Distortion-Loop-Aware Amplify-and-Forward Full-Duplex Relaying with Multiple Antennas

Omid Taghizadeh*, Tianyu Yang*, Ali Cagatay Cirik† and Rudolf Mathar*
* Institute for Theoretical Information Technology, RWTH Aachen University, Aachen, 52074, Germany
† Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T1Z4, Canada
Email: {taghizadeh, yang, mathar}@ti.rwth-aachen.de, cirik@ece.ubc.ca

***Abstract*—In this work we study the behavior of a full-duplex (FD) and amplify-and-forward (AF) relay with multiple antennas, where hardware impairments of the FD transceivers are taken into account. Due to the inter-dependency of the transmit relay power on each antenna and the residual self-interference intensity in an AF-FD relay, we observe a *distortion loop* which degrades the system performance. An optimization problem is formulated to maximize the end-to-end communication rate, under relay and source transmit power constraints. Due to the resulting problem complexity, we propose a gradient-projection-based algorithm to approach an optimum solution. An iterative relay transmit covariance shaping algorithm is also proposed, following the quadratic approximation of the relay function in each iteration. The aforementioned algorithm provides a convex optimization framework to approach an optimal solution with small number of optimization iterations. The performance of the proposed designs are then compared to a similar setup with a decode-and-forward (DF) process, where the distortion loop effect is eliminated due to decoding, via numerical simulations.**

I. Introduction

A full-duplex transceiver is known with the capability to transmit and receive at the same time and frequency, and hence has the potential to improve the spectral efficiency [1]. Nevertheless, such systems have been long considered to be practically infeasible due to the inherent self-interference. Recently, specialized cancellation techniques [2], [3] have provided an adequate level of isolation between Tx and Rx directions to facilitate a FD communication and motivated a wide range of related applications, see, e.g., [1]. Nevertheless, it is easy to observe that the obtained cancellation level may vary for various realistic conditions. This mainly includes i) aging and inaccuracy of the hardware components, e.g., quantization noise, power amplifier and oscillator phase noise in analog domain, as well as ii) inaccurate estimation of the interference paths due to the limited channel coherence time, noise and limited processing power in digital domain. As a result, it is essential to take into account the aforementioned inaccuracies to obtain a design which remains efficient under realistic situations.

In this work we are focusing on the application of FD capability on a relaying system, where the relay node has multiple antennas and suffers from the residual self-interference. A FD relay is capable of receiving the signal from a source while simultaneously communicating to a destination, and hence has the potential to achieve higher spectral efficiency with reduced latency. In the early work by Riihonen et al. [4], the relay operation with a generic processing protocol is modeled, and many insights are provided regarding the multiple-antenna strategies for reducing the self-interference power. The design methodologies and performance evaluation for FD relays with decode-and-forward (DF) operation have been then studied, e.g., in [5], and in [6] for a single antenna and AF relay operation, taking into account the effects of the hardware impairments.

While the aforementioned literature introduces the importance of an accurate transceiver modeling with respect to the effects of hardware impairments for an AF-FD relay, such works are not yet extended for the relays with multiple antennas. This stems from the fact that in an AF-FD relay, the inter-dependent behavior of the transmit power from the relay and the residual self-interference intensity, results in a distortion loop effect, see [6, Subsection II-C]. The aforementioned effect results in a rather complicated mathematical description when relay is equipped with multiple antennas. As a result, related studies resort to highly simplified models to reduce the consequent design complexity. In [7]–[10] a multiple-antenna AF-FD relay system is studied, assuming a perfect synchronization and accurate hardware operation, where behavior of the self-interference channel is perfectly known. As a result, it is assumed that the self-interference signal can be completely suppressed via estimating and subtracting the interference term in the receiver side [7], or can be nulled out via optimized transmit/receive strategies, e.g., spatial zero-forcing of the self-interference using transmit beamforming [8], [9]. In [10] a perfect cancellation is assumed on the condition that the self-interference power does not exceed a certain threshold.

In this work, we study the performance of a multiple-antenna AF-FD relay, where the impact of hardware inaccuracy at the receiver and transmit chains are jointly taken into account. In the first step, the relay operation is analyzed by taking into account the effects of transmit and receiver chain distortions, and a rate maximizing optimization problem is formulated. A gradient-projection-based solution is then proposed in order to act as a benchmark for the achievable performance. Furthermore, an iterative convex optimization framework is proposed, via quadratic approximation of the relay transmit covariance in each step. The numerical comparison of the proposed designs show the significance of a distortion-loop aware design as the intensity of the hardware impairments increase.

II. System Model

We investigate a scenario where a single-antenna half-duplex (HD) source communicates with a HD destination node with $M_{\rm d}$ antennas, with a help of a FD relay. The relay is assumed to have $M_{\rm t}$ transmit and $M_{\rm r}$ receive antennas, and

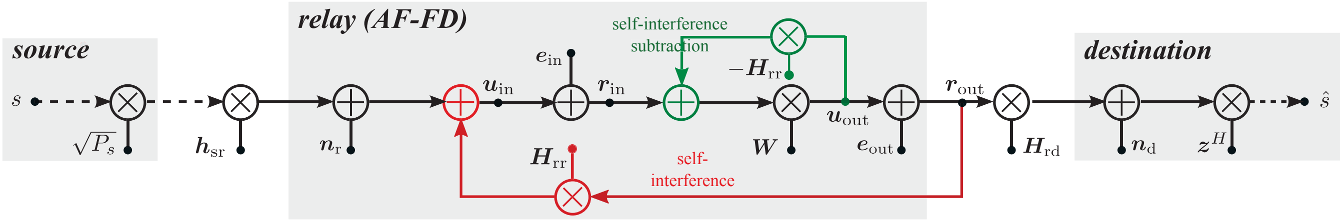

Fig. 1. The interaction of different signal components in an amplify-and-forward FD MIMO relay node. The loopback-self interference affects the residual interference intensity. The bold arrows represent the vectors while the dashed arrows represent the scalars.

