Proof of the Equivalency of MSE-Constrained and Rate-Constrained Power Optimization Approaches to Distributed Bi-Directional Beamforming

Razgar Rahimi, Shahram ShahbazPanahi, and Björn Ottersten

***Abstract*—Considered in this article is an *asynchronous* single-carrier two-way relay network, where two transceivers aim to communicate through a set of amplify-and-forward (AF) relays. This network is asynchronous in the sense that the signal transmitted by any of the two transceivers arrives the other transceiver through different relaying paths, with possibly different propagation delays, thereby materializing a multipath channel. At sufficiently high data rates, this multi-path end-to-end channel causes inter-symbol-interference (ISI) in the received signals. Considering such a network, this paper presents two contributions. The first contribution is the rigorous characterization of the region of the mean-squared errors (MSEs) of the symbol estimates at the two transceivers under a total power budget, and when linear block post-channel equalization is used at the receiver front-end of the two transceivers. The importance of this MSE region characterization resides in the fact that knowing this region allows for characterization the region of un-coded probabilities of error at the two transceivers. Also, this MSE region characterization paves the way towards presenting the second contribution in this paper. Indeed, in the second contribution, this article relies on this MSE region characterization to rigorously prove that an MSE-constrained total power minimization approach and a rate-constrained total power minimization approach to design transceiver power allocation and distributed beamforming are equivalent, if the MSE thresholds in the former approach and the rate thresholds in the latter approach are properly chosen. The equivalence of these two approaches implies that the un-coded MSE performance of the network can be inferred from the rate-constrained problem, and conversely, the coded rate performance of the network can be inferred from the MSE-constrained total power minimization problem.**

***Index Terms*—Asynchronous networks, cooperative communications, distributed beamforming, power control, power minimization, two-way relaying.**

Manuscript received June 13, 2019; revised February 5, 2020; accepted March 5, 2020. Date of publication March 26, 2020; date of current version June 16, 2020. This work was supported in part by the Natural Science and Engineering Research Council (NSERC) of Canada through its Discovery Grant and Collaborative Research and Development programs and by FNR, Luxembourg under the projects ECLECTIC and COHESAT and also through the FNR's Inter-Mobility program. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Philippe Ciblat. *(Corresponding author: Shahram ShahbazPanahi.)*

Razgar Rahimi is with the Department of Electrical, Software, and Computer Engineering, University of Ontario Institute of Technology, Oshawa, ON L1G 0C5, Canada (e-mail: Razgar.Rahimi@uoit.ca).

Shahram ShahbazPanahi is with the Department of Electrical, Software, and Computer Engineering, University of Ontario Institute of Technology, Oshawa, ON L1G 0C5, Canada, and also with the SnT Centre, Luxembourg City 1359, Luxembourg (e-mail: shahram.shahbazpanahi@uoit.ca).

Björn Ottersten is with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg City, 4365 Luxembourg (e-mail: bjorn.ottersten@uni.lu).

Digital Object Identifier 10.1109/TSIPN.2020.2981990

I. Introduction

RELAY-ASSISTED communication has been the center focus of numerous studies. Majority of the studies conducted on relay networks assume that the signal transmitted by any source in the network arrives at different relay nodes within the network with the same delays and that signals transmitted by different relay nodes arrive at any destination in the network with the same delays [1]–[5]. Obviously such assumptions are not realistic as source-relay and relay-destinations distances are different for different relays. As a result, in practice, the signal transmitted by a source arrives at different relays with different delays and the signals transmitted by different relays arrive at any destination with different delays. Hence, each relaying path causes its own delay in the signal traveling from a source, through the corresponding relay, to a destination. We herein refer to such a network as an asynchronous relay network. In an asynchronous relay network, the end-to-end channel is best modeled as a multipath link, which can cause inter-symbol interference. For example, in an LTE network with a rate of 20 Msymbols/second, only 15 meters difference between the distances travelled by the signal through two different relay nodes will cause one symbol difference in arrival times [6]–[8]. Larger distance differentials will cause longer delay spreads in the end-to-end channel.

Another possible use-case for asynchronous two-way relay networks is found in satellite communications. Satellites relay uplink messages from ground based transmitters back to ground based receivers. Large low-earth orbit (LEO) satellite constellations are being considered to provide world-wide high-speed internet access through mass produced and affordable satellites [9]. The possibility of coordinated relay transmission by several satellites can significantly improve link budgets, lower power consumption, and increase the spectral efficiency of the network [10].

Compared to the volume of studies conducted on synchronous relay networks, the results published on asynchronous relay networks are scarce. To put our work into the context of the available literature, we briefly review some of these results. The problem of timing estimation and relay synchronization has been studied in [11]–[14]. The authors of [15] and [16] develop a full-diversity achieving transmission scheme with optimal coding gain for an asynchronous decode-and-forward (DF) based relay network consisting of a source, a destination,

and multiple relays. In [17], the authors consider an asynchronous relay network, where each relay employs an FIR filter to obtain a filtered version of the relay's received signal, and then, forwards the so-obtained signal as the transmit signal. The authors of [17] also study the diversity of linear post-channel equalization schemes, such as zero-forcing and minimum mean squared error (MMSE) receivers. In [18], the authors consider an asynchronous bidirectional relay network, where a single multi-antenna relay enables a two-way information exchange between two transceivers, and devise a channel and timing offset estimation method to re-synchronize the relays. The study in [19] considers a bi-directional relay network, with multiple single-antenna relays. In the communication scheme of [19], the transceivers insert cyclic prefix (CP) in their transmit signals while each relay removes the CP from its received signal, circularly shifts the residual signal, inserts the CP again, and retransmits the so-obtained signal. The transceivers employ linear post-channel equalization schemes (such as zero forcing (ZF) and MMSE receivers) at their receiver front-ends. For such a communication scheme, the authors of [19] study the bit error rate (BER) performance and present a BER-optimal power allocation method. In [20], the authors consider a bi-directional DF relay network and present post-channel equalization schemes, such as MMSE and MMSE decision feedback equalizers for the case when multiple frequency offsets are present at the relays. By modeling the end-to-end channel as a multi-path link, the results in [6]–[8], [21]–[25] use different criteria to design distributed beamforming techniques for asynchronous two-way relay networks.

In this paper, we consider an asynchronous amplify-and-forward (AF) two-way relay network, which consists of two transceivers and multiple single-antenna relay nodes. For such a network, we present the following contributions:

- As the first contribution, under a total power budget and assuming linear post-channel block equalization at the receiver front-end of the two transceivers, this paper solves the problem of minimizing the mean-squared error (MSE) of the symbol estimates at the output of the linear block equalizer of one transceiver, while the MSE of the symbol estimates at the output of the linear block equalizer at the other transceiver is kept at a certain level. Carried over the transceivers' transmit powers and the relay complex beamforming weights, this minimization allows us to characterize the MSE region. This MSE region characterization is important as characterizing this region allows us to characterize the region of un-coded probabilities of error at the two transceivers. Also, this MSE region characterization paves the way towards presenting the next contribution in this paper.
- The second contribution of this paper is optimally finding the transceivers' transmit powers and relay complex beamforming weights by solving the total power minimization problem under MSE constraints and showing how this solution is related to the solution of the rate-constrained power minimization problem. More specifically, we rely on our MSE region characterization to rigorously prove that in designing transceiver power allocation and distributed beamforming, our MSE-constrained total power minimization approach and the rate-constrained power minimization technique of [24] are equivalent, provided that the MSE thresholds in the former approach and the rate thresholds in the latter approach satisfy a logarithmic (or an exponential) relationship. The equivalence of these two approaches allows us to infer the un-coded MSE performance of the network from the rate-constrained problem, and conversely, the coded rate performance of the network can be inferred from the MSE-constrained total power minimization problem.

Presenting the above two contributions heavily rests on a new formulation that we present for the end-to-end channel impulse response (CIR) between the two transceivers. Different from previous formulations appeared in [6]–[8], [21]–[25], this new formulation orders the relays based on the end-to-end delay they impose in the signal they are relaying. Doing so reveals a ceratin structure for the matrix which dictates which relays contribute to which tap of the end-to-end CIR. It is this interesting structure which facilitates the derivation of the subsequent contributions.

In the sequel, we differentiate our work from the relevant results on asynchronous relay networks. Unlike [11]–[14], we do not perform relay synchronization, thereby avoiding complexity at the relays. Indeed, as the relays use simple AF relaying protocols, requiring relay synchronization will increase the processing complexity at the relays. The networks considered in [15] and [16] use a one-way DF-based scheme, which accounts only for the different delays from different relays to the destination, while we herein assume an asynchronous bi-directional AF relay network, and thus, must and will take into account the delays corresponding to the channels between both transceivers and the relays. The relay network studied in [17] relies on a filter-and-froward relaying scheme, whereas in this paper, the relay nodes use simple AF relaying protocol. Moreover, the authors of [17] study the diversity of linear post-channel equalization schemes, such as zero forcing (ZF) and MMSE receivers, whereas in our work, we focus on obtaining the MSE-constrained power-optimal values of linear channel equalizers, the relays AF coefficients, and the transceivers' transmit powers with the aim to show the relationship between these values and rate-constrained power-optimal values of the same parameters. The goal in [18] is to develop a channel and timing offset estimation method that can be used to re-synchronization the relays, whereas we herein do not re-synchronize the relays but employ post-channel equalization to suppress the ISI at the two transceivers induced by the relays. Unlike [19], which relies on circularly shifting (cyclic delay) of the CP-free signal at the relays, we assume AF relaying without any cyclic delay for the sake of relay simplicity. The relaying scheme considered in [20] relies on a decode-and-forward protocol while the relaying scheme in this paper is an AF technique. The networks considered in [6], [21], [22] rely on OFDM to diagonalize the end-to-end channel (thereby eliminating ISI), while the scheme considered here uses single-carrier transmission and linear post-channel equalization to suppress ISI. In [7], [8], [25], the design approach is to minimize the sum-MSE at the two transceivers subject to a total power budget, while the results in this paper pertains to

minimizing the total power subject to MSE constraints. Also [8], [25] rely on pre-channel equalization and joint pre- and post-channel equalization, respectively, while our study here focuses only on post-channel equalization. The design approach in [23] is sum-rate maximization method subject to a total power budget, while we consider a total power minimization technique subject to MSE constraints. In [24], the design approach is total power minimization subject to rate constraints without assuming any type of equalizations, while here, assuming linear post-channel equalization, we study a total power minimization approach subject to MSE constraints.

Organization: Section II reviews the system and data model as well as the noise model, while presenting the new formulation for the end-to-end CIR. The MSE region characterization is presented in Section III. The MSE-constrained total power minimization problem is considered in Section IV, where the solution to this problem is carefully derived. The equivalence of the MSE-based power minimization approach and the rate-constrained power minimization technique is presented in Section V. The power-constrained sum-MSE minimization problem is studied in Section VI. Numerical results are presented in Section VII The paper is concluded in Section VIII.

Notations: We use bold upper- and lower-case letters to denote matrices and vectors, respectively. $E\{\cdot\}$ and $\mathrm{tr}(\cdot)$ represent the statistical expectation and the trace of a matrix, respectively. Transpose, Hermitian transpose, and complex conjugate operations are denoted as $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$, respectively. We use $\|\cdot\|$ to represent ℓ_2 norms. $\mathrm{diag}(\mathbf{w})$ yields a diagonal matrix whose diagonal entries are the elements of the vector $\mathbf{w}$. The notation $\mathrm{blkdiag}\{\mathbf{A}_n\}_{n\in\mathcal{N}}$ stands for a block diagonal matrix whose diagonal blocks belong to the matrix set $\{\mathbf{A}_n\}_{n\in\mathcal{N}}$, where $\mathcal{N}$ is an integer set. $\mathbf{I}_r$ is used to represent the $r\times r$ identity matrix, $\mathbf{0}_{m\times k}$ is an $m\times k$ matrix with all zero entries, and the (i,j)-th element of matrix $\mathbf{A}$ is denoted as $[\mathbf{A}]_{ij}$ or $\mathbf{A}(i,j)$. We also define $[a]^+ \triangleq \max\{0,a\}$. The principal eigenvalue and the normalized principal eigenvector of matrix $\mathbf{A}$ is denoted as $\lambda_{\max}\{\mathbf{A}\}$ and $\mathcal{P}\{\mathbf{A}\}$, respectively. Let the non-negative integers $\{l_n\}_{n=0}^{N-1}$ be such that $\sum_{n=0}^{N-1} l_n = L$ and define the $l_n\times L$ matrix $\mathbf{S}_n$, for $l_n\neq 0$, as

$$\mathbf{S}_n \triangleq [\mathbf{0}_{l_n\times i_n} \quad \mathbf{I}_{l_n} \quad \mathbf{0}_{l_n\times j_n}] \tag{1}$$

where we use the following definitions:

$$i_n \triangleq \sum_{k=0}^{n-1} l_k, \text{ for } n\geq 1 \text{ and } i_0 \triangleq 0 \tag{2}$$

$$j_n \triangleq \sum_{k=n+1}^{N-1} l_k, \text{ for } n\neq N-1, \text{ and } j_{N-1}\triangleq 0. \tag{3}$$

Then for any $L\times L$ matrix $\mathbf{A}$ and when $l_n\neq 0$, the matrix $\mathbf{S}_n\mathbf{A}\mathbf{S}_n^H$ is an $l_n\times l_n$ matrix extracted from matrix $\mathbf{A}$, where the matrix extraction starts from and includes the entry (i_n+1,i_n+1) of $\mathbf{A}$ and it ends at and includes the entry (i_n+l_n,i_n+l_n) of $\mathbf{A}$. That is, the upper-left and the lower-right corners of the extracted matrix are given by $\mathbf{A}(i_n+1,i_n+1)$ and $\mathbf{A}(i_n+l_n,i_n+l_n)$, respectively. Note that if $\mathbf{A}_1$ and $\mathbf{A}_2$ are two $L\times L$ block diagonal matrices whose diagonal block sizes are $\{l_n\times l_n\}_{n=0}^{N-1}$, then we can write

$$\mathbf{S}_n\mathbf{A}_1\mathbf{A}_2\mathbf{S}_n^H = \mathbf{S}_n\mathbf{A}_1\mathbf{S}_n^H\mathbf{S}_n\mathbf{A}_2\mathbf{S}_n^H \tag{4}$$

which is to say that, for $n=0,1,\ldots,N-1$, the $(n+1)$-th diagonal block (which is of size $l_n\times l_n$) of the product of $\mathbf{A}_1$ and $\mathbf{A}_2$ is the product of the $(n+1)$-th blocks (which are also of size $l_n\times l_n$) of $\mathbf{A}_1$ and $\mathbf{A}_2$. Note also that we can write

$$\mathbf{S}_n\mathbf{S}_n^H = \mathbf{I}_{l_n}. \tag{5}$$

We can also write

$$\mathbf{S}_n(\mathbf{A}^{-1})\mathbf{S}_n^H = (\mathbf{S}_n\mathbf{A}\mathbf{S}_n^H)^{-1} \tag{6}$$

which is to say that, for $n=0,1,\ldots,N-1$, the $(n+1)$-th diagonal block of the inverse of the $L\times L$ block diagonal matrix $\mathbf{A}$, whose diagonal block sizes are $\{l_n\times l_n\}_{n=0}^{N-1}$, is equal to the inverse of the $(n+1)$-th diagonal block of matrix $\mathbf{A}$, which is of size $l_n\times l_n$.

II. Preliminaries

Multipath Relay Channel: We consider a two-way relay-assisted communication between two single-antenna transceivers.[1] In the absence of a direct link between the two transceivers, the information exchange is materialized through a set of single-antenna AF relays. This implies that each information symbol transmitted by one of the transceivers travels through different relaying paths, and thus, the same symbol arrives at the other transceiver with different delays. Indeed we herein assume that different relaying paths have different propagation delays, implying that the end-to-end channel is time-dispersive. Such a time-dispersive behavior of the end-to-end channel causes different attenuated replicas of the signal transmitted by one transceiver arrive at the other transceiver with different delays. In other words, *this asynchronous relay channel* is a multipath link. Thus, at sufficiently high data rates, the inter-symbol interference (ISI) is inevitable in the received signal. In a block transmission and reception scheme, the ISI leads to inter-block-interference (IBI) and intra-block interference, which can affect the performance of the system. The IBI can be tackled using cyclic prefix (CP) insertion and removal at the transmitter and receiver sides, respectively. Indeed inserting a long enough CP ensures a sufficient guard time between two consecutive blocks.[2] In order to combat the intra-block interference, we use linear channel equalizers at the receiver front-ends of the transceivers. Fig. 1 depicts the considered two-way AF relay network with ISI inducing end-to-end channel. In this figure, CP insertion and removal operations are employed to eliminate IBI. Using such a communication setting, our goal is, to first find the achievable MSE region for a single-carrier asynchronous bi-directional (two-way) AF relay network under a total transmit budget, and then, use this region to solve the problem of total power minimization under two constraints on

[1]The scheme we study in this paper can be used for multiple peer-to-peer two-way relay-assisted communication between multiple transceiver pairs, provided that each transceiver pair uses a dedicated time-frequency resource block.

[2]Using CP will convert the linear convolution of the finite impulse response channel to the circular convolution.

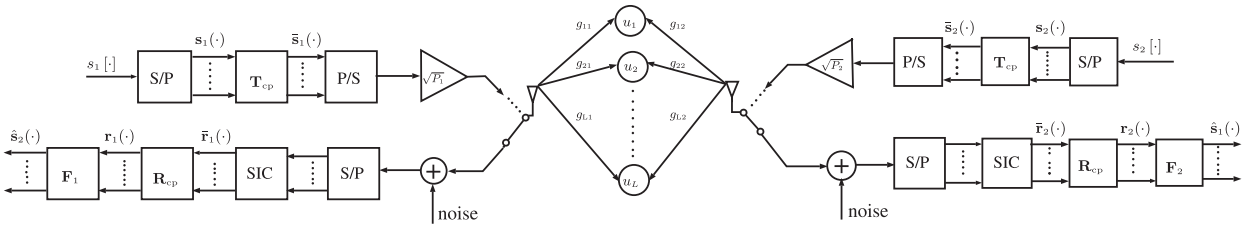

Fig. 1. System block diagram.

the mean squared error (MSE) at the two transceivers. We prove rigorously that this power minimization leads to the very same solution that minimizes the total power under two constraints on the data rates at the two transceivers [24], if the MSE thresholds in the MSE-constrained total power minimization and the rate thresholds in the rate-constrained power minimization technique satisfy a logarithmic (or an exponential) relationship. In our study, we consider linear post-channel equalization and aim to characterize the frontier of the MSE region through optimally obtaining the relay beamforming weights, the transceivers' transmit powers, and the post-channel equalizers. Doing so, we implicitly tackle the intra-block-interference. In the sequel, we present the end-to-end channel model and the received noise model.

Transmitted Signal Modeling: As shown in Fig. 1, at each transceiver, the sequences of the information symbols pass through a serial-to-parallel conversion block, denoted as "S/P," and are thus converted into blocks of $N_{\rm s}$ symbols. The i-th block of such information symbols at Transceiver q is represented as the following $N_{\rm s} \times 1$ vector:

$$\mathbf{s}_q(i) = [s_q[iN_{\rm s}]\ s_q[iN_{\rm s}+1]\ \cdots\ s_q[iN_{\rm s}+(N_{\rm s}-1)]]^T \quad (7)$$

where $s_q[iN_s + k]$ is the k-th unit-norm ($E\{|s_q[\cdot]|^2\} = 1$) symbol transmitted by Transceiver q at the i-th transmission block. The end-to-end channel is time-dispersive, and thus, results in IBI between adjacent symbol blocks. Adding a CP to each block can provide a guard interval between adjacent symbol blocks, thereby avoiding IBI. The CP is added to each block via multiplying that block by matrix $\mathbf{T}_{\rm cp} \triangleq [\mathbf{I}_{\rm cp}^T\ \mathbf{I}_{N_{\rm s}}^T]^T$. Here, $\mathbf{I}_{\rm cp}$ is a matrix consisting of the last $N_{\rm cp}$ rows of the identity matrix $\mathbf{I}_{N_{\rm s}}$, where $N_{\rm cp}$ is the length of the CP. It is worth mentioning that $N_{\rm cp}$ should be chosen at least equal to the length of the end-to-end multi-path CIR. Using $N_{\rm t}$ to represent the total length of each transmitted block (i.e., $N_{\rm t} = N_{\rm s} + N_{\rm cp}$), we can denote the output vector of the CP insertion block at Transceiver q as $\bar{\mathbf{s}}_q(i)$ and write it as

$$\begin{aligned}\bar{\mathbf{s}}_q(i) &\triangleq \mathbf{T}_{\rm cp}\mathbf{s}_q(i)\\ &= [s_q[(i+1)N_{\rm s}-N_{\rm CP}]\ \cdots\ s_q[(i+1)N_{\rm s}-1]\ s_q[iN_{\rm s}]\\ &\quad\cdots\ \ s_q[(i+1)N_{\rm s}-1]]^T. \end{aligned} \quad (8)$$

These blocks of signals then pass through a parallel-to-serial conversion block, denoted as "P/S," which converts these blocks to serial signals. These serial signals are then multiplied by $\sqrt{P_q}$ and are transmitted over the Transceiver's antenna. Here, P_q denotes the transmit power of Transceiver q.

