Abstract—In this paper we address the linear precoding and decoding design problem for a bidirectional orthogonal frequency-division multiplexing (OFDM) communication system, between two multiple-input multiple-output (MIMO) full-duplex (FD) nodes. The effects of hardware distortion, leading to the residual self-interference and inter-carrier leakage, as well as the channel state information (CSI) error are taken into account. In the first step, the operation of an FD MIMO OFDM transceiver is modeled, relying on the results of the available related system measurements. As a result, the explicit impact of hardware inaccuracies on the residual self-interference and inter-carrier leakage is formulated in relation to the intended transmit/received signal. Afterwards, linear precoding and decoding designs are proposed to enhance the system performance following minimum-mean-squared-error (MMSE) and sum rate maximization strategies, assuming the availability of perfect or erroneous CSI. The proposed designs are based on the application of alternating optimization over the system parameters, leading to a necessary convergence. Numerical results show that a significant gain is obtained as the transceiver inaccuracy increases, compared to the approaches where the impact of nonlinear hardware distortions, leading to inter-carrier leakage, is ignored.

I. INTRODUCTION

FULL-Duplex (FD) transceivers are known for their capability to transmit and receive at the same time and frequency, and hence have the potential to enhance the spectral efficiency [2]. Nevertheless, such systems suffer from the inherent self-interference from their own transmitter. Recently, specialized self-interference cancellation (SIC) techniques, e.g., [3]–[7], have demonstrated an adequate level of isolation between transmit (Tx) and receive (Rx) directions to facilitate an FD communication and motivated a wide range of related studies, see, e.g., [2], [8], [9]. A common idea of such SIC techniques is to subtract the dominant part of the self-interference signal, e.g., a line-of-sight (LOS) self-interference path or near-end reflections, in the radio frequency (RF) analog domain so that the remaining signal can be further processed in the baseband, i.e., digital domain. Nevertheless, such methods are still far from perfect in a realistic environment mainly due to \( i) \) aging and inherent inaccuracy of the hardware (analog) elements, as well as \( ii) \) inaccurate channel state information (CSI) in the self-interference path, due to noise and limited channel coherence time. In this regard, inaccuracy of the analog hardware elements used in subtracting the dominant self-interference path in RF domain may result in severe degradation of SIC quality. This issue becomes more relevant in a realistic scenario, where unlike the demonstrated setups in the lab environment, analog components are prone to aging, temperature fluctuations, and occasional physical damage. Moreover, an FD link is vulnerable to CSI inaccuracy at the self-interference path in environments with a small channel coherence time, see [6, Subsection 3.4.1]. A good example of such challenge is a high-speed vehicle that passes close to an FD device, and results in additional reflective self-interference paths.

In order to combat the aforementioned issues, an FD transceiver may adapt its transmit/receive strategy to the expected nature of CSI inaccuracy, e.g., by directing the transmit beams away from the moving objects or operating in the directions with smaller impact of CSI error. Moreover, the accuracy of the transmit/receiver chain elements can be considered, e.g., by dedicating less power, or ignoring the chains with noisier elements in the signal processing. In this regard, a widely used model for the operation of a multiple-antenna FD transceiver is proposed in [10], assuming a single carrier communication system, where CSI inaccuracy as well as the impact of hardware impairments are taken into account. A gradient-projection-based method is then proposed in the same work for maximizing the sum rate in an FD bidirectional setup. Building upon the proposed benchmark, a convex optimization design framework is introduced in [11]–[13] by defining a price/threshold for the self-interference power, assuming the availability of perfect CSI and accurate transceiver operation. While this approach reduces the design computational complexity, it does not provide a reliable performance for a scenario with erroneous CSI, particularly regarding the self-interference path [14]. Consequently, the consideration of CSI and transceiver error in an FD bidirectional system is further studied in [15]–[18] by maximizing the system sum rate, in [19] by minimizing the sum mean-squared-error (MSE), and in [20]–[22] for minimizing the system power consumption.
under a required quality of service.

