

Power-Efficient Distributed Beamforming for Full-Duplex MIMO Relaying Networks

Xiaofei Xu, Xiang Chen, *Member, IEEE*, Ming Zhao, *Member, IEEE*, Shidong Zhou, *Member, IEEE*, Chong-Yung Chi, *Senior Member, IEEE*, and Jing Wang, *Member, IEEE*

Abstract—Multiple-input–multiple-output (MIMO) full-duplex relaying (FDR) has been considered an efficient technique to provide coverage to users where their direct links from the base station (BS) are too weak for reliable signal reception. However, when multiple MIMO full-duplex relays are deployed in a network, the signal reception quality relies on the effective suppression of multiple types of interference. In this paper, the distributed beamforming is studied for the MIMO FDR network by formulating a power minimization problem with a nonstrict convex objective function (total transmit power of both the BS and all the relays in the network) under individual user rate constraints. We come up with two iterative distributed beamforming algorithms, Algorithm 1 for relays equipped with single receive antenna and Algorithm 2 with multiple receive antennas. The former can yield a global optimal solution of the power minimization problem, whereas the latter can yield only a local optimal solution due to conservative successive convex approximations (SCAs) performed at each iteration, and a rigorous analysis on the upper bounds of step sizes is proposed to guarantee their convergence. The proposed two algorithms only require local information exchange between relays and hence are scalable for different network sizes and topologies. An “early termination” strategy in the operation of the proposed two algorithms is also presented to acquire an acceptable transmit power solution with less computation time consumption and thus suitable for realistic applications. Finally, some simulation results

Manuscript received August 31, 2015; revised December 26, 2015, February 14, 2016, and March 15, 2016; accepted March 18, 2016. Date of publication April 8, 2016; date of current version February 10, 2017. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2012CB316002; by the National Natural Science Foundation of China under Grant 61501527 and Grant 61201192; by the State High-Tech Development Plan of China (863 Program) under Grant 2014AA01A707; by the National S&T Major Project under Grant 2014ZX03003003-002; by Huawei Technologies; by the National Science Council under Grant NSC-102-2221-E-007-019-MY3; by the China Electronics Technology Group Corporation No. 54 Institute; and by Shenzhen Basic Research Foundation under Grant JCYJ20150630153033410. This paper has been presented in part at the IEEE Global Communications Conference, San Diego, CA, USA, December 2015, and at the IEEE Military Communications Conference, Tampa, FL, USA, October 2015. The review of this paper was coordinated by Dr. S. Sun.

X. Xu, M. Zhao, S. Zhou, and J. Wang are with the Wireless and Mobile Communication Technology R&D Center, Research Institute of Information Technology, Tsinghua University, Beijing 100084, China (e-mail: xu-xf10@mails.tsinghua.edu.cn).

X. Chen is with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China, and also with the Key Laboratory of Electronic Design Automation (EDA), Research Institute of Tsinghua University in Shenzhen (RITS), Shenzhen 518075, China (e-mail: chenxiang@mail.sysu.edu.cn).

C.-Y. Chi is with the Institute of Communications Engineering and the Department of Electrical Engineering, National Tsinghua University, Hsinchu 30013, Taiwan (e-mail: cychi@ee.nthu.edu.tw).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVT.2016.2551376

are provided to demonstrate that the proposed two algorithms perform well and significantly better than the existing state-of-the-art scheme reported in the work of Lee and Shin.

Index Terms—Distributed beamforming, full-duplex relays, interference suppression, multiple-input multiple-output (MIMO), power efficiency.

I. INTRODUCTION

RELAYING is a power-efficient technique to guarantee uniform coverage in cellular systems for areas where the direct link from the base station (BS) is weak, e.g., cell edge and deep shadow fading areas [1], [2]. With no need of any wired backhaul from the BS, this cost-effective technique has been incorporated into several wireless communication standards, including Third-Generation Partnership Project Long Term Evolution Advanced and IEEE 802.16, and is expected to be deployed extensively in the upcoming fifth-generation systems [3], [4] and infrastructure-based vehicular networks [5]. To avoid self-interference (SI) from its own transmitter, traditional relay works in half-duplex mode, resulting in up to 50% spectral efficiency loss, which severely limits its realistic applications. In recent years, due to the advances of SI cancellation techniques [6], full-duplex relaying (FDR), which can recover the spectral efficiency loss of half-duplex relaying, has attracted extensive research interests.

For a single FDR node, its performance is highly determined by the capability of SI cancellation. Existing SI cancellation techniques can be roughly grouped into either time-domain or spatial-domain techniques. Time-domain techniques are performed in digital/analog hardware by subtracting the known SI from the received signal [7], [8], whereas residual SI (RSI) still remains after cancellation due to hardware imperfections. On the other research track, multiple-input–multiple-output (MIMO) FDR is proposed to suppress SI by exploiting different spatial-domain techniques at the relay. Antenna placement and cross-polarization are investigated in [9] and [10], respectively. A variety of beamforming-based techniques are proposed in [11]–[14]. By proper design of the relay’s transmit and/or receive antenna beam patterns, the intended signal is protected from the overwhelming SI in orthogonal signal subspaces. It is worth noting that with the help of massive arrays, such orthogonality can be achieved with simple beamforming designs [15], despite that the limited size of a cellular relay node might remain a problem. Recently, the two parallel research tracks have started to merge as in [16] and [17], where spatial-domain and time-domain SI cancellation are jointly considered

in the system design. This design makes better use of the spatial dimensions by allowing a low level of SI received at the relay antennas and further mitigated by subsequent time-domain SI cancelation modules, such that the beamformers can be optimized more flexibly to yield a better end-to-end performance.

With all the development in recent years, state-of-the-art MIMO FDR becomes one of the best options to guarantee uniform coverage in cellular systems that is hard to achieve with wired backhaul. In a realistic cellular system, multiple out-of-coverage areas may exist in each cell due to heavy path loss and/or complicated propagation environment; hence, multiple relays need to be deployed in these areas to guarantee coverage for the randomly distributed users. Typically, the BS communicates with its relays using the point-to-multipoint mode, and each relay serves its attached users, respectively [2], [18], [19], which constitutes a multirelay multiuser network. Due to the broadcast nature of wireless channel, interference management has been a critical issue to improve performance in cellular networks. When multiple FDRs are deployed in the network, the interference management is further complicated by new interference. In addition to SI at each FDR, both interrelay interference (IRI) and multiuser interference (MUI) should be considered [20], [21]. IRI occurs because the transmission of an FDR will interfere with the reception of nearby FDRs. This interference cannot be mitigated in the same fashion as time-domain SI cancelation since the transmitted signal of each FDR is only known to itself. Among existing research articles on FDR-aided cellular networks, [19] studies resource allocation in an MIMO-FDR-aided OFDMA system, where different end-to-end links are isolated via orthogonal frequency allocation. This scheme avoids IRI and MUI, but the gains from frequency reuse among geographically distributed relays are lost. In [22]–[24], different MIMO beamforming schemes are proposed to suppress SI and MUI in single-FDR systems. When these schemes are extended to practical multi-FDR multiuser networks, severe performance loss may occur because IRI is not considered. To the authors' knowledge, an efficient scheme for the suppression of IRI, SI, and MUI in the practical multi-FDR multiuser networks is still absent in the literature, thereby motivating the study of this paper.

On the other hand, most existing works on MIMO FDR networks provide centralized solutions for different design goals in the presence of SI and MUI. Designed for single-relay systems, these centralized solutions are less viable in realistic relay-aided cellular systems, where multiple relays are deployed at different locations in a cell. For such systems, the development of centralized algorithms requires a central node to aggregate all channel state information (CSI) and then find an optimal optimization, naturally not only incurring heavy signaling overhead but also the network scalability limitations. Therefore, a distributed algorithm is particularly favorable.

In this paper, the distributed beamforming design for decode-and-forward MIMO FDR networks with the presence of IRI, SI, and MUI is studied, with some benefits from existing time-domain SI cancelation. When the relays are deployed for coverage enhancement, the rate demand of each user should be maintained. Furthermore, the total transmit power at the BS

and relays is a proper metric of the overall system performance [25], [26] since reduced transmit power means not only reduced operational cost but also lower intercell interference. Therefore, an optimization problem for minimizing the transmit power at both the relays and the BS is formulated under individual user-rate constraints. Inspired by [27] and [28], a single-layer iterative proximal decomposition (SLIPD) method is proposed for solving nonstrict convex optimization problems by which two distributed beamforming algorithms are developed. These two algorithms, which are designed for maximum transmission power saving using the spatial dimensions of MIMO FDR, only require local information exchange among the relays and hence are robust to different network sizes and topologies. The main contributions of this paper are summarized as follows.

- Spatial interference suppression is investigated for MIMO FDR networks, where all the major types of interference, i.e., IRI, SI, and MUI, are efficiently suppressed via distributed beamforming. With the aim to improve overall system performance, an optimization problem for minimizing the total transmit power is formulated, and two iterative distributed beamforming algorithms are developed via the judicious use of spatial degrees of freedom of MIMO FDR.
- A SLIPD method is proposed to design the first distributed algorithm (Algorithm 1) for the case that each relay is equipped with single receive antenna. Algorithm 1 can yield a global optimal solution to the preceding power minimization problem. Then, the SLIPD method combined with successive convex approximations (SCAs) is further applied to the design of the second distributed algorithm (Algorithm 2) for the case of multiple receive antennas at each relay, which can yield at least a local optimal solution of the power minimization problem.
- A rigorous convergence analysis is conducted for the SLIPD method, providing an upper bound for the step sizes with which the convergence of the proposed two algorithms can be guaranteed, together with an early termination strategy for fast convergence and acceptable solution accuracy for practical applications.

The remainder of this paper is organized as follows. In Section II, the system model is presented and then the distributed beamforming problem is formulated. In Section III, we first study the case with single receive antenna at each relay and develop the SLIPD method from which Algorithm 1 is designed for solving this problem, followed by a rigorous convergence analysis. In Section IV, the SLIPD method is extended to the more general case with multiple receive antennas at each relay, and then in Algorithm 2, a dual-layer distributed algorithm is presented with the aid of SCAs. Simulation results are presented to support the efficacy of the proposed two algorithms in Section V. Finally, some conclusions are provided in Section VI.

Notation: Matrices and vectors are represented with uppercase and lowercase bold letters, respectively. $(\cdot)^C$, $(\cdot)^T$, and $(\cdot)^H$ denote conjugate, transpose, and conjugate transpose, respectively. $|\cdot|$ denotes the determinant of a matrix, the

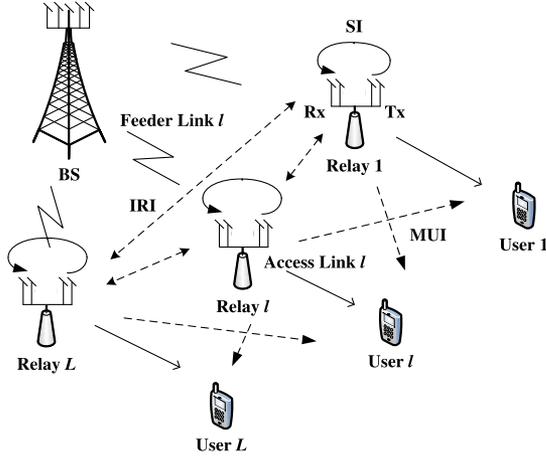


Fig. 1. MIMO-FDR-aided cellular network with IRI, SI, and MUI.

2-norm of a vector, or the absolute value of a complex or real number for notational simplicity. $\|\cdot\|_F$ and $\text{Tr}(\cdot)$ denote the Frobenius norm and the trace of a matrix, respectively. $\text{vec}(\cdot)$ denotes a column vector by stacking the columns of a matrix. \mathbf{I} and \mathbf{O} denote an identity matrix and a zero matrix of proper dimension, respectively. $E\{\cdot\}$ denotes the mathematical expectation of a random variable. $\text{Re}\{\cdot\}$ denotes the real part of a complex number. $\mathbb{C}^{M \times N}$ denotes the set of $M \times N$ complex matrices. $\mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$ denotes the circularly symmetric and zero-mean complex normal distribution with correlation matrix $\sigma_n^2 \mathbf{I}$. The generalized inequalities $\mathbf{X} \succeq \mathbf{0}$ and $\mathbf{X} \succ \mathbf{0}$ denote that the matrix \mathbf{X} is positive semi-definite, and positive definite, respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We study a relay-aided downlink cellular network with one N_B -antenna BS, L MIMO FDRs, and L single-antenna users, as shown in Fig. 1. Each FDR is equipped with N_t transmit antennas and N_r receive antennas, and performs decode-and-forward transmission to avoid noise/interference amplification. IRI is considered between any two relays. Similar to [18] and [19], we assume that the direct links between the BS and the users are ignored due to severe path loss and blockage. Typically, the users are dispersed geographically in a cell and associated with different relays. While more than one user may exist in each relay's coverage, we consider a simplified model in which each relay selects only one user to serve¹: the data intended for user i is first received and decoded by relay i on feeder link i , and then reencoded and forwarded to user i on access link i . To account for the case that users can possibly be located at the coverage overlap of adjacent relays, MUI is considered for each user from relays other than its serving relay. This model can be prospectively extended to the more general scenario with multiple users served by each relay in a multicell network, which is left for future work.

¹Our model can be extended to the scenario for a relay to serve multiple users, and scheduling is also necessary to improve the system performance. Selecting a "good" user for each relay, e.g., a user with strong access link, is a simple yet efficient scheme.

On each feeder link, multistream beamforming² is applied at BS. Assuming linear beamformers, the received signal at relay i is given by

$$\begin{aligned} \mathbf{y}_{R,i} = & \underbrace{\mathbf{H}_{BR,i} \mathbf{U}_{B,i} \mathbf{x}_{B,i}}_{\text{desired signal}} + \underbrace{\sum_{l=1, l \neq i}^L \mathbf{H}_{BR,i} \mathbf{U}_{B,l} \mathbf{x}_{B,l}}_{\text{interference across feeder links}} \\ & + \underbrace{\mathbf{H}_{RR,i,i} \mathbf{u}_{R,i} x_{R,i}}_{\text{self-interference}} + \underbrace{\sum_{l=1, l \neq i}^L \mathbf{H}_{RR,i,l} \mathbf{u}_{R,l} x_{R,l}}_{\text{interrelay interference}} + \mathbf{n}_{R,i} \end{aligned} \quad (1)$$

where $\mathbf{x}_{B,i} \in \mathbb{C}^{M_i \times 1}$ and $\mathbf{U}_{B,i} \in \mathbb{C}^{N_B \times M_i}$ are the transmitted symbol vector and the beamforming matrix at the BS for relay i , respectively; M_i is the number of streams transmitted on feeder link i ; $\mathbf{H}_{BR,i} \in \mathbb{C}^{N_r \times N_B}$ represents the channel matrix of the i th feeder link; $x_{R,i} \in \mathbb{C}$ and $\mathbf{u}_{R,i} \in \mathbb{C}^{N_t \times 1}$ are the transmitted symbol and the beamforming vector of the i th relay, respectively; and $\mathbf{H}_{RR,i,l} \in \mathbb{C}^{N_r \times N_t}$ represents the SI channel matrix of the i th relay for $i = l$ and the IRI channel matrix from relay l to relay i for $i \neq l$. Without loss of generality, assume $E\{\mathbf{x}_{B,i} \mathbf{x}_{B,i}^H\} = \mathbf{I}$ and $E\{|x_{R,i}|^2\} = 1$. $\mathbf{n}_{R,i} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$ is zero-mean additive white Gaussian noise (AWGN), where σ_n^2 is the noise power at any receive antenna.

