Robust Energy Efficiency Optimization for Amplify-and-Forward MIMO Relaying Systems

Shiqi Gong, *Student Member, IEEE*, Shuai Wang, *Member, IEEE*, Sheng Chen, *Fellow, IEEE*, Chengwen Xing, *Member, IEEE*, Lajos Hanzo, *Fellow, IEEE*

***Abstract*—We investigate the energy efficiency (EE) of multiple-input multiple-output (MIMO) amplify-and-forward relaying networks relying on the realistic imperfect channel state information (CSI). Specifically, the relay jointly optimizes the source covariance and relay beamforming matrices by maximizing EE under additive or multiplicative relay-destination CSI errors. The optimal channel-diagonalizing structure is derived for the source covariance and relay beamforming matrices under the spectral-norm constrained additive or multiplicative CSI error. Then the existence of a saddle point is proved, which shows that the channel-diagonalizing transmission strategy is optimal in the robust EE maximization under these two types of CSI errors, and the original matrix-valued fractional robust EE problem is transformed into a scalar fractional problem. We propose the Dinkelbach method based alternating optimization scheme for this transformed robust EE problem, which is capable of finding a locally optimal solution of the original robust EE problem efficiently, and show that the semi-closed-form solution to each of the two associated subproblems can be obtained. We then prove that the channel-diagonalizing transmission strategy remains optimal when the statistically imperfect source-relay channel is additionally imposed. We also extend our work into multi-hop MIMO relaying scenarios, and prove that the channel-diagonalizing structure is optimal for the source covariance matrix and multiple relay beamforming matrices.**

***Index Terms*—Robust energy efficiency optimization, additive and multiplicative CSI errors, channel-diagonalization**

I. Introduction

Cooperative relaying is a promising technique for improving the communication reliability and expanding the communication range [1]–[3]. Moreover, given the multiplexing and/or diversity gains provided by multiple-input multiple-output (MIMO) techniques, various relaying strategies have been proposed for MIMO relaying systems [3]–[8]. Among these existing relaying strategies, the amplify-and-forward (AF) strategy is popular since only a simple linear transformation is required for forwarding signals by relays [3], [5]. Most of the existing literature of MIMO AF relaying systems concentrates on the optimization of traditional performance metrics, such as the achievable capacity and the minimum mean square errors (MSE) of signal detection [4]–[8]. Recently, considerable attention has been focused on the energy efficiency (EE), which is an important system performance metric for promoting green communications. Traditionally, it is defined as the ratio of the achievable capacity to the total power consumption of signal transmission and circuit hardware dissipation [9]–[12]. There exist some works in the literature that investigate the EE optimization for MIMO AF relaying systems [10]–[12]. Interestingly, these works provide a common insight that the channel-diagonalizing transceiver structure is optimal, in terms of EE optimization, which implies that similar to the optimization of the traditional capacity and MSE metrics [7], [8], the eigenmode transmission strategy is still optimal for EE optimization. However, these works are based on the unrealistic assumption of perfect channel state information (CSI). Since the CSI estimation errors are generally unavoidable in practice, ignoring this uncertainty, as in the works [10]–[12], will lead to significant EE performance degradation for MIMO systems. Consequently, it is necessary to consider the influence of CSI errors on the EE optimization of MIMO AF relaying systems.

There are two types of imperfect CSI models. One is the statistical CSI model, in which only partial CSI is available, such as channel mean or covariance matrix. In this context, given the channel distribution, various designs based on system average performance were investigated [11], [13], [14]. For example, the works [11], [13] studied the robust EE maximization of two-hop MIMO relaying networks given the statistical source-relay channel or relay-destination channel. In this case, the eigenmode transmission strategy is optimal. The other is the approximate CSI model, which adopts an error model for the CSI approximation. Generally, various CSI error models are classified into the deterministic and stochastic ones [15]–[21]. The deterministic CSI error is often used for modeling quantization inaccuracy, in which only the estimated channel knowledge with bounded CSI error is available [15], [16]. In this case, the worst-case robustness optimization is mainly considered [17]–[19]. In other words, the system performance under the worst-case channel quality becomes an important criterion. Naturally, it is worth investigating whether channel diagonalization is still optimal for EE optimization subject to the deterministic CSI error. To the author's best knowledge, this important issue has not been addressed in the existing literature. For the stochastic CSI error, which is well suited

Manuscript received October 31, 2018; revised April 12, 2019; accepted June 05, 2019. This work was supported in part by the National Natural Science Foundation of China under Grants 61722104, 61671058 and 61620106001. The associate editor coordinating the review of this manuscript was Prof. Himal A. Suraweera. *(Corresponding author: Shuai Wang.)*

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/Noo4558/1, EP/PO34284/1, COALESCE, of the Royal Society's Global Challenges Research Fund Grant as well as of the European Research Council's Advanced Fellow Grant QuantCom.

S. Gong, S. Wang and C. Xing are with School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mails: gsqyx@163.com, swang@bit.edu.cn, xingchengwen@gmail.com).

S. Chen and L. Hanzo are with School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mails: sqc@ecs.soton.ac.uk, lh@ecs.soton.ac.u). S. Chen is also with King Abdulaziz University, Jeddah 21589, Saudi Arabia

for modeling estimation inaccuracy, existing works typically involve outage-type performance optimization [20], [21]. In this context, the system design is generally more difficult than that for the deterministic CSI error. There exist only a few works considering mean EE optimization under outage constraints, in which the channel diagonalization is generally unavailable [21].

This paper mainly investigates the robust EE maximization of MIMO AF relaying systems under imperfect CSI with the deterministically bounded CSI error. We aim to jointly optimizing the source covariance matrix and relay beamforming matrix/matrices to maximize the system's EE for MIMO AF relaying networks under both additive and multiplicative CSI errors. The main challenge of this robust EE maximization design is that it is essentially a two-objective optimization, namely, maximizing the achievable worst-case rate while minimizing the total power consumption. Clearly, these two objectives are conflicting. Therefore, the optimal diagonalization transmission strategy for traditional robust capacity maximization [17] is not applicable, since both the achievable worst-case rate and the total transmit power are simultaneously maximized by diagonalizing the channel matrices. The robust EE maximization design must find an optimal trade off between maximizing the worst-case rate and minimizing the total power consumption. Our main contributions are summarized as follows:

- Under the spectral-norm constrained additive and multiplicative relay-destination channel errors, we prove the existence of a saddle point for the robust max-min EE problem. The analytical structure of this saddle point is derived, which enables the scalarized reformulation of the robust EE problem to reduce the optimization complexity remarkably. An important insight provided by the optimal analytical solutions is that the eigenmode transmission strategy is optimal for the robust EE maximization under the deterministic CSI errors. We further show that this eigenmode transmission strategy remains optimal for the robust EE optimization when the statistically imperfect source-relay channel is additionally imposed.
- In order to effectively solve the scalarized EE optimization, we propose an alternating optimization of the source covariance matrix related subproblem and the relay beamforming related subproblem. For both these subproblems, we can jointly apply Dinkelbach's method [22] and the Lagrangian dual method to obtain the water-filling structured solutions. The convergence of the proposed alternating optimization is established. This approach is also applicable to the case with additional statistically imperfect source-relay channel.
- Furthermore, we extend our work to multi-hop MIMO AF relaying scenarios. The eigenmode transmission strategy is also proved to be optimal for the robust EE optimization subject to deterministic relay-destination CSI errors, and our proposed alternating optimization remains applicable. The extension to the robust EE optimization subject to additional statistically imperfect relay-relay channels is also discussed.

The bold-faced lower-case and upper-case letters stand for vectors and matrices, respectively. The transpose, Hermitian and inverse operators are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ and $(\cdot)^{-1}$, respectively, while $\mathrm{Tr}(\boldsymbol{A})$ and $\det(\boldsymbol{A})$ denote the trace and determinant of $\boldsymbol{A}$, respectively. $\mathsf{E}[\cdot]$ is the expectation, and $\boldsymbol{I}_n$ is the $n\times n$ identity matrix, while $\|\cdot\|_2$ denotes the matrix spectral norm, and $\boldsymbol{A}\succeq\boldsymbol{0}$ indicates that the square matrix $\boldsymbol{A}$ is positive semidefinite. $\boldsymbol{0}_{n\times m}$ and $\boldsymbol{1}_n$ denote the $n\times m$ zero matrix and the n-dimensional vector with all elements being one, respectively. The $n\times n$ square diagonal matrix with the diagonal elements $a_1,a_2,\cdots,a_n$ is denoted by $\mathrm{diag}\{a_1,a_2,\cdots,a_n\}$, and similarly for the $m\times n$ diagonal rectangular matrix, all the off-diagonal elements are zero. $\boldsymbol{A}\varnothing\boldsymbol{B}$ represents either $\boldsymbol{A}$ or $\boldsymbol{B}$ depending on which one is actually considered. The rank of $\boldsymbol{A}$ is denoted by $\mathrm{rank}(\boldsymbol{A})$, and $(a)^{+}=\max\{a,0\}$. The words 'independently and identically distributed' and 'with respect to' are abbreviated as 'i.i.d.' and 'w.r.t.', respectively.

II. System Model and Problem Formulation

A. Two-Hop MIMO AF Relaying Networks

Consider a MIMO AF relaying network consisting of an N_S-antenna source, an N_R-antenna relay and an N_D-antenna destination, which operates in half-duplex mode. In the first hop, the source transmits the data vector $\boldsymbol{s}\in\mathbb{C}^{N_S}$ having the covariance matrix $\mathsf{E}\big[\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\big]=\boldsymbol{V}_S\in\mathbb{C}^{N_S\times N_S}$ to the relay, whose received signal $\boldsymbol{y}_R\in\mathbb{C}^{N_R}$ is expressed as

$$\boldsymbol{y}_R=\boldsymbol{H}_{SR}\boldsymbol{s}+\boldsymbol{n}_R, \tag{1}$$

where $\boldsymbol{n}_R\in\mathbb{C}^{N_R}$ is the additive white Gaussian noise (AWGN) vector of the source-relay link with the covariance matrix $\sigma_r^2\boldsymbol{I}_{N_R}$, and $\boldsymbol{H}_{SR}\in\mathbb{C}^{N_R\times N_S}$ is the source-relay channel matrix. The source's transmit signal $\boldsymbol{s}$ has the power $P_S=\mathrm{Tr}\big(\boldsymbol{V}_S\big)$. In the second hop, the relay retransmits the signal received in the first hop by pre-multiplying $\boldsymbol{y}_R$ with the AF beamforming matrix $\boldsymbol{W}_R\in\mathbb{C}^{N_R\times N_R}$. Thus, the relay's transmitted signal $\boldsymbol{y}_R^{'}=\boldsymbol{W}_R\boldsymbol{y}_R$ has the power

$$P_R=\mathrm{Tr}\left(\boldsymbol{W}_R\big(\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}}+\sigma_r^2\boldsymbol{I}_{N_R}\big)\boldsymbol{W}_R^{\mathrm{H}}\right). \tag{2}$$

The signal $\boldsymbol{y}_D\in\mathbb{C}^{N_D}$ received at the destination is then given by

$$\boldsymbol{y}_D=\boldsymbol{H}_{RD}\boldsymbol{W}_R\boldsymbol{H}_{SR}\boldsymbol{s}+\boldsymbol{H}_{RD}\boldsymbol{W}_R\boldsymbol{n}_R+\boldsymbol{n}_D, \tag{3}$$

where $\boldsymbol{H}_{RD}\in\mathbb{C}^{N_D\times N_R}$ is the relay-destination channel matrix, and $\boldsymbol{n}_D\in\mathbb{C}^{N_D}$ is the AWGN vector of the relay-destination channel with the covariance matrix $\sigma_d^2\boldsymbol{I}_{N_D}$.

We adopt the EE metric of the MIMO AF relaying network as the optimization objective, which is defined as the ratio of the maximum achievable data rate to the total power consumed. The maximum achievable rate or capacity measured in [bit/s] is expressed as

$$\begin{aligned} R_{\mathrm{D}}=\frac{B}{2}\mathrm{logdet}\Big(&\boldsymbol{I}_{N_D}+\boldsymbol{H}_{RD}\boldsymbol{W}_R\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}}\boldsymbol{W}_R^{\mathrm{H}}\boldsymbol{H}_{RD}^{\mathrm{H}}\cdot \\ &\big(\sigma_r^2\boldsymbol{H}_{RD}\boldsymbol{W}_R\boldsymbol{W}_R^{\mathrm{H}}\boldsymbol{H}_{RD}^{\mathrm{H}}+\sigma_d^2\boldsymbol{I}_{N_D}\big)^{-1}\Big), \end{aligned} \tag{4}$$

where B denotes the allocated system bandwidth and the factor $\frac{1}{2}$ indicates the half-duplex loss. We model the total power consumption of the relaying network as the sum of the transmission powers of the source and relay, scaled by their respective power amplifier efficiencies, and the total circuit power consumption P_C, given by

$$\mathcal{P}\left(\boldsymbol{V}_S, \mathbf{W}_R\right) = \frac{P_S}{\tau_s} + \frac{P_R}{\tau_r} + P_C \text{ [Joule/s]}, \tag{5}$$

where $0 < \tau_s \leq 1$ and $0 < \tau_r \leq 1$ are the source and relay power amplifier efficiencies, respectively. According to [10], [23]–[25], the total circuit power consumption P_C can be modeled as $P_C = N_S P_{dy,s} + N_R P_{dy,r} + P_{st}$, where $P_{dy,s}$ and $P_{dy,r}$ are the dynamic power consumption of each RF chain of the source and relay, respectively, while $P_{st} = P_{st,s} + P_{st,r}$ is the total static power overhead of the source and relay, including baseband processing, power supply and cooling power consumption. Reception generally consumes less circuit power than transmission [23]. Therefore, we neglect the circuit power consumption at the destination. From (4) and (5), the EE metric is defined as

$$\mathrm{EE}\left(\boldsymbol{V}_S, \boldsymbol{W}_R\right) = \frac{R_{\mathrm{D}}}{\mathcal{P}\left(\boldsymbol{V}_S, \mathbf{W}_R\right)} \text{ [bit/Joule]}. \tag{6}$$

B. Robust EE Optimization Problem

At high signal-to-noise ratio conditions and with optimal pilot design, receiver can acquire an accurate CSI with training [26]. It is reasonable to assume that the CSI is perfectly available at receiver [11], [16], [17]. But transmitter can only acquire this estimated CSI through a finite-rate feedback channel, which introduces the quantization and feedback delay errors [16]. Consequently, the CSI at transmitter is inherently imperfect. Similar to most of the existing literature [15]–[17], we first assume that the relay has the perfect knowledge of the source-relay channel $\boldsymbol{H}_{SR}$ but it can only acquire an imperfect relay-destination channel $\boldsymbol{H}_{RD}$. However, we also consider the more generic senario where the perfect knowledge of the source-relay channel is also unavailable. As aforementioned, there exist two different types of imperfect CSI models, the statistically and deterministically imperfect CSI. Different from the works [11], [13], which study the robust EE optimization under the statistically imperfect CSI, we study the robust EE maximization for deterministically imperfect CSI. In general, the deterministic CSI errors take two different forms, additive CSI errors and multiplicative CSI errors. According to [18], [19], the CSI feedback and quantization errors are considered to be additive, while CSI calibration mismatch and the channel dynamic variations are regarded as multiplicative errors. By applying the two types of CSI errors to the relay-destination channel $\boldsymbol{H}_{RD}$, we have

$$\boldsymbol{H}_{RD} = \begin{cases} \widehat{\boldsymbol{H}}_{RD} + \boldsymbol{\Delta}_{RD}, & \|\boldsymbol{\Delta}_{RD}\|_2 \leq \epsilon_a, \\ \left(\boldsymbol{I}_{N_D} + \boldsymbol{E}_{RD}\right)\widehat{\boldsymbol{H}}_{RD}, & \|\boldsymbol{E}_{RD}\|_2 \leq \epsilon_m, \end{cases} \tag{7}$$

where $\widehat{\boldsymbol{H}}_{RD} \in \mathbb{C}^{N_D \times N_R}$ is the known nominal relay-destination channel, $\boldsymbol{\Delta}_{RD} \in \mathbb{C}^{N_D \times N_R}$ and $\boldsymbol{E}_{RD} \in \mathbb{C}^{N_D \times N_D}$ are the additive and multiplicative CSI errors, respectively, while ϵ_a and ϵ_m are the corresponding spectral norm bounds of the CSI errors. To focus on the underlying principles and without loss of generality, we consider additive CSI errors and multiplicative CSI errors separately[1]. Note that the spectral norm belongs to the unitarily-invariant norm sets, in which the norm-bounded terms are statistically independent and identical in all directions [15]. It also acts as the lower bound of all unitarily-invariant norms. Hence, for the same CSI errors, the spectral norm constrained case covers the largest uncertainty region [15]. Furthermore, when considering another popular Frobenius norm expression, we have $\|\boldsymbol{\Delta}_{RD}\|_2 \leq \|\boldsymbol{\Delta}_{RD}\|_F \leq \sqrt{N_D}\|\boldsymbol{\Delta}_{RD}\|_2$, which indicates that the spectral norm constrained CSI errors can also provide valuable insights for the Frobenius norm constrained case. Given the imperfect CSI specified by (7), the EE metric (6) also depends on $\boldsymbol{\Delta}_{RD} \varnothing \boldsymbol{E}_{RD}$ and, therefore, it is expressed as $\mathrm{EE}\left(\boldsymbol{V}_S, \boldsymbol{W}_R, \boldsymbol{\Delta}_{RD} \varnothing \boldsymbol{E}_{RD}\right)$.

Following the worst case robustness logic, the source covariance matrix $\boldsymbol{V}_S$ and relay beamforming matrix $\boldsymbol{W}_R$ are jointly designed by guaranteeing the maximum EE for all possible relay-destination channel realizations within the uncertainty region defined by (7). This robust EE optimization problem of MIMO AF relaying networks is formulated as

$$\begin{aligned} \max_{\boldsymbol{V}_S, \boldsymbol{W}_R} \min_{\boldsymbol{\Delta}_{RD} \varnothing \boldsymbol{E}_{RD}} \; & \mathrm{EE}\left(\boldsymbol{V}_S, \boldsymbol{W}_R, \boldsymbol{\Delta}_{RD} \varnothing \boldsymbol{E}_{RD}\right), \\ \text{s.t.} \; & \mathrm{Tr}(\boldsymbol{V}_S) \leq P_{S_{\max}}, \\ & \mathrm{Tr}\left(\boldsymbol{W}_R\left(\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}} + \sigma_r^2\boldsymbol{I}_{N_R}\right)\boldsymbol{W}_R^{\mathrm{H}}\right) \leq P_{R_{\max}}, \\ & \|\boldsymbol{\Delta}_{RD}\|_2 \leq \epsilon_a \text{ or } \|\boldsymbol{E}_{RD}\|_2 \leq \epsilon_m, \end{aligned} \tag{8}$$

where $P_{S_{\max}}$ and $P_{R_{\max}}$ are the maximum transmit powers of source and relay, respectively. As (8) contains the interrelated optimization variables and the semi-infinite CSI errors, the classical saddle point theory for concave-convex problems [27, Theorem 36.3] cannot be applied. According to [9], the max-min EE problem (8) can be reduced to a NP-hard sigmoidal programming [28, Theorem 1, page 15]. Therefore, it is also NP-hard and very difficult to solve directly.