Channel Modeling: We now model the end-to-end channels as linear time invariant (LTI) systems. To do so, we use $h_{pq}[\cdot]$ to denote the impulse response of the channel between Transceivers p and q for $p, q = 1, 2$. Note that $h_{pp}[\cdot]$ represents the channel associated with the self-interference at Transceiver p. As shown in Fig. 1, we use self-interference canceler block to remove the effect of such channels. Let $T_{\rm s}$ denote the symbol time duration, while τ_l^{pq} stands for the propagation delay that the signal undergoes when traveling from Transceiver p to Transceiver q through the l-th relay. Assuming different propagation delays for different relaying paths between the two transceivers, we can view the end-to-end CIR between Transceivers p and q as a multi-tap channel. Doing so allows us to write the Tap n of the discrete-time equivalent impulse response $h_{pq}[\cdot]$ as

$$h_{pq}[n] = \sum_{l=1}^{L} u_l g_{lp} g_{lq} \delta[n - m_l^{pq}], \quad \text{for } p, q \in \{1,2\},\ 0 \le n \le N-1, \quad (9)$$

where we use the following definitions: $\delta[n]$ is the discrete-time unit impulse response, i.e., $\delta[n] = 1$, for $n = 0$; and $\delta[n] = 0$, for $n \neq 0$; u_l is the complex weight of the l-th relay; g_{lp} is a complex coefficient representing the *reciprocal* frequency-flat channel between Transceiver p and the l-th relay; the integer m_l^{pq} represents the discrete-time propagation delay of the path between Transceivers p and q which goes through the l-th relay and satisfies $(m_l^{pq} - 1)T_{\rm s} < \tau_l^{pq} \le m_l^{pq} T_{\rm s}$; the integer N represents maximum of the lengths of the equivalent discrete-time multi-tap end-to-end CIRs $h_{pq}[\cdot]$, for $p, q = 1, 2$; that is,

$$N = 1 + \max_{p,q\in\{1,2\}, 1\le l\le L} \lceil \tau_l^{pq}/T_{\rm s} \rceil.$$

Note that $N_{\rm cp}$ should be chosen at least equal to N. To ensure this, we choose $N_{\rm cp} = N$. Let us index the relays, without loss of generality, in an ascending order based on the relaying path propagation delays they impose on the signal traveling from one transceiver to the other one. That is, if $l < \tilde{l}$, then $m_l^{12} \le m_{\tilde{l}}^{12}$. Doing so, we can index the relays in such a way that the first l_0 relays contribute to the first tap ($n = 0$) of the end-to-end CIR (i.e., $h_{12}[\cdot]$ which is equal[3] to $h_{21}[\cdot]$ due to the reciprocity of relay-channel links), the next l_1 relays contribute to the second tap ($n = 1$) of the end-to-end CIR and so on. That is, l_n is the number of relays which contribute to the $(n+1)$-th tap of the end-to-end CIR, for $n = 0, 1, 2, \ldots, N-1$. Note that

[3]We define $h[\cdot] \triangleq h_{21}[\cdot] = h_{12}[\cdot]$.

$\sum_{n=0}^{N-1} l_n = L$, as L stands for the total number of the relays. It is observed from (9) that the CIR depends on the relay weights and that *each relay contributes only to one tap of the end-to-end CIR*. In order to show this dependency explicitly, let us define the relay weight vector $\mathbf{u} \triangleq [u_1 \;\; u_2 \;\; \cdots \;\; u_L]^T$ and represent the vector of the multi-tap channel coefficients as

$$\mathbf{h}_{pq}(\mathbf{u}) \triangleq [h_{pq}[0] \;\; h_{pq}[1] \;\; \cdots \;\; h_{pq}[N-1]]^T. \tag{10}$$

Note that due to the reciprocity of the relay-transceiver links, we have $\mathbf{h}_{pq}(\mathbf{u}) = \mathbf{h}_{qp}(\mathbf{u})$, and hence, we hereafter drop the subscripts pq and use $\mathbf{h}(\mathbf{u})$ to represent the vector of the end-to-end CIR taps. Hence, based on (10), the vector of the end-to-end CIR taps is written as

$$\mathbf{h}(\mathbf{u}) = \mathbf{C}\mathbf{u}. \tag{11}$$

Here, the $N \times L$ matrix $\mathbf{C}$ is defined as

$$\mathbf{C} \triangleq [\mathbf{c}_1 \; \mathbf{c}_2 \; \cdots \; \mathbf{c}_N]^T \tag{12}$$

where, for $n \in \{0, \ldots, N-1\}$, the $1 \times L$ vector $\mathbf{c}_{n+1}^T$ is the $(n+1)$-th row of $\mathbf{C}$. Note that if no relay contributes to the $(n+1)$-th tap of the end-to-end CIR, then $\mathbf{c}_{n+1}^T = \mathbf{0}_{1\times L}$, otherwise

$$\mathbf{c}_{n+1}^T = [\mathbf{0}_{1\times i_n} \; \mathbf{a}_{n+1}^T \; \mathbf{0}_{1\times j_n}]. \tag{13}$$

Here, i_n and j_n are defined in (2) and (3), respectively. With $i_0 = 0$, for $\mathbf{c}_1$ and $j_{N-1} = 0$, for $\mathbf{c}_N$, we can write $\mathbf{c}_1^T = [\mathbf{a}_1^T \; \mathbf{0}_{1\times j_0}]$ and $\mathbf{c}_N^T = [\mathbf{0}_{1\times i_{(N-1)}} \; \mathbf{a}_N^T]$. The zero entries in $\mathbf{c}_{n+1}^T$ correspond to those relays which do not contribute to the $(n+1)$-th tap of the end-to-end CIR. Moreover, $\mathbf{a}_{n+1}$ is an $l_n \times 1$ vector of the products of the coefficients of the channels between the two transceivers and those relays which contribute to the $(n+1)$-th tap of the end-to-end CIR. More specifically, the m-th entry of $\mathbf{a}_{n+1}$ is given as $a_{(n+1),m} = g_{(m+i_n),1}\, g_{(m+i_n),2}$, where $m \in \{1, 2, \ldots, l_n\}$ and $n \in \{0, 1, \ldots, N-1\}$, only when $l_n \neq 0$. If $l_n = 0$, then $\mathbf{a}_{n+1}$ will be an empty vector. Using the above explanations, we can express $\mathbf{C}$ as in (14).[4]

Note that if no relay contributes to the $(n+1)$-th tap of $h[\cdot]$, i.e., if $l_n = 0$, then $h[n] = 0$, the $l_n \times 1$ vector $\mathbf{a}_{n+1}$ will be an empty vector, and hence, the $(n+1)$-th column block of $\mathbf{C}$ in (14) shown at the bottom of this page, will disappear, and the $(n+1)$-th row of $\mathbf{C}$ will be zero. In what follows, to deal with the indices of the taps that can be non-zero, we use the following definition: $\mathcal{N} \triangleq \{n | l_n \neq 0\}$ which represents a set consisting solely of the tap indices that can be non-zero. In other words, $\mathcal{N}$ is the set of the indices of those CIR taps to which at least one relay contributes. Note that the $L \times 1$ vector $\mathbf{u}$ can be partitioned into sub-vectors $\{\mathbf{u}_n\}_{n\in\mathcal{N}}$, where $\mathbf{u}_n$ is an $l_n \times 1$ vector of the beamforming weights of those relays which contribute to the $(n+1)$-th tap of the end-to-end CIR. To clarify our definitions, let us take a look at the following example for $\mathbf{C}$:

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ g_{11}g_{12} & g_{21}g_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_{31}g_{32} & g_{41}g_{42} & g_{51}g_{52} & 0 \\ 0 & 0 & 0 & 0 & 0 & g_{61}g_{62} \end{bmatrix} \tag{15}$$

In this 6-relay example, Relays 1 and 2 contribute to Tap $n = 1$ of the end-to-end CIR $h[\cdot]$, Relays 3, 4, and 5 contribute to Tap $n = 3$ of the end-to-end CIR $h[\cdot]$, and Relay 6 contributes to Tap $n = 4$ of the end-to-end CIR $h[\cdot]$. It can be observed that since no relay contributes to the first or to the third taps (i.e., Tap $n = 0$ and Tap $n = 2$) of the end-to-end CIR $h[\cdot]$, we have that $l_0 = l_2 = 0$, and thus, Rows 1 and 3 of $\mathbf{C}$ are zero. As a result, the set of the non-zero tap indices of the end-to-end CIR $h[\cdot]$ consists only of Taps 1, 3, and 4, that is $\mathcal{N} = \{1, 3, 4\}$, with $l_1 = 2$, $l_3 = 3$, and $l_4 = 1$. Moreover, the beamforming weight vector $\mathbf{u} = [w_1 \; w_2 \; w_3 \; w_4 \; w_5 \; w_6]^T$ can be partitioned into three sub-vectors: $\mathbf{u}_1 = [w_1 \; w_2]^T$, $\mathbf{u}_3 = [w_3 \; w_4 \; w_5]^T$, and $\mathbf{u}_4 = [w_6]^T$, that is $\mathbf{u} = [\mathbf{u}_1^T \; \mathbf{u}_3^T \; \mathbf{u}_4^T]^T$. Note that since no relay contributes to Taps $n = 0$ and $n = 2$ of the end-to-end CIR $h[\cdot]$, the corresponding sub-vectors are empty. That is, $\mathbf{u}_0 = \emptyset$ and $\mathbf{u}_2 = \emptyset$. Also, for this example, we can write $\mathbf{a}_1 = \emptyset$, $\mathbf{a}_3 = \emptyset$, $\mathbf{a}_2 = [g_{11}g_{12} \;\; g_{21}g_{22}]^T$, $\mathbf{a}_4 = [g_{31}g_{32} \;\; g_{41}g_{42} \;\; g_{51}g_{52}]^T$, $\mathbf{a}_5 = [g_{61}g_{62}]^T$. In general, we can write

$$\mathbf{a}_{n+1} = \begin{cases} \emptyset, & \text{if } n \notin \mathcal{N} \\ \begin{bmatrix} g_{(i_n+1),1}g_{(i_n+1),2} \\ g_{(i_n+2),1}g_{(i_n+2),2} \\ \vdots \\ g_{(i_n+l_n),1}g_{(i_n+l_n),2} \end{bmatrix}, & \text{if } n \in \mathcal{N}. \end{cases} \tag{16}$$

Based on the definition of $\mathbf{C}$ in (12) and its expansion in (14), we can write

$$\mathbf{C}^H\mathbf{C} = [\mathbf{c}_1^* \; \mathbf{c}_2^* \; \cdots \; \mathbf{c}_N^*] \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_N^T \end{bmatrix} = \sum_{n=0}^{N-1} \mathbf{c}_{n+1}^* \mathbf{c}_{n+1}^T$$

$$= \sum_{n\in\mathcal{N}} \mathbf{c}_{n+1}^* \mathbf{c}_{n+1}^T = \text{blkdiag}\{\mathbf{a}_{n+1}^* \mathbf{a}_{n+1}^T\}_{n\in\mathcal{N}} \tag{17}$$

[4]This novel expression of matrix $\mathbf{C}$ plays a key role in our derivations and is what makes our model formulation different from the model formulation in [7].

$$\begin{bmatrix} \mathbf{a}_1^T & \mathbf{0}_{1\times l_1} & \mathbf{0}_{1\times l_2} & \cdots & \mathbf{0}_{1\times l_{(N-1)}} \\ \mathbf{0}_{1\times l_0} & \mathbf{a}_2^T & \mathbf{0}_{1\times l_2} & \cdots & \mathbf{0}_{1\times l_{(N-1)}} \\ \mathbf{0}_{1\times l_0} & \mathbf{0}_{1\times l_1} & \mathbf{a}_3^T & \cdots & \mathbf{0}_{1\times l_{(N-1)}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \underbrace{\mathbf{0}_{1\times l_0}} & \underbrace{\mathbf{0}_{1\times l_1}} & \underbrace{\mathbf{0}_{1\times l_2}} & \cdots & \underbrace{\mathbf{a}_N^T} \end{bmatrix} \tag{14}$$

First Column Block | Second Column Block | Third Column Block | Last Column Block

where in the third equality, we use the fact that for $n \notin \mathcal{N}$, $\mathbf{c}_{n+1} = \mathbf{0}_{L\times 1}$, while the fourth equality follows from (13). The way we define matrix $\mathbf{C}$ reveals the block diagonal structure of $\mathbf{C}^H\mathbf{C}$ as in (17). This block diagonal structure of $\mathbf{C}^H\mathbf{C}$ plays an important role in allowing us to solve the problems that are of interest in this paper. Indeed, it is the revealing of this block diagonal structure of $\mathbf{C}^H\mathbf{C}$ that differentiates our channel model formulation here from the one presented in [7].

Using (1), we can also write

$$\mathbf{S}_n\mathbf{C}^H\mathbf{C}\mathbf{S}_n^H = \mathbf{a}_{n+1}^*\mathbf{a}_{n+1}^T, \text{ for } n \in \mathcal{N}. \tag{18}$$

For the example in (15), we can write

$$\mathbf{C}^H\mathbf{C} = \begin{bmatrix} \mathbf{a}_1^*\mathbf{a}_1^T & \mathbf{0}_{2\times 3} & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{3\times 2} & \mathbf{a}_3^*\mathbf{a}_3^T & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 3} & \mathbf{a}_4^*\mathbf{a}_4^T \end{bmatrix}$$

We later use (17) in the process of deriving the solution to the optimization problems of interest in this paper.

Received Noise Model: Let $z_l[\cdot]$ denote the temporally and spatially white noise at the l-th relay. This noise is assumed to be zero-mean with a variance of σ^2. The superposition of the amplified and delayed versions of the relay noises at Transceiver q is represented as the noise process $\chi_q[\cdot]$ and can be written as

$$\chi_q[n] = \sum_{l=1}^{L} u_l g_{lq} z_l[n - \hat{m}_{lq}] = \mathbf{z}_{n,q}^H \tilde{\mathbf{Q}}_q \mathbf{u},$$

where the following definitions are used $\mathbf{z}_{n,q} \triangleq [z_1[n - \hat{m}_{1q}] z_2[n - \hat{m}_{2q}] \cdots z_L[n - \hat{m}_{Lq}]]^T$ and $\tilde{\mathbf{Q}}_q \triangleq \text{diag}\{g_{lq}\}_{l=1}^L$, while $\hat{m}_{lq}$ represents the sample delay between the l-th relay and Transceiver q. Indeed, $\hat{m}_{lq}$ is an integer value which satisfies $\frac{\hat{\tau}_{lq}}{T_\text{s}} \leq \hat{m}_{lq} < \frac{\hat{\tau}_{lq}}{T_\text{s}} + 1$, and $\hat{\tau}_{lq}$ denotes the propagation delay between the l-th relay and Transceiver q. Thus, the noise vector received at Transceiver q that contaminates the i-th block of information symbols is given as

$$\boldsymbol{\upsilon}_q(i) = \boldsymbol{\chi}_q(i) + \hat{\boldsymbol{\upsilon}}_q(i) = \mathbf{Z}_q(i)\tilde{\mathbf{Q}}_q\mathbf{u} + \hat{\boldsymbol{\upsilon}}_q(i),$$

where we used the following definitions: $\boldsymbol{\upsilon}_q(i) \triangleq [\upsilon_q[iN_\text{t}] \upsilon_q[iN_\text{t}+1] \cdots \upsilon_q[iN_\text{t} + (N_\text{t}-1)]]^T$, $\boldsymbol{\chi}_q(i) \triangleq [\chi_q[iN_\text{t}]\ \chi_q[iN_\text{t}+1] \cdots \chi_q[iN_\text{t} + (N_\text{t}-1)]]^T$, $\hat{\boldsymbol{\upsilon}}_q(i) \triangleq [\hat{\upsilon}_q[iN_\text{t}] \hat{\upsilon}_q[iN_\text{t}+1] \cdots \hat{\upsilon}_q[iN_\text{t} + (N_\text{t}-1)]]^T$, $\upsilon_q[n] = \chi_q[n] + \hat{\upsilon}_q[n]$. Moreover, $\mathbf{Z}_q(i) \triangleq [\mathbf{z}_{iN_\text{t},q}\ \mathbf{z}_{iN_\text{t}+1,q} \cdots \mathbf{z}_{iN_\text{t}+(N_\text{t}-1),q}]^T$ is defined as an $N_\text{t} \times L$ matrix in which the l-th column represents the l-th relay noise delayed by the corresponding delay (i.e., by $\hat{m}_{lq}$ samples) at the i-th received block. Also, $\hat{\upsilon}_q[n]$ is the n-th sample of the receiver noise at Transceiver q.

Received Signal Model: At the receiver side of the transceivers, the received signal first passes through the serial-to-parallel conversion block, denoted as "S/P". As a result, the sequences of the received signals are transformed into blocks of signals. Afterwards, the self-interference cancellation is applied to the so-obtained blocks of signals. Using self-interference cancellation, denoted as "SIC," each transceiver can remove its own transmitted signal which has been relayed back to that transceiver. As explained before, due to the time dispersive nature of the end-to-end channel, IBI occurs between two consecutive blocks of received signals. In order to remove the induced IBI, $\bar{\mathbf{r}}_q(i)$ is passed through the CP removal block, represented by matrix $\mathbf{R}_\text{cp} \triangleq [\mathbf{0}_{N_\text{s}\times N_\text{cp}}\ \ \mathbf{I}_{N_\text{s}}]$. Therefore, the signal vector $\mathbf{r}_q(i)$ at the output of the CP removal block can be written as [26]

$$\mathbf{r}_q(i) \triangleq \mathbf{R}_\text{cp}\bar{\mathbf{r}}_q(i) = \sqrt{P_p}\tilde{\mathbf{H}}(\mathbf{u})\mathbf{s}_p(i) + \check{\boldsymbol{\upsilon}}_q(i). \tag{19}$$

Here, we define $\check{\boldsymbol{\upsilon}}_q(i) \triangleq \mathbf{R}_\text{cp}\boldsymbol{\upsilon}_q(i)$ and $\tilde{\mathbf{H}}(\mathbf{u})$ is an $N_\text{s} \times N_\text{s}$ circulant matrix whose (i,j)-th entry is given by $\tilde{h}[(i-j) \bmod N_\text{s}]$, where $\tilde{h}[n] = h[n]$, for $n = 0, 1, \ldots, N-1$; and $\tilde{h}[n] = 0$, for $n = N, N+1, \ldots, N_\text{s}-1$. Indeed, $\tilde{h}[\cdot]$ is the zero-padded version of $h[\cdot]$ which is obtained by zero-padding $h[\cdot]$ with $N_\text{s} - N$ zeros. It is worth mentioning that the length of each block (i.e., N_s), must be at least equal to the length of the end-to-end CIR N. In order to suppress or eliminate the residual ISI which still exists in each block of symbols, we multiply the received vector $\mathbf{r}_q(i)$ by the $N_\text{s} \times N_\text{s}$ matrix $\mathbf{F}_q$. Doing so, we obtain a $N_\text{s} \times 1$ linear estimate of the symbol vector transmitted by Transceiver p, denoted as $\hat{\mathbf{s}}_p(i)$, which can be written as

$$\hat{\mathbf{s}}_p(i) \triangleq \mathbf{F}_q\mathbf{r}_q(i) = \sqrt{P_p}\mathbf{F}_q\tilde{\mathbf{H}}(\mathbf{u})\mathbf{s}_p(i) + \mathbf{F}_q\check{\boldsymbol{\upsilon}}_q(i). \tag{20}$$

Note that matrices $\{\mathbf{F}_q\}_{q=1}^2$ along with $\mathbf{u}$ and $\{P_q\}_{q=1}^2$ are to be determined optimally.

