Full-Duplex Massive MIMO Multi-Pair Two-Way AF Relaying: Energy Efficiency Optimization

Ekant Sharma, Rohit Budhiraja, K Vasudevan and Lajos Hanzo, *Fellow IEEE*

Abstract

We consider two-way amplify and forward relaying, where multiple full-duplex user pairs exchange information via a shared full-duplex massive multiple-input multiple-output (MIMO) relay. Most of the previous massive MIMO relaying works maximize the spectral efficiency (SE). By contrast, we maximize the non-convex energy efficiency (EE) metric by approximating it as a pseudo-concave problem, which is then solved using the classic Dinkelbach approach. We also maximize the EE of the least energy-efficient user relying on the max-min approach. For solving these optimization problems, we derive closed-form lower bounds for the ergodic achievable rate both for maximal-ratio combining and zero-forcing processing at the relay, by using minimum mean squared error channel estimation. We numerically characterize the accuracy of the lower bounds derived. We also compare the SE and EE of the proposed design to those of the existing full-duplex systems and quantify the significant improvement achieved by the proposed algorithm. We also compare the EE of the proposed full-duplex system to that of its half-duplex counterparts, and characterize the self-loop and inter-user interference regimes, for which the proposed full-duplex system succeeds in outperforming the half-duplex ones.

Index Terms

Energy efficiency, full-duplex, relay.

I. Introduction

Relay-based communication has been extensively investigated to expand the coverage, improve the diversity, increase the data rate, and reduce the power consumption of wireless communication systems [1], [2]. The current generation of relays is predominantly half-duplex, as a benefit of their implementational simplicity. A half-duplex relay requires two channel uses to send a data packet from the transmitter to the receiver, since the relay cannot transmit and receive with in

Ekant Sharma, Rohit Budhiraja and K Vasudevan are with the Department of Electrical Engineering, Indian Institute of Technology, Kanpur, 208016, India. email: {ekant, rohitbr, vasu@}iitk.ac.in.

Lajos Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

the same time slot. Full-duplex technology is becoming popular as a benefit of its increased throughput [2]–[6]. A full-duplex one-way relay [4], [7] transmits and receives at the same time, hence theoretically it doubles the spectral efficiency (SE) of a half-duplex one-way relay [8], [9].

Full-duplex two-way relaying [10]–[12], wherein two users exchange their data over a single channel use via a relay, further improves the SE. Two-way full-duplex relaying has recently been extended to multi-pair two-way full-duplex relaying [13]–[15] wherein multiple user pairs exchange their data via a shared relay in a single channel use. A multi-pair two-way full-duplex relay system has the following interference sources: i) co-channel (inter-pair) interference due to multiple users simultaneously accessing the channel; ii) self-loop interference at the relay and at the users; and iii) inter-user interference caused due to simultaneous transmission and reception by full-duplex nodes.

Massive multiple-input multiple-output (MIMO) systems have become popular, since they cancel co-channel interference by using simple linear transmit processing schemes, such as, zero-forcing transmission (ZFT) and maximal-ratio transmission (MRT) [16]–[19], which significantly improve their SE. Massive MIMO technology is also being incorporated into multi-pair full-duplex relays for mitigating the self-loop interference at the relay, and the inter-pair co-channel interference [13]–[15], [20], [21]. Ngo *et al.* [13] derived the achievable rate and a power allocation scheme for maximizing the ergodic sum-rate for one-way decode and forward full-duplex massive MIMO-aided relaying. Zhang *et al.* [14] proposed four power scaling schemes for two-way full-duplex massive MIMO relaying to improve both its SE and its energy efficiency (EE). Zhang *et al.* [15] developed a power allocation scheme for maximizing the sum-rate of multi-pair two-way full-duplex massive MIMO amplify-and-forward (AF) relaying by using maximal-ratio combining (MRC)/MRT processing at the relay, and by using least squares (LS) channel estimation. Dai *et al.* [20] considered a half-duplex multi-pair two-way massive MIMO AF relay and derived closed-form achievable rate expressions and a power allocation scheme for maximizing the sum-rate under realistic imperfect channel state information (CSI). Cui *et al.* [21] developed power scaling schemes for half-duplex massive MIMO-aided one-way relay systems.

The energy cost of wireless network operation has increased to almost 50% of the total operational cost [22]. The EE metric, which relies on the Pareto-optimality between throughput and energy consumption, has recently drawn attention as a potent performance measure. The EE has recently been optimized for both conventional MIMO [23], [24] and for single-hop massive

MIMO systems [25], [26]. Nguyen *et al.* [25] optimized the downlink EE of a cell-free single-hop massive MIMO system using ZF precoding. Li *et al.* [26] optimized both the achievable rate and the EE of a small-cell based massive MIMO single-hop full-duplex system. It is anticipated that a paradigm shift towards multi-component Pareto-optimization is about to take place, leading to an entire optimal Pareto-front of solutions [27], [28].

The existing literature of massive MIMO relays, on the other hand, has either optimized the achievable rate [13], [15], [20] or analyzed the EE [14], [21]. The EE optimization for massive MIMO relaying has not been investigated at the time of writing, except for a recent study in [29] which optimized the *asymptotic* EE for a multi-pair *one-way decode and forward* massive MIMO *half-duplex* relay. To the best of our knowledge, the EE of multi-pair *two-way AF* massive MIMO *full-duplex* relay system has not been considered in the literature for a realistic finite-cardinality antenna-set. Hence we fill this gap. Due to the self-loop interference and the coupled channels encountered in *two-way full-duplex AF relaying*, the power allocation scheme of [29] cannot be applied to our system. Against this backdrop, we list the **main contributions** of this paper.

1) We derive closed-form lower bounds for the achievable rate of the multi-pair two-way AF full-duplex massive MIMO relay for an arbitrary number of relay antennas. We consider both MRC/MRT and zero-forcing reception (ZFR)/ZFT processing at the relay, whilst relying on the minimum mean square error (MMSE) relay channel estimation. In contrast to [15], which derives a closed-form lower bound for MRC/MRT processing alone, which is based on the LS channel estimation, we derive new bounds both for MRC/MRT and for ZFR/ZFT processing based on the MMSE channel estimation. These closed-form achievable-rate expressions have not been derived for an arbitrary number of relay antennas in the massive MIMO relaying literature.

2) We optimally allocate power to maximize the EE by using the closed-form achievable rate expressions derived. The EE maximization, which has a non-convex objective, is solved by proposing an algorithm wherein we first approximate the objective as a pseudo-convex function, and later choose Dinkelbach's approach. This contribution is significantly different from [29], which considers asymptotic EE optimization for decode and forward one-way half-duplex relay. The achievable rate expressions, and consequently the EE optimization, developed herein are applicable to any antenna configuration. Furthermore, the expressions and the analysis developed herein are significantly more complex due to the coupling of channels in AF relaying, and both the self-loop as well as the inter-user interference imposed by the full-duplex nodes. We numerically

compare the performance of the proposed EE algorithm to the equal-power approach of [14]. Furthermore, we show that the EE optimization algorithm can also be used for optimizing the EE under specific Quality-of-Service (QoS) constraints. We investigate the effect of QoS constraints on the EE performance.

3) We also maximize the EE by using the max-min fairness criterion; the problem has a non-differentiable objective. We solve this problem by first using the sequential convex programming approach to approximate the objective function by a quasi-concave function, and later by using the generalized Dinkelbach's method of [30].

4) The proposed EE maximization framework can also maximize the SE for both MRC/MRT and ZFR/ZFT processing. In contrast to [15], which derives the lower bound and maximizes the SE (not EE) for the MRC/MRT processing alone relying on LS channel estimation, we conceive a more general approach than that of [15]. We quantify the considerably improved SE of ZFR/ZFT processing and MMSE channel estimation over the LS-based MRC/MRT processing of [15].

5) The proposed full-duplex EE optimization framework can also be used for evaluating the EE of massive MIMO half-duplex AF systems which has not been investigated in the open literature. We compare the EE of both the full-duplex and of the half-duplex systems and numerically quantify the self-loop and inter-user interference values for which a full-duplex system has a better EE than a half-duplex system. We also show the significantly improved EE of optimal power allocation over the equal-power EE analyzed in [21].

The rest of the paper is organized as follows. We present our system model in Section II, and discuss the MMSE channel estimation in Section III. The relay processing is discussed in Section IV, while the achievable rates are analyzed in Section V. Our energy-efficient optimization problems are formulated in Section VI, while our performance improvements over [14], [15], [21] are presented in Section VII. The paper is concluded in Section VIII.

Notations: The boldface capital and small letter represents matrix and vector, respectively, $\mathbb{C}^{r \times s}$ denotes a complex matrix of dimension $r \times s$. The superscript $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$ denotes the transpose, Hermitian and conjugate operations, respectively. The diag($\mathbf{x}$) denotes a (square) diagonal matrix with elements $\mathbf{x}$ on its main diagonal, $\mathbf{I}_Q$ denotes an $Q \times Q$ identity matrix, and $\mathbf{1}_k$ denotes a $2K \times 1$ vector consisting of value one at kth row and zero otherwise. The expectation and trace operations are denoted by $\mathbb{E}[\cdot]$ and tr$\{\cdot\}$, respectively. The notation $\mathcal{CN}(\mathbf{0}, \mathbf{K})$ represents a circularly-symmetric complex Gaussian random vector with covariance matrix $\mathbf{K}$.

II. SYSTEM MODEL

We consider the multi-pair two-way AF full-duplex relaying shown in Fig. 1, where K full-duplex user pairs communicate via a single full-duplex relay within the same time-frequency resource. We assume that the user S_{2m-1} for $m = 1, \cdots, K$ on one side of the relay, wants to send as well as receive from the user S_{2m} on the other side of the relay. We also assume that there are no direct links between the user-pairs (S_{2m-1}, S_{2m}) due to a high path loss and heavy shadowing, which is commonly assumed in the multi-user two-way massive MIMO relaying literature [13], [14]. Furthermore, the relay has N transmit and N receive antennas, while each user has one transmit and one receive antenna. The users on either side of the relay interfere with each other, due to full-duplex architecture; the interference caused is termed as inter-user interference.
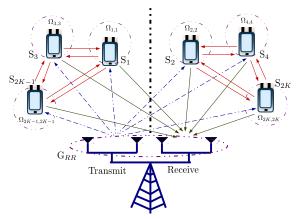

Fig. 1: Multi-pair two-way full-duplex AF massive MIMO relay system: All users and the relay simultaneously transmit and receive which leads to self-loop interference at the relay $\mathbf{G}_{RR}$ (shown by dotted ellipse), at the user $\Omega_{k,k}$ (shown by dotted circle) and inter-user interference (shown by solid red line). The users on either side of the relay (marked with bold dotted line at the center) are isolated.

At time instant n, each user S_k, $k = 1$ to $2K$,[1] transmits the signal $\sqrt{p_k}x_k(n)$ to the relay, and simultaneously the relay broadcasts a vector $\mathbf{x}_R(n) \in \mathbb{C}^{N \times 1}$ to all the users. Here the term p_k denotes the transmit power of the kth user. The signal received at the relay and at the user S_k are given by

$$\mathbf{y}_R(n) = \sum_{k=1}^{2K} \sqrt{p_k}\mathbf{g}_k x_k(n) + \mathbf{G_{RR}}\mathbf{x}_R(n) + \mathbf{z}_R(n) = \tilde{\mathbf{G}}\mathbf{x}(n) + \mathbf{G_{RR}}\mathbf{x}_R(n) + \mathbf{z}_R(n), \quad (1)$$

$$y_k(n) = \mathbf{f}_k^T\mathbf{x}_R(n) + \sum_{i,k \in U_k} \Omega_{k,i}\sqrt{p_i}x_i(n) + z_k(n). \quad (2)$$

Here $\mathbf{g}_k \in \mathbb{C}^{N \times 1}$ and $\mathbf{f}_k \in \mathbb{C}^{N \times 1}$ denote the channels spanning from the transmit antenna of the kth user to the relay's receive antenna array, and from the relay's transmit antenna

---

[1]To avoid repetition, we assume that $k = 1$ to $2K$ throughout this paper.

array to the receive antenna of the kth user, respectively. We now introduce the matrix $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3, \cdots, \mathbf{g}_{2K}] \in \mathbb{C}^{N \times 2K}$ and the matrix $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \cdots, \mathbf{f}_{2K}] \in \mathbb{C}^{N \times 2K}$ (to be used later in the sequel). Furthermore we have, $\tilde{\mathbf{G}} = \mathbf{GP}$ where $\mathbf{P} = \text{diag}\left\{\sqrt{p_1}, \sqrt{p_2}, \cdots \sqrt{p_{2K}}\right\}$ with $0 \leq p_k \leq P^{\max}$. The signal received at the relay and at the users are interfered by their own transmit signal, which is termed as the self-loop interference. In (1) and (2), $\mathbf{G}_{RR}$ and $\Omega_{k,k}$ denote the self-loop interference at the relay and at the user S_k, respectively. The entries of the matrix $\mathbf{G}_{RR}$ are independent and identically distributed (i.i.d.) with distribution $\mathcal{CN}(0, \sigma_{LIR}^2)$ [14]. Furthermore $\Omega_{k,k}$ is distributed as $\mathcal{CN}(0, \sigma_{k,k}^2)$. The terms $\Omega_{k,i}\,(k, i \in U_k, i \neq k)$, which denotes the inter-user interference channel, are independent and are distributed as $\mathcal{CN}(0, \sigma_{k,i}^2)$ [14], where the set obeys $U_k = [1, 3, 5, \cdots, 2K-1]$ for odd k and $U_k = [2, 4, 6, \cdots, 2K]$ for even k. We also define the vector $\mathbf{x}(n) = [x_1(n), x_2(n), x_3(n), \cdots, x_{2K}(n)]^T \in \mathbb{C}^{N \times 1}$ with $\mathbb{E}\left[\mathbf{x}(n)\mathbf{x}^H(n)\right] = \mathbf{I}_{2K}$. The vector $\mathbf{z}_R(n) \in \mathbb{C}^{N \times 1}$ and the scalar $z_k(n)$ are additive white Gaussian noise (AWGN) processes at the relay and at the user S_k. The elements of $\mathbf{z}_R$ and the scalar $z_k(n)$ are modeled as i.i.d. $\mathcal{CN}(0, \sigma_{nr}^2)$ and $\mathcal{CN}(0, \sigma_n^2)$, respectively.

Remark 1. The channel matrices account for both small-scale and large-scale fading; we therefore have $\mathbf{G} = \mathbf{H}_u \mathbf{D}_u^{1/2}$ and $\mathbf{F} = \mathbf{H}_d \mathbf{D}_d^{1/2}$. Here the small-scale fading matrices $\mathbf{H}_u$ and $\mathbf{H}_d$ have i.i.d. $\mathcal{CN}(0, 1)$ elements, while the kth element of the large-scale diagonal fading matrices $\mathbf{D}_u$ and $\mathbf{D}_d$ are denoted as $\sigma_{g,k}^2$ and $\sigma_{f,k}^2$, respectively.