In our proposal, block diagonalization (BD) [29] is adopted at the BS for transmission to different relays.³ Assuming that the BS has a sufficient number of antennas, the beamformer $\mathbf{U}_{B,i}$ can be decomposed as $\mathbf{U}_{B,i} = \mathbf{U}_{B,i}^{BD} \bar{\mathbf{U}}_{B,i}$, where $\mathbf{U}_{B,i}^{BD} \in \mathbb{C}^{N_B \times \bar{N}_{B,i}}$ and $\bar{\mathbf{U}}_{B,i} \in \mathbb{C}^{\bar{N}_{B,i} \times M_i}$, with $M_i \leq \bar{N}_{B,i} \leq N_B$, satisfy $\mathbf{H}_{BR,l} \mathbf{U}_{B,i}^{BD} = \mathbf{O}$ for $i \neq l$ and $(\mathbf{U}_{B,i}^{BD})^H \mathbf{U}_{B,i}^{BD} = \mathbf{I}$. In our design, only the low-dimension beamformer $\bar{\mathbf{U}}_{B,i}$ needs to be optimized at the relay nodes and fed back to the BS in the end of the design. The matrix $\mathbf{U}_{B,i}^{BD}$ only needs to be computed once and used in all transmissions since the channels on feeder links can be viewed as time invariant due to fixed antenna locations. For simplicity, we assume that $\bar{N}_{B,i} = \bar{N}_B$ for all i is the number of effective antennas at the BS for feeder link i . With BD at the BS, the interference across different feeder links are completely eliminated, and the transmit power at the BS for the i th feeder link is given by $E\{\mathbf{U}_{B,i} \mathbf{x}_{B,i} \mathbf{x}_{B,i}^H\} = \text{Tr}(\mathbf{Q}_{B,i})$, where $\mathbf{Q}_{B,i} = \bar{\mathbf{U}}_{B,i} \bar{\mathbf{U}}_{B,i}^H$. Then, (1) can be rewritten as

$$\begin{aligned} \mathbf{y}_{R,i} = & \bar{\mathbf{H}}_{BR,i} \bar{\mathbf{U}}_{B,i} \mathbf{x}_{B,i} + \mathbf{H}_{RR,i,i} \mathbf{u}_{R,i} x_{R,i} \\ & + \sum_{l=1, l \neq i}^L \mathbf{H}_{RR,i,l} \mathbf{u}_{R,l} x_{R,l} + \mathbf{n}_{R,i} \end{aligned} \quad (2)$$

where $\bar{\mathbf{H}}_{BR,i} = \mathbf{H}_{BR,i} \mathbf{U}_{B,i}^{BD}$ is the effective channel matrix of the i th feeder link.

²Also known as "precoding" in some of the literature. For simplicity, we use the term "beamforming" regardless of single-stream or multistream transmission in this paper.

³The performance loss of BD is limited compared with the optimal dirty paper coding scheme [30], but the complexity and overhead are much lower for the former.

Upon reception at relay i , certain SI cancelation processing (see [6] for different SI cancelation techniques) is performed to remove the SI part in the received signal. Note that the IRI [the last underbrace term in (1)] cannot be treated as SI since $x_{R,l} \forall l \neq i$ is not known to relay i . Due to hardware imperfections [16], SI cannot be completely eliminated, and the RSI, which is denoted by z_{RSI} , is modeled as AWGN with $z_{\text{RSI}} \sim \mathcal{CN}(\mathbf{0}, \gamma \mathbf{H}_{RR,i,i} \mathbf{Q}_{R,i} \mathbf{H}_{RR,i,i}^H)$, where $\mathbf{Q}_{R,i} = \mathbf{u}_{R,i} \mathbf{u}_{R,i}^H$, and the hardware-dependent parameter γ characterizes the power level of RSI with $0 < \gamma < 1$. This RSI model with proper value of γ has been considered a reasonable approximation of the models in [16] and [17]. Assuming that each relay can acquire accurate CSI on all links from/to itself (including SI and IRI channels), the data rate in terms of nats/s/Hz on the i th feeder link is known as

$$r_{BR,i} = \ln \frac{|\mathbf{I} + \tilde{\mathbf{H}}_{BR,i} \mathbf{Q}_{B,i} \tilde{\mathbf{H}}_{BR,i}^H + \mathbf{Z}_{R,i}|}{|\mathbf{I} + \mathbf{Z}_{R,i}|} \quad (3)$$

where

$$\mathbf{Z}_{R,i} = \sum_{l=1}^L \tilde{\mathbf{H}}_{RR,i,l} \mathbf{Q}_{R,l} \tilde{\mathbf{H}}_{RR,i,l}^H \quad (4)$$

is the covariance of SI and IRI received at relay i

$$\tilde{\mathbf{H}}_{RR,i,l} = \begin{cases} \sigma_n^{-1} \mathbf{H}_{RR,i,l}, & \text{if } l \neq i \text{ (Channel matrix of IRI)} \\ \sqrt{\gamma} \sigma_n^{-1} \mathbf{H}_{RR,i,i}, & \text{if } l = i \text{ (Channel matrix of SI)} \end{cases}$$

and $\tilde{\mathbf{H}}_{BR,i} = \sigma_n^{-1} \mathbf{H}_{BR,i}$ for all i are the equivalent channel gains normalized by the standard deviation of noise.

On the access link, the received signal by user i is

$$y_{U,i} = \underbrace{\mathbf{h}_{RU,i,i} \mathbf{u}_{R,i} x_{R,i}}_{\text{desired signal}} + \underbrace{\sum_{l=1, l \neq i}^L \mathbf{h}_{RU,i,l} \mathbf{u}_{R,l} x_{R,l}}_{\text{multiuser interference}} + n_{U,i} \quad (5)$$

where $\mathbf{h}_{RU,i,j} \in \mathbb{C}^{1 \times N_t}$ represents the MUI channel vector between relay j and user i for $j \neq i$, and the i th access channel vector between relay i and user i for $j = i$, respectively. $n_{U,i}$ is AWGN with $n_{U,i} \sim \mathcal{CN}(0, \sigma_n^2)$. Assume that the channel between each relay–user pair is time invariant within a block of end-to-end transmissions [16]. The achievable rate on access link i is given by

$$r_{RU,i} = \ln \left(1 + \frac{\tilde{\mathbf{h}}_{RU,i,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,i,i}^H}{1 + z_{U,i}} \right) \quad (6)$$

where

$$z_{U,i} = \sum_{l=1, l \neq i}^L \tilde{\mathbf{h}}_{RU,i,l} \mathbf{Q}_{R,l} \tilde{\mathbf{h}}_{RU,i,l}^H \quad (7)$$

is the power of the MUI received at user i , and $\tilde{\mathbf{h}}_{RU,i,l} = \sigma_n^{-1} \mathbf{h}_{RU,i,l}$ is the equivalent channel vector between relay l and user i normalized by the noise standard deviation.

The total transmit power of the system is given by $\sum_{i=1}^L (\text{Tr}(\mathbf{Q}_{B,i}) + \text{Tr}(\mathbf{Q}_{R,i}))$. By substituting (4) and (7) into

(3) and (6), respectively, it can be seen that a proper joint design of the beamformers $\{\mathbf{u}_{R,i}, i = 1, \dots, L\}$ and $\{\bar{\mathbf{U}}_{B,i}, i = 1, \dots, L\}$ will mitigate IRI, SI, and MUI simultaneously to improve power efficiency of the system. Given individual rate demands $\{r_i, i = 1, \dots, L\}$, the joint optimization problem can be formulated as

$$\min_{\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i} \forall i\}} \sum_{i=1}^L (\text{Tr}(\mathbf{Q}_{B,i}) + \text{Tr}(\mathbf{Q}_{R,i})) \quad (8a)$$

$$\text{s.t.} \quad r_{BR,i} \geq r_i \forall i \quad (8b)$$

$$r_{RU,i} \geq r_i \forall i \quad (8c)$$

$$\mathbf{Q}_{B,i} \succeq \mathbf{0}, \mathbf{Q}_{R,i} \succeq \mathbf{0} \forall i \quad (8d)$$

where (8b) and (8c) indicate that the data rates on feeder link i and access link i should be no less than user i 's demand, so that uninterrupted end-to-end transmission can be maintained; in (8d), semi-definite relaxation [31] has been applied to $\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i} \forall i\}$. Hence, as the optimal $\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i} \forall i\}$ is obtained, the associated beamformers $\bar{\mathbf{U}}_{B,i}$ and $\mathbf{u}_{R,i}$ can be obtained via either eigenvalue decomposition (EVD) or Gaussian randomization.

To design a distributed algorithm for solving (8), two difficulties should be tackled. First, it can be observed that the problem (8) is nonconvex due to the nonconvex constraints in (8b) for $N_r > 1$. Obtaining a globally optimal solution will involve sophisticated optimization theory in addition to prohibitively high complexity. Second, even with $N_r = 1$, direct application of conventional duality-based method [32] may fail to converge in the designed distributed algorithm, due to nonstrict convexity of the objective function in (8a) (with nonunique solutions); the iterates may oscillate between iterations, and even a local optimum can hardly be achieved (see [27], [28] for details). In Sections III and IV, we will first resolve the nonstrict convexity issue for the $N_r = 1$ case and then extend our results to the case of $N_r > 1$.

III. DISTRIBUTED ALGORITHM FOR $N_r = 1$

Here, a distributed algorithm based on a SLIPD method is proposed to solve the nonstrictly convex problem (8) for $N_r = 1$ and to pave the way for the $N_r > 1$ case.⁴

A. Algorithm Design: The SLIPD Method

For $N_r = 1$, the transmission on each feeder link reduces to single-stream beamforming, with transmitted symbol $x_{B,i}$ and beamformer $\mathbf{u}_{B,i} \in \mathbb{C}^{N_B \times 1}$ on feeder link i . We also use the row vectors $\tilde{\mathbf{h}}_{BR,i}$ and $\tilde{\mathbf{h}}_{RR,i,l}$, to replace $\tilde{\mathbf{H}}_{BR,i}$ and

⁴The study for $N_r = 1$ is not only essential for the more complex $N_r > 1$ case but also is supported by realistic reasons. As indicated by [7] and [8], each receive antenna of MIMO FDR entails an independent processing chain with relatively high hardware complexity for SI cancelation; therefore, $N_r = 1$ is the most cost-effective. Moreover, the computational complexity with $N_r > 1$ is much higher than that of $N_r = 1$, as can be seen later. Meanwhile, the benefits of more receive antennas are limited by highly correlated receive antennas (due to the size limitation of relay) and interference-limited environment [33]. All these reasons make $N_r = 1$ a favorable option in realistic systems.

$\tilde{\mathbf{H}}_{RR,i,l}$ involved in (8), respectively. With the use of (4) and (7) explicitly in problem (8), we obtain the following equivalent problem:

$$\min_{\substack{\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, \\ z_{R,i}, z_{U,i} \forall i\}}} \sum_{i=1}^L (\text{Tr}(\mathbf{Q}_{B,i}) + \text{Tr}(\mathbf{Q}_{R,i})) \quad (9a)$$

$$\text{s.t.} \quad \ln \left(1 + \frac{\tilde{\mathbf{h}}_{BR,i} \mathbf{Q}_{B,i} \tilde{\mathbf{h}}_{BR,i}^H}{1 + z_{R,i}} \right) \geq r_i \forall i \quad (9b)$$

$$\sum_{i=1}^L \tilde{\mathbf{h}}_{RR,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RR,l,i}^H = z_{R,l} \forall l \quad (9c)$$

$$\ln \left(1 + \frac{\tilde{\mathbf{h}}_{RU,i,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,i,i}^H}{1 + z_{U,i}} \right) \geq r_i \forall i \quad (9d)$$

$$\sum_{i=1, i \neq l}^L \tilde{\mathbf{h}}_{RU,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,l,i}^H = z_{U,l} \forall l \quad (9e)$$

$$\mathbf{Q}_{B,i} \succeq \mathbf{0}, \mathbf{Q}_{R,i} \succeq \mathbf{0} \forall i. \quad (9f)$$

Note that (9) is actually a convex problem since the nonconvex constraints (9b) and (9d) can be converted into convex constraints, and surely one can design a centralized algorithm by applying generic methods, e.g., an interior-point method. Furthermore, in the reformulated problem (9), all the constraints except (9c) and (9e) are uncoupled with respect to different i and therefore can be evaluated locally on individual end-to-end links. The coupling of variables belonging to different end-to-end links only exists in a linear form as in (9c) and (9e), which facilitates the design of a distributed algorithm to be presented in the following.

According to standard proximal point method [34], quadratic terms $\sum_{i=1}^L (c_i/2) (\|\mathbf{Q}_{B,i} - \mathbf{W}_{B,i}\|_F^2 + \|\mathbf{Q}_{R,i} - \mathbf{W}_{R,i}\|_F^2 + |z_{R,i} - v_{R,i}|^2 + |z_{U,i} - v_{U,i}|^2)$ are first added to the objective function in (9a) to recover strict convexity, where $\mathbf{W}_{B,i}$, $\mathbf{W}_{R,i}$, $v_{R,i}$, and $v_{U,i}$ are the auxiliary variables associated with the original variables $\mathbf{Q}_{B,i}$, $\mathbf{Q}_{R,i}$, $z_{R,i}$, and $z_{U,i}$, respectively, and $c_i/2 > 0$ is the weight of those quadratic terms associated with the i th end-to-end link. The resulting regularized problem is given by

$$\min \sum_{i=1}^L \left[\text{Tr}(\mathbf{Q}_{B,i}) + \text{Tr}(\mathbf{Q}_{R,i}) + \frac{c_i}{2} \left(\|\mathbf{Q}_{B,i} - \mathbf{W}_{B,i}\|_F^2 + \|\mathbf{Q}_{R,i} - \mathbf{W}_{R,i}\|_F^2 + |z_{R,i} - v_{R,i}|^2 + |z_{U,i} - v_{U,i}|^2 \right) \right] \quad (10a)$$

$$\text{s.t.} \quad \sum_{i=1}^L \tilde{\mathbf{h}}_{RR,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RR,l,i}^H = z_{R,l} \forall l \quad (10b)$$

$$\sum_{i=1, i \neq l}^L \tilde{\mathbf{h}}_{RU,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,l,i}^H = z_{U,l} \forall l \quad (10c)$$

$$\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, z_{R,i}, z_{U,i}\} \in \mathcal{D}_i \forall i \quad (10d)$$

where the minimization is performed with respect to $\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, z_{R,i}, z_{U,i}, \mathbf{W}_{B,i}, \mathbf{W}_{R,i}, v_{R,i}, v_{U,i} \forall i\}$, and \mathcal{D}_i is the convex constraint set for $\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, z_{R,i}, z_{U,i}\}$ defined by the uncoupling constraints (9b), (9d), (9f) for each i .

It can be easily verified that the optimal solution of (10) is also optimal to (9). As discussed in [34], an iterative dual-layer distributed algorithm can be directly derived to solve (10) and is outlined as Algorithm P.

