exist a significant mismatch between the true and the estimated channel state information (CSI), robust system designs were carried out under such an imperfect channel estimation [11], [12]. Extended from point-to-point relay system, multipoint-to-multipoint system enables a single MIMO relay node to serve multiple source-destination pairs concurrently [13]. According to the capacity scaling laws in [14], instead of single relay node, multiple parallel relay nodes can be employed to expand system capacity. With two source nodes exchanging information bi-directionally through assisting relay node(s), two-way networks were investigated in [15], [16], and the generalized N-way all-cast scenarios also received attention [17], while in this paper, we concentrate on one-way uplink communication.

Inspired from [18], which performed the joint transceiver design for one-hop multicarrier MIMO links under a unified framework, the authors in [19]–[21] studied two- or multi-hop (multicarrier) AF MIMO relay systems and drew similar conclusions that, for Schur-concave objective functions, the optimal source precoding, relay amplifying and minimal mean-squared error (MMSE) receiving matrices jointly diagonalize the MIMO channel into a set of parallel single-input single-output (SISO) ones, and the same applies to Schur-convex objectives along with a rotation of the source matrix. Aiming at 5G Internet of Things (IoT), [22] studied the optimal routing for multi-hop device-to-device (D2D) communications. Under the interference from Poisson distributed third-party devices, the performance of both HD and FD strategies for multi-hop networks was analyzed in [23].

Taking multiuser issues into consideration under multiaccess scenarios, with each user having single antenna, [24] explored the optimal closed-form solution of joint linear filter design, while for users equipped with multiple antennas, iterative algorithms were developed in [25] to solve a nonconvex problem. Note that iterative technique requires centralized processing, which involves a heavy load of computing time and signaling overhead, especially for multiuser multi-hop relay systems. In view of this, using linear MMSE receiver, [26] generalized the decomposable property, discovered in [27], of the mean-squared error (MSE) matrix for the signal waveform estimation at the destination node, and proposed simplified algorithms which can be implemented in distributed manners. The approach was further extended to the weighted MMSE (WMMSE)-based problem in [28], on the basis of the relationship between maximal mutual information (MMI) and WMMSE objectives.

Aiming to improve the quality of signal reception, the well-known decision feedback equalization (DFE) [29], also referred to as successive interference cancellation (SIC) [30] or vertical-Bell laboratories layered space-time (V-BLAST) technique [31], successively cancels the inter-symbol interference (ISI) produced by previously detected symbols for current decision. Different schemes for implementing the nonlinear DFE receivers are available, commonly evolving from their linear counterparts, zero-forcing (ZF) or the aforementioned MMSE. As shown in [32], [33], an MMSE-DFE receiver can provide a significantly better performance than ZF-DFE, particularly under severe channel conditions. Meanwhile, the MMSE-DFE receiver also has lower complexity than the optimal maximum likelihood (ML) detection [31], and is verified to be information lossless [30], thus preferred in practice. For a basic two-hop model fitted with this nonlinear receiver, the work in [34] designed two closed-form precoding solutions to be adaptively selected depending on channel conditions. Using parallel MIMO relay nodes, [35] developed a joint power loading algorithm. Considering both Schur-convex and Schur-concave objectives, [36] derived the optimal source and relay matrices together with feed-forward and feedback matrices for multi-hop systems.

This paper investigates multiuser multi-hop one-way HD AF MIMO relay communication with nonlinear MMSE-DFE receiver at the destination node. To our knowledge, it's the first time to consider both multiuser context and MMSE-DFE receiver for multi-hop AF MIMO relay research field. Under this new system architecture, inspired from the optimizing procedures in [26], [28], [34], [36], we propose two distributed transceiver optimization algorithms. To be specific, by utilizing MSE minimization criterion and transmission power constraints, we first formulate a transceiver optimization problem. Next, through further generalizing the aforesaid MSE matrix decomposition method, we obtain the optimal structures of relay amplifying matrices and demonstrate that the MSE matrix for the signal waveform estimation at the destination node can recursively degenerate into those at each relay node. Following that, under (moderately) high signal-to-noise ratio (SNR) environment, the original problem can be decomposed into separate subproblems. One is to jointly optimize source precoding and decision feedback matrices, for which an iterative process is developed, able to converge within reasonable time. The others aim at designing relay amplifying matrices, where two types of closed-form water-filling solutions are derived, with the latter being a simplified version of the former. The proposed algorithms are compared with existing approaches in terms of the system MSE, bit-error-rate (BER) and mutual information (MI), as well as the computing time and signaling overhead. We show that the proposed algorithms have a better performance than existing approaches.

The rest of this paper is organized as follows. In Section II, we describe the system model and formulate the optimization problem. The algorithms proposed to optimize the source precoding, relay amplifying and decision feedback matrices are developed in Section III. Following that, Section IV gives some comments and relevant applications. Numerical simulations and analysis results are presented in Section V. Finally, Section VI draws a conclusion.

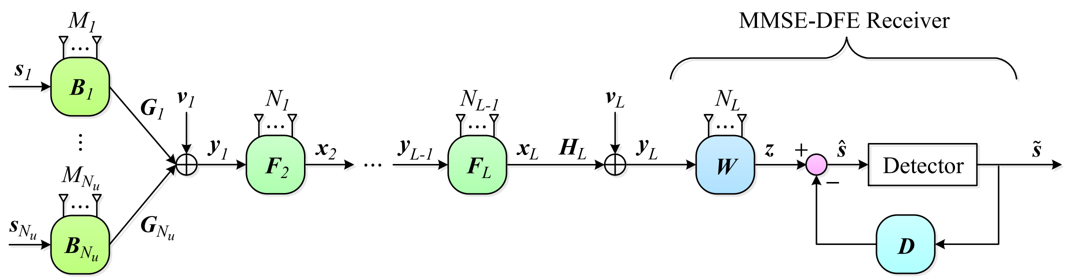


FIGURE 1. System model for N_u-user L-hop AF MIMO relay communication with MMSE-DFE receiver.

The following operators and notations are used throughout the paper: $\triangleq$ represents the phrase "is defined as"; $\mathbb{C}^n$, $\mathbb{C}^{m\times n}$ respectively denote the complex vector and matrix space with superscripts n, $m\times n$ being their dimensions; $(r)^+ \triangleq \max(r, 0)$ for real number r, and x^{-1}, x^*, $|x|$ stand for the reciprocal, complex conjugate, and modulus of scalar x; $(\cdot)^T$, $(\cdot)^H$ denote the transpose and Hermitian transpose of a vector or matrix; $\boldsymbol{X}^{-1}$, $\boldsymbol{X}^{\dagger}$, $\det(\boldsymbol{X})$, $\mathrm{rank}(\boldsymbol{X})$, $\mathrm{tr}(\boldsymbol{X})$, $\|\boldsymbol{X}\|_F$ stand for the inverse, pseudo-inverse, determinant, rank, trace, and Frobenius norm of matrix $\boldsymbol{X}$; $\prod$ represents sequential multiplication, i.e., for matrices $\boldsymbol{X}_k$, $k = m, \cdots, n$, $\prod_{k=m}^{n}\boldsymbol{X}_k \triangleq \boldsymbol{X}_m \cdots \boldsymbol{X}_n$; $[\boldsymbol{X}]_k$, $[\boldsymbol{X}]_{k,k}$, $[\boldsymbol{X}]_{m,n}$ indicate the kth column vector, the kth diagonal element, and the mth row and the nth column element of matrix $\boldsymbol{X}$; $[\boldsymbol{x}]_{1:k}$ denotes a subvector of vector $\boldsymbol{x}$, containing its first k elements; $[\boldsymbol{X}]_{1:k}$, $[\boldsymbol{X}]_{1:k,1:k}$ stand for submatrices of matrix $\boldsymbol{X}$, containing its leftmost k columns, and its first k rows and first k columns, respectively; $\mathcal{U}[\boldsymbol{X}]$ represents a matrix with its lower-triangular part filled by zeros and strictly upper-triangular part being the same as that of matrix $\boldsymbol{X}$; $\mathrm{bd}(\cdot)$ denotes a block diagonal matrix composed of the entries in parentheses; $\boldsymbol{I}_k$ is a kth-order identity matrix and $\boldsymbol{0}_{m\times n}$ stands for an $m \times n$ zero matrix; $\mathrm{E}[\cdot]$ represents statistical expectation; $\mathrm{I}(;)$ denotes the MI between two random vectors; $\mathrm{h}(\cdot)$, $\mathrm{h}(\cdot|\cdot)$ stand for the differential entropy and the conditional differential entropy of a continuous random vector, respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In the AF MIMO relay system under consideration, N_u users simultaneously transmit data towards the MMSE-DFE receiver at the destination node with the aid of $L-1$ sequential relay nodes as shown in Fig. 1, where each of the rounded rectangles contains a matrix to be optimized. Considering propagation path loss, we assume only adjacent nodes can establish direct communication links. Hence the source signals travel through L hops and need to be amplified for $L-1$ times before reaching the receiver. Moreover, by adopting the HD strategy, each relay node utilizes orthogonal channels (in time and/or frequency domain) for separate signal transmission and reception, thus avoiding loop-back interference. Here proper frequency reuse methods can be applied to take full advantage of the limited channel resources.

For $i = 1, \cdots, N_u$, there are M_i independent data streams transmitted from the ith user via M_i antennas, so the total number of independent data streams from all users is $N_0 = \sum_{i=1}^{N_u} M_i$. At each source node, the modulated signal vector $\boldsymbol{s}_i \in \mathbb{C}^{M_i}$ is linearly precoded by $\boldsymbol{B}_i \in \mathbb{C}^{M_i\times M_i}$, the source precoding matrix, and the precoded signal vector $\boldsymbol{u}_i = \boldsymbol{B}_i\boldsymbol{s}_i$ is then transmitted to the first relay node, where the received signal vector is given by

$$\boldsymbol{y}_1 = \sum_{i=1}^{N_u} \boldsymbol{G}_i\boldsymbol{u}_i + \boldsymbol{v}_1 = \boldsymbol{H}_1\boldsymbol{x}_1 + \boldsymbol{v}_1 \tag{1}$$

with $\boldsymbol{G}_i \in \mathbb{C}^{N_1\times M_i}$ being the MIMO channel matrix between the ith source node and the first relay node, $\boldsymbol{v}_1 \in \mathbb{C}^{N_1}$ being the independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) vector of the first hop, and

$$\boldsymbol{H}_1 \triangleq \left[\boldsymbol{G}_1, \cdots, \boldsymbol{G}_{N_u}\right], \qquad \boldsymbol{x}_1 \triangleq \left[\boldsymbol{u}_1^T, \cdots, \boldsymbol{u}_{N_u}^T\right]^T. \tag{2}$$

Here $\boldsymbol{H}_1 \in \mathbb{C}^{N_1\times N_0}$ stands for the equivalent first-hop MIMO channel and $\boldsymbol{x}_1 \in \mathbb{C}^{N_0}$ contains all users' precoded symbols.

For convenience in analysis, we specify

$$\boldsymbol{F}_1 \triangleq \mathrm{bd}\left(\boldsymbol{B}_1, \cdots, \boldsymbol{B}_{N_u}\right) \tag{3}$$

as well as

$$\boldsymbol{s} \triangleq \left[\boldsymbol{s}_1^T, \cdots, \boldsymbol{s}_{N_u}^T\right]^T \tag{4}$$

where $\boldsymbol{F}_1 \in \mathbb{C}^{N_0\times N_0}$ represents the equivalent block diagonal source precoding matrix and $\boldsymbol{s} \in \mathbb{C}^{N_0}$ involves all users' modulated symbols. Therefore, we obtain

$$\boldsymbol{x}_1 = \boldsymbol{F}_1\boldsymbol{s}. \tag{5}$$

Besides, all users are assumed to be symbol-synchronous and the mean power of $\boldsymbol{s}$ is normalized, i.e., $\mathrm{E}\left[\boldsymbol{s}\boldsymbol{s}^H\right] = \boldsymbol{I}_{N_0}$.

For $l = 1, \cdots, L-1$, at the lth relay node, the relay amplifying matrix $\boldsymbol{F}_{l+1} \in \mathbb{C}^{N_l\times N_l}$ is adopted to filter the received signal vector $\boldsymbol{y}_l \in \mathbb{C}^{N_l}$ in a linear non-regenerative manner of the AF protocol, hence the output signal vector $\boldsymbol{x}_{l+1} \in \mathbb{C}^{N_l}$ is given by

$$\boldsymbol{x}_{l+1} = \boldsymbol{F}_{l+1}\boldsymbol{y}_l \tag{6}$$

where $\boldsymbol{y}_1$ has already been expressed in (1) and the other inputs can be similarly written as

$$\boldsymbol{y}_l = \boldsymbol{H}_l\boldsymbol{x}_l + \boldsymbol{v}_l, \qquad l = 2, \cdots, L. \tag{7}$$

Here $\boldsymbol{H}_l \in \mathbb{C}^{N_l \times N_{l-1}}$ and $\boldsymbol{v}_l \in \mathbb{C}^{N_l}$ are respectively the lth-hop MIMO channel matrix and i.i.d. AWGN vector. As commonly practiced, all noise vectors are assumed to satisfy circular symmetry property with zero mean and identity covariance matrix.

From (5), (1), (6), (7), we can rewrite the received signal vector at each relay node and the destination node in a more integrated way as follows,

$$\boldsymbol{y}_l = \boldsymbol{A}_l \boldsymbol{s} + \bar{\boldsymbol{v}}_l, \quad l = 1, \cdots, L \tag{8}$$

where $\boldsymbol{A}_l \in \mathbb{C}^{N_l \times N_0}$ stands for the equivalent MIMO channel across the first l hops, i.e., from the modulated source signal side to the lth relay node (or destination node if $l = L$), and likewise, $\bar{\boldsymbol{v}}_l \in \mathbb{C}^{N_l}$ represents the equivalent additive noise. They are given by

$$\boldsymbol{A}_l = \prod_{i=l}^{1} (\boldsymbol{H}_i \boldsymbol{F}_i), \quad l = 1, \cdots, L \tag{9}$$

and

$$\bar{\boldsymbol{v}}_1 = \boldsymbol{v}_1, \quad \bar{\boldsymbol{v}}_l = \sum_{j=2}^{l} \left[\prod_{i=l}^{j} (\boldsymbol{H}_i \boldsymbol{F}_i)\, \boldsymbol{v}_{j-1} \right] + \boldsymbol{v}_l, \quad l = 2, \cdots, L. \tag{10}$$

Moreover, we work out the covariance matrix of $\bar{\boldsymbol{v}}_l$, $\boldsymbol{C}_l = \mathrm{E}\left[\bar{\boldsymbol{v}}_l \bar{\boldsymbol{v}}_l^H\right]$, for $l = 1, \cdots, L$ as

$$\begin{aligned} \boldsymbol{C}_1 &= \boldsymbol{I}_{N_1}, \\ \boldsymbol{C}_l &= \sum_{j=2}^{l} \left[\prod_{i=l}^{j} (\boldsymbol{H}_i \boldsymbol{F}_i) \prod_{i=j}^{l} \left(\boldsymbol{F}_i^H \boldsymbol{H}_i^H \right) \right] + \boldsymbol{I}_{N_l}, \\ & \qquad l = 2, \cdots, L, \end{aligned} \tag{11}$$

which are positive definite matrices.

In this paper, we assume all nodes in the relay system are either static or have relatively low mobility. So it is reasonable to assume quasi-static block fading channels, where $\boldsymbol{H}_l$, $l = 1, \cdots, L$ remains constant over a certain time period before changing to other realizations. Given these, high-precision CSI estimations can be obtained with small mismatch against the real CSI. As requested by our algorithms proposed later, the first relay node should acquire $\boldsymbol{H}_1$, i.e., all $\boldsymbol{G}_i$, $i = 1, \cdots, N_u$, as well as $\boldsymbol{H}_2$, while the following relay nodes require only their respective forward channel matrices. To estimate them, either standard training methods [37] or those specific to relay-assisted MIMO channels like the least-squares based algorithm presented in [38] can be applied.