In the first time slot $(n = 1)$, the relay only receives the signal from the users, and does not transmit. The signals received at the relay and at the user S_k are given respectively as

$$\mathbf{y}_R(1) = \tilde{\mathbf{G}}\mathbf{x}(1) + \mathbf{z}_R(1), \quad y_k(1) = \sum_{i, k \in U_k} \Omega_{k,i}\sqrt{p_i}x_i(1) + z_k(1). \quad (3)$$

At the nth time slot, the relay linearly precodes its received signal $\mathbf{y}_R(n-1)$ using a matrix $\mathbf{W}$, such that

$$\mathbf{x}_R(n) = \alpha\mathbf{W}\mathbf{y}_R(n-1), \quad (4)$$

where α is the scaling factor chosen to satisfy the relay's power constraint. The relay's transmit signal $\mathbf{x}_R(n)$, before applying any self-loop interference cancellation technique, can be re-formulated by iteratively substituting (1) into (4) as

$$\mathbf{x}_R(n) = s\left[\mathbf{x}(n-\nu) + \mathbf{x}(n-2\nu) + \cdots + \mathbf{z}_R(n-\nu) + \mathbf{z}_R(n-2\nu) + \cdots\right]. \quad (5)$$

Here $s[\cdot]$, as discussed in [14], is a function involving both vector and matrix operations, while ν is the relay's processing delay ($\nu = 1$ in this paper). The relay transmit signal, in the above form which assumes no interference cancellation, is difficult to analyze [14]. The recent full-

duplex studies e.g. in [4], [6], [14] have shown that the self-loop interference can be significantly suppressed, and the residual self-loop interference at the relay can be modeled as an additional Gaussian noise source. We also apply these cancellation techniques, and replace $\mathbf{x}_R(n)$ in the self-loop interference term $\mathbf{G}_{RR}\mathbf{x}_R(n)$ in (1) with a Gaussian noise $\tilde{\mathbf{x}}_R(n)$, which represents the residual self-loop interference with the power constraint $\left\{\mathbb{E}\left[\tilde{\mathbf{x}}_R(n)\tilde{\mathbf{x}}_R^H(n)\right]\right\} = \frac{P_R}{N}\mathbf{I}_N$ [14]. We therefore have

$$\tilde{\mathbf{y}}_R(n) = \tilde{\mathbf{G}}\mathbf{x}(n) + \mathbf{G}_{\mathbf{RR}}\tilde{\mathbf{x}}_R(n) + \mathbf{z}_R(n). \tag{6}$$

After loop interference suppression, the relay's transmit signal in (4) is re-expressed using (6) as

$$\mathbf{x}_R(n) = \alpha\mathbf{W}\tilde{\mathbf{y}}_R(n-1). \tag{7}$$

We observe from (7) that $\mathbf{x}_R(n)$ now does not have infinite memory of $\mathbf{x}(n)$, which was the case earlier in (5). It is a function of $\mathbf{x}(n-1)$. The time labels from the model in (7) can now be dropped for the sake of brevity. We re-write (7), using (6), after dropping the labels as

$$\mathbf{x}_R = \alpha\mathbf{W}\tilde{\mathbf{y}}_R = \alpha\mathbf{W}\tilde{\mathbf{G}}\mathbf{x} + \alpha\mathbf{W}\mathbf{G}_{\mathbf{RR}}\tilde{\mathbf{x}}_R + \alpha\mathbf{W}\mathbf{z}_R. \tag{8}$$

The relay's transmit signal should satisfy its transmit power constraint, so we have

$$P_R = \mathrm{Tr}\left\{\mathbb{E}\left[\mathbf{x}_R\mathbf{x}_R^H\right]\right\} = \mathbb{E}\left[\|\alpha\mathbf{W}\tilde{\mathbf{G}}\mathbf{x}\|^2\right] + \mathbb{E}\left[\|\alpha\mathbf{W}\mathbf{G}_{\mathbf{RR}}\tilde{\mathbf{x}}_R\|^2\right] + \mathbb{E}\left[\|\alpha\mathbf{W}\mathbf{z}_R\|^2\right], \tag{9}$$

which leads to

$$\alpha = \sqrt{\frac{P_R}{\mathbb{E}\left[\|\mathbf{W}\tilde{\mathbf{G}}\mathbf{x}\|^2\right] + \mathbb{E}\left[\|\mathbf{W}\mathbf{G}_{\mathbf{RR}}\tilde{\mathbf{x}}_R\|^2\right] + \mathbb{E}\left[\|\mathbf{W}\mathbf{z}_R\|^2\right]}}. \tag{10}$$

We next re-formulate the signal received at the user S_k in (2), by using (8), as

$$y_k = \alpha\mathbf{f}_k^T\mathbf{W}\tilde{\mathbf{G}}\mathbf{x} + \alpha\mathbf{f}_k^T\mathbf{W}\mathbf{G}_{\mathbf{RR}}\tilde{\mathbf{x}}_R + \alpha\mathbf{f}_k^T\mathbf{W}\mathbf{z}_R + \sum_{i,k\in Uk}\Omega_{k,i}\sqrt{p_i}x_i + z_k$$

$$= \underbrace{\alpha\mathbf{f}_k^T\mathbf{W}\sqrt{p_{k'}}\mathbf{g}_{k'}x_{k'}}_{\text{desired signal}} + \underbrace{\alpha\mathbf{f}_k^T\mathbf{W}\sqrt{p_k}\mathbf{g}_k x_k}_{\text{self-interference}} + \underbrace{\alpha\mathbf{f}_k^T\mathbf{W}\sum_{i\neq k,k'}^{2K}\sqrt{p_i}\mathbf{g}_i x_i}_{\text{inter-pair interference}} + \underbrace{\alpha\mathbf{f}_k^T\mathbf{W}\mathbf{G}_{\mathbf{RR}}\tilde{\mathbf{x}}_R}_{\text{amplified self-loop interference}}$$

$$+ \underbrace{\alpha\mathbf{f}_k^T\mathbf{W}\mathbf{z}_R}_{\text{amplified noise from relay}} + \underbrace{\sum_{i,k\in Uk}\Omega_{k,i}\sqrt{p_i}x_i}_{\substack{\text{self-loop interference} \\ \text{and inter-user interference}}} + \underbrace{z_k}_{\text{AWGN at } S_k}. \tag{11}$$

Here we have $(k, k') = (2m-1, 2m)$ or $(2m, 2m-1)$, where $m = 1, \cdots, K$ denotes the user pair, which exchange information with one another.

In this work we assume that the relay estimates the channels $\mathbf{G}$ and $\mathbf{F}$ and uses them to design the precoder $\mathbf{W}$. The relay then transmits the self-interference cancellation (SIC) coefficient $\hat{\mathbf{f}}_k^T\mathbf{W}\hat{\mathbf{g}}_k$ for each user, where $\hat{\mathbf{f}}_k$ and $\hat{\mathbf{g}}_k$ are the estimated channel coefficients. The signal received

at user S_k after SIC is

$$\tilde{y}_k = \underbrace{\alpha \mathbf{f}_k^T \mathbf{W} \sqrt{p_{k'}} \mathbf{g}_{k'} x_{k'}}_{\text{desired signal}} + \underbrace{\alpha \sqrt{p_k} \lambda_k x_k}_{\text{residual interference}} + \underbrace{\alpha \mathbf{f}_k^T \mathbf{W} \sum_{i \neq k,k'}^{2K} \sqrt{p_i} \mathbf{g}_i x_i}_{\text{inter-pair interference}} + \underbrace{\alpha \mathbf{f}_k^T \mathbf{W} \mathbf{G_{RR}} \tilde{\mathbf{x}}_R}_{\text{amplified self-loop interference}}$$

$$+ \underbrace{\alpha \mathbf{f}_k^T \mathbf{W} \mathbf{z}_R}_{\text{amplified noise from relay}} + \underbrace{\sum_{i,k \in Uk} \Omega_{k,i} \sqrt{p_i} x_i}_{\substack{\text{self-loop interference} \\ \text{and inter-user interference}}} + \underbrace{z_k}_{\text{AWGN at } S_k} . \tag{12}$$

Here $\lambda_k = \mathbf{f}_k^T \mathbf{W} \mathbf{g}_k - \hat{\mathbf{f}}_k^T \mathbf{W} \hat{\mathbf{g}}_k$ is the residual self-interference. Before designing the relay's precoder $\mathbf{W}$, we briefly digress to discuss the MMSE channel estimation process.

III. CHANNEL ESTIMATION

During the channel estimation phase, all $2K$ users simultaneously transmit pilot sequences of length $\tau \leq T$ symbols to the relay, where T is the channel's coherence interval. Let $\sqrt{\tau P_\rho} \boldsymbol{\varphi} \in \mathbb{C}^{2K \times \tau}$ denote the pilot symbols transmitted from $2K$ users with P_ρ being the transmit power of each pilot symbol. The pilots are assumed to be orthogonal so that $\boldsymbol{\varphi} \boldsymbol{\varphi}^H = \mathbf{I}_{2K}$, which requires that $\tau \geq 2K$ [31]. The MMSE channel estimates $\hat{\mathbf{G}}$ and $\hat{\mathbf{F}}$ are given as [13], [32]

$$\mathbf{G} = \hat{\mathbf{G}} + \mathbf{E}_g, \text{ and } \mathbf{F} = \hat{\mathbf{F}} + \mathbf{E}_f. \tag{13}$$

Here $\mathbf{E}_g$ and $\mathbf{E}_f$ represents the estimation error matrices of $\mathbf{G}$ and $\mathbf{F}$, respectively. The estimated channel matrices $\hat{\mathbf{G}}$ and $\hat{\mathbf{F}}$ are independent of the error matrices $\mathbf{E}_g$ and $\mathbf{E}_f$, respectively [32]. The rows of the matrices $\hat{\mathbf{G}}$ and $\hat{\mathbf{F}}$ are distributed as $\mathcal{CN}(0, \hat{\mathbf{D}}_u)$ and $\mathcal{CN}(0, \hat{\mathbf{D}}_d)$ respectively, where $\hat{\mathbf{D}}_u = \text{diag}\{\hat{\sigma}_{g,1}^2, \hat{\sigma}_{g,2}^2, \cdots, \hat{\sigma}_{g,2K}^2\}$ and $\hat{\mathbf{D}}_d = \text{diag}\{\hat{\sigma}_{f,1}^2, \hat{\sigma}_{f,2}^2, \cdots, \hat{\sigma}_{f,2K}^2\}$, with $\hat{\sigma}_{g,k}^2 = \frac{\tau P_\rho \sigma_{g,k}^4}{\tau P_\rho \sigma_{g,k}^2 + 1}$ and $\hat{\sigma}_{f,k}^2 = \frac{\tau P_\rho \sigma_{f,k}^4}{\tau P_\rho \sigma_{f,k}^2 + 1}$ [32]. Hence $\mathbf{E}_g \sim \mathcal{CN}(0, \mathbf{D}_u - \hat{\mathbf{D}}_u)$ and $\mathbf{E}_f \sim \mathcal{CN}(0, \mathbf{D}_d - \hat{\mathbf{D}}_d)$, with $\mathbf{D}_u - \hat{\mathbf{D}}_u = \text{diag}\{\sigma_{\xi,g,1}^2, \sigma_{\xi,g,2}^2, \sigma_{\xi,g,3}^2, \cdots, \sigma_{\xi,g,2K}^2\}$ and $\mathbf{D}_d - \hat{\mathbf{D}}_d = \text{diag}\{\sigma_{\xi,f,1}^2, \sigma_{\xi,f,2}^2, \sigma_{\xi,f,3}^2, \cdots, \sigma_{\xi,f,2K}^2\}$ with $\sigma_{\xi,g,k}^2 = \frac{\sigma_{g,k}^2}{\tau P_\rho \sigma_{g,k}^2 + 1}$ and $\sigma_{\xi,f,k}^2 = \frac{\sigma_{f,k}^2}{\tau P_\rho \sigma_{f,k}^2 + 1}$.

IV. RELAY PRECODER DESIGN

We design our relay precoder based on: i) MRC/MRT; and ii) ZFR/ZFT processing.

A. MRC/MRT processing

The MRC/MRT precoder using the estimated CSI is formulated as

$$\mathbf{W} = \hat{\mathbf{F}}^* \mathbf{T} \hat{\mathbf{G}}^H, \tag{14}$$

where $\mathbf{T} = \text{blkdiag}\{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \cdots, \mathbf{T}_K\}$. The permutation matrix $\mathbf{T}_m = [0\,1; 1\,0]$ permutes the transmit data of each user pair to ensure that the transmit data reaches the intended receiver. We next state the following proposition, whose proof is relegated to Appendix A, to simplify the scaling factor α in (10).

Proposition 1. For the MRC/MRT precoder, we have

$$\alpha = \sqrt{\frac{P_R}{N^2 \left(\Psi + \sigma_{nr}^2 + P_R \sigma_{LIR}^2\right)\hat{\hat{\Phi}} + N^3 \hat{\Upsilon}}}, \tag{15}$$

where $\Psi = \sum_{i=1}^{K}\left(p_{2i-1}\sigma_{g,2i-1}^2 + p_{2i}\sigma_{g,2i}^2\right), \hat{\hat{\Phi}} = \sum_{i=1}^{K}\left(\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2 + \hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2\right)$ and

$\hat{\Upsilon} = \sum_{i=1}^{K}\left(p_{2i-1}\hat{\sigma}_{g,2i-1}^4\hat{\sigma}_{f,2i}^2 + p_{2i}\hat{\sigma}_{g,2i}^4\hat{\sigma}_{f,2i-1}^2\right).$

B. ZFR/ZFT processing

The ZFR/ZFT precoder using the estimated CSI is formulated as

$$\mathbf{W} = \hat{\tilde{\mathbf{F}}}^* \mathbf{T} \hat{\tilde{\mathbf{G}}}^H, \tag{16}$$

where $\hat{\tilde{\mathbf{F}}} = \hat{\mathbf{F}}\left(\hat{\mathbf{F}}^H\hat{\mathbf{F}}\right)^{-1}$ and $\hat{\tilde{\mathbf{G}}} = \hat{\mathbf{G}}\left(\hat{\mathbf{G}}^H\hat{\mathbf{G}}\right)^{-1}$. In the next proposition, which is proved in Appendix B, we simplify the scaling factor α in (10).

Proposition 2. For the ZFR/ZFT precoder, we have

$$\alpha = \sqrt{\frac{P_R}{\hat{\lambda} + \hat{\eta}\left(\sum_{i=1}^{2K} p_i\sigma_{\xi,g,i}^2 + \sigma_{nr}^2 + P_R\sigma_{LIR}^2\right)}}, \tag{17}$$

where $\hat{\lambda} = \sum_{i=1}^{2K}\frac{p_{i'}}{(N-2K-1)\hat{\sigma}_{f,i}^2}$, $\hat{\eta} = \sum_{j=1}^{2K}\frac{1}{(N-2K-1)^2\hat{\sigma}_{f,j}^2\hat{\sigma}_{g,j'}^2}.$

V. ACHIEVABLE SUM-RATE OF MRC/MRT AND ZFR/ZFT PRECODERS

In this section, we calculate lower bounds on the instantaneous sum-rate for both MRC/MRT and ZFR/ZFT precoders. The instantaneous SNR_k at the user S_k can be expressed using (12) as in

$$\text{SNR}_k = \frac{p_{k'}|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}|}{p_k|\lambda_k|^2 + \sum_{i\neq k,k'}^{2K}|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_i|^2 + \|\mathbf{f}_k^T\mathbf{W}\mathbf{G}_{RR}\|^2\frac{P_R}{N} + \|\mathbf{f}_k^T\mathbf{W}\|^2\sigma_{n_R}^2 + \frac{1}{\alpha^2}\sum_{i,k\in Uk}\sigma_{k,i}^2 p_i + \frac{1}{\alpha^2}\sigma_n^2}. \tag{18}$$

The sum-rate of the system is $R = \mathbb{E}\left\{\sum_{k=1}^{2K}\log_2\left(1 + \text{SNR}_k\right)\right\}. \tag{19}$

Next we derive a lower bound on the achievable rate using the method of [33]. For the (k, k') pair, the signal received by the kth user can be written as (see (12))

$$\tilde{y}_k = \underbrace{\alpha\sqrt{p_{k'}}\mathbb{E}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right]x_{k'}}_{\text{desired signal}} + \underbrace{\tilde{n}_k}_{\text{effective noise}}, \tag{20}$$

where $\tilde{n}_k = \alpha\sqrt{p_{k'}}\left(\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'} - \mathbb{E}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right]\right)x_{k'} + \alpha\sqrt{p_k}\lambda_k x_k + \alpha\mathbf{f}_k^T\mathbf{W}\sum_{i\neq k,k'}^{2K}\sqrt{p_i}\mathbf{g}_i x_i$

$\qquad + \alpha\mathbf{f}_k^T\mathbf{W}\mathbf{G}_{\mathbf{RR}}\tilde{\mathbf{x}}_R + \alpha\mathbf{f}_k^T\mathbf{W}\mathbf{z}_R + \sum_{i,k\in Uk}\Omega_{k,i}\sqrt{p(k)}x_i + z_k. \tag{21}$