Algorithm P Primitive Dual-Layer Algorithm for $N_r = 1$

Initialize the auxiliary variables arbitrarily and the dual variables to zero.

For outer iteration n ,

Step 1. Fix the auxiliary variables in (10), and solve it with respect to the original variables, using the subgradient method [32]. The subgradient method, which is an iterative method by itself, carries out the inner iterations of this algorithm.

Step 2. Update the auxiliary variables by assigning the values of the corresponding original variables in Step 1 to them.

Stop until the total transmit power of (9) satisfies a predefined convergence criterion.

In Step 1, the dual decomposition [32] is applied to deal with the coupling constraints. The corresponding partial Lagrangian of (10) subject to constraint (10d) is given by

$$\begin{aligned} \mathcal{L}(\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, z_{R,i}, z_{U,i}, \varphi_{R,i}, \varphi_{U,i}, \mathbf{W}_{B,i}, \mathbf{W}_{R,i}, v_{R,i}, v_{U,i} \forall i) \\ \triangleq \sum_{i=1}^L \mathcal{L}_i(\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, z_{R,i}, z_{U,i}, \varphi_{R,i}, \varphi_{U,i}, \\ \mathbf{W}_{B,i}, \mathbf{W}_{R,i}, v_{R,i}, v_{U,i}) \end{aligned} \quad (11)$$

where

$$\begin{aligned} \mathcal{L}_i(\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, z_{R,i}, z_{U,i}, \varphi_{R,i}, \varphi_{U,i}, \mathbf{W}_{B,i}, \mathbf{W}_{R,i}, v_{R,i}, v_{U,i}) \\ = \text{Tr}(\mathbf{Q}_{B,i}) + \text{Tr}(\mathbf{Q}_{R,i}) + \sum_{l=1}^L \varphi_{R,l} \tilde{\mathbf{h}}_{RR,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RR,l,i}^H \\ - \varphi_{R,i} z_{R,i} + \sum_{l=1, l \neq i}^L \varphi_{U,l} \tilde{\mathbf{h}}_{RU,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,l,i}^H - \varphi_{U,i} z_{U,i} \\ + \frac{c_i}{2} \left(\|\mathbf{Q}_{B,i} - \mathbf{W}_{B,i}\|_F^2 + \|\mathbf{Q}_{R,i} - \mathbf{W}_{R,i}\|_F^2 \right. \\ \left. + |z_{R,i} - v_{R,i}|^2 + |z_{U,i} - v_{U,i}|^2 \right). \end{aligned} \quad (12)$$

With the above decomposition, Step 1 solves (10) (with the auxiliary variables fixed) by solving the following problem in an iterative manner:

$$\max_{\{\varphi_{R,i}, \varphi_{U,i} \forall i\}} \left\{ \min_{\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, z_{R,i}, z_{U,i}\} \in \mathcal{D}_i \forall i} \mathcal{L}(\cdot) \right\} \quad (13)$$

where the arguments in the partial Lagrangian (11) are omitted for simplicity. In each inner iteration, the inner minimization of (13) (a convex problem) is solved first, and then the dual variables $\varphi_{R,l}$ and $\varphi_{U,l}$ are updated by the subgradient method (with the subgradient $\sum_{i=1}^L \tilde{\mathbf{h}}_{RR,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RR,l,i}^H - z_{R,l}$ for the former and $\sum_{i=1, i \neq l}^L \tilde{\mathbf{h}}_{RU,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,l,i}^H - z_{U,l}$ for the latter). The inner iterations repeat until the iterates for the original variables converge. Each outer iteration ends up by assigning the iterates for the original variables to the corresponding iterates for the auxiliary variables.

Remark 1: Algorithm P allows a distributed implementation. In Step 1, the inner minimization in (13) can be performed at respective relays [cf. (12)], and the following update of the dual variables can be performed by exchanging $\{\mathbf{Q}_{R,i} \forall i\}$ with neighboring relays. Then, the updated dual variables are broadcast to neighboring relays for the next inner iteration. In Step 2, the auxiliary variables can be also updated at respective relays.

Nevertheless, such a dual-layer iterative algorithm may not be very suitable for online distributed implementation since a great many inner iterations may be required for convergence of the dual variables. In view of this, we propose a SLIPD method to construct a single-layer distributed algorithm, for which the auxiliary variables are updated without the need of convergence of the dual variables (i.e., the number of the inner iterations is set to one).

Before proceeding, we introduce some notations to facilitate description of our proposal. Let $\mathbf{Q}_B = [\mathbf{Q}_{B,1}, \dots, \mathbf{Q}_{B,L}]$, $\mathbf{Q}_R = [\mathbf{Q}_{R,1}, \dots, \mathbf{Q}_{R,L}]$, $\mathbf{z}_R = [z_{R,1}, \dots, z_{R,L}]^T$, and $\mathbf{z}_U = [z_{U,1}, \dots, z_{U,L}]^T$. Similarly, $\mathbf{W}_B, \mathbf{W}_R, \mathbf{v}_R, \mathbf{v}_U$ can be defined for the associated auxiliary variables. Let $\mathcal{D} = \mathcal{D}_1 \times \dots \times \mathcal{D}_L$. Define the vectors

$$\mathbf{q} = \begin{bmatrix} \text{vec}(\mathbf{Q}_B) \\ \text{vec}(\mathbf{Q}_R) \\ \mathbf{z}_R \\ \mathbf{z}_U \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} \text{vec}(\mathbf{W}_B) \\ \text{vec}(\mathbf{W}_R) \\ \mathbf{v}_R \\ \mathbf{v}_U \end{bmatrix}$$

$\varphi = [\varphi_{R,1}, \dots, \varphi_{R,L}, \varphi_{U,1}, \dots, \varphi_{U,L}]^T$, and the function

$$f(\mathbf{q}) = \begin{cases} \sum_{i=1}^L (\text{Tr}(\mathbf{Q}_{B,i}) + \text{Tr}(\mathbf{Q}_{R,i})), & \text{if } \mathbf{q} \in \mathcal{D} \\ +\infty, & \text{otherwise} \end{cases} \quad (14)$$

where we slightly abuse the use of the set \mathcal{D} as the constraint set for the elements in \mathbf{q} . The function (14) can be viewed as an extended-value function of the original objective function in (9a). Define the $2L \times (\tilde{N}_B^2 + N_t^2 + 2)L$ coefficient matrix as

$$\mathbf{E} = \begin{bmatrix} \mathbf{O}_{L \times \tilde{N}_B^2} & \mathbf{E}_{RR} & -\mathbf{I}_{L \times L} & \mathbf{O}_{L \times L} \\ \mathbf{O}_{L \times \tilde{N}_B^2} & \mathbf{E}_{RU} & \mathbf{O}_{L \times L} & -\mathbf{I}_{L \times L} \end{bmatrix} \quad (15)$$

where $\mathbf{E}_{RR} \in \mathbb{C}^{L \times N_t^2 L}$ is given by

$$\begin{aligned} & \mathbf{E}_{RR} \\ &= \begin{bmatrix} \text{vec}(\tilde{\mathbf{h}}_{RR,1,1}^H \tilde{\mathbf{h}}_{RR,1,1}) & \dots & \text{vec}(\tilde{\mathbf{h}}_{RR,L,1}^H \tilde{\mathbf{h}}_{RR,L,1}) \\ \vdots & & \vdots \\ \text{vec}(\tilde{\mathbf{h}}_{RR,1,L}^H \tilde{\mathbf{h}}_{RR,1,L}) & \dots & \text{vec}(\tilde{\mathbf{h}}_{RR,L,L}^H \tilde{\mathbf{h}}_{RR,L,L}) \end{bmatrix}^H \end{aligned} \quad (16)$$

and $\mathbf{E}_{RU} \in \mathbb{C}^{L \times N_t^2 L}$ can be divided into $L \times L$ equal-sized blocks, where the (l, i) th block, i.e., $e_{RU,l,i} \in \mathbb{C}^{1 \times N_t^2}$, is given by

$$e_{RU,l,i} = \begin{cases} \text{vec}(\tilde{\mathbf{h}}_{RU,l,i}^H \tilde{\mathbf{h}}_{RU,l,i})^H, & \text{if } l \neq i \\ \mathbf{0}_{1 \times N_t^2}, & \text{otherwise.} \end{cases} \quad (17)$$

With the given notations, the partial Lagrangian (11) can be rewritten as

$$\mathcal{L}(\mathbf{q}, \varphi, \mathbf{w}) = f(\mathbf{q}) + \mathbf{q}^T \mathbf{E}^T \varphi^C + \frac{1}{2}(\mathbf{q} - \mathbf{w})^T \mathbf{V}(\mathbf{q} - \mathbf{w})^C \quad (18)$$

where \mathbf{V} is a diagonal matrix consisting of the weights $\{c_i \forall i\}$. The j th diagonal element \mathbf{V} is the weight c_i if the j th element in \mathbf{q} is an element in the matrices $\mathbf{Q}_{B,i}$ or $\mathbf{Q}_{R,i}$, or an element of $\{z_{R,i}, z_{U,i}\}$.

Next, we propose a proximal decomposition method, which solves (10) by updating the auxiliary variables and the dual variables alternatively. Specifically, the auxiliary variables are updated immediately after every single update of the dual variables, exhibiting a single-layer iterative structure. This is different from Algorithm P, where dual variables are updated until convergence before updating the auxiliary variables. This SLIPD method is used to construct Algorithm 1, with \mathbf{w} treated as the primal variable as in the standard proximal point method, and the original variable \mathbf{q} acts as an intermediate variable. Similar to Algorithm P, Algorithm 1 can also be implemented in a distributed manner: In Step 1, (19) is equivalent to minimizing the partial Lagrangian given by (11) and can be decomposed as in (12) and solved at respective relays. Next, $\{\mathbf{Q}_{R,i} \forall i\}$ is exchanged between neighboring relays such that the subgradients of the dual variables [as presented below (13)] can be calculated. By the subgradient method, the dual variables are updated using (20), completing the update of the dual variables. In Step 2, the updated dual variables are first broadcast to neighboring relays. With the new iterates for the dual variables, (21) is solved distributedly as (19), and then the obtained solution for the original variables are assigned to the corresponding iterates for the primal variables at each relay as given by (22).

Algorithm 1 SLIPD-based Distributed Beamforming Algorithm for $N_r = 1$

Initialize $\mathbf{w}(0)$ arbitrarily and $\varphi(0)$ to zero.

For iteration t ,

Step 1. Dual variable update:

Given $\varphi(t)$ and fixing $\mathbf{w}(t)$, solve

$$\mathbf{q}(t) = \arg \min_{\mathbf{q}} \mathcal{L}(\mathbf{q}, \varphi(t), \mathbf{w}(t)) \quad (19)$$

distributedly (cf. Remark 1) at the relays.

Update the dual variables

$$\varphi(t+1) = \varphi(t) + \mathbf{A} \mathbf{E} \mathbf{q}(t) \quad (20)$$

where \mathbf{A} is a diagonal matrix with positive diagonal elements $\alpha_{R,1}, \dots, \alpha_{R,L}, \alpha_{U,1}, \dots, \alpha_{U,L}$, with $\alpha_{R,l}$ and $\alpha_{U,l}$ denoting the step sizes used at relay l to update the corresponding dual variables.

Step 2. Primal variable update:

With updated dual variables, solve

$$\mathbf{q}'(t) = \arg \min_{\mathbf{q}} \mathcal{L}(\mathbf{q}, \varphi(t+1), \mathbf{w}(t)) \quad (21)$$

distributedly (cf. Remark 1). Update the primal variables

$$\mathbf{w}(t+1) = \mathbf{q}'(t). \quad (22)$$

Stop until the total transmit power of (9) satisfies a predefined convergence criterion.

Quite different from the standard proximal point method (Algorithm P), the conditions for the convergence of Algorithm 1

remain to be established. Another related issue is the choice of the step sizes in (20). In realistic applications, larger step sizes are favorable for faster convergence, while still preserving stable convergence. For these purposes, a convergence analysis is conducted in the sequel for the proposed distributed beamforming algorithm.

B. Convergence Analysis

In [35], it has been proven that for a Hermitian matrix, all the components have *independent* differentials, and formal derivatives/gradients should be used in optimization problems with complex-valued matrix variables. Real scalars can be viewed as 1-D Hermitian matrices and the above results also apply. Following the mathematical results in [35], for $\mathbf{q} = \arg \min_{\mathbf{q}} L(\mathbf{q}, \boldsymbol{\varphi}, \mathbf{w})$, by the first-order optimality condition, there exists a subgradient $\nabla f(\mathbf{q})$ [27], satisfying

$$\nabla f(\mathbf{q}) + \mathbf{E}^H \boldsymbol{\varphi} + \mathbf{V}(\mathbf{q} - \mathbf{w}) = \mathbf{0} \quad (23)$$

and that

$$\nabla f(\mathbf{w}^*) + \mathbf{E}^H \boldsymbol{\varphi}^* = \mathbf{0} \quad (24)$$

holds for any stationary point $(\boldsymbol{\varphi}^*, \mathbf{w}^*)$ of Algorithm 1 since $\mathbf{w}^* = \mathbf{q}^*$.

The convergence of Algorithm 1 to the global optimum of (9) is guaranteed by the following theorem.

Theorem 1: In the proposed SLIPD method for solving the nonstrict convex problem (9), if the step sizes $\alpha_{R,l}$, $\alpha_{U,l}$ satisfy

$$\max_l \alpha_{R,l} < \frac{2}{3\|\mathbf{E}\|_F^2} \min_l c_l \quad (25)$$

$$\max_l \alpha_{U,l} < \frac{2}{3\|\mathbf{E}\|_F^2} \min_l c_l \quad (26)$$

where \mathbf{E} is defined in (15), Algorithm 1 converges with the obtained \mathbf{q} being a global optimum of problem (9).

The proof of Theorem 1 relies on the results given in the following lemma.

Lemma 1: For the optimization framework in (19)–(22), the two inequalities hold

$$\begin{aligned} & (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}(t))^H \mathbf{E} \mathbf{V}^{-1} \mathbf{E}^H (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}(t)) \\ & \geq (\mathbf{q}'(t) - \mathbf{q}(t))^H \mathbf{V} (\mathbf{q}'(t) - \mathbf{q}(t)). \end{aligned} \quad (27)$$

$$\begin{aligned} & (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}(t))^H \mathbf{E} \mathbf{V}^{-1} \mathbf{E}^H (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}(t)) \\ & \geq -4\text{Re} \left\{ (\nabla f(\mathbf{q}'(t)) - \nabla f(\mathbf{w}^*))^H (\mathbf{q}(t) - \mathbf{w}^*) \right\}. \end{aligned} \quad (28)$$

The proof of Lemma 1 is given in Appendix A.

Remark 2: The first inequality (27) indicates that, within one iteration of Algorithm 1, if the step sizes are sufficiently small (meaning $\boldsymbol{\varphi}(t+1)$ and $\boldsymbol{\varphi}(t)$ are not far apart), then $\mathbf{q}'(t)$ and $\mathbf{q}(t)$ are not far apart. Furthermore, once $\boldsymbol{\varphi}(t)$ converges, then $\mathbf{q}'(t) = \mathbf{q}(t)$ (cf. Algorithm 1), and then (28) reduces to the common result (53) in Appendix A for convex functions.

Now, we are ready to prove Theorem 1.