Calculation of the Total Transmit Power: We now express the average total power, consumed in the entire network, in terms of the relay beamforming vector $\mathbf{u}$ and the transceivers' transmit powers P_1 and P_2. Let us define the $N_\text{t} \times 1$ vector $\mathbf{y}_l(i)$ as the i-th block of the signals relayed by the l-th relay. Doing so, we can write

$$\mathbf{y}_l(i) = u_l\left(\sqrt{P_1}g_{l1}\bar{\mathbf{s}}_1(i) + \sqrt{P_2}g_{l2}\bar{\mathbf{s}}_2(i) + \mathbf{z}_l(i)\right). \tag{21}$$

Here, $\mathbf{z}_l(i) \triangleq [z_l(iN_\text{t})\, z_l(iN_\text{t}+1) \cdots z_l(iN_\text{t}+N_\text{t}-1)]^T$ is defined as the i-th block of measurement noise at the l-th relay, and each of its entries is assumed to be drawn from an i.i.d. random process with zero-mean and a variance of σ^2. Furthermore, we assume that the vectors $\bar{\mathbf{s}}_1(i)$, $\bar{\mathbf{s}}_2(i)$ and $\mathbf{z}_l(i)$ are mutually independent stationary random processes. We can use (21) to show that, the transmit power of the l-th relay is given by[5] $|u_l|^2(P_1|g_{l1}|^2 + P_2|g_{l2}|^2 + \sigma^2)$. Therefore, the total transmit power consumed in the entire network can be expressed as

$$\begin{aligned} P_\text{t}(P_1, P_2, \mathbf{u}) &\triangleq P_1 + P_2 + \sum_{l=1}^{L}|u_l|^2(|g_{l1}|^2P_1 + |g_{l2}|^2P_2 + \sigma^2) \\ &= P_1h_1(\mathbf{u}) + P_2h_2(\mathbf{u}) + \sigma^2\|\mathbf{u}\|^2 \end{aligned} \tag{22}$$

where we define $h_q(\mathbf{u}) \triangleq \mathbf{u}^H\mathbf{Q}_q\mathbf{u} + 1$, $\mathbf{Q}_q \triangleq \text{diag}\{|g_{lq}|^2\}_{l=1}^L$, for $q = 1, 2$. In the next section, we use the data and channel models presented in this section to formulate our optimization problem for characterizing the MSE region, and then, show how to solve this problem.

[5]Note that the derivation of (22) relies on the assumption that the communication time frame is much longer than the maximum of the time differences between arrivals of transceiver signals at the relays [6].

III. The MSE Region

Problem Formulation: In order to analytically find the achievable MSE region, we solve the following optimization problem:

$$\min_{P_1\geq 0, P_2\geq 0, \mathbf{u}} \ \min_{\mathbf{F}_1,\mathbf{F}_2} \ \mathrm{MSE}_1(\mathbf{u},\mathbf{F}_1,P_2)$$
$$\text{s.t.} \quad \mathrm{MSE}_2(\mathbf{u},\mathbf{F}_2,P_1)=\varepsilon_2, \ \ P_{\mathrm{t}}(P_1,P_2,\mathbf{u})\leq P, \quad (23)$$

where ε_2 is the MSE at Transceiver 2, P is the total available power budget and $\mathrm{MSE}_q(\mathbf{u},\mathbf{F}_q,P_p)$ is the normalized mean squared error of symbol estimates at Transceiver q and is defined as

$$\mathrm{MSE}_q(\mathbf{u},\mathbf{F}_q,P_p)\triangleq\frac{1}{N_{\mathrm{s}}}\mathrm{E}\left\{\|\mathbf{e}_q(i)\|^2\right\},$$
$$\text{for } p,q\in\{1,2\}, p\neq q. \quad (24)$$

Here, $\mathbf{e}_q(i)\triangleq\hat{\mathbf{s}}_p(i)-\mathbf{s}_p(i)$ is the $N_{\mathrm{s}}\times 1$ vector of the symbol estimation errors at Transceiver q. Solving (23) leads us to find the range of the MSEs at the end-nodes under a total power budget. In the rest of the paper, we use the convention that $q=2$, when $p=1$; and $q=1$, when $p=2$. Note that solving the optimization problem is important for three reasons: 1) Solving (23) allows us to characterize the range of un-coded probabilities of symbol error at the two transceivers under a given total power budget. As the un-coded probability of error at each transceiver is a monotonically non-decreasing function of the MSE at that transceiver, solving (23) allows us to find out what is the lowest un-coded probability of error at Transceiver 1, when the un-coded probability of error at Transceiver 2 is kept at a certain level (determined by ε_2), under a total power budget P. Indeed, under a given total power budget, the boundary of the achievable MSE region determines the boundary of the achievable region of the un-coded probabilities of symbol error at the two transceivers. The latter boundary allows us to design strong enough error correction codes. 2) As shown in the next section, the solution to (23) allows us to solve the MSE-constrained total power minimization problem for the considered scheme. Solving this power minimization problem in turn allows us to show the relationship between the solution to this problem and the solution to the rate-constrained power minimization problem. 3) Last but not least, having the boundary of the achievable MSE region provides an easy way to solve the problem of sum-MSE minimization under a total power budget. Indeed, the minimization of the sum-MSE occurs on the boundary of the achievable MSE region. This interesting result is presented in Section VI.

One can easily show that the MSE in (24) can be written as

$$\mathrm{MSE}_q(\mathbf{u},\mathbf{F}_q,P_p)=N_{\mathrm{s}}{}^{-1}\mathrm{tr}\left\{\mathbf{F}_q\mathbf{R}_q(\mathbf{u},P_p)\mathbf{F}_q^H\right\}+1$$
$$-N_{\mathrm{s}}{}^{-1}\sqrt{P_p}\mathrm{tr}\left\{\mathbf{F}_q\tilde{\mathbf{H}}(\mathbf{u})+\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{F}_q^H\right\}, \quad (25)$$

where the correlation matrix of the received signal $\mathbf{R}_q(P_p,\mathbf{u})$ is given as

$$\mathbf{R}_q(\mathbf{u},P_p)\triangleq\mathrm{E}\left\{\mathbf{r}_q(i)\mathbf{r}_q^H(i)\right\}$$
$$=P_p\tilde{\mathbf{H}}(\mathbf{u})\tilde{\mathbf{H}}^H(\mathbf{u})+\sigma^2h_q(\mathbf{u})\mathbf{I}_{N_{\mathrm{s}}}. \quad (26)$$

In the optimization problem (23), given P and ε_2, we aim to jointly optimize the post-channel equalization matrices, the relay beamforming weights, and the power allocation schemes at the two transceivers such that the MSE_1 is minimized. Solving (23) allows us to find the smallest possible value of the MSE at Transceiver 1, for given P and ε_2, thereby characterizing the MSE region in terms of P and ε_2. It is also worth stressing that the applications and the advantages of using a total power constraint in the context of asynchronous relay networks have been well justified in the literature. We avoid repeating these justifications for a total transmit power constraint and refer our reader to [6]–[8], [21], [23], [25] for these justifications in the context under consideration.

Optimal Channel Equalizer Design: For fixed $\mathbf{u}$, P_1, and P_2, let us consider the inner minimization in (23) as

$$\eta(\mathbf{u},P_1,P_2)=\min_{\mathbf{F}_1,\mathbf{F}_2}\ \mathrm{MSE}_1(\mathbf{u},\mathbf{F}_1,P_2)$$
$$\text{s.t.} \quad \mathrm{MSE}_2(\mathbf{u},\mathbf{F}_2,P_1)=\varepsilon_2\,. \quad (27)$$

To solve the inner minimization in (27), we can use Lagrange multipliers method to find the optimal solution to this minimization problem. To do so, we can write the Lagrangian function as

$$L(\mathbf{F}_1,\mathbf{F}_2,\nu)\triangleq\ \mathrm{MSE}_1(\mathbf{u},\mathbf{F}_1,P_2)$$
$$+\nu\left(\mathrm{MSE}_2(\mathbf{u},\mathbf{F}_2,P_1)-\varepsilon_2\right), \quad (28)$$

where ν is the Lagrange multiplier. The optimal $\mathbf{F}_1$ and $\mathbf{F}_2$ (denoted as $\mathbf{F}_1^{\mathrm{o}}$ and $\mathbf{F}_2^{\mathrm{o}}$, respectively) can be obtained by differentiating the Lagrangian function in (28) with respect to $\mathbf{F}_1$ and $\mathbf{F}_2$ and equating these derivatives to zero, and then, solving for $\mathbf{F}_1$ and $\mathbf{F}_2$. Doing so leads us to the following two Wiener filter solutions:

$$\mathbf{F}_1^{\mathrm{o}}=\sqrt{P_2}\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_1^{-1}(\mathbf{u},P_2),\mathbf{F}_2^{\mathrm{o}}=\sqrt{P_1}\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_2^{-1}(\mathbf{u},P_1). \quad (29)$$

Note that $\mathbf{F}_2^{\mathrm{o}}$ must satisfy the constraint in (27). Inserting $\mathbf{F}_2^{\mathrm{o}}$ in the constraint in (27) results in a constraint on P_1 and $\mathbf{u}$ that has to be taken into account when the outer minimization in (23) is solved. The details of solving the outer minimization is provided in the sequel.

Optimal Joint Relay Beamforming and Power Control: Using (29), we show in Appendix A that $\eta(\mathbf{u},P_1,P_2)$ can be written as

$$\eta(\mathbf{u},P_1,P_2)=N_{\mathrm{s}}^{-1}\sum_{k=1}^{N_{\mathrm{s}}}\left[\frac{P_2N_{\mathrm{s}}\alpha_k(\mathbf{u})}{\sigma^2h_1(\mathbf{u})}+1\right]^{-1} \quad (30)$$

where we use the following definitions: $\alpha_k(\mathbf{u})\triangleq|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2$, $\mathbf{f}_k\triangleq\frac{1}{\sqrt{N_{\mathrm{s}}}}[1\ e^{j\frac{2\pi(k-1)}{N_{\mathrm{s}}}}\ \cdots\ e^{j\frac{2\pi(N-1)(k-1)}{N_{\mathrm{s}}}}]^T$, for $k=1,2,\ldots,N_{\mathrm{s}}$. It is worth stressing that $\alpha_k(\mathbf{u})$ is the magnitude-squared of the frequency response of the end-to-end CIR $h[\cdot]$ at a normalized, frequency of $(k-1)/N_{\mathrm{s}}$. Similarly, one can use (29) and rewrite the first constraint in (23) as

$$N_{\mathrm{s}}^{-1}\sum_{k=1}^{N_{\mathrm{s}}}\left[\frac{P_1N_{\mathrm{s}}|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2}{\sigma^2h_2(\mathbf{u})}+1\right]^{-1}=\varepsilon_2\,. \quad (31)$$

Using (30) and (31), we can simplify the optimization problem in (23) as

$$\min_{P_1\geq 0,P_2\geq 0,\mathbf{u}} \; N_{\mathrm{s}}^{-1}\sum_{k=1}^{N_{\mathrm{s}}}\left[\frac{P_2N_{\mathrm{s}}\alpha_k(\mathbf{u})}{\sigma^2h_1(\mathbf{u})}+1\right]^{-1}$$
$$\text{s.t. } N_{\mathrm{s}}^{-1}\sum_{k=1}^{N_{\mathrm{s}}}\left[\frac{P_1N_{\mathrm{s}}\alpha_k(\mathbf{u})}{\sigma^2h_2(\mathbf{u})}+1\right]^{-1}=\varepsilon_2$$
$$P_1h_1(\mathbf{u})+P_2h_2(\mathbf{u})+\sigma^2\|\mathbf{u}\|^2\leq P. \quad (32)$$

Note that the first constraint in (32) is feasible only when $\varepsilon_2\in(0,1]$. The reason is that the N_{s} terms under the summation in the left-hand side (LHS) of this constraint are nonnegative and less than 1 (see (31)). Similarly, the values of cost function in (32) are in the interval $(0,1]$. That is, the achievable MSE region is restricted to a subset of the square area $(0,1]\times(0,1]$. The optimization problem (32) amounts to designing jointly optimal relay beamforming and transceiver power allocation. To solve this problem, we first solve another optimization problem which provides a lower bound to the optimum value of the optimization problem (32). This new optimization problem is obtained the by first replacing the objective function in (32) with another function which is a lower bound to the objective function in (32) and by then relaxing the second constraint in (32). As a result, solving this new optimization problem yields a lower bound to the optimization problem (32). Once the new optimization problem is solved, we show that this lower bound is tight, meaning that the solution to this problem is feasible for the problem in (32), and that, for this solution, the objective function in (32) touches the lower bound.

Relaxation: Using the fact that the arithmetic mean of any set of positive real numbers is greater than or equal to their harmonic mean, we can write

$$\frac{1}{N_{\mathrm{s}}}\sum_{k=1}^{N_{\mathrm{s}}}\left[\frac{P_1N_{\mathrm{s}}\alpha_k(\mathbf{u})}{\sigma^2h_2(\mathbf{u})}+1\right]^{-1}\geq N_{\mathrm{s}}\left[\sum_{k=1}^{N_{\mathrm{s}}}\left(\frac{P_1N_{\mathrm{s}}\alpha_k(\mathbf{u})}{\sigma^2h_2(\mathbf{u})}+1\right)\right]^{-1}$$
$$=N_{\mathrm{s}}\left[\frac{P_1N_{\mathrm{s}}}{\sigma^2h_2(\mathbf{u})}\left(\sum_{k=1}^{N_{\mathrm{s}}}\alpha_k(\mathbf{u})\right)+N_{\mathrm{s}}\right]^{-1}$$
$$=N_{\mathrm{s}}\left[\frac{P_1N_{\mathrm{s}}\|\mathbf{Cu}\|^2}{\sigma^2h_2(\mathbf{u})}+N_{\mathrm{s}}\right]^{-1}=\left[\frac{P_1\|\mathbf{Cu}\|^2}{\sigma^2h_2(\mathbf{u})}+1\right]^{-1} \quad (33)$$

where in the second equality in (33), we used Parseval's theorem, i.e., $\sum_{k=1}^{N_{\mathrm{s}}}\alpha_k(\mathbf{u})=\sum_{k=1}^{N_{\mathrm{s}}}|\mathbf{f}_k^H\mathbf{Cu}|^2=\|\mathbf{Cu}\|^2$. Note that the inequality in (33) is satisfied with equality if and only if $\alpha_k(\mathbf{u})=\alpha_{\tilde{k}}(\mathbf{u})$ (or $|\mathbf{f}_k^H\mathbf{Cu}|=|\mathbf{f}_{\tilde{k}}^H\mathbf{Cu}|$) holds true, for $k,\tilde{k}\in\{1,2,\ldots,N_{\mathrm{s}}\}$. The right-hand side of (33) provides a lower bound for the left-hand side of the first constraint in (32). In a similar fashion, we can find a lower bound for the objective function in (32) and write

$$\frac{1}{N_{\mathrm{s}}}\sum_{k=1}^{N_{\mathrm{s}}}\left[\frac{P_2N_{\mathrm{s}}\alpha_k(\mathbf{u})}{\sigma^2h_1(\mathbf{u})}+1\right]^{-1}\geq\left[\frac{P_2\|\mathbf{Cu}\|^2}{\sigma^2h_1(\mathbf{u})}+1\right]^{-1} \quad (34)$$

Using (33) and (34), we can write a relaxed form of the minimization problem in (32) as

$$\min_{P_1\geq 0,P_2\geq 0,\mathbf{u}} \left[\frac{P_2\|\mathbf{Cu}\|^2}{\sigma^2h_1(\mathbf{u})}+1\right]^{-1}\text{s.t.}\left[\frac{P_1\|\mathbf{Cu}\|^2}{\sigma^2h_2(\mathbf{u})}+1\right]^{-1}\leq\varepsilon_2$$
$$\text{and } P_1h_1(\mathbf{u})+P_2h_2(\mathbf{u})+\sigma^2\|\mathbf{u}\|^2\leq P. \quad (35)$$

Doing so, we now deal with a relaxed problem with a "larger" feasible set. The solution to the relaxed problem in (35) yields a lower bound for the optimum value of the objective function of the original optimization problem in (32). The reason is that in light of (34), the cost function in (35) is a lower bound to the objective function in (32), and because of (33), the first constraint in (35) is a relaxed version of the first constraint in (32). We will later show that the solution to the relaxed problem in (35) satisfies the inequalities in (33) and (34) with equality, that is, we will show that at the optimum of (32), $|\mathbf{f}_k^H\mathbf{Cu}|=|\mathbf{f}_{\tilde{k}}^H\mathbf{Cu}|$ holds true, for $k,\tilde{k}\in\{1,2,\ldots,N_{\mathrm{s}}\}$. In other words, we later show that the solution to the relaxed problem in (35) is a solution to the original optimization problem in (32). To solve (35), we can easily show that, at the optimum, both constraints in the optimization problem (35) are satisfied with equality. As a result, we can use the first constraint in (35) to express P_1, in terms of $\mathbf{u}$, as

$$P_1=\frac{\sigma^2h_2(\mathbf{u})}{\|\mathbf{Cu}\|^2}\left(\frac{1}{\varepsilon_2}-1\right)\geq 0, \quad (36)$$

where $P_1\geq 0$ holds simply because we assumed that $\varepsilon_2\leq 1$ holds. Using (36) and the second constraint in (35), one can obtain a closed form for P_2, in terms of $\mathbf{u}$, as

$$P_2=\frac{P-\sigma^2\|\mathbf{u}\|^2}{h_2(\mathbf{u})}-\frac{\sigma^2h_1(\mathbf{u})}{\|\mathbf{Cu}\|^2}\left(\frac{1}{\varepsilon_2}-1\right)\geq 0 \quad (37)$$

Using (36) and (37), we can recast the problem in (35) as a maximization problem written as

$$\max_{\mathbf{u}} \quad \psi(\mathbf{u})-\frac{1}{\varepsilon_2}+2, \quad \text{s.t. to} \quad \psi(\mathbf{u})-\frac{1}{\varepsilon_2}+1>0 \quad (38)$$

where we define $\psi(\mathbf{u})\triangleq\dfrac{\|\mathbf{Cu}\|^2(P-\sigma^2\|\mathbf{u}\|^2)}{\sigma^2h_1(\mathbf{u})h_2(\mathbf{u})}$. Note that as $\varepsilon_2\leq 1$, the constraint in (38), requires that $\psi(\mathbf{u})\geq 0$ hold true. This in turn means that $\|\mathbf{u}\|^2\leq P$ must be ensured. Note that if $\max_{\mathbf{u}}\psi(\mathbf{u})<\frac{1}{\varepsilon_2}-1$, then the optimization problem (38) is not feasible. Hence, to solve the optimization problem (38), we only need to solve the following optimization problem:

$$\rho\triangleq\max_{\mathbf{u}}\psi(\mathbf{u}) \quad (39)$$

and verify whether the feasibility condition $\rho\geq\dfrac{1}{\varepsilon_2}-1$ holds true or not. If $\rho<\dfrac{1}{\varepsilon_2}-1$ holds true, then the relaxed lower bound problem is infeasible, so is the original problem. In this case, a maximum MSE of ε_2 for Transceiver 2 can not be guaranteed for the given total transmit power P. In other words, the total available transmit power is not sufficient to ensure that the MSE at Transceiver 2 is equal to ε_2. In the sequel, we solve (39).