On account of the inherent physical property of MIMO channels, a linear non-regenerative MIMO relay system, in order to support N_0 active independent data streams in each relay transmission, should satisfy

$$N_0 \leqslant \min\{\mathrm{rank}(\boldsymbol{H}_1), \cdots, \mathrm{rank}(\boldsymbol{H}_L)\}. \tag{12}$$

Then from the fact that the dimensions of a matrix are greater than its rank, the inequality $N_0 \leqslant \min\{N_1, \cdots, N_L\}$ is further obtained. Besides, noticing that $\boldsymbol{A}_l$, $l = 1, \cdots, L$ represents an equivalent MIMO channel matrix, we should also control $N_0 \leqslant \mathrm{rank}(\boldsymbol{A}_l)$ based on the same reason aforesaid, which, together with the restriction of dimensions, $\mathrm{rank}(\boldsymbol{A}_l) \leqslant \min\{N_l, N_0\} = N_0$, leads to $\mathrm{rank}(\boldsymbol{A}_l) = N_0$. Following that, $\mathrm{rank}(\boldsymbol{F}_l) \geqslant N_0$, $l = 1, \cdots, L$ is acquired from (9) as well as the right half of the Sylvester inequality [39]:

$$\begin{aligned} \mathrm{rank}(\boldsymbol{X}) + \mathrm{rank}(\boldsymbol{Y}) - k &\leqslant \mathrm{rank}(\boldsymbol{XY}) \\ &\leqslant \min\{\mathrm{rank}(\boldsymbol{X}), \mathrm{rank}(\boldsymbol{Y})\} \end{aligned} \tag{13}$$

where matrices $\boldsymbol{X} \in \mathbb{C}^{m \times k}$, $\boldsymbol{Y} \in \mathbb{C}^{k \times n}$. In order to facilitate the analysis and design of system parameters, we tend to control $\mathrm{rank}(\boldsymbol{F}_l) = N_0$, i.e., to hold the minimum requirement for the transmission of N_0 independent data streams. Lastly, in case any hop allows more than one potential forward relay nodes, this L-hop communication link should be set up by choosing those nodes equipped with enough number of antennas, based on appropriate upper layer protocols like the ones in [40], and its specific configurations at physical layer will be determined by the algorithms developed later in Section III.

To recover source symbols, within the nonlinear MMSE-DFE receiver at the destination node, the received signal vector first goes through a linear feed-forward filter, resulting in

$$\boldsymbol{z} = \boldsymbol{W} \boldsymbol{y}_L \tag{14}$$

where $\boldsymbol{W} \in \mathbb{C}^{N_0 \times N_L}$ stands for the feed-forward matrix and $\boldsymbol{z} \triangleq \left[z_1, \cdots, z_{N_0}\right]^T$ denotes the filtered signal vector. Then for $i = 1, \cdots, N_0$, a weighted linear combination of previous symbol decisions is fed back and subtracted from z_i, followed by a hard decision detector to successively determine the current symbol output.

Here the order in which the N_0 filtered symbols are detected has a significant influence on detection performance, and several ordering algorithms are summarized in [41]. Generally speaking, those symbols received more reliably should be detected earlier. In this paper, we detect the N_0th symbol first and the first symbol last, based on the assumption of adopting a proper coding scheme whose subcodes at growing symbol indexes are increasingly more powerful. So we have the N_0th source symbol estimated as $\hat{s}_{N_0} = z_{N_0}$, then detected as $\tilde{s}_{N_0}$, followed by the other source symbols being estimated as

$$\hat{s}_i = z_i - \sum_{j=i+1}^{N_0} d_{i,j} \tilde{s}_j, \quad i = 1, \cdots, N_0 - 1 \tag{15}$$

where $d_{i,j}$, $j = i+1, \cdots, N_0$ represents the decision feedback coefficient employed to cancel the ISI produced by the jth previously detected symbol $\tilde{s}_j$ from the ith data stream. It lies at the ith row and the jth column of the decision feedback matrix $\boldsymbol{D} \in \mathbb{C}^{N_0 \times N_0}$, which is a strictly upper triangular matrix owing to the detection order set above.

Following common practice as in [29]–[36], [41], etc., we assume the DFE receiver is free of error propagation, i.e., $\tilde{s}_j = s_j$. This assumption can be justified by the information theory [42]: there always exists a coding scheme to achieve an arbitrarily small probability of error for each data stream with any rate below subchannel capacity, meaning that the influence of error propagation can be minimized through well-designed channel coding. Hence, via introducing the detected signal vector $\tilde{\boldsymbol{s}} \triangleq \left[\tilde{s}_1, \cdots, \tilde{s}_{N_0}\right]^T$, which is assumed to be equal to $\boldsymbol{s}$, and the source signal estimator $\hat{\boldsymbol{s}} \triangleq \left[\hat{s}_1, \cdots, \hat{s}_{N_0}\right]^T$, we can reformulate (15) in a matrix form as

$$\hat{\boldsymbol{s}} = \boldsymbol{z} - \boldsymbol{D}\tilde{\boldsymbol{s}} = \boldsymbol{W}\boldsymbol{y}_L - \boldsymbol{D}\boldsymbol{s}. \tag{16}$$

Therefore, the estimation error vector $\boldsymbol{e} \triangleq \hat{\boldsymbol{s}} - \boldsymbol{s}$ is given by

$$\boldsymbol{e} = \left(\boldsymbol{W}\boldsymbol{A}_L - \boldsymbol{D} - \boldsymbol{I}_{N_0}\right)\boldsymbol{s} + \boldsymbol{W}\bar{\boldsymbol{v}}_L \tag{17}$$

from which we are able to compute the error covariance, or MSE, matrix $\boldsymbol{E}_L \triangleq \mathrm{E}\left[\boldsymbol{e}\boldsymbol{e}^H\right]$ as

$$\boldsymbol{E}_L = \left(\boldsymbol{W}\boldsymbol{A}_L - \boldsymbol{D} - \boldsymbol{I}_{N_0}\right)\left(\boldsymbol{W}\boldsymbol{A}_L - \boldsymbol{D} - \boldsymbol{I}_{N_0}\right)^H + \boldsymbol{W}\boldsymbol{C}_L\boldsymbol{W}^H. \tag{18}$$

The feed-forward matrix $\boldsymbol{W}$ is scheduled to be optimized in the first place by supposing that the other system parameters, $\boldsymbol{D}$ and $\boldsymbol{F}_l$, $l = 1, \cdots, L$, have already been configured. Applying the orthogonality principle detailed by Theorem 9.1 in [43] for MMSE estimation, we obtain $\mathrm{E}\left[\boldsymbol{e}\boldsymbol{y}_L^H\right] = \boldsymbol{0}_{N_0 \times N_L}$, which leads to

$$\boldsymbol{W} = \left(\boldsymbol{D} + \boldsymbol{I}_{N_0}\right)\boldsymbol{A}_L^H \left(\boldsymbol{A}_L\boldsymbol{A}_L^H + \boldsymbol{C}_L\right)^{-1}. \tag{19}$$

Now substituting (19) back into (18) and utilizing the matrix inversion lemma [44]:

$$(\boldsymbol{A} + \boldsymbol{X}\boldsymbol{B}\boldsymbol{Y})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{X}\left(\boldsymbol{B}^{-1} + \boldsymbol{Y}\boldsymbol{A}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{Y}\boldsymbol{A}^{-1} \tag{20}$$

for nonsingular $\boldsymbol{A}$ and $\boldsymbol{B}$, we have

$$\begin{aligned}\boldsymbol{E}_L &= \left(\boldsymbol{D} + \boldsymbol{I}_{N_0}\right)\left[\boldsymbol{I}_{N_0} - \boldsymbol{A}_L^H\left(\boldsymbol{A}_L\boldsymbol{A}_L^H + \boldsymbol{C}_L\right)^{-1}\boldsymbol{A}_L\right]\left(\boldsymbol{D} + \boldsymbol{I}_{N_0}\right)^H \\ &= \boldsymbol{U}\left(\boldsymbol{I}_{N_0} + \boldsymbol{A}_L^H\boldsymbol{C}_L^{-1}\boldsymbol{A}_L\right)^{-1}\boldsymbol{U}^H \end{aligned} \tag{21}$$

where, given for the convenience in mathematical derivations,

$$\boldsymbol{U} \triangleq \boldsymbol{D} + \boldsymbol{I}_{N_0} \in \mathbb{C}^{N_0 \times N_0} \tag{22}$$

is also called a decision feedback matrix, which is an upper triangular matrix with all its diagonal elements equal to 1.

To establish constraints on transmission power, we restrict the power consumed at the ith source node for $i = 1, \cdots, N_u$ to be no more than budget q_i, i.e.,

$$\mathrm{tr}\left\{\mathrm{E}\left[\boldsymbol{u}_i\boldsymbol{u}_i^H\right]\right\} = \mathrm{tr}\left(\boldsymbol{B}_i\boldsymbol{B}_i^H\right) \leqslant q_i, \tag{23}$$

and the power consumed at the $(l-1)$th relay node for $l = 2, \cdots, L$ to be no more than budget p_l, i.e.,

$$\begin{aligned}\mathrm{tr}\left\{\mathrm{E}\left[\boldsymbol{x}_l\boldsymbol{x}_l^H\right]\right\} &= \mathrm{tr}\left[\boldsymbol{F}_l\left(\boldsymbol{A}_{l-1}\boldsymbol{A}_{l-1}^H + \boldsymbol{C}_{l-1}\right)\boldsymbol{F}_l^H\right] \\ &= \mathrm{tr}\left(\boldsymbol{F}_l\boldsymbol{D}_{l-1}\boldsymbol{F}_l^H\right) \leqslant p_l \end{aligned} \tag{24}$$

where the matrix

$$\boldsymbol{D}_l \triangleq \boldsymbol{A}_l\boldsymbol{A}_l^H + \boldsymbol{C}_l \in \mathbb{C}^{N_l \times N_l}, \quad l = 1, \cdots, L-1 \tag{25}$$

is positive definite and essentially the covariance matrix of $\boldsymbol{y}_l$, $\mathrm{E}\left[\boldsymbol{y}_l\boldsymbol{y}_l^H\right]$. These hold true as well for $l = L$. At this point, employing the definitions of $\boldsymbol{U}$ and $\boldsymbol{D}_L$ for (19) results in

$$\boldsymbol{W} = \boldsymbol{U}\boldsymbol{A}_L^H\boldsymbol{D}_L^{-1}. \tag{26}$$

Ultimately, via minimizing $\mathrm{tr}(\boldsymbol{E}_L)$, which is the sum of the MSEs in all data streams for their signal waveform estimations at the destination node, along with constraints (23) and (24), we can formulate the transceiver optimization problem as

$$\min_{\{\boldsymbol{B}_i\},\{\boldsymbol{F}_l\},\boldsymbol{U}} \ \mathrm{tr}\left[\boldsymbol{U}\left(\boldsymbol{I}_{N_0} + \boldsymbol{A}_L^H\boldsymbol{C}_L^{-1}\boldsymbol{A}_L\right)^{-1}\boldsymbol{U}^H\right] \tag{27}$$

$$\text{s.t.} \ \mathrm{tr}\left(\boldsymbol{B}_i\boldsymbol{B}_i^H\right) \leqslant q_i, \quad i = 1, \cdots, N_u \tag{28}$$

$$\mathrm{tr}\left(\boldsymbol{F}_l\boldsymbol{D}_{l-1}\boldsymbol{F}_l^H\right) \leqslant p_l, \quad l = 2, \cdots, L \tag{29}$$

$$[\boldsymbol{U}]_{m,n} = \begin{cases} 0, & m > n \\ 1, & m = n \end{cases} \tag{30}$$

where sets

$$\{\boldsymbol{B}_i\} \triangleq \left\{\boldsymbol{B}_1, \cdots, \boldsymbol{B}_{N_u}\right\}, \quad \{\boldsymbol{F}_l\} \triangleq \{\boldsymbol{F}_2, \cdots, \boldsymbol{F}_L\} \tag{31}$$

are made up of source precoding and relay amplifying matrix variables, respectively.

The problem (27)–(30) is a complicated nonconvex problem with matrix variables, of which the globally optimal solution is intractable through any practicable non-exhaustive searching technique. So we take a step back and turn to looking for its locally optimal solutions. Here iterative procedures like those in [25] may be applied, but their demands for centralized processing imply a high computational complexity as well as a large signaling overhead. What inspires us is the MSE matrix decomposition method adopted in [26], where the algorithms presented could be performed in much more economic distributed manners. Note that [28] extended the results in [26] to decompose a weighted MSE matrix. In the following, we shall further generalize this decomposition method to the DFE-based MSE matrix $\boldsymbol{E}_L$ and accordingly develop two distributed transceiver design algorithms.

III. ALGORITHMS DESIGN

The proposed transceiver design algorithms have three major steps. Firstly, we show that $\boldsymbol{E}_L$ can be decomposed into the cascade of L MSE matrices, which enables the problem (27)–(30) to be decomposed into L subproblems. Then, an iterative process is exploited to jointly optimize $\{\boldsymbol{B}_i\}$ and $\boldsymbol{U}$. Lastly, closed-form water-filling solutions are derived for $\{\boldsymbol{F}_l\}$.

A. DFE-BASED MSE MATRIX DECOMPOSITION

By exploring the optimal structures of the relay amplifying matrices, we can recursively decompose the MSE matrix at the receiver into those at respective relay nodes, which forms a cascading construction and consequently enables distributed parameter optimizations.

Theorem 1: The optimal structures of $\{\boldsymbol{F}_l\}$ for the problem (27)–(30) are given by

$$\boldsymbol{F}_l = \boldsymbol{T}_l \boldsymbol{U} \boldsymbol{A}_{l-1}^H \boldsymbol{D}_{l-1}^{-1}, \quad l = 2, \cdots, L \tag{32}$$

where matrices $\boldsymbol{T}_l \in \mathbb{C}^{N_{l-1}\times N_0}$ remain to be solved. Using (32), $\boldsymbol{E}_L$ can be decomposed as

$$\boldsymbol{E}_L = \boldsymbol{U}\left(\boldsymbol{I}_{N_0} + \boldsymbol{F}_1^H \boldsymbol{H}_1^H \boldsymbol{H}_1 \boldsymbol{F}_1\right)^{-1} \boldsymbol{U}^H + \sum_{l=2}^{L}\left(\boldsymbol{R}_l^{-1} + \boldsymbol{T}_l^H \boldsymbol{H}_l^H \boldsymbol{H}_l \boldsymbol{T}_l\right)^{-1} \tag{33}$$

with

$$\boldsymbol{R}_l \triangleq \boldsymbol{U}\boldsymbol{A}_{l-1}^H \boldsymbol{D}_{l-1}^{-1} \boldsymbol{A}_{l-1} \boldsymbol{U}^H \in \mathbb{C}^{N_0\times N_0}, \quad l = 2, \cdots, L. \tag{34}$$

Hence, through substituting $\{\boldsymbol{T}_l\} \triangleq \{\boldsymbol{T}_2, \cdots, \boldsymbol{T}_L\}$ for $\{\boldsymbol{F}_l\}$, the problem (27)–(30) can be equivalently rewritten as

$$\min_{\{\boldsymbol{B}_i\},\{\boldsymbol{T}_l\},\boldsymbol{U}} \operatorname{tr}\left[\boldsymbol{U}\left(\boldsymbol{I}_{N_0} + \boldsymbol{F}_1^H \boldsymbol{H}_1^H \boldsymbol{H}_1 \boldsymbol{F}_1\right)^{-1} \boldsymbol{U}^H + \sum_{l=2}^{L}\left(\boldsymbol{R}_l^{-1} + \boldsymbol{T}_l^H \boldsymbol{H}_l^H \boldsymbol{H}_l \boldsymbol{T}_l\right)^{-1}\right] \tag{35}$$

$$\text{s.t. } \operatorname{tr}\left(\boldsymbol{B}_i \boldsymbol{B}_i^H\right) \leqslant q_i, \quad i = 1, \cdots, N_u \tag{36}$$

$$\operatorname{tr}\left(\boldsymbol{T}_l \boldsymbol{R}_l \boldsymbol{T}_l^H\right) \leqslant p_l, \quad l = 2, \cdots, L \tag{37}$$

$$[\boldsymbol{U}]_{m,n} = \begin{cases} 0, & m > n \\ 1, & m = n \end{cases} \tag{38}$$

Proof: See Appendix A. ∎

Here an interesting, also reasonable, point to note about (32) is that, by referring back to (26), the optimal $\boldsymbol{F}_l$ can be alternatively rewritten as

$$\boldsymbol{F}_l = \boldsymbol{T}_l \boldsymbol{W}_{l-1}, \quad l = 2, \cdots, L \tag{39}$$

where, at the $(l-1)$th relay node, specific to the received signal vector $\boldsymbol{y}_{l-1} = \boldsymbol{A}_{l-1}\boldsymbol{s} + \bar{\boldsymbol{v}}_{l-1}$, matrix $\boldsymbol{W}_{l-1} \triangleq \boldsymbol{U}\boldsymbol{A}_{l-1}^H \boldsymbol{D}_{l-1}^{-1}$ can be equivalently viewed as the feed-forward matrix of the MMSE-DFE receiver, with consistent feedback matrix $\boldsymbol{U}$ for $l = 2, \cdots, L$, and the additional linear weighting matrix $\boldsymbol{T}_l$ remains to be determined later. It is worth mentioning that $\boldsymbol{R}_l$ is exactly the covariance matrix of $\boldsymbol{W}_{l-1}\boldsymbol{y}_{l-1}$, i.e., $\boldsymbol{R}_l = \boldsymbol{W}_{l-1}\mathrm{E}\left[\boldsymbol{y}_{l-1}\boldsymbol{y}_{l-1}^H\right]\boldsymbol{W}_{l-1}^H$.