Proof for Theorem 1: First, a metric is needed to measure the “distance” between the intermediate solution $(\boldsymbol{\varphi}(t), \mathbf{w}(t))$ and

a stationary point $(\boldsymbol{\varphi}^*, \mathbf{w}^*)$ of Algorithm 1. For this purpose, we define the Lyapunov function

$$\lambda(t) = \|\mathbf{w}(t) - \mathbf{w}^*\|_V + \|\boldsymbol{\varphi}(t) - \boldsymbol{\varphi}^*\|_A \quad (29)$$

where $\|\mathbf{w}\|_V = \mathbf{w}^H \mathbf{V} \mathbf{w}$ and $\|\boldsymbol{\varphi}\|_A = \boldsymbol{\varphi}^H \mathbf{A}^{-1} \boldsymbol{\varphi}$. Similar to [27] and [28], it can be proved that Algorithm 1 converges and yields an optimal solution of problem (9) if $\lambda(t)$ converges.

Next, let us derive the condition that Lyapunov function (29) converges. First, consider

$$\begin{aligned} & \|\boldsymbol{\varphi}(t) - \boldsymbol{\varphi}^*\|_A - \|\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}^*\|_A \\ & = \|\boldsymbol{\varphi}(t) - \boldsymbol{\varphi}(t+1) + \boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}^*\|_A - \|\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}^*\|_A \\ & = \|\boldsymbol{\varphi}(t) - \boldsymbol{\varphi}(t+1)\|_A \\ & \quad - 2\text{Re} \left\{ (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}^*)^H \mathbf{A}^{-1} (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}(t)) \right\} \\ & = \|\boldsymbol{\varphi}(t) - \boldsymbol{\varphi}(t+1)\|_A - 2\text{Re} \left\{ (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}^*)^H \mathbf{E} \mathbf{q}(t) \right\} \\ & \quad (\text{by (20)}). \end{aligned} \quad (30)$$

Then, we have

$$\begin{aligned} & \lambda(t+1) - \lambda(t) \\ & = \|\mathbf{w}(t+1) - \mathbf{w}^*\|_V + \|\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}^*\|_A - \|\mathbf{w}(t) - \mathbf{w}^*\|_V \\ & \quad - \|\boldsymbol{\varphi}(t) - \boldsymbol{\varphi}^*\|_A \\ & = -\|\boldsymbol{\varphi}(t) - \boldsymbol{\varphi}(t+1)\|_A + 2\text{Re} \left\{ (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}^*)^H \mathbf{E} \mathbf{q}(t) \right\} \\ & \quad + \|\mathbf{q}'(t) - \mathbf{w}^*\|_V - \|\mathbf{w}(t) - \mathbf{w}^*\|_V \quad (\text{by (30) and (22)}) \\ & = -\|\boldsymbol{\varphi}(t) - \boldsymbol{\varphi}(t+1)\|_A + \|\mathbf{q}'(t) - \mathbf{w}^*\|_V - \|\mathbf{w}(t) - \mathbf{w}^*\|_V \\ & \quad + 2\text{Re} \left\{ (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}^*)^T \mathbf{E} (\mathbf{q}(t) - \mathbf{w}^*) \right\} \end{aligned} \quad (31)$$

where we have used the result $\mathbf{E} \mathbf{w}^* = \mathbf{0}$ [due to constraints (10b) and (10c) and the definition of \mathbf{E} in (15)] in the last equality. From (21), (23), and (24), we have

$$\begin{aligned} & \mathbf{E}^H (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}^*) \\ & = -(\nabla f(\mathbf{q}'(t)) - \nabla f(\mathbf{w}^*)) - \mathbf{V} (\mathbf{q}'(t) - \mathbf{w}(t)). \end{aligned} \quad (32)$$

Substituting (32) into (31), it follows that

$$\begin{aligned} & \lambda(t+1) - \lambda(t) \\ & = -\|\boldsymbol{\varphi}(t) - \boldsymbol{\varphi}(t+1)\|_A + \|\mathbf{q}'(t) - \mathbf{w}^*\|_V - \|\mathbf{w}(t) - \mathbf{w}^*\|_V \\ & \quad - 2\text{Re} \left\{ (\mathbf{q}'(t) - \mathbf{w}(t))^H \mathbf{V} (\mathbf{q}(t) - \mathbf{w}^*) \right\} \\ & \quad - 2\text{Re} \left\{ (\nabla(\mathbf{q}'(t)) - \nabla f(\mathbf{w}^*))^H (\mathbf{q}(t) - \mathbf{w}^*) \right\}. \end{aligned} \quad (33)$$

With proper mathematical manipulations, it can be proved that the sum of the second, third, and fourth terms in (33) can be simplified as follows:

$$\begin{aligned} & \|\mathbf{q}'(t) - \mathbf{w}^*\|_V - \|\mathbf{w}(t) - \mathbf{w}^*\|_V \\ & \quad - 2\text{Re} \left\{ (\mathbf{q}'(t) - \mathbf{w}(t))^H \mathbf{V} (\mathbf{q}(t) - \mathbf{w}^*) \right\} \\ & = \|\mathbf{q}'(t) - \mathbf{q}(t) + \mathbf{q}(t) - \mathbf{w}^*\|_V - \|\mathbf{w}(t) - \mathbf{q}(t) + \mathbf{q}(t) - \mathbf{w}^*\|_V \\ & \quad - 2\text{Re} \left\{ (\mathbf{q}'(t) - \mathbf{w}(t))^H \mathbf{V} (\mathbf{q}(t) - \mathbf{w}^*) \right\} \\ & = \|\mathbf{q}'(t) - \mathbf{q}(t)\|_V - \|\mathbf{w}(t) - \mathbf{q}(t)\|_V. \end{aligned} \quad (34)$$

Hence, (33) becomes

$$\begin{aligned} \lambda(t+1) - \lambda(t) &= -\|\varphi(t) - \varphi(t+1)\|_A + \|\mathbf{q}'(t) - \mathbf{q}(t)\|_V \\ &\quad - \|\mathbf{w}(t) - \mathbf{q}(t)\|_V \\ &\quad - 2\text{Re} \left\{ (\nabla f(\mathbf{q}'(t)) - \nabla f(\mathbf{w}^*))^H (\mathbf{q}(t) - \mathbf{w}^*) \right\}. \end{aligned} \quad (35)$$

Invoking Lemma 1, we have

$$\begin{aligned} \lambda(t+1) - \lambda(t) &\leq -\|\varphi(t) - \varphi(t+1)\|_A - \|\mathbf{w}(t) - \mathbf{q}(t)\|_V \\ &\quad + \frac{3}{2} (\varphi(t+1) - \varphi(t))^H \mathbf{E} \mathbf{V}^{-1} \mathbf{E}^H (\varphi(t+1) - \varphi(t)) \\ &= -(\varphi(t+1) - \varphi(t))^H \mathbf{B} (\varphi(t+1) - \varphi(t)) \\ &\quad - \|\mathbf{w}(t) - \mathbf{q}(t)\|_V \end{aligned} \quad (36)$$

where $\mathbf{B} = \mathbf{A}^{-1} - (3/2)\mathbf{E}\mathbf{V}^{-1}\mathbf{E}^H$. Obviously, if \mathbf{B} is positive definite, then $\lambda(t)$ decreases monotonically by (36), implying that it converges since $\lambda(t) \geq 0$. Furthermore, when $\lambda(t)$ converges, it can be easily seen from (36) that $\varphi(t)$ converges to some $\tilde{\varphi}$ as $t \rightarrow \infty$, and that $\mathbf{w}(t) = \mathbf{q}'(t) = \mathbf{q}(t)$ converges to some $\tilde{\mathbf{w}}$ [cf. (21) and (22)]. Therefore, $(\tilde{\varphi}, \tilde{\mathbf{w}})$ is a stationary point of Algorithm 1, and we can take $\varphi^* = \tilde{\varphi}$ and $\mathbf{w}^* = \tilde{\mathbf{w}}$, which give an optimal solution to (9). A sufficient condition for positive definite \mathbf{B} is provided in the following lemma (with the proof presented in Appendix B).

Lemma 2: For arbitrary positive constant a , if

$$\max_j \alpha_j < \frac{a}{\|\mathbf{E}\|_F^2} \min_j c_j$$

where $\{\alpha_j \forall j\}$ and $\{c_j \forall j\}$ are the diagonal elements of positive definite diagonal matrices \mathbf{A} and \mathbf{V} , respectively, then

$$a\mathbf{A}^{-1} \succ \mathbf{E}\mathbf{V}^{-1}\mathbf{E}^H.$$

Applying the given inequality with $a = 2/3$ to the matrix \mathbf{B} in (36), Theorem 1 is proved immediately.

Remark 3: Theorem 1 provides a sufficient condition [cf. (25) and (26)] under which Algorithm 1 converges. By our simulation experience, in some cases, step sizes larger than those in (25) and (26) can also yield a global optimum of (9). However, it is difficult to find a weaker condition under which Algorithm 1 can converge uniformly for all the channel realizations. Nevertheless, a proper selection of step sizes, e.g., decreasing step sizes, that satisfy (25) and (26) after a certain number of iterations, can be used for faster convergence in practice.

Remark 4: It can be proved that $\|\mathbf{E}\|_F^2$ is indeed the sum of squared power gains on all interference links, plus a constant due to the identity matrices involved. This means the maximum convergence speed of Algorithm 1 is inversely proportional to the total squared interference power gains in the network. Moreover, it provides a simpler way to calculate the step sizes. While the SI and IRI channels have a larger coherence time due to the antennas' height and fixed positions, the MUI channels will be different in successive blocks of end-to-end transmissions, requiring frequent step size calculations. However, an estimate of the MUI power gains can be used for a conservative step size for each relay, so that the convergence speed can be maintained with much lower overhead.

Remark 5: In practice, the coherence time of the channels between relays and users can be small; therefore, it is highly preferred if we can obtain a feasible solution with acceptable total transmission power in a limited number of iterations. This can be done by periodically solving an extra set of optimization problems during the operation of Algorithm 1. Suppose that $\{\hat{\mathbf{Q}}_{B,i}, \hat{\mathbf{Q}}_{R,i}\}$ is the local solution at relay i after some iterations, and that

$$\begin{aligned} \hat{z}_{\text{IRI},i} &= \sum_{l=1, l \neq i}^L \tilde{\mathbf{h}}_{RR,i,l} \hat{\mathbf{Q}}_{R,l} \tilde{\mathbf{h}}_{RR,i,l}^H \\ \hat{z}_{U,i} &= \sum_{l=1, l \neq i}^L \tilde{\mathbf{h}}_{RU,i,l} \hat{\mathbf{Q}}_{R,l} \tilde{\mathbf{h}}_{RU,i,l}^H \end{aligned}$$

respectively, are acquired by collecting the corresponding variables $(\hat{\mathbf{Q}}_{R,l}$ and $\tilde{\mathbf{h}}_{RU,i,l})$ from neighboring relays. Then, the extra convex problem at relay i is given by

$$\begin{aligned} \min_{\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}\}} & \text{Tr}(\mathbf{Q}_{B,i}) + \text{Tr}(\mathbf{Q}_{R,i}) \\ \text{s.t.} & \ln \left(1 + \frac{\tilde{\mathbf{h}}_{BR,i} \mathbf{Q}_{B,i} \tilde{\mathbf{h}}_{BR,i}^H}{1 + \hat{z}_{\text{IRI},i} + \tilde{\mathbf{h}}_{RR,i,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RR,i,i}^H} \right) \geq r_i \\ & \tilde{\mathbf{h}}_{RR,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RR,l,i}^H \leq \tilde{\mathbf{h}}_{RR,l,i} \hat{\mathbf{Q}}_{R,i} \tilde{\mathbf{h}}_{RR,l,i}^H \quad \forall l \neq i \\ & \ln \left(1 + \frac{\tilde{\mathbf{h}}_{RU,i,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,i,i}^H}{1 + \hat{z}_{U,i}} \right) \geq r_i \\ & \tilde{\mathbf{h}}_{RU,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,l,i}^H \leq \tilde{\mathbf{h}}_{RU,l,i} \hat{\mathbf{Q}}_{R,i} \tilde{\mathbf{h}}_{RU,l,i}^H \quad \forall l \neq i \\ & \mathbf{Q}_{B,i} \succeq \mathbf{0}, \mathbf{Q}_{R,i} \succeq \mathbf{0}. \end{aligned} \quad (37)$$

If all the relays declare feasibility for their own local problem (37), then the solutions of (37) $\forall i$, provide an acceptable feasible solution of (9). Otherwise, the operation of Algorithm 1 should continue.

Remark 6: The proposed beamforming design for the MIMO FDR network can be extended to a more general case of multiple users served by each relay. For this case, extra MUI from each relay need to be added in the received signal model given by (5), followed by some modifications of the coefficient matrix \mathbf{E} in (15) that is used in the partial Lagrangian in (18), and then the ensuing convergence analysis also needs to be modified to finish the corresponding distributed beamforming algorithm (counterpart of the proposed Algorithm 1) for this case.

C. Complexity and Overhead Analysis

The overall complexity of Algorithm 1 is almost fully contributed by the solution of (19) and (21), both of which are equivalent to minimizing (11) with respect to the original variables $\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, z_{R,i}, z_{U,i} \forall i\}$ under constraint (10d). For the distributed implementation, the i th relay minimizes (12) with respect to its local original variables $\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, z_{R,i}, z_{U,i}\}$ under the local constraint set \mathcal{D}_i . Each local optimization problem belongs to the category of nonlinear semi-definite programming, and a solution can be achieved with complexity $O((\bar{m}\bar{n}^2 + \bar{m}^2\bar{n}^2 + \bar{n}^3)/\varepsilon)$, where \bar{m} is the maximum dimension of the constraints involving generalized inequalities, \bar{n} is the total number of components (scalars)

in the optimization variables, and ε is the accuracy tolerance [36]. Specifically, for each local optimization problem, $\bar{m} = \max\{\bar{N}_B, N_t\}$ and $\bar{n} = \bar{N}_B^2 + N_t^2 + 2$. The convergence rate of Algorithm 1 is difficult to analyze theoretically; however, a linear or faster convergence rate can be observed in all our simulations. Therefore, with ε' as the accuracy tolerance for Algorithm 1, the overall complexity of Algorithm 1 is estimated as $O(L(\bar{m}\bar{n}^2 + \bar{m}^2\bar{n}^2 + \bar{n}^3)/\varepsilon')$. It can be seen that the complexity of Algorithm 1 is heavily dependent on \bar{N}_B and N_t . Benefiting from BD at the BS, \bar{N}_B can be designed much smaller than N_B . As for N_t , although more transmit antennas at each relay mean more spatial degrees of freedom to suppress interference, the complexity increases quite rapidly. It implies that the number of transmit antennas at each relay should be properly selected as a tradeoff between complexity and end-to-end performance of the network.