The aforementioned works focus on modeling and design methodologies for single-carrier FD bidirectional systems, under frequency-flat channel assumptions. In this regard, the importance of extending the previous designs for a multi-carrier (MC) system with a frequency selective channel is threefold. Firstly, due to the increasing rate demand of the wireless services, and following the same rationale for the promotion of FD systems, the usage of larger bandwidths becomes necessary. This, in turn, invalidates the usual frequency-flat assumption and calls for updated design methodologies. Secondly, unlike the half-duplex (HD) systems where the operation of different subcarriers can be safely separated in the digital domain, an FD system is highly prone to the inter-carrier leakage due to the impact of hardware distortions on the strong self-interference channel. This, in particular, calls for a proper modeling of the inter-carrier leakage as a result of nonlinear hardware distortions for FD transceivers. And finally, the frequency diversity on the frequency selective channels shall be properly exploited to enhance the system performance.

A. Related works on FD MC systems

In the early work by Riihonen et al. [23], the performance of a combined analog/digital SIC scheme is evaluated for an FD orthogonal-frequency-division-multiplexing (OFDM) transceiver, taking into account the impact of hardware distortions, e.g., limited analog-to-digital converter (ADC) accuracy. The problem of resource allocation and performance analysis for FD MC communication systems is then addressed in [24]–[30], however, assuming a single antenna transceiver. Specifically, an FD MC system is studied in [24]–[26] in the context of FD relaying, in [28], [29] and [27] in the context of FD cellular systems with non-orthogonal multiple access (NOMA) capability, and in [30] for rate region analysis of a hybrid HD/FD link. Moreover, an MC relaying system with hybrid decode/amplify-and-forward operation is studied in [31], with the goal of maximizing the system sum rate via scheduling and resource allocation. However, in all of the aforementioned designs, the behavior of the residual self-interference signal is modeled as a purely linear system. As a result, the impacts of the hardware distortions leading to inter-carrier leakage, as observed in [23], are neglected.

B. Contribution and paper organization

In this paper we study a bidirectional FD MIMO OFDM system, where the impacts of hardware distortions leading to imperfect SIC and inter-carrier leakage are taken into account. Our main contributions, together with the paper organization are summarized as follows:

- In the seminal work by Bliss et al. [10], a FD MIMO transceiver is modeled considering the impacts of limited dynamic range in transmit/receiver chains, however, assuming a single-carrier setup with a frequency-flat channel. In this work, we utilize the similar experimental results on MIMO OFDM systems [32]–[34], indicating the Gaussian and statistically independent nature of distortion signals at different chains and different time samples, and extend the available model to a FD MC system. As a result, in Section II the explicit impact of transmit/receiver chain distortions on residual self-interference and inter-carrier leakage is formulated in relation to the intended transmit/received signals. This is in contrast to the available prior works on MC systems [24]–[31], which model the residual self-interference signal via a linear system.
- Building on the proposed model, linear transmit/receive strategies are proposed in order to enhance the system performance. In Section III an alternating quadratic convex program (QCP), denoted as AltQCP, is proposed in order to obtain a minimum weighted MSE transceiver design. The known weighted-minimum-MSE (WMMSE) method [35] is then utilized to extend the AltQCP framework for maximizing the system sum rate. For both algorithms, a monotonic performance improvement is observed at each step, leading to a necessary convergence.
- In Section IV the proposed design in Section III is extended by also taking into account the impact of CSI error. This is done by updating the system model proposed in Section II. Moreover, a worst-case MMSE design is proposed as an alternating semi definite program (SDP), denoted as AltSDP. Similar to the previous methods, a monotonic performance improvement is observed at each step, leading to a necessary convergence.
- FD systems are vulnerable to the CSI error due to, e.g., additional reflections from a moving object, particularly due to the impact of residual self-interference. It is hence beneficial to obtain the worst-case CSI error conditions, for the purpose of worst-case performance evaluation. Moreover, other than the usual approach of optimizing system parameters independently for each frame [15]–[22], such knowledge can be used for prevention purposes, e.g., updating the channel training sequence to prevent destructive CSI error conditions. To facilitate this, in Section V a methodology to obtain the least favorable CSI error matrices is obtained, by transforming the resulting non-convex quadratic problem into a convex problem. Moreover, the computational complexity of the proposed AltSDP algorithm is analytically obtained in relation to the system dimensions.