The value of $\mathbb{E}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right]$ can be calculated by exploiting the knowledge of channel distribution. We observe that the desired signal and the effective noise are uncorrelated. Similar to [34], [35],

we only exploit the knowledge of $\mathbb{E}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right]$ in the detection. We use the central limit theorem to treat the uncorrelated additive noise $\tilde{n}(k)$ as the worst-case Gaussian noise, when computing the sum-rate. For massive MIMO systems, the central limit theorem provides a tight statistical lower bound on the achievable rate. This fact is extensively exploited for deriving the SE and EE expressions in massive MIMO systems [13], [20], [35]. The lower bound on the achievable sum-rate, consequently, becomes

$$R_{\text{lower}} = \left[\sum_{k=1}^{2K} \log_2\left(1 + \text{SNR}_{k,\text{lower}}\right)\right], \tag{22}$$

where

$$\text{SNR}_{k,\text{lower}} = \frac{\alpha^2 p_{k'}\left|\mathbb{E}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right]\right|^2}{\alpha^2 p_{k'}\text{var}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right] + \alpha^2 p_k\text{SI}_k + \alpha^2\text{IP}_k + \alpha^2\text{NR}_k + \alpha^2\text{LIR}_k + \text{UI}_k + \text{NU}_k}. \tag{23}$$

In (23), the residual self-interference after SIC (SI), the inter-pair interference (IP), the amplified noise from the relay (NR), the amplified self-loop interference at relay (LIR), self-loop interference and inter-user interference (UI), and the noise at the user (NU), are given as follows:

$$\begin{aligned}
\text{SI}_k &= \mathbb{E}\left[|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_k - \hat{\mathbf{f}}_k^T\mathbf{W}\hat{\mathbf{g}}_k|^2\right], \ \text{IP}_k = \sum_{\substack{i \neq k,k'}}^{2K} p_i\mathbb{E}\left[|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_i|^2\right], \\
\text{NR}_k &= \mathbb{E}\left[|\mathbf{f}_k^T\mathbf{W}\mathbf{z}_R|^2\right], \ \text{LIR}_k = \mathbb{E}\left[|\mathbf{f}_k^T\mathbf{W}\mathbf{G}_{RR}\tilde{\mathbf{x}}|^2\right], \\
\text{UI}_k &= \sum_{i,k\in Uk} p_i\mathbb{E}\left[|\Omega_{k,i}x_i|^2\right], \ \text{NU}_k = \mathbb{E}\left[|z_k|^2\right].
\end{aligned} \tag{24}$$

We further simplify the $\text{SNR}_{k,\text{lower}}$ expressions for both the MRC/MRT and ZFR/ZFT precoders in the following theorems.

Theorem 1. The achievable rate of user S_k for a finite number of receive antennas at the relay relying on MMSE channel estimate based MRC/MRT processing is lower bounded as follows: $\log_2\left\{1 + \text{SNR}_k^{\text{mrc}}(p_k, P_R)\right\}$, where we have:

$$\text{SNR}_k^{\text{mrc}}(p_k, P_R) =$$

$$\frac{a_k p_{k'}}{\sum_{i=1}^{2K}\left(b_{k,i}^{(1)} + b_{k,i}^{(2)}P_R^{-1} + \sum_{i,k\in U_k} p_i P_R^{-1}b_{k,i}^{(3)}\right)p_i + c_k p_k + \left(d_k^{(1)} + d_k^{(2)}P_R + d_k^{(3)}P_R^{-1}\right) + \sum_{i,k\in U_k} p_i f_{k,i}(P_R)}, \tag{25}$$

where $a_k = N^2\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,k'}^4$, $b_{k,i}^{(1)} = \eta_{k,i} = \hat{\Phi}\sigma_{f,k}^2\sigma_{g,i}^2 + N\left(\sigma_{f,k}^2\hat{\sigma}_{g,i}^4\hat{\sigma}_{f,i'}^2 + \sigma_{g,i}^2\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,k'}^2\right)$,
$b_{k,i}^{(2)} = \sigma_n^2\left(\hat{\Phi}\sigma_{g,i}^2 + N\hat{\sigma}_{g,i}^4\hat{\sigma}_{f,i'}^2\right)$, $b_{k,i}^{(3)} = \sigma_{k,i}^2\left(\hat{\Phi}\sigma_{g,i}^2 + N\hat{\sigma}_{g,i}^4\hat{\sigma}_{f,i'}^2\right)$,
$c_k = -\left(\hat{\Phi}\hat{\sigma}_{f,k}^2\hat{\sigma}_{g,k}^2 + N\left(\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,k}^2\hat{\sigma}_{g,k'}^2 + \hat{\sigma}_{f,k}^2\hat{\sigma}_{g,k}^4\hat{\sigma}_{f,k'}^2\right)\right)$,
$d_k^{(1)} = \left(\sigma_{LIR}^2\sigma_n^2 + \sigma_{nr}^2\sigma_{f,k}^2\right)\hat{\Phi} + N\sigma_{nr}^2\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,k'}^2$, $d_k^{(2)} = \sigma_{LIR}^2\left(\sigma_{f,k}^2\hat{\Phi} + N\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,k'}^2\right)$, $d_k^{(3)} = \sigma_{nr}^2\sigma_n^2\hat{\Phi}$,
$f_{k,i}(P_R) = \left(P_R^{-1}e_{k,i}^{(1)} + e_{k,i}^{(2)}\right)$, $e_{k,i}^{(1)} = \sigma_{nr}^2\sigma_{k,i}^2\hat{\Phi}$, $e_{k,i}^{(2)} = \sigma_{LIR}^2\sigma_{k,i}^2\hat{\Phi}$.

Proof: Refer to Appendix C. ∎

Theorem 2. The achievable rate of user S_k for a finite number of receive antennas at the relay with MMSE channel estimate based ZFR/ZFT processing is lower bounded as $\log_2\left\{1 + \mathrm{SNR}_k^{\mathrm{zf}}(p_k, P_R)\right\}$, where

$$\mathrm{SNR}_k^{\mathrm{zf}}(p_k, P_R) =$$

$$\frac{u_k p_{k'}}{\sum_{i=1}^{2K}\left(d_{k,i}^{(1)} + d_{k,i}^{(2)} P_R^{-1} + \sum_{i,k \in U_k} p_i P_R^{-1} d_{k,i}^{(3)}\right) p_i + \left(v_k^{(1)} + v_k^{(2)} P_R + v_k^{(3)} P_R^{-1}\right) + \sum_{i,k \in U_k} p_i u_{k,i}(P_R)}, \quad (26)$$

where $u_k = 1$, $d_{k,i}^{(1)} = \frac{1}{(N-2K-1)}\left(\frac{\sigma_{\xi,f,k}^2}{\hat{\sigma}_{f,i'}^2} + \frac{\sigma_{\xi,g,i}^2}{\hat{\sigma}_{g,k'}^2}\right) + \sigma_{\xi,f,k}^2 \sigma_{\xi,g,i}^2 \hat{\eta}$,

$d_{k,i}^{(2)} = \sigma_n^2\left(\frac{1}{(N-2K-1)\hat{\sigma}_{f,i'}^2} + \hat{\eta}\sigma_{\xi,g,i}^2\right)$, $d_{k,i}^{(3)} = \sigma_{k,i}^2\left(\frac{1}{(N-2K-1)\hat{\sigma}_{f,i'}^2} + \hat{\eta}\sigma_{\xi,g,i}^2\right)$,

$v_k^{(1)} = \sigma_{nr}^2\left(\frac{1}{(N-2K-1)\hat{\sigma}_{g,k'}^2} + \sigma_{\xi,f,k}^2 \hat{\eta}\right) + \hat{\eta}\sigma_{LIR}^2\sigma_n^2$, $v_k^{(2)} = \sigma_{LIR}^2\left(\frac{1}{(N-2K-1)\hat{\sigma}_{g,k'}^2} + \sigma_{\xi,f,k}^2 \hat{\eta}\right)$,

$v_k^{(3)} = \hat{\eta}\sigma_{nr}^2\sigma_n^2$, $u_{k,i}(P_R) = \left(w_{k,i}^{(1)} + P_R^{-1} w_{k,i}^{(2)}\right)$, $w_{k,i}^{(1)} = \hat{\eta}\sigma_{k,i}^2\sigma_{LIR}^2$, $w_{k,i}^{(2)} = \hat{\eta}\sigma_{k,i}^2\sigma_{nr}^2$

Proof: Refer to Appendix D. ∎

VI. ENERGY-EFFICIENT OPTIMIZATION

We now optimally allocate the power for maximizing the EE subject to the rate required by each user. The EE (in bits/Joule/Hz) is defined as [14], [24]

$$\mathrm{EE} = \frac{\tilde{R}(p_k, P_R)}{P_T(p_k, P_R)}. \quad (27)$$

The numerator in the EE is the SE, which also includes the channel estimation overhead, and is given by

$$\tilde{R}(p_k, P_R) = \left(1 - \frac{\tau}{T}\right)\sum_{k=1}^{2K}\log_2\left(1 + \mathrm{SNR}_k^{\zeta}(p_k, P_R)\right). \quad (28)$$

The expression of SNR_k^{ζ}, where $\zeta \in (\mathrm{mrc}, \mathrm{zf})$ are given in (25) and (26), respectively. The denominator in the EE denotes the overall power consumed by the system [26]

$$P_T(p_k, P_R) = \sum_{k=1}^{2K} p_k + P_R + P_c. \quad (29)$$

The terms p_k and P_R denote the transmit power of user S_k and of the relay, respectively. The term P_c represents the fixed circuit power used by the system.

Remark 2. We assume the power amplifiers efficiency to be unity both at the user and at the relay as in [26], [36], [37] for mathematical simplicity. Also, P_c takes into account the power required by the different components, namely by the transceiver's radio-frequency chain, the oscillator and the power consumption of channel estimation [26], [36], [37]. This assumption is commonly exploited in the massive MIMO literature for mathematical simplicity [26], and it does not affect the overall behavior of the system considered.

Before formulating the related optimization problem, we briefly discuss the terminologies used in geometric and fractional programming from [38] and [39] respectively, which will be used in the sequel. Geometric and fractional programming have earlier been used for power allocation in [39]–[42].

A real valued function $f(x)$ of the form $f(x) = c\, x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$, where $c > 0$, $a_i \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n_{++}$, is referred to as a monomial function. The sum of one or more monomials, i.e. $f(x) = \sum_{m=1}^{K} c_m\, x_1^{a_{1m}} x_2^{a_{2m}} \cdots x_n^{a_{nm}}$, where $c_m > 0$ is termed as a posynomial. Monomials are closed both under multiplication and division, whereas posynomials are closed under addition and multiplication, but not under division. The ratio of a posynomial and monomial is a posynomial. A geometric program (GP) has a posynomial objective and upper bounded posynomial inequality constraints.

A fractional program is of the form $u(x)/v(x)$, so that the optimization variable obeys $x \in \mathcal{X}$, where $u : \mathcal{C} \subset \mathbb{R}^n \to \mathbb{R}$, $v : \mathcal{C} \subset \mathbb{R}^n \to \mathbb{R}_{x+}$ and $\mathcal{X} \subset \mathcal{C} \subset \mathbb{R}^n$. Since the objective is a fraction, the problem is not guaranteed to be always convex, even if both u and v are affine functions. For maximizing fractions, mostly two classes of generalized concave functions, namely quasi-concave (QC) functions and pseudo-concave (PC) functions are used. If $\mathcal{C} \subset \mathbb{R}^n$ is a convex set, then $r : \mathcal{C} \to \mathbb{R}$ is QC if $r(\lambda x_1 + (1 - \lambda)x_2) \geq \min\{r(x_1, x_2)\}$ for all $x_1, x_2 \in \mathcal{C}$ and $\lambda \in [0; 1]$. Similarly, $r : \mathcal{C} \to \mathbb{R}$ is PC if it is differentiable and, $r(x_2) < r(x_1) \Rightarrow \nabla(r(x_2))^T(x_1 - x_2) > 0$ for all $x_1, x_2 \in \mathcal{C}$. A local maximum of optimization associated with the PC objective constitutes a global maximum, whereas under a QC objective it is not necessarily a global maximum. For example, when the objective has a concave numerator and a convex denominator, the fractional program is a PC and its stationary point is its global maximizer. Such a problem belongs to the class of concave-convex fractional programs (CCFP) [39].

With this information in mind, we optimize the EE in the next section.

A. EE maximization

The EE maximization problem is formulated as

$$\textbf{P1 :} \underset{p_k, P_R}{\text{Max}} \; \frac{\left(1 - \frac{\tau}{T}\right) \sum\limits_{k=1}^{2K} \log_2 \left(1 + \text{SNR}_k^\zeta(p_k, P_R)\right)}{\sum\limits_{k=1}^{2K} p_k + P_R + P_c} \tag{30a}$$

$$\text{s.t. } 0 \leq p_k \leq P^{\max},\, 0 \leq P_R \leq P_R^{\max} \tag{30b}$$

$$\sum_{k=1}^{2K} p_k + P_R \leq P_t^{\max}. \tag{30c}$$

The first two constraint specify the peak transmit power of the user and the relay i.e. P^{max} and P_R^{max} respectively. The last constraint impose a constraint of P_t^{max} on the total system transmit power. The problem **P1** can be re-cast as

$$\textbf{P2}: \underset{p_k, P_R, \Gamma_k}{\text{Max}} \quad \frac{\left(1 - \frac{\tau}{T}\right) \log_2 \prod_{k=1}^{2K} \left(1 + \Gamma_k\right)}{\sum_{k=1}^{2K} p_k + P_R + P_c} \tag{31a}$$

$$\text{s.t.} \quad 0 \leq p_k \leq P^{\text{max}}, 0 \leq P_R \leq P_R^{\text{max}} \tag{31b}$$

$$\sum_{k=1}^{2K} p_k + P_R \leq P_t^{\text{max}} \tag{31c}$$

$$\text{ISNR}_k^\zeta(p_k, P_R) \leq \Gamma_k^{-1}. \tag{31d}$$

The symbol $\Gamma_k = \text{SNR}_k^\zeta$ denotes an auxiliary variable and the term ISNR_k^ζ denotes the inverse of SNR_k^ζ, i.e. $\text{ISNR}_k^\zeta = 1/\text{SNR}_k^\zeta$. We observe from (25) and (26) that the SNR_k^ζ is a ratio of a monomial and of a posynomial, the numerator of the objective in **P2** becomes a ratio of two posynomials, which is not a posynomial and hence non-convex. We also note that the constraints in **P2** are upper-bounded posynomials and are therefore convex. We now approximate the numerator as a monomial such that **P2** becomes a CCFP. To this end, we use the following lemma from [43].

Lemma 1. Consider a monomial function $q(\nu_k) = \delta_k \nu_k^{\alpha_k}$ $(\nu_k > 0)$, which is used for approximating $s(\nu_k) = 1 + \nu_k$ near an arbitrary point $\tilde{\nu}_k > 0$. For the above approximation, the following two conditions hold.

1) For the best monomial local approximation, the parameters α_k and δ_k are given by

$$\alpha_k = \tilde{\nu}_k(1 + \tilde{\nu}_k)^{-1}, \ \delta_k = \tilde{\nu}_k^{-\alpha_k}(1 + \tilde{\nu}_k). \tag{32}$$

2) For all $\nu_k > 0$, $s(\nu_k) \geq q(\nu_k)$.