The communication overhead of Algorithm 1 is another concern, particularly when the exchange of local information occupies wireless resources. In each iteration, the i th relay needs to collect $\{\mathbf{Q}_{R,l} \forall l \neq i\}$ to update its local dual variables $\varphi_{R,i}$ and $\varphi_{U,i}$, and then broadcast the local dual variables to other relays. In fact, the update of dual variables requires only the interference power, i.e., relay i needs only $\tilde{\mathbf{h}}_{RR,i,l} \mathbf{Q}_{R,l} \tilde{\mathbf{h}}_{RR,i,l}^H$ and $\tilde{\mathbf{h}}_{RU,i,l} \mathbf{Q}_{R,l} \tilde{\mathbf{h}}_{RU,i,l}^H \forall l \neq i$. Suppose that $\tilde{\mathbf{h}}_{RR,i,l}$ and $\tilde{\mathbf{h}}_{RU,i,l}$ are known to relay l (e.g., by channel reciprocity), relay l may send only the interference power terms (2 scalars) to relay i ; therefore, the overhead for the update of dual variables is $2L(L-1)$ scalars for L relays. Moreover, the overhead for broadcasting the local dual variables is $2L$ scalars for L relays. Therefore, the total overhead in each iteration of Algorithm 1 is $2L^2$ scalars. In realistic systems with a large number of relays, IRI and MUI mainly come from adjacent relays, and the interference from nonadjacent relays can be ignored, thereby further reducing the overhead.

IV. DISTRIBUTED ALGORITHM FOR $N_r > 1$

Here, we aim to design a distributed algorithm to solve problem (8) for the case of $N_r > 1$ for better end-to-end performance. Following the reformulation similar to (9) for the uncoupling variables, the resulting problem for $N_r > 1$ is given by

$$\min_{\substack{\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, \\ \mathbf{Z}_{R,i}, z_{U,i} \forall i\}}} \sum_{i=1}^L (\text{Tr}(\mathbf{Q}_{B,i}) + \text{Tr}(\mathbf{Q}_{R,i})) \quad (38a)$$

$$\text{s.t.} \quad \ln \frac{\mathbf{I} + \tilde{\mathbf{H}}_{BR,i} \mathbf{Q}_{B,i} \tilde{\mathbf{H}}_{BR,i}^H + \mathbf{Z}_{R,i}}{|\mathbf{I} + \mathbf{Z}_{R,i}|} \geq r_i \forall i \quad (38b)$$

$$\sum_{i=1}^L \tilde{\mathbf{H}}_{RR,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{H}}_{RR,l,i}^H = \mathbf{Z}_{R,l} \forall l \quad (38c)$$

$$\ln \left(1 + \frac{\tilde{\mathbf{h}}_{RU,i,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,i,i}^H}{1 + z_{U,i}} \right) \geq r_i \forall i \quad (38d)$$

$$\sum_{i=1, i \neq l}^L \tilde{\mathbf{h}}_{RU,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,l,i}^H = z_{U,l} \forall l \quad (38e)$$

$$\mathbf{Q}_{B,i} \succeq \mathbf{0}, \mathbf{Q}_{R,i} \succeq \mathbf{0} \forall i. \quad (38f)$$

Based on this formulation, we present how to solve the problem via SCA and the SLIPD method developed in Section III, in Section IV-A and B, respectively.

A. Successive Convex Approximation

It can be easily seen that problem (38) is nonconvex due to the nonconvex rate constraint (38b). Here, we apply SCA (SCA) to approximate (38) by a series of convex problems, so that a suboptimal solution of (38) can be obtained by successively solving these convex problems. Moreover, the SCA facilitates the application of the preceding SLIPD method to the distributed beamforming algorithm design for $N_r > 1$ to be presented in the following.

Because $\ln|\mathbf{I} + \mathbf{X}|$ is a concave function of $\mathbf{X} \succeq \mathbf{0}$, the first-order inequality

$$\ln|\mathbf{I} + \mathbf{X}| \leq \ln|\mathbf{I} + \mathbf{X}_0| + \text{Tr}((\mathbf{I} + \mathbf{X}_0)^{-1}(\mathbf{X} - \mathbf{X}_0)) \quad (39)$$

is true for $\mathbf{X} \succeq \mathbf{0}$, and the equality holds when $\mathbf{X} = \mathbf{X}_0$. The right-hand side of (39) is an affine function and a tight upper bound of the concave function $\ln|\mathbf{I} + \mathbf{X}|$ that will be used for a conservative convex approximation to the nonconvex constraint (38b) in the ensuing iterative beamforming algorithm design. Let $\{\mathbf{Q}_{B,i}^{(n-1)}, \mathbf{Q}_{R,i}^{(n-1)}, \mathbf{Z}_{R,i}^{(n-1)}, z_{U,i}^{(n-1)} \forall i\}$ denote the feasible point of (38) obtained at the $(n-1)$ th iteration. Since (38b) can be equivalently represented as

$$\ln|\mathbf{I} + \tilde{\mathbf{H}}_{BR,i} \mathbf{Q}_{B,i} \tilde{\mathbf{H}}_{BR,i}^H + \mathbf{Z}_{R,i}| - \ln|\mathbf{I} + \mathbf{Z}_{R,i}| \geq r_i \quad (40)$$

applying the upper bound in (39) with $\mathbf{X} = \mathbf{Z}_{R,i}$ and $\mathbf{X}_0 = \mathbf{Z}_{R,i}^{(n-1)}$ to $\ln|\mathbf{I} + \mathbf{Z}_{R,i}|$ in (40) yields the following conservative convex constraint:

$$\ln|\mathbf{I} + \tilde{\mathbf{H}}_{BR,i} \mathbf{Q}_{B,i} \tilde{\mathbf{H}}_{BR,i}^H + \mathbf{Z}_{R,i}| - \ln|\mathbf{I} + \mathbf{Z}_{R,i}^{(n-1)}| - \text{Tr} \left[(\mathbf{I} + \mathbf{Z}_{R,i}^{(n-1)})^{-1} (\mathbf{Z}_{R,i} - \mathbf{Z}_{R,i}^{(n-1)}) \right] \geq r_i \forall i. \quad (41)$$

Thus, we come up with a new convex problem given by

$$\begin{aligned} & g^{(n)} \left(\mathbf{Q}_{B,i}^{(n)}, \mathbf{Q}_{R,i}^{(n)}, \mathbf{Z}_{R,i}^{(n)}, z_{U,i}^{(n)} \forall i \right) \\ & \triangleq \min_{\substack{\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, \\ \mathbf{Z}_{R,i}, z_{U,i} \forall i\}}} \sum_{i=1}^L (\text{Tr}(\mathbf{Q}_{B,i}) + \text{Tr}(\mathbf{Q}_{R,i})) \\ & \text{s.t.} \quad (41), (38c), (38d), (38e), (38f). \end{aligned} \quad (42)$$

Note that problem (42) is convex with a feasible set, which is denoted $\mathcal{C}^{(n)}$, that is also a subset of the feasible set, which is denoted as \mathcal{C} , of problem (38); therefore, $\{\mathbf{Q}_{B,i}^{(n-1)}, \mathbf{Q}_{R,i}^{(n-1)}, \mathbf{Z}_{R,i}^{(n-1)}, z_{U,i}^{(n-1)} \forall i\} \in \mathcal{C}^{(n)} \subset \mathcal{C}$. Let $g(\cdot)$ denote the objective function of (38) (i.e., the total power). Then, we have

$$\begin{aligned} & g \left(\mathbf{Q}_{B,i}^{(n-1)}, \mathbf{Q}_{R,i}^{(n-1)}, \mathbf{Z}_{R,i}^{(n-1)}, z_{U,i}^{(n-1)} \forall i \right) \\ & \geq g^{(n)} \left(\mathbf{Q}_{B,i}^{(n)}, \mathbf{Q}_{R,i}^{(n)}, \mathbf{Z}_{R,i}^{(n)}, z_{U,i}^{(n)} \forall i \right) \quad (\text{by (42)}) \\ & = g \left(\mathbf{Q}_{B,i}^{(n)}, \mathbf{Q}_{R,i}^{(n)}, \mathbf{Z}_{R,i}^{(n)}, z_{U,i}^{(n)} \forall i \right) \quad (\text{since } \mathcal{C}^{(n)} \subset \mathcal{C}). \end{aligned} \quad (43)$$

With the obtained solution of (42), $\{\mathbf{Q}_{B,i}^{(n)}, \mathbf{Q}_{R,i}^{(n)}, \mathbf{Z}_{R,i}^{(n)}, z_{U,i}^{(n)} \forall i\}$, used as a new feasible point, one can update the convex approximation constraint (41) with respect to this feasible point, and then solve (42) at the next iteration, thereby making $\{g(\mathbf{Q}_{B,i}^{(n)}, \mathbf{Q}_{R,i}^{(n)}, \mathbf{Z}_{R,i}^{(n)}, z_{U,i}^{(n)} \forall i)\}$ to monotonically decrease according to the given inequality. Hence, by repeating this procedure, one can obtain a local optimal solution of (38) as $\{g(\mathbf{Q}_{B,i}^{(n)}, \mathbf{Q}_{R,i}^{(n)}, \mathbf{Z}_{R,i}^{(n)}, z_{U,i}^{(n)} \forall i)\}$ converges.

As for the convex optimization problem (42) at each iteration, the preceding SLIPD method can also be applied to solve it in a similar distributed manner and will be detailed in the ensuing Section IV-B. The resulting distributed beamforming design for $N_r > 1$ is summarized in Algorithm 2.

Algorithm 2 SLIPD-Based Distributed Beamforming Algorithm for $N_r > 1$

For outer iteration n ,

Step 1. Solve problem (42) with the SLIPD method presented in Section IV-B, which carries out the inner iterations of this algorithm. Let $\mathbf{Z}_{R,i}^{(n)} \forall i$ be the obtained optimal IRI covariance.

Step 2. Update the conservative convex constraint (41) by replacing $\mathbf{Z}_{R,i}^{(n-1)} \forall i$ with $\mathbf{Z}_{R,i}^{(n)} \forall i$.

Stop until the total transmit power of (38) satisfies a predefined convergence criterion.

Remark 7: An initial feasible point of problem (38) is needed to initialize Algorithm 2 that can yield a local optimum of (38). The simple zero-forcing (ZF) beamformer, which nullifies strong MUI and meanwhile satisfies the rate constraint of its own user on the access link, can be used for initialization since $\mathbf{Z}_{R,l}^{(0)} \forall l$ can be easily calculated via local information exchange.

B. SLIPD Method

Here, we apply the SLIPD method to solve (42) distributedly, which is the core step (Step 1) in Algorithm 2.

Denote \mathcal{D}_i as the constraint set for $\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, \mathbf{Z}_{R,i}, z_{U,i}\}$ defined by the decoupled constraints (41), (38d), and (38f) for all i . Similar to (10) in the $N_r = 1$ case, problem (42) can also be handled by adding a quadratic regularization term, although the only difference is that the matrix variable $\mathbf{V}_{R,i}$ is the auxiliary variable for $\mathbf{Z}_{R,i}$. The corresponding partial Lagrangian can be expressed as

$$\begin{aligned} \mathcal{L}(\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, \mathbf{Z}_{R,i}, z_{U,i}, \Phi_{R,i}, \varphi_{U,i}, \mathbf{W}_{B,i}, \mathbf{W}_{R,i}, \mathbf{V}_{R,i}, v_{U,i} \forall i) \\ \triangleq \sum_{i=1}^L \mathcal{L}_i(\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, \mathbf{Z}_{R,i}, z_{U,i}, \Phi_{R,i}, \varphi_{U,i}, \\ \mathbf{W}_{B,i}, \mathbf{W}_{R,i}, \mathbf{V}_{R,i}, v_{U,i}) \end{aligned} \quad (44)$$

under the decoupled constraints $\{\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, \mathbf{Z}_{R,i}, z_{U,i}\} \in \mathcal{D}_i \forall i$, where $\Phi_{R,l}$ and $\varphi_{U,l}$, $l = 1, \dots, L$, are the Lagrangian

multipliers for (38c) and (38e), respectively, and

$$\begin{aligned} \mathcal{L}_i(\mathbf{Q}_{B,i}, \mathbf{Q}_{R,i}, \mathbf{Z}_{R,i}, z_{U,i}, \Phi_{R,i}, \varphi_{U,i}, \mathbf{W}_{B,i}, \mathbf{W}_{R,i}, \mathbf{V}_{R,i}, v_{U,i}) \\ = \text{Tr}(\mathbf{Q}_{B,i}) + \text{Tr}(\mathbf{Q}_{R,i}) + \sum_{l=1}^L \text{Tr}(\Phi_{R,l}^H \tilde{\mathbf{H}}_{RR,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{H}}_{RR,l,i}^H) \\ - \text{Tr}(\Phi_{R,i}^H \mathbf{Z}_{R,i}) + \sum_{l=1, l \neq i}^L \varphi_{U,l} \tilde{\mathbf{h}}_{RU,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{h}}_{RU,l,i}^H \\ - \varphi_{U,i} z_{U,i} + \frac{C_i}{2} \left(\|\mathbf{Q}_{B,i} - \mathbf{W}_{B,i}\|_F^2 + \|\mathbf{Q}_{R,i} - \mathbf{W}_{R,i}\|_F^2 \right. \\ \left. + \|\mathbf{Z}_{R,i} - \mathbf{V}_{R,i}\|_F^2 + |z_{U,i} - v_{U,i}|^2 \right). \end{aligned} \quad (45)$$

Hence, for the SLIPD method in Step 1 of Algorithm 2, the update direction of the dual variable $\Phi_{R,l}$ is given by

$$\sum_{i=1}^L \tilde{\mathbf{H}}_{RR,l,i} \mathbf{Q}_{R,i} \tilde{\mathbf{H}}_{RR,l,i}^H - \mathbf{Z}_{R,l} \quad [35].$$

We reuse the notations in Section III to facilitate derivations of the algorithm for $N_r > 1$. Define $\mathbf{Z}_r = [\mathbf{Z}_{R,1}, \dots, \mathbf{Z}_{R,L}]$, $\mathbf{V}_r = [\mathbf{V}_{R,1}, \dots, \mathbf{V}_{R,L}]$, and $\varphi = [\text{vec}(\Phi_{R,1})^T, \dots, \text{vec}(\Phi_{R,L})^T, \varphi_{U,1}, \dots, \varphi_{U,L}]^T$. The vectors \mathbf{q} and \mathbf{w} are defined similarly as in Section III, but \mathbf{z}_r and \mathbf{v}_r need to be replaced with $\text{vec}(\mathbf{Z}_r)$ and $\text{vec}(\mathbf{V}_r)$, respectively. The $(N_r^2 + 1)L \times (\bar{N}_B^2 + N_t^2 + N_r^2 + 1)L$ matrix \mathbf{E} is in the same partition form as in (15), given by

$$\mathbf{E} = \begin{bmatrix} \mathbf{O}_{N_r^2 L \times \bar{N}_B^2 L} & \mathbf{E}_{RR} & -\mathbf{I}_{N_r^2 L \times N_r^2 L} & \mathbf{O}_{N_r^2 L \times L} \\ \mathbf{O}_{L \times \bar{N}_B^2 L} & \mathbf{E}_{RU} & \mathbf{O}_{L \times N_r^2 L} & -\mathbf{I}_{L \times L} \end{bmatrix} \quad (46)$$

where the $L \times N_r^2 L$ matrix \mathbf{E}_{RU} is the same as the one presented in Section III [cf. (17)], and the matrix $\mathbf{E}_{RR} \in \mathbb{C}^{N_r^2 L \times N_r^2 L}$ also consists of $L \times L$ equal-sized block matrices, with the (l, i) th block given by

$$\mathbf{E}_{RR,l,i} = \begin{bmatrix} \text{vec} \left(\tilde{\mathbf{H}}_{RR,l,i}^H(1, :) \tilde{\mathbf{H}}_{RR,l,i}(1, :) \right)^H \\ \vdots \\ \text{vec} \left(\tilde{\mathbf{H}}_{RR,l,i}^H(1, :) \tilde{\mathbf{H}}_{RR,l,i}(N_r, :) \right)^H \\ \vdots \\ \text{vec} \left(\tilde{\mathbf{H}}_{RR,l,i}^H(N_r, :) \tilde{\mathbf{H}}_{RR,l,i}(1, :) \right)^H \\ \vdots \\ \text{vec} \left(\tilde{\mathbf{H}}_{RR,l,i}^H(N_r, :) \tilde{\mathbf{H}}_{RR,l,i}(N_r, :) \right)^H \end{bmatrix} \quad (47)$$

where $\tilde{\mathbf{H}}_{RR,l,i}(j, :)$ denotes the j th row of $\tilde{\mathbf{H}}_{RR,l,i}$. With the notations defined earlier, the partial Lagrangian (44) can be expressed exactly as (18), and the distributed algorithm that solves (42) turns out to take the same form as Algorithm 1, except that the diagonal elements of \mathbf{A} are

$$\underbrace{\alpha_{R,1}, \dots, \alpha_{R,1}}_{N_r^2 \text{ terms}}, \dots, \underbrace{\alpha_{R,L}, \dots, \alpha_{R,L}}_{N_r^2 \text{ terms}}, \alpha_{U,1}, \dots, \alpha_{U,L}.$$

The convergence of the distributed algorithm that solves (42) is guaranteed by the following corollary.