operates in AF mode. The channels between the source and the relay, and between the relay and the destination are denoted as $\mathbf{h}_{\rm sr} \in \mathbb{C}^{M_{\rm r}}$ and $\mathbf{H}_{\rm rd} \in \mathbb{C}^{M_{\rm d}\times M_{\rm t}}$, respectively. The self-interference channel, which is the channel between the relay's transmit and receive ends is denoted as $\mathbf{H}_{\rm rr} \in \mathbb{C}^{M_{\rm r}\times M_{\rm t}}$. All channels are known[1] and following the quasi-static[2] flat-fading model. Furthermore, it is assumed that the path between the source and destination nodes is ignorable. The received signal at the relay is expressed as

$$\mathbf{r}_{\rm in} = \underbrace{\mathbf{h}_{\rm sr}\sqrt{P_s}s + \mathbf{H}_{\rm rr}\mathbf{r}_{\rm out} + \mathbf{n}_{\rm r}}_{=:\mathbf{u}_{\rm in}} + \mathbf{e}_{\rm in}, \quad (1)$$

where $\mathbf{r}_{\rm in} \in \mathbb{C}^{M_{\rm r}}, \mathbf{r}_{\rm out} \in \mathbb{C}^{M_{\rm t}}$ respectively represent the received and transmitted signal from the relay and $\mathbf{n}_{\rm r} \sim \mathcal{CN}\{\mathbf{0}, \sigma_{\rm nr}^2\mathbf{I}_{M_{\rm r}}\}$ represents the zero-mean additive white complex Gaussian (ZMAWCG) noise at the relay. The transmitted data symbol from the source is denoted as $s \in \mathbb{C}$, $\mathbb{E}\{ss^*\} = 1$. $P_{\rm s}$ is the source transmit power and $\mathbf{u}_{\rm in} \in \mathbb{C}^{M_{\rm r}}$ represents the *undistorted* received signal at the relay. The receiver distortion is denoted as

$$\mathbf{e}_{\rm in} \sim \mathcal{CN}\left\{\mathbf{0}, \beta\text{diag}\left(\mathbb{E}\left\{\mathbf{u}_{\rm in}\mathbf{u}_{\rm in}^H\right\}\right)\right\}, \quad (2)$$

and represents the combined effects of the receiver chain impairments where $\beta \in \mathbb{R}^+$ relates the intensity of the signal power to the error variance at each receiver chain. The known, i.e., distortion-free, part of the self-interference signal is then suppressed in the receiver and the resulting signal is amplified to constitute the relay's output:[3]

$$\mathbf{r}_{\rm out} = \mathbf{u}_{\rm out} + \mathbf{e}_{\rm out}, \quad \mathbf{u}_{\rm out} := \mathbf{W}\mathbf{r}_{\rm supp}, \quad (3)$$

$$\mathbf{r}_{\rm supp} = \mathbf{r}_{\rm in} - \mathbf{H}_{\rm rr}\mathbf{u}_{\rm out}, \quad (4)$$

where $\mathbf{r}_{\rm supp} \in \mathbb{C}^{M_{\rm r}}$ and $\mathbf{W} \in \mathbb{C}^{M_{\rm t}\times M_{\rm r}}$ respectively represent the interference-suppressed version of the received signal and the relay amplification matrix. The *intended* transmit signal is denoted as $\mathbf{u}_{\rm out}$. Similar to the defined additive distortion in the receiver chains, the combined effects of the transmit chain impairments is denoted as

$$\mathbf{e}_{\rm out} \sim \mathcal{CN}\left\{\mathbf{0}, \kappa\text{diag}\left(\mathbb{E}\left\{\mathbf{u}_{\rm out}\mathbf{u}_{\rm out}^H\right\}\right)\right\}, \quad (5)$$

where $\kappa \in \mathbb{R}^+$ relates the intensity of the signal power to the error variance at each transmit chain. Note that the defined statistics in (2) and (5) convey the intuition that unlike the traditional additive white noise model, the higher transmit (receive) signal power result in higher transmit (receive) distortion intensity at the corresponding chain. For further elaborations on the used distortion model please see [5], [11], [12], and the references therein. Moreover, in order to take into account the transmit power limitations we impose

$$\mathbb{E}\{\|\mathbf{r}_{\rm out}\|_2^2\} \leq P_{\rm r,max}, \quad P_{\rm s} \leq P_{\rm s,max}, \quad (6)$$

where $P_{\rm r,max}$ and $P_{\rm s,max}$ respectively represents the maximum transmit power from the relay and the source. The transmitted signal from the relay node passes through the relay to destination channel and constitutes the received signal at the destination:

$$\mathbf{y} = \mathbf{H}_{\rm rd}\mathbf{r}_{\rm out} + \mathbf{n}_{\rm d}, \quad \hat{s} = \mathbf{z}^H\mathbf{y}, \quad (7)$$

where $\mathbf{y} \in \mathbb{C}^{M_{\rm d}}$ is the received signal at the destination, and $\mathbf{n}_{\rm d} \sim \mathcal{CN}\{\mathbf{0}, \sigma_{\rm nd}^2\mathbf{I}_{M_{\rm d}}\}$ is the ZMCSCG noise. The linear receiver filter and the estimated received symbol is denoted as $\mathbf{z} \in \mathbb{C}^{M_{\rm d}}$ and $\hat{s}$, respectively.