III. Worst Case EE Maximization for Two-Hop MIMO AF Relaying

A. Derivation of Saddle Point

To simplify the intricate relationships among the optimization variables $\left\{\boldsymbol{V}_S, \boldsymbol{W}_R, \boldsymbol{\Delta}_{RD} \varnothing \boldsymbol{E}_{RD}\right\}$, we utilize the Woodbury matrix identity to equivalently transform (4) into

$$\begin{aligned} R_{\mathrm{D}} = & \frac{B}{2} \log\det\Big(\boldsymbol{I}_{N_R} + \sigma_r^{-2}\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}} - \sigma_r^{-2}\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}}\big(\\ & \qquad \boldsymbol{I}_{N_R} + \sigma_r^2\sigma_d^{-2}\boldsymbol{W}_R^{\mathrm{H}}\boldsymbol{H}_{RD}^{\mathrm{H}}\boldsymbol{H}_{RD}\boldsymbol{W}_R\big)^{-1}\Big) \\ = & \, \mathcal{R}\big(\boldsymbol{V}_S, \boldsymbol{W}_R, \boldsymbol{\Delta}_{RD} \varnothing \boldsymbol{E}_{RD}\big). \end{aligned} \tag{9}$$

Then the EE metric in (8) is rewritten as $\mathrm{EE}\left(\boldsymbol{V}_S, \boldsymbol{W}_R, \boldsymbol{\Delta}_{RD} \varnothing \boldsymbol{E}_{RD}\right) = \mathcal{R}\big(\boldsymbol{V}_S, \boldsymbol{W}_R, \boldsymbol{\Delta}_{RD} \varnothing \boldsymbol{E}_{RD}\big) / \mathcal{P}\big(\boldsymbol{V}_S, \boldsymbol{W}_R\big)$. Since (8) is not concave-convex in $\left\{\boldsymbol{V}_S, \boldsymbol{W}_R, \boldsymbol{\Delta}_{RD} \varnothing \boldsymbol{E}_{RD}\right\}$, it is difficult to solve it directly.

[1]Our work can easily be extended to the case having both additive and multiplicative CSI errors, namely, $\boldsymbol{H}_{RD} = \left(\boldsymbol{I}_{N_D} + \boldsymbol{E}_{RD}\right)\widehat{\boldsymbol{H}}_{RD} + \boldsymbol{\Delta}_{RD}$. This is because in this case, similar worst-case channel-diagonalizing structure can easily be derived by applying the results of this work for the additive CSI errors ($\boldsymbol{\Delta}_{RD}$) and the multiplicative CSI errors ($\boldsymbol{E}_{RD}\widehat{\boldsymbol{H}}_{RD}$).

Instead, we consider its counterpart, i.e., the following min-max EE problem,

$$\begin{aligned}\min_{\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}}\ \max_{\boldsymbol{V}_S,\boldsymbol{W}_R}\ & \mathrm{EE}\left(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\right)\\ \text{s.t.}\quad & \mathrm{Tr}(\boldsymbol{V}_S)\leq P_{S_{\max}},\\ & \mathrm{Tr}\left(\boldsymbol{W}_R\big(\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}}+\sigma_r^2\boldsymbol{I}_{N_R}\big)\boldsymbol{W}_R^{\mathrm{H}}\right)\leq P_{R_{\max}},\\ & \|\boldsymbol{\Delta}_{RD}\|_2\leq\epsilon_a \text{ or } \|\boldsymbol{E}_{RD}\|_2\leq\epsilon_m.\end{aligned}\tag{10}$$

Generally, $\max_x\min_y f(x,y)\leq\min_y\max_x f(x,y)$ [29, Section 5.4] holds implying that the max-min problem and the min-max problem are not identical. However, according to [30, Corollary 9.16], if there exists a saddle point for EE$(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD})$, then it is globally optimal for both problems (8) and (10). Hence, we first study the problem (10) and derive its optimal solution, and then prove that the obtained solution is indeed a saddle point of EE$(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD})$. Let's define the singular value decomposition (SVD) of $\boldsymbol{H}_{SR}$ and $\widehat{\boldsymbol{H}}_{RD}$ as

$$\boldsymbol{H}_{SR}=\boldsymbol{U}_{SR}\boldsymbol{\Sigma}_{SR}\boldsymbol{Q}_{SR}^{\mathrm{H}},\tag{11}$$

$$\widehat{\boldsymbol{H}}_{RD}=\widehat{\boldsymbol{U}}_{RD}\widehat{\boldsymbol{\Sigma}}_{RD}\widehat{\boldsymbol{Q}}_{RD}^{\mathrm{H}},\tag{12}$$

where $\boldsymbol{U}_{SR}\in\mathbb{C}^{N_R\times N_R}$ and $\boldsymbol{Q}_{SR}\in\mathbb{C}^{N_S\times N_S}$ as well as $\widehat{\boldsymbol{U}}_{RD}\in\mathbb{C}^{N_D\times N_D}$ and $\widehat{\boldsymbol{Q}}_{RD}\in\mathbb{C}^{N_R\times N_R}$ are the unitary singular matrices for $\boldsymbol{H}_{SR}$ and $\widehat{\boldsymbol{H}}_{RD}$, while the diagonal rectangular matrices $\boldsymbol{\Sigma}_{SR}\in\mathbb{C}^{N_R\times N_S}$ and $\widehat{\boldsymbol{\Sigma}}_{RD}\in\mathbb{C}^{N_D\times N_R}$ take $N_P=\min\{N_R,N_S\}$ singular values (SVs) $\left\{\sigma_{sr,1},\cdots,\sigma_{sr,N_P}\right\}$ of $\boldsymbol{H}_{SR}$ and $N_C=\min\{N_D,N_R\}$ SVs $\left\{\widehat{\sigma}_{rd,1},\cdots,\widehat{\sigma}_{rd,N_C}\right\}$ of $\widehat{\boldsymbol{H}}_{RD}$ as diagonal elements.

Theorem 1. *For the min-max EE problem* (10)*, the optimal source covariance matrix $\boldsymbol{V}_S^\star$, the optimal relay beamforming matrix $\boldsymbol{W}_R^\star$ and the worst-case CSI errors $\boldsymbol{\Delta}_{RD}^\star\varnothing\boldsymbol{E}_{RD}^\star$ satisfy*

$$\boldsymbol{V}_S^\star=\boldsymbol{Q}_{SR}\boldsymbol{\Sigma}_S\boldsymbol{Q}_{SR}^{\mathrm{H}},\tag{13}$$

$$\boldsymbol{W}_R^\star=\widehat{\boldsymbol{Q}}_{RD}\boldsymbol{\Sigma}_X\left(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\boldsymbol{\Sigma}_{SR}\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_{SR}^{\mathrm{H}}\right)^{-\frac{1}{2}}\boldsymbol{U}_{SR}^{\mathrm{H}},\tag{14}$$

$$\boldsymbol{\Delta}_{RD}^\star=-\widehat{\boldsymbol{U}}_{RD}\boldsymbol{\Lambda}_{RD}\widehat{\boldsymbol{Q}}_{RD}^{\mathrm{H}} \text{ or } \boldsymbol{E}_{RD}^\star=-\epsilon_m\boldsymbol{I}_{N_D},\tag{15}$$

where $\boldsymbol{\Sigma}_S=\mathrm{diag}\left\{\lambda_{s,1},\cdots,\lambda_{s,N_S}\right\}$ and $\boldsymbol{\Sigma}_X=\mathrm{diag}\{\sigma_{x,1},\cdots,\sigma_{x,N_C},0,\cdots,0\}\in\mathbb{C}^{N_R\times N_R}$, in which $\lambda_{s,i}$ for $1\leq i\leq N_S$ and $\sigma_{x,j}$ for $1\leq j\leq N_C$ replace $\boldsymbol{V}_S$ and $\boldsymbol{W}_R$ as the optimization scalar variables for the min-max EE problem (10), while the diagonal rectangular matrix $\boldsymbol{\Lambda}_{RD}\in\mathbb{C}^{N_D\times N_R}$ has the N_C diagonal elements $\min\{\widehat{\sigma}_{rd,1},\epsilon_a\},\cdots,\min\{\widehat{\sigma}_{rd,N_C},\epsilon_a\}$.

Proof. See Appendix A. □

Theorem 2. *The optimal solution $\left\{\boldsymbol{V}_S^\star,\boldsymbol{W}_R^\star,\boldsymbol{\Delta}_{RD}^\star\varnothing\boldsymbol{E}_{RD}^\star\right\}$ of the min-max EE problem (10) provided by Theorem 1 is the saddle point of the EE metric $\mathrm{EE}\big(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big)$, i.e.,*

$$\begin{aligned}&\mathrm{EE}\big(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}^\star\varnothing\boldsymbol{E}_{RD}^\star\big)\leq\mathrm{EE}\big(\boldsymbol{V}_S^\star,\boldsymbol{W}_R^\star,\boldsymbol{\Delta}_{RD}^\star\varnothing\boldsymbol{E}_{RD}^\star\big)\\ &\leq\mathrm{EE}\big(\boldsymbol{V}_S^\star,\boldsymbol{W}_R^\star,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big),\end{aligned}\tag{16}$$

holds for any feasible $\boldsymbol{V}_S$, $\boldsymbol{W}_R$ and $\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}$. According to [30, Corollary 9.16], it is also optimal for the original max-min EE problem (8).

Proof. See Appendix B. □

According to Theorems 1 and 2, the optimal $\boldsymbol{V}_S^\star$, $\boldsymbol{W}_R^\star$ and $\boldsymbol{\Delta}_{RD}^\star\varnothing\boldsymbol{E}_{RD}^\star$ that solve the max-min EE problem (8) all have the channel-diagonalizing structure. Calculating the optimal $\boldsymbol{V}_S^\star$ and $\boldsymbol{W}_R^\star$ becomes determining the values of $\boldsymbol{\lambda}_s=\left[\lambda_{s,1}\cdots\lambda_{s,N_S}\right]^{\mathrm{T}}$ and $\boldsymbol{\sigma}_x=\left[\sigma_{x,1}^2\cdots\sigma_{x,N_C}^2\right]^{\mathrm{T}}$.

B. Proposed Alternating Optimization Algorithm

Based on Theorem 1, the original max-min EE problem (8) with matrix variables can be equivalently transformed into the problem (17) with scalar variables, as shown at the top of the next page, where $N_L=\min\left\{N_P,N_C\right\}$ and for $1\leq i\leq N_C$,

$$\widetilde{\sigma}_{rd,i}=\begin{cases}\left(\widehat{\sigma}_{rd,i}-\epsilon_a\right)^+ & \text{for additive CSI errors,}\\ (1-\epsilon_m)^+\widehat{\sigma}_{rd,i} & \text{for multiplicative CSI errors,}\end{cases}\tag{18}$$

Compared to the original problem (8), the number of optimization variables in the problem (17) is significantly reduced, namely, from $N_S^2+N_R^2+N_DN_R$ to N_S+N_C. In order to efficiently solve the problem (17), the fractional programming theory [22], [30] is first introduced.

Lemma 1. *([22], [30]) Given a fractional function $f(\boldsymbol{A})=\frac{N(\boldsymbol{A})}{G(\boldsymbol{A})}$, provided that $N(\boldsymbol{A})$ and $G(\boldsymbol{A})$ are concave and convex w.r.t. $\boldsymbol{A}$, respectively, then $f(\boldsymbol{A})$ is quasi-concave. By introducing an auxiliary variable η, a single-parameter subtractive function is defined as*

$$F(\eta)=\max_{\boldsymbol{A}}\ N(\boldsymbol{A})-\eta G(\boldsymbol{A}).\tag{19}$$

The inner maximization problem of (19) *is concave w.r.t. $\boldsymbol{A}$ for any fixed η, and $F(\eta)$ is a decreasing function of η. Moreover, the problem of maximizing $f(\boldsymbol{A})$ is equivalent to finding the zero point of $F(\eta)$, and Dinkelbach's method can be invoked for finding $F(\eta)=0$, which is guaranteed to converge to a globally optimal solution of maximizing $f(\boldsymbol{A})$ [22].*

Taking the second derivative of the numerator of the objective function in (17) w.r.t. $\boldsymbol{\lambda}_s$ for fixed $\boldsymbol{\sigma}_x$, it is seen that the numerator of the objective function is concave w.r.t. $\boldsymbol{\lambda}_s$. The denominator of the objective function in (17) is linear w.r.t. $\boldsymbol{\lambda}_s$ given $\boldsymbol{\sigma}_x$. According to Lemma 1, the problem (17) is quasi-concave for $\boldsymbol{\lambda}_s$ given $\boldsymbol{\sigma}_x$. Similarly, the problem (17) is quasi-concave for $\boldsymbol{\sigma}_x$ given $\boldsymbol{\lambda}_s$. Therefore, it can be efficiently tackled by an alternating optimization between the subproblem of optimizing $\boldsymbol{\lambda}_s$ for fixed $\boldsymbol{\sigma}_x$ and that of optimizing $\boldsymbol{\sigma}_x$ for fixed $\boldsymbol{\lambda}_s$.

By introducing the auxiliary variable η based on Lemma 1, we transform (17) into the following single-parameter subtractive problem

$$\begin{aligned}\max_{\boldsymbol{\lambda}_s,\boldsymbol{\sigma}_x}\quad & \sum\nolimits_{i=1}^{N_L}\Big(\log\big(\frac{1+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}{1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}\big)\\ &+\log\big(1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}\big)\Big)\\ &-\eta\Big(\sum\nolimits_{i=1}^{N_S}\lambda_{s,i}+\sum\nolimits_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2+P_C\Big),\\ \text{s.t.}\quad & \sum\nolimits_{i=1}^{N_S}\lambda_{s,i}\leq P_{S_{\max}},\ \ \sum\nolimits_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2\leq P_{R_{\max}}.\end{aligned}\tag{20}$$

For given η, since (20) is strictly concave w.r.t $\boldsymbol{\lambda}_s$ for fixed $\boldsymbol{\sigma}_x$ and vice versa, the Lagrangian dual method can be adopted for obtaining the corresponding optimal solutions to

$$\max_{\boldsymbol{\lambda}_s,\boldsymbol{\sigma}_x} \quad \frac{\sum_{i=1}^{N_L}\log\left(\frac{1+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}{1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}\right)+\sum_{i=1}^{N_L}\log\left(1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}\right)}{\sum_{i=1}^{N_S}\lambda_{s,i}+\sum_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2+P_C},$$
$$\text{s.t.} \quad \sum_{i=1}^{N_S}\lambda_{s,i}\le P_{S_{\max}}, \quad \sum_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2\le P_{R_{\max}}, \tag{17}$$

$$\sigma_{x,i}^2(\boldsymbol{\lambda}_s;\eta)=\begin{cases}\frac{\sqrt{\sigma_r^{-4}\sigma_{sr,i}^4\lambda_{s,i}^2+\frac{4\sigma_d^{-2}\sigma_{sr,i}^2\lambda_{s,i}\widetilde{\sigma}_{rd,i}^2}{\ln 2(\beta+\eta)\sigma_r^2}}-\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}-2}{2\sigma_r^2\sigma_d^{-2}\widetilde{\sigma}_{rd,i}^2}, & \widetilde{\sigma}_{rd,i}>0 \text{ and } 1\le i\le N_L,\\ 0, & \widetilde{\sigma}_{rd,i}=0 \text{ or } N_L+1\le i\le N_C,\end{cases}, \tag{22}$$

$$\lambda_{s,i}(\boldsymbol{\sigma}_x;\eta)=\begin{cases}\frac{\sqrt{\sigma_r^4\sigma_d^{-4}\sigma_{x,i}^4\widetilde{\sigma}_{rd,i}^4+\frac{4\sigma_d^{-2}\sigma_{sr,i}^2\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}{\ln 2(\mu+\eta)}}-\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2-2}{2\sigma_r^{-2}\sigma_{sr,i}^2}, & \sigma_{sr,i}>0 \text{ and } 1\le i\le N_L,\\ 0, & \sigma_{sr,i}=0 \text{ or } N_L+1\le i\le N_S,\end{cases} \tag{23}$$

$$\eta(\boldsymbol{\lambda}_s,\boldsymbol{\sigma}_x)=\frac{\sum_{i=1}^{N_L}\log\left(\frac{1+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}{1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}\right)+\sum_{i=1}^{N_L}\log\left(1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}\right)}{\sum_{i=1}^{N_S}\lambda_{s,i}+\sum_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2+P_C}. \tag{24}$$

the respective subproblems. Specifically, given η, we introduce the Lagrangian dual function of (20) as

$$\begin{aligned}&L(\boldsymbol{\lambda}_s,\boldsymbol{\sigma}_x,\mu,\beta;\eta)\\&=\sum_{i=1}^{N_L}\log\left(\frac{1+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}{1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}\right)\\&+\sum_{i=1}^{N_L}\log\left(1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}\right)\\&-(\eta+\mu)\sum_{i=1}^{N_S}\lambda_{s,i}-(\eta+\beta)\sum_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2+C,\end{aligned} \tag{21}$$

where μ and β are the Lagrangian multipliers for the source and relay transmit power constraints, respectively, and $C=\mu P_{S_{\max}}+\beta P_{R_{\max}}-\eta P_C$. For the subproblem of optimizing $\boldsymbol{\sigma}_x$ given $\boldsymbol{\lambda}_s$, we take the derivative of $L(\boldsymbol{\lambda}_s,\boldsymbol{\sigma}_x,\mu,\beta;\eta)$ w.r.t. $\sigma_{x,i}$ and utilize Karush-Kuhn-Tucker (KKT) conditions to obtain the optimal $\boldsymbol{\sigma}_x$ for fixed $\boldsymbol{\lambda}_s$, denoted by $\boldsymbol{\sigma}_x(\boldsymbol{\lambda}_s;\eta)$, as (22) at the top of this page. where β satisfies $\beta\Big(\sum_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2-P_{R_{\max}}\Big)=0$ and it can be determined by the bisection search owing to the monotonically decreasing property of $\sigma_{x,i}^2(\boldsymbol{\lambda}_s;\eta)$ w.r.t. β. Similarly, for fixed $\boldsymbol{\sigma}_x$, we have the optimal $\boldsymbol{\lambda}_s$, denoted by $\boldsymbol{\lambda}_s(\boldsymbol{\sigma}_x;\eta)$, in (23) at the top of this page. where owing to the monotonically decreasing property of $\boldsymbol{\lambda}_s(\boldsymbol{\sigma}_x;\eta)$ w.r.t μ, μ is chosen by the bisection search to ensure $\mu\Big(\sum_{i=1}^{N_S}\lambda_{s,i}-P_{S_{\max}}\Big)=0$.