Solving the Relaxed Lower Bound Problem: Let us rewrite the optimization problem (39) as

$$\max_{\mathbf{u}} \frac{\|\mathbf{C}\mathbf{u}\|^2(P-\sigma^2\|\mathbf{u}\|^2)}{\sigma^2 h_1(\mathbf{u})h_2(\mathbf{u})}. \tag{40}$$

Defining a new optimization variable $z \triangleq \dfrac{P-\sigma^2\|\mathbf{u}\|^2}{h_2(\mathbf{u})} = \dfrac{P-\sigma^2\|\mathbf{u}\|^2}{\mathbf{u}\mathbf{Q}_2\mathbf{u}+1}$, we write the optimization problem (40) as[6]

$$\begin{aligned} \max_{z\geq 0} \quad & z\left(\max_{\mathbf{u}} \frac{\mathbf{u}^H\mathbf{C}^H\mathbf{C}\mathbf{u}}{\mathbf{u}^H\mathbf{Q}_1\mathbf{u}+1}\right) \\ \text{s.t.} \quad & \mathbf{u}^H(z\mathbf{Q}_2+\sigma^2\mathbf{I}_L)\mathbf{u} \leq P-z \end{aligned} \tag{41}$$

where the constraint $z \geq 0$ follows from the condition $\|\mathbf{u}\|^2 \leq P$ which must hold, as explained earlier. Let us define $\tilde{\mathbf{u}}$ as an $L \times 1$ unit-norm ($\|\tilde{\mathbf{u}}\|^2 = 1$) vector and let p be a real positive scalar playing the role of the normalization factor such that

$$\mathbf{B}_1^{-1}(z)\mathbf{u} = \sqrt{p}\tilde{\mathbf{u}} \tag{42}$$

where we define the $L \times L$ diagonal matrix $\mathbf{B}_1(z)$ as $\mathbf{B}_1(z) \triangleq (z\mathbf{Q}_2+\sigma^2\mathbf{I}_L)^{-\frac{1}{2}}$. Doing so, we can rewrite the inner maximization problem in (41) as

$$\max_{p,\tilde{\mathbf{u}}} \frac{p\tilde{\mathbf{u}}^H\mathbf{B}_2(z)\tilde{\mathbf{u}}}{p\tilde{\mathbf{u}}^H\mathbf{B}_3(z)\tilde{\mathbf{u}}+1}, \quad \text{s.t. } \|\tilde{\mathbf{u}}\|^2 = 1,\ 0 \leq p \leq P-z \tag{43}$$

where we used the following definitions:

$$\mathbf{B}_2(z) \triangleq \mathbf{B}_1(z)\mathbf{C}^H\mathbf{C}\mathbf{B}_1(z) \tag{44}$$

$$\mathbf{B}_3(z) \triangleq \mathbf{B}_1(z)\mathbf{Q}_1\mathbf{B}_1(z) = \mathbf{B}_1^2(z)\mathbf{Q}_1 \tag{45}$$

The second constraint in (43) is derived from the constraint in (41) as we can write $0 \leq \mathbf{u}^H(z\mathbf{Q}_2\mathbf{u}+\sigma^2\mathbf{I}_L)\mathbf{u} = \mathbf{u}^H\mathbf{B}_1^{-2}(z)\mathbf{u} = p\tilde{\mathbf{u}}^H\tilde{\mathbf{u}} = p \leq P-z$. Note that $\mathbf{B}_2(z)$ is an $L \times L$ block diagonal matrix (because $\mathbf{B}_1(z)$ is an $L \times L$ block diagonal matrix and $\mathbf{C}^H\mathbf{C}$ is an $L \times L$ block diagonal matrix (see (17)). The sizes of the diagonal blocks of $\mathbf{B}_2(z)$ are $\{l_n \times l_n\}_{n\in\mathcal{N}}$. Also $\mathbf{B}_3(z)$ is an $L \times L$ diagonal matrix (because both $\mathbf{B}_1(z)$ and $\mathbf{Q}_1$ are $L \times L$ diagonal matrices). Since for any value of $\tilde{\mathbf{u}}$, the objective function in (43) is monotonically increasing with respect to p, the maximum value of the objective function is attained when $p = P-z$. Denoting the optimal value of p as $p^{\mathrm{o}} \triangleq P-z$, we can rewrite the maximization problem in (43) as

$$\max_{\tilde{\mathbf{u}}} \frac{p^{\mathrm{o}}\tilde{\mathbf{u}}^H\mathbf{B}_2(z)\tilde{\mathbf{u}}}{\tilde{\mathbf{u}}^H(\mathbf{I}_L+p^{\mathrm{o}}\mathbf{B}_3(z))\tilde{\mathbf{u}}}, \quad \text{subject to } \|\tilde{\mathbf{u}}\|^2 = 1 \tag{46}$$

which is the well-known Rayleigh quotient problem and has a closed-form solution for $\tilde{\mathbf{u}}$. To present this solution, let us define the $L \times L$ diagonal matrix $\mathbf{B}_4(z)$ as

$$\begin{aligned} \mathbf{B}_4(z) &\triangleq \mathbf{I}_L + p^{\mathrm{o}}\mathbf{B}_3(z) = \mathbf{I}_L + (P-z)\mathbf{B}_3(z) \\ &= \mathbf{I}_L + (P-z)\mathbf{B}_1(z)\mathbf{Q}_1\mathbf{B}_1(z) \end{aligned} \tag{47}$$

[6]Here, we drop the constant σ^2 in the denominator of the objective function without loss of optimality.

Denoting the largest eigenvalue of $\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)$ as $\lambda^{\mathrm{o}}(z)$, we can express the objective function in (46), at the optimum, as $p^{\mathrm{o}}\lambda^{\mathrm{o}}(z)$, where

$$p^{\mathrm{o}}\lambda^{\mathrm{o}}(z) = \max_{\|\tilde{\mathbf{u}}\|^2=1} \frac{p^{\mathrm{o}}\tilde{\mathbf{u}}^H\mathbf{B}_2(z)\tilde{\mathbf{u}}}{\tilde{\mathbf{u}}^H\mathbf{B}_4(z)\tilde{\mathbf{u}}}. \tag{48}$$

Here, $\lambda^{\mathrm{o}}(z)$ is expressed as

$$\lambda^{\mathrm{o}}(z) = \lambda_{\max}\left\{\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)\right\} \tag{49}$$

where $\lambda_{\max}\{\cdot\}$ is used to denote the principal eigenvalue of a matrix. Moreover, for any given z, the optimal value of $\tilde{\mathbf{u}}$, denoted as $\tilde{\mathbf{u}}^{\mathrm{o}}(z)$ is equal to the normalized principal eigenvector of $\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)$. That is, we can express $\tilde{\mathbf{u}}^{\mathrm{o}}(z)$ as

$$\tilde{\mathbf{u}}^{\mathrm{o}}(z) = \mathcal{P}\left\{\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)\right\} \tag{50}$$

where $\mathcal{P}\{\cdot\}$ denotes the normalized principal eigenvector of a matrix. We show in Appendix B that

$$\lambda^{\mathrm{o}}(z) = \max_{n\in\mathcal{N}} \ \mathbf{a}_{n+1}^H\mathbf{D}_n^{-1}(z)\mathbf{a}_{n+1} \tag{51}$$

where the following definitions are used: $\mathbf{D}_n(z) \triangleq z\mathbf{Q}_2^n + \sigma^2\mathbf{I}_{l_n} + (P-z)\mathbf{Q}_1^n$, $\mathbf{Q}_1^n \triangleq \mathbf{S}_n\mathbf{Q}_1\mathbf{S}_n^H$, $\mathbf{Q}_2^n \triangleq \mathbf{S}_n\mathbf{Q}_2\mathbf{S}_n^H$. Note that $\mathbf{Q}_q^n$ is an $l_n \times l_n$ diagonal matrix whose diagonal entries are amplitude-squared of the channel coefficients between Transceiver q and those l_n relays which contribute to the n-th tap of the end-to-end CIR $h[\cdot]$. That is

$$\mathbf{Q}_q^n \triangleq \mathrm{diag}\left(\left\{|g_{lq}|^2\right\}_{l=i_n+1}^{i_n+l_n}\right). \tag{52}$$

Using (48) and (51), we can rewrite the optimization problem (41) as

$$\max_{z} \max_{n\in\mathcal{N}} \ \phi_n(z), \text{ subject to } 0 \leq z \leq P \tag{53}$$

where we define $\phi_n(z) \triangleq z(P-z)\mathbf{a}_{n+1}^H\mathbf{D}_n^{-1}(z)\mathbf{a}_{n+1}$, for $n \in \mathcal{N}$, and use the fact that $p^{\mathrm{o}} = P-z$. Swapping the order of the two maximizations, we can rewrite (53) as

$$\max_{n\in\mathcal{N}} \max_{z} \ \phi_n(z), \text{ subject to } \ 0 \leq z \leq P. \tag{54}$$

As shown in [27], for any $n \in \mathcal{N}$, the objective function of the inner maximization problem in (54) is guaranteed to have a unique maximum. Note that when $z = 0$ or $z = P$, the objective function in (54) is zero. As a result, the maximum of the objective function of the inner maximization problem in (54) is attained somewhere within the interval $(0, P)$, for any $n \in \mathcal{N}$. To solve the inner maximization problem in (54) for each $n \in \mathcal{N}$, we can equate the derivative of the objective function with respect to z to zero. We can use a simple bisection method (or a Newton-Raphson technique) to find the optimal value of z which results in the value of the derivative of the function $\phi_n(z)$ being equal to 0. The derivative of $\phi_n(z)$ with respect to z is given by [27]

$$\begin{aligned} f_n(z) &\triangleq \frac{\partial\phi_n(z)}{\partial z} = \\ &\mathbf{a}_{n+1}^H\mathbf{D}_n^{-2}(z)\left((P-2z)\sigma^2\mathbf{I}_{l_n} + (P-z)^2\mathbf{Q}_1^n - z^2\mathbf{Q}_2^n\right)\mathbf{a}_{n+1}. \end{aligned} \tag{55}$$

Using a bisection method for each $n \in \mathcal{N}$, we can obtain the optimal value of z in the interval $[0, P]$ for that n. Denoting this optimal value as z_n, we can then write $f_n(z_n) = 0$. Denoting the index of the optimal block as $n_{\rm o}$, we can use (54) to obtain $n_{\rm o}$ as

$$n_{\rm o} = \arg\max_{n\in\mathcal{N}} \ \phi_n(z_n) \tag{56}$$

and thus write

$$\rho = \sigma^{-2}\ \phi_{n_{\rm o}}(z_{n_{\rm o}}) \tag{57}$$

Indeed, by conducting a search on $\mathcal{N}$, we can find $n_{\rm o}$ as the value of n which results in the maximum value of the objective function in (56). With $n_{\rm o}$ and $z_{n_{\rm o}}$ obtained, we show in Appendix C that the optimal value of the relay beamforming weight vector, denoted as $\mathbf{u}^{\rm o}$, can be obtained as

$$\mathbf{u}^{\rm o} = \kappa\sqrt{P - z_{n_{\rm o}}}\left[\mathbf{0}^T_{i_{n_{\rm o}}\times 1} \quad \mathbf{b}^T_{n_{\rm o}+1} \quad \mathbf{0}^T_{j_{n_{\rm o}}\times 1}\right]^T, \tag{58}$$

where the following definitions are used:

$$\mathbf{b}_{n_{\rm o}+1} \triangleq \mathbf{D}^{-1}_{n_{\rm o}}(z_{n_{\rm o}})\mathbf{a}^*_{n_{\rm o}+1}$$
$$\kappa = 1/\sqrt{\mathbf{b}^H_{n_{\rm o}+1}(\sigma^2\mathbf{I}_{l_{n_{\rm o}}} + z_{n_{\rm o}}\mathbf{Q}_2^{n_{\rm o}})\mathbf{b}_{n_{\rm o}+1}}. \tag{59}$$

To summarize, we note that $\mathbf{u}^{\rm o}$ is the optimal value of the beamforming vector for the relaxed lower bound problem in (35). In what follows, we prove that $\mathbf{u}^{\rm o}$ is the optimal value of the beamforming weight vector for the original optimization problem (23).

Solution to the Original Problem: It follows from the structure of $\mathbf{u}^{\rm o}$ in (58) that the first $i_{n_{\rm o}}$ relays and the last $j_{n_{\rm o}}$ relays will have to be switched off and only the $l_{n_{\rm o}}$ relays which contribute to Tap $n_{\rm o}$ of the end-to-end CIR $h[\cdot]$ will be active and participate in the information exchange between the two transceivers. This implies that only Tap $n_{\rm o}$ of the end-to-end CIR $h[\cdot]$ is nonzero at the optimum and all other taps are zero. Indeed, at the optimum, we can use (58) in (11) to write the vector of the CIR taps as

$$\mathbf{h}(\mathbf{u}^{\rm o}) = \mathbf{C}\mathbf{u}^{\rm o}$$
$$= \kappa\sqrt{P - z_{n_{\rm o}}}[\mathbf{0}^T_{(n_{\rm o}-1)\times 1} \ \ \mathbf{a}^T_{n_{\rm o}+1}\mathbf{b}_{n_{\rm o}+1} \ \ \mathbf{0}^T_{(N-n_{\rm o})\times 1} \ ]^T. \tag{60}$$

As (60) implies, the *optimal end-to-end channel* has only one non-zero tap, and thus, this channel will have a frequency-flat amplitude response over the $N_{\rm s}$-point DFT grid, that is

$$|\mathbf{f}^H_k\mathbf{C}\mathbf{u}^{\rm o}| = |\mathbf{f}^H_{\tilde{k}}\mathbf{C}\mathbf{u}^{\rm o}|, \qquad \text{for } k, \tilde{k} \in \{1, 2, \ldots, N_{\rm s}\} \tag{61}$$

It follows from (61) that at the optimum of the relaxed optimization problem (35), the two inequalities in (33) and (34) are satisfied with equality. As a result, the solution to the relaxed optimization problem (35) is feasible for the original optimization problem in (32) and since the relaxed optimization problem (35) yields a lower bound for the original optimization problem in (32), we conclude that the solution to the relaxed problem (35) for $\mathbf{u}$, given in (58), is optimal for the original optimization problem in (32). The optimal values of P_1 and P_2 can then be obtained from (36) and (37), by replacing $\mathbf{u}$ with $\mathbf{u}^{\rm o}$, as

$$P_1^{\rm o} = \frac{\sigma^2 h_2(\mathbf{u}^{\rm o})}{\|\mathbf{C}\mathbf{u}^{\rm o}\|^2}(\varepsilon_2^{-1} - 1) \tag{62}$$

$$P_2^{\rm o} = \frac{P - \sigma^2\|\mathbf{u}^{\rm o}\|^2}{h_2(\mathbf{u}^{\rm o})} - \frac{\sigma^2 h_1(\mathbf{u}^{\rm o})}{\|\mathbf{C}\mathbf{u}^{\rm o}\|^2}\left(\varepsilon_2^{-1} - 1\right). \tag{63}$$

Next, we characterize the achievable MSE region.

MSE Region Characterization: With the results obtained so far, we are now well-positioned to characterize the MSE region and the boundary thereof. The minimum MSE of Transceiver 1 is the optimal value of the objective function of the optimization problem (32), which is equal to the optimal value of the objective function of the relaxed optimization problem (35), as proven earlier. The minimum MSE at Transceiver 1 can be obtained by replacing $\mathbf{u}$ and P_2 in (35), respectively, with $\mathbf{u}^{\rm o}$ and $P_2^{\rm o}$, given in (58) and (63). That is

$$\mathrm{MSE}_1(\mathbf{u}^{\rm o}, \mathbf{F}_1^{\rm o}, P_2^{\rm o}) = \frac{1}{\frac{1}{\sigma^2}\left(\frac{(P - \sigma^2\|\mathbf{u}^{\rm o}\|^2)\|\mathbf{C}\mathbf{u}^{\rm o}\|^2}{h_1(\mathbf{u}^{\rm o})h_2(\mathbf{u}^{\rm o})} - \sigma^2(\frac{1}{\varepsilon_2} - 1)\right) + 1}. \tag{64}$$

Note that it follows from (41) that by eliminating variable z we can write

$$\rho = \max_{\mathbf{u}} \frac{(P - \sigma^2\|\mathbf{u}\|^2)\|\mathbf{C}\mathbf{u}\|^2}{\sigma^2(\mathbf{u}^H\mathbf{Q}_1\mathbf{u} + 1)(\mathbf{u}^H\mathbf{Q}_2\mathbf{u} + 1)}$$
$$= \frac{(P - \sigma^2\|\mathbf{u}^{\rm o}\|^2)\|\mathbf{C}\mathbf{u}^{\rm o}\|^2}{\sigma^2 h_1(\mathbf{u}^{\rm o})h_2(\mathbf{u}^{\rm o})}. \tag{65}$$

Using (57) and (65), we can rewrite (64) as

$$\mathrm{MSE}_1(\mathbf{u}^{\rm o}, \mathbf{F}_1^{\rm o}, P_2^{\rm o}) = \frac{1}{\frac{1}{\sigma^2}\left(z_{n_{\rm o}}(P - z_{n_{\rm o}})\mathbf{a}^H_{n_{\rm o}+1}\mathbf{D}^{-1}_{n_{\rm o}}(z_{n_{\rm o}})\mathbf{a}_{n_{\rm o}+1} - \sigma^2(\frac{1}{\varepsilon_2} - 1)\right) + 1}. \tag{66}$$

Given that $\mathrm{MSE}_2(\mathbf{u}^{\rm o}, \mathbf{F}_2^{\rm o}, P_1^{\rm o}) = \varepsilon_2$, we can rewrite (66) as

$$\mathrm{MSE}_1^{-1}(\mathbf{u}^{\rm o}, \mathbf{F}_1^{\rm o}, P_2^{\rm o}) + \mathrm{MSE}_2^{-1}(\mathbf{u}^{\rm o}, \mathbf{F}_2^{\rm o}, P_1^{\rm o})$$
$$= \sigma^{-2}z_{n_{\rm o}}(P - z_{n_{\rm o}})\mathbf{a}^H_{n_{\rm o}+1}\mathbf{D}^{-1}_{n_{\rm o}}(z_{n_{\rm o}})\mathbf{a}_{n_{\rm o}+1} + 2 \tag{67}$$

which describes the boundary of the MSE region. As proved above, on this boundary the end-to-end CIR $h[\cdot]$ has only one non-zero tap, and thus, is frequency-flat. That is, the set

$$\mathcal{R}_{\rm MSE} \triangleq \Big\{(\varepsilon_1, \varepsilon_2) \in (0, 1]^2 \Big| \varepsilon_1^{-1} + \varepsilon_2^{-1} \leq \sigma^{-2}z_{n_{\rm o}}(P - z_{n_{\rm o}})\mathbf{a}^H_{n_{\rm o}+1}\mathbf{D}^{-1}_{n_{\rm o}}(z_{n_{\rm o}})\mathbf{a}_{n_{\rm o}+1} + 2\Big\} \tag{68}$$

is the set of the values of $(\mathrm{MSE}_1(\mathbf{u}, \mathbf{F}_1, P_2), \mathrm{MSE}_2(\mathbf{u}, \mathbf{F}_2, P_1))$ that can be achieved with a given total transmit power budget P.

IV. MSE-Constrained Total Power Optimization

Problem Formulation:[7] We now aim to solve the MSE-constrained total power minimization problem, expressed as

$$\min_{P_1\geq 0,P_2\geq 0,\mathbf{u}} \min_{\mathbf{F}_1,\mathbf{F}_2} \; P_{\mathrm{t}}(P_1,P_2,\mathbf{u})$$
$$\text{s.t. } \mathrm{MSE}_1(\mathbf{u},\mathbf{F}_1,P_2)\leq \varepsilon_1,\;\; \mathrm{MSE}_2(\mathbf{u},\mathbf{F}_2,P_1)\leq \varepsilon_2\,. \quad (69)$$

Note that for fixed P_1, P_2 and $\mathbf{u}$, the inner minimization in (69) is convex in $\mathbf{F}_1$ and $\mathbf{F}_2$ due to the fact that the objective function does not depend on $\mathbf{F}_1$ and $\mathbf{F}_2$, while the inequality constraints are convex in $\mathbf{F}_1$ and $\mathbf{F}_2$ (see (25)). Indeed, this inner minimization is a convex feasibility problem. Thus, we can use Karush-Kuhn-Tucker (KKT) conditions to find the optimal solution to this minimization problem. To do so, we can write the Lagrangian function as

$$L(\mathbf{F}_1,\mathbf{F}_2,\nu)\triangleq P_{\mathrm{t}}(P_1,P_2,\mathbf{u})+\nu_1(\mathrm{MSE}_1(\mathbf{u},\mathbf{F}_1,P_2)-\varepsilon_1) + \nu_2(\mathrm{MSE}_2(\mathbf{u},\mathbf{F}_2,P_2)-\varepsilon_2). \quad (70)$$

where ν_1 and ν_2 are the Lagrange multipliers. The optimal $\mathbf{F}_1$ and $\mathbf{F}_2$ (denoted as $\mathbf{F}_1^{\mathrm{o}}$ and $\mathbf{F}_2^{\mathrm{o}}$, respectively) can be obtained by differentiating the Lagrangian function in (70) with respect to $\mathbf{F}_1$ and $\mathbf{F}_2$ and equating these derivatives to zero, and then, solving for $\mathbf{F}_1$ and $\mathbf{F}_2$. Doing so leads us to the following two Wiener filter solutions:

$$\mathbf{F}_1^{\mathrm{o}}=\sqrt{P_2}\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_1^{-1}(\mathbf{u}),\;\; \mathbf{F}_2^{\mathrm{o}}=\sqrt{P_1}\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_2^{-1}(\mathbf{u}). \quad (71)$$

Note that for given P_1, P_2, and $\mathbf{u}$, the values of $\mathbf{F}_1^{\mathrm{o}}$ and $\mathbf{F}_2^{\mathrm{o}}$ must satisfy the constraints in (69), otherwise those values of P_1, P_2, and $\mathbf{u}$ are not feasible. Inserting $\mathbf{F}_1^{\mathrm{o}}$ and $\mathbf{F}_2^{\mathrm{o}}$ into the constraints in (69) results in two constraints, one on P_2 and $\mathbf{u}$, and one on P_1 and $\mathbf{u}$, which have to be taken into account when the outer minimization in (69) is solved. The details of solving the outer minimization is provided in the sequel.