III. Distortion Loop Aware Analysis for Multiple Antenna AF-FD Relaying

In this part, we analyze the end-to-end performance as a function of the relay amplification matrix, $\mathbf{W}$, receive linear filter at the destination, $\mathbf{z}$, as well as the transmit power from the source, $P_{\rm s}$. By incorporating (1), (2) and (5) into (3) and (4) we have [3]

$$\mathbf{Q} = \mathbf{W}\Big(P_{\rm s}\mathbf{h}_{\rm sr}\mathbf{h}_{\rm sr}^H + \sigma_{\rm nr}^2\mathbf{I}_{M_{\rm r}} + \beta\text{diag}\big(\mathbb{E}\left\{\mathbf{u}_{\rm in}\mathbf{u}_{\rm in}^H\right\}\big) + \kappa\mathbf{H}_{\rm rr}\text{diag}(\mathbf{Q})\mathbf{H}_{\rm rr}^H\Big)\mathbf{W}^H, \quad (8)$$

where $\mathbf{Q} \in \mathbb{C}^{M_{\rm t}\times M_{\rm t}}$ is the covariance matrix of the undistorted transmit signal from the relay, i.e., $\mathbf{Q} := \mathbb{E}\{\mathbf{u}_{\rm out}\mathbf{u}_{\rm out}^H\}$. Furthermore, the undistorted receive covariance matrix can be formulated from (1)-(4) as

$$\mathbb{E}\{\mathbf{u}_{\rm in}\mathbf{u}_{\rm in}^H\} = P_{\rm s}\mathbf{h}_{\rm sr}\mathbf{h}_{\rm sr}^H + \sigma_{\rm nr}^2\mathbf{I}_{M_{\rm r}} + \mathbf{H}_{\rm rr}\mathbb{E}\{\mathbf{r}_{\rm out}\mathbf{r}_{\rm out}^H\}\mathbf{H}_{\rm rr}^H. \quad (9)$$

By recalling (3) and (5) the relay transmit covariance matrix can be formulated as

$$\mathbb{E}\{\mathbf{r}_{\rm out}\mathbf{r}_{\rm out}^H\} = \mathbf{Q} + \kappa\text{diag}\left(\mathbf{Q}\right), \quad (10)$$

and consequently from (8) and (9) it follows

$$\mathbf{Q} = \mathbf{W}\mathcal{R}\left(\mathbf{Q}\right)\mathbf{W}^H, \quad (11)$$

[1] In this work we focus on the effects of hardware impairments, and assume that the channel state information (CSI), representing all linear signal dependencies in the system, are known by dedicating an adequately long training sequence at the relay, see [5, Subsection III.A]. Therefore, the studied framework can serve best for the scenarios with a stationary channel where long training sequences can be utilized.

[2] It indicates that the channel remains constant in each communication frame, but may vary from frame to frame.

[3] The relay output signals, i.e., $\mathbf{u}_{\rm out}$ and $\mathbf{r}_{\rm out}$, are generated from the received signals in the previous symbol duration. The subsequent communicated symbols are assumed zero-mean and independent. The time (symbol) index is eliminated to simplify the notations.

where

$$\mathcal{R}(\mathbf{Q}) := P_{\text{s}}\mathbf{h}_{\text{sr}}\mathbf{h}_{\text{sr}}^{H} + \sigma_{\text{nr}}^{2}\mathbf{I}_{M_{\text{r}}} + \beta\text{diag}\left(P_{\text{s}}\mathbf{h}_{\text{sr}}\mathbf{h}_{\text{sr}}^{H} + \sigma_{\text{nr}}^{2}\mathbf{I}_{M_{\text{r}}}\right) + \beta\text{diag}\left(\mathbf{H}_{\text{rr}}\left[\mathbf{Q} + \kappa\text{diag}(\mathbf{Q})\right]\mathbf{H}_{\text{rr}}^{H}\right) + \kappa\mathbf{H}_{\text{rr}}\text{diag}(\mathbf{Q})\,\mathbf{H}_{\text{rr}}^{H}. \tag{12}$$

Note that (8)-(11) hold as the noise, the desired signal at subsequent symbol durations, and the distortion components are zero-mean and mutually independent. By applying the famous matrix equality $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^{\text{T}} \otimes \mathbf{A})\text{vec}(\mathbf{B})$, we can write (10) as

$$\text{vec}\left(\mathbb{E}\{\mathbf{r}_{\text{out}}\mathbf{r}_{\text{out}}^{H}\}\right) = \left(\mathbf{I}_{M_{\text{t}}^{2}} + \kappa\mathbf{S}_{\text{D}}\right)\text{vec}(\mathbf{Q}), \tag{13}$$

where $\mathbf{S}_{\text{D}}$ is a selection matrix with one or zero elements such that $\mathbf{S}_{\text{D}}\text{vec}(\mathbf{Q}) = \text{vec}(\text{diag}(\mathbf{Q}))$. Similarly from (11) we obtain

$$\text{vec}(\mathbf{Q}) = \left(\mathbf{I}_{M_{\text{t}}^{2}} - (\mathbf{W}^{*} \otimes \mathbf{W})\,\mathbf{A}\right)^{-1}(\mathbf{W}^{*} \otimes \mathbf{W})\,\mathbf{a}, \tag{14}$$