For efficiently realizing the worst-case EE maximization, we apply Dinkelbach's method to both the subproblems of (20) to update η by utilizing the optimal $\boldsymbol{\sigma}_x$ in (22) for fixed $\boldsymbol{\lambda}_s$ and by utilizing the optimal $\boldsymbol{\lambda}_s$ in (23) for fixed $\boldsymbol{\sigma}_x$, respectively. Dinkelbach's method is an iterative optimization process, which converges when the zero objective value of the problem (20) is realized. Specifically, the update of η in Dinkelbach's method based on $\{\boldsymbol{\lambda}_s,\boldsymbol{\sigma}_x\}$ is given by (24) at the top of this page. Integrating (22) to (24), the proposed alternating optimization for the worst-case EE maximization under additive or multiplicative CSI errors is summarized in Algorithm 1.

C. Convergence and the speedup strategy

For characterizing the convergence, let us consider the arbitrary feasible initial value $\boldsymbol{\lambda}_s^{(n)}$ and $\eta^{(n)}$ for the nth iteration of Algorithm 1. According to Lemma 1, given $\boldsymbol{\lambda}_s^{(n)}$, the $\boldsymbol{\sigma}_x$ related subproblem of (20) is strictly concave, and we can obtain the unique and globally optimal $\boldsymbol{\sigma}_x^{(n+1)}$ in (22), which implies that after step 2 of Algorithm 1, $\mathrm{EE}\big(\boldsymbol{V}_S^{(n)},\boldsymbol{W}_R^{(n)},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)\le\mathrm{EE}\big(\boldsymbol{V}_S^{(n)},\boldsymbol{W}_R^{(n+1)},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)$. Given $\boldsymbol{\sigma}_x^{(n+1)}$, the $\boldsymbol{\lambda}_s$ related subproblem of (20) is also concave, and the globally optimal and unique $\boldsymbol{\lambda}_s^{(n+1)}$ is derived in (23), which implies that after step 3 of Algorithm 1, we have $\mathrm{EE}\big(\boldsymbol{V}_S^{(n)},\boldsymbol{W}_R^{(n+1)},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)\le\mathrm{EE}\big(\boldsymbol{V}_S^{(n+1)},\boldsymbol{W}_R^{(n+1)},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)$. Combining these two non-strict inequalities, we generally have $\mathrm{EE}\big(\boldsymbol{V}_S^{(n)},\boldsymbol{W}_R^{(n)},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)<\mathrm{EE}\big(\boldsymbol{V}_S^{(n+1)},\boldsymbol{W}_R^{(n+1)},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)$ after the $(n+1)$th iteration. It is widely exploited that the worst-case robust maximization is generally upper bounded by the corresponding perfect-case maximization [11], [13]. By setting $\sigma_{rd,i}=\widehat{\sigma}_{rd,i}$, $1\le i\le N_C$, in the problem (20) to indicate that the relay's knowledge of $\boldsymbol{H}_{RD}$ is perfect, it becomes the perfect-case EE maximization, which has been solved in [11] and the solution provides an effective upper bound for our worst-case EE. Since the achievable worst-case EE of Algorithm 1 is non-decreasing and upper bounded, we conclude that Algorithm 1 is guaranteed to converge to a stationary point $\{\boldsymbol{\lambda}_s^{\mathrm{st}},\boldsymbol{\sigma}_x^{\mathrm{st}}\}$ of the problem (17).

Substituting this stationary point of the problem (17) into (13) and (14) of Theorem 1 yields a locally optimal solution $\{\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}}\}$ to the problem (8) given $\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}$, which satisfies $\mathrm{EE}\big(\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)\ge\mathrm{EE}\big(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)$ and $(\boldsymbol{V}_S,\boldsymbol{W}_R)\in\mathcal{U}(\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}};\delta_d)$ (Here, $\mathcal{U}(\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}};\delta_d)$ denotes the vicinity sphere of point $\{\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}}\}$ with radius δ_d). Moreover, given $\{\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}}\}$, it is observed that the worst-case additive/multiplicative error to the problem (8) is still $\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}$, and thus we have $\mathrm{EE}\big(\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)\le\mathrm{EE}\big(\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}},\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big)$. Therefore, for any $(\boldsymbol{V}_S,\boldsymbol{W}_R)\in\mathcal{U}(\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}};\delta_d)$, it yields

Algorithm 1 The proposed alternating optimization for solving (17)

Input: The initial $\boldsymbol{\lambda}_s^{(0)}$ and $\eta^{(0)}$; a sufficiently small tolerance threshold $\zeta > 0$; the iteration index $n = 0$;

1: **repeat**

2: Fix $\boldsymbol{\lambda}_s = \boldsymbol{\lambda}_s^{(n)}$, start from $\eta = \eta^{(n)}$, apply Dinkelbach method to iteratively optimize between $\boldsymbol{\sigma}_x\big(\boldsymbol{\lambda}_s^{(n)};\eta\big)$ of (22) and $\eta\big(\boldsymbol{\lambda}_s^{(n)},\boldsymbol{\sigma}_x\big)$ of (24) to obtain $\boldsymbol{\sigma}_x^{(n+1)}$ and $\widetilde{\eta}$ that realize the zero objective value of (20);

3: Fix $\boldsymbol{\sigma}_x = \boldsymbol{\sigma}_x^{(n+1)}$, start from $\eta = \widetilde{\eta}$, apply Dinkelbach method to iteratively optimize between $\boldsymbol{\lambda}_s\big(\boldsymbol{\sigma}_x^{(n+1)};\eta\big)$ of (23) and $\eta\big(\boldsymbol{\lambda}_s,\boldsymbol{\sigma}_x^{(n+1)}\big)$ of (24) to obtain $\boldsymbol{\lambda}_s^{(n+1)}$ and $\eta^{(n+1)}$ that realize the zero objective value of (20);

4: $n = n + 1$;

5: **until** $\big|\mathrm{EE}\big(\boldsymbol{V}_S^{(n)},\boldsymbol{W}_R^{(n)},\boldsymbol{\Delta}_{RD}^{(n)}\varnothing\boldsymbol{E}_{RD}^{(n)}\big) - \mathrm{EE}\big(\boldsymbol{V}_S^{(n-1)},\boldsymbol{W}_R^{(n-1)},\boldsymbol{\Delta}_{RD}^{(n-1)}\varnothing\boldsymbol{E}_{RD}^{(n-1)}\big)\big| \le \zeta$;

Output: $\mathrm{EE}\big(\boldsymbol{V}_S^{\star},\boldsymbol{W}_R^{\star},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)$ with $\boldsymbol{V}_S^{\star} = \boldsymbol{V}_S^{(n)}$, $\boldsymbol{W}_R^{\star} = \boldsymbol{W}_R^{(n)}$ and $\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star} = \boldsymbol{\Delta}_{RD}^{(n)}\varnothing\boldsymbol{E}_{RD}^{(n)}$;

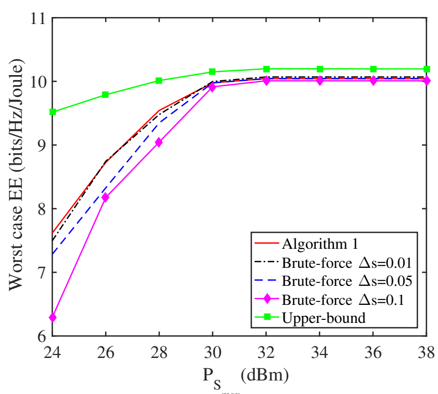


Fig. 1. The worst-case EE performance as the functions of the source maximum transmit power $P_{S_{\max}}$ for Algorithm 1 and the brute-force search, in comparison with the upper bound.

$\mathrm{EE}\big(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big) \le \mathrm{EE}\big(\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big) \le \mathrm{EE}\big(\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}},\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big)$, based on which we can conclude that $\{\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}}\ \boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\}$ is a local saddle point (stationary point) of the max-min EE problem (8). In addition, it is readily inferred that by setting $\eta = \eta^{(n)}$ and $\eta = \widetilde{\eta}$ for Step 2 and Step 3, respectively, the objective function value of problem (17) is also non-decreasing between the inner iteration of Step 2 and that of Step 3. In other words, with the obtained $\{\boldsymbol{W}_R^{(n)},\boldsymbol{V}_S^{(n)}\}$ after the nth outer iteration, we further have $\mathrm{EE}\big(\boldsymbol{V}_S^{(n)},\boldsymbol{W}_R^{(n)},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big) \le \cdots \le \mathrm{EE}\big(\boldsymbol{V}_S^{(n)},\boldsymbol{W}_R^{(n_{in,k})},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big) \le \cdots \le \mathrm{EE}\big(\boldsymbol{V}_S^{(n)},\boldsymbol{W}_R^{(n+1)},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big) \le \cdots \le \mathrm{EE}\big(\boldsymbol{V}_S^{(n_{in,k})},\boldsymbol{W}_R^{(n+1)},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big) \le \cdots \le \mathrm{EE}\big(\boldsymbol{V}_S^{(n+1)},\boldsymbol{W}_R^{(n+1)},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)$, where $\boldsymbol{W}_R^{(n_{in,k})}$ and $\boldsymbol{V}_S^{(n_{in,k})}$ denote the optimized $\boldsymbol{W}_R$ and $\boldsymbol{V}_S$ after the kth inner iteration of Step 2 and Step 3 in the $n{+}1$th outer iteration, respectively. Overall, by defining $\eta = \eta^{(n)}$ and $\eta = \widetilde{\eta}$ for Step 2 and Step 3 of Algorithm 1, respectively, the achieved EE value for the problem (17) is guaranteed to be non-decreasing both in the inner iteration of Step 2/Step 3 and in the outer iteration between Step 2 and Step 3, which is beneficial for speeding up the convergence of Algorithm 1.

D. Optimality and Complexity

Although a local stationary point $\{\boldsymbol{\lambda}_s^{\mathrm{st}},\boldsymbol{\sigma}_x^{\mathrm{st}}\}$ of the problem (17) found by Algorithm 1 is not necessary an optimal solution, it can be seen from the above discussions that the solution $\big\{\boldsymbol{V}_S^{\mathrm{st}},\boldsymbol{W}_R^{\mathrm{st}},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big\}$ associated with $\{\boldsymbol{\lambda}_s^{\mathrm{st}},\boldsymbol{\sigma}_x^{\mathrm{st}}\}$ as in Theorem 1 is also a local saddle point to the max-min EE problem (8). A further advantage of Algorithm 1 for solving the non-convex problem (17) is its low computational complexity, on the order of $\mathcal{O}\big(I_{\mathrm{ite}}\big(N_S\log_2(N_S) + N_C\log_2(N_C)\big)\big)$ due to water-filling solution in each step, where I_{ite} denotes the total number of outer-inner iterations. As presented in Section V, Algorithm 1 converges within 10 iterations for both outer and inner loops. This should be contrasted with the computational complexity of brute-force search for finding an optimal solution to the problem (17), namely, $\mathcal{O}\left(\left(\frac{P_{S_{\max}}}{\Delta s}\right)^{N_S}\left(\frac{P_{R_{\max}}}{\Delta s}\right)^{N_C}\right)$, where Δs is the step length. To obtain an accurate solution, a small step length Δs is required, which imposes extremely high complexity. In addition, an upper bound for the objective function of problem (17), denoted as $f_{\mathrm{up}}(\boldsymbol{\lambda}_s,\boldsymbol{\sigma}_x)$, is given by (25) at the top of the next page. The inequality in (25) is derived according to the identity $a+b \ge 2\sqrt{ab}$. Note that this upper bound is jointly quasi-concave w.r.t $\{\boldsymbol{\lambda}_s,\boldsymbol{\sigma}_x\}$, to which the globally optimal solution is available. Moreover, we readily find that this upper bound becomes tight when the sufficiently high transmit powers at source and relay are considered.

To demonstrate the effectiveness of our proposed alternating optimization, Fig. 1 firstly shows the worst-case EE performance as the functions of the source transmit power $P_{S_{\max}}$ for Algorithm 1 and the brute-force search, in comparison with the upper bound (25), where a small-scale system setup with $N_S/N_R/N_D$=2/2/2 and $P_{R_{\max}}$=30 dBm is considered. Then Table I compares the execution times of Algorithm 1 and brute-force search for solving the problem (17). It can be seen from Fig. 1 and Table I that there is almost no loss of optimality by using Algorithm 1, which imposes a dramatically lower complexity than the brute-force search. We also observe from Fig. 1 that the gap between the upper bound and optimal solution to the problem (17) is much reduced at high source transmit power.

E. Extension to Imperfect Source-Relay Channel

Since the relay can estimate $\boldsymbol{H}_{SR}$ with higher accuracy, the deterministically imperfect model, such as the one adopted for $\boldsymbol{H}_{RD}$ in (7), is inappropriate for $\boldsymbol{H}_{SR}$. It is more appropriate to adopt the statistically imperfect CSI model [11], [13], [14] to express $\boldsymbol{H}_{SR}$ as

$$\boldsymbol{H}_{SR} = \boldsymbol{R}_R^{\frac{1}{2}}\widetilde{\boldsymbol{H}}_{SR}\boldsymbol{R}_S^{\frac{1}{2}}, \tag{26}$$

where the positive semidefinite $\boldsymbol{R}_R \in \mathbb{C}^{N_R\times N_R}$ and $\boldsymbol{R}_S \in \mathbb{C}^{N_S\times N_S}$ are the relay and source spatial correlation matrices, respectively, while $\widetilde{\boldsymbol{H}}_{SR} \in \mathbb{C}^{N_R\times N_S}$ is a random matrix whose elements are i.i.d. complex Gaussian variables with the distribution $\mathcal{CN}(0,1)$. Both $\boldsymbol{R}_R$ and $\boldsymbol{R}_S$ are available at the relay but the instantaneous $\widetilde{\boldsymbol{H}}_{SR}$ is unknown. Thus the

$$f_{\rm up}(\boldsymbol{\lambda}_s,\boldsymbol{\sigma}_x)=\frac{\log\left(1+\frac{\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2\sigma_{sr,i}^2\lambda_{s,i}}{1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}\right)}{\sum_{i=1}^{N_S}\lambda_{s,i}+\sum_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2+P_C}\leq\frac{\log\left(1+\frac{\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2\sigma_{sr,i}^2\left(\lambda_{s,i}+\sigma_r^2\sigma_{x,i}^2\right)}{\sigma_r^{-2}\sigma_{sr,i}^2+\sigma_r^2\sigma_d^{-2}\widetilde{\sigma}_{rd,i}^2+2\sqrt{\sigma_{sr,i}^2\sigma_d^{-2}\widetilde{\sigma}_{rd,i}^2}}\right)}{\sum_{i=1}^{N_S}\lambda_{s,i}+\sum_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2+P_C}. \tag{25}$$

TABLE I
COMPARISON OF EXECUTION TIMES BY THE PROPOSED ALTERNATING OPTIMIZATION AND BRUTE-FORCE SEARCH

Time (s) \ Power(dBm) / Method	24	26	28	30	32	34	36	38
Algorithm 1	0.033	0.038	0.061	0.067	0.070	0.071	0.081	0.062
Brute-force search $\Delta s=0.1$	0.039	0.042	0.0642	0.072	0.147	0.278	1.363	7.558
Brute-force search $\Delta s=0.05$	0.042	0.057	0.071	0.191	1.076	7.272	110.524	846.584
Brute-force search $\Delta s=0.01$	1.381	8.116	79.445	744.259	4.874E+3	1.718E+4	5.305E+4	1.599E+5

instantaneous robust EE optimization (8) is infeasible, and the following average EE metric can be considered [13]

$$\begin{aligned}\widetilde{\rm EE}\left(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\right)&=\frac{\mathsf{E}\left[R_{\rm D}\left(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\right)\right]}{\mathsf{E}\left[\mathcal{P}\left(\boldsymbol{V}_S,\mathbf{W}_R\right)\right]}\\&=\frac{\mathsf{E}\left[R_{\rm D}\left(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\right)\right]}{\frac{P_S}{\tau_s}+\frac{\mathsf{E}[P_R]}{\tau_r}+P_C},\end{aligned} \tag{27}$$

where the expectation is w.r.t. the distribution of $\widetilde{\boldsymbol{H}}_{SR}$. Then the robust average EE maximization problem for the two-hop MIMO relaying network is given by

$$\begin{aligned}\max_{\boldsymbol{V}_S,\boldsymbol{W}_R}\ \min_{\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}}\ &\widetilde{\rm EE}\left(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\right),\\ \text{s.t.}\ \ &{\rm Tr}(\boldsymbol{V}_S)\leq P_{S_{\max}},\\ &\mathsf{E}\left[{\rm Tr}\left(\boldsymbol{W}_R\left(\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\rm H}+\sigma_r^2\boldsymbol{I}_{N_R}\right)\boldsymbol{W}_R^{\rm H}\right)\right]\leq P_{R_{\max}},\\ &\|\boldsymbol{\Delta}_{RD}\|_2\leq\epsilon_a\ \text{or}\ \|\boldsymbol{E}_{RD}\|_2\leq\epsilon_m.\end{aligned} \tag{28}$$

By defining the eigenvalue decompositions (EVDs) of $\boldsymbol{R}_R$ and $\boldsymbol{R}_S$ as $\boldsymbol{R}_R=\boldsymbol{U}_R\boldsymbol{\Lambda}_R\boldsymbol{U}_R^{\rm H}$ and $\boldsymbol{R}_S=\boldsymbol{U}_S\boldsymbol{\Lambda}_S\boldsymbol{U}_S^{\rm H}$, respectively, where the unitary matrices $\boldsymbol{U}_R\in\mathbb{C}^{N_R\times N_R}$ and $\boldsymbol{U}_S\in\mathbb{C}^{N_S\times N_S}$ consist of the eigenvectors of $\boldsymbol{R}_R$ and $\boldsymbol{R}_S$, respectively, we have the following Theorem.

Theorem 3. *For the robust average EE maximization* (28)*, the worst-case error $\boldsymbol{\Delta}_{RD}^\star\varnothing\boldsymbol{E}_{RD}^\star$ is the same as that given in Theorem 1, with the structures of the optimal $\boldsymbol{V}_S^\star$ and $\boldsymbol{W}_R^\star$ given by*

$$\begin{aligned}\boldsymbol{V}_S^\star&=\boldsymbol{U}_S\boldsymbol{\Sigma}_S\boldsymbol{U}_S^{\rm H},\\ \boldsymbol{W}_R^\star&=\widehat{\boldsymbol{Q}}_{RD}\boldsymbol{\Sigma}_X\left(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\boldsymbol{\Sigma}_{SR}\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_{SR}^{\rm H}\right)^{-\frac{1}{2}}\boldsymbol{U}_R^{\rm H}.\end{aligned} \tag{29}$$

Proof. See Appendix C. □

Based on Theorem 3, the matrix-variable robust average EE problem (28) can be equivalently transformed into a scalar-variable one. As shown in [11], this scalar-variable problem consists of two concave subproblems due to the concavity and monotonicity of the function $\mathsf{E}[\log(\cdot)]$. However, evaluating $\mathsf{E}[\log(\cdot)]$ imposes high-complexity. To solve (28) efficiently, a deterministic approximation of the average EE is required. We apply Jensen's inequality to the concave function $\mathsf{E}[\log(\cdot)]$ to derive the analytical upper bound of the average EE [11]. Then the proposed alternating optimization can readily be applied to this upper-bound average EE optimization.