Numerical simulations show that the application of a distortion-aware design is essential, as transceiver accuracy degrades, and inter-carrier leakage becomes a dominant factor. However, this improvement is obtained at the expense of a higher computational complexity.

C. Mathematical Notation

Throughout this paper, column vectors and matrices are denoted as lower-case and upper-case bold letters, respectively. Mathematical expectation, trace, inverse, determinant,
transpose, conjugate and Hermitian transpose are denoted by $\mathbb{E}\{\cdot\}$, $\text{tr}(\cdot)$, $\cdot^{-1}$ | $\cdot$, $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$, respectively. The Kronecker product is denoted by $\otimes$. The identity matrix with dimension $K$ is denoted as $I_K$ and $\text{vec}(\cdot)$ operator stacks the elements of a matrix into a vector. $\mathbf{0}_{m \times n}$ represents an all-zero matrix with size $m \times n$. $\| \cdot \|_2$ and $\| \cdot \|_F$ respectively represent the Euclidean and Frobenius norms. diag($\cdot$) returns a diagonal matrix by putting the off-diagonal elements to zero. $|A|_{i=1,...,K}$ denotes a tall matrix, obtained by stacking the matrices $A_i$, $i = 1, ..., K$. $\mathcal{R}(A)$ represents the range (column space) of the matrix $A$. The set $\mathcal{F}_K$ is defined as $\{1, \ldots, K\}$. The set of real, positive, and complex numbers are respectively denoted as $\mathbb{R}, \mathbb{R}^+, \mathbb{C}$.

II. SYSTEM MODEL

We consider a bidirectional OFDM communication system between two MIMO FD transceivers. Each communication direction is associated with $N_t$ transmit and $M_t$ receive antennas, where $i \in \mathbb{I}$, and $\mathbb{I} := \{1,2\}$ represents the set of the communication directions. The desired channel in the communication direction $i$ and subcarrier $k \in \mathcal{F}_K$ is denoted as $\mathbf{H}_{ik}^d \in \mathbb{C}^{M_t \times N_i}$, where $K$ is the number of subcarriers. The self-interference channel from $i$ to $j$-th communication direction is denoted as $\mathbf{H}_{ik}^s \in \mathbb{C}^{M_t \times N_j}$. All channels are quasi-static and frequency-flat in each subcarrier. The transmitted signal in the direction $i$, subcarrier $k$ is formulated as

$$x_i^k = \mathbf{V}_{ik}^k s_i^k + e_{i,k}^t, \quad \sum_{k \in \mathcal{F}_K} \mathbb{E}\{\|x_i^k\|_2^2\} \leq P_i, \quad (1)$$

where $s_i^k \in \mathbb{C}^{d_i}$, $\mathbf{V}_{ik}^k \in \mathbb{C}^{N_i \times d_i}$, and $P_i \in \mathbb{R}^+$ respectively represent the vector of the data symbols, the transmit precoding matrix, and the maximum affordable transmit power. The number of the data streams in each subcarrier, and in direction $i$ is denoted as $d_i$, and $\mathbb{E}\{s_i^k s_i^{k H}\} = \mathbf{I}_{d_i}$. Moreover, $\mathbf{v}_i^k \in \mathbb{C}^{N_i}$ represents the desired signal to be transmitted, where $e_{i,k}^t$ models the inaccurate behavior of the transmit chain elements, i.e., transmit distortion, see Subsection II-A for more details.