where

$$\mathbf{A} := \beta\mathbf{S}_{\text{D}}\left(\mathbf{H}_{\text{rr}}^{*} \otimes \mathbf{H}_{\text{rr}}\right)\left(\mathbf{I}_{M_{\text{t}}^{2}} + \kappa\mathbf{S}_{\text{D}}\right) + \kappa\left(\mathbf{H}_{\text{rr}}^{*} \otimes \mathbf{H}_{\text{rr}}\right)\mathbf{S}_{\text{D}}, \tag{15}$$

$$\mathbf{a} := \left(\mathbf{I}_{M_{\text{r}}^{2}} + \beta\mathbf{S}_{\text{D}}\right)\text{vec}\left(P_{\text{s}}\mathbf{h}_{\text{sr}}\mathbf{h}_{\text{sr}}^{H} + \sigma_{\text{nr}}^{2}\mathbf{I}_{M_{\text{r}}}\right). \tag{16}$$

The direct dependence of the relay transmit covariance matrix and $\mathbf{W}$ can be consequently obtained from (14) and (13) as

$$\text{vec}\left(\mathbb{E}\{\mathbf{r}_{\text{out}}\mathbf{r}_{\text{out}}^{H}\}\right) = \left(\mathbf{I}_{M_{\text{t}}^{2}} + \kappa\mathbf{S}_{\text{D}}\right)\left(\mathbf{I}_{M_{\text{t}}^{2}} - (\mathbf{W}^{*} \otimes \mathbf{W})\,\mathbf{A}\right)^{-1} \times (\mathbf{W}^{*} \otimes \mathbf{W})\,\mathbf{a}. \tag{17}$$

In order to formulate the end-to-end link quality, we observe that the received signal power at the destination, after application of $\mathbf{z}$, can be separated as

$$P_{\text{desired}} = \mathbb{E}\{|\mathbf{z}^{H}\mathbf{H}_{\text{rd}}\mathbf{W}\mathbf{h}_{\text{sr}}\sqrt{P_{s}}s|^{2}\},\ P_{\text{error}} = P_{\text{tot}} - P_{\text{desired}}, \quad P_{\text{tot}} = \mathbb{E}\{|\mathbf{z}^{H}\mathbf{H}_{\text{rd}}\mathbf{r}_{\text{out}} + \mathbf{z}^{H}\mathbf{n}_{\text{d}}|^{2}\}, \tag{18}$$

where P_{desired} and P_{error} respectively represent the power of the desired and distortion-plus-noise parts of the estimated signal $\hat{s}$, and $P_{\text{tot}} := \mathbb{E}\{\hat{s}\hat{s}^{*}\}$. Due to the defined single stream communication setup, the end-to-end rate maximization can be equivalently considered as a signal-to-error-ratio (SER) maximization problem

$$\max_{P_{\text{s}},\mathbf{z},\mathbf{W}} \quad \frac{P_{\text{desired}}}{P_{\text{error}}} \tag{19a}$$

$$\text{s.t.} \quad \mathbf{Q} \in \mathcal{H},\ \ \text{tr}(\mathbf{Q}) \leq \tilde{P}_{\text{r,max}},\ \ P_{\text{s}} \leq P_{\text{s,max}},\ \ (14), \tag{19b}$$

where $\mathcal{H}$ represents the set of positive semi-definite matrices, and (19b) limits the feasible set of $\mathbf{W}$ to those resulting in a feasible $\mathbf{Q}$. Note that $\tilde{P}_{\text{r,max}} := \frac{P_{\text{r,max}}}{1+\kappa}$, and the power constraint in (19b) follows as $\text{tr}(\mathbf{Q} + \kappa\text{diag}(\mathbf{Q})) = (1+\kappa)\text{tr}(\mathbf{Q})$. It can be observed that the source transmit power constraint is always tight at the optimality[4]. The defined problem in (19a)-(19b) is a non-convex optimization problem and can not be solved in a closed form. In order to approach the solution we propose a gradient projection based optimization method in the following section.

[4]This is grounded on the fact that for any $P_{\text{s}} < P_{\text{s,max}}$, the joint variable update $P_{\text{s}} \leftarrow P_{\text{s,max}}$ and $\mathbf{W} \leftarrow \mathbf{W}\sqrt{\frac{P_{\text{s}}}{P_{\text{s,max}}}}$, result in the same P_{desired}, see (18), while decreasing the P_{error}, see (18) in connection to (17).

IV. Gradient Projection for SER maximization

In this part we propose an iterative solution to (19a)-(19b) based on gradient projection method [5]. The idea is to update the variables in the improving directions. When a variable update violates any of the problem constraints, it will be projected into the feasible solution set. The update rule for $\mathbf{W}$ can be formulated as

$$\tilde{\mathbf{W}}^{(\ell)} = \mathcal{P}\left(\mathbf{W}^{(\ell)} + \tau \cdot \nabla_{\mathbf{W}^{*}}\left(\frac{P_{\text{desired}}}{P_{\text{error}}}\right)\right), \quad \mathbf{W}^{(\ell+1)} = \delta^{(\ell)}\mathbf{W}^{(\ell)} + (1-\delta^{(\ell)})\tilde{\mathbf{W}}^{(\ell)}, \tag{20}$$