Remark 1: The relay needs to feed back the optimal covariance matrix $\boldsymbol{V}_S^\star$ to the source, which introduces the feedback errors to $\boldsymbol{V}_S^\star$. When the source covariance matrix error $\Delta\boldsymbol{V}_S\in\mathbb{C}^{N_S\times N_S}$ is taken into account, the proof of Appendix A is not applicable, since both the numerator and denominator of the EE metric contain the semi-infinite $\Delta\boldsymbol{V}_S$. A possible solution is to consider a lower bound optimization of this truly robust EE design, where the possible maximum total power consumption under the spectral norm constrained $\Delta\boldsymbol{V}_S$ is adopted. Then $\Delta\boldsymbol{V}_S$ is only contained in the rate function and the corresponding minimum achievable rate in (4) is readily observed at $\Delta\boldsymbol{V}_S^\star=-\epsilon\boldsymbol{I}_{N_S}$. Since the resultant lower-bound robust EE optimization is similar to that of (8) or (28), the channel-diagonalizing structured new relay beamforming $\boldsymbol{W}_R^\star$ can still be proved following the proof in Appendix A.

IV. Extension to Multihop MIMO AF Relaying Networks

An N_S-antenna source transmits signals to an N_D-antenna destination via K N_R-antenna relays R_k, for $1\leq k\leq K$. Denote the source-relay channel by $\boldsymbol{H}_{SR_1}\in\mathbb{C}^{N_R\times N_S}$, the relay-relay channels by $\boldsymbol{H}_{R_kR_{k+1}}\in\mathbb{C}^{N_R\times N_R}$ for $1\leq k\leq K-1$, and the relay-destination channel by $\boldsymbol{H}_{R_KD}\in\mathbb{C}^{N_D\times N_R}$. The received signals at each relay and the destination are given by (30) and (31), respectively, at the top of the next page. where $\boldsymbol{n}_{R_k}\in\mathbb{C}^{N_R}$, $1\leq k\leq K$, is the AWGN vector at relay k with the covariance matrix $\sigma_r^2\boldsymbol{I}_{N_R}$, and $\boldsymbol{n}_{R_D}\in\mathbb{C}^{N_D}$ is the AWGN vector at the destination with the covariance matrix $\sigma_d^2\boldsymbol{I}_{N_D}$, while $\boldsymbol{W}_{R_k}$, $1\leq k\leq K$, is the beamforming matrix of relay k. Since typically the relays chosen are static or at most slowly mobile w.r.t. the source, the source-relay channel and all relay-relay channels can be acquired with high precision through training [31]. We further assume that the estimated $\boldsymbol{H}_{SR_1}$ and $\boldsymbol{H}_{R_kR_{k+1}}$, $1\leq k\leq K-2$, can be transmitted to relay K perfectly. Thus we assume that the perfect $\{\boldsymbol{H}_{SR_1},\boldsymbol{H}_{R_kR_{k+1}},1\leq k\leq K-1\}$ are available at relay K, and consider the deterministically imperfect $\boldsymbol{H}_{R_KD}$ with the additive or multiplicative CSI errors

$$\boldsymbol{H}_{R_KD}=\begin{cases}\widehat{\boldsymbol{H}}_{R_KD}+\boldsymbol{\Delta}_{R_KD}, & \|\boldsymbol{\Delta}_{R_KD}\|_2\leq\epsilon_a,\\ \left(\boldsymbol{I}_{N_D}+\boldsymbol{E}_{R_KD}\right)\widehat{\boldsymbol{H}}_{R_KD}, & \|\boldsymbol{E}_{R_KD}\|_2\leq\epsilon_m,\end{cases} \tag{32}$$

where $\widehat{\boldsymbol{H}}_{R_KD}\in\mathbb{C}^{N_D\times N_R}$ is the known nominal relay-destination channel, $\boldsymbol{\Delta}_{R_KD}\in\mathbb{C}^{N_D\times N_R}$ and $\boldsymbol{E}_{R_KD}\in\mathbb{C}^{N_D\times N_D}$ are the corresponding additive and multiplicative

$$\boldsymbol{y}_{R_k} = \begin{cases} \boldsymbol{H}_{SR_1}\boldsymbol{s}+\boldsymbol{n}_{R_1}, & k=1,\\ \left(\prod_{i=1}^{k-1}\boldsymbol{H}_{R_iR_{i+1}}\boldsymbol{W}_{R_i}\right)\boldsymbol{H}_{SR_1}\boldsymbol{s}+\sum_{m=1}^{k-1}\prod_{i=m}^{k-1}\boldsymbol{H}_{R_iR_{i+1}}\boldsymbol{W}_{R_i}\boldsymbol{n}_{R_i}+\boldsymbol{n}_{R_k}, & k=2,\cdots,K, \end{cases} \tag{30}$$

$$\begin{aligned} \boldsymbol{y}_D = & \boldsymbol{H}_{R_KD}\boldsymbol{y}_{R_K}+\boldsymbol{n}_{R_D} = \boldsymbol{H}_{R_KD}\boldsymbol{W}_{R_K}\left(\prod_{i=1}^{K-1}\boldsymbol{H}_{R_iR_{i+1}}\boldsymbol{W}_{R_i}\right)\boldsymbol{H}_{SR_1}\boldsymbol{s}\\ & +\boldsymbol{H}_{R_KD}\boldsymbol{W}_{R_K}\left(\sum_{m=1}^{K-1}\prod_{i=m}^{K-1}\boldsymbol{H}_{R_iR_{i+1}}\boldsymbol{W}_{R_i}\boldsymbol{n}_{R_i}\right)+\boldsymbol{H}_{R_KD}\boldsymbol{W}_{R_K}\boldsymbol{n}_{R_K}+\boldsymbol{n}_{R_D}, \end{aligned} \tag{31}$$

$$P_{R_k} = \begin{cases} \mathrm{Tr}\left(\boldsymbol{W}_{R_1}\left(\boldsymbol{H}_{SR_1}\boldsymbol{V}_S\boldsymbol{H}_{SR_1}^{\mathrm{H}}+\sigma_r^2\boldsymbol{I}_{N_R}\right)\boldsymbol{W}_{R_1}^{\mathrm{H}}\right), & k=1,\\ \mathrm{Tr}\left(\boldsymbol{W}_{R_k}\left(\prod_{i=1}^{k-1}\boldsymbol{H}_{R_iR_{i+1}}\boldsymbol{W}_{R_i}\right)\boldsymbol{H}_{SR_1}\boldsymbol{V}_S\boldsymbol{H}_{SR_1}^{\mathrm{H}}\left(\prod_{i=1}^{k-1}\boldsymbol{H}_{R_iR_{i+1}}\boldsymbol{W}_{R_i}\right)^{\mathrm{H}}\right. & \\ \left.+\sigma_r^2\sum_{m=1}^{k-1}\left(\prod_{i=m}^{k-1}\boldsymbol{H}_{R_iR_{i+1}}\boldsymbol{W}_{R_i}\right)\left(\prod_{i=m}^{k-1}\boldsymbol{H}_{R_iR_{i+1}}\boldsymbol{W}_{R_i}\right)^{\mathrm{H}}+\sigma_r^2\boldsymbol{I}_{N_R}\right)\boldsymbol{W}_{R_k}^{\mathrm{H}}\Bigg), & 2\le k\le K, \end{cases} \tag{37}$$

CSI errors. The robust EE optimization problem under the uncertainty model (32) is formulated as

$$\begin{aligned} &\max_{\boldsymbol{V}_S,\widetilde{\boldsymbol{W}}_R}\ \min_{\boldsymbol{\Delta}_{R_KD}\varnothing\boldsymbol{E}_{R_KD}} \mathrm{EE}_M\left(\boldsymbol{V}_S,\widetilde{\boldsymbol{W}}_R,\boldsymbol{\Delta}_{R_KD}\varnothing\boldsymbol{E}_{R_KD}\right)=\frac{R_{\mathrm{mul}}}{P_{\mathrm{mul}}},\\ &\ \text{s.t.}\quad \mathrm{Tr}\left(\mathrm{V_S}\right)\le \mathrm{P_{S_{max}}},\quad \mathrm{P_{R_k}}\le \mathrm{P_{R_{max}}},\ 1\le \mathrm{k}\le \mathrm{K},\\ &\qquad\quad \|\boldsymbol{\Delta}_{R_KD}\|\le\epsilon_a \text{ or } \|\boldsymbol{E}_{R_KD}\|\le\epsilon_m, \end{aligned} \tag{33}$$

where $\widetilde{\boldsymbol{W}}_R=\left\{\boldsymbol{W}_{R_1},\boldsymbol{W}_{R_2},\cdots,\boldsymbol{W}_{R_K}\right\}$, the maximum achievable rate R_{mul} is given by

$$\begin{aligned} & R_{\mathrm{mul}} \\ & =\frac{B}{K+1}\log\det\left(\boldsymbol{I}_{N_D}+\boldsymbol{H}_{\mathrm{mul},1}\boldsymbol{H}_{SR_1}\boldsymbol{V}_S\boldsymbol{H}_{SR_1}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{mul},1}^{\mathrm{H}}\boldsymbol{N}_{\mathrm{mul}}^{-1}\right), \end{aligned} \tag{34}$$

with

$$\boldsymbol{N}_{\mathrm{mul}}=\sigma_r^2\sum_{m=1}^{K-1}\boldsymbol{H}_{\mathrm{mul,m}}\boldsymbol{H}_{\mathrm{mul,m}}^{\mathrm{H}}+\sigma_d^2\boldsymbol{I}_{N_D}, \tag{35}$$

$$\boldsymbol{H}_{\mathrm{mul,m}}=\boldsymbol{H}_{R_KD}\boldsymbol{W}_{R_K}\left(\prod_{i=m}^{K-1}\boldsymbol{H}_{R_iR_{i+1}}\boldsymbol{W}_{R_i}\right), \tag{36}$$

and the transmit signal power P_{R_k} of relay k, $1\le k\le K$, is given by (37) at the top of this page, while the total power consumption P_{mul} is expressed as

$$P_{\mathrm{mul}}=\frac{P_S}{\tau_s}+\sum_{k=1}^{K}\frac{P_{R_k}}{\tau_r}+P_C^{'}, \tag{38}$$

in which $P_C^{'}$ is the total circuit power consumption. Similarly to (5), $P_C^{'}=N_SP_{dy,s}+N_R\sum_{k=1}^{K}P_{dy,r,k}+P_{st}^{'}$ with $P_{dy,r,k}$ denoting the dynamic power consumption of the kth relay's RF chain, $1\le k\le K$, and the total static power consumption of the source and all relays is $P_{st}^{'}=P_{st,s}+\sum_{k=1}^{K}P_{st,r,k}$.

A. Proposed Robust EE Design

Clearly, the max-min EE problem (33) is more challenging than the problem (8) but the former has the similar structure to the latter and, therefore, it can be solved with the similar approach as detailed in Section III. Specifically, let us define the following SVDs

$$\boldsymbol{H}_{SR_1}=\boldsymbol{U}_{SR_1}\boldsymbol{\Sigma}_{SR_1}\boldsymbol{Q}_{SR_1}^{\mathrm{H}}, \tag{39}$$

$$\boldsymbol{H}_{R_kR_{k+1}}=\boldsymbol{U}_{R_kR_{k+1}}\boldsymbol{\Sigma}_{R_kR_{k+1}}\boldsymbol{Q}_{R_kR_{k+1}}^{\mathrm{H}}, \tag{40}$$

$$\widehat{\boldsymbol{H}}_{R_KD}=\widehat{\boldsymbol{U}}_{R_KD}\widehat{\boldsymbol{\Sigma}}_{R_KD}\widehat{\boldsymbol{Q}}_{R_KD}^{\mathrm{H}},\ 1\le k\le K-1. \tag{41}$$

where $\boldsymbol{U}_{SR_1}\in\mathbb{C}^{N_R\times N_R}$ and $\boldsymbol{Q}_{SR_1}\in\mathbb{C}^{N_S\times N_S}$, $\boldsymbol{U}_{R_kR_{k+1}}\in\mathbb{C}^{N_R\times N_R}$ and $\boldsymbol{Q}_{R_kR_{k+1}}\in\mathbb{C}^{N_R\times N_R}$, as well as $\widehat{\boldsymbol{U}}_{R_KD}\in\mathbb{C}^{N_D\times N_D}$ and $\widehat{\boldsymbol{Q}}_{R_KD}\in\mathbb{C}^{N_R\times N_R}$ are the unitary matrices for $\boldsymbol{H}_{SR_1}$, $\boldsymbol{H}_{R_kR_{k+1}}$ and $\widehat{\boldsymbol{H}}_{R_KD}$, respectively, while the diagonal matrices $\boldsymbol{\Sigma}_{SR_1}\in\mathbb{C}^{N_R\times N_S}$, $\boldsymbol{\Sigma}_{R_kR_{k+1}}\in\mathbb{C}^{N_R\times N_R}$, and $\widehat{\boldsymbol{\Sigma}}_{R_KD}\in\mathbb{C}^{N_D\times N_R}$ contain $N_P=\min\{N_R,\,N_S\}$ SVs of $\boldsymbol{H}_{SR_1}$, N_R SVs of $\boldsymbol{H}_{R_kR_{k+1}}$, and $N_C=\min\{N_D,\,N_R\}$ SVs $\left\{\widehat{\sigma}_{r_Kd,1},\cdots,\widehat{\sigma}_{r_Kd,N_C}\right\}$ of $\widehat{\boldsymbol{H}}_{R_KD}$ at their diagonal positions, respectively.

Theorem 4. *For the max-min EE problem* (33)*, the optimal source covariance matrix $\boldsymbol{V}_S^{\star}$, the optimal relay beamforming matrices $\boldsymbol{W}_{R_k}^{\star}, 1\le k\le K$, and the worst-case errors $\boldsymbol{\Delta}_{R_KD}^{\star}\varnothing\boldsymbol{E}_{R_KD}^{\star}$ have the following structures*

$$\boldsymbol{V}_S^{\star}=\boldsymbol{Q}_{SR_1}\boldsymbol{\Sigma}_S\boldsymbol{Q}_{SR_1}^{\mathrm{H}}, \tag{42}$$

$$\boldsymbol{W}_{R_k}^{\star}=\begin{cases} \boldsymbol{Q}_{R_1R_2}\boldsymbol{\Sigma}_{\boldsymbol{W}_{R_1}}\boldsymbol{U}_{SR_1}^{\mathrm{H}}, & k=1,\\ \boldsymbol{Q}_{R_kR_{k+1}}\boldsymbol{\Sigma}_{\boldsymbol{W}_{R_k}}\boldsymbol{U}_{R_{k-1}R_k}^{\mathrm{H}}, & 2\le k\le K-1,\\ \widehat{\boldsymbol{Q}}_{R_KD}\boldsymbol{\Sigma}_{\boldsymbol{W}_{R_K}}\boldsymbol{U}_{R_{K-1}R_K}^{\mathrm{H}}, & k=K, \end{cases} \tag{43}$$

$$\begin{aligned} &\boldsymbol{\Delta}_{R_KD}^{\star}\\ &=-\widehat{\boldsymbol{U}}_{R_KD}\begin{bmatrix}\boldsymbol{\Lambda}_{R_KD} & \boldsymbol{0}_{N_C\times(N_R-N_C)}\\ \boldsymbol{0}_{(N_D-N_C)\times N_C} & \boldsymbol{0}_{(N_D-N_C)\times(N_R-N_C)}\end{bmatrix}\widehat{\boldsymbol{Q}}_{R_KD}^{\mathrm{H}}\\ &\textit{or}\ \ \boldsymbol{E}_{R_KD}^{\star}=-\epsilon_m\boldsymbol{I}_{N_D}, \end{aligned} \tag{44}$$

where $\boldsymbol{\Lambda}_{R_KD}=\mathrm{diag}\left\{\min\left\{\widehat{\sigma}_{r_Kd,1},\epsilon_a\right\},\cdots,\min\left\{\widehat{\sigma}_{r_Kd,N_C},\epsilon_a\right\}\right\}$,

$$\boldsymbol{\Sigma}_{\boldsymbol{W}_{R_1}}=\boldsymbol{\Sigma}_{X_1}\left(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\boldsymbol{\Sigma}_{SR_1}\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_{SR_1}^{\mathrm{H}}\right)^{-\frac{1}{2}}, \tag{45}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{W}_{R_k}}=&\boldsymbol{\Sigma}_{X_k}\Bigg(\boldsymbol{I}_{N_R}+\sum_{m=1}^{k-1}\overline{\overline{\boldsymbol{\Sigma}}}_k^m+\sigma_r^{-2}\Big(\prod_{i=1}^{k-1}\widetilde{\boldsymbol{\Sigma}}_{i,i+1}\Big)\boldsymbol{\Sigma}_{SR_1}\boldsymbol{\Sigma}_S\\ &\cdot\boldsymbol{\Sigma}_{SR_1}^{\mathrm{H}}\Big(\prod_{i=1}^{k-1}\widetilde{\boldsymbol{\Sigma}}_{i,i+1}\Big)^{\mathrm{H}}\Bigg)^{-\frac{1}{2}},2\le k\le K, \end{aligned} \tag{46}$$

with $\boldsymbol{\Sigma}_{X_k}=\mathrm{diag}\big\{\sigma_{x_k,1},\cdots,\sigma_{x_k,N_C},\underbrace{0,0,\cdots,0}_{N_R-N_C}\big\}$ *for* $1\le i\le K$, $\overline{\overline{\boldsymbol{\Sigma}}}_k^m=\big(\prod_{i=m}^{k-1}\widetilde{\boldsymbol{\Sigma}}_{i,i+1}\big)\big(\prod_{i=m}^{k-1}\widetilde{\boldsymbol{\Sigma}}_{i,i+1}\big)^{\mathrm{H}}$ *and* $\widetilde{\boldsymbol{\Sigma}}_{i,i+1}=\boldsymbol{\Sigma}_{R_iR_{i+1}}\boldsymbol{\Sigma}_{\boldsymbol{W}_{R_i}}$.