Optimal Relay Beamforming: Using (71), we show in Appendix A that $\mathrm{MSE}_1(\mathbf{u},\mathbf{F}_1^{\mathrm{o}},P_2)$ can be written as

$$\mathrm{MSE}_1(\mathbf{u},\mathbf{F}_1^{\mathrm{o}},P_2)=N_{\mathrm{s}}^{-1}\sum_{k=1}^{N_{\mathrm{s}}}\left[\frac{P_2N_{\mathrm{s}}|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2}{\sigma^2h_1(\mathbf{u})}+1\right]^{-1}, \quad (72)$$

where $\mathbf{f}_k\triangleq\frac{1}{\sqrt{N_{\mathrm{s}}}}[1\;\; e^{j\frac{2\pi(k-1)}{N_{\mathrm{s}}}}\;\cdots\; e^{j\frac{2\pi(N-1)(k-1)}{N_{\mathrm{s}}}}]^T$. Similarly, one can use (71) to write

$$\mathrm{MSE}_2(\mathbf{u},\mathbf{F}_2^{\mathrm{o}},P_1)=N_{\mathrm{s}}^{-1}\sum_{k=1}^{N_{\mathrm{s}}}\left[\frac{P_1N_{\mathrm{s}}|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2}{\sigma^2h_2(\mathbf{u})}+1\right]^{-1}. \quad (73)$$

Note that the optimization problem (69) is feasible only when $\varepsilon_1,\varepsilon_2\in(0,1]$. The reason is that each of the N_{s} terms under the summations in (72) and (73) is less than 1 (see also (68)). Using MSE region characterization results from the previous section, we can use the following lemma to solve the optimization problem (69).

[7]We redefine some of the notations that we used in the previous section.

Lemma 1: Without loss of optimality, one can assume that at the optimum of (69), $|\mathbf{f}_k\mathbf{C}\mathbf{u}|^2=|\mathbf{f}_{\tilde{k}}\mathbf{C}\mathbf{u}|^2$, holds true, for $k,\tilde{k}\in\{1,2,\ldots,N_{\mathrm{s}}\}$. That is, the optimal end-to-end channel has a constant magnitude over the N_{s}-point DFT grid.

Proof: The proof relies on the results we presented in the previous section, see Appendix D.

Based on Lemma 1 and using Parseval's theorem,[8] i.e., $\sum_{k=1}^{N_{\mathrm{s}}}|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2=\|\mathbf{C}\mathbf{u}\|^2$, we can write $N_{\mathrm{s}}|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2=\|\mathbf{C}\mathbf{u}\|^2$, for $k=1,2,\ldots,N_{\mathrm{s}}$, and thus, rewrite the minimization problem in (69) as

$$\min_{P_1\geq 0,P_2\geq 0,\mathbf{u}} \; P_1(\mathbf{u}^H\mathbf{Q}_1\mathbf{u}+1)+P_2(\mathbf{u}^H\mathbf{Q}_2\mathbf{u}+1)+\sigma^2\|\mathbf{u}\|^2$$
$$\text{s.t.}\;\; \left[\frac{P_2\|\mathbf{C}\mathbf{u}\|^2}{\sigma^2h_1(\mathbf{u})}+1\right]^{-1}\leq\varepsilon_1\,,\;\; \left[\frac{P_1\|\mathbf{C}\mathbf{u}\|^2}{\sigma^2h_2(\mathbf{u})}+1\right]^{-1}\leq\varepsilon_2. \quad (74)$$

One can easily prove that at the optimum, the two inequality constraints in (74) must be satisfied with equality. Otherwise, if any of the two constraints is satisfied with inequality, then the corresponding power can be reduced without violating that constraint while resulting in a lower value for the objective function, thereby contradicting the optimality. Based on the fact that at the optimum, the two inequality constraints in (74) must be satisfied with equality, we can express P_1 and P_2, in terms of $\mathbf{u}$, as

$$P_1=\frac{\sigma^2h_2(\mathbf{u})}{\|\mathbf{C}\mathbf{u}\|^2}(\varepsilon_2^{-1}-1),\quad P_2=\frac{\sigma^2h_1(\mathbf{u})}{\|\mathbf{C}\mathbf{u}\|^2}(\varepsilon_1^{-1}-1)\,. \quad (75)$$

We now use (75) to eliminate P_1 and P_2 in the optimization problem (74) and rewrite this problem as

$$\min_{\mathbf{u}}\;\; \frac{\sigma^2h_1(\mathbf{u})h_2(\mathbf{u})}{\|\mathbf{C}\mathbf{u}\|^2}((\varepsilon_1^{-1}+\varepsilon_2^{-1})-2)+\sigma^2\|\mathbf{u}\|^2\,. \quad (76)$$

Let us define a new optimization variable z as $z\triangleq\frac{h_2(\mathbf{u})}{\|\mathbf{C}\mathbf{u}\|^2}((\varepsilon_1^{-1}+\varepsilon_2^{-1})-2)=\frac{(\mathbf{u}^H\mathbf{Q}_2\mathbf{u}+1)}{\|\mathbf{C}\mathbf{u}\|^2}((\varepsilon_1^{-1}+\varepsilon_2^{-1})-2)$. Note that as ε_1 and ε_2 are both smaller than 1 (see also (68)), $z\geq 0$ holds true. As a result, we can write the optimization problem (76) as

$$\min_{z\geq 0}\min_{\mathbf{u}}\;\; (\mathbf{u}^H\mathbf{Q}_1\mathbf{u}+1)z+\|\mathbf{u}\|^2$$
$$\text{subject to}\;\; z=\frac{(\mathbf{u}^H\mathbf{Q}_2\mathbf{u}+1)}{\|\mathbf{C}\mathbf{u}\|^2}((\varepsilon_1^{-1}+\varepsilon_2^{-1})-2) \quad (77)$$

or, equivalently, as

$$\min_{z\geq 0} z+\min_{\mathbf{u}}\;\mathbf{u}^H(z\mathbf{Q}_1+\mathbf{I})\mathbf{u},\;\text{s.t. }\mathbf{u}^H(z\mathbf{C}^H\mathbf{C}-\gamma\mathbf{Q}_2)\mathbf{u}=\gamma, \quad (78)$$

where we define $\gamma\triangleq((\varepsilon_1^{-1}+\varepsilon_2^{-1})-2)\geq 0$. Note that we can write the constraint in (78) as $\mathbf{u}^H\mathbf{Q}_2^{H/2}(z\mathbf{Q}_2^{-H/2}\mathbf{C}^H\mathbf{C}\mathbf{Q}_2^{-1/2}-\gamma\mathbf{I})\mathbf{Q}_2^{1/2}\mathbf{u}=\gamma$. As a result, the inner minimization in (78) is feasible iff

$$\lambda_{\max}\{z\mathbf{Q}_2^{-H/2}\mathbf{C}^H\mathbf{C}\mathbf{Q}_2^{-1/2}-\gamma\mathbf{I}\}>0\,. \quad (79)$$

[8]Note that $\mathbf{f}_k\triangleq\frac{1}{\sqrt{N_{\mathrm{s}}}}[1\;\; e^{j\frac{2\pi(k-1)}{N_{\mathrm{s}}}}\;\cdots\; e^{j\frac{2\pi(N-1)(k-1)}{N_{\mathrm{s}}}}]^T$, for $k=1,2,\ldots,N_{\mathrm{s}}$ is the truncated version of k-th column of the $N_{\mathrm{s}}\times N_{\mathrm{s}}$ unitary inverse DFT matrix.

Using the fact that $\mathbf{C}^H\mathbf{C}$ is a block diagonal matrix with rank-1 diagonal blocks (see (17)), we can show that (see Appendix E)

$$\lambda_{\max}\{z\mathbf{Q}_2^{-H/2}\mathbf{C}^H\mathbf{C}\mathbf{Q}_2^{-1/2} - \gamma\mathbf{I}\} = \max_{n\in\mathcal{N}} z\,\mathbf{a}_{n+1}^H(\mathbf{Q}_2^n)^{-1}\mathbf{a}_{n+1} - \gamma. \quad (80)$$

Hence, the feasibility condition in (79) is equivalent to

$$z > \frac{\gamma}{\max_{n\in\mathcal{N}}\ \mathbf{a}_{n+1}^H(\mathbf{Q}_2^n)^{-1}\mathbf{a}_{n+1}} = \frac{\gamma}{\max_{n\in\mathcal{N}}\|\mathbf{g}_1^n\|^2}. \quad (81)$$

where we define $\mathbf{g}_1^n \triangleq [g_{(i_n+1),1} \quad g_{(i_n+2),1} \quad \cdots \quad g_{(i_n+l_n),1}]^T$ and the equality follows from (16) and (52). Let $\tilde{\mathbf{u}}(z)$ represent the normalized (i.e., $\|\tilde{\mathbf{u}}(z)\|^2 = 1$) principal eigenvector of matrix

$$\mathbf{P}(z) \triangleq (z\mathbf{Q}_1 + \mathbf{I})^{-1/2}(z\mathbf{C}^H\mathbf{C} - \gamma\mathbf{Q}_2)(z\mathbf{Q}_1 + \mathbf{I})^{-1/2}. \quad (82)$$

That is, $\tilde{\mathbf{u}}(z) \triangleq \mathcal{P}\{\mathbf{P}(z)\}$. Then, for any feasible z, the inner minimization in (78) has a closed-form solution given by

$$\begin{aligned}\mathbf{u}^{\mathrm{o}}(z) &= \kappa\mathcal{P}\left\{(z\mathbf{Q}_1 + \mathbf{I})^{-1}(z\mathbf{C}^H\mathbf{C} - \gamma\mathbf{Q}_2)\right\} \\ &= \kappa(z\mathbf{Q}_1 + \mathbf{I})^{-1/2}\mathcal{P}\left\{\mathbf{P}(z)\right\} = \kappa(z\mathbf{Q}_1 + \mathbf{I})^{-1/2}\tilde{\mathbf{u}}(z).\end{aligned} \quad (83)$$

Here, κ is a normalization factor which guarantees that the constraint in (78) is satisfied, that is $\kappa^2 = \frac{\gamma}{(\tilde{\mathbf{u}}^H(z)\mathbf{P}(z)\tilde{\mathbf{u}}(z))} = \frac{\gamma}{\lambda_{\max}\{\mathbf{P}(z)\}}$, Also, the optimum value of the inner minimization in (77) is given by $\frac{\gamma}{\lambda_{\max}\{\mathbf{P}(z)\}}$. Note that $\mathbf{P}(z)$ is a block diagonal matrix, because $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are diagonal matrices, while $\mathbf{C}^H\mathbf{C} = \mathrm{blkdiag}\{\mathbf{a}_{n+1}^*\mathbf{a}_{n+1}^T\}_{n\in\mathcal{N}}$ is a block diagonal matrix (see (17)). Hence, the principal eigenvalue of $\mathbf{P}(z)$ is equal to the largest of the principal eigenvalues of different blocks. That is

$$\lambda(z) \triangleq \lambda_{\max}\{\mathbf{P}(z)\} = \max_{n\in\mathcal{N}}\lambda_{\max}\{\mathbf{S}_n\mathbf{P}(z)\mathbf{S}_n^H\} = \max_{n\in\mathcal{N}}\lambda_n(z) \quad (84)$$

where we define $\lambda_n(z) \triangleq \lambda_{\max}\{(z\mathbf{Q}_1^n + \mathbf{I})^{-1/2}(z\mathbf{a}_{n+1}^*\mathbf{a}_{n+1}^T - \gamma\mathbf{Q}_2^n)(z\mathbf{Q}_1^n + \mathbf{I})^{-1/2}\}$. Using (84), we can rewrite the optimization problem (78) as $\min_{z\geq 0} z + \frac{\gamma}{\max_{n\in\mathcal{N}}\lambda_n(z)}$ or, equivalently, as[9] $\min_{z\geq 0} z + \min_{n\in\mathcal{N}}\frac{\gamma}{[\lambda_n(z)]^+}$ or as

$$\min_{n\in\mathcal{N}}\min_{z>0}\ \ z + \frac{\gamma}{[\lambda_n(z)]^+}. \quad (85)$$

For any $n \in \mathcal{N}$, the solution to the inner minimization in (85) occurs when $\lambda_n(z) > 0$ (i.e., when[10] $z > \frac{\gamma}{\|\mathbf{g}_1^n\|^2}$). The reason is that when $\lambda_n(z) \leq 0$, this cost function approaches $+\infty$. We can thus write (85) as

$$\min_{n\in\mathcal{N}}\min_{z>\frac{\gamma}{\|\mathbf{g}_1^n\|^2}}\ \ z + \frac{\gamma}{\lambda_n(z)}. \quad (86)$$

[9]Note that for any given z, $\lambda_n(z)$ could be negative for some n but the feasibility condition in (80) guarantees that $\lambda_n(z)$ cannot be negative for all $n \in \mathcal{N}$.

[10]Note that $\lambda_n(z) = \lambda_{\max}\{(z\mathbf{Q}_1^n + \mathbf{I})^{-1/2}(z\mathbf{a}_{n+1}^*\mathbf{a}_{n+1}^T - \gamma\mathbf{Q}_2^n)(z\mathbf{Q}_1^n + \mathbf{I})^{-1/2}\}$ is positive only when $\lambda_{\max}\{(z\mathbf{a}_{n+1}^*\mathbf{a}_{n+1}^T - \gamma\mathbf{Q}_2^n)\} = z\,\mathbf{a}_{n+1}^H(\mathbf{Q}_2^n)^{-1}\mathbf{a}_{n+1} - \gamma > 0$ holds true.

One can show that for any feasible $z > \frac{\gamma}{\|\mathbf{g}_1^n\|^2}$, the value of $\lambda_n(z) = \lambda_{\max}\{(z\mathbf{Q}_1^n + \mathbf{I})^{-1/2}(z\mathbf{a}_{n+1}^*\mathbf{a}_{n+1}^T - \gamma\mathbf{Q}_2^n)(z\mathbf{Q}_1^n + \mathbf{I})^{-1/2}\}$ is equal to the provably unique positive value of λ which satisfies the following equation:

$$z\mathbf{a}_{n+1}^H\mathbf{A}_n^{-1}(\lambda, z)\mathbf{a}_{n+1} = 1, \quad (87)$$

where we define $\mathbf{A}_n(\lambda, z) \triangleq \gamma\mathbf{Q}_2^n + \lambda(z\mathbf{Q}_1^n + \mathbf{I}_{l_n})$. Moreover, the solution to the inner minimization in (86) in the interval $(\frac{\gamma}{\|\mathbf{g}_1^n\|^2}, +\infty)$ is provably unique and can be obtained as the root of the following equation:

$$\hbar_n(z) \triangleq \frac{1/z^2 - \lambda\mathbf{a}_{n+1}^H\mathbf{A}_n^{-2}(\lambda, z)\mathbf{Q}_1^n\mathbf{a}_{n+1}}{\lambda^2\mathbf{a}_{n+1}^H\mathbf{A}_n^{-2}(\lambda, z)(z\mathbf{Q}_1^n + \mathbf{I}_{l_n})\mathbf{a}_{n+1}} - \frac{1}{\gamma} = 0 \quad (88)$$

where λ is obtained as the provably unique positive solution to (87), for any $z \in (\frac{\gamma}{\|\mathbf{g}_1^n\|^2}, +\infty)$. Denoting the solution to the inner minimization in (86) as z_n, we can obtain the optimal value of n, denoted as n_{o}, as

$$n_{\mathrm{o}} = \arg\min_{n\in\mathcal{N}}\ \ z_n + \frac{\gamma}{\lambda_n(z_n)}. \quad (89)$$

It is worth stressing that the value of n_{o} dictates which diagonal block of the block diagonal matrix $\mathbf{P}(z_{n_{\mathrm{o}}})$ has the largest eigenvalue. This block in turn yields the value of the principal eigenvector of matrix $\mathbf{P}(z_{n_{\mathrm{o}}})$. With n_{o} and $z_{n_{\mathrm{o}}}$ obtained, we show in Appendix F that the optimal value of the relay beamforming weight vector, $\mathbf{u}^{\mathrm{o}}$ can be obtained as

$$\mathbf{u}^{\mathrm{o}} = \begin{bmatrix}\mathbf{0}_{i_{n_{\mathrm{o}}}\times 1}^T & \mathbf{b}_{n_{\mathrm{o}}+1}^T & \mathbf{0}_{j_{n_{\mathrm{o}}}\times 1}^T\end{bmatrix}^T \quad (90)$$

where we used the following definitions:

$$\begin{aligned}\mathbf{b}_{n_{\mathrm{o}}+1} &\triangleq \kappa\alpha\sqrt{z_{n_{\mathrm{o}}}}\mathbf{h} = \frac{\sqrt{\gamma}}{\sqrt{\lambda_{n_{\mathrm{o}}}(z_{n_{\mathrm{o}}})\mathbf{h}^H(z_{n_{\mathrm{o}}}\mathbf{Q}_1^{n_{\mathrm{o}}} + \mathbf{I}_{l_{n_{\mathrm{o}}}})\mathbf{h}}}\mathbf{h} \\ \mathbf{h} &= \mathbf{A}_{n_{\mathrm{o}}}^{-1}(\lambda_{n_{\mathrm{o}}}(z_{n_{\mathrm{o}}}), z_{n_{\mathrm{o}}})\mathbf{a}_{n_{\mathrm{o}}+1}^*\end{aligned} \quad (91)$$

Using (90) in (11), we can obtain the vector of the end-to-end CIR taps as

$$\mathbf{h}(\mathbf{u}^{\mathrm{o}}) = \mathbf{C}\mathbf{u}^{\mathrm{o}} = [\mathbf{0}_{(n_{\mathrm{o}}-1)\times 1}^T \ \ \mathbf{a}_{n_{\mathrm{o}}+1}^T\mathbf{b}_{n_{\mathrm{o}}+1} \ \ \mathbf{0}_{(N-n_{\mathrm{o}})\times 1}^T \ \ ]^T. \quad (92)$$

It now becomes clear from (92) that at the optimum, the end-to-end channel has only a single non-zero tap.

V. Discussions

We now have all the derivations to show how the solution to the MSE-constrained total power minimization problem (69) is related to the solution to the rate-constrained power minimization problem. The latter problem is expressed as

$$\begin{aligned}&\min_{P_1\geq 0, P_2\geq 0, \mathbf{u}}\ P_{\mathrm{t}}(P_1, P_2, \mathbf{u}) \\ &\text{subject to}\ \ R_1(\mathbf{u}, P_2) \geq r_1, \ \ R_2(\mathbf{u}, P_1) > r_2.\end{aligned} \quad (93)$$

where $R_1(\mathbf{u}, P_2)$ ($R_2(\mathbf{u}, P_1)$) is the achievable data rate at Transceiver 1 (2) and r_1 (r_2) is the minimum required rate data rate at Transceiver 1 (2). It needs to be stressed that the MSE-constrained total power minimization (69) and the rate-constrained total power minimization (93) have two important

differences, as spelled out in the sequel: 1) While solving (93), no linear receiver is assumed, rather the optimal decoder is assumed at the receiving front-ends of the two transceivers. However, in the MSE-constrained total power minimization (69), linear receiver (equalizer) is assumed. 2) Moreover, the minimum required rates (93) are achieved only when infinitely long Gaussian channel source codes are used at the two transceivers, while the MSE constraints in (69) guarantees non-coded performance. Note that in the existing literature, one can find the relationship between MSE and the data rate when the data rate is measured at the output of the linear receiver. Such a relationship between the MSE at the output of a linear receiver and the data rate in the absence of a linear receiver does not hold in general. Using the notations used here, we can simplify the optimization problem (93) as [24]

$$\min_{\mathbf{u}} \quad \frac{\sigma^2 h_1(\mathbf{u}) h_2(\mathbf{u})}{\|\mathbf{C}\mathbf{u}\|^2}(2^{\frac{r_1}{N_\mathrm{s}}}+2^{\frac{r_2}{N_\mathrm{s}}}-2)+\sigma^2\|\mathbf{u}\|^2$$
$$\text{s.t.} \quad \mathbf{u}\in\bigcup_{n\in\mathcal{N}}\mathcal{B}_n. \tag{94}$$

Here, $\mathcal{B}_n$ is the set of those values of $\mathbf{u}$ which result in only $h[n]$ being non-zero, for some $n\in\mathcal{N}$, while all other taps of $h[\cdot]$ are zero.[11] The sets $\{\mathcal{B}_n\}_{n=0}^{N-1}$ are mutually exclusive, that is $\mathcal{B}_{n_1}\bigcap\mathcal{B}_{n_2}=\emptyset$, for $n_1\neq n_2$. The optimization problem (94) is similar to the one in (76) except for the constraint $\mathbf{u}\in\bigcup_{n\in\mathcal{N}}\mathcal{B}_n$, which guarantees that the end-to-end CIR has only one non-zero tap. Note however that, as shown in the previous section, minimizing the cost function in (94) without this constraint already leads to a single-tap end-to-end CIR. Hence, dropping this constraint from (94) will not cause any loss of optimality. As a result, the MSE-constrained total power minimization problem (69) and the rate-constrained power minimization (93) lead to the same result if we choose $\varepsilon_1=2^{\frac{-r_1}{N_\mathrm{s}}}$ and $\varepsilon_2=2^{\frac{-r_2}{N_\mathrm{s}}}$. In other words, the solution to the rate-constrained power minimization, with threshold pair (r_1,r_2) is identical to the solution to the MSE-constrained total power minimization problem with maximum MSE pair of $(2^{-\frac{r_1}{N_\mathrm{s}}},2^{-\frac{r_2}{N_\mathrm{s}}})$. Such a relationship between these two solutions is the direct result of the fact that at the optimum, both solutions lead to a single-tap end-to-end channel, and hence, the following relationship between the achieved rates and the MSEs holds true:

$$r_p=N_\mathrm{s}\log_2\varepsilon_p^{-1},\quad \text{for } p=1,2 \tag{95}$$

Note that without going through the rigorous derivation presented in this paper, there was no way to conclude that choosing r_1 and r_2 as in (95) and solving the rate-constrained power minimization would be also power-optimal when the MSEs are constrained to be less that ε_1 and ε_2.