The received signal at the destination can be consequently written as

$$y_i^k = \mathbf{H}_{ik}^r x_i^k + \mathbf{H}_{ik}^s s_j^k + n_i^k + e_{r,i}^k, \quad (2)$$

where $n_i^k \sim \mathcal{CN}\left(0, \sigma_i^2 \mathbf{I}_{M_t}\right)$ is the additive noise. Similar to the transmit signal model, $e_{r,i}^k$, represents the receiver distortion, and models the inaccuracies of the receive chain elements. The known, i.e., distortion-free, part of the self-interference signal is then subtracted from the received signal. This is formulated as

$$y_i^k = y_i^k - \mathbf{H}_{ij}^r s_j^k = \mathbf{H}_{ik}^r \mathbf{v}_i^k s_i^k + \nu_i^k, \quad (3)$$

where $\hat{y}_i^k$ is the received signal in direction $i$ and subcarrier $k$, after the self-interference cancellation. Moreover, the aggregate interference-plus-noise term is denoted as $\nu_i^k \in \mathbb{C}^{M_t}$, where

$$\nu_i^k = \mathbf{H}_{ij}^r e_{i,j}^t + \mathbf{H}_{ik}^s e_{i,k}^t + e_{i,k}^t + n_i^k, \quad j \neq i. \quad (4)$$

Finally, the estimated data vector is obtained at the receiver as

$$\hat{s}_i^k = \left(\mathbf{U}_i^k\right)^H \tilde{y}_i^k, \quad (5)$$

where $\mathbf{U}_i^k \in \mathbb{C}^{M_t \times d_i}$ is the linear receive filter.

A. Limited dynamic range in an FD OFDM system

In the seminal work by [10], [36], a model for the operation of a FD MIMO system is proposed, relying on the experimental results and modeling on the impact of hardware distortions. In this work, we rely on the similar experimental results conducted on a MIMO OFDM system, see [33] regarding the measurements on receiver chain distortions and [12], [34] for transmit distortions, as well as the recent characterizations of FD transceivers [23], [37] in order to provide a reasonable abstraction. The results of the aforementioned experiments hold four major indications:

1) The collective distortion signal in each chain can be approximated by an additive zero-mean Gaussian distortion,
2) The variance of the distortion signal is proportional to the power of the intended transmit/received signal,
3) The distortion signal is statistically independent to the intended transmit/receive signal at each chain,
4) The distortion signal is statistically independent for different chains, and at different time samples,

where (1) is substantiated in [32] and [23], Section II], and (3)-(4) are obtained in [32], Section IV], [33], also see [10], Section II] for a similar set of conclusions. Please note that the accuracy of the above-mentioned assumptions vary for different implementations of FD transceivers, depending on the complexity and the used SIC method. In this regard, the statistical independence of distortion elements defined in (3) and (4) hold also for an advanced implementation of a FD transceiver, assuming a high signal processing capability. This is since, any correlation structure in the distortion signal can be exploited and removed in order to reduce the residual self-interference via advanced signal processing. However, the linear dependence of the remaining distortion signal to the signal strength may vary for different SIC techniques, see [37], and hence is considered as an approximation.

Following the above arguments, the inaccurate function of the transmit chain elements, e.g., digital-to-analog converter (DAC) error, power amplifier noise and oscillator phase noise, are jointly modeled for each antenna as an additive distortion, and (4) hold also for an advanced implementation of a FD transceiver, assuming a high signal processing capability. This is since, any correlation structure in the distortion signal can be exploited and removed in order to reduce the residual self-interference via advanced signal processing. However, the linear dependence of the remaining distortion signal to the signal strength may vary for different SIC techniques, see [37], and hence is considered as an approximation.
where \( u_t, x_t, \) and \( e_{tx,i} \in \mathbb{C} \) respectively represent the intended transmit signal, the actual transmit signal, and the additive transmit distortion at the \( l \)-th transmit chain, and \( t \) denotes the instance of time. The set \( L_T \) represents the set of all transmit chains. Moreover, \( \theta_{tx,i} \in \mathbb{R}^+ \) represents the distortion coefficient for the \( l \)-th transmit chain, relating the collective power of the distortion signal, over the active spectrum, to the intended transmit power.