Proof. See Appendix D. □

$$\begin{aligned}\max_{\boldsymbol{\Sigma}_S,\forall\boldsymbol{\Sigma}_{X_k}} \quad & \log\det\Bigg(\boldsymbol{I}_{N_D}+\bigg(\widetilde{\boldsymbol{\Sigma}}_{R_KD}\bigg(\prod_{i=1}^{K-1}\widetilde{\boldsymbol{\Sigma}}_{i,i+1}\bigg)\boldsymbol{\Sigma}_{SR_1}\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_{SR_1}^{\mathrm{H}}\bigg(\prod_{i=1}^{K-1}\widetilde{\boldsymbol{\Sigma}}_{i,i+1}\bigg)^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{R_KD}^{\mathrm{H}}\bigg) \\ & \times\bigg(\sigma_r^2\sum_{m=1}^{K-1}\widetilde{\boldsymbol{\Sigma}}_{R_KD}\overline{\widetilde{\boldsymbol{\Sigma}}}_K^m\widetilde{\boldsymbol{\Sigma}}_{R_KD}^{\mathrm{H}}+\sigma_d^2\boldsymbol{I}_{N_D}\bigg)^{-1}\Bigg)-\eta_{\mathrm{mul}}\bigg(\mathrm{Tr}\big(\boldsymbol{\Sigma}_S\big)+\sum_{k=1}^{K}\mathrm{Tr}\big(\boldsymbol{\Sigma}_{X_k}\big)+P_C\bigg), \\ \text{s.t.} \quad & \mathrm{Tr}\big(\boldsymbol{\Sigma}_S\big)\le P_{S_{\max}},\ \mathrm{Tr}\big(\boldsymbol{\Sigma}_{X_k}\big)\le P_{R_{k_{\max}}},\ 1\le k\le K,\end{aligned} \tag{47}$$

Similarly to the two-hop case, based on Theorem 4 and Lemma 1, we can equivalently transform the robust EE problem (33) into the follow optimization problem with scalar variables where $\widetilde{\boldsymbol{\Sigma}}_{R_KD}=\Big(\widehat{\boldsymbol{\Sigma}}_{R_KD}-\boldsymbol{\Lambda}_{R_KD}\Big)\boldsymbol{\Sigma}_{\boldsymbol{W}_{R_K}}$ and $\widetilde{\boldsymbol{\Sigma}}_{R_KD}=(1-\epsilon_m)\widehat{\boldsymbol{\Sigma}}_{R_KD}\boldsymbol{\Sigma}_{\boldsymbol{W}_{R_K}}$ are defined for the additive and multiplicative CSI errors, respectively, and η_{mul} is the auxiliary variable. For convenience, denote $\boldsymbol{\Sigma}_S=\boldsymbol{\Sigma}_{X_0}$. Observe that the problem (47) is convex w.r.t $\boldsymbol{\Sigma}_{X_k}$ when the remaining variables $\{\boldsymbol{\Sigma}_{X_0},\boldsymbol{\Sigma}_{X_1},\cdots,\boldsymbol{\Sigma}_{X_K}\}\backslash\boldsymbol{\Sigma}_{X_k}$ are fixed. Therefore, we can apply the alternating optimization of Section III-B to efficiently solve the problem (47) by decomposing it into $(K+1)$ alternating subproblems, in order to obtain a locally optimal solution.

B. Extension to Imperfect $\boldsymbol{H}_{SR_1}$ and $\boldsymbol{H}_{R_kR_{k+1}},\forall k$

We now consider the most generic case, where the source-relay channel $\boldsymbol{H}_{SR_1}$ and all the relay-relay channels $\boldsymbol{H}_{R_kR_{k+1}}$, $\forall k$, are also imperfect at relay R_K. Similarly to the two-hop case, by adopting the statistically imperfect $\boldsymbol{H}_{SR_1}$ and $\boldsymbol{H}_{R_kR_{k+1}}$, $\forall k$, and the deterministically imperfect $\boldsymbol{H}_{R_KD}$, the robust average EE optimization for this generic multihop AF relaying network can be formed. Following the same philosophy in proving Theorem 3, a similar conclusion can also be obtained for the multihop scenario by proving the optimal channel-diagonalizing structure one by one for the optimal source covariance matrix $\boldsymbol{V}_S^\star$ and the optimal relay beamforming matrices $\boldsymbol{W}_{R_k}^*$, $1\le k\le K$. Specifically, the eigenspaces of $\boldsymbol{V}_S^\star$ and $\boldsymbol{W}_{R_k}^*$, $1\le k\le K-1$, are aligned with that of the source spatial correlation matrix and that of the relay R_k spatial correlation matrix, respectively, while $\boldsymbol{W}_{R_K}^*$ is jointly determined by the eigenspace of the relay R_K spatial correlation matrix and the right singular matrix of the relay-destination channel $\boldsymbol{H}_{R_KD}$. The corresponding robust average EE problem can also be transformed into a robust optimization with scalar variables. However, due to the simultaneous expectations for multiple statistically imperfect channels, this scalar-variable problem is generally intractable and cannot be decomposed into a series of convex subproblems [13]. A possible solution is to apply successive Jensen's inequalities to the expectations on $\{\boldsymbol{H}_{SR_1},\boldsymbol{H}_{R_kR_{k+1}},\forall k\}$ to find an upper-bound of the average EE [13], which can then be solved by the proposed alternating optimization.

Remark 2: The last relay R_K needs to feed back the optimal source covariance matrix $\boldsymbol{V}_S^\star$ and optimal relay beamforming matrices $\boldsymbol{W}_{R_k}^\star$, $1\le k\le K-1$, perfectly to the source and corresponding relays. If the feedback errors for $\{\boldsymbol{V}_S^\star,\boldsymbol{W}_{R_k}^\star,1\le k\le K-1\}$ are serious, they should be additionally imposed on the robust EE design, and the optimal beamforming matrix $\boldsymbol{W}_{R_K}^\star$ of relay K should also be redesigned accordingly. Unfortunately, even if the spectral norm constrained errors for $\{\boldsymbol{V}_S^\star,\boldsymbol{W}_{R_k}^\star,1\le k\le K-1\}$ are jointly considered, the channel-diagonalizing structured optimal $\boldsymbol{W}_{R_K}^\star$ is not guaranteed, since the multiple relay beamforming errors are coupled in both the objective and the transmit power constraints. Future research is warranted to develop the low-complexity suboptimal algorithms to effectively address this issue.

V. Simulation Study

In the simulation, the source is a base station (BS), while the destination and relays are mobile stations (MSs). The default parameters of the simulated MIMO AF relaying network are listed in Table II. Unless otherwise stated, these default values are used. To demonstrate the excellent performance of our robust EE design, we adopt the non-robust EE maximization (NREE) and the naive AF based EE maximization (NAF) [17] for comparison. For the NREE scheme, the optimization problem (8)/(32) is firstly solved by assuming no CSI errors, i.e., $\epsilon_a=\epsilon_m=0$. Then the resultant optimal solution is applied to the imperfect CSI scenario for calculating the worst-case EE. For the NAF scheme [17], the relay scales the received signal transmitted by source with the maximum power by a constant to realize the maximum relay power transfer. All simulation results are obtained by averaging over 100 channel realizations.

A. Two-hop MIMO AF Relaying Networks

The convergence of the proposed alternative optimization algorithm is investigated under the two sets of the initial values $\{\boldsymbol{\lambda}_s^{(0)},\eta^{(0)}\}$, given by $\{\boldsymbol{\lambda}_{s,\mathrm{ini1}}^{(0)},\eta_1^{(0)}\}=\{\frac{P_{S_{\max}}}{N_S}\mathbf{1}_{N_S},0\}$ and $\{\boldsymbol{\lambda}_{s,\mathrm{ini2}}^{(0)},\eta_2^{(0)}\}=\{\frac{P_{S_{\max}}}{2N_S}\big[\mathbf{1}_{N_S/2}^{\mathrm{T}}\ \mathbf{0}_{N_S/2}^{\mathrm{T}}\big]^{\mathrm{T}},0.1\}$. Algorithm 1 consists of an outer alternating optimization loop, and within each alternating iteration, there are two inner Dinkelbach iterative loops at step 2 and step 3. To demonstrate the convergence of the two inner Dinkelbach iterative loops, Fig. 2 (a) plots the objective value of the problem (20) after each Dinkelbach iteration for optimizing $\boldsymbol{\sigma}_x$ given $\boldsymbol{\lambda}_s$ at the first outer iteration. Observe from Fig. 2 (a) that this very first Dinkelbach iterative procedure of Algorithm 1 takes no more than 6 iterations to converge. Since any subsequent Dinkelbach iterative procedure is unlikely to take more iterations to converge, we conclude that for this example, any Dinkelbach iterative procedure of Algorithm 1 takes no more than 6 iterations to converge. The convergence of Algorithm 1 is illustrated in Fig. 2 (b), where it is seen that for this example, Algorithm 1 takes no more than 7 outer iterations to converge. Fig. 2 (c) depicts the curve of Fig. 2 (a) corresponding to the additive CSI errors with the initial condition $\{\boldsymbol{\lambda}_{s,\mathrm{ini1}}^{(0)},\eta_1^{(0)}\}$ but with the logarithmic scale in y-axis and with the error bars. It can be seen that after 6 iterations, the objective value becomes smaller than 10^{-3}.

[2] In our simulations, the large-scale fading coefficient σ_h^2 is determined according to the LTE path-loss model $-\sigma_h^2+30=L_P(\mathrm{dB})=128.1+37.6\log(d\ [\mathrm{Km}]\ )$ with $d=330\mathrm{m}$ [33].

TABLE II
DEFAULT SYSTEM PARAMETERS

Parameter	Value	Parameter	Value
The number of BS/Relay/MS antennas $N_S/N_R/N_D$	4/6/4 (4/4/4; 6/4/4; 4/4/6)	Relay dynamic power consumption $P_{dy,r}$	35dBm [32]
Transmission bandwidth B	50MHz	Relay static power consumption $P_{st,r}$	30dBm [32]
Power amplifier efficiency τ_s, τ_r	0.5	Additive errors threshold (two-hop) ϵ_a	$\epsilon_a = p\|\widehat{\boldsymbol{H}}_{RD}\|_2$, $p=0.4$
BS maximum transmit power $P_{S_{\max}}$	46dBm [33]	Additive errors threshold (multihop) ϵ_a	$\epsilon_a = p\|\widehat{\boldsymbol{H}}_{R_K D}\|_2$, $p=0.4$
Relay maximum transmit power $P_{R_{\max}}$	40dBm [33]	Multiplicative errors threshold ϵ_m	$\epsilon_m = p = 0.4$
BS dynamic power consumption $P_{dy,s}$	40dBm [32]	Rayleigh MIMO relaying channels with large-scale fading coefficient σ_h^2. [2]	$\mathcal{CN}(0, \sigma_h^2 \boldsymbol{I})$, $\sigma_h^2 = -80$dBm [33]
BS static power consumption $P_{st,s}$	35dBm [32]	Noise power $\sigma_r^2 = \sigma_d^2 = N_0 B$	-90dBm ($N_0 = -167$dBm/Hz) [32]

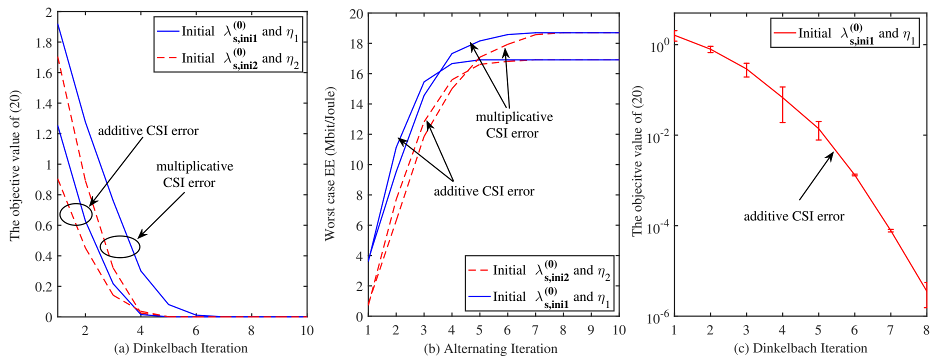


Fig. 2. Convergence of Algorithm 1: (a) inner Dinkelbach loop of step 2 at the first outer iteration, (b) outer alternating optimization loop, given two sets of different initial points and for both additive and multiplicative CSI errors, and (c) inner Dinkelbach loop of step 2 at the first outer iteration with error bars, for one set of initial points and additive CSI errors. $N_S/N_R/N_D = 4/6/4$.

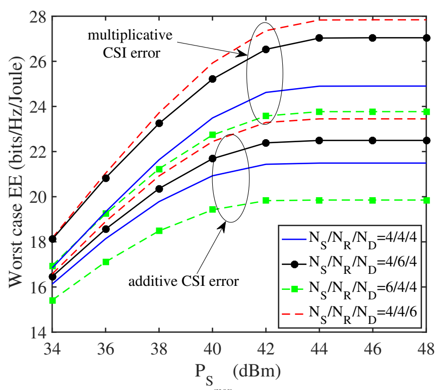


Fig. 3. Comparison of the worst-case EE performance as the functions of the source maximum transmit power $P_{S_{\max}}$ given different antenna configurations $N_S/N_R/N_D$.

The influence of antenna configuration to the worst-case EE performance is investigated in Fig. 3, which plots the worst-case EE performance of the proposed robust EE design as the functions of the source maximum transmit power $P_{S_{\max}}$ under different antenna configurations $N_S/N_R/N_D$. We observe that the antenna configuration $4/4/6$ attains the highest worst-case EE, and the antenna configuration $4/6/4$ attains the second highest worst-case EE, which is considerably larger than the $4/4/4$ configuration. While the $6/4/4$ configuration achieves the lowest worst-case EE. We now explain these phenomena based on the following facts. First, the circuit power consumption at receiver is neglected in our work, since reception generally consumes much less circuit power than transmission [23]. Second, the existing literature [7], [34] have shown that increasing the number of source antennas N_S without optimizing the source covariance matrix generally causes the capacity shrinking phenomenon for MIMO AF relaying systems. Even considering the design of source covariance matrix, the capacity gain is close to zero as N_S increases [7], [34]. By contrast, adding more relay and/or destination antennas is helpful to improve system capacity, of which the enhancement is more evident when increasing the number of relay antennas [34].

These known conclusions explain why the best worst-case EE performance for both multiplicative and additive CSI errors is observed at $N_S/N_R/N_D = 4/4/6$, which is because the achievable data rate increases with the number of destination antennas N_D, while the transmit power consumption remains unchanged. These conclusions also agree with our observation that the achievable worst-case EE under $N_S/N_R/N_D = 4/6/4$ is significantly higher than that under $N_S/N_R/N_D = 4/4/4$, because when the number of relay antennas N_R increases, the remarkable increase of data rate outweighs the increase in dynamic relay circuit power consumption. However, for the case of $N_S/N_R/N_D = 6/4/4$, the achievable worst-case EE is much reduced, since the increased source dynamic circuit power consumption outweighs the slight data rate gain due to the increasing number of source antennas N_S. Our results therefore support the existing literature and provide insights to design MIMO AF relaying systems, namely, increasing the number of relay and/or destination antennas rather than the number of source antennas is beneficial to improve system's worst-case EE performance.

Fig. 4 (a) compares the worst-case EE performance as the functions of $P_{S_{\max}}$ for the proposed robust EE design, NREE and NAF schemes, while Fig. 4 (b) shows the worst-case EE performance as the functions of $P_{R_{\max}}$ for the three designs. As expected, the proposed robust EE design achieves the highest worst-case EE, while the NAF scheme has the worst-

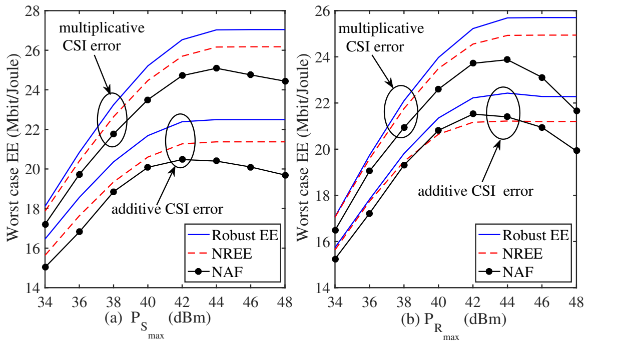


Fig. 4. Comparison of the worst-case EE performance as the functions of: (a) the source maximum transmit power $P_{S_{\max}}$, and (b) the relay maximum transmit power $P_{R_{\max}}$, for three designs. $N_S/N_R/N_D = 4/6/4$.

case EE. For our robust EE design and NREE scheme, the worst-case EE first increases with $P_{S_{\max}}$ ($P_{R_{\max}}$) but becomes saturated for large $P_{S_{\max}}$ ($P_{R_{\max}}$). This is because when $P_{S_{\max}}$ ($P_{R_{\max}}$) is small compared to the circuit power consumption P_C, the EE metric is mainly determined by the maximum achievable rate in the numerator of the EE metric. Increasing $P_{S_{\max}}$ ($P_{R_{\max}}$) increases the source (relay) transmit power too, which in turn increases the maximum achievable rate. Therefore, for relatively small $P_{S_{\max}}$ ($P_{R_{\max}}$), increasing $P_{S_{\max}}$ ($P_{R_{\max}}$) increases the worst-case EE. In this region, the power constraint is active, i.e., the source (relay) transmit power reaches $P_{S_{\max}}$ ($P_{R_{\max}}$). By contrast, when $P_{S_{\max}}$ ($P_{R_{\max}}$) becomes large, the EE metric is also determined by the source (relay) transmit power in the denominator of the EE metric. Therefore, to maximize the worst-case EE, a trade-off between the maximum achievable rate and the total transmit power must be made. As a result, the maximum worst-case EE metric reaches a saturated value. In this region, the power constraint is inactive and the source (relay) transmit power remains constant. For the NAF, the worst-case EE decreases with $P_{S_{\max}}$ ($P_{R_{\max}}$) for large $P_{S_{\max}}$ ($P_{R_{\max}}$). This is because the maximum achievable rate of the NAF strategy cannot be arbitrarily enhanced by increasing $P_{S_{\max}}$ ($P_{R_{\max}}$), since the relay simultaneously amplifies the received signal and noise. Therefore, for large $P_{S_{\max}}$ ($P_{R_{\max}}$), the NAF's worst-case EE decreases with $P_{S_{\max}}$ ($P_{R_{\max}}$) due to the limited maximum achievable rate and the increase in the total power consumption.

The two thresholds are fairly set for the multiplicative and additive CSI errors in Table II, as they correspond to the same quantitative measure for the spectral norm constrained multiplicative and additive CSI errors. According to Theorem 1, we can infer that the SVs of the worst-case relay-destination channel for multiplicative CSI errors are larger or equal to those for additive CSI errors. Therefore, we can conclude that under the same size of the spectral norm constrained CSI errors, the quality of the relay-destination channel under multiplicative CSI errors is better than that under additive errors. Naturally, the achievable worst-case EE under multiplicative CSI errors is also higher than that under additive CSI errors. This is also confirmed by Fig. 4.

Fig. 5 investigates the impact of the CSI uncertainty threshold p on the achievable worst-case EE. Not surprisingly, as p increases, the worst-case EE performance decreases for every scheme. Again our robust design achieves the best worst-case

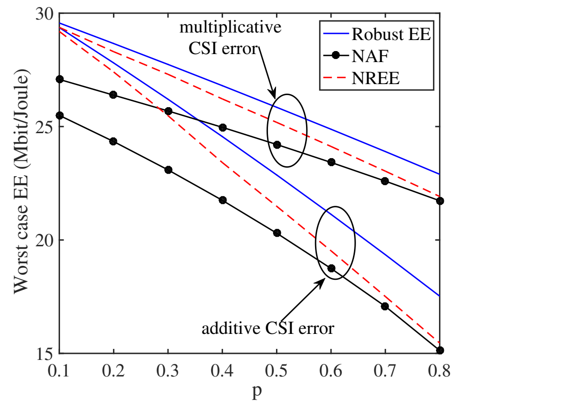


Fig. 5. Influence of CSI uncertainty p on the achievable worst-case EE for three desig

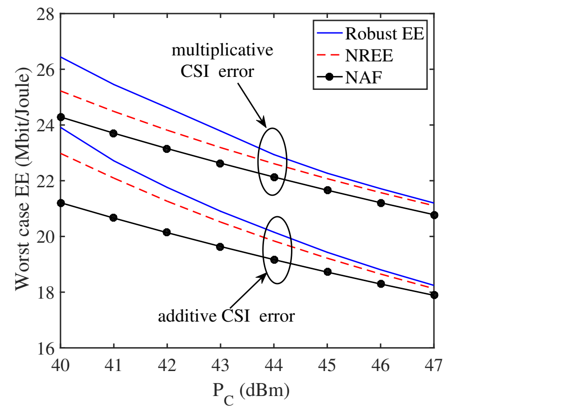


Fig. 6. Influence of circuit power consumption P_C on the achievable worst-case EE for three designs. $N_S/N_R/N_D = 4/6/4$.