Using Lemma 1, the numerator of the objective function can be approximated as $\log_2 \prod_{k=1}^{2K} \left[\delta_k(p_k, P_R)\Gamma_k^{\alpha_k(p_k, P_R)}\right]$, where $\alpha_k(p_k, P_R) = \tilde{\Gamma}_k \left(1 + \tilde{\Gamma}_k\right)^{-1}$ and $\delta_k(p_k, P_R) = \left(\tilde{\Gamma}_k\right)^{-\alpha_k(p_k, P_R)} \left(1 + \tilde{\Gamma}_k\right)$, where $\tilde{\Gamma}_k$ is an initial value approximation for Γ_k. Given the approximated objective, the optimization **P2** can be formulated as follows.

$$\textbf{P3}: \underset{p_k, P_R, \Gamma_k}{\text{Max}} \quad \frac{\log_2 \prod_{k=1}^{2K} \left[\delta_k(p_k, P_R)\Gamma_k^{\alpha_k(p_k, P_R)}\right]}{\sum_{k=1}^{2K} p_k + P_R + P_c} \tag{33}$$

$$\text{s.t.} \quad (31b), (31c), (31d).$$

Here we have dropped the constant $\left(1 - \frac{\tau}{T}\right)$ from the objective. The optimization **P3** is now a CCFP. We use the following result, proved in [39], [44], to solve it.

Proposition 3. Consider the CCFP $g(x) = u(x)/v(x)$, with u being non-negative, differentiable and concave, while v being positive, differentiable and convex. Then the function $g(x)$ is a PC and a stationary point x^* of $g(x)$ is its global maximizer. The problem of maximizing $g(x)$ is equivalent to finding the positive zero of $D(\lambda)$, which is defined as

$$D(\lambda) \triangleq \max_x \left\{ u(x) - \lambda v(x) \right\}. \tag{34}$$

The function $D(\lambda)$ is convex, continuous and strictly monotonically decreasing and its zero is found using Dinkelbach's algorithm [44].

Let us now exploit the monomial approximation and Dinkelbach's algorithm to solve the EE problem **P3**, as illustrated in Algorithm 1[2].

Remark 3. The first inequality constraint in Algorithm 1, also termed as the trust region constraint [38], is added to confine the domain of variable Γ to a region around the current guess $\tilde{\Gamma}^{[m]}$. In most practical cases, $\alpha = 1.1$ provides a good accuracy/speed trade-off for the monomial approximations [38], [43].

B. Energy efficient optimization relying on the max-min approach

We next maximize the EE of the least energy-efficient user using the classic fractional programming approach. The max-min optimization can be cast as

$$\mathbf{P4} : \underset{p_k, P_R, \Gamma_k}{\text{Max}} \quad \underset{1 \le k \le 2K}{\min} \frac{\log_2 \left[\delta_k(p_k, P_R) \Gamma_k^{\alpha_k(p_k, P_R)} \right]}{p_k + \dfrac{P_R + P_c}{2K}} \tag{36}$$

$$\text{s.t.} \quad (31b), (31c), (31d).$$

Here $\delta_k(p_k, P_R)$ and $\alpha_k(p_k, P_R)$ are calculated similar to (32). The problem **P4** now belongs to a class of a max-min fractional program (MMFP). We next use the following proposition from [30], [39] to solve our MMFP.

---

[2]The Algorithm 1 uses the monomial, obtained using Lemma 1, to cast the problem **P3** as a GP. Note that this algorithm is not optimal as it approximates the posynomial with a monomial. It is a heuristic that often computes the globally optimal power allocation [45]. The monotonic optimization, which obtains global optimal solution in finite time [45], can also be used to solve problem P2 without any approximation. But the worst-case complexity of monotonic optimization increases at least exponentially in $2K$ [45]. The proposed sequential fractional solution can solve the problem in polynomial time with affordable complexity. Further the proposed algorithm – as concluded by exhaustive numerical studies in [45] – is an excellent heuristic that achieves a global optimum solution.

Algorithm 1: EE maximization algorithm

Input: Given a tolerance $\epsilon > 0$ and the maximum number of iterations L, set $m = 1$ and $\lambda^{[1]} = 0$. Calculate the initial values p_k, P_R and $\tilde{\Gamma}_k^{[1]}$ by allocating equal power to all users i.e. $p_k = P_R/2K$, and $P_R = P/2$.

Output: p_k^* and P_R^* as the solutions.

1 **for** $m \leftarrow 1$ **to** L **do**

2 $\quad$ Given a feasible p_k, P_R, compute $\alpha_k^{[m]}(p_k, P_R)$ and $\delta_k^{[m]}(p_k, P_R)$.

3 $\quad$ Solve the GP to calculate p_k, P_R and Γ_k.

$$\underset{p_k, P_R, \Gamma_k}{\text{Max}} \quad \left\{ \log_2 \prod_{k=1}^{2K} \delta_k^{[m]}(p_k, P_R) \Gamma_k^{\alpha_k^{[m]}(p_k, P_R)} - \lambda^{[m]} \left(\sum_{k=1}^{2K} p_k + P_R + P_c \right) \right\}$$

$$\text{s.t.} \quad \beta^{-1} \tilde{\Gamma}_k^{[m]} \leq \Gamma_k \leq \beta \tilde{\Gamma}_k^{[m]}$$

$$(31b), (31c), (31d).$$

(35)

4 $\quad$ Based on the solutions P_R and p_k of GP, compute $\tilde{\tilde{\Gamma}}_k$, and

$$D(\lambda^{[m]}) = \log_2 \prod_{k=1}^{2K} \delta_k^{[m]}(p_k, P_R) \tilde{\tilde{\Gamma}}_k^{\alpha_k^{[m]}(p_k, P_R)} - \lambda^{[m]} \left(\sum_{k=1}^{2K} p_k + P_R + P_c \right).$$

5 $\quad$ Do until convergence

$\quad$ **if** $D(\lambda_k^{[m]}) > \epsilon$ **then** $\tilde{\Gamma}_k^{[m+1]} = \tilde{\tilde{\Gamma}}_k$, and

$$\lambda^{[m+1]} = \frac{\log_2 \prod_{k=1}^{2K} \delta_k^{[m]}(p_k, P_R) \tilde{\tilde{\Gamma}}_k^{\alpha_k^{[m]}(p_k, P_R)}}{\sum_{k=1}^{2K} p_k + P_R + P_c}.$$

6 $\quad$ **else** break.

7 **return** p_k^* and P_R^*.

Proposition 4. The MMFP optimization problem is stated as

$$\underset{x}{\text{Max}} \ \underset{1 \leq k \leq 2K}{\min} \frac{u_k(x)}{v_k(x)}$$

$$\text{s.t.} \ c_i(x), \ i = 1, 2, 3, \cdots, I.$$

(37)

For a non-negative and concave u_k, positive and convex v_k, $\forall k = 1, 2, 3, \cdots, 2K$, as well as convex c_i, $\forall i = 1, 2, 3, \cdots, I$, each ratio of the objective is a QC function. Furthermore, if $z(x) = \underset{1 \leq k \leq 2K}{\min} y_k(x)$, so that $y_k(x)$ is a QC, $\forall k = 1, 2, 3, \cdots, 2K$, then $z(x)$ is also a QC function. The auxiliary function of our real variable is defined as

$$D(\lambda) \triangleq \underset{x}{\text{Max}} \left\{ \underset{1 \leq k \leq 2K}{\min} \{u(x) - \lambda v(x)\} \right\}.$$

(38)

The generalized Dinkelbach algorithm [30] solves a sequence of convex problems to obtain the global solution of MMFP, as illustrated in Algorithm 2.

Algorithm 2: Max-min EE optimization algorithm

Input: Given a tolerance $\epsilon > 0$ and the maximum number of iterations L, set $m = 1$ and $\lambda^{[1]} = 0$, calculate the initial values p_k, P_R and $\tilde{\Gamma}_k^{[1]}$ by allocating equal power to all users i.e. $p_k = P_R/2K$, and $P_R = P/2$.

Output: p_k^* and P_R^* as the solutions.

1 for $m \leftarrow 1$ **to** L **do**

2   Given a feasible p_k, P_R, compute $\alpha_k^{[m]}(p_k, P_R)$ and $\delta_k^{[m]}(p_k, P_R)$.

3   Solve the GP to calculate p_k, P_R and Γ_k.

$$\underset{p_k, P_R, \Gamma_k}{\text{Max}} \quad \min \left\{ \log_2 \left[\delta_k^{[m]}(p_k, P_R) \Gamma_k^{\alpha_k^{[m]}(p_k, P_R)} \right] - \lambda^{[m]} \left(p_k + \frac{P_R + P_c}{2K} \right) \right\}$$
$$\text{s.t.} \quad (31b), (31c), (31d)$$
$$\beta^{-1} \tilde{\Gamma}_k^{[m]} \leq \Gamma_k \leq \beta \tilde{\Gamma}_k^{[m]}.$$

4   Based on the solutions P_R and p_k of GP, compute $\tilde{\Gamma}_k$, and

$$D(\lambda^{[m]}) = \min_{1 \leq k \leq 2K} \left\{ \log_2 \left[\delta_k^{[m]}(p_k, P_R) \tilde{\Gamma}_k^{\alpha_k^{[m]}(p_k, P_R)} \right] - \lambda^{[m]} \left(p_k + \frac{P_R + P_c}{2K} \right) \right\}.$$

5   Update $\tilde{\Gamma}_k^{[m+1]} = \tilde{\Gamma}_k$, and

$$\lambda^{[m+1]} = \min_{1 \leq k \leq 2K} \frac{\log_2 \left[\delta_k^{[m]}(p_k, P_R) \tilde{\Gamma}_k^{\alpha_k^{[m]}(p_k, P_R)} \right]}{p_k + \frac{P_R + P_c}{2K}}.$$

6 return p_k^* and P_R^*.

VII. SIMULATION RESULTS

Let us now investigate the EE and the max-min algorithms using Monte-Carlo simulations to highlight their advantage. For this study, we set the length of the coherence interval to $T = 200$ symbols and the pilot length to $\tau = 2K$. We also set the large scale fading matrices $\mathbf{D}_u$ and $\mathbf{D}_d$, similar to [17], [20] as

$$\mathbf{D}_u = \text{diag} [0.749, 0.045, 0.246, 0.121, 0.125, 0.142, 0.635, 0.256, 0.021, 0.123], \text{ and}$$
$$\mathbf{D}_d = \text{diag} [0.257, 0.856, 1.000, 0.899, 0.014, 0.759, 0.315, 0.432, 0.195, 0.562]. \quad (39)$$

Furthermore, the maximum transmit power of each user is $P^{\max}$=10 dBm, the maximum relay transmit power is $P_R^{\max}$=23 dBm, and the circuit power is P_c=30 dBm [20]. For this analysis, we consider K=5 user pairs and $N = 500$ relay antennas. The noise variances are set to $\sigma_n^2 = \sigma_{nr}^2 = \sigma^2$ and we define $\eta = P_t^{\max}/\sigma^2$, where $P_t^{\max}$ is the maximum total transmit power of the system.

A. EE maximization

We investigate the performance of the EE maximization algorithm for both MRC/MRT and ZFR/ZFT processing. The values of the self-loop interference σ_{LIR}^2 and inter-user interference

$\sigma_{UI}^2 \triangleq \sigma_{k,j}^2$ for $k, j = 1, \cdots, 2K$ are assumed to be 0 dB with respect to σ^2. Before investigating the performance, we have to decide the pilot transmission power P_ρ, which affects the channel estimation and consequently the overall performance. To determine P_ρ, we maximize the EE by varying P_ρ from -10 dBm to 40 dBm with two different values of η, i.e. 0 dB and 20 dB. We see from Fig. 2a that $P_\rho = 20$ dBm is capable of achieving the maximum EE for both MRC/MRT and ZFR/ZFT designs. We therefore set $P_\rho = 20$ dBm for the rest of the EE analysis. We also see that a lower pilot power degrades the ZFR/ZFT EE performance more than that of the MRC/MRT design. This is because the ZFR/ZFT design depends on the composite channels of the users.
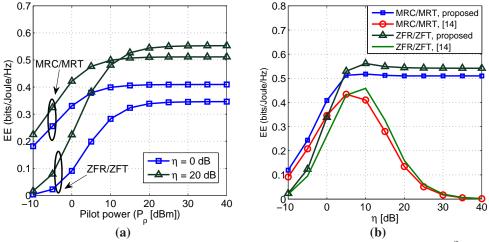
Fig. 2: EE comparison of MRC/MRT and ZFR/ZFT processing versus a) P_ρ; b) $\eta = P_t^{\max}/\sigma^2$ for $P_\rho = 20$ dBm pilot transmit power. For both these figures we have $K = 5$ user pairs and $N = 500$ relay antennas.

We plot in Fig. 2b the EE versus η for both MRC/MRT and ZFR/ZFT processing. We compare the performance of the proposed EE algorithm to [14], where Zhang *et al.* have derived the asymptotic SE and EE for two-way amplify-and-forward full-duplex relaying systems assuming that all the $2K$ users are allocated equal power. We see that the proposed algorithm yields a higher EE for both MRC/MRT and ZFR/ZFT than that in [14]. We also see that the achieved EE remains constant for $\eta \geq 7$ dB for the MRC/MRT scheme, since with $\eta = 7$ dB, the system achieves the maximum EE. After attaining the maximum EE, the system does not use any additional power, since any additional power usage would reduce the EE. We observed this undesired behavior in [14], where the system keeps using the available power beyond $\eta = 7$ dB and hence the EE reduces. We also see that for $\eta \leq 0$ dB the MRC/MRT has better EE than ZFR/ZFT.

The proposed EE optimization algorithm (Algorithm 1) can also optimize the EE under QoS

constraints that are specified as the rate r_k required by each user

$$\log_2\left(1 + \text{SNR}_k^\zeta(p_k, P_R)\right) \geq r_k. \tag{40}$$

It is readily seen that these constraints are posynomial and the GP in Step-3 of the algorithm is still a GP. We now maximize in Fig. 3a the EE for ZFR/ZFT processing under per-user QoS requirements of 0.2 bps/Hz, 0.5 bps/Hz and 0.7 bps/Hz. We see that the QoS constraints of 0.2 bps/Hz, 0.5 bps/Hz and 0.7 bps/Hz cannot be met for $\eta \leq -5$ dB, $\eta \leq 0$ dB and $\eta \leq 5$ dB, respectively. We observe that for $\eta \geq 5$ dB, the EE is the same, both with and without QoS. This implies that the system is capable of satisfying the QoS constraints, while maximizing the EE.
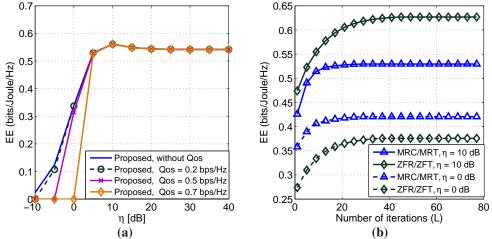

Fig. 3: a) EE versus $\eta = P_t^{\max}/\sigma^2$ for the proposed optimal power allocation subject to the QoS constraint, considering ZFR/ZFT processing associated with QoS requirements of 0.2 bps/Hz, 0.5 bps/Hz and 0.7 bps/Hz; b) EE versus the number of iterations. For both these figures we have $K = 5$, $N = 500$, $P_\rho = 20$ dBm

B. Convergence of the proposed algorithm

We now investigate the number of iterations required for the algorithm to converge for two different η values, namely for $\eta = 0$ dB, and $\eta = 10$ dB. We observe from Fig. 3b that the algorithm converges within 20 iterations for MRC/MRT and within 40 iterations for ZFR/ZFT. The algorithm has a similar convergence behavior for other η values as well.