In the receiver side, the combined effect of the inaccurate hardware elements, i.e., ADC error, oscillator phase noise and automatic gain control noise, are presented as additive distortion terms and written as \( y_l(t) = u_l(t) + e_{rx,l}(t) \) such that

\[
e_{t,l}(t) \sim \mathcal{CN}\left(0, \theta_{tx,i} \mathbb{E}\{|u_l(t)|^2\}\right),
\]

\[
e_{t,l}(t) \perp u_l(t), \quad e_{t,l}(t) \perp e_{t,l}(t'), \quad l \neq l' \in L_R, \quad t \neq t',
\]

where \( u_l, e_{t,l}, \) and \( y_l \in \mathbb{C} \) respectively represent the intended (distortion-free) received signal, additive receive distortion, and the received signal from the \( l \)-th receive antenna. The set \( L_R \) represents the set of all receive chains. Similar to the transmit chain characterization, \( \theta_{rx,i} \in \mathbb{R}^+ \) is the distortion coefficient for the \( l \)-th receive chain, see Fig. 1.

In this work we consider a general framework where the transmit (receive) distortion coefficients are not necessarily identical for all transmit (receive) chains belonging to the same transceiver, i.e., different chains may hold different accuracy due to occasional damage and aging. This assumption is important in practice since it enables the design algorithms to reduce communication task on the chains with noisier elements. The statistics of the distortion terms, introduced in [1], [2] can be hence inferred as

\[
e_{t,i}^k \sim \mathcal{CN}\left(0, \frac{B}{B_{tot}} \Theta_{tx,i} \mathbf{P}_{tx,i}\right),
\]

\[
e_{t,i}^k \sim \mathcal{CN}\left(0, \frac{B}{B_{tot}} \Theta_{rx,i} \mathbf{P}_{rx,i}\right),
\]

and

\[
\mathbf{P}_{tx,i} := \sum_{k \in \mathcal{K}_x} \text{diag}\left(\mathbb{E}\left\{\mathbf{v}_i^k \mathbf{v}_i^k H\right\}\right),
\]

\[
\mathbf{P}_{rx,i} := \sum_{k \in \mathcal{K}_x} \text{diag}\left(\mathbb{E}\left\{\mathbf{u}_i^k \mathbf{u}_i^k H\right\}\right),
\]

where \( B \) and \( B_{tot} \) respectively represent the bandwidth associated with each subcarrier, and the total bandwidth of the system. In the above formulations, \( \Theta_{tx,i} \in \mathbb{R}^{N_t \times N_t}, (\Theta_{rx,i} \in \mathbb{R}^{M_r \times M_r}) \) is a diagonal matrix including distortion coefficients \( \theta_{tx,i} (\theta_{rx,i}) \) for the corresponding chains. Similarly, \( \mathbf{P}_{tx,i} \) (\( \mathbf{P}_{rx,i} \)) is a diagonal matrix with each diagonal element representing the intended transmit (receive) signal power at the corresponding chain.

\[\text{A simpler mathematical presentation can be obtained by assuming the same transceiver accuracy over all antennas. In such a case, the defined diagonal matrices can be replaced by a scalar.}\]

---

**Figure 1.** A FD transceiver model. Limited dynamic range is modeled by injecting additive distortion terms at each transmit or receive chain. \( e_{tx,i} \) and \( e_{rx,i} \) denote the distortion terms, and \( n_i \) represent the additive thermal noise.