Besides, the first term in (33) is equivalent to the DFE-based MSE matrix of the signal waveform estimation at the first relay node. As each user's transmission power goes up, the Frobenius norm of $\boldsymbol{F}_1^H \boldsymbol{H}_1^H \boldsymbol{H}_1 \boldsymbol{F}_1$ tends towards infinity, hence making the first term in (33) converge to the zero matrix $\boldsymbol{0}_{N_0\times N_0}$. For $l = 2, \cdots, L$, respectively, the lth hop brings about an MSE increment in (33), i.e., $\left(\boldsymbol{R}_l^{-1} + \boldsymbol{T}_l^H \boldsymbol{H}_l^H \boldsymbol{H}_l \boldsymbol{T}_l\right)^{-1}$. Here $\boldsymbol{R}_l$ can be approximated as $\boldsymbol{U}\boldsymbol{U}^H$ when SNR is (moderately) high, which will be elaborated in the next paragraph. As SNR increases due to the growth of the transmission power at each source and relay node, the Frobenius norm of $\boldsymbol{T}_l^H \boldsymbol{H}_l^H \boldsymbol{H}_l \boldsymbol{T}_l$ gets close to infinity, thereby vanishing that MSE increment.

Making use of the matrix inversion lemma (20), we can reformulate $\boldsymbol{R}_l$ for $l = 2, \cdots, L$ as

$$\begin{aligned}\boldsymbol{R}_l &= \boldsymbol{U}\boldsymbol{A}_{l-1}^H \Big[\boldsymbol{C}_{l-1}^{-1} - \boldsymbol{C}_{l-1}^{-1}\boldsymbol{A}_{l-1}\left(\boldsymbol{I}_{N_0} + \boldsymbol{A}_{l-1}^H \boldsymbol{C}_{l-1}^{-1}\boldsymbol{A}_{l-1}\right)^{-1} \boldsymbol{A}_{l-1}^H \boldsymbol{C}_{l-1}^{-1}\Big]\boldsymbol{A}_{l-1}\boldsymbol{U}^H \\ &= \boldsymbol{U}\boldsymbol{A}_{l-1}^H \boldsymbol{C}_{l-1}^{-1}\boldsymbol{A}_{l-1}\left(\boldsymbol{I}_{N_0} + \boldsymbol{A}_{l-1}^H \boldsymbol{C}_{l-1}^{-1}\boldsymbol{A}_{l-1}\right)^{-1}\boldsymbol{U}^H.\end{aligned} \tag{40}$$

In the case of (moderately) high SNR environment where

$$\left\|\boldsymbol{A}_{l-1}^H \boldsymbol{C}_{l-1}^{-1}\boldsymbol{A}_{l-1}\right\|_F \gg \left\|\boldsymbol{I}_{N_0}\right\|_F, \tag{41}$$

we have the approximation $\boldsymbol{R}_l \approx \boldsymbol{U}\boldsymbol{U}^H$, irrelative to variables $\{\boldsymbol{B}_i\}$ and $\{\boldsymbol{T}_l\}$, which implies that, provided $\boldsymbol{U}$ is a constant, both of the objective function (35) and constraints (36)–(37) can be split in terms of respective $\{\boldsymbol{B}_i\}$ and $\{\boldsymbol{T}_l\}$. Hence, assuming relatively good channel state and a given decision feedback matrix, we are able to decompose the problem (35)–(38) into separate source precoding matrices optimization problem

$$\min_{\{\boldsymbol{B}_i\},\boldsymbol{U}} \operatorname{tr}\left[\boldsymbol{U}\left(\boldsymbol{I}_{N_0} + \boldsymbol{F}_1^H \boldsymbol{H}_1^H \boldsymbol{H}_1 \boldsymbol{F}_1\right)^{-1}\boldsymbol{U}^H\right] \tag{42}$$

$$\text{s.t. } \operatorname{tr}\left(\boldsymbol{B}_i \boldsymbol{B}_i^H\right) \leqslant q_i, \quad i = 1, \cdots, N_u \tag{43}$$

$$[\boldsymbol{U}]_{m,n} = \begin{cases} 0, & m > n \\ 1, & m = n \end{cases} \tag{44}$$

and relay amplifying matrix optimization problems, with each handling one $\boldsymbol{T}_l$ for $l = 2, \cdots, L$ as follows

$$\min_{\boldsymbol{T}_l} \operatorname{tr}\left[\left(\boldsymbol{R}_l^{-1} + \boldsymbol{T}_l^H \boldsymbol{H}_l^H \boldsymbol{H}_l \boldsymbol{T}_l\right)^{-1}\right] \tag{45}$$

$$\text{s.t. } \operatorname{tr}\left(\boldsymbol{T}_l \boldsymbol{R}_l \boldsymbol{T}_l^H\right) \leqslant p_l. \tag{46}$$

Here for simplifying algorithm design, we only consider optimizing $\boldsymbol{U}$ in the problem (42)–(44), and directly use the locally optimal $\{\boldsymbol{B}_i\}$ and $\boldsymbol{U}$ in the calculations of relay amplifying matrices. This practice is equivalent to design the decision feedback matrix merely within the first hop, which makes sense in that, commonly the links between users and the first relay node are more critical to the system performance than those between other (relay or destination) nodes, so it is reasonable to assign this extra assistance to the former. In the next two subsections, we shall focus on solving the problem (42)–(44) and the problem (45)–(46), respectively.

B. JOINT SOURCE AND FEEDBACK MATRICES OPTIMIZATION

It's worth mentioning that each of the algorithms developed in [26] and [28] has handled a degraded counterpart of the problem (42)–(44) to optimize source precoding matrices. In particular, for the case of $\boldsymbol{U} = \boldsymbol{I}_{N_0}$ in [26], via introducing a positive semidefinite matrix and using the Schur complement [45, sec. A.5.5], the problem (42)–(44) could be converted into a convex semidefinite programming (SDP) problem and thus solved through the interior-point method [45, ch. 11]. When $\boldsymbol{U}$ was replaced by a weighting matrix in [28], with its value depending on other variables, the problem, though unable to be transformed into a convex one, could be regarded as the WMMSE problem for a multiuser single-hop MIMO system with linear receiver. Note that the problem (42)–(44) is harder than that in [28], as $\boldsymbol{U}$ is an unknown upper triangular matrix with unit diagonal elements, while the weighting matrix in [28] does not have such constraint. Interestingly, the problem (42)–(44) can be rewritten as the MMSE problem for a multiuser single-hop MIMO system with nonlinear DFE receiver. Namely, the objective function in (42) satisfies

$$\begin{aligned}&\mathrm{tr}\left[\boldsymbol{U}\left(\boldsymbol{I}_{N_0}+\boldsymbol{F}_1^H\boldsymbol{H}_1^H\boldsymbol{H}_1\boldsymbol{F}_1\right)^{-1}\boldsymbol{U}^H\right]\\&=\min_{\boldsymbol{L}}\mathrm{tr}\left\{\mathrm{E}\left[\left(\boldsymbol{L}^H\boldsymbol{y}_1-\boldsymbol{D}\boldsymbol{s}-\boldsymbol{s}\right)\left(\boldsymbol{L}^H\boldsymbol{y}_1-\boldsymbol{D}\boldsymbol{s}-\boldsymbol{s}\right)^H\right]\right\}\\&=\min_{\boldsymbol{L}}\mathrm{tr}\left\{\mathrm{E}\left[\left(\boldsymbol{L}^H\boldsymbol{y}_1-\boldsymbol{U}\boldsymbol{s}\right)\left(\boldsymbol{L}^H\boldsymbol{y}_1-\boldsymbol{U}\boldsymbol{s}\right)^H\right]\right\}\end{aligned}\tag{47}$$

where, for system across the first hop in Fig. 1, $\boldsymbol{L}^H \in \mathbb{C}^{N_0\times N_1}$ acts as the feed-forward filter of the MMSE-DFE receiver, and $\boldsymbol{D}$ is still the original feedback matrix awaiting to be optimized.

To see the equivalence of (47) to (42), let's work out the MMSE problem (47), which, via utilizing $\boldsymbol{y}_1 = \boldsymbol{A}_1\boldsymbol{s} + \bar{\boldsymbol{v}}_1 = \boldsymbol{H}_1\boldsymbol{F}_1\boldsymbol{s} + \boldsymbol{v}_1$ from (8)–(10), can be written as

$$\begin{aligned}\min_{\boldsymbol{L}}\mathrm{tr}\Big[&\boldsymbol{L}^H\left(\boldsymbol{H}_1\boldsymbol{F}_1\boldsymbol{F}_1^H\boldsymbol{H}_1^H+\boldsymbol{I}_{N_1}\right)\boldsymbol{L}\\&-\boldsymbol{L}^H\boldsymbol{H}_1\boldsymbol{F}_1\boldsymbol{U}^H-\boldsymbol{U}\boldsymbol{F}_1^H\boldsymbol{H}_1^H\boldsymbol{L}+\boldsymbol{U}\boldsymbol{U}^H\Big].\end{aligned}\tag{48}$$

Apparently, with positive definite $\boldsymbol{D}_1 = \boldsymbol{H}_1\boldsymbol{F}_1\boldsymbol{F}_1^H\boldsymbol{H}_1^H + \boldsymbol{I}_{N_1}$ as $l = 1$ in (25), the above is an unconstrained convex problem. The problem (48) can be solved by employing the well-known optimality condition [45, eqn. (4.22)] to make its gradient, i.e., $2\left(\boldsymbol{H}_1\boldsymbol{F}_1\boldsymbol{F}_1^H\boldsymbol{H}_1^H + \boldsymbol{I}_{N_1}\right)\boldsymbol{L} - 2\boldsymbol{H}_1\boldsymbol{F}_1\boldsymbol{U}^H$ [46], [47], equal to zero. Hence we obtain the following solution

$$\boldsymbol{L}=\left(\boldsymbol{H}_1\boldsymbol{F}_1\boldsymbol{F}_1^H\boldsymbol{H}_1^H+\boldsymbol{I}_{N_1}\right)^{-1}\boldsymbol{H}_1\boldsymbol{F}_1\boldsymbol{U}^H\tag{49}$$

with full column rank. Then substituting (49) back into (48) results in (42). So we are able to equivalently reformulate the problem (42)–(44) as

$$\begin{aligned}\min_{\{\boldsymbol{B}_i\},\boldsymbol{U},\boldsymbol{L}}\ &\mathrm{tr}\Big[\left(\boldsymbol{L}^H\boldsymbol{H}_1\boldsymbol{F}_1-\boldsymbol{U}\right)\\&\quad\times\left(\boldsymbol{L}^H\boldsymbol{H}_1\boldsymbol{F}_1-\boldsymbol{U}\right)^H+\boldsymbol{L}^H\boldsymbol{L}\Big]\end{aligned}\tag{50}$$

$$\text{s.t. }\mathrm{tr}\left(\boldsymbol{B}_i\boldsymbol{B}_i^H\right)\leqslant q_i,\quad i=1,\cdots,N_u\tag{51}$$

$$[\boldsymbol{U}]_{m,n}=\begin{cases}0, & m>n\\1, & m=n\end{cases}\tag{52}$$

which can be solved by an iterative process described below.

Firstly, we initialize $\boldsymbol{F}_1$ via setting its diagonal submatrices $\boldsymbol{B}_i$ equal to $\sqrt{q_i/M_i}\boldsymbol{I}_{M_i}$ for $i = 1, \cdots, N_u$, which makes full use of the power budget q_i.

Next, with given $\boldsymbol{F}_1$, let's introduce the QR factorization of the following full column rank matrix

$$\begin{bmatrix}\boldsymbol{H}_1\boldsymbol{F}_1\\\boldsymbol{I}_{N_0}\end{bmatrix}=\boldsymbol{Q}\boldsymbol{R}=\begin{bmatrix}\bar{\boldsymbol{Q}}\\\ddot{\boldsymbol{Q}}\end{bmatrix}\boldsymbol{R}\tag{53}$$

where $\boldsymbol{Q} \in \mathbb{C}^{(N_1+N_0)\times N_0}$ is semi-unitary, i.e., $\boldsymbol{Q}^H\boldsymbol{Q} = \boldsymbol{I}_{N_0}$, its first N_1 and last N_0 rows respectively constitute submatrices $\bar{\boldsymbol{Q}} \in \mathbb{C}^{N_1\times N_0}$ and $\ddot{\boldsymbol{Q}} \in \mathbb{C}^{N_0\times N_0}$, and the other factor $\boldsymbol{R} \in \mathbb{C}^{N_0\times N_0}$ is a nonsingular upper-triangular matrix with all its diagonal elements being nonzero (or even positive if the Gram-Schmidt orthogonalization method [48] is applied). Thus we have

$$\boldsymbol{H}_1\boldsymbol{F}_1=\bar{\boldsymbol{Q}}\boldsymbol{R},\quad \ddot{\boldsymbol{Q}}=\boldsymbol{R}^{-1}.\tag{54}$$

Via utilizing (53) and (54), we are able to derive the optimal $\boldsymbol{U}$ and $\boldsymbol{L}$ as follows.

Theorem 2: Based on the QR factorization (53) with given $\boldsymbol{F}_1$, the optimal feedback matrix $\boldsymbol{U}$ and matrix $\boldsymbol{L}$ can be expressed as

$$\boldsymbol{U}=\boldsymbol{D}_{\boldsymbol{R}}^{-1}\boldsymbol{R},\quad \boldsymbol{L}=\bar{\boldsymbol{Q}}\boldsymbol{D}_{\boldsymbol{R}}^{-H}\tag{55}$$

where $\boldsymbol{D}_{\boldsymbol{R}} \in \mathbb{C}^{N_0\times N_0}$ stands for a diagonal matrix extracting all the diagonal elements from matrix $\boldsymbol{R}$.

Proof: See Appendix B. ■

Expression (55) is inspired by Theorem 1 in [36], which aims at the MMSE-DFE receiver design of a single-user multi-hop MIMO relay system, while here, Theorem 2 is specific to that of a multiuser single-hop MIMO system, instead.

Then, with $\boldsymbol{U}$ and $\boldsymbol{L}$ obtained in the current iteration, $\{\boldsymbol{B}_i\}$ can be optimized by solving

$$\min_{\{\boldsymbol{B}_i\}}\mathrm{tr}\left[\left(\boldsymbol{L}^H\boldsymbol{H}_1\boldsymbol{F}_1-\boldsymbol{U}\right)\left(\boldsymbol{L}^H\boldsymbol{H}_1\boldsymbol{F}_1-\boldsymbol{U}\right)^H\right]\tag{56}$$

$$\text{s.t. }\mathrm{tr}\left(\boldsymbol{B}_i\boldsymbol{B}_i^H\right)\leqslant q_i,\quad i=1,\cdots,N_u\tag{57}$$

which is derived directly from the problem (50)–(52). Here, in order to facilitate analysis, we define $\boldsymbol{Z} \triangleq \boldsymbol{L}^H\boldsymbol{H}_1 \in \mathbb{C}^{N_0\times N_0}$, followed by introducing submatrices $\boldsymbol{Z}_i \in \mathbb{C}^{N_0\times M_i}$ and $\boldsymbol{U}_i \in \mathbb{C}^{N_0\times M_i}$ for $i = 1, \cdots, N_u$, which are respectively composed of M_i columns in $\boldsymbol{Z}$ and $\boldsymbol{U}$, from the $\left(\sum_{j=0}^{i-1} M_j + 1\right)$th to the $\left(\sum_{j=0}^{i} M_j\right)$th one, with M_0 set equal to 0. Then, the objective function in (56) can be reformulated as

$$\sum_{i=1}^{N_u} \mathrm{tr}\left[(\boldsymbol{Z}_i\boldsymbol{B}_i - \boldsymbol{U}_i)(\boldsymbol{Z}_i\boldsymbol{B}_i - \boldsymbol{U}_i)^H\right] \tag{58}$$

which enables us to decompose the problem (56)–(57) into N_u subproblems, each related to one user's precoding matrix $\boldsymbol{B}_i$ for $i = 1, \cdots, N_u$ as follows

$$\min_{\boldsymbol{B}_i} \ \mathrm{tr}\left[(\boldsymbol{Z}_i\boldsymbol{B}_i - \boldsymbol{U}_i)(\boldsymbol{Z}_i\boldsymbol{B}_i - \boldsymbol{U}_i)^H\right] \tag{59}$$

$$\text{s.t.} \ \ \mathrm{tr}\left(\boldsymbol{B}_i\boldsymbol{B}_i^H\right) \leqslant q_i. \tag{60}$$

This convex quadratically constrained quadratic programming (QCQP) problem can be solved, with the aid of a Lagrange multiplier $\lambda_i \in \mathbb{R}$, via Karush-Kuhn-Tucker (KKT) optimality conditions [45, sec. 5.5.3], which are necessary and sufficient, appearing here as

$$\begin{aligned} \mathrm{tr}\left(\boldsymbol{B}_i\boldsymbol{B}_i^H\right) - q_i \leqslant 0, \quad \lambda_i\left[\mathrm{tr}\left(\boldsymbol{B}_i\boldsymbol{B}_i^H\right) - q_i\right] &= 0, \\ \lambda_i \geqslant 0, \quad \left(\boldsymbol{Z}_i^H\boldsymbol{Z}_i + \lambda_i\boldsymbol{I}_{M_i}\right)\boldsymbol{B}_i &= \boldsymbol{Z}_i^H\boldsymbol{U}_i. \end{aligned} \tag{61}$$

Depending on the value of λ_i, (61) can be handled under two separate cases. As $\lambda_i = 0$, we have $\boldsymbol{B}_i = \left(\boldsymbol{Z}_i^H\boldsymbol{Z}_i\right)^{\dagger}\boldsymbol{Z}_i^H\boldsymbol{U}_i$, which is just the optimal solution, provided that it also satisfies the constraint in (60). Otherwise, as $\lambda_i > 0$, we have

$$\boldsymbol{B}_i = \left(\boldsymbol{Z}_i^H\boldsymbol{Z}_i + \lambda_i\boldsymbol{I}_{M_i}\right)^{-1}\boldsymbol{Z}_i^H\boldsymbol{U}_i \tag{62}$$

where λ_i is determined by substituting (62) into the equation $\mathrm{tr}\left(\boldsymbol{B}_i\boldsymbol{B}_i^H\right) - q_i = 0$, with the result able to be solved through the bisection method [49].