C. EE optimization under the max-min approach

We now evaluate the performance of the max-min algorithm for ZFR/ZFT processing. For this study, we set both the self-loop interference σ_{LIR}^2 and the inter-user interference σ_{UI}^2 to be 5 dB below σ^2. Again, we have to decide the pilot transmission power P_ρ, which affects the channel estimation. To decide its value, we first solve our max-min optimization problem by varying P_ρ from -10 dBm to 40 dBm for a fixed $\eta = 10$ dB. We observe from Fig. 4a that $P_\rho = 20$ dBm attains the maximum EE for the worst user. We therefore set $P_\rho = 20$ dBm for the max-min

analysis. The max EE provided by the proposed optimization is higher than [14] for all values of pilot power. We see from Fig. 4b that the proposed algorithm significantly improves the EE of the worst user, when compared to equal power allocation. We also see that the system is capable of satisfying the QoS of 0.5 bps/Hz above $\eta = 10$ dB without compromising the EE of the worst user.
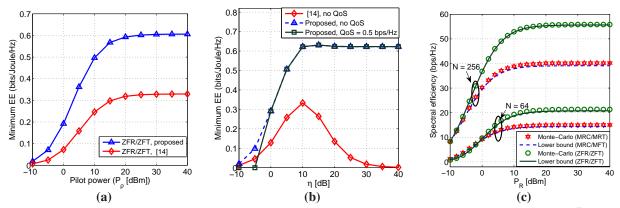

Fig. 4: Minimum EE for both equal and optimal power allocation for ZFR/ZFT processing a) versus P_ρ, where $\eta = 10$ dB; and b) versus η, where $P_\rho = 20$ dBm. For both these figures we have $K = 5$ and $N = 500$; c) SE efficiency of MRC/MRT and ZFR/ZFT versus P_R, where we have $K = 10$, $P_\rho = 10$ dBm.

In this paper, we also derived the analytical lower bounds of the achievable rate for both MRC/MRT and ZFR/ZFT processing with MMSE channel estimation which were formulated in Theorem 1 and Theorem 2. In this section, we compare the SE[3] lower bounds with their exact expressions in (19) for $N = 64$ and $N = 256$ relay antennas and for $K = 10$ user pairs. We set $P_\rho = 10$ dBm and allocate equal power to all users i.e. we have $p_i = P_R/2K, \forall i = 1, 2, 3, \cdots, 2K$. The large scale fading coefficients $\sigma_{g,i}^2 = \sigma_{f,i}^2, \forall i, j = 1, 2, 3, \cdots, 2K$, and the self-loop interference σ_{LIR}^2 as well as the inter-user interference $\sigma_{UI}^2 = \sigma^2$ are set to 0 dB with respect to σ^2. We see from Fig. 4c that the derived lower bound and the exact expression overlap for ZFR/ZFT processing for $N = 256$ relay antennas. For MRC/MRT, the lower bound marginally differs from the exact expression. We also observe that the SE, of high P_R values saturates for both MRC/MRT and ZFR/ZFT. This is because the relay self-loop interference also increases proportionally upon increasing P_R.

D. SE comparison with existing full-duplex designs

We note that the EE maximization framework considered in this work can also maximize the SE by maximizing the numerator of the Problem **P3** – the modified problem is a GP. We

---

[3]Recall from (28) that SE is obtained by scaling the achievable rate with pilot overhead.

can consequently also maximize the SE both for MRC/MRT and for ZFR/ZFT processing. In contrast to [15], which derives the lower bound and maximizes the SE (and not the EE and max-min EE), for the MRC/MRT processing alone, whilst relying on LS channel estimation. The current work is therefore more general than [15]. The objective of this study is to compare the SE of the proposed system to that of the existing full-duplex relaying system in [15], and to that of [14] where the latter has developed equal-power MRC/MRT and ZFR/ZFT for full-duplex relaying designs.
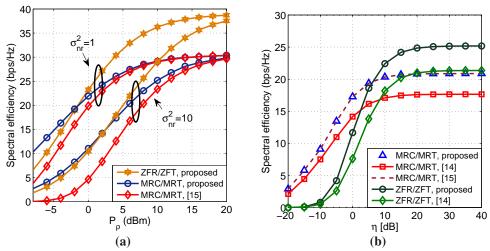

Fig. 5: SE of MRC/MRT and ZFR/ZFT versus: a) P_ρ, comparing the performance of MMSE and LS estimators, where $K = 5$, $N = 256$ and $P_R = 10$ dBm; b) η, where $K = 5$, $P_\rho = 20$ dBm, $N = 500$.

We commence by comparing the SE to that of the design in [15]. We plot in Fig. 5a the SE versus P_ρ, where we observe that for MRC/MRT processing the MMSE estimator outperforms the LS estimator [15] both for a lower pilot power P_ρ, and a higher noise power σ_{nr}^2. We also see that the ZFR/ZFT associated with $P_\rho > 0$ dB performs much better than the MRC/MRT processing.

We now evaluate the SE by varying η for the proposed algorithm considering both MRC/MRT and ZFR/ZFT processing and compare it to the designs in references [14], [15]. We see from Fig. 5b that the algorithm proposed for both the MRC/MRT and ZFR/ZFT designs considerably improves the SE, when compared to their counterparts from [14], which considers an equal power allocation. We also see that the MRC/MRT associated with the proposed algorithm outperforms even ZFR/ZFT for $\eta < 5$ dB. Furthermore, the performance of the MRC/MRT processing associated with the proposed algorithm overlaps with that of [15]. This fact can also be justified by observing Fig. 5a, where for $P_\rho = 20$ dBm, the performance of the proposed system and [15] is the same for the MRC/MRT processing.

E. Full-duplex versus half-duplex comparison

In this section, we compare both the SE and the EE of the proposed full-duplex system to those of the existing half-duplex systems [20], [21] for different values of self-loop interference σ_{LIR}^2 and inter-user interference σ_{UI}^2. The objective is to characterize the values of σ_{LIR}^2 and σ_{UI}^2 for which the proposed full-duplex relaying system using MMSE channel estimation as well as MRC/MRT and ZFR/ZFT processing has better SE and EE than their half-duplex counterparts. Reference [21] derives the asymptotic spectral and energy efficiencies of the half-duplex system by allocating equal power to all users, whereas reference [20] allocates power to maximize the SE of the half-duplex system. We observe from Fig. 6a that for the full-duplex system using MRC/MRT processing, the σ_{LIR}^2 dB and σ_{UI}^2 dB should be around 0 dB for it to achieve a better SE than the half-duplex system with optimal power allocation [20]. As expected for ZFR/ZFT processing, the full-duplex system requires much higher self-loop interference and inter-user interference suppression. We see from Fig. 6b that both σ_{LIR}^2 and σ_{UI}^2 should have values around -10 dB.
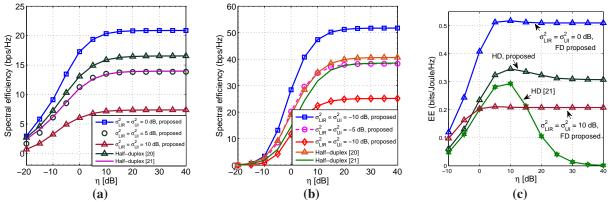
Fig. 6: Performance comparison of full- and half-duplex systems: a) SE versus $\eta = P_t^{\max}/\sigma^2$ for MRC/MRT processing b) SE versus $\eta = P_t^{\max}/\sigma^2$ for ZFR/ZFT processing; c) EE versus η for MRC/MRT processing. For all of these figures we have $K = 5$, $N = 500$ and $P_\rho = 20$ dB.

We now investigate the EE of full- and half-duplex relaying for different values of σ_{LIR}^2 and σ_{UI}^2 in conjunction with MRC/MRT processing. We note that the EE of an AF half-duplex massive MIMO relaying has not been investigated in the literature. In the proposed framework, the EE of half-duplex massive MIMO relaying can be evaluated by setting the self-loop interference and inter-user interference to zero. We observe from Fig. 6c that for $\sigma_{LIR}^2 = \sigma_{UI}^2 = 0$ dB the full duplex system has higher EE than a half-duplex system. However, the half-duplex performs better for $\sigma_{LIR}^2 = \sigma_{LIR}^2 = 10$ dB. The proposed EE power allocation algorithm, which works for a

half-duplex system, has better EE than the equal-power half-duplex MRC/MRT scheme of [21].

VIII. CONCLUSION

We considered a multi-pair AF full-duplex massive MIMO two-way relay associated with full-duplex users. We derived closed-form lower bounds for the achievable rate of both MRC/MRT and ZFR/ZFT relay processing in conjunction with MMSE channel estimation and an arbitrary number of relay antennas. We numerically showed that these lower bounds match with their exact expressions. Furthermore, we used these bounds to design algorithms for jointly allocating power for the users and the relay based on the EE maximization and max-min fairness criterion. We investigated the performance of both algorithms, with and without QoS constraints, and demonstrated their improved performance over existing full-duplex systems. We exhaustively compared the spectral and energy efficiencies of the proposed relaying system to the existing state-of-the-art half-duplex relaying systems, and characterized the values of self-loop and inter-user interference, for which the proposed system becomes capable of outperforming the existing ones.

APPENDIX A

The term $\mathbb{E}\left[\|\mathbf{W}\tilde{\mathbf{G}}\mathbf{x}\|^2\right]$ in the denominator of (10) can be expressed using (14) as

$$
\begin{aligned}
\mathbb{E}\left[\left\|\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H\tilde{\mathbf{G}}\mathbf{x}\right\|^2\right] &= \mathrm{Tr}\left\{\mathbb{E}\left[(\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H\mathbf{G}\mathbf{P}\mathbf{P}^H\mathbf{G}^H\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{F}}^T)\right]\right\} \\
&= \mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H(\hat{\mathbf{G}}+\mathbf{E}_g)\mathbf{P}\mathbf{P}^H(\hat{\mathbf{G}}^H+\mathbf{E}_g^H)\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{F}}^T\right]\right\} \\
&\overset{(a)}{=} \mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{G}}\mathbf{P}\mathbf{P}^H\hat{\mathbf{G}}^H\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{F}}^T\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H\right]\right\} \qquad (41) \\
&+ \mathrm{Tr}\left\{\mathbb{E}\left[\mathbf{E}_g\mathbf{P}\mathbf{P}^H\mathbf{E}_g^H\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{F}}^T\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H\right]\right\}. \qquad (42)
\end{aligned}
$$

The equality in (a) holds, because $\hat{\mathbf{G}}$ and $\mathbf{E}_g$ are independent and $\mathrm{Tr}(\mathbf{A}\mathbf{B}) = \mathrm{Tr}(\mathbf{B}\mathbf{A})$. We now want to simplify (41) and (42). We first expand (41) as in

$$
\begin{aligned}
&\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{G}}\mathbf{P}\mathbf{P}^H\hat{\mathbf{G}}^H\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{F}}^T\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H\right]\right\} \\
&= \mathrm{Tr}\left\{\sum_{k=1}^{2K}p_k\mathbb{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\sum_{i=1}^{K}\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{f}}_{2i}^T+\hat{\mathbf{g}}_{2i}\hat{\mathbf{f}}_{2i-1}^T\right)\left(\hat{\mathbf{f}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{f}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\right]\right\}. \qquad (43)
\end{aligned}
$$

To simplify (43), we decompose the summation therein for $k \neq 2i$ or $2i-1$, $k = 2i-1$ and $k = 2i$. For these k values, Eq. (43) can respectively be simplified as

$$
\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{f}}_{2i}^T+\hat{\mathbf{g}}_{2i}\hat{\mathbf{f}}_{2i-1}^T\right)\left(\hat{\mathbf{f}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{f}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\right]\right\} = N^2\hat{\sigma}_{g,k}^2\left(\hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2+\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2\right), \qquad (43)
$$

$$
\begin{aligned}
&\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{g}}_{2i-1}^H(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{f}}_{2i}^T+\hat{\mathbf{g}}_{2i}\hat{\mathbf{f}}_{2i-1}^T)(\hat{\mathbf{f}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{f}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H)\right]\right\} \\
&\qquad\qquad\qquad = N^2\hat{\sigma}_{g,2i-1}^2\left(\hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2+(N+1)\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2\right), \qquad (44)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{g}}_{2i}\hat{\mathbf{g}}_{2i}^H(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{f}}_{2i}^T+\hat{\mathbf{g}}_{2i}\hat{\mathbf{f}}_{2i-1}^T)(\hat{\mathbf{f}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H+\hat{\mathbf{f}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H)\right]\right\} \\
&\qquad\qquad\qquad = N^2\hat{\sigma}_{g,2i}^2\left(\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2+(N+1)\hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2\right). \qquad (45)
\end{aligned}
$$

In deriving the equalities in (43), (44), and (45), we exploit the following facts: i) for any arbitrary vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x}^* \mathbf{y}^T = \mathbf{y}\mathbf{x}^H$; and ii) $\mathbb{E}\left[\hat{\mathbf{g}}_i^H \hat{\mathbf{g}}_i\right] = N\hat{\sigma}_{g,i}^2$, $\mathbb{E}\left[\hat{\mathbf{g}}_i^H \hat{\mathbf{g}}_j\right] = 0$, $\mathbb{E}\left[\left|\hat{\mathbf{g}}_i^H \hat{\mathbf{g}}_j\right|^2\right] = N\hat{\sigma}_{g,i}^2\hat{\sigma}_{g,j}^2$, $\mathbb{E}\left[\hat{\mathbf{g}}_i \hat{\mathbf{g}}_i^H \hat{\mathbf{g}}_i \hat{\mathbf{g}}_i^H\right] = (N+1)\hat{\sigma}_{g,i}^4 \mathbf{I}_N$ [46]. Furthermore, we can simplify (43) as

$$
\begin{aligned}
&\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{G}}\mathbf{P}\mathbf{P}^H\hat{\mathbf{G}}^H\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{F}}^T\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H\right]\right\} \\
&\overset{(a)}{=} \mathrm{Tr}\left\{\sum_{k=1}^{2K} p_k \mathbb{E}\left[\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H \sum_{i=1}^{K}\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{f}}_{2i}^T + \hat{\mathbf{g}}_{2i}\hat{\mathbf{f}}_{2i-1}^T\right)\left(\hat{\mathbf{f}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H + \hat{\mathbf{f}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\right]\right\} \\
&\overset{(b)}{=} N^2 \sum_{k \neq 2i, 2i-1}^{2K} p_k \sum_{i=1}^{K} \hat{\sigma}_{g,k}^2\left(\hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2 + \hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2\right) + N^2 \sum_{i=1}^{K}\Big\{p_{2i}\hat{\sigma}_{g,2i}^2\left(\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2\right. \\
&\quad \left. +(N+1)\hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2\right)\Big\} + N^2 \sum_{i=1}^{K} p_{2i-1}\hat{\sigma}_{g,2i-1}^2\left(\hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2 + (N+1)\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2\right) \\
&\overset{(c)}{=} N^2 \sum_{k \neq 2i, 2i-1}^{2K}\sum_{i=1}^{K} p_k \hat{\sigma}_{g,k}^2 \hat{\Phi}_i + N^2 \sum_{i=1}^{K}\left(p_{2i-1}\hat{\sigma}_{g,2i-1}^2 + p_{2i}\hat{\sigma}_{g,2i}^2\right)\hat{\Phi}_i \\
&\quad + N^3 \sum_{i=1}^{K}\left(p_{2i-1}\hat{\sigma}_{g,2i-1}^4\hat{\sigma}_{f,2i}^2 + p_{2i}\hat{\sigma}_{g,2i}^4\hat{\sigma}_{f,2i-1}^2\right) \\
&\overset{(d)}{=} N^2 \sum_{j=1, j\neq i}^{K}\sum_{i=1}^{K}\left(p_{2j-1}\hat{\sigma}_{g,2j-1}^2 + p_{2j}\hat{\sigma}_{g,2j}^2\right)\hat{\Phi}_i + N^2 \sum_{i=1}^{K}\left(p_{2i-1}\hat{\sigma}_{g,2i-1}^2 + p_{2i}\hat{\sigma}_{g,2i}^2\right)\hat{\Phi}_i + N^3\hat{\Upsilon} \\
&\overset{(e)}{=} N^2 \sum_{j=1}^{K}\sum_{i=1}^{K}\left(p_{2j-1}\hat{\sigma}_{g,2j-1}^2 + p_{2j}\hat{\sigma}_{g,2j}^2\right)\hat{\Phi}_i + N^3\hat{\Upsilon} = N^2\sum_{j=1}^{K}\sum_{i=1}^{K}\hat{\Psi}_j\hat{\Phi}_i + N^3\hat{\Upsilon} = N^2\hat{\Psi}\hat{\Phi} + N^3\hat{\Upsilon}. \quad (46)
\end{aligned}
$$