Via the application of (10)-(13) on (8) we conclude

\[
\Sigma_k^x := \mathbb{E}\left\{\mathbf{v}_i^k \mathbf{v}_i^k H\right\}
\]

\[
= \sum_{j \in \mathcal{I}} \mathbf{H}_{ij}^k \Theta_{tx,j} \text{diag}\left(\sum_{l \in \mathcal{K}_x} \mathbf{V}_j^l \mathbf{V}_j^l H\right) \mathbf{H}_{ij}^k
\]

\[
+ \Theta_{tx,j} \text{diag}\left(\sum_{l \in \mathcal{K}_x} (\sigma^2_{ij} \mathbf{I}_{M_t} + \sum_{j \in \mathcal{I}} \mathbf{H}_{ij}^l \mathbf{V}_j^l \mathbf{V}_j^l H \mathbf{H}_{ij}^l H)\right) + \sigma^2_{ij} \mathbf{I}_{M_t}, \quad k \in \mathcal{K}_x,
\]

where \( \Sigma_k^x \in \mathbb{C}^{M_x \times M_x} \) is the covariance of the received collective interference-plus-noise signal, and is obtained considering \( 0 \leq \theta_{tx,i} < 1, 0 \leq \theta_{rx,i} < 1, \) and hence ignoring the terms containing higher orders of the distortion coefficients in (14).

**B. Remarks**

- In this section we have assumed the availability of perfect CSI, and focused on the impact of non-linear transceiver distortions. This assumption is relevant for the scenarios with stationary channel, e.g., a backhaul directive link with zero mobility [38], where an adequately long training sequence can be applied, see Subsection III.A. The impact of the CSI inaccuracy is later addressed in Section IV.
- As expected, the role of the distortion signals on the residual self-interference, including the resulting inter-subcarrier leakage, is evident from (14). This is the main goal of the remaining parts of this paper to incorporate and evaluate this impact on the design of the defined MC system.

---

**III. LINEAR TRANSCEIVER DESIGN FOR MULTI-CARRIER COMMUNICATIONS**

Via the application of \( \mathbf{V}_i^k \) and \( \mathbf{U}_i^k \), as the linear transmit precoder and receive filters, the mean-squared-error (MSE) matrix of the defined system is calculated as

\[
\mathbf{E}_i^k := \mathbb{E}\left\{\left(\hat{s}_i^k - s_i^k\right)\left(\hat{s}_i^k - s_i^k\right)^H\right\}
\]

\[
= \left(\mathbf{U}_i^k \mathbf{H}_{ii}^k \mathbf{V}_i^k - \mathbf{I}_{d_i}\right)\left(\mathbf{U}_i^k H_{ii}^k V_i^k - \mathbf{I}_{d_i}\right)^H
\]

\[
+ \mathbf{U}_i^k \Sigma_k^x \mathbf{U}_i^k,
\]

where \( \Sigma_k^x \) is given in (14). In the following we propose two design strategies for the defined system, proposing an alternating QCP framework.
A. Weighted MSE minimization via Alternating QCP (AltQCP)

An optimization problem for minimizing the weighted sum MSE is written as

$$\min_{\mathbf{V}, \mathbf{U}} \sum_{i \in I} \sum_{k \in F_k} \text{tr} \left( \mathbf{S}_i^k \mathbf{E}_i^k \right)$$

(16a)

s.t. \( \text{tr} \left( (\mathbf{I}_{N_i} + \Theta_{i,x,i}) \sum_{l \in F_k} \mathbf{V}_i^l \mathbf{V}_i^l H_{ii}^l \right) \leq P_i, \forall i \in I \),

(16b)

where \( \mathbf{X} := \{ \mathbf{X}_i^k, \forall i \in I, \forall k \in F_k \} \), with \( \mathbf{X} \in \{ \mathbf{U}, \mathbf{V} \} \), and (16b) represents the transmit power constraint. It is worth mentioning that the application of \( \mathbf{S}_i^k \succ 0 \), as a weight matrix associated with \( \mathbf{E}_i^k \) is two-folded. Firstly, it may appear as a diagonal matrix, emphasizing the importance of different data streams and different users. Secondly, it can be applied as an auxiliary variable which later relates the defined weighted MSE minimization to a sum-rate maximization problem, see Subsection III-B.