During the above iterative process, each conditional update of respective $\boldsymbol{U}$, $\boldsymbol{L}$ and $\{\boldsymbol{B}_i\}$ solves a convex optimization problem, which indicates monotonic convergence towards a Nash point as the iterations progress [50]. Note that for a block multiconvex problem like (50)–(52), a stationary point must be a Nash point, but a Nash point is not necessarily a stationary point. Although here we can only guarantee to achieve a Nash point, neither the globally optimal point nor a stationary point, the Monte Carlo simulation results in Section V show much better MSE and BER performance of our proposed algorithms under (moderately) high SNR environment than other existing approaches.

C. RELAY AMPLIFYING MATRICES OPTIMIZATION

On the basis of $\boldsymbol{U}$ and $\{\boldsymbol{B}_i\}$ obtained in Subsection III-B, a closed-form water-filling solution of each problem (45)–(46) with respect to variable $\boldsymbol{T}_l$ for $l = 2, \cdots, L$ can be derived as follows.

Here, through introducing $\tilde{\boldsymbol{T}}_l \triangleq \boldsymbol{T}_l\boldsymbol{R}_l^{1/2} \in \mathbb{C}^{N_{l-1}\times N_0}$, we first reformulate the problem (45)–(46) as

$$\min_{\tilde{\boldsymbol{T}}_l} \ \mathrm{tr}\left[\boldsymbol{R}_l^{1/2}\left(\boldsymbol{I}_{N_0} + \tilde{\boldsymbol{T}}_l^H\boldsymbol{H}_l^H\boldsymbol{H}_l\tilde{\boldsymbol{T}}_l\right)^{-1}\boldsymbol{R}_l^{1/2}\right] \tag{63}$$

$$\text{s.t.} \ \ \mathrm{tr}\left(\tilde{\boldsymbol{T}}_l\tilde{\boldsymbol{T}}_l^H\right) \leqslant p_l. \tag{64}$$

Now let us introduce the eigenvalue decomposition (EVD) of $\boldsymbol{H}_l^H\boldsymbol{H}_l$ and that of $\boldsymbol{R}_l$ as

$$\boldsymbol{H}_l^H\boldsymbol{H}_l = \boldsymbol{V}_l\boldsymbol{\Lambda}_l\boldsymbol{V}_l^H, \quad \boldsymbol{R}_l = \boldsymbol{P}_l\boldsymbol{\Sigma}_l\boldsymbol{P}_l^H \tag{65}$$

where $\boldsymbol{V}_l \in \mathbb{C}^{N_{l-1}\times N_{l-1}}$ and $\boldsymbol{P}_l \in \mathbb{C}^{N_0\times N_0}$ are unitary matrices, $\boldsymbol{\Lambda}_l \in \mathbb{C}^{N_{l-1}\times N_{l-1}}$ and $\boldsymbol{\Sigma}_l \in \mathbb{C}^{N_0\times N_0}$ are diagonal eigenvalue matrices with their diagonal elements both sorted in increasing orders. From Lemma 2 in [26], the singular value decomposition (SVD) structure of the optimal $\tilde{\boldsymbol{T}}_l$ for the problem (63)–(64) is given by

$$\tilde{\boldsymbol{T}}_l = \boldsymbol{V}_{l,1}\boldsymbol{\Omega}_l\boldsymbol{P}_l^H \tag{66}$$

where $\boldsymbol{V}_{l,1} \in \mathbb{C}^{N_{l-1}\times N_0}$ is a submatrix of $\boldsymbol{V}_l$, containing its rightmost N_0 columns, and $\boldsymbol{\Omega}_l \in \mathbb{C}^{N_0\times N_0}$ is an unknown diagonal singular value matrix. Thus we have

$$\boldsymbol{T}_l = \tilde{\boldsymbol{T}}_l\boldsymbol{R}_l^{-1/2} = \boldsymbol{V}_{l,1}\boldsymbol{\Omega}_l\boldsymbol{P}_l^H\boldsymbol{P}_l\boldsymbol{\Sigma}_l^{-1/2}\boldsymbol{P}_l^H = \boldsymbol{V}_{l,1}\boldsymbol{\Delta}_l\boldsymbol{P}_l^H. \tag{67}$$

Here $\boldsymbol{\Delta}_l \triangleq \boldsymbol{\Omega}_l\boldsymbol{\Sigma}_l^{-1/2} \in \mathbb{C}^{N_0\times N_0}$ remains to be optimized.

Further, for $j = 1, \cdots, N_0$, we denote the jth diagonal element of $\boldsymbol{\Delta}_l$, $\boldsymbol{\Sigma}_l$ by $\delta_{l,j}$, $\sigma_{l,j}$, respectively, and each of the N_0 increasing diagonal elements lying at the right-bottom corner of $\boldsymbol{\Lambda}_l$ by $\lambda_{l,j}$. Apparently, $\delta_{l,j} \geqslant 0$, $\sigma_{l,j} > 0$, and on account of $\mathrm{rank}(\boldsymbol{H}_l) \geqslant N_0$, $\lambda_{l,j} > 0$. Hence by substituting (67) back into (45)–(46), this matrix optimization problem can be equivalently rewritten as the following convex power loading problem with respect to the squares of scalar variables $\delta_{l,j}$,

$$\min_{\{\delta_{l,j}^2\}} \ \sum_{j=1}^{N_0} \frac{1}{\sigma_{l,j}^{-1} + \delta_{l,j}^2\lambda_{l,j}} \tag{68}$$

$$\text{s.t.} \ \ \sum_{j=1}^{N_0} \delta_{l,j}^2\sigma_{l,j} \leqslant p_l \tag{69}$$

where $\left\{\delta_{l,j}^2\right\} \triangleq \left\{\delta_{l,1}^2, \cdots, \delta_{l,N_0}^2\right\}$.

Similar to Example 5.2 in [45], via applying KKT optimality conditions, the problem (68)–(69) has a closed-form water-filling solution given by

$$\delta_{l,j}^2 = \frac{1}{\lambda_{l,j}}\left(\sqrt{\frac{\lambda_{l,j}}{\mu_l\sigma_{l,j}}} - \frac{1}{\sigma_{l,j}}\right)^+, \quad j = 1, \cdots, N_0. \tag{70}$$

Here $\mu_l > 0$ is a Lagrangian multiplier satisfying

$$\sum_{j=1}^{N_0} \frac{\sigma_{l,j}}{\lambda_{l,j}}\left(\sqrt{\frac{\lambda_{l,j}}{\mu_l\sigma_{l,j}}} - \frac{1}{\sigma_{l,j}}\right)^+ = p_l. \tag{71}$$

Its left-hand side can be reformulated as

$$\sum_{j=1}^{N_0}\left(\sqrt{\frac{\sigma_{l,j}}{\lambda_{l,j}}}\frac{1}{\sqrt{\mu_l}} - \frac{1}{\lambda_{l,j}}\right)^+ \tag{72}$$

which is a piecewise-linear non-decreasing function of $1/\sqrt{\mu_l}$, with turning points at $1/\sqrt{\lambda_{l,j}\sigma_{l,j}}$ for $j = 1, \cdots, N_0$.

So there is a unique solution meeting (71), able to be readily determined.

Eventually, via combining both (32) and (67), we obtain our resultant relay amplifying matrices

$$\boldsymbol{F}_l = \boldsymbol{V}_{l,1}\boldsymbol{\Delta}_l\boldsymbol{P}_l^H\boldsymbol{U}\boldsymbol{A}_{l-1}^H\boldsymbol{D}_{l-1}^{-1}, \quad l = 2, \cdots, L. \tag{73}$$

Moreover, within (moderately) high SNR range, inspired from (40)–(41), we can simplify the optimization process of $\{\boldsymbol{F}_l\}$ by approximating $\boldsymbol{R}_l$ as $\boldsymbol{U}\boldsymbol{U}^H$, regardless of the index l, which is still positive definite. Thus the problem (45)–(46) for $l = 2, \cdots, L$ becomes

$$\min_{\boldsymbol{T}_l} \operatorname{tr}\left\{\left[\left(\boldsymbol{U}\boldsymbol{U}^H\right)^{-1} + \boldsymbol{T}_l^H\boldsymbol{H}_l^H\boldsymbol{H}_l\boldsymbol{T}_l\right]^{-1}\right\} \tag{74}$$

$$\text{s.t.} \ \operatorname{tr}\left[\boldsymbol{T}_l\left(\boldsymbol{U}\boldsymbol{U}^H\right)\boldsymbol{T}_l^H\right] \leqslant p_l. \tag{75}$$

In a similar way, we first introduce the EVD: $\boldsymbol{U}\boldsymbol{U}^H = \boldsymbol{P}\boldsymbol{\Gamma}\boldsymbol{P}^H$, where $\boldsymbol{P} \in \mathbb{C}^{N_0\times N_0}$ is a unitary matrix, and $\boldsymbol{\Gamma} \in \mathbb{C}^{N_0\times N_0}$ is a diagonal eigenvalue matrix whose diagonal elements, denoted by γ_j for $j = 1, \cdots, N_0$, are positive and sorted in increasing order. Following that, by using Lemma 2 in [26], the SVD structure of the optimal $\boldsymbol{T}_l$ for the problem (74)–(75) can be derived as

$$\boldsymbol{T}_l = \boldsymbol{V}_{l,1}\boldsymbol{\Theta}_l\boldsymbol{P}^H. \tag{76}$$

Here variable $\boldsymbol{\Theta}_l \in \mathbb{C}^{N_0\times N_0}$ is a diagonal singular value matrix, with its jth nonnegative diagonal element denoted by $\theta_{l,j}$ for $j = 1, \cdots, N_0$. Substituting (76) back into the problem (74)–(75) and applying KKT optimality conditions, we obtain the following closed-form water-filling solution

$$\theta_{l,j} = \left[\frac{1}{\lambda_{l,j}}\left(\sqrt{\frac{\lambda_{l,j}}{\nu_l\gamma_j}} - \frac{1}{\gamma_j}\right)^+\right]^{1/2}, \quad j = 1, \cdots, N_0 \tag{77}$$

where $\nu_l > 0$ is the Lagrangian multiplier, determined by

$$\sum_{j=1}^{N_0}\frac{\gamma_j}{\lambda_{l,j}}\left(\sqrt{\frac{\lambda_{l,j}}{\nu_l\gamma_j}} - \frac{1}{\gamma_j}\right)^+ = p_l. \tag{78}$$

At last, via combining both (32) and (76), a new design of $\{\boldsymbol{F}_l\}$ is given below,

$$\boldsymbol{F}_l = \boldsymbol{V}_{l,1}\boldsymbol{\Theta}_l\boldsymbol{P}^H\boldsymbol{U}\boldsymbol{A}_{l-1}^H\boldsymbol{D}_{l-1}^{-1}, \quad l = 2, \cdots, L. \tag{79}$$

IV. COMMENTS AND APPLICATIONS

So far, we have finished the theoretical derivations of our two newly proposed algorithms. For convenience, we denote the algorithm which computes the relay amplifying matrices as (73) by the MMSE-DFE algorithm, while its simplified version based on (79) is named as the SMMSE-DFE algorithm. Their complete procedures of transceiver design are summarized in Table 1, where we initialize the feedback matrix $\boldsymbol{U}$ as $\boldsymbol{I}_{N_0}$ to indicate the original linear receiver, and the precoding matrices $\boldsymbol{B}_i$ as $\sqrt{q_i/M_i}\boldsymbol{I}_{M_i}$ to keep the source signals unchanged with maximum permissible transmission power q_i. For the loop of joint $\boldsymbol{U}$ and $\boldsymbol{F}_1$ optimization,

TABLE 1. Procedures of (S)MMSE-DFE algorithms.

1) Initialize $\boldsymbol{U}^{(0)} = \boldsymbol{I}_{N_0}$ and $\boldsymbol{B}_i^{(0)} = \sqrt{q_i/M_i}\boldsymbol{I}_{M_i}, i = 1, \cdots, N_u$. Construct $\boldsymbol{F}_1^{(0)}$ as (3). Specify tolerance $\epsilon > 0$. Set counter $n = 0$.
   a) Let $n = n + 1$.
      Compute $\boldsymbol{U}^{(n)}$, $\boldsymbol{L}^{(n)}$ as (55) based on fixed $\boldsymbol{F}_1^{(n-1)}$.
   b) For $i = 1, \cdots, N_u$: Solve the problem (59)–(60) to obtain $\boldsymbol{B}_i^{(n)}$ based on fixed $\boldsymbol{U}^{(n)}$, $\boldsymbol{L}^{(n)}$; Construct $\boldsymbol{F}_1^{(n)}$ as (3).
   c) Continue to Step 1a until
      $\left(\|\boldsymbol{U}^{(n)} - \boldsymbol{U}^{(n-1)}\|_F + \|\boldsymbol{F}_1^{(n)} - \boldsymbol{F}_1^{(n-1)}\|_F\right)/2N_0 < \epsilon$.
2) For $l = 2, \cdots, L$, with fixed $\boldsymbol{U}^{(n)}$, $\boldsymbol{F}_1^{(n)}$:
   MMSE-DFE: Acquire $\boldsymbol{\Delta}_l$ via (70)–(71); Compute $\boldsymbol{F}_l$ as (73).
   SMMSE-DFE: Acquire $\boldsymbol{\Theta}_l$ via (77)–(78); Compute $\boldsymbol{F}_l$ as (79).
3) Compute $\boldsymbol{W}$ as (26) and $\boldsymbol{D} = \boldsymbol{U}^{(n)} - \boldsymbol{I}_{N_0}$.

$\boldsymbol{L}$ acts as an auxiliary variable, the variable with superscript (n) denotes that it is at the nth iteration, and tolerance ϵ takes a small enough positive value, only lower than which the convergence error, expressed as

$$\left(\|\boldsymbol{U}^{(n)} - \boldsymbol{U}^{(n-1)}\|_F + \|\boldsymbol{F}_1^{(n)} - \boldsymbol{F}_1^{(n-1)}\|_F\right)/2N_0, \tag{80}$$

is acceptable. (80) is essentially an average of the differences of all entries in both $\boldsymbol{U}$ and $\boldsymbol{F}_1$ over two consecutive iterations, which works well through testing.

Similar to the algorithms in [26], the proposed algorithms can also be implemented in distributed manners, with most system parameters optimized locally. For details about the MMSE-DFE algorithm, at the first relay node, we carry out the iterative process of joint source and feedback matrices design, with the resulting $\boldsymbol{B}_i$ for $i = 1, \cdots, N_u$ sent back to each respective user. Next, from $\boldsymbol{A}_1$, $\boldsymbol{D}_1$ and $\boldsymbol{U}$, $\boldsymbol{R}_2$ is computed so as to obtain the first relay amplifying matrix $\boldsymbol{F}_2$, after which we prepare $\boldsymbol{A}_2$, $\boldsymbol{D}_2$ and deliver them along with $\boldsymbol{U}$ to the second relay node. Then for $l = 2, \cdots, L-1$, based on $\boldsymbol{A}_l$, $\boldsymbol{D}_l$ and $\boldsymbol{U}$ forwarded by the $(l-1)$th relay node, the lth relay node, in turns, computes $\boldsymbol{R}_{l+1}$, obtains $\boldsymbol{F}_{l+1}$, prepares $\boldsymbol{A}_{l+1}$, $\boldsymbol{D}_{l+1}$ and delivers them as well as $\boldsymbol{U}$ to the next node in need. During the algorithm execution, preparations for both $\boldsymbol{A}_l$ and $\boldsymbol{D}_l$ are facilitated by their recursive relationships. In particular, from (9) and (11), we have

$$\boldsymbol{A}_l = \boldsymbol{H}_l\boldsymbol{F}_l\boldsymbol{A}_{l-1}, \tag{81}$$

$$\boldsymbol{C}_l = \boldsymbol{H}_l\boldsymbol{F}_l\boldsymbol{C}_{l-1}\boldsymbol{F}_l^H\boldsymbol{H}_l^H + \boldsymbol{I}_{N_l} \tag{82}$$

for $l = 2, \cdots, L$, and by further considering definition (25), we obtain

$$\boldsymbol{D}_l = \boldsymbol{H}_l\boldsymbol{F}_l\boldsymbol{D}_{l-1}\boldsymbol{F}_l^H\boldsymbol{H}_l^H + \boldsymbol{I}_{N_l}, \quad l = 2, \cdots, L. \tag{83}$$

For the SMMSE-DFE algorithm, $\boldsymbol{R}_l$ is simplified and fixed to $\boldsymbol{U}\boldsymbol{U}^H$ despite the index l, which, compared with (34), leaves out one matrix inversion and several matrix multiplications. Particularly, we may also perform the computation of $\boldsymbol{R}_l = \boldsymbol{U}\boldsymbol{U}^H$ and its EVD for only one time in this case, with the results, $\boldsymbol{\Gamma}$ and $\boldsymbol{P}$, achieved at the first relay node and forwarded to subsequent ones, together with $\boldsymbol{U}$

and other parameters, to be directly used there. Hence the SMMSE-DFE algorithm may significantly reduce the computational complexity.