Corollary 1: The SLIPD method used in Algorithm 1 for $N_r = 1$ also applies to the inner iterations of Algorithm 2 for $N_r > 1$, and the latter also converges to the global optimum of (42), if the step sizes satisfy

$$\max_l \alpha_{R,l} < \frac{2}{3\|\mathbf{E}\|_F^2} \min_l c_l \quad (48)$$

and

$$\max_l \alpha_{U,l} < \frac{2}{3\|\mathbf{E}\|_F^2} \min_l c_l. \quad (49)$$

Because the partial Lagrangian given by (18) remains the same regardless of the value of N_r , the convergence proof of Corollary 1 is almost the same as that for Theorem 1. Due to this convergence result, together with the convergence of SCA in Section IV-A, we can conclude that Algorithm 2, i.e., the two-layer iterative algorithm, achieves at least a local optimum of (38).

Remark 8: Similar to Algorithm 1, in the inner iterations of Algorithm 2, any initial values $\varphi(0), w(0)$ can be used. By our experience, the corresponding values for these variables obtained at the previous outer iteration are usually a good initial state for warm-starting the inner layer computations because they generally do not vary significantly after the first few outer iterations; therefore, Algorithm 2 can converge much faster.

Remark 9: In Algorithm 2, the “early termination” described in Remark 5 also applies to the inner iterations with minor modifications so that the inner loop can be ended with an acceptable solution quickly.

C. Complexity and Overhead Analysis

Similar to the analysis in Section III-C, the complexity of inner iterations can be estimated as $O(L(\bar{m}\bar{n}^2 + \bar{m}^2\bar{n}^2 + \bar{n}^3)/\varepsilon\varepsilon')$, with $\bar{m} = \max\{N_B, N_t\}$ and $\bar{n} = N_B^2 + N_t^2 + N_r + 1$, which is higher than Algorithm 1. However, the overall complexity of Algorithm 2 can be increased by one or more orders of magnitude due to the outer iterations. With respect to the communication overhead, given the similar assumption that $\mathbf{H}_{RR,i,l}$ and $\mathbf{h}_{RU,i,l}$ are known to relay l , we can also reduce the overhead in each inner iteration to $(N_r^2 + 1)L^2$ scalars for L relays, where N_r has a dominant effect. Again, the total overhead of Algorithm 2 is increased greatly due to the outer iterations. To summarize, the complexity and overhead involved in Algorithm 2 can be much higher than those of Algorithm 1, making Algorithm 2 advisable only when relays with $N_r = 1$ cannot meet the system demands.

V. SIMULATION RESULTS

Here, some simulation results are presented to evaluate the performance of the proposed distributed beamforming algorithms. The considered MIMO FDR network consists of a BS with $N_B = 4$ antennas, and $L = 2$ or $L = 3$ FDRs and users. The channels used in our simulations are composed of a large-scale path-loss component and a small-scale Rayleigh fading component. The parameters for large-scale path loss are taken from the measurements in [37], where the path loss of the feeder link is 105 dB at a BS–relay distance of 600 m, and the

path loss of access link is 100 dB at a relay–user distance of about 100 m. The typical multiuser interference link is 10 dB weaker than the access link, if not specified otherwise. For the time-domain SI cancellation modules, an attenuation of 100 dB with respect to the transmit power has been reported in [8]. This attenuation can be viewed as the equivalent path loss of SI. With proper protection between the transmit and receive antennas of a relay, an attenuation of up to 105 dB is reasonable for SI strength without performing the spatial interference suppression. The large-scale channel gain on the IRI link ranges from -120 to -95 dB, to account for varying distances and isolation conditions between adjacent relays. The noise power $\sigma_n^2 = -100$ dBm. The rate constraints are $r_1 = r_2 = 3$ b/s/Hz for $L = 2$, and $r_1 = r_2 = r_3 = 2$ b/s/Hz for $L = 3$. $c_i = 5$ [the weights in problem (10)] is used for all relays. Three different antenna settings at each relay are considered.

- AS1) $N_t = 3$, and $N_r = 1$;
- AS2) $N_t = 4$, and $N_r = 1$;
- AS3) $N_t = 4$, and $N_r = 2$.

The FDR gain of the proposed algorithms is demonstrated in Fig. 2 by comparing with the half-duplex relaying scheme in [18] under different IRI path gains. The half-duplex scheme uses the same antenna configuration as the full-duplex scheme. For Algorithm 1 under antenna setting (AS1), it can be observed in Fig. 2(a) that, when the IRI is weak (the lower two solid curves), FDR outperforms half-duplex relaying even with relatively weak time-domain SI cancellation. The curves tend to be flat as SI becomes stronger, indicating that our algorithm tends to nullify strong SI in the spatial domain when the capability of time-domain SI cancellation is insufficient. Such flexibility is in fact the benefits of combining time-domain and spatial-domain SI suppression. On the other hand, when IRI is strong (the upper solid curve), the performance of FDR can be inferior to half-duplex relaying, even with a good time-domain SI cancellation capability, as shown in Fig. 2(a) for the IRI path gain equal to -95 dB and the SI path gain higher than -95 dB. Therefore, when multiple FDRs are applied in a network, the performance gain over half-duplex relaying is possibly harder to achieve than the deployment of a single relay node; careful site and/or frequency planning of FDRs are necessary to lower the strength of IRI links. For Algorithm 2 under antenna setting (AS3), it can be observed in Fig. 2(b) that, with more antennas at each relay, the restriction on IRI strength to achieve a full-duplex gain is relaxed compared with Fig. 2(a). This leads to another way, namely adding more relay antennas, to improve overall performance in FDR networks. However, as is discussed earlier, the application of Algorithm 2 requires higher computational complexity.

Due to few advanced distributed beamforming algorithms for FDR networks reported in the open literature, a reference distributed beamforming scheme in [22], which nullifies SI and MUI via BD/ZF, is also evaluated for performance comparison with the proposed algorithms. It can be observed, in Fig. 3, that the total power performance of Algorithm 1 is superior to that of the reference scheme for (AS1) and (AS2) and their performance gap significantly increases with the gain of IRI links,

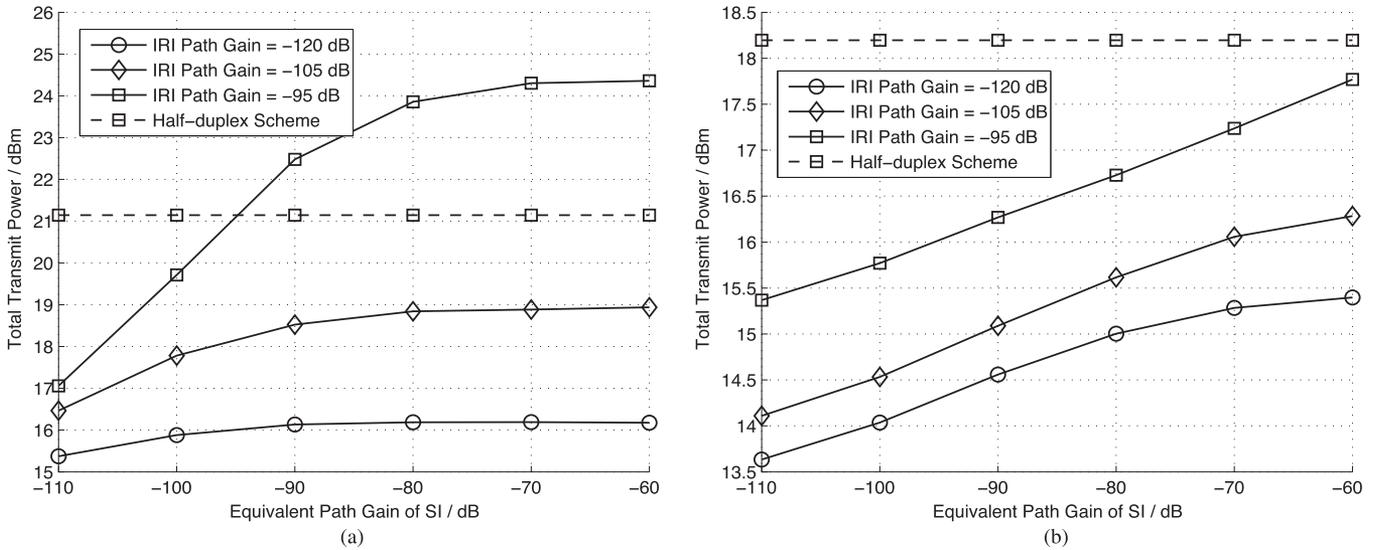


Fig. 2. Gain of FDR over half-duplex relaying for $L = 2$ with different large-scale gains on the IRI links. Part (a) for Algorithm 1 under antenna setting (AS1) and Part (b) for Algorithm 2 under antenna setting (AS3).

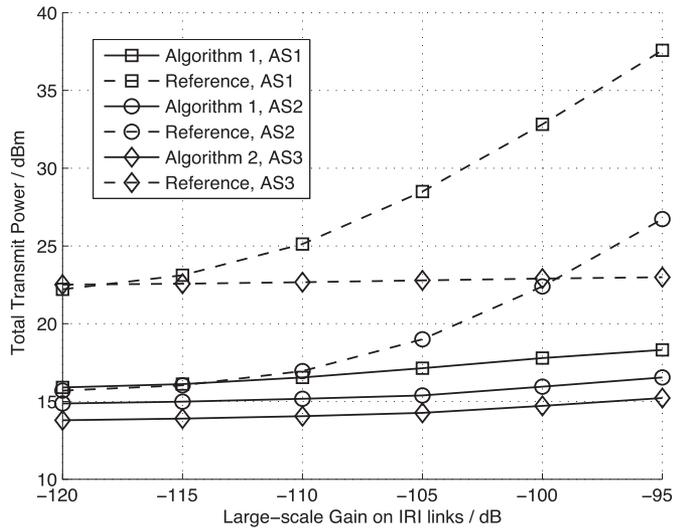


Fig. 3. Total transmit power (versus large-scale gain on IRI links) of the proposed algorithms and the reference scheme [22] for $L = 2$ under three different antenna settings.

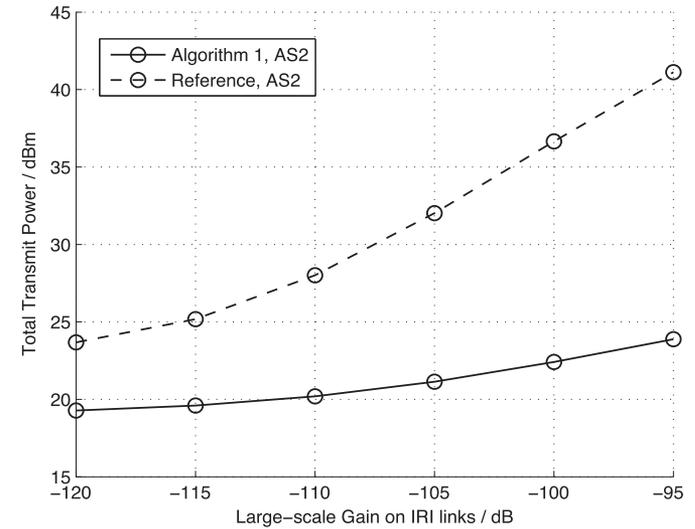


Fig. 4. Total transmit power (versus large-scale gain on IRI links) of the proposed algorithm (Algorithm 1) and the reference scheme [22] for $L = 3$ under the antenna setting (AS2). Note that the reference scheme [22] is not applicable for (AS1) and (AS3) due to an insufficient number of transmit antennas at the relays; therefore, no simulation results are provided for (AS1) and (AS3).

demonstrating that, in the considered MIMO FDR network, Algorithm 1 can suppress the IRI well, whereas the reference scheme actually fails to combat IRI even for moderate gain of IRI links due to improper use of spatial dimensions of MIMO FDR. However, under antenna setting (AS3), Algorithm 2 also performs better than the reference scheme by about 6–8 dB, and their total power performances do not vary with the IRI level too much. The reason for this is that the number of receive antennas of each relay ($N_r = 2$) is larger than the number of interfering relays ($L - 1 = 1$) for (AS3), thereby providing an extra spatial degree of freedom for the IRI suppression on the feeder links. However, when more relays coexist, it may not be very practical for each relay equipped with more antennas than the number of interfering relays due to physical size limitations of relays. The simulation results for the case of $L = 3$ are shown in Fig. 4, where only the results under (AS2)

are displayed since the ZF/BD-based reference scheme requires more transmit antennas at the relays for the other two antenna settings. The same observations in Fig. 3 for the case of $L = 2$ also apply to Fig. 4. These simulation results demonstrate that the proposed algorithms are able to effectively suppress IRI, SI, and MUI simultaneously and achieve significant transmit power savings in the considered MIMO FDR network.