We now discuss the steps used for simplifying (43). Equality in (b) is obtained by using the results in (43), (44) and (45). Equality in (c) is obtained by defining $\hat{\Phi}_i = \left(\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2 + \hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2\right)$, and by rearranging the terms. Equality in (d) is obtained by exploiting the fact that $\sum_{k=1}^{2K} x_k = \sum_{j=1}^{K}\left(x_{2j-1} + x_{2j}\right)$, and by defining $\hat{\Upsilon} = \sum_{i=1}^{K}\left(p_{2i-1}\hat{\sigma}_{g,2i-1}^4\hat{\sigma}_{f,2i}^2 + p_{2i}\hat{\sigma}_{g,2i}^4\hat{\sigma}_{f,2i-1}^2\right)$. Equality in (e) is obtained by defining $\hat{\Phi} = \sum_{i=1}^{K}\hat{\Phi}_i = \sum_{i=1}^{2K}\hat{\sigma}_{g,i}^2\hat{\sigma}_{f,i'}^2$ and $\hat{\Psi} = \sum_{i=1}^{K}\hat{\Psi}_i = \sum_{i=1}^{K}\left(p_{2i-1}\hat{\sigma}_{g,2i-1}^2 + p_{2i}\hat{\sigma}_{g,2i}^2\right) = \sum_{i=1}^{2K} p_i\hat{\sigma}_{g,i}^2$. By using an approach similar to (41), we simplify (42) as

$$
\begin{aligned}
&\mathrm{Tr}\left\{\mathbb{E}\left[\mathbf{E}_g\mathbf{P}\mathbf{P}^H\mathbf{E}_g^{\ H}\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{F}}^T\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H\right]\right\} = \mathrm{Tr}\left\{\mathbb{E}\left[\mathbf{E}_g\mathbf{P}\mathbf{P}^H\mathbf{E}_g^H\right]\mathbb{E}\left[\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{F}}^T\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H\right]\right\} \\
&= \mathrm{Tr}\left\{\sum_{k=1}^{2K} p_k\hat{\sigma}_{\xi g,k}^2 \mathbb{E}\left[\sum_{i=1}^{K}\left(\hat{\mathbf{g}}_{2i-1}\hat{\mathbf{f}}_{2i}^T + \hat{\mathbf{g}}_{2i}\hat{\mathbf{f}}_{2i-1}^T\right)\left(\hat{\mathbf{f}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H + \hat{\mathbf{f}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\right]\right\} \\
&= \sum_{k=1}^{2K} p_k\hat{\sigma}_{\xi g,k}^2 \sum_{i=1}^{K} N^2\left(\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2 + \hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2\right) = N^2\hat{\Phi}\sum_{k=1}^{2K} p_k\hat{\sigma}_{\xi g,k}^2. \quad (47)
\end{aligned}
$$

We now simplify the second term in the denominator of (10).

$$
\begin{aligned}
\mathbb{E}\left[\|\mathbf{W}\mathbf{G}_{RR}\hat{\mathbf{x}}_R\|^2\right] &= \frac{P_R}{N}\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H\mathbf{G}_{RR}\mathbf{G}_{RR}^H\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{F}}^T\right]\right\} \\
&= P_R\sigma_{LIR}^2 \sum_{i=1}^{K} N^2\left(\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2 + \hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2\right) = N^2 P_R\sigma_{LIR}^2\hat{\Phi}. \quad (48)
\end{aligned}
$$

The last term $\mathbb{E}\left[\|\mathbf{W}\mathbf{z}_R\|^2\right]$ in the denominator of (10) can be simplified as

$$\mathbb{E}\left[\|\mathbf{W}\mathbf{z}_R\|^2\right] = \mathrm{Tr}\left\{\sigma_{nr}^2 \mathbb{E}\left[\hat{\mathbf{F}}^*\mathbf{T}\hat{\mathbf{G}}^H\hat{\mathbf{G}}\mathbf{T}^H\hat{\mathbf{F}}^T\right]\right\}$$

$$= \sigma_{nr}^2 \sum_{i=1}^{K} N^2\left(\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2 + \hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2\right) = N^2\sigma_{nr}^2\hat{\Phi}. \tag{49}$$

By substituting the expression of $\mathbb{E}\left[\|\mathbf{W}\tilde{\mathbf{G}}\mathbf{x}\|^2\right]$ from (46) as well as (47), and the expressions of $\mathbb{E}\left[\|\mathbf{W}\mathbf{G}_{RR}\hat{\mathbf{x}}_R\|^2\right]$ and $\mathbb{E}\left[\|\mathbf{W}\mathbf{z}_R\|^2\right]$ from (48) and (49) respectively in (10), we get (15).

APPENDIX B

To derive this result, we will first simplify $\mathbb{E}\left[\|\mathbf{W}\tilde{\mathbf{G}}\mathbf{x}\|^2\right]$ in the denominator of (10) using (16).

$$\mathbb{E}\left[\|\mathbf{W}\tilde{\mathbf{G}}\mathbf{x}\|^2\right] = \mathbb{E}\left[\|\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\tilde{\mathbf{G}}\mathbf{x}\|^2\right] = \mathrm{Tr}\left\{\mathbb{E}\left[(\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\mathbf{G}\mathbf{P}\mathbf{P}^H\mathbf{G}^H\hat{\bar{\mathbf{G}}}\mathbf{T}^H\hat{\bar{\mathbf{F}}}^T)\right]\right\}$$

$$= \mathrm{Tr}\left\{\mathbb{E}\left[(\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H(\hat{\mathbf{G}} + \mathbf{E}_g)\mathbf{P}\mathbf{P}^H(\hat{\mathbf{G}}^H + \mathbf{E}_g^H)\hat{\bar{\mathbf{G}}}\mathbf{T}^H\hat{\bar{\mathbf{F}}}^T)\right]\right\}$$

$$= \mathrm{Tr}\left\{\mathbb{E}\left[\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\hat{\mathbf{G}}\mathbf{P}\mathbf{P}^H\hat{\mathbf{G}}^H\hat{\bar{\mathbf{G}}}\mathbf{T}^H\hat{\bar{\mathbf{F}}}^T\right]\right\} + \mathrm{Tr}\left\{\mathbb{E}\left[\mathbf{E}_g\mathbf{P}\mathbf{P}^H\mathbf{E}_g{}^H\hat{\bar{\mathbf{G}}}\mathbf{T}^H\hat{\bar{\mathbf{F}}}^T\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\right]\right\} \tag{50}$$

$$\overset{(a)}{=} \mathrm{Tr}\left\{\mathbb{E}\left[\hat{\bar{\mathbf{F}}}^*\mathbf{T}\mathbf{P}\mathbf{P}^H\mathbf{T}\hat{\bar{\mathbf{F}}}^T\right]\right\} + \sum_{i=1}^{2K} p_i\sigma_{\xi,g,i}^2\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\bar{\mathbf{G}}}\mathbf{T}^H\hat{\bar{\mathbf{F}}}^T\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\right]\right\}$$

$$\overset{(b)}{=} \sum_{i=1}^{2K} p_{i'}\mathbb{E}\left[\hat{\bar{\mathbf{f}}}_i^H\hat{\bar{\mathbf{f}}}_i\right] + \sum_{i=1}^{2K} p_i\sigma_{\xi,g,i}^2\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{\Lambda}}_F^*\mathbf{T}\hat{\mathbf{\Lambda}}_G\mathbf{T}\right]\right\} \overset{(c)}{=} \hat{\lambda} + \sum_{i=1}^{2K} p_i\sigma_{\xi,g,i}^2\hat{\eta}. \tag{51}$$

The equality in (a) is obtained by exploiting the fact that $\hat{\bar{\mathbf{G}}}^H\hat{\mathbf{G}} = \hat{\mathbf{G}}^H\hat{\bar{\mathbf{G}}} = \mathbf{I}_{2K}$, and $\mathbb{E}\left[\mathbf{E}_g\mathbf{P}\mathbf{P}^H\mathbf{E}_g{}^H\right] = \sum_{i=1}^{2K} p_i\sigma_{\xi,g,i}^2\mathbf{I}_N$. In the equality (b), we define $\hat{\mathbf{\Lambda}}_F \triangleq (\hat{\bar{\mathbf{F}}}^H\hat{\bar{\mathbf{F}}}) = (\bar{\mathbf{F}}^H\bar{\mathbf{F}})^{-1}$ as well as $\hat{\mathbf{\Lambda}}_G \triangleq (\hat{\bar{\mathbf{G}}}^H\hat{\bar{\mathbf{G}}}) = (\bar{\mathbf{G}}^H\bar{\mathbf{G}})^{-1}$ and exploits the fact that $\mathbf{T}\mathbf{P}\mathbf{P}^H\mathbf{T} = \mathrm{diag}\{p_2, p_1, \cdots, p_{2K}, p_{2K-1}\}$. To derive the equality in (c), we first note that the random matrices $\hat{\mathbf{\Lambda}}_F$ and $\hat{\mathbf{\Lambda}}_G$ have inverse Wishart distribution, i.e. we have $\hat{\mathbf{\Lambda}}_F \sim \mathcal{W}^{-1}(\hat{\mathbf{D}}_d^{-1}, 2K)$, $\hat{\mathbf{\Lambda}}_G \sim \mathcal{W}^{-1}(\hat{\mathbf{D}}_u^{-1}, 2K)$, where $\hat{\mathbf{D}}_d$ and $\hat{\mathbf{D}}_u$ are covariance matrices of the rows of the estimated channels $\hat{\mathbf{G}}$ and $\hat{\mathbf{F}}$, which are given after (13). We denote $\hat{w}_{f,i,j} = \left(\hat{\mathbf{\Lambda}}_F\right)_{i,j}$, $\hat{w}_{g,i,j} = \left(\hat{\mathbf{\Lambda}}_G\right)_{i,j}$, $\forall i, j = 1, 2, 3, ..., 2K$ and $\mathbb{E}\left[\hat{\mathbf{\Lambda}}_F\right] = \frac{\hat{\mathbf{D}}_d^{-1}}{N-2K-1}$, $\mathbb{E}\left[\hat{\mathbf{\Lambda}}_G\right] = \frac{\hat{\mathbf{D}}_u^{-1}}{N-2K-1}$ [47]. We also have $\mathbb{E}\left[\hat{\bar{\mathbf{f}}}_i^H\hat{\bar{\mathbf{f}}}_j\right] = \mathbb{E}\left[\hat{w}_{f,i,j}\right] = \frac{1}{(N-2K-1)\hat{\sigma}_{f,i}^2}, \forall i = j$, and 0 otherwise. Similarly, $\mathbb{E}\left[\hat{\bar{\mathbf{g}}}_i^H\hat{\bar{\mathbf{g}}}_j\right] = \mathbb{E}\left[\hat{w}_{g,i,j}\right] = \frac{1}{(N-2K-1)\hat{\sigma}_{g,i}^2}, \forall i = j$, and 0 otherwise. Given the above equalities, we obtain the equality in (c), where $\hat{\lambda} = \sum_{i=1}^{2K} \frac{p_{i'}}{(N-2K-1)\hat{\sigma}_{f,i}^2}$. The expression $\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{\Lambda}}_F^*\mathbf{T}\hat{\mathbf{\Lambda}}_G\mathbf{T}\right]\right\}$ is simplified as

$$\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{\Lambda}}_F^*\mathbf{T}\hat{\mathbf{\Lambda}}_G\mathbf{T}\right]\right\} = \sum_{j=1}^{2K}\left(\mathbb{E}\left[w_{f,j,j'}^*\right]\mathbb{E}\left[w_{g,j,j'}\right] + \mathbb{E}\left[w_{f,j,j}^*\right]\mathbb{E}\left[w_{g,j',j'}\right]\right)$$

$$= \sum_{j=1}^{2K}\frac{1}{(N-2K-1)^2\,\hat{\sigma}_{f,j}^2\hat{\sigma}_{g,j'}^2} \triangleq \hat{\eta}. \tag{52}$$

Following similar lines, the second term in the denominator of (10) can be simplified as

$$\mathbb{E}\left[\|\mathbf{W}\mathbf{G}_{RR}\tilde{\mathbf{x}}_R\|^2\right] = \mathbb{E}\left[\|\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\mathbf{G}_{RR}\tilde{\mathbf{x}}_R\|^2\right] = \frac{P_R}{N}\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\mathbf{G}_{RR}\mathbf{G}_{RR}^H\hat{\bar{\mathbf{G}}}\mathbf{T}^H\hat{\bar{\mathbf{F}}}^T\right]\right\}$$

$$= P_R\sigma_{LIR}^2\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\bar{\mathbf{F}}}^T\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\hat{\bar{\mathbf{G}}}\mathbf{T}^H\right]\right\} = P_R\sigma_{LIR}^2\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{\Lambda}}_F^*\mathbf{T}\hat{\mathbf{\Lambda}}_G\mathbf{T}\right]\right\} = P_R\sigma_{LIR}^2\hat{\eta}. \qquad (53)$$

The last term in the denominator of (10) can be simplified to

$$\mathbb{E}\left[\|\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\mathbf{z}_R\|^2\right] = \sigma_{nr}^2\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\hat{\bar{\mathbf{G}}}\mathbf{T}^H\hat{\bar{\mathbf{F}}}^T\right]\right\} = \sigma_{nr}^2\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\mathbf{\Lambda}}_F^*\mathbf{T}\hat{\mathbf{\Lambda}}_G\mathbf{T}\right]\right\} = \sigma_{nr}^2\hat{\eta}. \qquad (54)$$

By using (51), (53) and (54), we get the simplified expression of α in (17).

APPENDIX C

We will start by simplifying the numerator of (23), with $\mathbf{f}_k = \hat{\mathbf{f}}_k + \mathbf{e}_{f,k}$, $\mathbf{g}_k = \hat{\mathbf{g}}_k + \mathbf{e}_{g,k}$, and using the fact that $\hat{\mathbf{f}}_k$, $\hat{\mathbf{g}}_k$, $\mathbf{e}_{f,k}$, $\mathbf{e}_{g,k}$ are independent, as follows

$$\mathbb{E}\left[|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}|\right] = \mathbb{E}\left[\left(\hat{\mathbf{f}}_k + \mathbf{e}_{f,k}\right)^T\mathbf{W}\left(\hat{\mathbf{g}}_{k'} + \mathbf{e}_{g,k'}\right)\right] = \mathbb{E}\left[\hat{\mathbf{f}}_k^T\hat{\mathbf{F}}_k^*\mathbf{T}\hat{\mathbf{G}}^H\hat{\mathbf{g}}_{k'}\right]$$

$$= \mathbb{E}\left[\hat{\mathbf{f}}_k^T\sum_{i=1}^K\left(\hat{\mathbf{f}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H + \hat{\mathbf{f}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\hat{\mathbf{g}}_{k'}\right] = \mathbb{E}\left[\hat{\mathbf{f}}_k^T\hat{\mathbf{f}}_k^*\hat{\mathbf{g}}_{k'}^H\hat{\mathbf{g}}_{k'} + \hat{\mathbf{f}}_k^T\hat{\mathbf{f}}_{k'}^*\hat{\mathbf{g}}_k^H\hat{\mathbf{g}}_{k'}\right]$$

$$= \mathbb{E}\left[\hat{\mathbf{f}}_k^H\hat{\mathbf{f}}_k\right]\mathbb{E}\left[\hat{\mathbf{g}}_{k'}^H\hat{\mathbf{g}}_{k'}\right] + \mathbb{E}\left[\hat{\mathbf{f}}_{k'}^H\hat{\mathbf{f}}_k\right]\mathbb{E}\left[\hat{\mathbf{g}}_k^H\hat{\mathbf{g}}_{k'}\right] \overset{(a)}{=} N^2\hat{\sigma}_{f,k}^2\hat{\sigma}_{g,k'}^2. \qquad (55)$$

The equality in (a) is because $\mathbb{E}\left[\hat{\mathbf{g}}_i^H\hat{\mathbf{g}}_i\right] = N\hat{\sigma}_{g,i}^2$, $\mathbb{E}\left[\hat{\mathbf{g}}_i^H\hat{\mathbf{g}}_j\right] = 0$ [46].