Noteworthily, in view of notational convenience, this paper focuses on narrow-band single-carrier systems, while the generalized results for broadband multicarrier relay systems can be directly obtained, following either subcarrier-independent or subcarrier-cooperative approaches as analyzed in [19].

In this paper, we make comparisons among four algorithms, including those two proposed here, the first algorithm designed in [26], denoted by the LMMSE algorithm since it uses linear receiver, and the first algorithm in [28], named as the WMMSE algorithm for its constructed objective function. The reasons for selecting them are that they all target at multiuser multi-hop AF MIMO relay systems and utilize certain MSE matrix decomposition methods. In the LMMSE algorithm, $\{\boldsymbol{B}_i\}$ are obtained via solving a convex SDP problem through the well-developed interior-point method, and similar to our treatments, $\{\boldsymbol{F}_l\}$ are optimized hop-by-hop in a distributed manner, except no need to optimize and deliver $\boldsymbol{U}$. The criterion employed by the WMMSE algorithm is to maximize the MI between source and destination nodes, which can be transformed into a weighted MSE minimization problem with definite weighting matrix. Two nested iterative loops constitute its optimization procedures, with the inner loop optimizing $\{\boldsymbol{B}_i\}$ and the outer loop focusing on $\{\boldsymbol{F}_l\}$ as well as the weighting matrix, which require centralized processing. Therefore, the WMMSE algorithm has a higher complexity than the other three distributed approaches, among which the LMMSE algorithm owns the lowest complexity as it only needs to optimize the source and relay matrices. Compared with the LMMSE and WMMSE algorithms, the utilization of DFE receivers may further reduce the system MSE, with the SMMSE-DFE algorithm slightly inferior to the MMSE-DFE algorithm. Section V will elaborate their performance comparisons in more detail via numerical simulations.

We anticipate that the newly proposed algorithms could be applicable in a few practical scenarios, where fixed relay nodes help to deliver multiple users' aggregate data streams over long distances. Typical application examples include: the multi-hop wireless backhaul network for multimedia broadband access [51], the multi-hop wireless sensor network for mine tunnel monitoring [52], the multi-hop underwater acoustic network for reliable communication [53], etc. Besides, several industry standards take into consideration the relay issues as well, e.g., IEEE 802.16j-2009 [54] specifying both physical and medium access control (MAC) layer enhancements to enable relay operations for high-speed wireless access systems. Also worth mentioning, the millimeter-wave distribution network (MDN) [55] may benefit from our work to establish dense gigabit-speed backhaul. Note that the IEEE P802.11 Task Group AY has recently made MDN a use case and been adding features to support it [56].

V. NUMERICAL SIMULATIONS

This section presents the performance of the MMSE-DFE and SMMSE-DFE algorithms in optimizing multiuser multi-hop AF MIMO relay systems through numerical simulations. Here we compare them with the aforementioned LMMSE and WMMSE algorithms over several criteria, including MSE, BER, MI and computational complexity.

Intel® Core™2 Duo Processor E7500 [57] is employed to run MATLAB® R2015b under 32-bit Windows 10 operating system. When programming to solve the SDP problem in the LMMSE algorithm, we utilize CVX [58], a MATLAB-based software for convex optimization.

Following Monte Carlo method, the simulation results shown below are set to be averaged over 1000 independent channel realizations, where a flat Rayleigh fading environment is assumed, with all the entries in channel matrices having zero means. Besides, to normalize the impact from the number of transmitting antennas at each source and relay node upon the received SNR at the destination node, we configure the variances of entries in $\boldsymbol{G}_i$ as $1/M_i$, $i = 1, \cdots, N_u$, and those in $\boldsymbol{H}_l$ as $1/N_{l-1}$, $l = 2, \cdots, L$. The distribution of noise $\boldsymbol{v}_l$ for $l = 1, \cdots, L$ is subject to $\mathcal{CN}\big(\boldsymbol{0}, \boldsymbol{I}_{N_l}\big)$, denoting a circularly symmetric complex Gaussian (CSCG) random vector with zero mean and identity covariance matrix $\boldsymbol{I}_{N_l}$. Without loss of generality, we assume unified transmission power budgets for both source and relay nodes, i.e., $q_i = Q$, $i = 1, \cdots, N_u$ and $p_l = P$, $l = 2, \cdots, L$. Note that the propagation path loss is implicitly considered through P and Q, as we normalize the noise power to be unit. Based on several trials, 10^{-3} may act as a suitable value of the tolerance ϵ to stop the iterations in Table 1. For respective performance indicators versus variable P, comparisons among the four algorithms are implemented over three different examples of system settings as listed in Table 2, where "Ex." is short for the word "Example", $\{M_i\} \triangleq \big\{M_1, \cdots, M_{N_u}\big\}$ and $\{N_l\} \triangleq \{N_1, \cdots, N_L\}$.

TABLE 2. Examples of system settings.

Ex.	N_u	$\{M_i\}$	L	$\{N_l\}$	Q (dB)
1	2	$\{3, 2\}$	2	$\{8, 7\}$	20
2	3	$\{2, 3, 2\}$	3	$\{8, 10, 9\}$	20
3	3	$\{2, 3, 2\}$	5	$\{8, 10, 9, 11, 8\}$	25

A. MSE PERFORMANCE

For all data streams, the arithmetic average of their MSE is computed as $\mathrm{tr}(\boldsymbol{E}_L)/N_0$. Fig. 2 shows the MSE comparisons among all four algorithms under different examples, with P ranging from 0 dB to 60 dB. It's verified by all three examples that, when the value of P is relatively low, signifying a poor SNR environment where to decompose the problem (35)–(38) into subproblems (42)–(44) and (45)–(46) is not suitable, our two algorithms perform worse than the

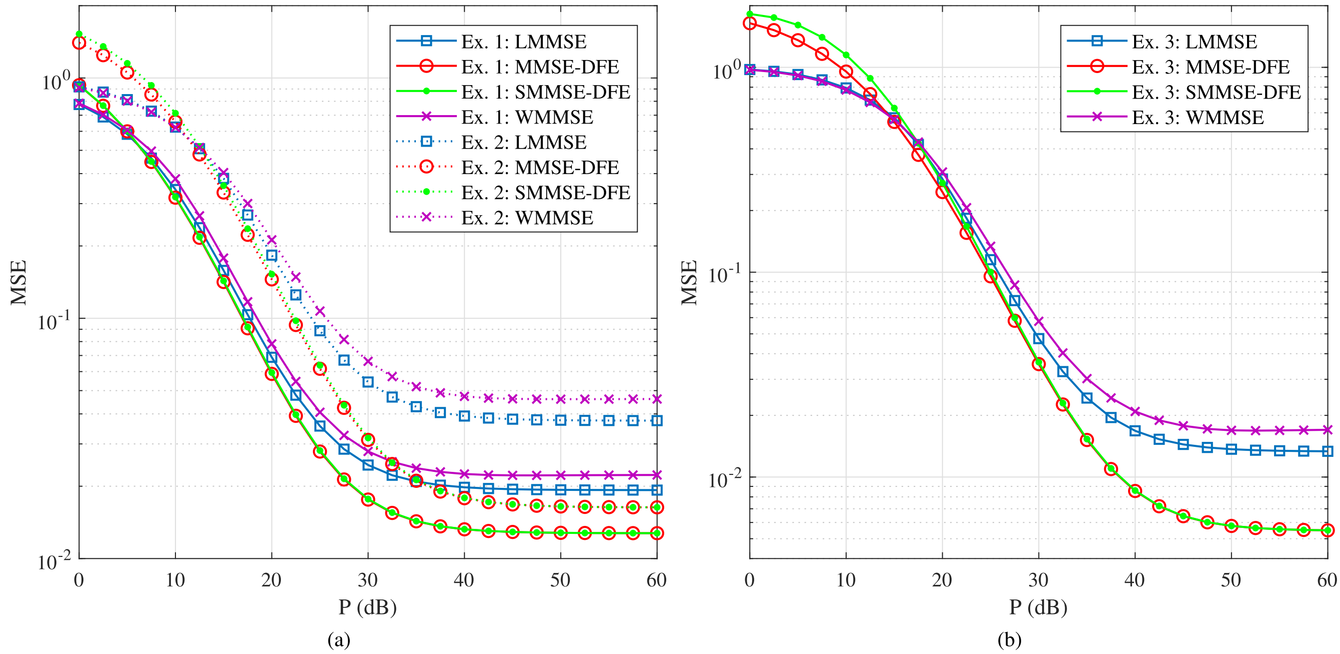


FIGURE 2. MSE versus P performance comparisons for (a) Ex. 1-2 and (b) Ex. 3.

others, while under (moderately) high SNR environment, in contrast to the LMMSE and WMMSE algorithms, we observe much enhanced performance of both MMSE-DFE and SMMSE-DFE algorithms. Moreover, the MSE curves of the two proposed algorithms almost overlap with each other so that we can substitute the latter for the former with only a negligible performance degradation. Remarkably, in each curve, there's always a rapid fall of MSE as P varies in medium range, which is mainly attributed to the growth of SNR, as well as the high spatial diversity order obtained by the underloaded utilization of available data streams, since we have $N_0 = \sum_{i=1}^{N_u} M_i < N_l$ for $l = 1, \cdots, L$ in Table 2. Besides, all the simulations are executed over fixed Q, which restricts the performance to be further improved once P becomes large enough. This saturation effect can be readily noticed in fig.2a for P greater than 40 dB. Regarding each algorithm, Ex. 2 always underperforms Ex. 1 owing to the increased number of users, which adds system load, and the increased number of hops, which brings in more noise. Meanwhile, Ex. 3 has the maximum numbers of both users and hops, leading to the highest MSE among the three examples for P lower than approximately 30 dB. However, its performance turns out to be the best after P exceeds 40 dB, thanks to the raising of Q from 20 dB to 25 dB.

B. BER PERFORMANCE

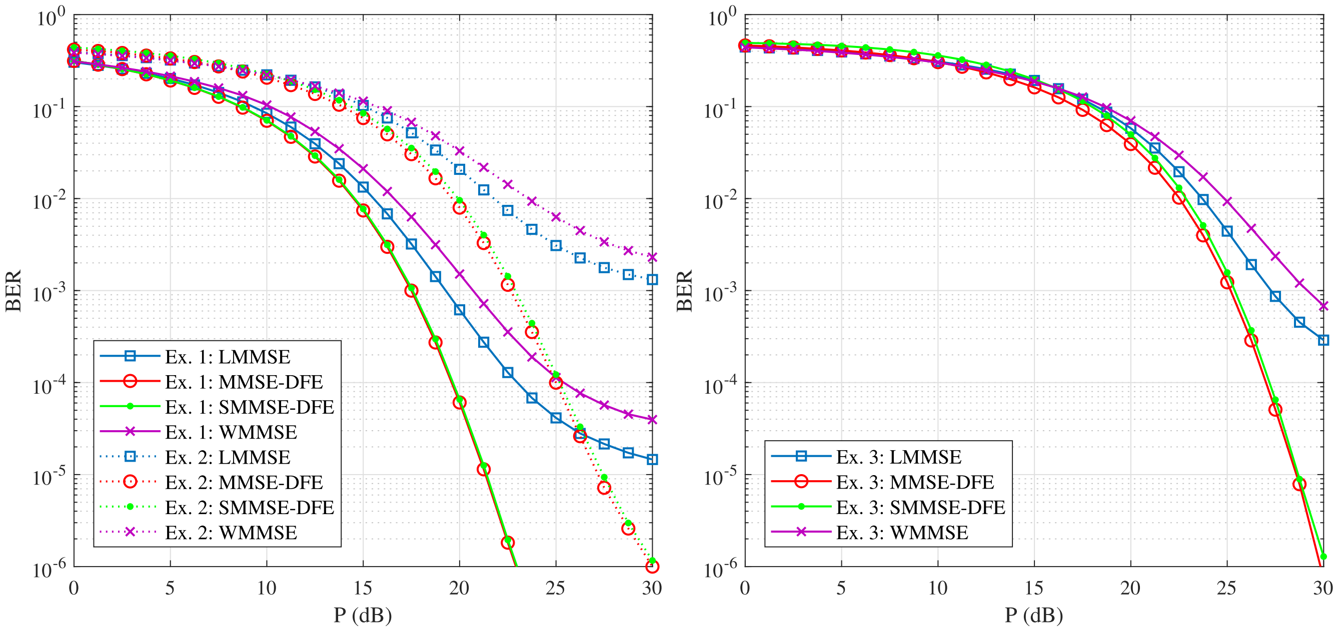
Here we illustrate the BER performance of the proposed algorithms under two different simulation schemes in Fig. 3 and Fig. 4, respectively.

The first scheme follows the assumption of no error propagation in DFE receivers to see the ideal algorithm performance without the aid of channel coding. For each channel realization, ten thousand QPSK symbols per data stream are transmitted through the channel, then are demodulated with the resulting error bits counted up for computing the average BER. Remarkably, in Fig. 3, it can be seen that the proposed algorithms have a better BER performance than existing approaches, e.g., for Ex. 2 at $P = 30$ dB, the (S)MMSE-DFE algorithms outperform the others by even more than three orders of magnitude. Moreover, via observing both Fig. 2 and Fig. 3, one can recognize that an algorithm having a better MSE also has a lower system BER, indicating that MSE is a sensible design criterion. We can also observe that the SMMSE-DFE algorithm has only a negligible higher BER than the MMSE-DFE algorithm.

The second simulation scheme, which is closer to a practical communication system, adopts Hamming codes [59] and considers the error propagation in DFE receivers, i.e., to feed back the QPSK symbols regenerated from previously decoded information bits for each data stream. The simulation results in Fig. 4 are obtained by transmitting

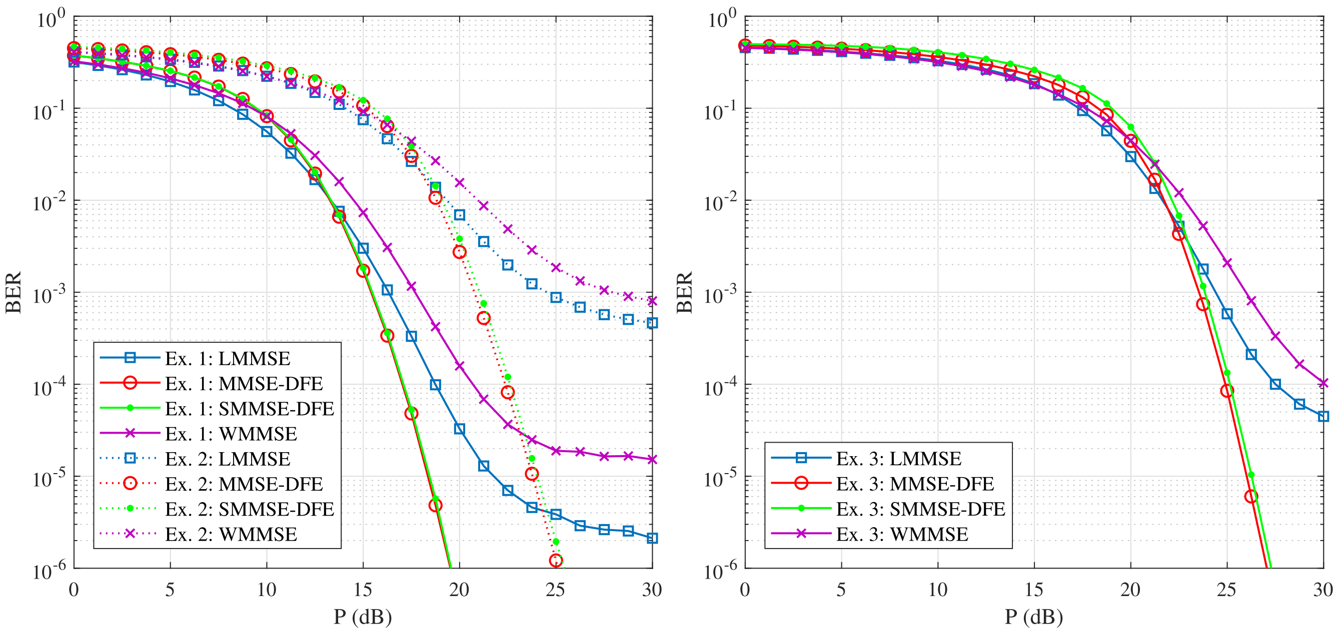
TABLE 3. The Hamming codes in use.