where $\mathcal{P}(\cdot)$ represents the projection to the feasible solution space, ℓ represent the iteration index and $\nabla_{\mathbf{W}^{*}}(\cdot)$ represents the steepest descent direction, i.e., the conjugate gradient with respect to $\mathbf{W}^{*}$, which is obtained from (24)-(25) and the fact that $\nabla_{\mathbf{W}^{*}}\left(\frac{P_{\text{desired}}}{P_{\text{error}}}\right) = \frac{1}{P_{\text{error}}{}^{2}}\Big(\nabla_{\mathbf{W}^{*}}(P_{\text{desired}}) \cdot P_{\text{error}} - \nabla_{\mathbf{W}^{*}}(P_{\text{error}}) \cdot P_{\text{desired}}\Big)$. The variables $\tau, \delta \in \mathbb{R}^{+}$ act as step sizes. In the sequel we set $\tau = 0.1$ and δ is chosen following the Armijo step size rules [13], also see [5, Section IV.B] for similar elaborations. For each update iteration, if the updated $\mathbf{W}$ violates the problem constraints in (19b), it will be projected into the feasible solution space. In order to define the projection rule, let $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{H}$ be the eigen decomposition of the matrix $\mathbf{Q}$, corresponding to the updated (infeasible) $\mathbf{W}$, where $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues. A feasible relay undistorted transmit covariance matrix, i.e., $\mathbf{Q}_{\text{new}}$ can be obtained as

$$\mathbf{Q}_{\text{new}} \leftarrow \mathbf{U}\underbrace{\left(\mathbf{\Lambda} - \nu\mathbf{I}_{M_{\text{r}}}\right)^{+}}_{=:\mathbf{\Lambda}_{\text{new}}}\mathbf{U}^{H}, \quad \nu \in \mathbb{R}_{+}, \tag{21}$$

where $(\cdot)^{+}$ substitutes the negative elements by zero, and $\nu \in \mathbb{R}$ is the minimum non-negative value that satisfies $\text{tr}\{\mathbf{\Lambda}_{\text{new}}\} \leq \tilde{P}_{\text{r,max}}$. The $\mathbf{W}_{\text{new}}$, as the projected version of $\mathbf{W}$ can be then calculated as

$$\mathbf{W}_{\text{new}} \leftarrow \mathbf{Q}_{\text{new}}^{\frac{1}{2}}\mathbf{V}\mathcal{R}^{\frac{-1}{2}}(\mathbf{Q}_{\text{new}}), \tag{22}$$

where $\mathcal{R}(\mathbf{Q})$ is defined in (12). In the above formulations $\mathbf{V}$ is a unitary matrix and $\mathbf{Q}_{\text{new}}^{\frac{1}{2}} = \mathbf{U}\mathbf{\Lambda}_{\text{new}}^{\frac{1}{2}}\mathbf{U}^{H}$, $\mathcal{R}^{\frac{-1}{2}}(\mathbf{Q}_{\text{new}}) = \mathbf{U}_{\text{r}}\mathbf{\Sigma}_{\text{r}}^{\frac{-1}{2}}\mathbf{U}_{\text{r}}^{H}$, where $\mathbf{U}_{\text{r}}\mathbf{\Sigma}_{\text{r}}\mathbf{U}_{\text{r}}^{H}$ represents the eigenvalue decomposition of $\mathcal{R}(\mathbf{Q}_{\text{new}})$. After each successful (feasible) update of $\mathbf{W}$, the value of the receive linear filter, i.e., $\mathbf{z}$, will be updated. The optimal choice of $\mathbf{z}$ can be obtained for a given $\mathbf{Q}$ as

$$\mathbf{z}^{\star} = \left(\mathbf{H}_{\text{rd}}\left(\mathbf{Q} + \kappa\text{diag}(\mathbf{Q})\right)\mathbf{H}_{\text{rd}}^{H} + \sigma_{\text{nd}}^{2}\mathbf{I}_{M_{\text{d}}}\right)^{-1}\mathbf{H}_{\text{rd}}\mathbf{W}\mathbf{h}_{\text{sr}}. \tag{23}$$

Please note that (23) does not affect the problem constraints (19b) and hence does not lead to infeasibility. The update iterations for $\mathbf{W}$ and $\mathbf{z}$ are continued until a stable feasible point is achieved. Please note that this algorithm leads to a local optimum solution due to the monotonically increasing SER following (23), and (20) following the Armijo step size rule [13]. Nevertheless, it does not guarantee a global optimality. Hence, we repeat the algorithm with multiple initial points to ensure, with higher confidence, that the obtained performance is close to optimality.

$$(\nabla_{\mathbf{w}^*} P_{\text{error}})^T = \text{vec}\left(\left(\mathbf{H}_{\text{rd}}^H \mathbf{z}\mathbf{z}^H \mathbf{H}_{\text{rd}}\right)^T\right)^T \left(\mathbf{I}_{M_t^2} + \kappa \mathbf{S}_{\text{D}}\right) \left(\left[\mathbf{a} + \mathbf{A}\left(\mathbf{I}_{M_r^2} - (\mathbf{W}^* \otimes \mathbf{W})\,\mathbf{A}\right)^{-1} (\mathbf{W}^* \otimes \mathbf{W})\,\mathbf{a}\right]^T \otimes \left(\mathbf{I}_{M_r^2} - (\mathbf{W}^* \otimes \mathbf{W})\,\mathbf{A}\right)^{-1} \right) \mathbf{S}_{\text{K}} \left(\mathbf{w} \otimes \mathbf{I}_{M_r^2}\right) - \left(\mathbf{z}^T \mathbf{H}_{\text{rd}}^*\right) \otimes \left(P_{\text{s}} \mathbf{z}^H \mathbf{H}_{\text{rd}} \mathbf{W} \mathbf{h}_{\text{sr}} \mathbf{h}_{\text{sr}}^H\right) \mathbf{S}_{\text{T}}, \tag{24}$$