Remark 1: In the studied scheme, it is assumed that the two transceivers know which relays, if any, contribute to each tap of the end-to-end CIR. This assumption is realistic as the propagation delay corresponding to each transceiver-relay link depends on the geographical distance between the respective transceiver and the relay. If, prior to data communication between the two transceivers, each relay provides its own coordinates, through a control channel, to each of the two transceivers, then each transceiver can calculate the propagation delay between itself and all relays. As a result, both transceivers can have the knowledge of the propagation delays between both transceivers and the relays, and thus, can reconstruct the end-to-end delay corresponding to each relaying path. With the knowledge of these relaying delays, each transceiver can find out which relays contribute to each tap of the end-to-end CIR. Note also that the calculation of the relaying delays can be performed at the relays, provided that each relay acquires the coordinates of the two transceivers over the control channel. Then, each relay can send its corresponding relaying delay to the two transceivers.

[11] In other words, for $\mathbf{u}\in\bigcup_{n\in\mathcal{N}}\mathcal{B}_n$, the end-to-end CIR $h[\cdot]$ has only a single non-zero tap and the beamforming weights of those relays which do not subscribe to this non-zero tap will be zero.

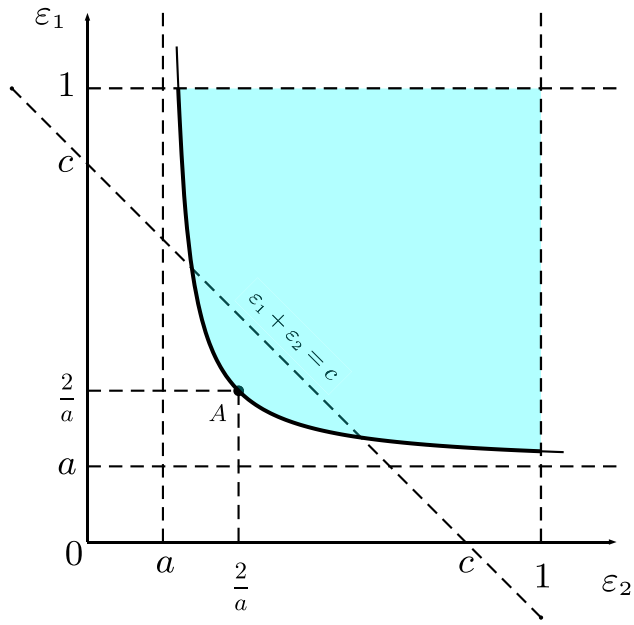

Fig. 2. The achievable MSE region $\mathcal{R}_{\mathrm{MSE.with}}$ with $a\triangleq(\sigma^{-2}z_{n_\mathrm{o}}(P-z_{n_\mathrm{o}})\mathbf{a}_{n_\mathrm{o}+1}^H\mathbf{D}_{n_\mathrm{o}}^{-1}(z_{n_\mathrm{o}})\mathbf{a}_{n_\mathrm{o}+1}+2)^{-1}$.

VI. Sum-MSE minimization

We now show how the characterization of the boundary of the achievable MSE region can be used to solve the problem of sum-MSE (or total MSE) minimization under a total power budget. The latter minimization problem is formulated as

$$\min_{P_1\geq0,P_2\geq0,\mathbf{u}} \quad \min_{\mathbf{F}_1,\mathbf{F}_2} \quad \mathrm{MSE}_1(\mathbf{u},\mathbf{F}_1,P_2)+\mathrm{MSE}_2(\mathbf{u},\mathbf{F}_2,P_1)$$
$$\text{subject to} \quad P_\mathrm{t}(P_1,P_2,\mathbf{u})\leq P. \tag{96}$$

The achievable MSE region $\mathcal{R}_{\mathrm{MSE}}$, characterized in (68), is shown in Fig. 2 as the area with light blue color. In this figure, we define $a\triangleq(\sigma^{-2}z_{n_\mathrm{o}}(P-z_{n_\mathrm{o}})\mathbf{a}_{n_\mathrm{o}+1}^H\mathbf{D}_{n_\mathrm{o}}^{-1}(z_{n_\mathrm{o}})\mathbf{a}_{n_\mathrm{o}+1}+2)^{-1}$. As one can easily see from this figure, the minimum sum-MSE is achieved, when the line $\varepsilon_1+\varepsilon_2=c$ touches the boundary at point A, where $\varepsilon_1=\varepsilon_2=2a$ holds true, i.e., when MSE balancing occurs. As we proved in Section III, at any point on

this boundary, the end-to-end channel will have only one single non-zero tap, and thus, the sum-MSE-optimal value of the relay beamforming vector is given as in (58). The corresponding optimal value of the transceiver's transmit powers can be obtained by replacing ε_1 and ε_2 in (62) and (63) with $2a$, thereby arriving at

$$P_1^{\mathrm{o}} = \frac{h_2(\mathbf{u}^{\mathrm{o}})}{2\|\mathbf{C}\mathbf{u}^{\mathrm{o}}\|^2}(z_{n_{\mathrm{o}}}(P - z_{n_{\mathrm{o}}})\mathbf{a}_{n_{\mathrm{o}}+1}^H\mathbf{D}_{n_{\mathrm{o}}}^{-1}(z_{n_{\mathrm{o}}})\mathbf{a}_{n_{\mathrm{o}}+1})$$

$$\begin{aligned} P_2^{\mathrm{o}} &= \frac{P - \sigma^2\|\mathbf{u}^{\mathrm{o}}\|^2}{h_2(\mathbf{u}^{\mathrm{o}})} - \frac{\sigma^2 h_1(\mathbf{u}^{\mathrm{o}})}{\|\mathbf{C}\mathbf{u}^{\mathrm{o}}\|^2}\left(\varepsilon_2^{-1} - 1\right) \\ &= \frac{\sigma^2 h_1(\mathbf{u}^{\mathrm{o}})}{\|\mathbf{C}\mathbf{u}^{\mathrm{o}}\|^2} \times \left(\frac{(P - \sigma^2\|\mathbf{u}^{\mathrm{o}}\|^2)\|\mathbf{C}\mathbf{u}^{\mathrm{o}}\|^2}{h_1(\mathbf{u}^{\mathrm{o}})h_2(\mathbf{u}^{\mathrm{o}})} - \right. \\ &\qquad \left. 0.5\sigma^{-2}z_{n_{\mathrm{o}}}(P - z_{n_{\mathrm{o}}})\mathbf{a}_{n_{\mathrm{o}}+1}^H\mathbf{D}_{n_{\mathrm{o}}}^{-1}(z_{n_{\mathrm{o}}})\mathbf{a}_{n_{\mathrm{o}}+1}\right) \\ &= \frac{h_1(\mathbf{u}^{\mathrm{o}})}{2\|\mathbf{C}\mathbf{u}^{\mathrm{o}}\|^2}(z_{n_{\mathrm{o}}}(P - z_{n_{\mathrm{o}}})\mathbf{a}_{n_{\mathrm{o}}+1}^H\mathbf{D}_{n_{\mathrm{o}}}^{-1}(z_{n_{\mathrm{o}}})\mathbf{a}_{n_{\mathrm{o}}+1}) \quad (97) \end{aligned}$$

where we have used (57) and (65) to arrive at (97).

VII. Numerical Results

The network consists of two transceivers which are located at the Cartesian coordinates of $(-1000, 0)$ m and (1000,0) m. The relay nodes are randomly located over an area of 1000 meters by 1000 meters centered at (0,0) m. The small-scale fading coefficients are modeled as independent and identically distributed complex Gaussian random variables with zero mean and unit variance. The Okumura-Hata model for urban environment is used to model the path loss as $137.03 + 35\log_{10}(\text{distance})$, with the path loss exponent being 3.5. We assume shadowing standard deviation to be 8 dB. The small scale fading coefficients are modeled as independent and identically distributed complex Gaussian random variables with zero mean and unit variance. The noise at the transceivers and the relays are also modeled as zero-mean white Gaussian random variables with variance $\sigma^2 = -104$ dBm, which is equivalent to the noise level in a 10 MHz transmission bandwidth.

Fig. 3 shows two sets of curves: 1) the average minimum total transmit power versus $\varepsilon_1 = \varepsilon_2$ (red curves and red axes), and 2) average minimum total transmit power versus the rate thresholds (black curves and axes). As can seen in these curves, doubling the number of relays will reduce the average minimum total power consumption by about 3.5 dB, while a 10 fold increase in the number of relays, reduces the total power consumption by 10 dB. The power-vs-rate curves also show that the power increases exponentially with rate requirements for moderate to high rate regimes.

VIII. Conclusion

A relay network is referred to as asynchronous, when the signal travel times between any end-node (source or destination) and different relays are significantly different from each other. Such a relay channel will cause inter-symbol interference at the receiving side of the link. Considering such asynchronous relay networks, in this paper, we presented two contributions: 1) under a total power constraint and using linear post-channel block (single-carrier) equalization at the two transceivers, we solved

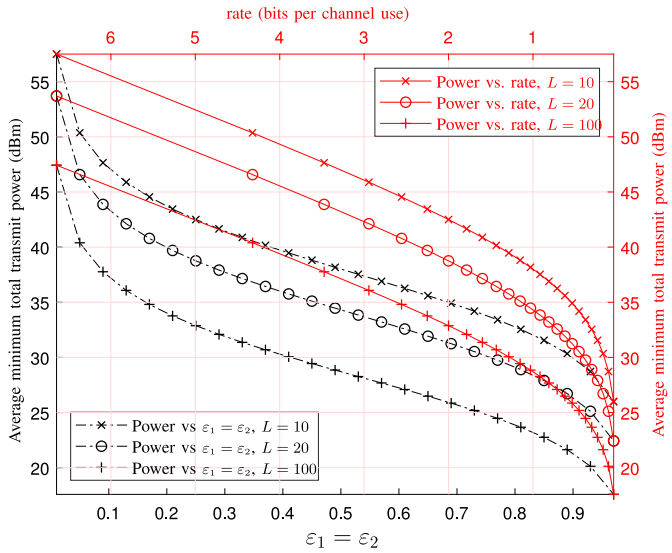

Fig. 3. Average minimum total transmit power versus $\varepsilon_1 = \varepsilon_2$ (red curves) and average minimum total transmit power versus rate thresholds (black curves).

the problem of minimizing the MSE at the output of the linear block equalizer of one transceiver, while the MSE at the output of the linear block equalizer at the other transceiver is kept at a certain level. Carried over the transceivers' transmit powers, the linear post-channel block equalizers and the relay complex beamforming weights, this minimization aims to characterize the MSE region. Such an MSE region characterization allows us to characterize the region of un-coded probabilities of error at the two transceivers. 2) As the second contribution, we optimized the transceivers' transmit powers, the linear post-channel block equalizers, and the relay complex beamforming weights by solving the total power minimization problem under MSE constraints and proved how this solution is related to the solution of the rate-constrained power minimization problem. To solve this problem, we relied on our MSE region characterization to rigorously prove that in designing transceiver power allocation and distributed beamforming, our MSE-constrained total power minimization approach and the rate-constrained power minimization technique of [24] are equivalent, provided that the MSE thresholds in the former approach and the rate thresholds in the latter approach are properly chosen. The equivalence of these two approaches allows us to infer the un-coded MSE performance of the network from the rate-constrained problem, and conversely, the coded rate performance of the network can be inferred from the MSE-constrained total power minimization problem.

Appendix A
Derivation of (30)

In light of (29), $\eta(\mathbf{u}, P_1, P_2) = \mathrm{MSE}_1(\mathbf{u}, \mathbf{F}_1^{\mathrm{o}}, P_2)$ can be written as in (A.1), shown at the bottom of the next page, where the last equality follows from the fact that $\mathbf{R}_1(\mathbf{u}, P_2) = \mathbf{R}_1^H(\mathbf{u}, P_2)$. Let $\mathbf{W}$ denote the $N_{\mathrm{s}} \times N_{\mathrm{s}}$ inverse DFT matrix and define $\mathbf{f}_k \triangleq \frac{1}{\sqrt{N_{\mathrm{s}}}}[1 \;\; e^{j\frac{2\pi(k-1)}{N_{\mathrm{s}}}} \;\; \cdots \; e^{j\frac{2\pi(N-1)(k-1)}{N_{\mathrm{s}}}}]^T$. Let us now define the $N_{\mathrm{s}} \times N_{\mathrm{s}}$ diagonal matrix $\mathbf{J}(\mathbf{u})$, whose diagonal entries are the frequency response of the end-to-end

channel given by $H(e^{j2\pi f}) \triangleq \sum_{n=0}^{N-1} h[n]e^{-j2\pi fn}$ at normalized frequencies of integer multiples of $1/N_{\rm s}$. Hence, we can write

$$\begin{aligned}\mathbf{J}(\mathbf{u}) &\triangleq \text{diag}\{H(e^{j0}), H(e^{j\frac{2\pi}{N_{\rm s}}}), \ldots, H(e^{j\frac{2\pi(N_{\rm s}-1)}{N_{\rm s}}})\} \\ &= \sqrt{N_{\rm s}}\text{diag}\left\{\mathbf{f}_k^H\mathbf{C}\mathbf{u}\right\}_{k=1}^{N_{\rm s}}. \end{aligned} \tag{A.2}$$

Note that the circulant matrix $\tilde{\mathbf{H}}(\mathbf{u})$ can be decomposed as $\tilde{\mathbf{H}}(\mathbf{u}) = \mathbf{W}^H\mathbf{J}(\mathbf{u})\mathbf{W}$, and hence, $\eta(\mathbf{u}, P_1, P_2)$ is written in terms of $\mathbf{W}$ and $\mathbf{J}(\mathbf{u})$ as in (A.3), shown at the bottom of this page. Now, using (A.2), we can simplify $\eta(\mathbf{u}, P_1, P_2)$ as in (A.4) shown at the bottom of the next page.

Appendix B
Derivation of (51)

As $\mathbf{B}_2(z)$, given in (44), is a block diagonal matrix with block sizes $\{l_n \times l_n\}_{n\in\mathcal{N}}$, while $\mathbf{B}_4(z)$ is a diagonal matrix, $\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)$ is also a block-diagonal matrix with block sizes $\{l_n \times l_n\}_{n\in\mathcal{N}}$. Note that one can obtain the principal eigenvector and the largest eigenvalue of a block-diagonal matrix via dealing with each block of such a matrix separately. That is, aiming to find the principal eigenvalue as well as the principal eigenvector of the block diagonal matrix $\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)$, we can first obtain the principal eigenvalue of each block and then introduce the largest of these principal eigenvalues, obtained for different blocks, as the principal eigenvalue of the block diagonal matrix $\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)$. One can then easily find the principal eigenvector of the block diagonal matrix $\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)$ from the principal eigenvector of that block of this matrix which has the largest principal eigenvalue among all different blocks. Based on these explanations and using the fact that different blocks of $\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)$ can be extracted as $\mathbf{S}_n\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)\mathbf{S}_n^H$, for $n \in \mathcal{N}$, we can rewrite (49) as[12] in (B.1), shown at the bottom of the next page, where in the second equality, we use (44); in the third equality, we use (4); in the fourth equality we use (18); and in the fifth equality, we use the fact that the only non-zero eigenvalue of $\mathbf{x}\mathbf{y}^H$ is $\mathbf{y}^H\mathbf{x}$; in the seventh equality, we use the fact that $\mathbf{B}_1(z) = (z\mathbf{Q}_2 + \sigma^2\mathbf{I}_L)^{-\frac{1}{2}}$; and in the eighth equality, we use (6). The derivation of (51) is now complete.

Appendix C
Derivation of (58)

With $n_{\rm o}$ and $P_2^{n_{\rm o}}$ obtained, we now use (50) to obtain the optimal value of $\tilde{\mathbf{u}}$ as

$$\begin{aligned}\tilde{\mathbf{u}}^{\rm o} &= \tilde{\mathbf{u}}(z_{n_{\rm o}}) = \mathcal{P}\left\{\mathbf{B}_4^{-1}(z_{n_{\rm o}})\mathbf{B}_2(z_{n_{\rm o}})\right\} \\ &= \left[\mathbf{0}_{i_{n_{\rm o}}\times 1}^T \;\; \mathcal{P}^T\left\{\mathbf{S}_{n_{\rm o}}\mathbf{B}_4^{-1}(z_{n_{\rm o}})\mathbf{B}_2(z_{n_{\rm o}})\mathbf{S}_{n_{\rm o}}^H\right\} \;\; \mathbf{0}_{j_{n_{\rm o}}\times 1}^T\right]^T\end{aligned} \tag{C.1}$$

[12]Note that the sizes of the diagonal blocks of $\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)$ are dictated by and are equal to the sizes of the diagonal blocks of $\mathbf{C}^H\mathbf{C}$.