N_u	Users	LMMSE/WMMSE algorithms	(S)MMSE-DFE algorithms
2	1	$(7, 4)$	$(15, 11)$
	2	$(7, 4)$	$(7, 4)$
3	1	$(7, 4)$	$(31, 26)$
	2	$(7, 4)$	$(15, 11)$
	3	$(7, 4)$	$(7, 4)$



(a) (b)

FIGURE 3. BER versus *P* performance comparisons for (a) Ex. 1-2 and (b) Ex. 3 with no error propagation in DFE receivers and no channel coding.



(a) (b)

FIGURE 4. BER versus *P* performance comparisons for (a) Ex. 1-2 and (b) Ex. 3 with error propagation in DFE receivers and Hamming codes.

twenty thousand information bits per data stream through the system for each channel realization. As shown in Table 3, for the (S)MMSE-DFE algorithms, in order to minimize the error propagation from earlier detected symbols to later detected ones, we resort to the idea of unequal error protection (UEP) [41], [60] and employ increasingly more powerful Hamming codes from the first to the last user. Besides, for all the users in the LMMSE and WMMSE algorithms, we employ the $(7, 4)$ Hamming code, which is the most powerful code used for the (S)MMSE-DFE algorithms. Namely, the channel coding for the LMMSE and WMMSE algorithms is more advantageous. Even so, within relatively high range of P in Fig. 4, the BER performance of our two proposed algorithms is still better than that of the LMMSE

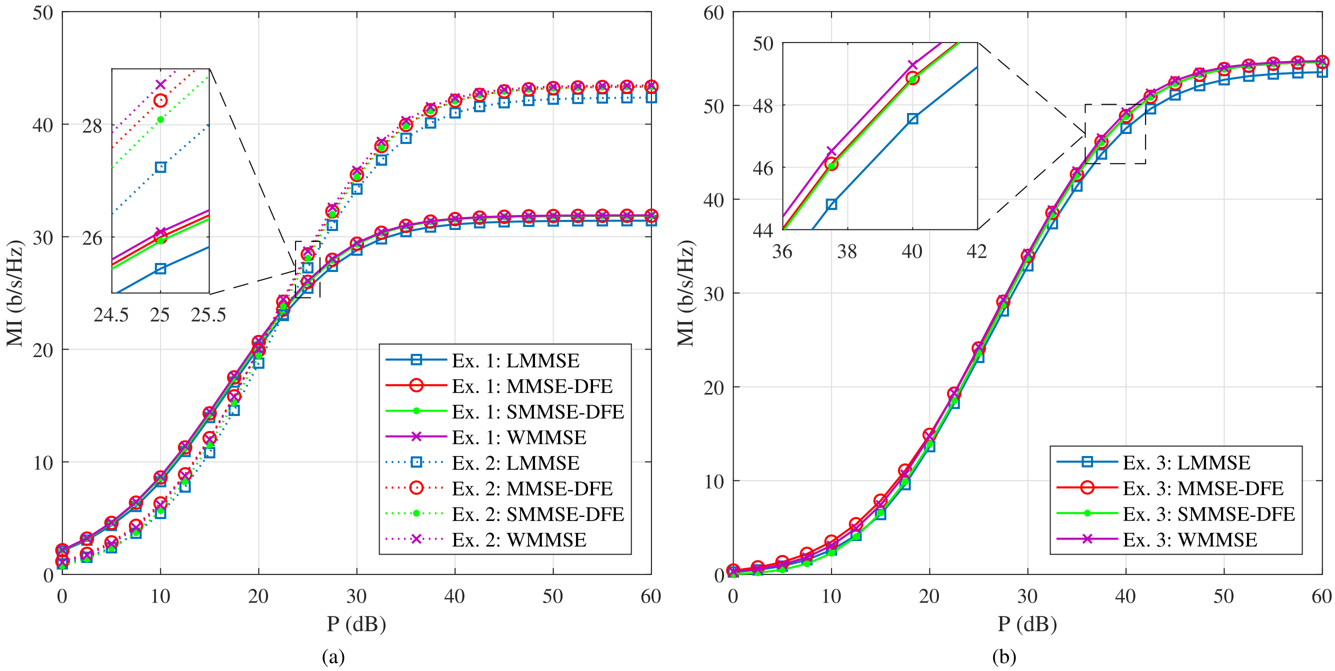


FIGURE 5. MI versus P performance comparisons for (a) Ex. 1-2 and (b) Ex. 3.

and WMMSE algorithms. It can be also observed that as the value of P is low, e.g., as $P < 19$ dB for Ex. 2, our algorithms perform slightly worse than the others, but the disadvantage is not obvious. Thus overall, the (S)MMSE-DFE algorithms, despite suffering from the error propagation, have an outstanding BER performance.

C. MI PERFORMANCE

The MI performance is also an important indicator used for evaluating the effectiveness of data transmission. Note that the capacity of a multi-hop MIMO relay channel with arbitrary signal processing methods adopted at each relay node is still a challenging problem under research. Nevertheless, aiming at the specific system model as shown in (8) with $l = L$, i.e., $\boldsymbol{y}_L = \boldsymbol{A}_L\boldsymbol{s} + \bar{\boldsymbol{v}}_L$, which is established on the basis of AF protocol and HD strategy, we can yet derive the upper boundary MI between source and destination nodes by assuming CSCG input,

$$\mathrm{I}\big(\boldsymbol{y}_L; \boldsymbol{s}\big) = \log\det\Big(\boldsymbol{I}_{N_0} + \boldsymbol{A}_L^H \boldsymbol{C}_L^{-1} \boldsymbol{A}_L\Big). \tag{84}$$

This result can be directly obtained through a similar approach as that in [61], for completeness, details of which are shown in Appendix C.

Observed from Fig. 5, the four algorithms have similar MI performance, but within the close-up views, we are still able to see the small differences. Here the WMMSE algorithm and the LMMSE algorithm respectively own the best and the worst behavior, while the curves of the (S)MMSE-DFE algorithms lying in the middle and approximating to that of the WMMSE algorithm, so the DFE-based algorithms also ensure relatively good MI performance. Note that we compute the total MI for all data streams. Hence, from Ex. 1 to Ex. 2, despite the increased number of hops, which would result in the decline of MI, we eventually obtain a higher MI due to the increase of N_0 from 5 to 7. Besides, compared with Ex. 2, the system in Ex. 3 has the same number of data streams. However, the latter system has more hops as well as higher transmission power Q for each user, which have opposite effects on the system MI. Similar to the case in Fig. 2, only at about 35 dB for the value of P, the curves of Ex. 3 start to exceed those owned by Ex. 2, i.e., the second setting plays a dominant role herefrom.

D. COMPUTING TIME AND SIGNALLING OVERHEAD

To reflect the computational complexity of each algorithm, which depends on the number of iterations that is hard to be predicted beforehand, we record the execution time instead as shown in Table 4. Note that the unit of time adopted by its fourth column is millisecond (ms), and our simulation platform can only be accurate to 1 microsecond (µs), hence the digits at the last decimal place of this column are not significant, but the outcomes of averaging operations. Here the first three distributed algorithms optimize $\{\boldsymbol{B}_i\}$ through iterative loops, which are irrelevant to variable P, then directly work out $\boldsymbol{F}_l$ for $l = 2, \cdots, L$ one after another, while the WMMSE algorithm adopts nested inner and outer loops to jointly optimize $\{\boldsymbol{B}_i\}$ and $\{\boldsymbol{F}_l\}$, with both of their time costs strongly depending on the value of P. Consequently, for the WMMSE algorithm, the entries in Table 4 are presented in ranges, rather than averages. Among those three algorithms whose computing time is insensitive to P, the LMMSE algorithm is the fastest, and as the numbers of users and hops increase, its

TABLE 4. Computing time comparisons.

Ex.	Algorithms	Time needed for inner loop (s)	Time needed to obtain $\{F_l\}$ (ms)	Iterations of outer loop	Total time overhead (s)
1	LMMSE	0.5318 [a]	0.3781	— [b]	0.5322
	MMSE-DFE	0.5510	0.4038	— [b]	0.5515
	SMMSE-DFE	0.5510	0.3286	— [b]	0.5514
	WMMSE [c]	0.0097–0.1310	— [d]	4–185	0.4378–1.9278
2	LMMSE	0.6055 [a]	0.8314	— [b]	0.6064
	MMSE-DFE	0.8961	0.8690	— [b]	0.8971
	SMMSE-DFE	0.8961	0.6092	— [b]	0.8968
	WMMSE [c]	0.0136–0.1223	— [d]	6–233	0.4034–3.5646
3	LMMSE	0.6166 [a]	1.5172	— [b]	0.6181
	MMSE-DFE	1.0918	1.5673	— [b]	1.0934
	SMMSE-DFE	1.0918	0.9944	— [b]	1.0928
	WMMSE [c]	0.0182–0.2982	— [d]	6–319	0.9323–6.8591

[a] The inner loop of the LMMSE algorithm is performed by the CVX tool to solve an SDP problem.
[b] None of the three distributed algorithms, except for the WMMSE algorithm, has outer loop.
[c] The computing time of the WMMSE algorithm varies (not always monotonically) with P ranging from 0 dB to 60 dB.
[d] In the WMMSE algorithm, locally optimal $\{F_l\}$ are obtained, along with $\{B_i\}$, via nested inner and outer loops.

high efficiency over the other algorithms becomes even more apparent. For our two proposed algorithms, their respective loops to obtain both $\{\boldsymbol{B}_i\}$ and $\boldsymbol{U}$ are exactly the same and thus consume equal time, while the distinctions are in the procedures to obtain $\{\boldsymbol{F}_l\}$, where, since the SMMSE-DFE algorithm only needs to compute the EVD of $\boldsymbol{R}_l = \boldsymbol{U}\boldsymbol{U}^H$ for one time, it has less matrix operations and, therefore, is faster than the MMSE-DFE algorithm. It's worth mentioning that, towards all three distributed algorithms, the first relay node undertakes the heaviest amount of computations, while the tasks of the other relay nodes are much easier. This should be considered in resource allocation.

As to the issue on signalling overhead, detailed analysis has already been presented in Section IV, so here we intend to make a brief summary. The WMMSE algorithm requires centralized processing, which has a high signalling overhead to collect and deliver system parameters, while the other three algorithms are able to be implemented in distributed manners, hence having lighter signalling overhead. Among them, the LMMSE algorithm is the most economic, as the DFE-based algorithms have additional demands for transmitting $\boldsymbol{U}$ hop-by-hop from the first relay node to the destination node.

VI. CONCLUSION

Targeting at uplink multiaccess scenarios in applications like high-speed wireless backhaul, we propose two promising algorithms to optimize the multiuser multi-hop AF MIMO relay system, with an MMSE-DFE receiver at the destination node for improving the reliability of long distance communication. In particular, we demonstrate that the previously discovered MSE matrix decomposition method can be generalized to this new system architecture, enabling separate optimizations for source precoding and relay amplifying matrices, of which the former, along with the decision feedback matrix, is achieved by an iterative process, while for the latter, two types of closed-form water-filling solutions are derived. From simulation and analysis in terms of MSE, BER, MI as well as computing time and signalling overhead, the outstanding performance of our two proposed algorithms is verified by comparisons with existing works.

APPENDIX A
PROOF OF THEOREM 1

The DFE-based MSE matrix $\boldsymbol{E}_L$ at the destination node can be reformulated as

$$\boldsymbol{E}_L = \boldsymbol{U}\left[\boldsymbol{I}_{N_0} - \boldsymbol{A}_L^H\left(\boldsymbol{C}_L + \boldsymbol{A}_L\boldsymbol{A}_L^H\right)^{-1}\boldsymbol{A}_L\right]\boldsymbol{U}^H \tag{85}$$

$$= \boldsymbol{U}\Big[\boldsymbol{I}_{N_0} - \boldsymbol{A}_{L-1}^H\boldsymbol{F}_L^H\boldsymbol{H}_L^H\Big(\boldsymbol{I}_{N_L} + \boldsymbol{H}_L\boldsymbol{F}_L\boldsymbol{D}_{L-1}\boldsymbol{F}_L^H\boldsymbol{H}_L^H\Big)^{-1}\boldsymbol{H}_L\boldsymbol{F}_L\boldsymbol{A}_{L-1}\Big]\boldsymbol{U}^H \tag{86}$$

$$= \boldsymbol{U}\Big\{\boldsymbol{I}_{N_0} - \boldsymbol{A}_{L-1}^H\Big[\boldsymbol{D}_{L-1}^{-1} - \Big(\boldsymbol{D}_{L-1} + \boldsymbol{D}_{L-1}\boldsymbol{F}_L^H\boldsymbol{H}_L^H\boldsymbol{H}_L\boldsymbol{F}_L\boldsymbol{D}_{L-1}\Big)^{-1}\Big]\boldsymbol{A}_{L-1}\Big\}\boldsymbol{U}^H \tag{87}$$

$$= \boldsymbol{U}\left(\boldsymbol{I}_{N_0} + \boldsymbol{A}_{L-1}^H\boldsymbol{C}_{L-1}^{-1}\boldsymbol{A}_{L-1}\right)^{-1}\boldsymbol{U}^H + \tilde{\boldsymbol{A}}_{L-1}^H\Big(\boldsymbol{D}_{L-1} + \boldsymbol{D}_{L-1}\boldsymbol{F}_L^H\boldsymbol{H}_L^H\boldsymbol{H}_L\boldsymbol{F}_L\boldsymbol{D}_{L-1}\Big)^{-1}\tilde{\boldsymbol{A}}_{L-1} \tag{88}$$

where the matrix inversion lemma (20) is applied to obtain (85) and (88), the recursive relationships of $\boldsymbol{A}_l$ and $\boldsymbol{C}_l$, i.e., (81) and (82), are used to achieve (86), the equation $\boldsymbol{X}^H\left(\boldsymbol{I} + \boldsymbol{X}\boldsymbol{Y}\boldsymbol{X}^H\right)^{-1}\boldsymbol{X} = \boldsymbol{Y}^{-1} - \left(\boldsymbol{Y} + \boldsymbol{Y}\boldsymbol{X}^H\boldsymbol{X}\boldsymbol{Y}\right)^{-1}$ for identity matrix $\boldsymbol{I}$ and nonsingular matrix $\boldsymbol{Y}$ of arbitrary orders is employed to derive (87), and $\tilde{\boldsymbol{A}}_{L-1} \triangleq \boldsymbol{A}_{L-1}\boldsymbol{U}^H \in \mathbb{C}^{N_{L-1}\times N_0}$. Noteworthily, in (88), the first term can be equivalently viewed as the degenerate DFE-based MSE matrix for the signal transmission across the first $L-1$ hops, and the second term represents the MSE increment introduced by the last hop. Besides, the first term in (88) is unrelated to $\boldsymbol{F}_L$, so extracted from the problem (27)–(30), the optimization problem only towards $\boldsymbol{F}_L$ with other variables regarded as constants is given by

$$\min_{\boldsymbol{F}_L} \operatorname{tr}\Big[\tilde{\boldsymbol{A}}_{L-1}^H \Big(\boldsymbol{D}_{L-1} + \boldsymbol{D}_{L-1} \times \boldsymbol{F}_L^H \boldsymbol{H}_L^H \boldsymbol{H}_L \boldsymbol{F}_L \boldsymbol{D}_{L-1}\Big)^{-1} \tilde{\boldsymbol{A}}_{L-1}\Big] \quad (89)$$

$$\text{s.t.} \ \operatorname{tr}\Big(\boldsymbol{F}_L \boldsymbol{D}_{L-1} \boldsymbol{F}_L^H\Big) \leqslant p_L. \quad (90)$$

Via further introducing $\tilde{\boldsymbol{F}}_L \triangleq \boldsymbol{F}_L \boldsymbol{D}_{L-1}^{1/2} \in \mathbb{C}^{N_{L-1}\times N_{L-1}}$, the problem (89)–(90) can be rewritten as

$$\min_{\tilde{\boldsymbol{F}}_L} \operatorname{tr}\Big[\boldsymbol{\Psi}_{L-1}^H \Big(\boldsymbol{I}_{N_{L-1}} + \tilde{\boldsymbol{F}}_L^H \boldsymbol{H}_L^H \boldsymbol{H}_L \tilde{\boldsymbol{F}}_L\Big)^{-1} \boldsymbol{\Psi}_{L-1}\Big] \quad (91)$$

$$\text{s.t.} \ \operatorname{tr}\Big(\tilde{\boldsymbol{F}}_L \tilde{\boldsymbol{F}}_L^H\Big) \leqslant p_L \quad (92)$$

where matrix $\boldsymbol{\Psi}_{L-1} \triangleq \boldsymbol{D}_{L-1}^{-1/2}\tilde{\boldsymbol{A}}_{L-1} = \boldsymbol{D}_{L-1}^{-1/2}\boldsymbol{A}_{L-1}\boldsymbol{U}^H \in \mathbb{C}^{N_{L-1}\times N_0}$, and from the left half of the Sylvester inequality (13), we have $\operatorname{rank}(\boldsymbol{\Psi}_{L-1}) = \operatorname{rank}\big(\tilde{\boldsymbol{A}}_{L-1}\big) = N_0$.