It is observed that (16) is not a jointly convex problem. Nevertheless, it holds a QCP structure separately over the sets \( \mathbf{V} \) and \( \mathbf{U} \), in each case when other variables are fixed. In this regard, the objective (16a) can be decomposed over \( \mathbf{U} \) for different communication directions, and for different subcarriers. The optimal minimum MSE (MMSE) receive filter can be hence calculated in closed form as

$$\mathbf{U}_{i,mmse}^k = \left( \mathbf{\Sigma}_i^k + \mathbf{H}_{ii}^k \mathbf{V}_i^k \mathbf{H}_{ii}^k \right)^{-1} \mathbf{H}_{ii}^k \mathbf{V}_i^k. \quad (17)$$

Nevertheless, the defined problem is coupled over \( \mathbf{V}^k \), due to the impact of inter-carrier leakage, as well as the power constraint (16b). The Lagrangian function, corresponding to the optimization (16) over \( \mathbf{V} \) is expressed as

$$\mathcal{L} (\mathbf{V}, \lambda) := \sum_{i \in I} \left( \lambda_i P_i (\mathbf{V}) + \sum_{k \in F_k} \text{tr} \left( \mathbf{S}_i^k \mathbf{E}_i^k \right) \right), \quad (18)$$

$$\mathcal{P}_i (\mathbf{V}) := -P_i + \text{tr} \left( (\mathbf{I}_{N_i} + \Theta_{i,x,i}) \sum_{l \in F_k} \mathbf{V}_i^l \mathbf{V}_i^l H_{ii}^l \right), \quad (19)$$

where \( \lambda := \{ \lambda_i, \forall i \in I \} \) is the set of dual variables. The dual function, corresponding to the above Lagrangian is defined as

$$\mathcal{F} (\lambda) := \min_{\mathbf{V}} \mathcal{L} (\mathbf{V}, \lambda) \quad (20)$$

where the optimal \( \mathbf{V}^* \) is obtained as

$$\mathbf{V}_{i}^* = \left( \mathbf{J}_i^k + \lambda_i (\mathbf{I}_{N_i} + \Theta_{i,x,i}) + \mathbf{H}_{ii}^k \mathbf{U}_{i}^k \mathbf{S}_{i}^k \mathbf{U}_{i}^k \mathbf{H}_{ii}^k \right)^{-1}$$

$$\times \mathbf{H}_{ii}^k \mathbf{U}_{i}^k \mathbf{S}_{i}^k, \quad (21)$$

and

$$\mathbf{J}_i^k := \sum_{l \in F_k} \sum_{j \in \Omega_l} \left( \mathbf{H}_{ji}^l \text{diag} \left( \mathbf{U}_j^l \mathbf{S}_j^l \mathbf{U}_j^l \Theta_{j,x,j}^l \right) \mathbf{H}_{ji}^l \right.$$

$$\left. + \text{diag} \left( \mathbf{H}_{ji}^l \Theta_{j,x,j} \mathbf{U}_j^l \mathbf{S}_j^l \mathbf{U}_j^l \mathbf{H}_{ji}^l \right) \right). \quad (22)$$

Due to the convexity of the original problem (16) over \( \mathbf{V} \), the defined dual problem is a concave function over \( \lambda \), with \( \mathcal{P}_i (\mathbf{V}) \) as a subgradient, see [39, Eq. (6.1)]. As a result, the optimal \( \lambda \) is obtained from the maximization

$$\lambda^* = \arg \max \mathcal{F} (\lambda), \quad (23)$$

following a standard subgradient update, [39, Subsection 6.3.1].