$$(\nabla_{\mathbf{w}^*} P_{\text{desired}})^T = \left(\mathbf{z}^T \mathbf{H}_{\text{rd}}^*\right) \otimes \left(P_{\text{s}} \mathbf{z}^H \mathbf{H}_{\text{rd}} \mathbf{W} \mathbf{h}_{\text{sr}} \mathbf{h}_{\text{sr}}^H\right) \mathbf{S}_{\text{T}}, \text{ where } \mathbf{w} := \text{vec}(\mathbf{W}),\ \mathbf{S}_{\text{K}} \in \{0,1\}^{M_r^2 M_t^2 \times M_r^2 M_t^2},\ \mathbf{S}_{\text{T}} \in \{0,1\}^{M_r M_t \times M_r M_t}, \tag{25}$$

$$\text{such that: } \text{vec}\left(\mathbf{W}^* \otimes \mathbf{W}\right) = \mathbf{S}_{\text{K}} \text{vec}\left(\mathbf{w}^* \mathbf{w}^T\right), \text{ and } \mathbf{S}_{\text{T}} \text{vec}(\mathbf{W}) = \text{vec}(\mathbf{W}^T). \tag{26}$$

V. Iterative Covariance Shaping via Quadratic Approximation

The proposed gradient-based optimization requires high iteration counts and imposes high computational complexity. In this section we propose an iterative method which is based on quadratic approximation of relay transmit covariance matrix, with respect to $\mathbf{W}$. In this way the resulting optimization problem can be cast as a separately convex (and not a jointly convex) optimization problem in each iteration. In each optimization iteration, $\mathbf{Q}^{(\ell)}$ is approximated as

$$\mathbf{Q}^{(\ell)} \approx \mathbf{W}^{(\ell)} \mathcal{R}\left(\mathbf{Q}^{\star(\ell-1)}\right) \mathbf{W}^{(\ell)H},\ \|\mathbf{W}^{(\ell)} - \mathbf{W}^{(\ell-1)}\|_{\text{fro}} \leq \zeta^{(\ell)}, \tag{27}$$

where ζ represents a trust region in which the covariance approximation is valid, and will be chosen numerically. The index ℓ represents the approximation iteration, and $\mathbf{W}^{\star(\ell-1)}, \mathbf{Q}^{\star(\ell-1)}$ are the obtained $\mathbf{W}$ and $\mathbf{Q}$ in the $(\ell-1)$-th optimization step.[5] It is worth mentioning, that the defined approximation (27) becomes more accurate as the distortion coefficients, i.e, β and κ, become smaller and the distortion components in (12) become negligible. This consequently allows for larger trust region, i.e, ζ, and faster convergence. Applying $P_{\text{s}}^\star = P_{\text{s,max}}$ and exploiting the fact that the ratio $\frac{P_{\text{desired}}}{P_{\text{error}}}$ is strictly monotonic with respect to $\epsilon := \frac{P_{\text{desired}}}{P_{\text{tot}}}$, the optimization problem (19a)-(19b) can be approximated in iteration ℓ as

$$\max_{\mathbf{z}^{(\ell)}, \mathbf{W}^{(\ell)}, \epsilon^{(\ell)}} \quad \epsilon^{(\ell)} \tag{28a}$$

$$\text{s.t.} \quad \text{tr}\left(\mathbf{Q}^{(\ell)}\right) \leq \tilde{P}_{\text{r,max}}, \text{ and (27)}, \tag{28b}$$

where the last constraint enforces the covariance approximation (27), as well as the positive semi-definiteness constraint for the relay undistorted transmit covariance. A feasibility check, corresponding to the ratio $\epsilon^{(\ell)}$ can be hence written as

$$\max_{\mathbf{z}^{(\ell)}, \mathbf{W}^{(\ell)}, t^{(\ell)}} \quad P_{\text{desired}} - t^{(\ell)} \epsilon^{(\ell)} \tag{29a}$$

$$\text{s.t.} \quad P_{\text{tot}} \leq t^{(\ell)}, \quad \text{and (28b)}, \tag{29b}$$

where a value $\epsilon^{(\ell)}$ with a corresponding $\mathbf{z}^{(\ell)}, \mathbf{W}^{(\ell)}, t^{(\ell)}$ is feasible iff the maximized objective in (29a) is not negative. Note that we look for the maximum value of $\epsilon^{(\ell)}$ for which the feasibility holds. The defined feasibility check (29a)-(29b) is reformulated as

$$\max_{\mathbf{z}^{(\ell)}, \mathbf{W}^{(\ell)}, t^{(\ell)}} \quad \mathbf{w}^{(\ell)H} \mathbf{C}^{(\ell)} \mathbf{w}^{(\ell)} - t^{(\ell)} \epsilon \tag{30a}$$

$$\text{s.t.} \quad \mathbf{z}^{(\ell)H} \left(\mathbf{H}_{\text{rd}} \mathbf{W}^{(\ell)} \mathcal{R}\left(\mathbf{Q}^{\star(\ell-1)}\right) \mathbf{W}^{(\ell)H} \mathbf{H}_{\text{rd}}{}^H + \sigma_{\text{nd}}^2 \mathbf{I}_{M_d}\right) \mathbf{z}^{(\ell)} \leq t^{(\ell)},\ \|\mathbf{W}^{(\ell)} - \mathbf{W}^{(\ell-1)}\|_{\text{fro}} \leq \zeta^{(\ell)}, \tag{30b}$$