Now let us introduce the following EVD:

$$\boldsymbol{H}_L^H \boldsymbol{H}_L = \boldsymbol{V}_L \boldsymbol{\Lambda}_L \boldsymbol{V}_L^H, \quad (93)$$

and SVD:

$$\boldsymbol{\Psi}_{L-1} = \boldsymbol{U}_{\boldsymbol{\Psi}} \boldsymbol{\Sigma}_{\boldsymbol{\Psi}} \boldsymbol{V}_{\boldsymbol{\Psi}}^H. \quad (94)$$

Here matrices $\boldsymbol{V}_L$, $\boldsymbol{\Lambda}_L \in \mathbb{C}^{N_{L-1}\times N_{L-1}}$, $\boldsymbol{U}_{\boldsymbol{\Psi}} \in \mathbb{C}^{N_{L-1}\times N_0}$, $\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}$, $\boldsymbol{V}_{\boldsymbol{\Psi}} \in \mathbb{C}^{N_0\times N_0}$, and the diagonal elements of $\boldsymbol{\Lambda}_L$, $\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}$ are both sorted in increasing orders. Based on Lemma 2 in [26], the SVD of the optimal $\tilde{\boldsymbol{F}}_L$ for the problem (91)–(92) is given by

$$\tilde{\boldsymbol{F}}_L = \boldsymbol{V}_{L,1} \boldsymbol{\Omega}_L \boldsymbol{U}_{\boldsymbol{\Psi}}^H \quad (95)$$

where $\boldsymbol{V}_{L,1} \in \mathbb{C}^{N_{L-1}\times N_0}$ contains the rightmost N_0 columns of $\boldsymbol{V}_L$, and $\boldsymbol{\Omega}_L \in \mathbb{C}^{N_0\times N_0}$ is the diagonal singular value matrix. Then, via simple manipulations, we have

$$\tilde{\boldsymbol{F}}_L = \boldsymbol{V}_{L,1}\boldsymbol{\Omega}_L\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}^{-1}\boldsymbol{V}_{\boldsymbol{\Psi}}^H\boldsymbol{V}_{\boldsymbol{\Psi}}\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}\boldsymbol{U}_{\boldsymbol{\Psi}}^H = \boldsymbol{T}_L\boldsymbol{\Psi}_{L-1}^H \quad (96)$$

where $\boldsymbol{T}_L \triangleq \boldsymbol{V}_{L,1}\boldsymbol{\Omega}_L\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}^{-1}\boldsymbol{V}_{\boldsymbol{\Psi}}^H \in \mathbb{C}^{N_{L-1}\times N_0}$, so

$$\boldsymbol{F}_L = \tilde{\boldsymbol{F}}_L\boldsymbol{D}_{L-1}^{-1/2} = \boldsymbol{T}_L\boldsymbol{\Psi}_{L-1}^H\boldsymbol{D}_{L-1}^{-1/2} = \boldsymbol{T}_L\boldsymbol{U}\boldsymbol{A}_{L-1}^H\boldsymbol{D}_{L-1}^{-1}. \quad (97)$$

From (97) and once again the matrix inversion lemma (20), the second term in (88) can be reformulated as

$$\begin{aligned}
&\tilde{\boldsymbol{A}}_{L-1}^H\Big(\boldsymbol{D}_{L-1} + \tilde{\boldsymbol{A}}_{L-1}\boldsymbol{T}_L^H\boldsymbol{H}_L^H\boldsymbol{H}_L\boldsymbol{T}_L\tilde{\boldsymbol{A}}_{L-1}^H\Big)^{-1}\tilde{\boldsymbol{A}}_{L-1}\\
&= \tilde{\boldsymbol{A}}_{L-1}^H\Big[\boldsymbol{D}_{L-1}^{-1} - \boldsymbol{D}_{L-1}^{-1}\tilde{\boldsymbol{A}}_{L-1}\boldsymbol{T}_L^H\boldsymbol{H}_L^H\\
&\quad\times\Big(\boldsymbol{I}_{N_L} + \boldsymbol{H}_L\boldsymbol{T}_L\tilde{\boldsymbol{A}}_{L-1}^H\boldsymbol{D}_{L-1}^{-1}\tilde{\boldsymbol{A}}_{L-1}\boldsymbol{T}_L^H\boldsymbol{H}_L^H\Big)^{-1}\\
&\quad\times\boldsymbol{H}_L\boldsymbol{T}_L\tilde{\boldsymbol{A}}_{L-1}^H\boldsymbol{D}_{L-1}^{-1}\Big]\tilde{\boldsymbol{A}}_{L-1}\\
&= \Big[\Big(\tilde{\boldsymbol{A}}_{L-1}^H\boldsymbol{D}_{L-1}^{-1}\tilde{\boldsymbol{A}}_{L-1}\Big)^{-1} + \boldsymbol{T}_L^H\boldsymbol{H}_L^H\boldsymbol{H}_L\boldsymbol{T}_L\Big]^{-1}. \quad (98)
\end{aligned}$$

Substituting (98) back into (88) and making use of (34), we obtain

$$\boldsymbol{E}_L = \boldsymbol{E}_{L-1} + \Big(\boldsymbol{R}_L^{-1} + \boldsymbol{T}_L^H\boldsymbol{H}_L^H\boldsymbol{H}_L\boldsymbol{T}_L\Big)^{-1} \quad (99)$$

where

$$\boldsymbol{E}_{L-1} \triangleq \boldsymbol{U}\Big(\boldsymbol{I}_{N_0} + \boldsymbol{A}_{L-1}^H\boldsymbol{C}_{L-1}^{-1}\boldsymbol{A}_{L-1}\Big)^{-1}\boldsymbol{U}^H \quad (100)$$

is regarded as the equivalent DFE-based MSE matrix at the $(L-1)$th relay node, which can also be decomposed in a similar way as (85)–(99).

Hence the decomposition process of $\boldsymbol{E}_L$ is recursive, during which, by introducing $\boldsymbol{T}_l$ for $l = 2, \cdots, L$, we can successively get the optimal structures of $\{\boldsymbol{F}_l\}$, i.e., $\boldsymbol{F}_l = \boldsymbol{T}_l\boldsymbol{U}\boldsymbol{A}_{l-1}^H\boldsymbol{D}_{l-1}^{-1}$, resulting in a cascading construction:

$$\boldsymbol{E}_l = \boldsymbol{E}_{l-1} + \Big(\boldsymbol{R}_l^{-1} + \boldsymbol{T}_l^H\boldsymbol{H}_l^H\boldsymbol{H}_l\boldsymbol{T}_l\Big)^{-1}, \quad l = 2, \cdots, L \quad (101)$$

with

$$\boldsymbol{E}_{l-1} \triangleq \boldsymbol{U}\Big(\boldsymbol{I}_{N_0} + \boldsymbol{A}_{l-1}^H\boldsymbol{C}_{l-1}^{-1}\boldsymbol{A}_{l-1}\Big)^{-1}\boldsymbol{U}^H \quad (102)$$

and particularly,

$$\boldsymbol{E}_1 = \boldsymbol{U}\Big(\boldsymbol{I}_{N_0} + \boldsymbol{F}_1^H\boldsymbol{H}_1^H\boldsymbol{H}_1\boldsymbol{F}_1\Big)^{-1}\boldsymbol{U}^H. \quad (103)$$

Therefore at last, $\boldsymbol{E}_L = \boldsymbol{U}\big(\boldsymbol{I}_{N_0} + \boldsymbol{F}_1^H\boldsymbol{H}_1^H\boldsymbol{H}_1\boldsymbol{F}_1\big)^{-1}\boldsymbol{U}^H + \sum_{l=2}^{L}\Big(\boldsymbol{R}_l^{-1} + \boldsymbol{T}_l^H\boldsymbol{H}_l^H\boldsymbol{H}_l\boldsymbol{T}_l\Big)^{-1}$, and the transmission power consumed by each relay node in (29) can be rewritten as

$$\operatorname{tr}\Big(\boldsymbol{T}_l\boldsymbol{R}_l\boldsymbol{T}_l^H\Big), \quad l = 2, \cdots, L, \quad (104)$$

which turn the problem (27)–(30) into (35)–(38).

VII. APPENDIX B PROOF OF THEOREM 2

For the hypothetical multiuser single-hop MIMO system, assuming correct decisions are made in its DFE receiver, the source signal estimator is given by $\check{\boldsymbol{s}} = \boldsymbol{L}^H\boldsymbol{y}_1 - \boldsymbol{D}\boldsymbol{s}$. Therefore, as to the kth data stream,

$$\check{s}_k = \begin{cases} \boldsymbol{l}_k^H\boldsymbol{y}_1 - \sum_{n=k+1}^{N_0} d_{k,n}s_n, & k = 1, \cdots, N_0 - 1, \quad (105a)\\ \boldsymbol{l}_k^H\boldsymbol{y}_1, & k = N_0 \quad (105b)\end{cases}$$

where $\check{\boldsymbol{s}} \triangleq [\check{s}_1, \cdots, \check{s}_{N_0}]^T$, $\boldsymbol{L} \triangleq [\boldsymbol{l}_1, \cdots, \boldsymbol{l}_{N_0}]$, $d_{k,n}$ indicates the feedback coefficient at the kth row and the nth column of matrix $\boldsymbol{D}$, and s_n denotes the nth source symbol. Towards each data stream, we measure its performance in terms of MSE, which is derived as

$$MSE_k \triangleq \mathrm{E}\left[\left|\check{s}_k - s_k\right|^2\right] =$$

$$\begin{cases} \left|\boldsymbol{l}_k^H [\boldsymbol{H}_1\boldsymbol{F}_1]_k - 1\right|^2 + \sum_{n=1}^{k-1}\left|\boldsymbol{l}_k^H [\boldsymbol{H}_1\boldsymbol{F}_1]_n\right|^2 + \boldsymbol{l}_k^H\boldsymbol{l}_k \\ \quad + \sum_{n=k+1}^{N_0}\left|\boldsymbol{l}_k^H [\boldsymbol{H}_1\boldsymbol{F}_1]_n - d_{k,n}\right|^2, \quad k = 1,\cdots,N_0-1, & \text{(106a)} \\ \left|\boldsymbol{l}_k^H [\boldsymbol{H}_1\boldsymbol{F}_1]_k - 1\right|^2 + \sum_{n=1}^{k-1}\left|\boldsymbol{l}_k^H [\boldsymbol{H}_1\boldsymbol{F}_1]_n\right|^2 + \boldsymbol{l}_k^H\boldsymbol{l}_k, \\ \quad k = N_0. & \text{(106b)} \end{cases}$$

To minimize the MSE_k within (106a), obviously, $d_{k,n}$ should satisfy $\left|\boldsymbol{l}_k^H [\boldsymbol{H}_1\boldsymbol{F}_1]_n - d_{k,n}\right| = 0$ for $k = 1,\cdots,N_0-1$ and $n = k+1,\cdots,N_0$, hence

$$d_{k,n} = \boldsymbol{l}_k^H [\boldsymbol{H}_1\boldsymbol{F}_1]_n, \tag{107}$$

or equivalently shown in matrix form as

$$\boldsymbol{D} = \mathcal{U}\left[\boldsymbol{L}^H\boldsymbol{H}_1\boldsymbol{F}_1\right]. \tag{108}$$

Then, substituting (107) back into (105) and (106), we have

$$\begin{aligned} \check{s}_k &= \boldsymbol{l}_k^H\left(\sum_{n=1}^{k}[\boldsymbol{H}_1\boldsymbol{F}_1]_n s_n + \boldsymbol{v}_1\right) \\ &= \boldsymbol{l}_k^H\left([\boldsymbol{H}_1\boldsymbol{F}_1]_{1:k}[\boldsymbol{s}]_{1:k} + \boldsymbol{v}_1\right) \end{aligned} \tag{109}$$

and

$$\begin{aligned} MSE_k &= \left|\boldsymbol{l}_k^H [\boldsymbol{H}_1\boldsymbol{F}_1]_k - 1\right|^2 \\ &\quad + \sum_{n=1}^{k-1}\left|\boldsymbol{l}_k^H [\boldsymbol{H}_1\boldsymbol{F}_1]_n\right|^2 + \boldsymbol{l}_k^H\boldsymbol{l}_k \end{aligned} \tag{110}$$

for $k = 1,\cdots,N_0$.

Note that the Hessian matrix of (110) with respect to $\boldsymbol{l}_k$,

$$\nabla_{\boldsymbol{l}_k}^2 MSE_k = 2\left(\sum_{n=1}^{k}[\boldsymbol{H}_1\boldsymbol{F}_1]_n[\boldsymbol{H}_1\boldsymbol{F}_1]_n^H + \boldsymbol{I}_{N_1}\right), \tag{111}$$

is positive definite. Thus the second-order convexity condition [45, sec. 3.1.4] confirms (110) as a convex function for variable $\boldsymbol{l}_k$. Based on the MMSE criterion, according to the optimality condition [45, eqn. (4.22)], the gradient of MSE_k,

$$\begin{aligned} \nabla_{\boldsymbol{l}_k} MSE_k &= 2\left(\sum_{n=1}^{k}[\boldsymbol{H}_1\boldsymbol{F}_1]_n[\boldsymbol{H}_1\boldsymbol{F}_1]_n^H + \boldsymbol{I}_{N_1}\right) \\ &\quad \times \boldsymbol{l}_k - 2[\boldsymbol{H}_1\boldsymbol{F}_1]_k, \end{aligned} \tag{112}$$

should be equal to zero. Hence the optimal $\boldsymbol{l}_k$ is given by

$$\boldsymbol{l}_k = \left(\sum_{n=1}^{k}[\boldsymbol{H}_1\boldsymbol{F}_1]_n[\boldsymbol{H}_1\boldsymbol{F}_1]_n^H + \boldsymbol{I}_{N_1}\right)^{-1}[\boldsymbol{H}_1\boldsymbol{F}_1]_k \tag{113}$$

$$= \left[\left([\boldsymbol{H}_1\boldsymbol{F}_1]_{1:k}[\boldsymbol{H}_1\boldsymbol{F}_1]_{1:k}^H + \boldsymbol{I}_{N_1}\right)^{-1}[\boldsymbol{H}_1\boldsymbol{F}_1]_{1:k}\right]_k \tag{114}$$

for $k = 1,\cdots,N_0$, which, from the matrix inversion lemma (20), can be rewritten as

$$\boldsymbol{l}_k = \left[[\boldsymbol{H}_1\boldsymbol{F}_1]_{1:k}\left([\boldsymbol{H}_1\boldsymbol{F}_1]_{1:k}^H[\boldsymbol{H}_1\boldsymbol{F}_1]_{1:k} + \boldsymbol{I}_k\right)^{-1}\right]_k. \tag{115}$$

Via applying (53) and (54), we have

$$[\boldsymbol{H}_1\boldsymbol{F}_1]_{1:k} = \left[\bar{\boldsymbol{Q}}\right]_{1:k}[\boldsymbol{R}]_{1:k,1:k}. \tag{116}$$

Therefore, substituting (116) back into (115) and noticing that $\boldsymbol{I}_k$ can be reformulated as

$$\begin{aligned} \boldsymbol{I}_k &= \left[\boldsymbol{I}_{N_0}\right]_{1:k}^H\left[\boldsymbol{I}_{N_0}\right]_{1:k} = \left[\ddot{\boldsymbol{Q}}\boldsymbol{R}\right]_{1:k}^H\left[\ddot{\boldsymbol{Q}}\boldsymbol{R}\right]_{1:k} \\ &= [\boldsymbol{R}]_{1:k,1:k}^H\left[\ddot{\boldsymbol{Q}}\right]_{1:k}^H\left[\ddot{\boldsymbol{Q}}\right]_{1:k}[\boldsymbol{R}]_{1:k,1:k}, \end{aligned} \tag{117}$$

we obtain

$$\begin{aligned} \boldsymbol{l}_k &= \Bigg[\left[\bar{\boldsymbol{Q}}\right]_{1:k}[\boldsymbol{R}]_{1:k,1:k}\Big[[\boldsymbol{R}]_{1:k,1:k}^H\Big(\left[\bar{\boldsymbol{Q}}\right]_{1:k}^H\left[\bar{\boldsymbol{Q}}\right]_{1:k} \\ &\quad + \left[\ddot{\boldsymbol{Q}}\right]_{1:k}^H\left[\ddot{\boldsymbol{Q}}\right]_{1:k}\Big)[\boldsymbol{R}]_{1:k,1:k}\Big]^{-1}\Bigg]_k \end{aligned} \tag{118}$$

$$\begin{aligned} &= \Bigg[\left[\bar{\boldsymbol{Q}}\right]_{1:k}[\boldsymbol{R}]_{1:k,1:k}\Big([\boldsymbol{R}]_{1:k,1:k}^H[\boldsymbol{Q}]_{1:k}^H[\boldsymbol{Q}]_{1:k} \\ &\quad \times [\boldsymbol{R}]_{1:k,1:k}\Big)^{-1}\Bigg]_k = \left[\left[\bar{\boldsymbol{Q}}\right]_{1:k}[\boldsymbol{R}]_{1:k,1:k}^{-H}\right]_k \end{aligned} \tag{119}$$

$$= \left[\bar{\boldsymbol{Q}}\right]_k[\boldsymbol{R}]_{k,k}^{-*}, \quad k = 1,\cdots,N_0. \tag{120}$$