$$\begin{aligned}\eta(\mathbf{u}, P_1, P_2) &= \text{MSE}_1(\mathbf{u}, \mathbf{F}_1^{\rm o}, P_2) = N_{\rm s}^{-1}\text{tr}\left\{\mathbf{F}_1^{\rm o}\mathbf{R}_1(\mathbf{u}, P_2)\mathbf{F}_1^{{\rm o}H}\right\} - N_{\rm s}^{-1}\sqrt{P_2}\text{tr}\left\{\mathbf{F}_1^{\rm o}\tilde{\mathbf{H}}(\mathbf{u}) + \tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{F}_1^{{\rm o}H}\right\} + 1 \\ &= N_{\rm s}^{-1}\text{tr}\left\{(\sqrt{P_2}\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_1^{-1}(\mathbf{u}, P_2))\mathbf{R}_1(\mathbf{u}, P_2)(\sqrt{P_2}\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_1^{-1}(\mathbf{u}, P_2))^H\right\} \\ &- N_{\rm s}^{-1}\sqrt{P_2}\text{tr}\left\{(\sqrt{P_2}\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_1^{-1}(\mathbf{u}, P_2)\tilde{\mathbf{H}}(\mathbf{u}) + \tilde{\mathbf{H}}^H(\mathbf{u})(\sqrt{P_2}\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_1^{-1}(\mathbf{u}, P_2))^H\right\} + 1 \\ &= N_{\rm s}^{-1}P_2\text{tr}\left\{\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_1^{-1}(\mathbf{u}, P_2)\mathbf{R}_1(\mathbf{u}, P_2)\mathbf{R}_1^{-1}(\mathbf{u}, P_2)^H\tilde{\mathbf{H}}(\mathbf{u})\right\} - 2N_{\rm s}^{-1}P_2\text{tr}\left\{\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_1^{-1}(\mathbf{u}, P_2)\tilde{\mathbf{H}}(\mathbf{u})\right\} + 1 \\ &= 1 - N_{\rm s}^{-1}P_2\text{tr}\left\{\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{R}_1^{-1}(\mathbf{u}, P_2)\tilde{\mathbf{H}}(\mathbf{u})\right\},\end{aligned} \tag{A.1}$$

$$\begin{aligned}&\eta(\mathbf{u}, P_1, P_2) = 1 - N_{\rm s}^{-1}P_2\text{tr}\left\{\mathbf{W}^H\mathbf{J}^H(\mathbf{u})\mathbf{W}\mathbf{R}_1^{-1}(\mathbf{u}, P_2)\mathbf{W}^H\mathbf{J}(\mathbf{u})\mathbf{W}\right\} \\ &= 1 - N_{\rm s}^{-1}P_2\text{tr}\left\{\underbrace{\mathbf{W}\mathbf{W}^H}_{\mathbf{I}}\mathbf{J}^H(\mathbf{u})\mathbf{W}\mathbf{R}_1^{-1}(\mathbf{u}, P_2)\mathbf{W}^H\mathbf{J}(\mathbf{u})\right\} = 1 - N_{\rm s}^{-1}P_2\text{tr}\left\{\mathbf{J}^H(\mathbf{u})\left(\mathbf{W}\mathbf{R}_1(\mathbf{u}, P_2)\mathbf{W}^H\right)^{-1}\mathbf{J}(\mathbf{u})\right\} \\ &= 1 - N_{\rm s}^{-1}P_2\text{tr}\left\{\mathbf{J}^H(\mathbf{u})\left(\mathbf{W}\left(P_2\tilde{\mathbf{H}}(\mathbf{u})\tilde{\mathbf{H}}^H(\mathbf{u}) + \sigma^2(\|\mathbf{Q}_1\mathbf{u}\|^2 + 1)\mathbf{I}_{N_{\rm s}}\right)\mathbf{W}^H\right)^{-1}\mathbf{J}(\mathbf{u})\right\} \\ &= 1 - N_{\rm s}^{-1}P_2\text{tr}\left\{\mathbf{J}^H(\mathbf{u})\left(P_2\underbrace{\mathbf{W}\tilde{\mathbf{H}}(\mathbf{u})}_{\mathbf{J}(\mathbf{u})\mathbf{W}}\underbrace{\tilde{\mathbf{H}}^H(\mathbf{u})\mathbf{W}^H}_{\mathbf{W}^H\mathbf{J}^H(\mathbf{u})} + \sigma^2(\|\mathbf{Q}_1\mathbf{u}\|^2 + 1)\underbrace{\mathbf{W}\mathbf{W}^H}_{\mathbf{I}}\right)^{-1}\mathbf{J}(\mathbf{u})\right\} \\ &= 1 - N_{\rm s}^{-1}P_2\text{tr}\left\{\mathbf{J}^H(\mathbf{u})\left(P_2\mathbf{J}(\mathbf{u})\mathbf{J}^H(\mathbf{u}) + \sigma^2(\|\mathbf{Q}_1\mathbf{u}\|^2 + 1)\mathbf{I}_{N_{\rm s}}\right)^{-1}\mathbf{J}(\mathbf{u})\right\}.\end{aligned} \tag{A.3}$$

where the third equality follows from the fact that the principal eigenvector of the block diagonal matrix $\mathbf{B}_4^{-1}(z_{n_\mathrm{o}})\mathbf{B}_2(z_{n_\mathrm{o}})$ can be obtained from the principal eigenvector of the block which has the largest principal eigenvalue among all blocks. To further simplify (C.1), we can use (C.2), shown at the top of the next page, where in the first equality, we use (44); in the second equality, we use (4); in the third equality, we use (18); and in the fourth equality, we use the fact that the principal eigenvector of $\mathbf{x}\mathbf{y}^H$ is $\mathbf{x}$; and κ is a normalizing factor guaranteeing that $\|\tilde{\mathbf{u}}^\mathrm{o}\|^2 = 1$ holds true. More specifically, we obtain κ as in (C.3), shown at the top of the next page, where the following definition is used:

$$\mathbf{b}_{n_\mathrm{o}+1} \triangleq \left(z_{n_\mathrm{o}}\mathbf{Q}_2^{n_\mathrm{o}} + \sigma^2\mathbf{I}_{l_{n_\mathrm{o}}} + (P - z_{n_\mathrm{o}})\mathbf{Q}_1^{n_\mathrm{o}}\right)^{-1}\mathbf{a}_{n_\mathrm{o}+1}^*. \quad \text{(C.4)}$$

Using (C.1) and (C.2) in (42), we can obtain the optimal value of $\mathbf{u}$, denoted as $\mathbf{u}^\mathrm{o}$, as in (C.5), shown at the top of the next page, where in the third equality, we use (47); in the fourth equality, we use the fact that $\mathbf{B}_1(z) = (z\mathbf{Q}_2 + \sigma^2\mathbf{I}_L)^{-\frac{1}{2}}$; in the fifth equality, we use (6); and in the seventh equality, we use (C.4). The derivation of (58) is now complete.

Appendix D
Proof of Lemma 1

Let $P_\mathrm{t}^{\min}$ stand for the minimum value of $P_\mathrm{t}(P_1, P_2, \mathbf{u})$ obtained by solving (69), while the corresponding optimal values of the optimization variables are given by $(P_1^\mathrm{o}, P_2^\mathrm{o}, \mathbf{u}^\mathrm{o}, \mathbf{F}_1^\mathrm{o}, \mathbf{F}_2^\mathrm{o})$. We now consider the following minimization problem:

$$\min_{P_1\geq 0, P_2\geq 0, \mathbf{u}} \ \min_{\mathbf{F}_1, \mathbf{F}_2} \ \mathrm{MSE}_1(\mathbf{u}, \mathbf{F}_1, P_2)$$
$$\text{subject to } \mathrm{MSE}_2(\mathbf{u}, \mathbf{F}_2, P_1) = \varepsilon_2, \quad P_\mathrm{t}(P_1, P_2, \mathbf{u}) \leq P_\mathrm{t}^{\min} \quad \text{(D.1)}$$

Let $(\hat{P}_1^\mathrm{o}, \hat{P}_2^\mathrm{o}, \hat{\mathbf{u}}^\mathrm{o}, \hat{\mathbf{F}}_1^\mathrm{o}, \hat{\mathbf{F}}_2^\mathrm{o})$ be the solution to the optimization problem (D.1), while $\hat{\varepsilon}_1$ denotes the minimum achievable value for the MSE at Transceiver 1, obtained by solving (D.1), for the given budget $P_\mathrm{t}^{\min}$. We now show that $\hat{\varepsilon}_1 = \varepsilon$ holds true. To show this, we rely on contradiction: if $\hat{\varepsilon}_1 > \varepsilon_1$, then $(P_1^\mathrm{o}, P_2^\mathrm{o}, \mathbf{u}^\mathrm{o}, \mathbf{F}_1^\mathrm{o}, \mathbf{F}_2^\mathrm{o})$ leads to a smaller value for the cost function in (D.1). Indeed, $(P_1^\mathrm{o}, P_2^\mathrm{o}, \mathbf{u}^\mathrm{o}, \mathbf{F}_1^\mathrm{o}, \mathbf{F}_2^\mathrm{o})$ being a solution to (69), leads to a smaller value for MSE, i.e., ε_1, at Transceiver 1, while the MSE at Transceiver 2 is equal to ε_2, and at the same time, $P_\mathrm{t}(P_1^\mathrm{o}, P_2^\mathrm{o}, \mathbf{u}^\mathrm{o}) = P_\mathrm{t}^{\min}$ and this contradicts the optimality of $(\hat{P}_1^\mathrm{o}, \hat{P}_2^\mathrm{o}, \hat{\mathbf{u}}^\mathrm{o}, \hat{\mathbf{F}}_1^\mathrm{o}, \hat{\mathbf{F}}_2^\mathrm{o})$ for (D.1). On the other hand, we can easily show that $\hat{\varepsilon}_1$ cannot be smaller than ε_1 either. Otherwise, if $\hat{\varepsilon}_1 < \varepsilon_1$, then one can scale down the optimal value $\hat{P}_2^\mathrm{o}$ by some real $\alpha < 1$ (see (72)) such that $\mathrm{MSE}_1(\hat{\mathbf{u}}^\mathrm{o}, \hat{\mathbf{F}}_1^\mathrm{o}, \alpha\hat{P}_2^\mathrm{o}) = \varepsilon_1$ holds true, without violating the constraint in (D.1), i.e., $P_\mathrm{t}(\hat{P}_1^\mathrm{o}, \alpha\hat{P}_2^\mathrm{o}, \hat{\mathbf{u}}^\mathrm{o}) < P_\mathrm{t}^{\min}$. In other words, the tuple $(\hat{P}_1^\mathrm{o}, \alpha\hat{P}_2^\mathrm{o}, \hat{\mathbf{u}}^\mathrm{o}, \mathbf{F}_1^\mathrm{o}, \mathbf{F}_2^\mathrm{o}))$ results in a lower total power, while satisfying the two constraints in (69).

$$\begin{aligned}
\eta(\mathbf{u}, P_1, P_2) &= 1 - N_\mathrm{s}^{-1}\mathrm{tr}\left\{\mathrm{diag}\left\{\frac{N_\mathrm{s}P_2|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2}{N_\mathrm{s}P_2|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2 + \sigma^2(\|\mathbf{Q}_1\mathbf{u}\|^2+1)}\right\}_{k=1}^{N_\mathrm{s}}\right\} \\
&= 1 - N_\mathrm{s}^{-1}\sum_{k=1}^{N_\mathrm{s}}\frac{N_\mathrm{s}P_2|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2}{N_\mathrm{s}P_2|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2 + \sigma^2(\|\mathbf{Q}_1\mathbf{u}\|^2+1)} = \sum_{k=1}^{N_\mathrm{s}}\left(N_\mathrm{s}^{-1} - \frac{P_2|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2}{N_\mathrm{s}P_2|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2 + \sigma^2(\|\mathbf{Q}_1\mathbf{u}\|^2+1)}\right) \\
&= \sum_{k=1}^{N_\mathrm{s}}\frac{N_\mathrm{s}^{-1}\sigma^2(\|\mathbf{Q}_1\mathbf{u}\|^2+1)}{N_\mathrm{s}P_2|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2 + \sigma^2(\|\mathbf{Q}_1\mathbf{u}\|^2+1)} = N_\mathrm{s}^{-1}\sum_{k=1}^{N_\mathrm{s}}\frac{1}{\frac{N_\mathrm{s}P_2|\mathbf{f}_k^H\mathbf{C}\mathbf{u}|^2}{\sigma^2(\mathbf{u}^H\mathbf{Q}_1\mathbf{u}+1)}+1}
\end{aligned} \quad \text{(A.4)}$$

$$\begin{aligned}
\lambda^\mathrm{o}(z) &= \max_{n\in\mathcal{N}}\ \lambda_{\max}\left\{\mathbf{S}_n\mathbf{B}_4^{-1}(z)\mathbf{B}_2(z)\mathbf{S}_n^H\right\} = \max_{n\in\mathcal{N}}\ \lambda_{\max}\left\{\mathbf{S}_n\mathbf{B}_4^{-1}(z)\mathbf{B}_1(z)\mathbf{C}^H\mathbf{C}\mathbf{B}_1(z)\mathbf{S}_n^H\right\} \\
&= \max_{n\in\mathcal{N}}\ \lambda_{\max}\left\{\mathbf{S}_n\mathbf{B}_4^{-1}(z)\mathbf{B}_1(z)\mathbf{S}_n^H\mathbf{S}_n\mathbf{C}^H\mathbf{C}\mathbf{S}_n^H\mathbf{S}_n\mathbf{B}_1(z)\mathbf{S}_n^H\right\} = \max_{n\in\mathcal{N}}\ \lambda_{\max}\left\{\mathbf{S}_n\mathbf{B}_4^{-1}(z)\mathbf{B}_1(z)\mathbf{S}_n^H\mathbf{a}_{n+1}^*\mathbf{a}_{n+1}^T\mathbf{S}_n\mathbf{B}_1(z)\mathbf{S}_n^H\right\} \\
&= \max_{n\in\mathcal{N}}\ \mathbf{a}_{n+1}^T\mathbf{S}_n\mathbf{B}_1(z)\underbrace{\mathbf{S}_n^H\mathbf{S}_n}_{\mathbf{I}_{l_n}}\left(\mathbf{I}_L + p^\mathrm{o}\mathbf{B}_1(z)\mathbf{Q}_1\mathbf{B}_1(z)\right)^{-1}\mathbf{B}_1(z)\mathbf{S}_n^H\mathbf{a}_{n+1}^* = \max_{n\in\mathcal{N}}\ \mathbf{a}_{n+1}^T\mathbf{S}_n\left(\mathbf{B}_1^{-2}(z) + p^\mathrm{o}\mathbf{Q}_1\right)^{-1}\mathbf{S}_n^H\mathbf{a}_{n+1}^* \\
&= \max_{n\in\mathcal{N}}\ \mathbf{a}_{n+1}^T\mathbf{S}_n\left(z\mathbf{Q}_2 + \sigma^2\mathbf{I}_L + (P-z)\mathbf{Q}_1\right)^{-1}\mathbf{S}_n^H\mathbf{a}_{n+1}^* \\
&= \max_{n\in\mathcal{N}}\ \mathbf{a}_{n+1}^T\left(z\underbrace{\mathbf{S}_n\mathbf{Q}_2\mathbf{S}_n^H}_{\mathbf{Q}_2^n} + \sigma^2\mathbf{S}_n\mathbf{S}_n^H + (P-z)\underbrace{\mathbf{S}_n\mathbf{Q}_1\mathbf{S}_n^H}_{\mathbf{Q}_1^n}\right)^{-1}\mathbf{a}_{n+1}^* \\
&= \max_{n\in\mathcal{N}}\ \mathbf{a}_{n+1}^T\left(\underbrace{z\mathbf{Q}_2^n + \sigma^2\mathbf{I}_{l_n} + (P-z)\mathbf{Q}_1^n}_{\triangleq\mathbf{D}_n(z)}\right)^{-1}\mathbf{a}_{n+1}^* = \max_{n\in\mathcal{N}}\ \mathbf{a}_{n+1}^H\mathbf{D}_n^{-1}(z)\mathbf{a}_{n+1}
\end{aligned} \quad \text{(B.1)}$$

$$\begin{aligned}\mathcal{P}\left\{\mathbf{S}_{n_\mathrm{o}}\mathbf{B}_4^{-1}(z_{n_\mathrm{o}})\mathbf{B}_2(z_{n_\mathrm{o}})\mathbf{S}_{n_\mathrm{o}}^H\right\} &= \mathcal{P}\left\{\mathbf{S}_{n_\mathrm{o}}\mathbf{B}_4^{-1}(z_{n_\mathrm{o}})\mathbf{B}_1(z_{n_\mathrm{o}})\mathbf{C}^H\mathbf{C}\mathbf{B}_1(z_{n_\mathrm{o}})\mathbf{S}_{n_\mathrm{o}}^H\right\}\\ &= \mathcal{P}\left\{\mathbf{S}_{n_\mathrm{o}}\mathbf{B}_4^{-1}(z_{n_\mathrm{o}})\mathbf{B}_1(z_{n_\mathrm{o}})\mathbf{S}_{n_\mathrm{o}}^H\mathbf{S}_{n_\mathrm{o}}\mathbf{C}^H\mathbf{C}\mathbf{S}_{n_\mathrm{o}}^H\mathbf{S}_{n_\mathrm{o}}\mathbf{B}_1(z_{n_\mathrm{o}})\mathbf{S}_{n_\mathrm{o}}^H\right\}\\ &= \mathcal{P}\left\{\mathbf{S}_{n_\mathrm{o}}\mathbf{B}_4^{-1}(z_{n_\mathrm{o}})\mathbf{B}_1(z_{n_\mathrm{o}})\mathbf{S}_{n_\mathrm{o}}^H\mathbf{a}_{n_\mathrm{o}+1}^*\mathbf{a}_{n_\mathrm{o}+1}^T\mathbf{S}_{n_\mathrm{o}}\mathbf{B}_1(z_{n_\mathrm{o}})\mathbf{S}_{n_\mathrm{o}}^H\right\} = \kappa\mathbf{S}_{n_\mathrm{o}}\mathbf{B}_4^{-1}(z_{n_\mathrm{o}})\mathbf{B}_1(z_{n_\mathrm{o}})\mathbf{S}_{n_\mathrm{o}}^H\mathbf{a}_{n_\mathrm{o}+1}^* \end{aligned} \tag{C.2}$$

$$\begin{aligned}\kappa &= 1/\sqrt{\mathbf{a}_{n_\mathrm{o}+1}^T\mathbf{S}_{n_\mathrm{o}}\mathbf{B}_4^{-2}(z_{n_\mathrm{o}})\mathbf{B}_1^2(z_{n_\mathrm{o}})\mathbf{S}_{n_\mathrm{o}}^H\mathbf{a}_{n_\mathrm{o}+1}^*} = 1/\sqrt{\mathbf{a}_{n_\mathrm{o}+1}^T\mathbf{S}_{n_\mathrm{o}}\left(\mathbf{B}_1^{-1}(z_{n_\mathrm{o}})+(P-z_{n_\mathrm{o}})\mathbf{B}_1(z_{n_\mathrm{o}})\mathbf{Q}_1\right)^{-2}\mathbf{S}_{n_\mathrm{o}}^H\mathbf{a}_{n_\mathrm{o}+1}^*}\\ &= 1/\sqrt{\mathbf{a}_{n_\mathrm{o}+1}^T\mathbf{S}_{n_\mathrm{o}}\mathbf{B}_1^{-2}(z_{n_\mathrm{o}})\left(\mathbf{B}_1^{-2}(z_{n_\mathrm{o}})+(P-z_{n_\mathrm{o}})\mathbf{Q}_1\right)^{-2}\mathbf{S}_{n_\mathrm{o}}^H\mathbf{a}_{n_\mathrm{o}+1}^*}\\ &= 1/\sqrt{\mathbf{a}_{n_\mathrm{o}+1}^T\left(\sigma^2\mathbf{I}_{l_{n_\mathrm{o}}}+z_{n_\mathrm{o}}\mathbf{Q}_2^{n_\mathrm{o}}\right)\left(z_{n_\mathrm{o}}\mathbf{Q}_2^{n_\mathrm{o}}+\sigma^2\mathbf{I}_{l_{n_\mathrm{o}}}+(P-z_{n_\mathrm{o}})\mathbf{Q}_1^{n_\mathrm{o}}\right)^{-2}\mathbf{a}_{n_\mathrm{o}+1}^*} = 1/\sqrt{\mathbf{b}_{n_\mathrm{o}+1}^H(\sigma^2\mathbf{I}_{l_{n_\mathrm{o}}}+z_{n_\mathrm{o}}\mathbf{Q}_2^{n_\mathrm{o}})\mathbf{b}_{n_\mathrm{o}+1}}\end{aligned} \tag{C.3}$$

$$\begin{aligned}\mathbf{u}^\mathrm{o} &= \sqrt{p^\mathrm{o}}\mathbf{B}_1(z_{n_\mathrm{o}})\tilde{\mathbf{u}}^\mathrm{o} = \kappa\sqrt{P-z_{n_\mathrm{o}}}\left[\mathbf{0}_{i_{n_\mathrm{o}}\times 1}^T \quad \mathbf{S}_{n_\mathrm{o}}\mathbf{B}_1(z_{n_\mathrm{o}})\mathbf{B}_4^{-1}(z_{n_\mathrm{o}})\mathbf{B}_1(z_{n_\mathrm{o}})\mathbf{S}_{n_\mathrm{o}}^H\mathbf{a}_{n_\mathrm{o}+1}^* \quad \mathbf{0}_{j_{n_\mathrm{o}}\times 1}^T\right]^T\\ &= \kappa\sqrt{P-z_{n_\mathrm{o}}}\left[\mathbf{0}_{i_{n_\mathrm{o}}\times 1}^T \quad \mathbf{S}_{n_\mathrm{o}}\left(\mathbf{B}_1^{-2}(z_{n_\mathrm{o}})+(P-z_{n_\mathrm{o}})\mathbf{Q}_1\right)^{-1}\mathbf{S}_{n_\mathrm{o}}^H\mathbf{a}_{n_\mathrm{o}+1}^* \quad \mathbf{0}_{j_{n_\mathrm{o}}\times 1}^T\right]^T\\ &= \kappa\sqrt{P-z_{n_\mathrm{o}}}\left[\mathbf{0}_{i_{n_\mathrm{o}}\times 1}^T \quad \mathbf{S}_{n_\mathrm{o}}\left(\sigma^2\mathbf{I}_L+z_{n_\mathrm{o}}\mathbf{Q}_2+(P-z_{n_\mathrm{o}})\mathbf{Q}_1\right)^{-1}\mathbf{S}_{n_\mathrm{o}}^H\mathbf{a}_{n_\mathrm{o}+1}^* \quad \mathbf{0}_{j_{n_\mathrm{o}}\times 1}^T\right]^T\\ &= \kappa\sqrt{P-z_{n_\mathrm{o}}}\left[\mathbf{0}_{i_{n_\mathrm{o}}\times 1}^T \quad \left(\sigma^2\underbrace{\mathbf{S}_{n_\mathrm{o}}\mathbf{S}_{n_\mathrm{o}}^H}_{\mathbf{I}_{l_{n_\mathrm{o}}}}+z_{n_\mathrm{o}}\underbrace{\mathbf{S}_{n_\mathrm{o}}\mathbf{Q}_2\mathbf{S}_{n_\mathrm{o}}^H}_{\mathbf{Q}_2^{n_\mathrm{o}}}+(P-z_{n_\mathrm{o}})\underbrace{\mathbf{S}_{n_\mathrm{o}}\mathbf{Q}_1\mathbf{S}_{n_\mathrm{o}}}_{\mathbf{Q}_2^{n_\mathrm{o}}}\right)^{-1}\mathbf{a}_{n_\mathrm{o}+1}^* \quad \mathbf{0}_{j_{n_\mathrm{o}}\times 1}^T\right]^T\\ &= \kappa\sqrt{P-z_{n_\mathrm{o}}}\left[\mathbf{0}_{i_{n_\mathrm{o}}\times 1}^T \quad \left(\sigma^2\mathbf{I}_{l_{n_\mathrm{o}}}+z_{n_\mathrm{o}}\mathbf{Q}_2^{n_\mathrm{o}}+(P-z_{n_\mathrm{o}})\mathbf{Q}_1^{n_\mathrm{o}}\right)^{-1}\mathbf{a}_{n_\mathrm{o}+1}^* \quad \mathbf{0}_{j_{n_\mathrm{o}}\times 1}^T\right]^T\\ &= \kappa\sqrt{P-z_{n_\mathrm{o}}}\left[\mathbf{0}_{i_{n_\mathrm{o}}\times 1}^T \quad \mathbf{b}_{n_\mathrm{o}+1}^T \quad \mathbf{0}_{j_{n_\mathrm{o}}\times 1}^T\right]^T\end{aligned} \tag{C.5}$$

$$\begin{aligned}&\lambda_{\max}\{z\mathbf{Q}_2^{-H/2}\mathbf{C}^H\mathbf{C}\mathbf{Q}_2^{-1/2}-\gamma\mathbf{I}\} = \max_{n\in\mathcal{N}}\ \lambda_{\max}\{\mathbf{S}_n(z\mathbf{Q}_2^{-H/2}\mathbf{C}^H\mathbf{C}\mathbf{Q}_2^{-1/2}-\gamma\mathbf{I})\mathbf{S}_n^H\} =\\ &\max_{n\in\mathcal{N}}\ \lambda_{\max}\{z(\mathbf{Q}_2^n)^{-H/2}\mathbf{a}_{n+1}^*\mathbf{a}_{n+1}^T(\mathbf{Q}_2^n)^{-1/2}-\gamma\mathbf{I}_{l_n}\} = \max_{n\in\mathcal{N}}\ z\mathbf{a}_{n+1}^T(\mathbf{Q}_2^n)^{-1}\mathbf{a}_{n+1}^*-\gamma = \max_{n\in\mathcal{N}}\ z\mathbf{a}_{n+1}^H(\mathbf{Q}_2^n)^{-1}\mathbf{a}_{n+1}-\gamma\end{aligned} \tag{E.1}$$