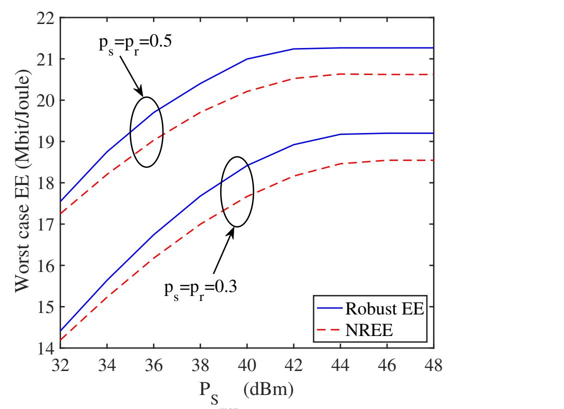


Fig. 7. Influence of the source maximum transmit power $P_{S_{\max}}$ on the achievable worst-case EE for two designs. The statistically imperfect source-relay channel and additive CSI errors for the relay-destination channel are considered. $N_S/N_R/N_D = 4/6/4$.

EE, while the NAF has the worst performance. Moreover, the performance gap between our robust EE design and the NREE increases with p, which further indicates the effectiveness of our design. Fig. 6 shows the worst-case EE performance as the functions of the circuit power consumption P_C for three schemes. As expected, our robust design attains the highest worst-case EE. Observe that the worst-case EE decreases as P_c increases, since P_C only appears in the denominator of the EE metric. Next we consider the statistically imperfect source-relay channel $\boldsymbol{H}_{SR}$. The source and relay spatial correlation matrices $\boldsymbol{R}_R$ and $\boldsymbol{R}_S$ are simulated using the exponential model [35]. Specifically, for $i, j = 1, \cdots, N_R\,(N_S)$ and $j \geq i$, the (i, j)-th elements of $\boldsymbol{R}_R$ and $\boldsymbol{R}_S$ are given respectively by $\left[\boldsymbol{R}_R\right]_{i,j} = p_r^{j-i}$ and $\left[\boldsymbol{R}_S\right]_{i,j} = p_s^{j-i}$, where $|p_r| < 1$ and

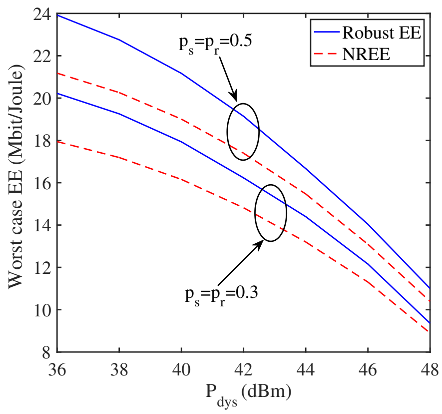


Fig. 8. Influence of the the BS dynamic power consumption $P_{dy,s}$ on the achievable worst-case EE for two designs. The statistically imperfect source-relay channel and additive CSI errors for the relay-destination channel are considered. $N_S/N_R/N_D = 4/6/4$.

$|p_s| < 1$ [35]. Without loss of generality, only additive CSI error is considered for the relay-destination channel. Fig. 7 depicts the worst-case EE performance versus $P_{S_{\max}}$ for the upper-bound robust average EE design of Section III-E and the NREE scheme. Observe from Fig. 7 that the relationship between the worst-case EE and $P_{S_{\max}}$ for the two designs is similar to that shown in Fig. 4 (a). Compared with the case of $p_s = p_r = 0.3$, the stronger correlation of $p_s = p_r = 0.5$ leads to higher achievable worst-case EE, because higher correlation means higher source-relay channel energy, which is beneficial to improve the ergodic rate. For the two designs considered, Fig. 8 shows that the achievable worst-case EE decreases with the BS dynamic power consumption $P_{dy,s}$, which is a portion of the total circuit power consumption P_C. The reason is similar to that given for Fig. 6.

B. Three-hop ($K = 2$) MIMO AF Relaying Networks

A three-hop MIMO AF relaying network is simulated, and we only consider additive CSI errors for the relay-destination channel. Fig. 9 (a) compares the worst-case EE performance versus the relays' maximum transmit power $P_{R_{\max}}$ for the three designs. Compared with Fig. 4 (b), similar trends between the worst-case EE and $P_{R_{\max}}$ for the three schemes can also be observed from Fig. 9 (a). Clearly, the achievable worst-case EE in the three-hop case is lower than that in the two-hop case due to the greater channel fading and higher power consumption. Fig. 9 (b) depicts the influence of p on the worst-case EE for the three designs. Both Fig. 9 (a) and Fig. 9 (b) confirm that the proposed robust EE design attains the best worst-case EE performance.

VI. Conclusions

We have optimized the EE of two-hop MIMO AF relaying networks under the deterministically imperfect CSI. By considering the additive and multiplicative CSI errors for the relay-destination channel, the source covariance and relay beamforming matrices are jointly optimized to maximize the worst-case EE. We have proved the existence of a saddle point for this robust EE problem, and have derived the channel-diagonalizing structure of the optimal source covariance and relay beamforming matrices as well as the worst-case errors

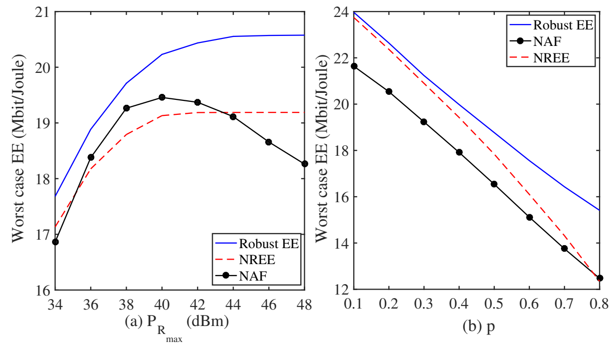


Fig. 9. Comparison of the worst-case EE performance as the functions of: (a) the relays' maximum transmit power $P_{R_{\max}}$, and (b) the CSI uncertainty p, for three designs. The three-hop MIMO AF relaying network under additive CSI uncertainty is considered. $N_S = 4$, $N_{R_1} = N_{R_2} = 6$ and $N_D = 4$.

under the spectral-norm constrained additive and multiplicative CSI errors. Based on this structure, the original robust EE problem is transformed into an optimization problem with scalar variables, which can be efficiently solved by the proposed alternating optimization. We have also proved that all these results are applicable when the statistically imperfect source-relay channel is additionally imposed. Furthermore, we have extended our work to multihop MIMO AF relaying networks, and have proved that the channel-diagonalizing structure remains optimal for the source covariance matrix and all the relays' beamforming matrices under deterministically imperfect relay-destination CSI.

Appendix

A. Proof of **Theorem 1**

Proof. Unless otherwise stated, the eigenvalues (EVs)/SVs of an EVD/SVD for a matrix are always arranged in a decreasing order. First we have the following lemma [36].c

Lemma 2. *For the two $N \times N$ Hermitian matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ whose EVs are denoted by $\lambda_i(\boldsymbol{A})$ and $\lambda_i(\boldsymbol{B})$, respectively, for $i = 1, \cdots, N$, we have*

$$\prod_{i=1}^{N} \big(\lambda_i(\boldsymbol{A}) + \lambda_i(\boldsymbol{B})\big) \leq \det(\boldsymbol{A} + \boldsymbol{B}) \leq \prod_{i=1}^{N} \big(\lambda_i(\boldsymbol{A}) + \lambda_{N+1-i}(\boldsymbol{B})\big), \tag{48}$$

$$\sum_{i=1}^{N} \lambda_i(\boldsymbol{A})\lambda_{N+1-i}(\boldsymbol{B}) \leq \mathrm{Tr}(\boldsymbol{AB}) \leq \sum_{i=1}^{N} \lambda_i(\boldsymbol{A})\lambda_i(\boldsymbol{B}). \tag{49}$$

All the equalities in (48) *and* (49) *hold only when $\boldsymbol{A}$ and $\boldsymbol{B}$ are simultaneously diagonalizable.*

1) Optimal $\boldsymbol{W}_R^{\star}$: Re-express the rate formulation (9) as

$$2^{\frac{2R_D}{B}} = \frac{\det\big(\boldsymbol{I}_{N_R} + \sigma_r^2 \sigma_d^{-2} \big(\boldsymbol{I}_{N_R} + \sigma_r^{-2} \boldsymbol{H}_{SR} \boldsymbol{V}_S \boldsymbol{H}_{SR}^{\mathrm{H}}\big) \boldsymbol{W}_R^{\mathrm{H}} \boldsymbol{H}_{RD}^{\mathrm{H}} \boldsymbol{H}_{RD} \boldsymbol{W}_R\big)}{\det\big(\boldsymbol{I}_{N_R} + \sigma_r^2 \sigma_d^{-2} \boldsymbol{W}_R^{\mathrm{H}} \boldsymbol{H}_{RD}^{\mathrm{H}} \boldsymbol{H}_{RD} \boldsymbol{W}_R\big)}. \tag{50}$$

According to the identity $\det(\boldsymbol{I}_N + \boldsymbol{AB}) = \det(\boldsymbol{I}_K + \boldsymbol{BA})$, where $\boldsymbol{A} \in \mathbb{C}^{N\times K}$ and $\boldsymbol{B} \in \mathbb{C}^{K\times N}$, the achievable EE metric $\mathrm{EE}\big(\boldsymbol{V}_S, \boldsymbol{W}_R, \boldsymbol{\Delta}_{RD} \varnothing \boldsymbol{E}_{RD}\big)$ can be reformulated as Next perform the EVDs of the nonnegative definite matrices $\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}}$ and $\boldsymbol{H}_{RD}^{\mathrm{H}}\boldsymbol{H}_{RD}$:

$$\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}} = \widetilde{\boldsymbol{U}}_{SR}\widetilde{\boldsymbol{\Sigma}}_{SR}\widetilde{\boldsymbol{U}}_{SR}^{\mathrm{H}}, \tag{52}$$

$$\boldsymbol{H}_{RD}^{\mathrm{H}}\boldsymbol{H}_{RD} = \widetilde{\boldsymbol{Q}}_{RD}\widetilde{\boldsymbol{\Sigma}}_{RD}\widetilde{\boldsymbol{Q}}_{RD}^{\mathrm{H}}, \tag{53}$$

$$\mathrm{EE}\big(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big)=\frac{\frac{B}{2}\log\det\big(\boldsymbol{I}_{N_D}+\sigma_r^2\sigma_d^{-2}\boldsymbol{H}_{RD}\boldsymbol{W}_R\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}}\big)\boldsymbol{W}_R^{\mathrm{H}}\boldsymbol{H}_{RD}^{\mathrm{H}}\big)}{\mathrm{Tr}\big(\boldsymbol{V}_S\big)+\mathrm{Tr}\big(\boldsymbol{W}_R(\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}}+\sigma_r^2\boldsymbol{I}_{N_R})\boldsymbol{W}_R^{\mathrm{H}}\big)+P_C} - \frac{\frac{B}{2}\log\det\big(\boldsymbol{I}_{N_R}+\sigma_r^2\sigma_d^{-2}\boldsymbol{W}_R^{\mathrm{H}}\boldsymbol{H}_{RD}^{\mathrm{H}}\boldsymbol{H}_{RD}\boldsymbol{W}_R\big)}{\mathrm{Tr}\big(\boldsymbol{V}_S\big)+\mathrm{Tr}\big(\boldsymbol{W}_R(\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}}+\sigma_r^2\boldsymbol{I}_{N_R})\boldsymbol{W}_R^{\mathrm{H}}\big)+P_C}. \tag{51}$$

$$\frac{2}{B}\cdot\mathrm{EE}\big(\boldsymbol{V}_S,\boldsymbol{X},\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big)=\frac{\log\det\big(\boldsymbol{I}_{N_R}+\sigma_r^2\sigma_d^{-2}\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}\big)}{\mathrm{Tr}\big(\boldsymbol{V}_S\big)+\mathrm{Tr}\big(\sigma_r^2\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\big)+P_C} + \frac{\log\det\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR}\big)}{\mathrm{Tr}\big(\boldsymbol{V}_S\big)+\mathrm{Tr}\big(\sigma_r^2\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\big)+P_C}-\frac{\log\det\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR}+\sigma_r^2\sigma_d^{-2}\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}\big)}{\mathrm{Tr}\big(\boldsymbol{V}_S\big)+\mathrm{Tr}\big(\sigma_r^2\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\big)+P_C}. \tag{56}$$

---

where $\widetilde{\boldsymbol{\Sigma}}_{SR}=\mathrm{diag}\big\{\widetilde{\sigma}_{sr,1}^2,\cdots,\widetilde{\sigma}_{sr,N_P}^2,0,\cdots,0\big\}$ and $\widetilde{\boldsymbol{\Sigma}}_{RD}=\mathrm{diag}\big\{\widetilde{\sigma}_{rd,1}^2,\cdots,\widetilde{\sigma}_{rd,N_C}^2,0,\cdots,0\big\}$ contain the N_P nonzero EVs of $\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}}$ and the N_C nonzero EVs of $\boldsymbol{H}_{RD}^{\mathrm{H}}\boldsymbol{H}_{RD}$, respectively, while $\widetilde{\boldsymbol{U}}_{SR}\in\mathbb{C}^{N_R\times N_R}$ and $\widetilde{\boldsymbol{Q}}_{RD}\in\mathbb{C}^{N_R\times N_R}$ are the associated unitary matrices. Note that $\widetilde{\boldsymbol{Q}}_{RD}$ and $\widetilde{\boldsymbol{\Sigma}}_{RD}$ are unknown, while $\widetilde{\boldsymbol{\Sigma}}_{SR}$ depends on the matrix variable $\boldsymbol{V}_S$ whose optimal structure is yet to be determined. Clearly, $\widetilde{\boldsymbol{\Sigma}}_{SR}\succeq\boldsymbol{0}$ and $\widetilde{\boldsymbol{\Sigma}}_{RD}\succeq\boldsymbol{0}$. By defining $\boldsymbol{X}\in\mathbb{C}^{N_R\times N_R}$ as

$$\boldsymbol{X}=\widetilde{\boldsymbol{Q}}_{RD}^{\mathrm{H}}\boldsymbol{W}_R\widetilde{\boldsymbol{U}}_{SR}\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR}\big)^{\frac{1}{2}}, \tag{54}$$

we can express the relay beamforming matrix $\boldsymbol{W}_R$ as

$$\boldsymbol{W}_R=\widetilde{\boldsymbol{Q}}_{RD}\boldsymbol{X}\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR}\big)^{-\frac{1}{2}}\widetilde{\boldsymbol{U}}_{SR}^{\mathrm{H}}. \tag{55}$$

Then $\boldsymbol{X}$ is the new optimization matrix variable. Substituting (52), (53) and (55) into (51) (at the top of this page) yields (56) (also at the top of this page). Denote $\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}=\boldsymbol{U}_T\boldsymbol{\Sigma}_T\boldsymbol{U}_T^{\mathrm{H}}$, where the unitary matrix $\boldsymbol{U}_T\in\mathbb{C}^{N_R\times N_R}$ and the $N_R\times N_R$ diagonal matrix $\boldsymbol{\Sigma}_T$ contains N_C nonzero EVs of $\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}$. For any $\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}$, by introducing $\widetilde{\boldsymbol{X}}=\boldsymbol{X}\boldsymbol{U}_T$, we have $\widetilde{\boldsymbol{X}}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\widetilde{\boldsymbol{X}}=\boldsymbol{\Sigma}_T$, i.e., $\widetilde{\boldsymbol{X}}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\widetilde{\boldsymbol{X}}$ is diagonal, $\mathrm{Tr}\big(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\big)=\mathrm{Tr}\big(\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{H}}\big)$ and

$$\begin{aligned}&\det\big(\boldsymbol{I}_{N_R}+\sigma_r^2\sigma_d^{-2}\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}\big)\\&\quad=\det\big(\boldsymbol{I}_{N_R}+\sigma_r^2\sigma_d^{-2}\widetilde{\boldsymbol{X}}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\widetilde{\boldsymbol{X}}\big).\end{aligned} \tag{57}$$

Furthermore, according to the left-hand part of the identity (48), we have

$$\begin{aligned}&\det\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR}+\sigma_r^2\sigma_d^{-2}\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}\big)\\&\geq\det\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR}+\sigma_r^2\sigma_d^{-2}\boldsymbol{\Sigma}_T\big)\\&=\det\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR}+\sigma_r^2\sigma_d^{-2}\widetilde{\boldsymbol{X}}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\widetilde{\boldsymbol{X}}\big).\end{aligned} \tag{58}$$

The inequality in (58) becomes equality when $\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}$ is diagonal. By substituting (57) and (58) into (56), we have

$$\begin{aligned}&\frac{2}{B}\mathrm{EE}\big(\boldsymbol{V}_S,\boldsymbol{X},\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big)\\&\leq\frac{\log\det\big(\boldsymbol{I}_{N_R}+\sigma_r^2\sigma_d^{-2}\widetilde{\boldsymbol{X}}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\widetilde{\boldsymbol{X}}\big)}{\mathrm{Tr}\big(\boldsymbol{V}_S\big)+\mathrm{Tr}\big(\sigma_r^2\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{H}}\big)+P_C}+\frac{\log\det\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR}\big)}{\mathrm{Tr}\big(\boldsymbol{V}_S\big)+\mathrm{Tr}\big(\sigma_r^2\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{H}}\big)+P_C}\\&\quad-\frac{\log\det\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR}+\sigma_r^2\sigma_d^{-2}\widetilde{\boldsymbol{X}}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\widetilde{\boldsymbol{X}}\big)}{\mathrm{Tr}\big(\boldsymbol{V}_S\big)+\mathrm{Tr}\big(\sigma_r^2\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^{\mathrm{H}}\big)+P_C}\\&=\frac{2}{B}\mathrm{EE}\big(\boldsymbol{V}_S,\widetilde{\boldsymbol{X}},\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big).\end{aligned} \tag{59}$$

Since the inequality in (59) becomes equality for the diagonal $\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}$, to maximize the EE metric, the optimal $\boldsymbol{X}$ must satisfy $\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}=\boldsymbol{\Sigma}_T$. By introducing the $N_C\times N_C$ diagonal matrix $\overline{\boldsymbol{\Sigma}}_{RD}=\mathrm{diag}\big\{\widetilde{\sigma}_{rd,1}^2,\cdots,\widetilde{\sigma}_{rd,N_C}^2\big\}$, which is positive definite since the first N_C diagonal elements of $\widetilde{\boldsymbol{\Sigma}}_{RD}$ are positive, and denoting $\boldsymbol{X}^{\mathrm{H}}=\big[\boldsymbol{X}_1^{\mathrm{H}}\ \boldsymbol{X}_2^{\mathrm{H}}\big]$ with $\boldsymbol{X}_1\in\mathbb{C}^{N_C\times N_R}$ and $\boldsymbol{X}_2\in\mathbb{C}^{(N_R-N_C)\times N_R}$, we re-express $\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}=\boldsymbol{\Sigma}_T$ as

$$\begin{aligned}&\big[\boldsymbol{X}_1^{\mathrm{H}}\ \boldsymbol{X}_2^{\mathrm{H}}\big]\begin{bmatrix}\overline{\boldsymbol{\Sigma}}_{RD}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_1\\\boldsymbol{X}_2\end{bmatrix}=\begin{bmatrix}\overline{\boldsymbol{\Sigma}}_T&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{0}\end{bmatrix}\\&\Rightarrow\boldsymbol{X}_1^{\mathrm{H}}\overline{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}_1=\begin{bmatrix}\overline{\boldsymbol{\Sigma}}_T&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{0}\end{bmatrix},\end{aligned} \tag{60}$$

where $\overline{\boldsymbol{\Sigma}}_T\in\mathbb{C}^{N_C\times N_C}$ is a positive semidefinite diagonal submatrix of $\boldsymbol{\Sigma}_T$. (60) indicates that $\boldsymbol{X}_2$ has no effect on realizing $\boldsymbol{X}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RD}\boldsymbol{X}=\boldsymbol{\Sigma}_T$. Similarly to [6], we can infer from (60) that

$$\boldsymbol{X}=\big[\boldsymbol{X}_1^T,\boldsymbol{X}_2^T\big]^T=\Big[\big(\overline{\boldsymbol{\Sigma}}_{RD}^{-\frac{1}{2}}\boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}_T^{\frac{1}{2}}\big)^T,\boldsymbol{X}_2^T\Big]^T, \tag{61}$$

where $\boldsymbol{Q}\in\mathbb{C}^{N_C\times N_C}$ is an arbitrary unitary matrix and $\widetilde{\boldsymbol{\Sigma}}_T^{\frac{1}{2}}=\big[\overline{\boldsymbol{\Sigma}}_T^{\frac{1}{2}}\ \boldsymbol{0}\big]\in\mathbb{C}^{N_C\times N_R}$.