We now simplify the first term in the denominator of (23), the variance of $\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'} = \mathrm{var}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right] = \mathbb{E}\left[|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}|^2\right] - \left|\mathbb{E}\left[|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}|\right]\right|^2$, where we have:

$$\mathbb{E}\left[|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}|^2\right] = \mathbb{E}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\mathbf{g}_{k'}^H\mathbf{W}^H\mathbf{f}_k^*\right]$$

$$= \underbrace{\mathbb{E}\left[\hat{\mathbf{f}}_k^T\mathbf{W}\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^H\mathbf{W}^H\hat{\mathbf{f}}_k^*\right]}_{(I)} + \underbrace{\mathbb{E}\left[\hat{\mathbf{f}}_k^T\mathbf{W}\mathbf{e}_{g,k'}\mathbf{e}_{g,k'}^H\mathbf{W}^H\hat{\mathbf{f}}_k^*\right]}_{(II)}$$

$$+ \underbrace{\mathbb{E}\left[\mathbf{e}_{f,k}^T\mathbf{W}\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^H\mathbf{W}^H\mathbf{e}_{f,k}^*\right]}_{(III)} + \underbrace{\mathbb{E}\left[\mathbf{e}_{f,k}^T\mathbf{W}\mathbf{e}_{g,k'}\mathbf{e}_{g,k'}^H\mathbf{W}^H\mathbf{e}_{f,k}^*\right]}_{(IV)}. \qquad (56)$$

The term-(I) in (56) is simplified as following.

$$\mathbb{E}\left[\hat{\mathbf{f}}_k^T\mathbf{W}\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^H\mathbf{W}^H\hat{\mathbf{f}}_k^*\right] \overset{(a)}{=} \mathbb{E}\left[\hat{\mathbf{f}}_k^T\sum_{i=1}^K\left(\hat{\mathbf{f}}_{2i-1}^*\hat{\mathbf{g}}_{2i}^H + \hat{\mathbf{f}}_{2i}^*\hat{\mathbf{g}}_{2i-1}^H\right)\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^H\left(\hat{\mathbf{g}}_{2i}\hat{\mathbf{f}}_{2i-1}^T + \hat{\mathbf{g}}_{2i-1}\hat{\mathbf{f}}_{2i}^T\right)\hat{\mathbf{f}}_k^*\right]$$

$$\overset{(b)}{=} N^2\sum_{i=1,i\neq\lceil\frac{k}{2}\rceil}^K\hat{\sigma}_{f,k}^2\hat{\sigma}_{g,k'}^2\left(\hat{\sigma}_{f,2i-1}^2\hat{\sigma}_{g,2i}^2 + \hat{\sigma}_{f,2i}^2\hat{\sigma}_{g,2i-1}^2\right) + N^2\hat{\sigma}_{f,k}^2\hat{\sigma}_{f,k'}^2\hat{\sigma}_{g,k}^2\hat{\sigma}_{g,k'}^2 + N^2(N+1)^2\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,k'}^4$$

$$\overset{(c)}{=} N^2\hat{\sigma}_{f,k}^2\hat{\sigma}_{g,k'}^2\sum_{i=1}^K\left(\hat{\sigma}_{f,2i-1}^2\hat{\sigma}_{g,2i}^2 + \hat{\sigma}_{f,2i}^2\hat{\sigma}_{g,2i-1}^2\right) + N^3(N+2)\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,k'}^4$$

$$\overset{(d)}{=} N^2\hat{\sigma}_{f,k}^2\hat{\sigma}_{g,k'}^2\left(N(N+2)\hat{\sigma}_{f,k}^2\hat{\sigma}_{g,k'}^2 + \hat{\Phi}\right). \qquad (57)$$

The equality in (b) is obtained by expanding (a) for two different cases: i) $\{k, k'\} = \{2i, 2i-1\}$;

ii) $\{k, k'\} \neq \{2i, 2i-1\}$, and by using the fact that $\mathbb{E}\left[|\hat{\mathbf{g}}_i^H \hat{\mathbf{g}}_j|^2\right] = N\hat{\sigma}_{g,i}^2 \hat{\sigma}_{g,j}^2$ and $\mathbb{E}\left[\hat{\mathbf{g}}_i\hat{\mathbf{g}}_i^H \hat{\mathbf{g}}_i\hat{\mathbf{g}}_i^H\right] = (N+1)\hat{\sigma}_{g,i}^4 \mathbf{I}_N$ [46]. The equality in (c) is obtained by simple arithmetic manipulations and by re-arranging terms. The equality in (d) is obtained by defining $\hat{\Phi} = \sum_{i=1}^{K}\left(\hat{\sigma}_{g,2i-1}^2\hat{\sigma}_{f,2i}^2 + \hat{\sigma}_{g,2i}^2\hat{\sigma}_{f,2i-1}^2\right)$. The term-(II) in (56) is simplified using similar ideas as follows:

$$\mathbb{E}\left[\hat{\mathbf{f}}_k^T \mathbf{W} \mathbf{e}_{g,k'}\mathbf{e}_{g,k'}^H \mathbf{W}^H \hat{\mathbf{f}}_k^*\right]$$
$$= N^2\left(\hat{\sigma}_{f,k}^2 \hat{\sigma}_{f,k'}^2 \hat{\sigma}_{g,k}^2 \hat{\sigma}_{\xi g,k'}^2 + (N+1)\hat{\sigma}_{f,k}^4 \hat{\sigma}_{g,k'}^2 \hat{\sigma}_{\xi g,k'}^2 + \sum_{i\neq\lceil\frac{k}{2}\rceil}^{K}\hat{\sigma}_{\xi g,k'}^2 \hat{\sigma}_{f,k}^2 \left(\hat{\sigma}_{f,2i-1}^2 \hat{\sigma}_{g,2i}^2 + \hat{\sigma}_{f,2i}^2 \hat{\sigma}_{g,2i-1}^2\right)\right)$$
$$= N^2\hat{\sigma}_{f,k}^2 \hat{\sigma}_{\xi g,k'}^2 \sum_{i=1}^{K}\left(\hat{\sigma}_{f,2i-1}^2 \hat{\sigma}_{g,2i}^2 + \hat{\sigma}_{f,2i}^2 \hat{\sigma}_{g,2i-1}^2\right) + N^3\hat{\sigma}_{f,k}^4 \hat{\sigma}_{\xi g,k'}^2 \hat{\sigma}_{g,k'}^2$$
$$= N^2\hat{\sigma}_{f,k}^2 \hat{\sigma}_{\xi g,k'}^2 \left(N\hat{\sigma}_{f,k}^2 \hat{\sigma}_{g,k'}^2 + \hat{\Phi}\right). \tag{58}$$

The term-(III) in (56) is simplified as following.

$$\mathbb{E}\left[\mathbf{e}_{f,k}^T \mathbf{W} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H \mathbf{W}^H \mathbf{e}_{f,k}^*\right]$$
$$= N^2\left(\hat{\sigma}_{f,k'}^2 \hat{\sigma}_{\xi f,k}^2 \hat{\sigma}_{g,k}^2 \hat{\sigma}_{g,k'}^2 + (N+1)\hat{\sigma}_{g,k'}^4 \hat{\sigma}_{f,k}^2 \hat{\sigma}_{\xi f,k}^2 + \sum_{i\neq\lceil\frac{k}{2}\rceil}^{K}\hat{\sigma}_{\xi f,k}^2 \hat{\sigma}_{g,k'}^2 \left(\hat{\sigma}_{f,2i-1}^2 \hat{\sigma}_{g,2i}^2 + \hat{\sigma}_{f,2i}^2 \hat{\sigma}_{g,2i-1}^2\right)\right)$$
$$= N^2\hat{\sigma}_{\xi f,k}^2 \hat{\sigma}_{g,k'}^2 \sum_{i=1}^{K}\left(\hat{\sigma}_{f,2i-1}^2 \hat{\sigma}_{g,2i}^2 + \hat{\sigma}_{f,2i}^2 \hat{\sigma}_{g,2i-1}^2\right) + N^3\hat{\sigma}_{\xi f,k}^2 \hat{\sigma}_{g,k'}^4 \hat{\sigma}_{f,k}^2$$
$$= N^2\hat{\sigma}_{\xi f,k}^2 \hat{\sigma}_{g,k'}^2 \left(N\hat{\sigma}_{g,k'}^2 \hat{\sigma}_{f,k}^2 + \hat{\Phi}\right). \tag{59}$$

Finally the term-(IV) in (56) is simplified as follows:

$$\mathbb{E}\left[\mathbf{e}_{f,k}^T \mathbf{W} \mathbf{e}_{g,k'}\mathbf{e}_{g,k'}^H \mathbf{W}^H \mathbf{e}_{f,k}^*\right] = N^2\sum_{i=1}^{K}\hat{\sigma}_{\xi f,k}^2 \hat{\sigma}_{\xi g,k'}^2 \left(\hat{\sigma}_{f,2i-1}^2 \hat{\sigma}_{g,2i}^2 + \hat{\sigma}_{f,2i}^2 \hat{\sigma}_{g,2i-1}^2\right) = N^2\hat{\sigma}_{\xi f,k}^2 \hat{\sigma}_{\xi g,k'}^2 \hat{\Phi}. \tag{60}$$

Further using (57)-(60), (56) can be re-written as

$$\mathbb{E}\left[|\mathbf{f}_k^T \mathbf{W} \mathbf{g}_{k'}|^2\right] = N^2\hat{\Phi}\left(\hat{\sigma}_{f,k}^2 \hat{\sigma}_{g,k'}^2 + \hat{\sigma}_{f,k}^2 \hat{\sigma}_{\xi g,k'}^2 + \hat{\sigma}_{\xi f,k}^2 \hat{\sigma}_{g,k'}^2 + \hat{\sigma}_{\xi f,k}^2 \hat{\sigma}_{\xi g,k'}^2\right)$$
$$+ N^3\hat{\sigma}_{f,k}^2 \hat{\sigma}_{g,k'}^2 \left((N+2)\hat{\sigma}_{f,k}^2 \hat{\sigma}_{g,k'}^2 + \hat{\sigma}_{f,k}^2 \hat{\sigma}_{\xi g,k'}^2 + \hat{\sigma}_{\xi f,k}^2 \hat{\sigma}_{g,k'}^2\right). \tag{61}$$

By using (55), (61), and considering the fact that $\sigma_{g,k}^2 = \hat{\sigma}_{\xi g,k}^2 + \hat{\sigma}_{g,k}^2$, $\sigma_{f,k}^2 = \hat{\sigma}_{\xi f,k}^2 + \hat{\sigma}_{f,k}^2$, we have

$$\text{var}\left[\mathbf{f}_k^T \mathbf{W} \mathbf{g}_{k'}\right] = N^2\left(\hat{\Phi}\sigma_{f,k}^2 \sigma_{g,k'}^2 + N\hat{\sigma}_{f,k}^2 \hat{\sigma}_{g,k'}^2 \left(\sigma_{f,k}^2 \hat{\sigma}_{g,k'}^2 + \hat{\sigma}_{f,k}^2 \sigma_{g,k'}^2\right)\right). \tag{62}$$

Similar to the above calculations, we can simplify other terms in the denominator of (23) as

$$\text{SI}_k = N^2\sigma_{\xi f,k}^2 \sigma_{\xi g,k}^2 \hat{\Phi} + N^2\hat{\sigma}_{f,k}^2 \sigma_{\xi g,k}^2 \left(N\hat{\sigma}_{f,k}^2 \hat{\sigma}_{g,k'}^2 + \hat{\Phi}\right) + N^2\sigma_{\xi f,k}^2 \hat{\sigma}_{g,k}^2 \left(N\hat{\sigma}_{g,k}^2 \hat{\sigma}_{f,k'}^2 + \hat{\Phi}\right), \tag{63}$$

$$\text{IP}_k = N^2\sum_{j\neq k,k'}^{2K} p_j\left[\sigma_{f,k}^2 \sigma_{g,j}^2 \hat{\Phi} + N\left(\sigma_{f,k}^2 \hat{\sigma}_{g,j}^4 \hat{\sigma}_{f,j'}^2 + \sigma_{g,j}^2 \hat{\sigma}_{f,k}^4 \hat{\sigma}_{g,k'}^2\right)\right], \tag{64}$$

$$\text{NR}_k = N^2\sigma_{nr}^2\left(N\hat{\sigma}_{f,k}^4 \hat{\sigma}_{g,k'}^2 + \sigma_{f,k}^2 \hat{\Phi}\right), \quad \text{LIR}_k = N^2 P_R, \sigma_{LIR}^2\left(N\hat{\sigma}_{f,k}^4 \hat{\sigma}_{g,k'}^2 + \sigma_{f,k}^2 \hat{\Phi}\right), \tag{65}$$

$$\text{UI}_k = \sum_{i,k\in Uk} p_i\mathbb{E}\left[|\Omega_{k,i}x_i|^2\right] = \sum_{i,k\in Uk} p_i\sigma_{k,i}^2, \quad \text{NU}_k = \mathbb{E}\left[|z_k|^2\right] = \sigma_n^2. \tag{66}$$

We can further simplify the denominator of (23) using (62)-(66), as

$$
\begin{aligned}
&p_{k'}\mathrm{var}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right] + p_k\mathrm{SI}_k + \mathrm{IP}_k + \mathrm{NR}_k + \mathrm{LIR}_k + \frac{1}{\alpha^2}\mathrm{UI}_k + \frac{1}{\alpha^2}\mathrm{NU}_k \\
&= N^2\left\{\sum_{i=1}^{2K} p_i\left[\eta_{k,i} + \left(\hat{\Phi}\sigma_{g,i}^2 + N\hat{\sigma}_{g,i}^4\hat{\sigma}_{f,i'}^2\right)P_R^{-1}\left(\sigma_n^2 + \sum_{i,k\in U_k} p_i\sigma_{k,i}^2\right)\right] + c_k p_k\right. \\
&\quad + P_R\sigma_{LIR}^2\left(\sigma_{f,k}^2\hat{\Phi} + N\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,k'}^2\right) + P_R^{-1}\sigma_{nr}^2\hat{\Phi}\left(\sigma_n^2 + \sum_{i,k\in U_k} p_i\sigma_{k,i}^2\right) + \sum_{i,k\in U_k} p_i\sigma_{k,i}^2\sigma_{LIR}^2\hat{\Phi} \\
&\quad \left. + \left(\sigma_{LIR}^2\sigma_n^2 + \sigma_{nr}^2\sigma_{f,k}^2\right)\hat{\Phi} + N\sigma_{nr}^2\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,k'}^2\right\},
\end{aligned}
\tag{67}
$$

where $\eta_{k,i} = \hat{\Phi}\sigma_{f,k}^2\sigma_{g,i}^2 + N\left(\sigma_{f,k}^2\hat{\sigma}_{g,i}^4\hat{\sigma}_{f,i'}^2 + \sigma_{g,i}^2\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,k'}^2\right)$ and,

$c_k = -\left(\hat{\Phi}\hat{\sigma}_{f,k}^2\hat{\sigma}_{g,j'}^2 + N\left(\hat{\sigma}_{f,k}^4\hat{\sigma}_{g,j}^2\hat{\sigma}_{g,k'}^2 + \hat{\sigma}_{f,k}^2\hat{\sigma}_{g,j}^4\hat{\sigma}_{f,k'}^2\right)\right)$. The proof of Theorem (1) is complete.

APPENDIX D

Starting with the numerator of (23) considering ZFR/ZFT processing, we have

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right] &= \mathbb{E}\left[\left(\hat{\mathbf{f}}_k + \mathbf{e}_{f,k}\right)^T\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\left(\hat{\mathbf{g}}_{k'} + \mathbf{e}_{g,k'}\right)\right] = \mathbb{E}\left[\hat{\mathbf{f}}_k^T\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\hat{\mathbf{g}}_{k'}\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\mathbf{1}_{f,k}^T\mathbf{T}\mathbf{1}_{g,k'}\right] \overset{(b)}{=} \mathbb{E}[1] = 1.
\end{aligned}
\tag{68}
$$

The equality in (a) is obtained by using the following results: $\hat{\mathbf{g}}_{k'}^H\hat{\bar{\mathbf{G}}} = \mathbf{1}_{g,k'}^T$, $\hat{\bar{\mathbf{G}}}^H\hat{\mathbf{g}}_{k'} = \mathbf{1}_{g,k'}$, $\hat{\mathbf{f}}_k^T\hat{\bar{\mathbf{F}}}^* = \mathbf{1}_{f,k}^T$, $\hat{\bar{\mathbf{F}}}^T\hat{\mathbf{f}}_k^* = \mathbf{1}_{f,k}$. Equality in (b) is attained because $\mathbf{1}_{f,k}^T\mathbf{T}\mathbf{1}_{g,k'} = 1$.