$$\text{tr}\left(\mathbf{W}^{(\ell)} \mathcal{R}\left(\mathbf{Q}^{\star(\ell-1)}\right) \mathbf{W}^{(\ell)H}\right) \leq P_{\text{r,max}}, \tag{30c}$$

where $\mathbf{w}^{(\ell)} := \text{vec}\left(\mathbf{W}^{(\ell)}\right)$, and $\mathbf{C}^{(\ell)} := P_{\text{s,max}} \left[\mathbf{h}_{\text{sr}}^T \otimes \left(\mathbf{z}^{(\ell)H} \mathbf{H}_{\text{rd}}\right)\right]^H \left[\mathbf{h}_{\text{sr}}^T \otimes \left(\mathbf{z}^{(\ell)H} \mathbf{H}_{\text{rd}}\right)\right]$. Our approach is to provide a maximization process for (30a)-(30c) together with an increasing update of $\epsilon^{(\ell)}$, until a joint stable point for $\epsilon^{(\ell)}$ and the optimization variables $\mathbf{z}^{(\ell)}, \mathbf{W}^{(\ell)}, t^{(\ell)}$ is achieved. Unfortunately, (30a)-(30c) is not a jointly (or separately) convex optimization problem. The following lemma will be consequently used to provide a separately-convex structure for (30a)-(30c).

Lemma 1. The maximization of $\mathbf{w}^H \mathbf{D} \mathbf{w}$ over $\mathbf{w}$, is equivalent to

$$\max_{\mathbf{w}, \mathbf{q}} \quad \mathbf{w}^H \mathbf{C} \mathbf{q} + \mathbf{q}^H \mathbf{C} \mathbf{w} - \mathbf{q}^H \mathbf{C} \mathbf{q}. \tag{31}$$

where $\mathbf{q}$ is with the same domain as $\mathbf{w}$ and $\mathbf{D} \succeq 0$.

Proof: The maximization of (31) over $\mathbf{q}$ is an unconstrained convex problem. By equalizing the derivative to zero we obtain an optimal $\mathbf{q}$ as $\mathbf{q}^\star = \mathbf{w}$. This consequently equalizes the objective in (31) to $\mathbf{w}^H \mathbf{C} \mathbf{w}$. ■

Consequently, the problem in (30a)-(30c) can be formulated as

$$\max_{\mathbf{z}^{(\ell)}, \mathbf{w}^{(\ell)}, \mathbf{q}^{(\ell)}, t^{(\ell)}} \quad \left(\mathbf{w}^{(\ell)H} \mathbf{C}^{(\ell)} \mathbf{q}^{(\ell)} + \mathbf{q}^{(\ell)H} \mathbf{C}^{(\ell)} \mathbf{w}^{(\ell)} - \mathbf{q}^{(\ell)H} \mathbf{C}^{(\ell)} \mathbf{q}^{(\ell)} - t^{(\ell)} \epsilon\right) \tag{32a}$$

$$\text{s.t.} \quad \text{(30b), (30c)}. \tag{32b}$$

Note that (32a)-(32b) is not a jointly convex optimization problem. Nevertheless it is separately convex over $\mathbf{w}^{(\ell)}, t^{(\ell)}$ or $\mathbf{q}^{(\ell)}$, if the other variables are kept constant. Furthermore, for a fixed value of $\mathbf{w}^{(\ell)}$, the optimal $\mathbf{z}^{(\ell)}$ can be calculated from (23). This facilitates an iterative optimization for (32a)-(32b) where in each step the problem is reduced to a convex problem. At the end of each optimization step, including the calculation of $\mathbf{w}^{(\ell)}, t^{(\ell)}$ and $\mathbf{q}^{(\ell)}$, the value of ϵ will be updated

[5] The value of $\mathbf{Q}^{\star(0)}$ can be initialized as an all-zero matrix.

to equalize the objective (32a) to zero:

$$\epsilon^{(\ell)} \leftarrow \frac{1}{t^{(\ell)}} \left(\mathbf{w}^{(\ell)^H} \mathbf{C}^{(\ell)} \mathbf{q}^{(\ell)} + \mathbf{q}^{(\ell)^H} \mathbf{C}^{(\ell)} \mathbf{w}^{(\ell)} - \mathbf{q}^{(\ell)^H} \mathbf{C}^{(\ell)} \mathbf{q}^{(\ell)} \right). \tag{33}$$

The above update results in a feasible $\epsilon^{(\ell)}$ for the corresponding values of $\mathbf{w}^{(\ell)}, \mathbf{z}^{(\ell)}, t^{(\ell)}$, see (29a) in connection to Lemma 1. The aforementioned iterative process, i.e., the approximation (27) and the optimization of (32a)-(32b), together with the update for ϵ will be continued until a stable point is achieved.

A. Convergence and Algorithm Description

Due to the separate convexity of (32a)-(32b), the objective value (32a) will be necessarily increased after each optimization step, for an adequately small trust region in (33). As a result, each update of $\epsilon^{(\ell)}$ results in a necessary increase. The aforementioned monotonicity results in a necessary convergence in the choice of $\mathbf{w}^{(\ell)}, \mathbf{z}^{(\ell)}, t^{(\ell)}$, which leads to a local optimum solution. Please note that while the obtained optimal solution is not necessarily a global optimum solution, it provides convergence within smaller number of iterations, and requires less computation compared to the proposed design in Section IV.