Consequently, the optimal $\boldsymbol{L}$ is acquired from (120) as $\boldsymbol{L} = \bar{\boldsymbol{Q}}\boldsymbol{D}_{\boldsymbol{R}}^{-H}$. Furthermore, through (54) and $\bar{\boldsymbol{Q}}^H\bar{\boldsymbol{Q}} = \boldsymbol{I}_{N_0} - \ddot{\boldsymbol{Q}}^H\ddot{\boldsymbol{Q}}$, we have

$$\begin{aligned} \boldsymbol{L}^H\boldsymbol{H}_1\boldsymbol{F}_1 &= \boldsymbol{D}_{\boldsymbol{R}}^{-1}\bar{\boldsymbol{Q}}^H\bar{\boldsymbol{Q}}\boldsymbol{R} = \boldsymbol{D}_{\boldsymbol{R}}^{-1}\left(\boldsymbol{I}_{N_0} - \boldsymbol{R}^{-H}\boldsymbol{R}^{-1}\right)\boldsymbol{R} \\ &= \boldsymbol{D}_{\boldsymbol{R}}^{-1}\boldsymbol{R} - \boldsymbol{D}_{\boldsymbol{R}}^{-1}\boldsymbol{R}^{-H}. \end{aligned} \tag{121}$$

Thus, after substituting (121) back into (108), the optimal feedback matrix $\boldsymbol{D}$ can be rewritten as

$$\boldsymbol{D} = \boldsymbol{D}_{\boldsymbol{R}}^{-1}\boldsymbol{R} - \boldsymbol{I}_{N_0}, \tag{122}$$

from which, $\boldsymbol{U}$ is given by $\boldsymbol{U} = \boldsymbol{D} + \boldsymbol{I}_{N_0} = \boldsymbol{D}_{\boldsymbol{R}}^{-1}\boldsymbol{R}$.

VIII. APPENDIX C PROOF OF FORMULA (84)

(84) is a generalized conclusion drawn from the relevant proofing process in [61], where Lemma 3 and 4 can verify $\bar{\boldsymbol{v}}_L$ to be CSCG with zero mean and covariance $\boldsymbol{C}_L$, thus we have

$$\begin{aligned} \mathrm{I}(\boldsymbol{y}_L;\boldsymbol{s}) &= \mathrm{h}(\boldsymbol{y}_L) - \mathrm{h}(\boldsymbol{y}_L|\boldsymbol{s}) = \mathrm{h}(\boldsymbol{y}_L) - \mathrm{h}(\bar{\boldsymbol{v}}_L) \\ &= \mathrm{h}(\boldsymbol{y}_L) - \log\det(\pi e\boldsymbol{C}_L). \end{aligned} \tag{123}$$

Since $\boldsymbol{s}$ and $\bar{\boldsymbol{v}}_L$ are independent, $\boldsymbol{y}_L$ has zero mean and covariance $\mathrm{E}\left[\boldsymbol{y}_L\boldsymbol{y}_L^H\right] = \boldsymbol{A}_L\boldsymbol{A}_L^H + \boldsymbol{C}_L$, however, is uncertainly distributed. Here, from Lemma 2 in [61], $\mathrm{h}(\boldsymbol{y}_L)$ can be bounded above only when $\boldsymbol{y}_L$ is CSCG, which is the case when $\boldsymbol{s}$ is also CSCG, i.e.,

$$\begin{aligned} \mathrm{I}(\boldsymbol{y}_L;\boldsymbol{s}) &\leqslant \log\det\left[\pi e\left(\boldsymbol{A}_L\boldsymbol{A}_L^H + \boldsymbol{C}_L\right)\right] - \log\det(\pi e\boldsymbol{C}_L) \\ &= \log\det\left[\left(\boldsymbol{A}_L\boldsymbol{A}_L^H + \boldsymbol{C}_L\right)\boldsymbol{C}_L^{-1}\right] \\ &= \log\det\left(\boldsymbol{I}_{N_0} + \boldsymbol{A}_L^H\boldsymbol{C}_L^{-1}\boldsymbol{A}_L\right) \end{aligned} \tag{124}$$

where the first equality makes use of $\det(\boldsymbol{C}^{-1}) = \det(\boldsymbol{C})^{-1}$ and $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$ with invertible matrix $\boldsymbol{C}$ and square matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and the second equality follows from $\det(\boldsymbol{I}_m + \boldsymbol{XY}) = \det(\boldsymbol{I}_n + \boldsymbol{YX})$ with $\boldsymbol{X} \in \mathbb{C}^{m\times n}$ and $\boldsymbol{Y} \in \mathbb{C}^{n\times m}$.

ACKNOWLEDGMENT

The authors would like to thank Prof. Kai Niu and Prof. Wenbo Xu for helpful discussions.

REFERENCES

[1] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications—A comprehensive literature survey," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 2, pp. 87–105, 2nd Quart., 2009.

[2] D. Muirhead, M. A. Imran, and K. Arshad, "A survey of the challenges, opportunities and use of multiple antennas in current and future 5G small cell base stations," *IEEE Access*, vol. 4, pp. 2952–2964, 2016.

[3] Y. Hua, "An overview of beamforming and power allocation for MIMO relays," in *Proc. Military Commun. Conf. (MILCOM)*, San Jose, CA, USA, Oct./Nov. 2010, pp. 375–380.

[4] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "A tutorial on the optimization of amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1331–1346, Sep. 2012.

[5] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[6] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1774–1786, May 2007.

[7] M. R. Bhatnagar and M. K. Arti, "Selection beamforming and combining in decode-and-forward MIMO relay networks," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1556–1559, Aug. 2013.

[8] Y. Rong, "Two-way compress-and-forward relaying with multiple MIMO relay nodes," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1387–1390, Aug. 2014.

[9] L. J. Rodríguez, N. H. Tran, and T. Le-Ngoc, "Performance of full-duplex AF relaying in the presence of residual self-interference," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1752–1764, Sep. 2014.

[10] 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.

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

[12] Z. He, W. Jiang, and Y. Rong, "Robust design for amplify-and-forward MIMO relay systems with direct link and imperfect channel information," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 353–363, Jan. 2015.

[13] H.-J. Choi, C. Song, H. Park, and I. Lee, "Transceiver designs for multipoint-to-multipoint MIMO amplify-and-forward relaying systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 198–209, Jan. 2014.

[14] H. Bölcskei, R. U. Nabar, Ö. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.

[15] Y. Rong, "Joint source and relay optimization for two-way MIMO multi-relay networks," *IEEE Commun. Lett.*, vol. 15, no. 12, pp. 1329–1331, Dec. 2011.

[16] Y. Rong, "Joint source and relay optimization for two-way linear non-regenerative MIMO relay communications," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6533–6546, Dec. 2012.

[17] T. Cui, J. Kliewer, and T. Ho, "Communication protocols for *N*-way all-cast relay networks," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3239–3251, Nov. 2012.

[18] 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.

[19] 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.

[20] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6068–6077, Dec. 2009.

[21] Y. Rong and Y. Hua, "Optimality of diagonalization of multicarrier multi-hop linear non-regenerative MIMO relays," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Sydney, NSW, Australia, Apr. 2010, pp. 1–6.

[22] G. Chen, J. Tang, and J. P. Coon, "Optimal routing for multihop social-based D2D communications in the Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1880–1889, Jun. 2018.

[23] G. Chen, J. P. Coon, A. Mondal, B. Allen, and J. A. Chambers, "Performance analysis for multi-hop full-duplex IoT networks subject to poisson distributed interferers," *IEEE Internet Things J.*, to be published. doi: 10.1109/JIOT.2018.2885756.

[24] G. Li, Y. Wang, T. Wu, and J. Huang, "Joint linear filter design in multi-user non-regenerative MIMO-relay systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dresden, Germany, Jun. 2009, pp. 1–6.

[25] M. R. A. Khandaker and Y. Rong, "Joint source and relay optimization for multiuser MIMO relay communication systems," in *Proc. 4th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Gold Coast, QLD, Australia, Dec. 2010, pp. 1–6.

[26] Y. Rong, "Simplified algorithms for optimizing multiuser multi-hop MIMO relay systems," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2896–2904, Oct. 2011.

[27] C. Song, K.-J. Lee, and I. Lee, "MMSE based transceiver designs in closed-loop non-regenerative MIMO relaying systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2310–2319, Jul. 2010.

[28] Z. He, S. Guo, Y. Ou, and Y. Rong, "Multiuser multihop MIMO relay system design based on mutual information maximization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5725–5733, Nov. 2014.

[29] G. E. Bottomley, *Channel Equalization for Wireless Communications: From Concepts to Detailed Mathematics* (Digital & Mobile Communication). Hoboken, NJ, USA: Wiley, 2011, ch. 3/5.

[30] D. Tse and P. Viswanath, *Fundamentals Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005, ch. 8, pp. 355–364.

[31] E. Biglieri and G. Taricco, *Transmission and Reception with Multiple Antennas: Theoretical Foundations* (Foundations and Trends (R) in Communications and Information Theory), vol. 1, no. 2. Delft, The Netherlands: Now Publishers, 2004, ch. 8, pp. 85–94.

[32] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding—Part I: Equalization results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.

[33] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding—Part 11: Coding results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2595–2604, Oct. 1995.

[34] M. Ahn, H.-B. Kong, T. Kim, C. Song, and I. Lee, "Precoding techniques for MIMO AF relaying systems with decision feedback receiver," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 446–455, Jan. 2015.

[35] A. Toding, M. R. A. Khandaker, and Y. Rong, "Joint source and relay optimization for parallel MIMO relays using MMSE-DFE receiver," in *Proc. 16th Asia-Pacific Conf. Commun. (APCC)*, Auckland, New Zealand, Oct./Nov. 2010, pp. 12–16.

[36] Y. Rong, "Optimal linear non-regenerative multi-hop MIMO relays with MMSE-DFE receiver at the destination," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2268–2279, Jul. 2010.

[37] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing* (Telecommunications Signal Process). New York, NY, USA: Wiley, 1997, chs. 14–15.

[38] P. Lioliou and M. Viberg, "Least-squares based channel estimation for MIMO relays," in *Proc. Int. ITG Workshop Smart Antennas (WSA)*, Vienna, Austria, Feb. 2008, pp. 90–95.

[39] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2012, p. 13.

[40] N. A. Pantazis, S. A. Nikolidakis, and D. D. Vergados, "Energy-efficient routing protocols in wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 551–591, 2nd Quart., 2013.

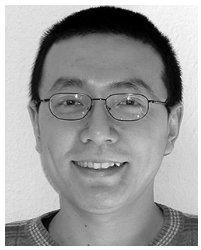
[41] E. Biglieri, G. Taricco, and A. Tulino, "Decoding space-time codes with BLAST architectures," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2547–2552, Oct. 2002.

[42] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Telecommunications and Signal Processing), 2nd ed. Hoboken, NJ, USA: Wiley, 2006, sec. 7.7, pp. 199–205.

[43] H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes*. Athens, Greece: LLC, 2014, secs. 9.1.4–9.1.7. [Online]. Available: https://www.probabilitycourse.com
[44] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *SIAM Rev.*, vol. 23, no. 1, pp. 53–60, Jan. 1981.
[45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[46] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.
[47] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. 2012, ch. 4, pp. 24–26. [Online]. Available: http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf
[48] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*. Philadelphia, PA, USA: SIAM, 1997, secs. 7–8.
[49] R. L. Burden, J. D. Faires, and A. M. Burden, *Numerical Analysis*, 10th ed. Boston, MA, USA: Cengage Learning, Jan. 2015, sec. 2.1, pp. 48–54.
[50] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, Sep. 2013.
[51] 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, no. 4, pp. 738–748, May 2007.
[52] H. Jiang, L. Chen, J. Wu, S. Chen, and H. Leung, "A reliable and high-bandwidth multihop wireless sensor network for mine tunnel monitoring," *IEEE Sensors J.*, vol. 9, no. 11, pp. 1511–1517, Nov. 2009.
[53] W. Zhang, M. Stojanovic, and U. Mitra, "Analysis of a linear multi-hop underwater acoustic network," *IEEE J. Ocean. Eng.*, vol. 35, no. 4, pp. 961–970, Oct. 2010.
[54] *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 1: Multihop Relay Specification*, IEEE Standard 802.16j-2009, Jun. 2009.
[55] M. Grigat, S. Sawhney, D. Tujkovic, S. Krauß, C. Lange, and O. Bonneß, *mmWave Distribution Network Usage Model*, document 802.11-17/1019r2, IEEE P802.11 Task Group ay, Jul. 2017. [Online]. Available: https://mentor.ieee.org/802.11/dcn/17/11-17-1019-02-00ay-mmwave-mesh-network-usage-model.pptx
[56] D. Tujkovic, S. Sawhney, and M. Grigat, *Changes to IEEE 802.11ay in Support of mmW Distribution Network Use Cases*, document 802.11-17/1022r0, IEEE P802.11 Task Group ay, Jul. 2017. [Online]. Available: https://mentor.ieee.org/802.11/dcn/17/11-17-1022-00-00ay-changes-toieee-802-11ay-in-support-of-mmw-mesh-network-use-cases.pptx
[57] *Intel Core 2 Duo Processor E7500 (3M Cache, 2.93 GHz, 1066 MHz FSB) Product Specifications*. Intel Corporation, Santa Clara, CA, USA. Accessed: Mar. 22, 2019. [Online]. Available: https://ark.intel.com/products/36503/Intel-Core2-Duo-Processor-E7500-3M-Cache-2_93-GHz-1066-MHz-FSB
[58] M. C. Grant and S. P. Boyd. (Dec. 2017). *The CVX Users' Guide, Rel. 2.1*. CVX Research. [Online]. Available: http://web.cvxr.com/cvx/doc/CVX.pdf
[59] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005, sec. 8.2.8, p. 244.
[60] A. R. Calderbank and N. Seshadri, "Multilevel codes for unequal error protection," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1234–1248, Jul. 1993.
[61] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

YANG LV (S'18) was born in Shenyang, China, in 1993. He received the B.E. degree in information engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree in information and communication engineering. His research interests include signal and information processing in wireless communications and underwater acoustic communications.

ZHIQIANG HE (S'01–M'04) received the B.E. degree in signal and information processing and the Ph.D. degree (Hons.) in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1999 and 2004, respectively. Since 2004, he has been with the School of Information and Communication Engineering, BUPT, where he is currently a Professor and the Director of the Center of Information Theory and Technology. He is also an Adjunct Professor with the College of Big Data and Information Engineering, Guizhou University, Guiyang, China. His research interests include signal and information processing in wireless communications, networking architecture and protocol design, machine learning, and underwater acoustic communications.



YUE RONG (S'03–M'06–SM'11) received the Ph.D. degree (*summa cum laude*) in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2005.

He was a Postdoctoral Researcher with the Department of Electrical Engineering, University of California at Riverside, Riverside, from 2006 to 2007. Since 2007, he has been with the School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Bentley, Australia, where he is currently a Professor. He has published over 160 journal and conference papers in his research areas. His research interests include signal processing for communications, wireless communications, underwater acoustic communications, the applications of linear algebra and optimization methods, and statistical and array signal processing.

Dr. Rong was a TPC Member of the IEEE ICC, the IEEE GlobalSIP, WCSP, IWCMC, EUSIPCO, and ChinaCom. He was a recipient of the Best Paper Award from the 2011 International Conference on Wireless Communications and Signal Processing, the Best Paper Award from the 2010 Asia-Pacific Conference on Communications, and the Young Researcher of the Year Award from the Faculty of Science and Engineering, Curtin University, in 2010. He was an Editor of the IEEE Wireless Communications Letters, from 2012 to 2014, and a Guest Editor of the IEEE Journal on Selected Areas in Communications with the special issue on theories and methods for advanced wireless relays. He is an Associate Editor of the IEEE Transactions on Signal Processing.