The expression of $\mathrm{var}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right]$ in the denominator of (23) is given by

$$
\begin{aligned}
\mathrm{var}&\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right] = \mathbb{E}\left[\left|\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right|^2\right] - \left|\mathbb{E}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right]\right|^2 \\
&\overset{(a)}{=} \mathbb{E}\left[\hat{\mathbf{f}}_k^T\mathbf{W}\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^H\mathbf{W}^H\hat{\mathbf{f}}_k^*\right] - 1 + \mathbb{E}\left[\mathbf{e}_{f,k}^T\mathbf{W}\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^H\mathbf{W}^H\mathbf{e}_{f,k}^*\right] + \mathbb{E}\left[\hat{\mathbf{f}}_k^T\mathbf{W}\mathbf{e}_{g,k'}\mathbf{e}_{g,k'}^H\mathbf{W}^H\hat{\mathbf{f}}_k^*\right] \\
&\quad + \mathbb{E}\left[\mathbf{e}_{f,k}^T\mathbf{W}\mathbf{e}_{g,k'}\mathbf{e}_{g,k'}^H\mathbf{W}^H\mathbf{e}_{f,k}^*\right] \\
&\overset{(b)}{=} \sigma_{\xi,f,k}^2\mathbb{E}\left[\hat{\mathbf{g}}_{k'}^H\mathbf{W}^H\mathbf{W}\hat{\mathbf{g}}_{k'}\right] + \sigma_{\xi,g,k'}^2\mathbb{E}\left[\hat{\mathbf{f}}_k^T\mathbf{W}\mathbf{W}^H\hat{\mathbf{f}}_k^*\right] + \sigma_{\xi,f,k}^2\sigma_{\xi,g,k'}^2\mathrm{Tr}\left\{\mathbb{E}\left[\mathbf{W}\mathbf{W}^H\right]\right\} \\
&\overset{(c)}{=} \sigma_{\xi,f,k}^2\mathbb{E}\left[\hat{\mathbf{g}}_{k'}^H\hat{\bar{\mathbf{G}}}\mathbf{T}\hat{\bar{\mathbf{F}}}^T\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\hat{\mathbf{g}}_{k'}\right] + \sigma_{\xi,g,k'}^2\mathbb{E}\left[\hat{\mathbf{f}}_k^T\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\hat{\bar{\mathbf{G}}}\mathbf{T}\hat{\bar{\mathbf{F}}}^T\hat{\mathbf{f}}_k^*\right] \\
&\quad + \sigma_{\xi,f,k}^2\sigma_{\xi,g,k'}^2\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\bar{\mathbf{F}}}^*\mathbf{T}\hat{\bar{\mathbf{G}}}^H\hat{\bar{\mathbf{G}}}\mathbf{T}\hat{\bar{\mathbf{F}}}^T\right]\right\} \\
&\overset{(d)}{=} \sigma_{\xi,f,k}^2\mathbb{E}\left[\mathbf{1}_{g,k'}^T\mathbf{T}\hat{\boldsymbol{\Lambda}}_F^*\mathbf{T}\mathbf{1}_{g,k'}\right] + \sigma_{\xi,g,k'}^2\mathbb{E}\left[\mathbf{1}_{f,k}^T\mathbf{T}\hat{\boldsymbol{\Lambda}}_G\mathbf{T}\mathbf{1}_{f,k}\right] + \sigma_{\xi,f,k}^2\sigma_{\xi,g,k'}^2\mathrm{Tr}\left\{\mathbb{E}\left[\hat{\boldsymbol{\Lambda}}_F^*\mathbf{T}\hat{\boldsymbol{\Lambda}}_G\mathbf{T}\right]\right\} \\
&\overset{(e)}{=} \sigma_{\xi,f,k}^2\mathbb{E}\left[\hat{w}_{f,k,k}^*\right] + \sigma_{\xi,g,k'}^2\left(\mathbb{E}\left[\hat{w}_{g,k',k'}\right] + \sigma_{\xi,f,k}^2\sum_{j=1}^{2K}\left(\mathbb{E}\left[w_{f,j,j'}^*\right]\mathbb{E}\left[w_{g,j,j'}\right] + \mathbb{E}\left[w_{f,j,j}^*\right]\mathbb{E}\left[w_{g,j',j'}\right]\right)\right) \\
&\overset{(f)}{=} \frac{\sigma_{\xi,f,k}^2}{(N-2K-1)\hat{\sigma}_{f,k}^2} + \frac{\sigma_{\xi,g,k'}^2}{(N-2K-1)\hat{\sigma}_{g,k'}^2} + \sigma_{\xi,f,k}^2\sigma_{\xi,g,k'}^2\hat{\eta}.
\end{aligned}
\tag{69}
$$

The equality in (b) therein is because $\hat{\mathbf{F}}^T\mathbf{W}\hat{\mathbf{G}} = \mathbf{T}$, i.e., $\hat{\mathbf{f}}_k^T\mathbf{W}\hat{\mathbf{g}}_j = 1, \forall j = k'$ and 0, otherwise. The equality in (c) is obtained by substituting the value of $\mathbf{W}$ from (16). Equalities in (d) and (e) are obtained by simple manipulations, and equality in (f) is because $\mathbb{E}\left[\hat{w}_{f,k,k}\right] = \frac{1}{(N-2K-1)\hat{\sigma}_{f,k}^2}$,

$\mathbb{E}\left[\hat{w}_{g,k',k}\right] = \frac{1}{(N-2K-1)\hat{\sigma}^2_{g,k'}}$, and $\hat{\eta} \triangleq \sum_{j=1}^{2K} \frac{1}{(N-2K-1)^2 \hat{\sigma}^2_{f,j} \hat{\sigma}^2_{g,j'}}$.

Remember that there is no need to perform SIC in the case ZFR/ZFT processing, and hence the self-interference term in the denominator of (23) can be expressed as

$$\mathrm{SI}_k = \frac{1}{(N-2K-1)}\left(\frac{\sigma^2_{\xi,f,k}}{\hat{\sigma}^2_{f,k'}} + \frac{\sigma^2_{\xi,g,k}}{\hat{\sigma}^2_{g,k'}}\right) + \sigma^2_{\xi,f,k}\sigma^2_{\xi,g,k}\hat{\eta}. \tag{70}$$

Similarly, the other terms in the denominator of (23) can be written as follows

$$\mathrm{IP}_k = \sum_{i\neq k,k'}^{2K} p_i\left[\frac{1}{(N-2K-1)}\left(\frac{\sigma^2_{\xi,f,k}}{\hat{\sigma}^2_{f,i'}} + \frac{\sigma^2_{\xi,g,i}}{\hat{\sigma}^2_{g,k'}}\right) + \sigma^2_{\xi,f,k}\sigma^2_{\xi,g,i}\hat{\eta}\right], \tag{71}$$

$$\mathrm{NR}_k = \sigma^2_{nr}\left(\frac{1}{(N-2K-1)\hat{\sigma}^2_{g,k'}} + \sigma^2_{\xi,f,k}\hat{\eta}\right),\ \mathrm{LIR}_k = P_R\sigma^2_{LIR}\left(\frac{1}{(N-2K-1)\hat{\sigma}^2_{g,k'}} + \sigma^2_{\xi,f,k}\hat{\eta}\right), \tag{72}$$

$$\mathrm{UI}_k = \sum_{i,k\in Uk} p_i\mathbb{E}\left[|\Omega_{k,i}x_i|^2\right] = \sum_{i,k\in Uk} p_i\sigma^2_{k,i}, \quad \mathrm{NU}_k = \mathbb{E}\left[|z_k|^2\right] = \sigma^2_n. \tag{73}$$

Substituting the values obtained from (69-73) in the denominator of (23), we obtain

$$p_{k'}\mathrm{var}\left[\mathbf{f}_k^T\mathbf{W}\mathbf{g}_{k'}\right] + p_k\mathrm{SI}_k + \mathrm{IP}_k + \mathrm{NR}_k + \mathrm{LIR}_k + \frac{1}{\alpha^2}\mathrm{UI}_k + \frac{1}{\alpha^2}\mathrm{NU}_k$$

$$= \sum_{i=1}^{2K}\theta_{k,i}p_i + \sigma^2_{nr}\left(\frac{1}{(N-2K-1)\hat{\sigma}^2_{g,k'}} + \sigma^2_{\xi,f,k}\hat{\eta}\right) + P_R\sigma^2_{LIR}\left(\frac{1}{(N-2K-1)\hat{\sigma}^2_{g,k'}} + \sigma^2_{\xi,f,k}\hat{\eta}\right)$$

$$+ \sum_{i=1}^{2K} p_i\left[P_R^{-1}\sigma^2_n\left(\frac{1}{(N-2K-1)\hat{\sigma}^2_{f,i'}} + \hat{\eta}\sigma^2_{\xi,g,i}\right) + \sum_{i,k\in U_k} p_i P_R^{-1}\sigma^2_{k,i}\left(\frac{1}{(N-2K-1)\hat{\sigma}^2_{f,i'}} + \hat{\eta}\sigma^2_{\xi,g,i}\right)\right]$$

$$+ \left(P_R^{-1}\hat{\eta}\sigma^2_{nr}\sigma^2_n + \sum_{i,k\in U_k} p_i P_R^{-1}\sigma^2_{k,i}\hat{\eta}\sigma^2_{nr} + \sum_{i,k\in U_k} p_i\sigma^2_{k,i}\hat{\eta}\sigma^2_{LIR} + \hat{\eta}\sigma^2_{LIR}\sigma^2_n\right), \tag{74}$$

where $\theta_{k,i} = \left(\frac{\sigma^2_{\xi,f,k}}{(N-2K-1)\hat{\sigma}^2_{f,i'}} + \frac{\sigma^2_{\xi,g,i}}{(N-2K-1)\hat{\sigma}^2_{g,k'}} + \sigma^2_{\xi,f,k}\sigma^2_{\xi,g,i}\hat{\eta}\right)$.

REFERENCES

[1] K. Lee and L. Hanzo, "Resource-efficient wireless relaying protocols," *IEEE Wireless Commun.*, vol. 17, no. 2, pp. 66–72, 2010.

[2] Z. Zhang, X. Chai, K. Long, A. V. 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, 2015.

[3] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE*, vol. 104, no. 7, pp. 1369–1409, 2016.

[4] 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.

[5] A. Sabharwal, P. Schniter, D. Guo, D. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.

[6] A. Nadh, J. Samuel, A. Sharma, S. Aniruddhan, and R. K. Ganti, "A taylor series approximation of self-interference channel in full-duplex radios," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4304–4316, 2017.

[7] Y. Y. Kang, B.-J. Kwak, and J. H. Cho, "An optimal full-duplex AF relay for joint analog and digital domain self-interference cancellation," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2758–2772, Aug. 2014.

[8] W. Zhang, U. Mitra, and M. Chiang, "Optimization of amplify-and-forward multicarrier two-hop transmission," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1434–1445, 2011.

[9] I. Hammerström and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2798–2802, 2007.

[10] G. Zheng, "Joint beamforming optimization and power control for full-duplex MIMO two-way relay channel," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 555–566, 2015.

[11] D. Choi and J. H. Lee, "Outage probability of two-way full-duplex relaying with imperfect channel state information," *IEEE Commun. Lett.*, vol. 18, no. 6, pp. 933–936, 2014.

[12] Z. Zhang, Z. Ma, Z. Ding, M. Xiao, and G. K. Karagiannidis, "Full-duplex two-way and one-way relaying: Average rate, outage probability, and tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3920–3933, 2016.

[13] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, 2014.

[14] Z. Zhang, Z. Chen, M. Shen, and B. Xia, "Spectral and energy efficiency of multipair two-way full-duplex relay systems with massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 848–863, 2016.

[15] Z. Zhang, Z. Chen, M. Shen, B. Xia, W. Xie, and Y. Zhao, "Performance analysis for training-based multipair two-way full-duplex relaying with massive antennas," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6130–6145, 2017.

[16] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[17] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[18] L. Lu, G. Y. Li, A. L. Swindlehurst, A. E. Ashikhmin, and R. Zhang, "An overview of massive MIMO: benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.

[19] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[20] Y. Dai and X. Dong, "Power allocation for multi-pair massive MIMO two-way AF relaying with linear processing," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 5932–5946, 2016.

[21] H. Cui, L. Song, and B. Jiao, "Multi-pair two-way amplify-and-forward relaying with very large number of relay antennas," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2636–2645, May 2014.

[22] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. V. der Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 66–72, 2010.

[23] Z. Hasan, H. Boostanimehr, and V. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, Fourth Quarter 2011.

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

[25] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, 2017.

[26] Y. Li, P. Fan, A. Leukhin, and L. Liu, "On the spectral and energy efficiency of full-duplex small-cell wireless systems with massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2339–2353, 2017.

[27] Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, and L. Hanzo, "A survey of multi-objective optimization in wireless sensor networks: Metrics, algorithms, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 550–586, 2017.

[28] D. Alanis, P. Botsinis, Z. Babar, S. X. Ng, and L. Hanzo, "Non-dominated quantum iterative routing optimization for wireless multihop networks," *IEEE Access*, vol. 3, pp. 1704–1728, 2015.

[29] F. Tan, T. Lv, and S. Yang, "Power allocation optimization for energy-efficient massive MIMO aided multi-pair decode-and-forward relay systems," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2368–2381, June 2017.

[30] J.-P. Crouzeix and J. A. Ferland, "Algorithms for generalized fractional programming," *Mathematical Programming*, vol. 52, no. 1, pp. 191–207, 1991.

[31] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, March 2006.

[32] S. M. Kay, "Fundamentals of statistical signal processing, volume I: estimation theory," 1993.

[33] J. Jose, A. E. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.

[34] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[35] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, 2013.

[36] A. Zappone, Z. Chong, E. A. Jorswieck, and S. Buzzi, "Energy-aware competitive power control in relay-assisted interference wireless networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1860–1871, 2013.

[37] A. Zappone, E. A. Jorswieck, and S. Buzzi, "Energy efficiency and interference neutralization in two-hop MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6481–6495, 2014.

[38] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optimization and engineering*, vol. 8, no. 1, p. 67, 2007.

[39] A. Zappone and E. A. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 3-4, pp. 185–396, 2015.

[40] R. Budhiraja, Karthik KS, and B. Ramamurthi, "Linear precoders for non-regenerative asymmetric two-way relaying in cellular systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5002–5014, Sep. 2014.

[41] R. Budhiraja and B. Ramamurthi, "Joint precoder and receiver design for AF non-simultaneous two-way MIMO relaying," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 2942–2955, Jun. 2015.

[42] ——, "Joint transceiver design for QoS-constrained MIMO two-way non-regenerative relaying using geometric programming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3453–3465, May 2016.

[43] P. C. Weeraddana, M. Codreanu, M. Latva-aho, and A. Ephremides, "Resource allocation for cross-layer utility maximization in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2790–2809, 2011.

[44] W. Dinkelbach, "On nonlinear fractional programming," *Management science*, vol. 13, no. 7, pp. 492–498, 1967.

[45] A. Zappone, E. Björnson, L. Sanguinetti, and E. A. Jorswieck, "Globally optimal energy-efficient power control and receiver design in wireless networks," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2844–2859, 2017.

[46] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends® in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[47] P. Graczyk, G. Letac, and H. Massam, "The complex wishart distribution and the symmetric group," *Annals of Statistics*, pp. 287–309, 2003.