An interesting but unanswered question about MIMO FDR network is: as IRI strength changes, how will SI and MUI affect the overall performance, respectively? In Fig. 5 we attempt to give a preliminary answer with Algorithm 1 under antenna setting (AS1). The performance for SI equivalent path gain = -105 dB and MUI path gain = -110 dB is used as a benchmark and compared with different values of SI and MUI path gains. Two extreme MUI path gains, i.e., the MUI-free case and the

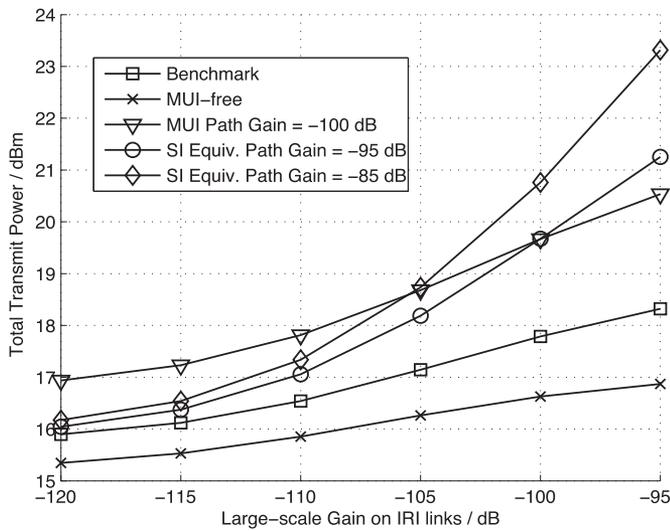


Fig. 5. Effects of SI and MUI on system performance as the large-scale gain on IRI links changes. Results for Algorithm 1 under antenna setting (AS1) are shown.

case that the MUI path gain equals the access link gain of -100 dB (which occurs when the users are located on the boundary of adjacent relays), are considered. It can be seen that the effects of MUI on the total power performance are relatively consistent as IRI path gain changes: In both cases, the performance curves show a similar slope as the benchmark. Moreover, the SI equivalent path gains of -95 and -85 dB are also considered for a performance comparison with the benchmark. The corresponding results in Fig. 5 show that the effects on the system performance by SI are different from those by MUI. When the IRI path gain is weak, the change of SI equivalent path gain has little influence on system performance. However, as the IRI path gain increases, the total power corresponding to different SI equivalent path gains increases quickly. In other words, the system performance becomes more sensitive to SI than to MUI as IRI becomes stronger. A possible explanation for this is that MUI has higher impact on the access links (i.e., a higher impact on the transmit power of relays), whereas both IRI and SI have higher impact on the feeder links (i.e., higher impact on transmit power of the BS), despite that the proposed algorithms try to balance their impacts such that the overall transmit power is minimized.

Fig. 6 shows the convergence behavior of the proposed SLIPD method (Algorithm 1 and the inner iterations of Algorithm 2) for a typical channel realization under antenna settings (AS2) and (AS3), respectively, with the large-scale channel gain on the IRI link being -105 dB, all the initial values for the unknown variables set to zero, and the step sizes chosen according to Theorem 1 and Corollary 1. Fig. 6(a) shows the dual-optimality gap, i.e., the difference between the optimal total transmit power [(10) or (42)] and the Lagrangian [(11) or (44)]. It can be seen that, in both cases, the dual gap diminishes with the iteration number, possibly with some damped oscillations in primal variables. Fig. 6(b) shows the monotonically decreasing Lyapunov function [(29) or its counterpart for $N_r > 1$] versus iteration number, demonstrating the convergence of the SLIPD method to the optimal solution (cf.

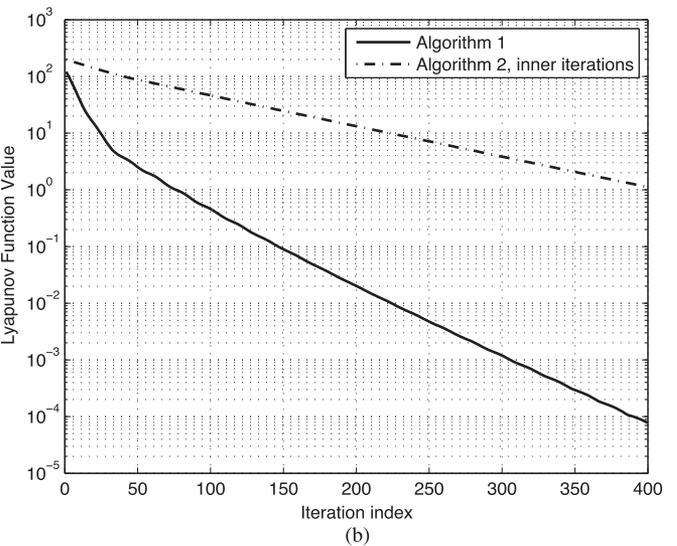
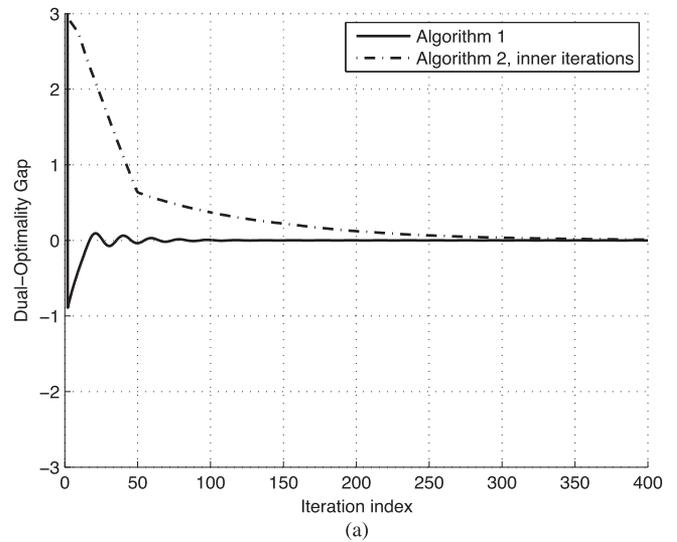


Fig. 6. Convergence behavior of Algorithm 1 with $L = 2$ and antenna setting (AS2), and the inner iterations of Algorithm 2 with $L = 2$ and antenna setting (AS3), for a typical channel realization with the large-scale channel gain on the IRI link being -105 dB. (a) Dual-optimality versus iteration number. (b) Lyapunov function versus iteration number.

Theorem 1 and Corollary 1 whose proofs are essentially due to the convergence of the Lyapunov function) and validating the effectiveness of the proposed algorithms.

The total transmit power performance of Algorithm 1 with the early termination used in the operation (cf. Remark 5) is also tested over 50 channel realizations with IRI link also set to -105 dB. For each realization, the problem (37) is solved every five iterations before the convergence of Algorithm 1, and the obtained feasible solution is recorded after 20 iterations and after 50 iterations. The simulation results are shown in Fig. 7(a) for $L = 2$ under antenna setting (AS1), and in Fig. 7(b) for $L = 3$ under antenna setting (AS2), respectively. It can be shown in Fig. 7 that the early termination yields near-optimal solutions even within 20 iterations, indicating that Algorithm 1 can well converge in several tens of iterations thus implying that it is suitable for realistic applications since faced with changing channel conditions, one may prefer to obtain a good (though not optimal) solution as early as possible.

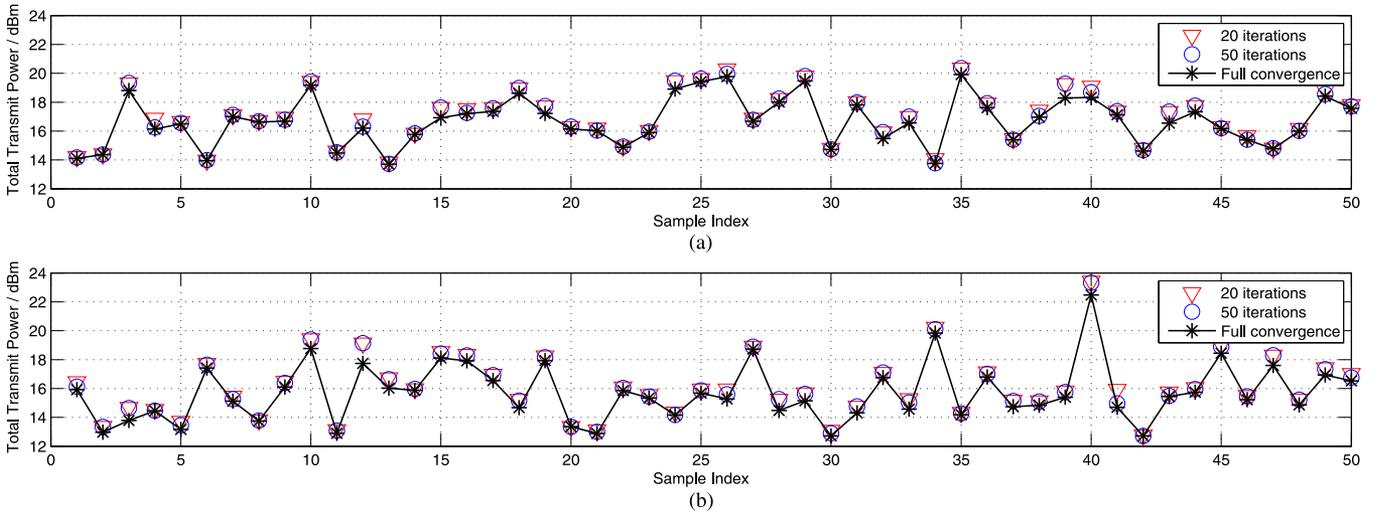


Fig. 7. Total transmit power performance of Algorithm 1 over 50 different channel realizations with the consideration of the early termination strategy as discussed in Remark 5, for the large-scale channel gain on the IRI link equal to -105 dB. (a) Results for $L = 2$ under antenna setting (AS1). (b) $L = 3$ under antenna setting (AS2).

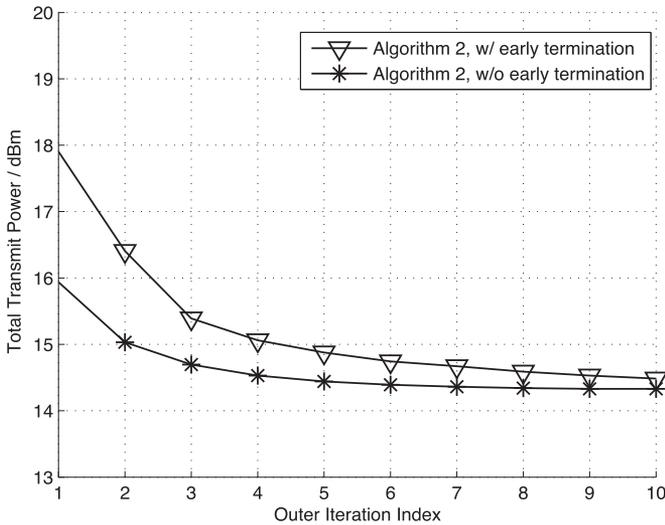


Fig. 8. With $L = 2$ and antenna setting (AS3) and the large-scale channel gain on the IRI link set to -105 dB, the convergence behavior of Algorithm 2 with/without considering the early termination in terms of average total transmit power over 50 realizations.

Finally, Algorithm 2 with/without early termination is tested with $L = 2$ and antenna setting (AS3) and the large-scale channel gain of the IRI link set to -105 dB, where early termination means that 50 inner iterations are performed in Step 1 for each outer iteration, and the obtained solution at the current outer iteration will be used for initializing the inner iterations in the next outer iteration, as is discussed in Remark 8. The average transmit power performance over 50 iterations is shown in Fig. 8. It can be seen that without early termination, Algorithm 2 converges in around seven outer iterations (after which small performance improvement is gained in the ensuing iterations), whereas with early termination, it converges in around ten outer iterations (after which small performance improvement is gained in the ensuing iterations as well). In other words, Algorithm 2 with early termination by running

only several more iterations than without early termination can also converge and achieve a near-optimal total transmit power performance. Therefore, the proposed distributed beamforming algorithms are applicable in practical systems for acceptable transmit power saving and latency.

VI. CONCLUSION

To achieve uniform coverage in cellular networks and required quality of service for multiple users in the MIMO FDR network, the transmit power at the BS and relays must be reliably controlled in facing IRI, SI, and MUI simultaneously so that the system can operate uninterruptedly and stably. We have formulated a power (sum of transmit power of the BS and relays in the network) minimization problem for the distributed beamforming design under individual user rate constraints, where the objective function turns out to be nonstrict convex. Then, we have presented a SLIPD method for efficiently solving this problem, by which the proposed two distributed beamforming algorithms (Algorithm 1 for the case of single receive antenna at relays and Algorithm 2 for the case of multiple receive antennas at relays) were developed. A globally optimal solution of the power minimization problem can be obtained by Algorithm 1, whereas Algorithm 2 can yield only a local optimal solution due to conservative SCA performed at each iteration, and their convergence can be guaranteed if the step sizes are below the upper bounds derived via a rigorous analysis. Moreover, the proposed two algorithms only require local information exchange between relays and hence are highly scalable on one hand. With the suggested early termination strategy in the operation of the two algorithms, near-optimal performance can be achieved within tens of iterations, making them suitable for realistic applications on the other hand. Finally, the efficacy of the proposed two algorithms was demonstrated by some simulation results, together with much superior total transmit power performance compared with the existing state-of-the-art algorithm reported in [22].

APPENDIX A
PROOF FOR LEMMA 1

Because both $\mathbf{q}(t)$ in (19) and $\mathbf{q}'(t)$ in (21) must satisfy (23), we have

$$\mathbf{E}^H \boldsymbol{\varphi}(t) = -\nabla f(\mathbf{q}(t)) - \mathbf{V}(\mathbf{q}(t) - \mathbf{w}(t)) \quad (50)$$

$$\mathbf{E}^H \boldsymbol{\varphi}(t+1) = -\nabla f(\mathbf{q}'(t)) - \mathbf{V}(\mathbf{q}'(t) - \mathbf{w}(t)). \quad (51)$$

Subtracting (50) from (51) followed by some mathematical derivations, we have the following inequality:

$$\begin{aligned} & (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}(t))^H \mathbf{E} \mathbf{V}^{-1} \mathbf{E}^H (\boldsymbol{\varphi}(t+1) - \boldsymbol{\varphi}(t)) \\ &= [\nabla f(\mathbf{q}'(t)) - \nabla f(\mathbf{q}(t))]^H \mathbf{V}^{-1} [\nabla f(\mathbf{q}'(t)) - \nabla f(\mathbf{q}(t))] \\ &+ 2\text{Re} \left\{ [\nabla f(\mathbf{q}'(t)) - \nabla f(\mathbf{q}(t))]^H [\mathbf{q}'(t) - \mathbf{q}(t)] \right\} \\ &+ [\mathbf{q}'(t) - \mathbf{q}(t)]^H \mathbf{V} [\mathbf{q}'(t) - \mathbf{q}(t)] \\ &\geq [\mathbf{q}'(t) - \mathbf{q}(t)]^H \mathbf{V} [\mathbf{q}'(t) - \mathbf{q}(t)] \end{aligned} \quad (52)$$

where we have used the following inequality (which can be easily proven by the first-order condition of convex functions):

$$\text{Re} \left\{ [\nabla f(\mathbf{z}_2) - \nabla f(\mathbf{z}_1)]^H \cdot (\mathbf{z}_2 - \mathbf{z}_1) \right\} \geq 0 \quad (53)$$

where \mathbf{z}_1 and \mathbf{z}_2 are any complex vectors. Hence, the first inequality (27) of Lemma 1 has been proven.