VI. Simulation Results

In this part, we study the behavior of the proposed optimal designs via Monte Carlo simulations. We average the results over 100 channel realizations, where all channels follow the uncorrelated Rayleigh flat-fading model. The resulting end-to-end rate is evaluated with respect to the transmit (receive) chain inaccuracy, i.e., κ (β). Our comparison includes the GP-based method in Section IV (GP-FD), the proposed method in Section V (MultiCVX-FD), the proposed method in [14, Section 3] for high and low dynamic range regime where the impact of distortion loop is ignored ([13]-HighSIC-FD), ([13]-LowSIC-FD), the performance of the equivalent AF-HD relaying system (AF-HD) and the optimized DF-FD relaying setup (DF-FD). The following values are used to define the simulation setup: $M_{\text{d}} = M_{\text{t}} = M_{\text{r}} = 4$, $\gamma_{\text{sr}} = \gamma_{\text{rd}} = 1$, $\gamma_{\text{rr}} = 1000$, $P_{\text{r,max}} = P_{\text{s,max}} = 1$ [Watt], $\sigma_{\text{nr}}^2 = \sigma_{\text{nd}}^2 = 0.1$ [Watt], where γ represents the variance of the channel coefficients for the respective channel. As it can be observed, the consideration of the distortion loop effect is essential as the chain inaccuracy increases, i.e., higher κ and β. Furthermore, the relative performance gain of the DF scheme is more significant for bigger κ, β. This is expected, as the discussed distortion loop effect for an AF-FD relay becomes more dominant as the chain inaccuracy increases.

VII. Conclusion

Due to the inter-dependency of the transmit relay power and the residual self-interference intensity, an AF-FD relay suffers from a distortion loop which degrades the performance. In this paper, we have proposed and evaluated optimization strategies to alleviate this effect in a multiple antenna system. The numerical comparison to the available designs and a DF-FD relaying system with a similar setup, has revealed the significant role of the distortion loop effect and shows the importance of a distortion-loop-aware design approach.

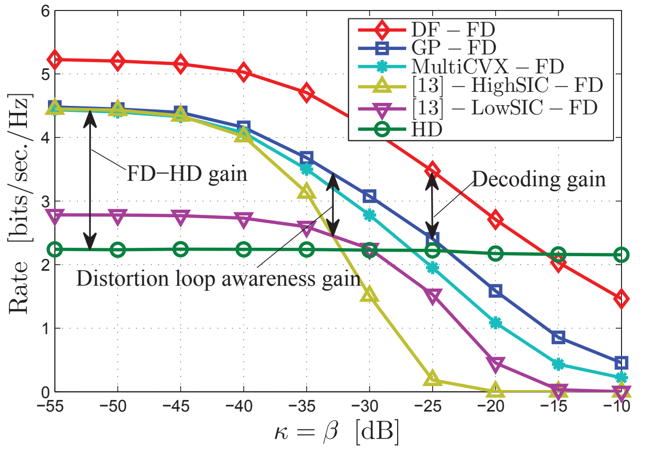

Fig. 2. Achievable rate vs. Tx and Rx chain inaccuracy κ, β). A distortion loop aware design is essential for higher chain inaccuracy.

References

[1] S. Hong, J. Brand, J. Choi, M. Jain, J. Mehlman, S. Katti, and P. Levis, "Applications of self-interference cancellation in 5G and beyond," *IEEE Communications Magazine*, February 2014.

[2] D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation*, ser. NSDI'14, Berkeley, CA, USA, 2014, pp. 359–372.

[3] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *Proceedings of the ACM SIGCOMM*, Aug. 2013.

[4] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Transactions on Signal Processing*, vol. 59, 2011.

[5] B. Day, A. Margetts, D. Bliss, and P. Schniter, "Full-duplex MIMO relaying: Achievable rates under limited dynamic range," *IEEE Journal on Selected Areas in Communications*, Sep. 2012.

[6] O. Taghizadeh, M. Rothe, A. C. Cirik, and R. Mathar, "Distortion-Loop analysis for Full-Duplex Amplify-and-Forward relaying in cooperative multicast scenarios," in *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Cairns, Australia, Dec. 2015, pp. 107–115.

[7] K. Lee, H. Kwon, M. Jo, H. Park, and Y. Lee, "MMSE-based optimal design of full-duplex relay system," in *IEEE Vehicular Technology Conference (VTC Fall)*, Sept 2012.

[8] H. Suraweera, I. Krikidis, G. Zheng, C. Yuen, and P. Smith, "Low-complexity end-to-end performance optimization in MIMO full-duplex relay systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 913–927, Feb. 2014.

[9] U. Ugurlu, T. Riihonen, and R. Wichman, "Optimized in-band full-duplex MIMO relay under single-stream transmission," *Vehicular Technology, IEEE Transactions on*, no. 99, pp. 1–1, Jan. 2015.

[10] J. Zhang, O. Taghizadeh, and M. Haardt, "Joint source and relay precoding design for one-way full-duplex MIMO relaying systems," *Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2013.

[11] B. Day, A. Margetts, D. Bliss, and P. Schniter, "Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range," *IEEE Transactions on Signal Processing*, Jul. 2012.

[12] X. Xia, D. Zhang, K. Xu, W. Ma, and Y. Xu, "Hardware impairments aware transceiver for full-duplex massive MIMO relaying," *Signal Processing, IEEE Transactions on*, vol. 63, no. 24, pp. 6565–6580, Dec 2015.

[13] A. Cohen, "Stepsize analysis for descent methods," in *Proceedings of IEEE 12th International Symposium on Adaptive Processes*, Dec. 1973.

[14] 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," *Communications, IEEE Transactions on*, vol. 62, no. 8, pp. 2758–2772, Aug 2014.