This contradicts the optimality of $(P_1^\mathrm{o}, P_2^\mathrm{o}, \mathbf{u}^\mathrm{o}, \mathbf{F}_1^\mathrm{o}, \mathbf{F}_2^\mathrm{o})$ for (69). Therefore, we can conclude that $\hat{\varepsilon}_1 = \varepsilon_1$ holds true, meaning that any solution to the optimization problem (D.1) is indeed a solution to the optimization problem (69) and any solution to the optimization (69) *inherits all properties of the solutions to (D.1)*. As shown in Section III, at the optimum of (D.1), the *end-to-end channel* has only one non-zero tap, and therefore, this channel will have a frequency-flat amplitude response over the N_s-point DFT grid (see (61), so will the end-to-end channel in (69). The proof is complete.

Appendix E
Derivation of (80)

Note that matrix $\mathbf{Q}_2$ is diagonal, while matrix $\mathbf{C}^H\mathbf{C}$ is block diagonal (see (17)). Hence, matrix $z\mathbf{Q}_2^{-H/2}\mathbf{C}^H\mathbf{C}\mathbf{Q}_2^{-1/2}-\gamma\mathbf{I}$ is block diagonal as well, and thus, the principal eigenvalue of this matrix is equal to the largest of the principal eigenvalues of different blocks, that is, we can write as in (E.1) shown at the top of this page. The derivation is complete.

Appendix F
Derivation of (90)

With n_o and z_{n_o} obtained, we can use (83) to obtain the optimal value of $\mathbf{u}$ as

$$\mathbf{u}^\mathrm{o} = \mathbf{u}^\mathrm{o}(z_{n_\mathrm{o}}) = \kappa(z_{n_\mathrm{o}}\mathbf{Q}_1+\mathbf{I})^{-1/2}\mathcal{P}\left\{\mathbf{P}(z_{n_\mathrm{o}})\right\} \tag{F.1}$$

Note that matrix $\mathbf{P}(z_{n_\mathrm{o}})$, as defined in (82), is block diagonal. The reason is that matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are diagonal while matrix $\mathbf{C}^H\mathbf{C}$ is block diagonal, see (17). Hence, the principal eigenvector of $\mathbf{P}(z)$ can be obtained from the principal eigenvector

$$\begin{aligned}\mathbf{w} &= \mathcal{P}\{\mathbf{S}_{n_o}(z_{n_o}\mathbf{Q}_1+\mathbf{I})^{-1/2}(z_{n_o}\mathbf{C}^H\mathbf{C}-\gamma\mathbf{Q}_2)(z_{n_o}\mathbf{Q}_1+\mathbf{I})^{-1/2}\mathbf{S}_{n_o}^H\}\\ &= \mathcal{P}\{\mathbf{S}_{n_o}(z_{n_o}\mathbf{Q}_1+\mathbf{I})^{-1/2}\mathbf{S}_{n_o}^H\mathbf{S}_{n_o}(z_{n_o}\mathbf{C}^H\mathbf{C}-\gamma\mathbf{Q}_2)\mathbf{S}_{n_o}^H\mathbf{S}_{n_o}(z_{n_o}\mathbf{Q}_1+\mathbf{I})^{-1/2}\mathbf{S}_{n_o}^H\}\\ &= \mathcal{P}\{(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})^{-1/2}(z_{n_o}\mathbf{a}_{n_o+1}^*\mathbf{a}_{n_o+1}^T-\gamma\mathbf{Q}_2^{n_o})(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})^{-1/2}\}\end{aligned} \tag{F.3}$$

$$\begin{aligned}\mathbf{w} &= \alpha(\mathbf{B}+\lambda_{n_o}(z_{n_o})\mathbf{I})^{-1}\mathbf{b} = \alpha\left[\gamma\mathbf{Q}_2^{n_o}(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})^{-1}+\lambda_{n_o}(z_{n_o})\mathbf{I}\right]^{-1}\times\sqrt{z_{n_o}}(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})^{-1/2}\mathbf{a}_{n_o+1}^*\\ &= \alpha\sqrt{z_{n_o}}\left[\gamma\mathbf{Q}_2^{n_o}+\lambda_{n_o}(z_{n_o})(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})\right]^{-1}(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})^{+1/2}\mathbf{a}_{n_o+1}^*\end{aligned} \tag{F.5}$$

$$\begin{aligned}\mathbf{u}^{\mathrm{o}} &= \mathbf{u}^{\mathrm{o}}(z_{n_o}) = \kappa(z_{n_o}\mathbf{Q}_1+\mathbf{I})^{-1/2}\mathcal{P}\left\{\mathbf{P}(z_{n_o})\right\}\\ &= \left[\mathbf{0}_{i_{n_o}\times 1}^T\quad \kappa\alpha\sqrt{z_{n_o}}\left(\left[\gamma\mathbf{Q}_2^{n_o}+\lambda_{n_o}(z_{n_o})(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})\right]^{-1}\mathbf{a}_{n_o+1}^*\right)^T\quad \mathbf{0}_{j_{n_o}\times 1}^T\right]^T = \left[\mathbf{0}_{i_{n_o}\times 1}^T\quad \mathbf{b}_{n_o+1}^T\quad \mathbf{0}_{j_{n_o}\times 1}^T\right]^T\end{aligned} \tag{F.6}$$

of that block of $\mathbf{P}(z)$ which has the largest eigenvalue among all blocks.[13] Therefore, we can write (F.1) as

$$\mathbf{u}^{\mathrm{o}} = \mathbf{u}^{\mathrm{o}}(z_{n_o}) = \kappa(z_{n_o}\mathbf{Q}_1+\mathbf{I})^{-1/2}\overbrace{\mathcal{P}\left\{\mathbf{P}(z_{n_o})\right\}}^{\tilde{\mathbf{u}}(z)} =$$

$$\kappa(z_{n_o}\mathbf{Q}_1+\mathbf{I})^{-1/2}\left[\mathbf{0}_{i_{n_o}\times 1}^T\ \underbrace{(\mathcal{P}\{\mathbf{S}_{n_o}\mathbf{P}(z_{n_o})\mathbf{S}_{n_o}^H\})^T}_{\triangleq\mathbf{w}}\ \mathbf{0}_{j_{n_o}\times 1}^T\right]^T \tag{F.2}$$

where the second equality follows from the fact that the value of n_{o} dictates which diagonal block of the block diagonal matrix $\mathbf{P}(z_{n_o})$ has the largest eigenvalue among all blocks. This block in turn yields the value of the principal eigenvector of matrix $\mathbf{P}(z_{n_o})$. We can now write $\mathbf{w}$ as in (F.3), shown at the top of this page, where we use (82) and (4) in the first and second equalities, respectively. Using the following definitions: $\mathbf{b}\triangleq\sqrt{z_{n_o}}(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})^{-1/2}\mathbf{a}_{n_o+1}^*$, $\mathbf{B}=\gamma\mathbf{Q}_2^{n_o}(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})^{-1}$, we can use (F.3) to write $(\mathbf{b}\mathbf{b}^H-\mathbf{B})\mathbf{w}=\lambda_{n_o}(z_{n_o})\mathbf{w}$ or

$$\mathbf{b}\mathbf{b}^H\mathbf{w} = (\mathbf{B}+\lambda_{n_o}(z_{n_o})\mathbf{I})\mathbf{w}\,. \tag{F.4}$$

Using (F.4), we can write as in (F.5), shown at the top of this page, where α is chosen such that $\|\mathbf{w}\|^2=1$, that is $\alpha=\frac{1}{\sqrt{z_{n_o}\mathbf{a}_{n_o+1}^T[\gamma\mathbf{Q}_2^{n_o}+\lambda_{n_o}(z_{n_o})(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})]^{-2}(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})\mathbf{a}_{n_o+1}^*}}$. We now use (F.5) in (F.2) to write as in (F.6), shown at the top of this page, where we used the following definitions:

$$\mathbf{h} = \left[\gamma\mathbf{Q}_2^{n_o}+\lambda_{n_o}(z_{n_o})(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})\right]^{-1}\mathbf{a}_{n_o+1}^* \tag{F.7}$$

$$\mathbf{b}_{n_o+1}\triangleq\kappa\alpha\sqrt{z_{n_o}}\mathbf{h} = \frac{\sqrt{\gamma}}{\sqrt{\lambda_{n_o}(z_{n_o})\mathbf{h}^H(z_{n_o}\mathbf{Q}_1^{n_o}+\mathbf{I}_{l_{n_o}})\mathbf{h}}}\mathbf{h}. \tag{F.8}$$

The derivation of (90) is now complete.

[13] With $\mathbf{A}_n$ being an $l_n\times l_n$ matrix, for $n\in\mathcal{N}$, the principal eigenvector of the block diagonal matrix $\mathrm{blkdiag}\{\mathbf{A}_n\}_{n\in\mathcal{N}}$ is equal to $[\mathbf{0}_{i_{n_o}\times 1}^T\ \ (\mathcal{P}\{\mathbf{A}_{n_o}\})^T\ \ \mathbf{0}_{j_{n_o}\times 1}]^T$. Here, $\mathbf{A}_{n_o}$ is the block which has the largest principal eigenvalue among all blocks $\{\mathbf{A}_n\}_{n\in\mathcal{N}}$ while i_{n_o} and j_{n_o} are given as in (2) and (3), respectively.

REFERENCES

[1] F. Jameel, S. Wyne, G. Kaddoum, and T. Q. Duong, "A comprehensive survey on cooperative relaying and jamming strategies for physical layer security," *IEEE Commun. Surv. Tut.*, vol. 21, no. 3, pp. 2734–2771, Jul.–Sep. 2019.

[2] R. Kumar and A. Hossain, "Survey on half- and full-duplex relay based cooperative communications and its potential challenges and open issues using Markov chains," *IET Commun.*, vol. 13, no. 11, pp. 1537–1550, 2019.

[3] Y. Jing and H. Jafarkhani, "Network beamforming using relays with perfect channel information," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2499–2517, Jun. 2009.

[4] V. Havary-Nassab, S. Shahbazpanahi, and A. Grami, "Optimal distributed beamforming for two-way relay networks," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1238–1250, Mar. 2010.

[5] V. H. Nassab, S. Shahbazpanahi, A. Grami, and Z. Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4306–4316, Sep. 2008.

[6] R. Vahidnia and S. Shahbazpanahi, "Multi-carrier asynchronous bi-directional relay networks: Joint subcarrier power allocation and network beamforming," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 3796–3812, Aug. 2013.

[7] R. Vahidnia and S. ShahbazPanahi, "Single-carrier equalization for asynchronous two-way relay networks," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5793–5808, Nov. 2014.

[8] R. Vahidnia, S. Shahbazpanahi, and A. Minasian, "Pre-channel equalization and distributed beamforming in asynchronous single-carrier bi-directional relay networks," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3968–3983, Aug. 2016.

[9] M. Harris, "Tech giants race to build orbital internet [news]," *IEEE Spectrum*, vol. 55, no. 6, pp. 10–11, Jun. 2018.

[10] Y. Xiao, T. Zhang, and L. Liu, "Addressing subnet division based on geographical information for satellite-ground integrated network," *IEEE Access*, vol. 6, pp. 75 824–75 833, 2018.

[11] X. Li, C. Xing, Y.-C. Wu, and S. Chan, "Timing estimation and resynchronization for amplify-and-forward communication systems," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2218–2229, Apr. 2010.

[12] Y. Yao and X. Dong, "Low-complexity timing synchronization for decode-and-forward cooperative communication systems with multiple relays," *IEEE Trans. Veh. Technol.*, vol. 62, no. 6, pp. 2865–2871, Jul. 2013.
[13] S. Chang and B. Kelley, "An efficient time synchronization scheme for broadband two-way relaying networks based on physical-layer network coding," *IEEE Commun. Lett.*, vol. 16, no. 9, pp. 1416–1419, Sep. 2012.
[14] A. Nasir, H. Mehrpouyan, S. Blostein, S. Durrani, and R. Kennedy, "Timing and carrier synchronization with channel estimation in multi-relay cooperative networks," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 793–811, Feb. 2012.
[15] M. Bhatnagar, A. Hjorungnes, and M. Debbah, "A simple scheme for delay-tolerant decode-and-forward based cooperative communication," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 498–502.
[16] M. Bhatnagar, A. Hjorungnes, and M. Debbah, "Delay-tolerant decode-and-forward based cooperative communication over Ricean channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1277–1282, Apr. 2010.
[17] H.-M. Wang, X.-G. Xia, and Q. Yin, "A linear analog network coding for asynchronous two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3630–3637, Dec. 2010.
[18] Z. Fang, F. Liang, Z. Wang, and D. Chen, "Low complexity timing estimation and resynchronization for asynchronous bi-directional communications with multiple antenna relay," *Int. J. Commun. Syst.*, vol. 28, pp. 1140–1150, Jan. 2014.
[19] Z. Fang, F. Liang, S. Zhang, and T. Peng, "A cyclic shift relaying scheme for asynchronous two-way relay networks," *Wireless Pers. Commun.*, vol. 71, pp. 2863–2876, Aug. 2013.
[20] H. Wang, X.-G. Xia, and Q. Yin, "Computationally efficient equalization for asynchronous cooperative communications with multiple frequency offsets," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 648–655, Feb. 2009.
[21] R. AliHemmati and S. Shahbazpanahi, "Sum-rate optimal network beamforming and subcarrier power allocation for multi-carrier asynchronous two-way relay networks," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 4129–4143, Aug. 2015.
[22] J. Mirzaei and S. Shahbazpanahi, "On achievable SNR region for multi-user multi-carrier asynchronous bidirectional relay networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3219–3230, Jun. 2015.
[23] M. Askari and S. Shahbazpanahi, "On sum-rate maximization approach to network beamforming and power allocation for asynchronous single-carrier two-way relay networks," *IEEE Access*, vol. 5, pp. 13 699–13 711, 2017.
[24] S. Bastanirad, S. Shahbazpanahi, R. Rahimi, and A. Grami, "On total transmission power minimization approach to decentralized beamforming in single-carrier asynchronous bidirectional relay-assisted communication networks," *IEEE Access*, vol. 7, pp. 30 966–30 979, 2019.
[25] F. E. Dorcheh and S. Shahbazpanahi, "Jointly optimal pre- and post-channel equalization and distributed beamforming in asynchronous bidirectional relay networks," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4593–4608, Sep. 2017.
[26] Z. Wang and G. Giannakis, "Wireless multicarrier communications," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 29–48, May 2000.
[27] S. Shahbazpanahi and M. Dong, "Achievable rate region under joint distributed beamforming and power allocation for two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 4026–4037, Nov. 2012.

Razgar Rahimi was born in Saqqez, Iran. He received the B.Sc. degree from the Iran University of Science and Technology in 2005, and the M.Sc. degree from Shahed University, Tehran, Iran, in 2010. He received the Ph.D. degree from the University Of Ontario Institute Of Technology, Oshawa, ON, Canada in 2018. He was a Telecommunication Engineer with Iran Control and Communications Systems Supply Company from 2005 to 2014. He is currently a Lecturer at the Faculty of Engineering and Applied Science, the University of Ontario Institute of Technology, Oshawa, Ontraio, Canada. His current research interests include signal processing, cognitive radio networks, and cooperative wireless communications.

Shahram ShahbazPanahi (Senior Member, IEEE) was born in Sanandaj, Kurdistan, Iran. He received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1992, 1994, and 2001, respectively. From September 1994 to September 1996, he was an Instructor with the Department of Electrical Engineering, Razi University, Kermanshah, Iran. From July 2001 to March 2003, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. From April 2003 to September 2004, he was a Visiting Researcher with the Department of Communication Systems, University of Duisburg-Essen, Duisburg, Germany. From September 2004 to April 2005, he was a Lecturer and Adjunct Professor with the Department of Electrical and Computer Engineering, McMaster University. In July 2005, he joined the Faculty of Engineering and Applied Science, University of Ontario Institute of Technology, Oshawa, ON, Canada, where he currently holds a Professor position. His research interests include statistical and array signal processing, space-time adaptive processing, detection and estimation, multi-antenna, multi-user, and cooperative communications, spread spectrum techniques, DSP programming, and hardware/real-time software design for telecommunication systems. Dr. Shahbazpanahi has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS. He has also served as a Senior Area Editor for the IEEE SIGNAL PROCESSING LETTERS. He was an elected member of the Sensor Array and Multichannel (SAM) Technical Committee of the IEEE Signal Processing Society. He has received several awards, including the Early Researcher Award from Ontario's Ministry of Research and Innovation, the NSERC Discovery Grant (three awards), the Research Excellence Award from the Faculty of Engineering and Applied Science, the University of Ontario Institute of Technology, and the Research Excellence Award, Early Stage, from the University of Ontario Institute of Technology.

Björn Ottersten (Fellow, IEEE) was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1990. He has held research positions with the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, the Katholieke Universiteit Leuven, Leuven, Belgium, and the University of Luxembourg, Luxembourg. From 1996 to 1997, he was the Director of Research with ArrayComm, Inc., a start-up in San Jose, CA, USA, based on his patented technology. In 1991, he was appointed Professor of signal processing with the Royal Institute of Technology (KTH), Stockholm, Sweden. Dr. Ottersten has been Head of the Department for Signals, Sensors, and Systems, KTH, and Dean of the School of Electrical Engineering, KTH. He is currently the Director for the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. He is a recipient of the IEEE Signal Processing Society Technical Achievement Award and the European Research Council advanced research grant twice. He has co-authored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, 2013, and 2019, and 8 IEEE conference papers best paper awards. He has been a Board Member of IEEE Signal Processing Society, the Swedish Research Council and currently serves of the boards of EURASIP and the Swedish Foundation for Strategic Research. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the Editorial Board of the *IEEE Signal Processing Magazine*. He is currently a member of the editorial boards of *IEEE Open Journal of Signal Processing, EURASIP Signal Processing Journal, EURASIP Journal of Advances Signal Processing and Foundations and Trends of Signal Processing. He is a fellow of EURASIP.*