For proving (28), we denote the j th element in \mathbf{q} and \mathbf{w} by q_j and w_j , respectively. The j th row of (23) is

$$f'(q_j) + \mathbf{e}_j^H \boldsymbol{\varphi} + V_j(q_j - w_j) = 0 \quad (54)$$

where $f'(q_j)$ denotes the partial derivative of $f(\mathbf{q})$ with respect to q_j , \mathbf{e}_j is the j th column of \mathbf{E} and V_j is the j th diagonal element of \mathbf{V} , respectively. Considering the two vectors $\mathbf{q}_1 = \arg \min_{\mathbf{q}} L(\mathbf{q}, \boldsymbol{\varphi}_1, \mathbf{w})$, $\mathbf{q}_2 = \arg \min_{\mathbf{q}} L(\mathbf{q}, \boldsymbol{\varphi}_2, \mathbf{w})$, with $q_{1,j}$ and $q_{2,j}$ denoting the j th component of the former and the latter, respectively, we have

$$\begin{aligned} & \text{Re} \left\{ (\nabla f(\mathbf{q}_1) - \nabla f(\mathbf{w}^*))^H (\mathbf{q}_2 - \mathbf{w}^*) \right\} \\ &= \text{Re} \left\{ \sum_j (f'(q_{1,j}) - f'(w_j^*))^C (q_{2,j} - w_j^*) \right\} \\ &\triangleq \sum_j \text{Re} \{ a_{1,j}^C b_{2,j} \} \end{aligned} \quad (55)$$

where $a_{1,j} = f'(q_{1,j}) - f'(w_j^*)$ and $b_{2,j} = q_{2,j} - w_j^*$. Similarly, $a_{2,j}$ and $b_{1,j}$ can be defined as $a_{2,j} = f'(q_{2,j}) - f'(w_j^*)$ and $b_{1,j} = q_{1,j} - w_j^*$. From (53), we have $\text{Re}\{a_{1,j}^C b_{1,j}\} \geq 0$ and $\text{Re}\{a_{2,j}^C b_{2,j}\} \geq 0$. Hence, $\text{Re}\{b_{1,j}^C/a_{1,j}\} \geq 0$ (where $a_{1,j} \neq 0$ is assumed). From (54), we have

$$\begin{aligned} & -\mathbf{e}_j^H (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2) \\ &= (\nabla_j f(q_{1,j}) - \nabla_j f(q_{2,j})) + V_j(q_{1,j} - q_{2,j}) \\ &= (a_{1,j} - a_{2,j}) + V_j(b_{1,j} - b_{2,j}). \end{aligned} \quad (56)$$

Hence

$$\begin{aligned} & \left(1 + \text{Re} \left\{ \frac{V_j b_{1,j}^C}{a_{1,j}^C} \right\} \right) \text{Re} \{ a_{1,j}^C b_{2,j} \} \\ &= \text{Re} \{ (a_{1,j}^C + V_j b_{1,j}^C) b_{2,j} \} \\ &= \text{Re} \{ [a_{2,j}^C + V_j b_{2,j}^C - \mathbf{e}_j^T (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2)^C] b_{2,j} \} \\ &\geq \text{Re} \{ V_j b_{2,j}^C b_{2,j} - \mathbf{e}_j^T (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2)^C b_{2,j} \} \quad (\text{by (54)}) \\ &= V_j \left\{ |b_{2,j}|^2 - 2\text{Re} \left\{ \frac{1}{2V_j} \mathbf{e}_j^T (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2)^C b_{2,j} \right\} \right\} \\ &= V_j \left\{ \left| b_{2,j} - \frac{1}{2V_j} \mathbf{e}_j^H (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2) \right|^2 \right. \\ &\quad \left. - \frac{1}{4V_j^2} (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2)^H \mathbf{e}_j \mathbf{e}_j^H (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2) \right\} \\ &\geq -\frac{1}{4V_j} (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2)^H \mathbf{e}_j \mathbf{e}_j^H (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2) \end{aligned} \quad (57)$$

where the first inequality is due to $\text{Re}\{a_{2,j}^C b_{2,j}\} \geq 0$. Since $1 + \text{Re}\{V_j b_{1,j}^C/a_{1,j}^C\} \geq 1$, (57) can be further simplified as

$$\begin{aligned} & -\text{Re} \{ a_{1,j}^C b_{2,j} \} \\ &\leq \frac{\frac{1}{4V_j} (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2)^H \mathbf{e}_j \mathbf{e}_j^H (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2)}{\left(1 + \text{Re} \left\{ \frac{V_j b_{1,j}^C}{a_{1,j}^C} \right\} \right)} \\ &\leq \frac{1}{4V_j} (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2)^H \mathbf{e}_j \mathbf{e}_j^H (\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2). \end{aligned} \quad (58)$$

By (58) and (55), we have

$$\begin{aligned} & -\text{Re} \left\{ (\nabla f(\mathbf{q}_1) - \nabla f(\mathbf{w}^*))^H (\mathbf{q}_2 - \mathbf{w}^*) \right\} \\ &\leq \frac{1}{4} (\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_1)^H \mathbf{E} \mathbf{V}^{-1} \mathbf{E}^H (\boldsymbol{\varphi}_2 - \boldsymbol{\varphi}_1). \end{aligned} \quad (59)$$

Hence, the second inequality (28) of Lemma 1 is proved.

APPENDIX B
PROOF FOR LEMMA 2

Let \mathbf{x} be any complex vector of proper dimension, \mathbf{e}_j denote the j th column of \mathbf{E} , and V_j denote the j th diagonal element of \mathbf{V} . It can be easily seen that

$$\begin{aligned} & \mathbf{x}^H \mathbf{E} \mathbf{V}^{-1} \mathbf{E}^H \mathbf{x} \\ &= \sum_j V_j^{-1} |\mathbf{e}_j^H \mathbf{x}|^2 \\ &\leq (\min_j V_j)^{-1} \sum_j |\mathbf{e}_j|^2 |\mathbf{x}|^2 \quad (\text{by Schwarz inequality}) \\ &= (\min_j V_j)^{-1} \|\mathbf{E}\|_F^2 |\mathbf{x}|^2. \end{aligned} \quad (60)$$

Similarly, we have

$$(\max_j \alpha_j)^{-1} |\mathbf{x}|^2 \leq \mathbf{x}^H \mathbf{A}^{-1} \mathbf{x}. \quad (61)$$

Obviously $\mathbf{E} \mathbf{V}^{-1} \mathbf{E}^H < a \mathbf{A}^{-1}$ holds when

$$\max_j \alpha_j < \frac{a(\min_j c_j)}{\|\mathbf{E}\|_F^2} \quad (\text{since } \{V_j \forall j\} = \{c_j \forall j\}). \quad (62)$$

Hence, Lemma 2 is proved.

REFERENCES

- [1] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
- [2] J. Sydir and R. Taori, "An evolved cellular system architecture incorporating relay stations," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 115–121, Jun. 2009.
- [3] Z. Zhang, X. Chai, K. Long, A. Vasilakos, and L. Hanzo, "Full duplex techniques for 5G networks: Self-interference cancellation, protocol design, and relay selection," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 128–137, May 2015.
- [4] R.-A. Pitaval *et al.*, "Full-duplex self-backhauling for small-cell 5G networks," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 83–89, Oct. 2015.
- [5] J. Jeong, S. Guo, Y. Gu, T. He, and D. Du, "Trajectory-based statistical forwarding for multihop infrastructure-to-vehicle data delivery," *IEEE Trans. Mobile Comput.*, vol. 11, no. 10, pp. 1523–1537, Oct. 2012.
- [6] A. Sabharwal *et al.*, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.
- [7] M. Duarte *et al.*, "Design and characterization of a full-duplex multi-antenna system for WiFi networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1160–1177, Mar. 2014.
- [8] D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proc. NSDI USENIX Assoc.*, 2014, pp. 359–372.
- [9] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. ACM Mobicom*, 2010, pp. 1–12.
- [10] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, Feb. 2014.
- [11] H. Ju, E. Oh, and D. Hong, "Improving efficiency of resource usage in two-hop full duplex relay systems based on resource sharing and interference cancellation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 3933–3938, Aug. 2009.
- [12] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983–5993, Dec. 2011.
- [13] H. Suraweera, I. Krikidis, G. Zheng, C. Yuen, and P. Smith, "Low-complexity end-to-end performance optimization in MIMO full-duplex relay systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 913–927, Feb. 2014.
- [14] U. Ugurlu, T. Riihonen, and R. Wichman, "Optimized in-band full-duplex MIMO relay under single-stream transmission," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 155–168, Jan. 2016.
- [15] H. Q. Ngo, H. Suraweera, M. Matthaiou, and E. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sep. 2014.
- [16] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex MIMO relaying: Achievable rates under limited dynamic range," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1541–1553, Sep. 2012.
- [17] A. C. Cirik, Y. Rong, and Y. Hua, "Achievable rates of full-duplex MIMO radios in fast fading channels with imperfect channel estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3874–3886, Aug. 2014.
- [18] A. Akhlaq, A. Sulyman, H. Hassanein, A. Alsanie, and S. Alshebeili, "Performance analysis of relay-multiplexing scheme in cellular systems employing massive multiple-input multiple-output antennas," *IET Commun.*, vol. 8, no. 10, pp. 1788–1799, Jul. 2014.
- [19] D. W. K. Ng, E. S. Lo, and R. Schober, "Dynamic resource allocation in MIMO-OFDMA systems with full-duplex and hybrid relaying," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1291–1304, May 2012.
- [20] G. Liu, F. Yu, H. Ji, V. Leung, and X. Li, "In-band full-duplex relaying: A survey, research issues and challenges," *IEEE Commun. Surveys Tut.*, vol. 17, no. 2, pp. 500–524, 2nd Quart. 2015.
- [21] O. Somekh, O. Simeone, H. Poor, and S. Shamai, "Cellular systems with non-regenerative relaying and cooperative base stations," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2654–2663, Aug. 2010.
- [22] J.-H. Lee and O.-S. Shin, "Full-duplex relay based on block diagonalisation in multiple-input multiple-output relay systems," *IET Commun.*, vol. 4, no. 15, pp. 1817–1826, Oct. 2010.
- [23] J.-H. Lee and O.-S. Shin, "Full-duplex relay based on distributed beamforming in multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1855–1860, May 2013.
- [24] G. Liu, F. Yu, H. Ji, and V. Leung, "Energy-efficient resource allocation in cellular networks with shared full-duplex relaying," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3711–3724, Aug. 2015.
- [25] R. Zhang, C. C. Chai, and Y.-C. Liang, "Joint beamforming and power control for multi-antenna relay broadcast channel with QoS constraints," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 726–737, Feb. 2009.
- [26] L. Sanguinetti and A. A. D'Amico, "Power allocation in two-hop amplify-and-forward MIMO relay systems with QoS requirements," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2494–2507, May 2012.
- [27] X. Lin and N. B. Shroff, "Utility maximization for communication networks with multipath routing," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 766–781, May 2006.
- [28] X. Zhang *et al.*, "Distributed power allocation for coordinated multi-point transmissions in distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2281–2291, May 2013.
- [29] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
- [30] Z. Shen, R. Chen, J. Andrews, R. Heath, and B. Evans, "Sum capacity of multiuser MIMO broadcast channels with block diagonalization," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2040–2045, Jun. 2007.
- [31] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
- [32] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.
- [33] Y. Wang and D.-W. Yue, "Capacity of MIMO Rayleigh fading channels in the presence of interference and receive correlation," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4398–4405, Oct. 2009.
- [34] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2013.
- [35] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [36] M. Stingl, *On the Solution of Nonlinear Semidefinite Programs by Augmented Lagrangian Methods*. Herzogenrath, Germany: Shaker Verlag GmbH, 2006.
- [37] Y. Yuan, *LTE-Advanced Relay Technology and Standardization*. Berlin, Germany: Springer-Verlag, 2012.



Xiaofei Xu received the B.E. degree from Shandong University, Jinan, China, in 2010. He is currently working toward the Ph.D. degree with the Wireless and Mobile Communication Technology R&D Center, Research Institute of Information Technology, Tsinghua University, Beijing, China.

His research interests include wireless communications and software radio.



Xiang Chen (M'08) received the B.E. and Ph.D. degrees, both from the Department of Electronics Engineering, Tsinghua University, Beijing, China, in 2002 and 2008, respectively.

From July 2008 to July 2012, he was with the Wireless and Mobile Communication Technology R&D Center, Research Institute of Information Technology, Tsinghua University. From August 2012 to December 2014, he was with Tsinghua Space Center, School of Aerospace, Tsinghua University. Since January 2015, he has been an Associate Professor

with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China. His research interests include fifth-generation wireless communications, Internet of Things, and software radio.



Ming Zhao (M'98) received the B.S. and Ph.D. degrees, both from the Department of Electronics Engineering, Tsinghua University, Beijing, China, in 1993 and 1998, respectively.

Since 1998, he has been with Tsinghua University, where he is currently a Professor. His research interests include topics on wireless digital communications, such as information theory, channel coding, multiuser detection, statistical signal processing, estimation theory, spread-spectrum communications, and multiple-antenna systems.



Shidong Zhou (M'98) received the B.S. and M.S. degrees in wireless communications from Southeast University, Nanjing, China, in 1991 and 1994, respectively, and the Ph.D. degree in communication and information systems from Tsinghua University, Beijing, China, in 1998.

From 1999 to 2001, he was in charge of several projects for the China 3G Mobile Communication R&D Project. He is currently a Professor with Tsinghua University. His research interests include wireless and mobile communications.



Chong-Yung Chi (S'83–M'83–SM'89) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1983.

From 1983 to 1988, he was with the Jet Propulsion Laboratory, Pasadena, CA. Since 1989, he has been a Professor with the Department of Electrical Engineering and, since 1999, with the Institute of Communications Engineering (also its Chair from 2002 to 2005), National Tsinghua University, Hsinchu, Taiwan. He is the author of more than 210

technical papers, including more than 80 journal papers (mostly in the IEEE TRANSACTIONS ON SIGNAL PROCESSING), four book chapters, more than 130 peer-reviewed conference papers, and a graduate-level textbook *Blind Equalization and System Identification* (Springer-Verlag, 2006). His current research interests include signal processing for wireless communications, convex analysis and optimization for blind source separation, and biomedical and hyperspectral image analysis.

Dr. Chi is a member of IEEE Signal Processing Committee on Signal Processing for Communications and Networking and the IEEE Signal Processing Committee on Sensor Array and Multichannel Signal Processing. He has been a Technical Program Committee member for many IEEE-sponsored and cosponsored workshops, symposia, and conferences on signal processing and wireless communications, including Co-organizer and General Cochair of the 2001 IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), the Cochair of the Signal Processing for Communications (SPC) Symposium of the 2008 EAI International Conference on Communications and Networking in China (ChinaCOM), and the Lead Cochair of SPC Symposium, 2009 ChinaCOM. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (May 2001–April 2006, January 2012–December 2015), the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II (January 2006–December 2007), the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I (January 2008–December 2009), and the IEEE SIGNAL PROCESSING LETTERS (June 2006–May 2010); a member of the Editorial Board of *EURASIP Signal Processing Journal* (May 2005–May 2008), and an Editor (July 2003–December 2005) and a Guest Editor (2006) of *EURASIP Journal on Applied Signal Processing*. From 2005 to 2010, he was a member of IEEE Signal Processing Committee on Signal Processing Theory and Methods.



Jing Wang (M'99) received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 1983 and 1986, respectively.

Since 1986, he has been with Tsinghua University, where he is currently a Professor and the Vice Dean of the Tsinghua National Laboratory for Information Science and Technology. His research interests include wireless digital communications, including modulation, channel coding, multiuser detection, and 2-D RAKE receivers. He is the author of more than 100 conference and journal papers. He

is a member of the Technical Group of the China Third-Generation Mobile Communication R&D Project, and he serves as an expert of communication technology in the National 863 Program.

Mr. Wang is a member of the Radio Communication Committee of the Chinese Institute of Communications and a Senior Member of the Chinese Institute of Electronics.