To further determine the optimal $\boldsymbol{X}_2$, we consider the relay power constraint

$$\begin{aligned}\mathrm{Tr}\big(\sigma_r^2\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\big)&=\mathrm{Tr}\big(\sigma_r^2\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{H}}\big)+\mathrm{Tr}\big(\sigma_r^2\boldsymbol{X}_2\boldsymbol{X}_2^{\mathrm{H}}\big)\\&\geq\mathrm{Tr}\big(\sigma_r^2\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{H}}\big),\end{aligned} \tag{62}$$

and the last inequality becomes the equality when $\boldsymbol{X}_2=\boldsymbol{0}_{(N_R-N_C)\times N_R}$. That is, $\boldsymbol{X}_2=\boldsymbol{0}_{(N_R-N_C)\times N_R}$ is the best choice in terms of minimizing relay power consumption. Furthermore,

$$\begin{aligned}&\mathrm{Tr}\big(\sigma_r^2\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{H}}\big)=\mathrm{Tr}\Big(\sigma_r^2\overline{\boldsymbol{\Sigma}}_{RD}^{-\frac{1}{2}}\boldsymbol{Q}\widetilde{\boldsymbol{\Sigma}}_T^{\frac{1}{2}}\big(\widetilde{\boldsymbol{\Sigma}}_T^{\frac{1}{2}}\big)^{\mathrm{H}}\boldsymbol{Q}^{\mathrm{H}}\big(\overline{\boldsymbol{\Sigma}}_{RD}^{-\frac{1}{2}}\big)^{\mathrm{H}}\Big)\\&=\mathrm{Tr}\big(\sigma_r^2\overline{\boldsymbol{\Sigma}}_{RD}^{-1}\boldsymbol{Q}\overline{\boldsymbol{\Sigma}}_T\boldsymbol{Q}^{\mathrm{H}}\big)\geq\mathrm{Tr}\big(\sigma_r^2\overline{\boldsymbol{\Sigma}}_{RD}^{-1}\overline{\boldsymbol{\Sigma}}_T\big),\end{aligned} \tag{63}$$

where the last inequality is due to the left-hand part of the identity (49) and the equality holds when $\boldsymbol{Q}=\boldsymbol{I}_{N_C}$ according to Lemma 2. Thus the optimal structure of $\boldsymbol{X}$ satisfies

$$\begin{aligned}\boldsymbol{X}^{\star}&=\begin{bmatrix}\overline{\boldsymbol{\Sigma}}_{RD}^{-\frac{1}{2}}\widetilde{\boldsymbol{\Sigma}}_T^{\frac{1}{2}}\\\boldsymbol{0}_{(N_R-N_C)\times N_R}\end{bmatrix}=\begin{bmatrix}\overline{\boldsymbol{\Sigma}}_{RD}^{-\frac{1}{2}}\overline{\boldsymbol{\Sigma}}_T^{\frac{1}{2}}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{0}\end{bmatrix}\\&=\boldsymbol{\Sigma}_X=\mathrm{diag}\{\sigma_{x,1},\cdots,\sigma_{x,N_C},0,\cdots,0\}.\end{aligned} \tag{64}$$

Using this optimal $\boldsymbol{X}^{\star}$ in (55), we obtain the optimal structure of the relay beamforming matrix

$$\boldsymbol{W}_R^{\star}=\widetilde{\boldsymbol{Q}}_{RD}\boldsymbol{\Sigma}_X\big(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR}\big)^{-\frac{1}{2}}\widetilde{\boldsymbol{U}}_{SR}^{\mathrm{H}}, \tag{65}$$

with

$$\mathrm{EE}\big(\boldsymbol{V}_S,\boldsymbol{X},\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big)\leq\mathrm{EE}\big(\boldsymbol{V}_S,\boldsymbol{\Sigma}_X,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big). \tag{66}$$

2) Optimal $\boldsymbol{V}_S^\star$: Denote the EVD of $\boldsymbol{V}_S$ by $\boldsymbol{V}_S = \boldsymbol{U}_S\boldsymbol{\Sigma}_S\boldsymbol{U}_S^{\mathrm{H}}$, where $\boldsymbol{U}_S \in \mathbb{C}^{N_S\times N_S}$ is the unitary matrix and $\boldsymbol{\Sigma}_S = \mathrm{diag}\{\lambda_{s,1},\cdots,\lambda_{s,N_S}\}$ has the N_S nonnegative diagonal elements. Using the SVD of $\boldsymbol{H}_{SR}$ given in (11), (52) can be rewritten as

$$\begin{aligned}\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}} &= \boldsymbol{U}_{SR}\boldsymbol{\Sigma}_{SR}\boldsymbol{Q}_{SR}^{\mathrm{H}}\boldsymbol{U}_S\boldsymbol{\Sigma}_S\boldsymbol{U}_S^{\mathrm{H}}\boldsymbol{Q}_{SR}\boldsymbol{\Sigma}_{SR}^{\mathrm{H}}\boldsymbol{U}_{SR}^{\mathrm{H}}\\ &= \widetilde{\boldsymbol{U}}_{SR}\widetilde{\boldsymbol{\Sigma}}_{SR}\widetilde{\boldsymbol{U}}_{SR}^{\mathrm{H}}. \end{aligned}\tag{67}$$

Furthermore, the EE metric in the problem (10) can be re-expressed as

$$\begin{aligned}&\frac{2}{B}\cdot\mathrm{EE}(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}) \qquad (68)\\ &= \frac{\log\det\Big(\boldsymbol{I}_{N_S}+\widetilde{\boldsymbol{V}}_S^{\mathrm{H}}\boldsymbol{H}_{SR}^{\mathrm{H}}\boldsymbol{W}_R^{\mathrm{H}}\boldsymbol{H}_{RD}^{\mathrm{H}}\boldsymbol{B}^{-1}\boldsymbol{H}_{RD}\boldsymbol{W}_R\boldsymbol{H}_{SR}\widetilde{\boldsymbol{V}}_S\Big)}{\mathrm{Tr}\big(\boldsymbol{V}_S\big)+\mathrm{Tr}\left(\boldsymbol{W}_R\big(\boldsymbol{H}_{SR}\boldsymbol{V}_S\boldsymbol{H}_{SR}^{\mathrm{H}}+\sigma_r^2\boldsymbol{I}_{N_R}\big)\boldsymbol{W}_R^{\mathrm{H}}\right)+P_C},\end{aligned}$$

where $\widetilde{\boldsymbol{V}}_S=\boldsymbol{U}_S\boldsymbol{\Sigma}_S^{\frac{1}{2}}$ and $\boldsymbol{B}=\sigma_r^2\boldsymbol{H}_{RD}\boldsymbol{W}_R\boldsymbol{W}_R^{\mathrm{H}}\boldsymbol{H}_{RD}^{\mathrm{H}}+\sigma_d^2\boldsymbol{I}_{N_D}$.

Substituting the optimal $\boldsymbol{W}_R^\star$ of (65) into (68) yields

$$\begin{aligned}&\frac{2}{B}\cdot\mathrm{EE}\left(\boldsymbol{V}_S,\boldsymbol{\Sigma}_X,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\right) \qquad (69)\\ &= \frac{\log\det\Big(\boldsymbol{I}_{N_S}+\widetilde{\boldsymbol{V}}_S^{\mathrm{H}}\boldsymbol{H}_{SR}^{\mathrm{H}}\widetilde{\boldsymbol{U}}_{SR}\widetilde{\boldsymbol{\Sigma}}_{RDX}\widetilde{\boldsymbol{U}}_{SR}^{\mathrm{H}}\boldsymbol{H}_{SR}\widetilde{\boldsymbol{V}}_S\Big)}{\mathrm{Tr}\big(\boldsymbol{\Sigma}_S\big)+\mathrm{Tr}\left(\sigma_r^2\boldsymbol{\Sigma}_X\boldsymbol{\Sigma}_X\right)+P_C},\end{aligned}$$

with

$$\widetilde{\boldsymbol{\Sigma}}_{RDX}=\boldsymbol{\Sigma}_{RDX}^{\mathrm{H}}\left(\sigma_d^2\boldsymbol{I}_{N_R}+\sigma_r^2\boldsymbol{\Sigma}_{RDX}\boldsymbol{\Sigma}_{RDX}^{\mathrm{H}}\right)^{-1}\boldsymbol{\Sigma}_{RDX},\tag{70}$$

where $\boldsymbol{\Sigma}_{RDX}=\widetilde{\boldsymbol{\Sigma}}_{RD}^{\frac{1}{2}}\boldsymbol{\Sigma}_X(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\widetilde{\boldsymbol{\Sigma}}_{SR})^{-\frac{1}{2}}$. Since $\mathrm{rank}(\boldsymbol{\Sigma}_X)$ $=\mathrm{rank}\big(\widetilde{\boldsymbol{\Sigma}}_{RD}\big)=N_C$, the diagonal matrix $\widetilde{\boldsymbol{\Sigma}}_{RDX}$ also has the rank of N_C. According to (67), we can re-express (69) as

$$\begin{aligned}&\frac{2}{B}\cdot\mathrm{EE}\left(\boldsymbol{V}_S,\boldsymbol{\Sigma}_X,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\right)\\ &= \frac{\log\det\Big(\boldsymbol{I}_{N_R}+\overline{\boldsymbol{U}}_{SR}^{\mathrm{H}}\boldsymbol{\Sigma}_{SR}\overline{\boldsymbol{Q}}_{SR}\boldsymbol{\Sigma}_S\overline{\boldsymbol{Q}}_{SR}^{\mathrm{H}}\boldsymbol{\Sigma}_{SR}^{\mathrm{H}}\overline{\boldsymbol{U}}_{SR}\widetilde{\boldsymbol{\Sigma}}_{RDX}\Big)}{\mathrm{Tr}\big(\boldsymbol{\Sigma}_S\big)+\mathrm{Tr}\left(\sigma_r^2\boldsymbol{\Sigma}_X\boldsymbol{\Sigma}_X\right)+P_C}\\ &\le \frac{\log\det\Big(\boldsymbol{I}_{N_R}+\boldsymbol{\Sigma}_{SR}\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_{SR}^{\mathrm{H}}\widetilde{\boldsymbol{\Sigma}}_{RDX}\Big)}{\mathrm{Tr}\big(\boldsymbol{\Sigma}_S\big)+\mathrm{Tr}\left(\sigma_r^2\boldsymbol{\Sigma}_X\boldsymbol{\Sigma}_X\right)+P_C}\\ &= \frac{2}{B}\cdot\mathrm{EE}\left(\boldsymbol{\Sigma}_S,\boldsymbol{\Sigma}_X,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\right),\end{aligned}\tag{71}$$

where $\overline{\boldsymbol{U}}_{SR}=\boldsymbol{U}_{SR}^{\mathrm{H}}\widetilde{\boldsymbol{U}}_{SR}$ and $\overline{\boldsymbol{Q}}_{SR}=\boldsymbol{Q}_{SR}^{\mathrm{H}}\boldsymbol{U}_S$ are unitary matrices. According to the right-hand part of identity (48), it can be inferred that $\det\left(\boldsymbol{I}_N+\boldsymbol{A}\boldsymbol{B}\right)\le\prod_{i=1}^N\left(1+\lambda_i(\boldsymbol{A})\lambda_i(\boldsymbol{B})\right)$ and the inequality in (71) becomes equality when $\overline{\boldsymbol{U}}_{SR}=\boldsymbol{I}_{N_R}$ and $\overline{\boldsymbol{Q}}_{SR}=\boldsymbol{I}_{N_S}$. The diagonal matrix $\widetilde{\boldsymbol{\Sigma}}_{RDX}$ (70) is satisfied at the maximum EE point. As a result, the optimal structures of $\widetilde{\boldsymbol{U}}_{SR}$ and $\boldsymbol{U}_S$ are $\widetilde{\boldsymbol{U}}_{SR}=\boldsymbol{U}_{SR}$ and $\boldsymbol{U}_S=\boldsymbol{Q}_{SR}$. Accordingly, the optimal structure of $\boldsymbol{V}_S$ is given by

$$\boldsymbol{V}_S^\star=\boldsymbol{Q}_{SR}\boldsymbol{\Sigma}_S\boldsymbol{Q}_{SR}^{\mathrm{H}}.\tag{72}$$

Combining (67) with (72) leads to $\boldsymbol{H}_{SR}\boldsymbol{V}_S^\star\boldsymbol{H}_{SR}^{\mathrm{H}}=\boldsymbol{U}_{SR}\boldsymbol{\Sigma}_{SR}\cdot\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_{SR}^{\mathrm{H}}\boldsymbol{U}_{SR}^{\mathrm{H}}=\widetilde{\boldsymbol{U}}_{SR}\widetilde{\boldsymbol{\Sigma}}_{SR}\widetilde{\boldsymbol{U}}_{SR}^{\mathrm{H}}$ and due to $\widetilde{\boldsymbol{U}}_{SR}=\boldsymbol{U}_{SR}$, we also have $\widetilde{\boldsymbol{\Sigma}}_{SR}=\boldsymbol{\Sigma}_{SR}\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_{SR}^{\mathrm{H}}$. Thus the optimal structure of the relay beamforming matrix in (65) can be re-expressed as

$$\boldsymbol{W}_R^\star=\widetilde{\boldsymbol{Q}}_{RD}\boldsymbol{\Sigma}_X\left(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\boldsymbol{\Sigma}_{SR}\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_{SR}^{\mathrm{H}}\right)^{-\frac{1}{2}}\boldsymbol{U}_{SR}^{\mathrm{H}}.\tag{73}$$

3) Worst-case $\boldsymbol{\Delta}_{RD}^\star\varnothing\boldsymbol{E}_{RD}^\star$: First, we introduce the following lemma [15], [37].

Lemma 3. *For $\boldsymbol{A}\in\mathbb{C}^{N_u\times N_l}$, $\boldsymbol{B},\boldsymbol{C}\in\mathbb{C}^{N_l\times N_m}$ with* $\mathrm{rank}(\boldsymbol{A})=\mathrm{rank}(\boldsymbol{B})=\mathrm{rank}(\boldsymbol{C})=N=\min\{N_u,N_l,N_m\}$, *whose SVs are $\sigma_i(\boldsymbol{A})$, $\sigma_i(\boldsymbol{B})$ and $\sigma_i(\boldsymbol{C})$, $i=1,\cdots,N$, respectively, we have*

$$\sigma_N(\boldsymbol{A})\sigma_i(\boldsymbol{B})\le\sigma_i(\boldsymbol{A}\boldsymbol{B}),\tag{74}$$

$$\big(\sigma_i(\boldsymbol{B})-\sigma_1(\boldsymbol{C})\big)^+\le\sigma_i(\boldsymbol{B}+\boldsymbol{C}).\tag{75}$$

(74) and (75) become equalities only if $\{\boldsymbol{A},\boldsymbol{B}\}$ and $\{\boldsymbol{B},\boldsymbol{C}\}$ are simultaneously diagonalizable.

Based on the definitions of the diagonal matrices $\boldsymbol{\Sigma}_S$, $\boldsymbol{\Sigma}_X$ and $\widetilde{\boldsymbol{\Sigma}}_{RDX}$ as well as the rectangular diagonal matrix $\boldsymbol{\Sigma}_{SR}$, at $\boldsymbol{V}_S^\star$ and $\boldsymbol{W}_R^\star$, the achievable EE metric in (71) can be expressed as (76) at the top of the next page, where $N_L=\min\{N_P,N_C\}$. It is readily observed from the first term of the numerator in (76) that the minimum value of $\mathrm{EE}(\boldsymbol{\Sigma}_S,\boldsymbol{\Sigma}_X,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD})$ is attained only when every $\widetilde{\sigma}_{rd,i}^2$ realizes its minimum, subject to the spectral norm constraint. Note that $\widetilde{\sigma}_{rd,i}^2$, $1\le i\le N_C$, are unknown since they are related to the unknown CSI $\boldsymbol{H}_{RD}$. However, for the known nominal CSI $\widehat{\boldsymbol{H}}_{RD}$, we have the SVD $\widehat{\boldsymbol{H}}_{RD}=\widehat{\boldsymbol{U}}_{RD}\widehat{\boldsymbol{\Sigma}}_{RD}\widehat{\boldsymbol{Q}}_{RD}^{\mathrm{H}}$, in which the $N_D\times N_R$ diagonal rectangular matrix $\widehat{\boldsymbol{\Sigma}}_{RD}$ contains N_C positive elements $\{\widehat{\sigma}_{rd,1},\cdots,\widehat{\sigma}_{rd,N_C}\}$.

3.1) Additive CSI errors: $\boldsymbol{H}_{RD}^{\mathrm{H}}\boldsymbol{H}_{RD}=\big(\widehat{\boldsymbol{H}}_{RD}+\boldsymbol{\Delta}_{RD}\big)^{\mathrm{H}}\cdot\big(\widehat{\boldsymbol{H}}_{RD}+\boldsymbol{\Delta}_{RD}\big)=\widetilde{\boldsymbol{Q}}_{RD}\widetilde{\boldsymbol{\Sigma}}_{RD}\widetilde{\boldsymbol{Q}}_{RD}^{\mathrm{H}}$. Applying Lemma 3 and considering $\big\|\boldsymbol{\Delta}_{RD}\big\|_2=\sigma_1\big(\boldsymbol{\Delta}_{RD}\big)\le\epsilon_a$, we conclude

$$\widetilde{\sigma}_{rd,i}\ge\big(\widehat{\sigma}_{rd,i}-\epsilon_a\big)^+,\ 1\le i\le N_C.\tag{77}$$

All the inequalities in (77) become the equalities when $\widetilde{\boldsymbol{Q}}_{RD}=\widehat{\boldsymbol{Q}}_{RD}$ according to Lemma 3, and the minimum $\mathrm{EE}\left(\boldsymbol{\Sigma}_S,\boldsymbol{\Sigma}_X,\boldsymbol{\Delta}_{RD}\right)$ is attained with $\widetilde{\sigma}_{rd,i}=\big(\widehat{\sigma}_{rd,i}-\epsilon_a\big)^+$, $1\le i\le N_C$. Consequently, the resultant worst-case CSI error is given by

$$\begin{aligned}\boldsymbol{\Delta}_{RD}^\star&=-\widehat{\boldsymbol{U}}_{RD}\begin{bmatrix}\widetilde{\boldsymbol{\Lambda}}_{RD} & \boldsymbol{0}_{N_C\times(N_R-N_C)}\\ \boldsymbol{0}_{(N_D-N_C)\times N_C} & \boldsymbol{0}_{(N_D-N_C)\times(N_R-N_C)}\end{bmatrix}\widehat{\boldsymbol{Q}}_{RD}^{\mathrm{H}}\\ &=-\widehat{\boldsymbol{U}}_{RD}\boldsymbol{\Lambda}_{RD}\widehat{\boldsymbol{Q}}_{RD}^{\mathrm{H}},\end{aligned}\tag{78}$$

where $\widetilde{\boldsymbol{\Lambda}}_{RD}=\mathrm{diag}\left\{\min\{\widehat{\sigma}_{rd,1},\epsilon_a\},\cdots,\min\{\widehat{\sigma}_{rd,N_C},\epsilon_a\}\right\}$, and we also have

$$\mathrm{EE}(\boldsymbol{\Sigma}_S,\boldsymbol{\Sigma}_X,\boldsymbol{\Delta}_{RD})\ge\mathrm{EE}(\boldsymbol{\Sigma}_S,\boldsymbol{\Sigma}_X,\boldsymbol{\Delta}_{RD}^\star).\tag{79}$$

Observe that $\widehat{\boldsymbol{U}}_{RD}$, $\boldsymbol{\Lambda}_{RD}$ and $\widehat{\boldsymbol{Q}}_{RD}^{\mathrm{H}}$ in (78) are all known.

3.2) Multiplicative CSI errors: $\boldsymbol{H}_{RD}^{\mathrm{H}}\boldsymbol{H}_{RD}=\widehat{\boldsymbol{H}}_{RD}^{\mathrm{H}}\big(\boldsymbol{I}_{N_D}+\boldsymbol{E}_{RD}\big)^{\mathrm{H}}\big(\boldsymbol{I}_{N_D}+\boldsymbol{E}_{RD}\big)\widehat{\boldsymbol{H}}_{RD}=\widetilde{\boldsymbol{Q}}_{RD}\widetilde{\boldsymbol{\Sigma}}_{RD}\widetilde{\boldsymbol{Q}}_{RD}^{\mathrm{H}}$. Similarly to the case of additive CSI errors, based on Lemma 3 and $\|\boldsymbol{E}_{RD}\|_2=\sigma_1\big(\boldsymbol{E}_{RD}\big)\le\epsilon_m$, the minimum values of $\widetilde{\sigma}_{rd,i},\forall i$, required for minimizing $\mathrm{EE}\left(\boldsymbol{\Sigma}_S,\boldsymbol{\Sigma}_X,\boldsymbol{E}_{RD}\right)$ are obtained as

$$\begin{aligned}&\widetilde{\sigma}_{rd,i}\ge\sigma_{N_D}\big(\boldsymbol{I}_{N_D}+\boldsymbol{E}_{RD}\big)\widehat{\sigma}_{rd,i}\ge(1-\epsilon_m)^+\widehat{\sigma}_{rd,i},\\ &1\le i\le N_C.\end{aligned}\tag{80}$$

The two inequalities in (80) simultaneously become the equalities when $\widetilde{\boldsymbol{Q}}_{RD}=\widehat{\boldsymbol{Q}}_{RD}$ according to Lemma 3. Therefore, the resulting worst-case $\boldsymbol{E}_{RD}^\star$ is given by $\boldsymbol{E}_{RD}^\star=$

$$\frac{2}{B}\cdot \mathrm{EE}\left(\boldsymbol{\Sigma}_S,\boldsymbol{\Sigma}_X,\boldsymbol{\Delta}_{RD}\varnothing \boldsymbol{E}_{RD}\right)=\frac{\sum_{i=1}^{N_L}\log\left(\frac{1+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}{1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}\right)+\sum_{i=1}^{N_L}\log\left(1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}\right)}{\sum_{i=1}^{N_S}\lambda_{s,i}+\sum_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2+P_C}$$
$$=\frac{\sum_{i=1}^{N_L}\log\left(1-\frac{\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}}{1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}+\sigma_r^2\sigma_d^{-2}\sigma_{x,i}^2\widetilde{\sigma}_{rd,i}^2}\right)+\sum_{i=1}^{N_L}\log\left(1+\sigma_r^{-2}\sigma_{sr,i}^2\lambda_{s,i}\right)}{\sum_{i=1}^{N_S}\lambda_{s,i}+\sum_{i=1}^{N_C}\sigma_r^2\sigma_{x,i}^2+P_C}, \tag{76}$$

$-\epsilon_m \boldsymbol{I}_{N_D}$, which only depends on the known ϵ_m, and we naturally have

$$\mathrm{EE}\left(\boldsymbol{\Sigma}_S,\boldsymbol{\Sigma}_X,\boldsymbol{E}_{RD}\right)\geq \mathrm{EE}\left(\boldsymbol{\Sigma}_S,\boldsymbol{\Sigma}_X,\boldsymbol{E}_{RD}^{\star}\right). \tag{81}$$

Because $\widetilde{\boldsymbol{Q}}_{RD}=\widehat{\boldsymbol{Q}}_{RD}$ holds for both additive and multiplicative CSI errors, (73) becomes

$$\boldsymbol{W}_R^{\star}=\widehat{\boldsymbol{Q}}_{RD}\boldsymbol{\Sigma}_X\left(\boldsymbol{I}_{N_R}+\sigma_r^{-2}\boldsymbol{\Sigma}_{SR}\boldsymbol{\Sigma}_S\boldsymbol{\Sigma}_{SR}^{\mathrm{H}}\right)^{-\frac{1}{2}}\boldsymbol{U}_{SR}^{\mathrm{H}}. \tag{82}$$

Observe that $\widehat{\boldsymbol{Q}}_{RD}$, $\boldsymbol{\Sigma}_{SR}$ and $\boldsymbol{U}_{SR}$ are all known, while $\boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_X$ are the new optimization variables. This completes the proof. □

B. *Proof of* Theorem 2

Proof. From (66) and (71) in Appendix A, it is easily seen that $\mathrm{EE}\big(\boldsymbol{V}_S^{\star},\boldsymbol{W}_R^{\star},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)\geq \mathrm{EE}\big(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)$ holds for any feasible $\boldsymbol{V}_S$ and $\boldsymbol{W}_R$. Similarly, from (79) and (81) in Appendix A, it is seen that $\mathrm{EE}\big(\boldsymbol{V}_S^{\star},\boldsymbol{W}_R^{\star},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\big)\leq \mathrm{EE}\big(\boldsymbol{V}_S^{\star},\boldsymbol{W}_R^{\star},\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\big)$ holds for any feasible $\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}$. Thus the optimal $\{\boldsymbol{V}_S^{\star}, \boldsymbol{W}_R^{\star},\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}\}$ is a saddle point of the original robust EE optimization problem (8). This completes the proof. □

C. *Proof of* Theorem 3

Proof. According to [11], for the statistically imperfect source-relay channel $\boldsymbol{H}_{SR}$, the channel-diagonalizing structure is optimal for the source covariance matrix $\boldsymbol{V}_S^{\star}$ and the relay beamforming $\boldsymbol{W}_R^{\star}$ for any relay-destination channel $\boldsymbol{H}_{RD}$. That is, the eigenvectors of the optimal $\boldsymbol{V}_S^{\star}$ are aligned with that of the source correlation matrix $\boldsymbol{R}_S$, while the left and right singular matrices of the optimal $\boldsymbol{W}_R^{\star}$ are aligned with the right singular matrix of the relay-destination channel $\boldsymbol{H}_{RD}$ and the eigenvectors of the relay correlation matrix $\boldsymbol{R}_R$, respectively [11]. Based on the optimal $\boldsymbol{V}_S^{\star}$ and $\boldsymbol{W}_R^{\star}$ with the channel-diagonalizing structure (29), we naturally obtain the same worst-case error $\boldsymbol{\Delta}_{RD}^{\star}\varnothing\boldsymbol{E}_{RD}^{\star}$ as that given in Theorem 1 by utilizing Lemma 3 of Appendix A. Moreover, we can also conclude that the solutions provided in Theorem 3 are the saddle point of $\widetilde{\mathrm{EE}}\left(\boldsymbol{V}_S,\boldsymbol{W}_R,\boldsymbol{\Delta}_{RD}\varnothing\boldsymbol{E}_{RD}\right)$ by referring to the proof of Theorem 2. □

D. *Proof of* Theorem 4

Proof. Construct the min-max EE counterpart problem to the problem (33). First by fixing the relay beamforming matrices of relays $k\in\{2,\cdots,K\}$ and referring to the proof of Theorem 1, we obtain the optimal beamforming matrix of relay $k'=1$ and the optimal source covariance matrix, both having channel-diagonalizing structure, as well as obtain the same worst-case CSI error as given in Theorem 1. In a similar manner, the remaining optimal beamforming matrices of relays k', $2\leq k'\leq K$, with channel-diagonalizing structure can be obtained one by one by fixing the relay beamforming matrices of relays $k\in\{1,2,\cdots,K\}\setminus k'$ and given the optimal source covariance matrix. This proves that the solution of (42) to (44) form the optimal solution of this min-max EE counterpart problem. Then referring to the proof of Theorem 2, we conclude that the solution of (42) to (44) is a saddle point of $\mathrm{EE}_M\Big(\boldsymbol{V}_S,\widetilde{\boldsymbol{W}}_R,\boldsymbol{\Delta}_{R_KD}\varnothing\boldsymbol{E}_{R_KD}\Big)$. □

References

[1] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[2] Y. Song, H. Shin, and E. K. Hong, "MIMO cooperative diversity with scalar-gain amplify-and-forward relaying," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 1932–1938, Jul. 2009.

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

[4] S. Gong, *et al.*, "Energy efficient transmission in multi-user MIMO relay channels with perfect and imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3885–3898, Jun. 2017.

[5] S. Gong, C. Xing, Z. Fei, and S. Ma, "Millimeter-wave secrecy beamforming designs for two-way amplify-and-forward MIMO relaying networks," *IEEE Trans. Veh. Techno.*, vol. 66, no. 3, pp. 2059–2071, Mar. 2017.

[6] 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.

[7] K. Lee, H. Sung, E. Park, and I. Lee, "Joint optimization for one and two-way MIMO AF multiple-relay systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3671–3681, Dec. 2010.

[8] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.

[9] Y. Li, *et al.*, "Max-min energy-efficient power allocation in interference-limited wireless networks," *IEEE Trans. Veh. Techno.*, vol. 64, no. 9, pp. 4321–4326, Sep. 2015.

[10] F. Héliot, "Low-complexity energy-efficient joint resource allocation for two-hop MIMO-AF systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3088–3099, Jun. 2014.

[11] A. Zappone, P. Cao, and E. A. Jorswieck, "Energy efficiency optimization in relay-assisted MIMO systems with perfect and statistical CSI," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 443–457, Jan. 2014.

[12] F. Héliot and R. Tafazolli, "Optimal energy-efficient joint resource allocation for multi-hop MIMO-AF systems," *IEEE Trans. Commun*, vol. 64, no. 9, pp. 3655–3668, Sep. 2016.

[13] A. Zappone, P. Cao, and E. A. Jorswieck, "Low-complexity energy efficiency optimization with statistical CSI in two-hop MIMO systems," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1398–1402, Nov. 2014.

[14] A. Soysal and S. Ulukus, "Optimum power allocation for single-user MIMO and multi-user MIMO-MAC with partial CSI," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1402–1412, Sep. 2007.

[15] S. Loyka and C. D. Charalambous, "On the compound capacity of a class of MIMO channels subject to normed uncertainty," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2048–2063, Apr. 2012.

[16] C. Jeong, *et al.*, "Relay precoding for non-regenerative MIMO relay systems with partial CSI feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1698–1711, May 2012.

[17] H. Shen, J. Wang, B. C. Levy, and C. Zhao, "Robust optimization for amplify-and-forward MIMO relaying from a worst-case perspective," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5458–5471, Nov. 2013.

[18] G. Zheng, K. K. Wong, A. Paulraj, and B. Ottersten, "Robust collaborative-relay beamforming," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3130–3143, Aug. 2009.

[19] Y. Cui, *et al.*, "CSI impaired precoding optimization for energy-efficient MIMO communications under total power constraint," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 514–517, Mar. 2016.

[20] L. Wang, *et al.*, "Mean energy efficiency maximization in cognitive radio channels with PU outage constraint," *IEEE Commun. Lett.*, vol. 19, no. 2, pp. 287–290, Feb. 2015.

[21] K. Wang, *et al.*, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp 5690–5705, Nov. 2014.

[22] K. Shen and W. Yu, "Fractional programming for communication systems - part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.

[23] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.

[24] O. Tervo, L. N. Tran, and M. Juntti, "Optimal energy-efficient transmit beamforming for multi-user MISO downlink," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5574–5588, Oct. 2015.

[25] J. Xu, L. Qiu, and C. Yu, "Improving energy efficiency through multimode transmission in the downlink MIMO systems," *EURASIP J. Wirel. Commun. Netw.*, vol. 2011, no. 200, pp. 1–12, Dec. 2011.

[26] A. Lapidoth and S. Shamai, "Fading channels: how perfect need 'perfect side information' be?" *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1118-1134, May 2002.

[27] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1970.

[28] M. Udell and S. Boyd, Maximizing a Sum of Sigmoids, 2013. [Online]. Available: http://www.stanford.edu/∼boyd /papers/max_sum_sigmoids.html

[29] S. Boyd, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2004.

[30] S. Schaible, "Fractional programming," *Zeitschrift für Operations Research*, vol. 27, no. 1, pp. 39–54, 1983.

[31] 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.

[32] Y. Qi, F. Héliot, M. A. Imran, and R. Tafazolli, "Green relay techniques in cellular systems," chapter 3 in F. R. Yu, X. Zhang, and V. C. M. Leung (eds.), *Green Communications and Networking*. Boca Raton, FL: CRC Press, 2013.

[33] F. Héliot and R. Tafazolli, "Optimal energy-efficient source and relay precoder design for cooperative MIMO-AF systems," *IEEE Trans. Signal Process.*, vol. 66, no. 3, pp. 573–588, Feb. 2018.

[34] K. Lee, J. Kim, G. Caire, and I. Lee, "Asymptotic ergodic capacity analysis for MIMO amplify-and-forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2712–2717, Sep. 2010.

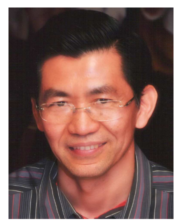
[35] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.

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

[37] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1991.

Shiqi Gong received her B.S. degree in Electronic Engineering from Beijing Institute of Technology, Beijing, China, in 2014. Currently, she is pursuing the Ph.D. degree with the School of Electronic and Information, Beijing Institute of Technology. From February 2017 to January 2019, she served as a Research Assistant in the Faculty of Science and Technology, University of Macau. Currently, she is a visiting Ph.D. student with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. Her research interests include array signal processing, physical-layer security, resource allocation, and convex optimization.

Shuai Wang received his B. Eng degree and Ph.D. degree from Zhengzhou University and Beijing Institute of Technology, P. R. China, in 2005 and 2012 respectively, both in the major of Communications Engineering and both with the highest honor. He won the award for "outstanding Ph.D. dissertation" granted by the Beijing Municipal Education Commission in 2013 with other 49 co-winners, nominated from all the Ph.D. graduates that received their degrees in Beijing that year. From 2010.09 to 2011.09, he was a visiting Ph.D. student in the School of Electronics and Computer Science, University of Southampton, U.K. He has been with the School of Information Science and Electronics, Beijing Institute of Technology since July 2012, where he now holds a post of Associate Professor. He is also a recipient of the (Second Class) Scientific and Technical Progress Award granted by the Ministry of Industry and Information Technology of China. His research interest includes channel estimation, anti-jamming transmission, synchronization techniques and beamforming.



Sheng Chen (M'90-SM'97-F'08) received his BEng degree from the East China Petroleum Institute, Dongying, China, in 1982, and his PhD degree from the City University, London, in 1986, both in control engineering. In 2005, he was awarded the higher doctoral degree, Doctor of Sciences (DSc), from the University of Southampton, Southampton, UK. From 1986 to 1999, He held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in UK. Since 1999, he has been with the School of Electronics and Computer Science, the University of Southampton, UK, where he holds the post of Professor in Intelligent Systems and Signal Processing. Dr Chen's research interests include adaptive signal processing, wireless communications, modelling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods and optimisation. He has published over 650 research papers. Dr. Chen is a Fellow of the United Kingdom Royal Academy of Engineering, a Fellow of IET, a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia, and an original ISI highly cited researcher in engineering (March 2004). Professor Chen has 13,000+ Web of Science citations and 27,000+ Google Scholar citations.

Chengwen Xing (S'08-M'10) received his B.Eng. degree from Xidian University, Xian, China, in 2005, and his Ph.D. degree from the University of Hong Kong, Hong Kong, China, in 2010. Since September 2010, he has been with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, where he is currently a Full Professor. From September 2012 to December 2012, he was a visiting scholar at the University of Macau. His current research interests include statistical signal processing, convex optimization, multivariate statistics, combinatorial optimization, massive MIMO systems, and high frequency band communication systems. Prof. Xing is an Associate Editor for the IEEE Transactions on Vehicular Technology, KSII Transactions on Internet and Information Systems, Transactions on Emerging Telecommunications Technologies, and China Communications.

Lajos Hanzo (F'04) received his 5-year degree in electronics in 1976 and his doctorate in 1983 from the Technical University of Budapest. In 2009, he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016, he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications, he has held various research and academic posts in Hungary, Germany and the U.K. Since 1986, he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 119 PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1800+ research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently, he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE ComSoc and VTS. He is a former Editor-in-Chief of the IEEE Press and a former Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications, please refer to http://www-mobile.ecs.soton.ac.uk