Utilizing the proposed optimization framework, the alternating optimization over \( \mathbf{V} \) and \( \mathbf{U} \) is continued until a stable point is obtained. Note that due to the monotonic decrease of the objective in each step, and the fact that (16a) is non-negative and hence bounded from below, the defined procedure leads to a necessary convergence. Algorithm 1 defines the necessary optimization steps.

**Algorithm 1** Alternating QCP (AltQCP) for weighted MSE minimization

1: \( \ell \leftarrow 0 \); (set iteration number to zero)
2: \( \mathbf{V} \leftarrow \) right singular matrix initialization, see [40, Appendix A]
3: \( \mathbf{U} \leftarrow \) solve (17)
4: repeat
5: \( \ell \leftarrow \ell + 1 \)
6: \( \mathbf{V} \leftarrow \) solve (21) or QCP (15), with fixed \( \mathbf{U} \)
7: \( \mathbf{U} \leftarrow \) solve (17) or QCP (15) with fixed \( \mathbf{V} \)
8: until a stable point, or maximum number of \( \ell \) reached
9: return \( \{ \mathbf{U}, \mathbf{V} \} \)

B. Weighted MMSE (WMMSE) design for sum rate maximization

Via the utilization of \( \mathbf{V}^k \) as the transmit precoders, the resulting communication rate for the \( k \)-th subcarrier and for the \( i \)-th communication direction is written as

$$I_{i}^k = \log_2 \left| \begin{bmatrix} \mathbf{I}_{d_i} + \mathbf{V}_i^k \mathbf{H}_{ii}^k (\mathbf{\Sigma}_i^k)^{-1} \mathbf{H}_{ii}^k \mathbf{V}_i^k \end{bmatrix} \right| \quad (24)$$

where \( B \) and \( \mathbf{\Sigma}_i^k \) are defined in (10) and (14). The sum rate maximization problem can be hence presented as

$$\max_{\mathbf{V}} \sum_{i \in I} \sum_{k \in F_k} I_{i}^k, \text{ s.t. (16b).} \quad (25)$$

The optimization problem (25) is intractable in the current form. In the following we propose an iterative optimization solution, following the WMMSE method [35].

Via the application of the MMSE receive linear filters from (17), the resulting MSE matrix is obtained as

$$\mathbf{E}_{i,mmse}^k = \left( \mathbf{I}_{d_i} + \mathbf{V}_i^k \mathbf{H}_{ii}^k (\mathbf{\Sigma}_i^k)^{-1} \mathbf{H}_{ii}^k \mathbf{V}_i^k \right)^{-1}. \quad (26)$$

By recalling (24), and upon utilization of \( \mathbf{U}_{i,mmse}^k \) we observe the following useful connection to the rate function

$$I_{i}^k = -\log_2 \left| \mathbf{E}_{i,mmse}^k \right|, \quad (27)$$

which facilitates the decomposition of rate function via the following lemma, see also [35, Eq. (9)].
Lemma III.1. Let $E \in \mathbb{C}^{d \times d}$ be a positive definite matrix. The maximization of the term $-\log |E|$ is equivalent to the maximization

$$\max_{E, S} \, tr(SE) + \log |S| + d,$$

where $S \in \mathbb{C}^{d \times d}$ is a positive definite matrix, and we have

$$S = E^{-1},$$

at the optimality.

Proof: See [41, Lemma 2].

By recalling (27), and utilizing Lemma III.1 the original optimization problem over $V$ can be equivalently formulated as

$$\max_{V, U, S} \sum_{k \in F_K} B \sum_{i \in \mathbb{I}} \left( \log |S_k^i| + d_i - tr(S_k^i E_k^i) \right) \text{ s.t. (16b)},$$

(30)

where $S := \{S_k^i > 0, \forall i \in \mathbb{I}, \forall k \in F_K\}$. The obtained optimization problem (30) is not a jointly convex problem. Nevertheless, it is a QCP over $V$ when other variables are fixed, and can be obtained with a similar structure as for (16). Moreover, the optimization over $U$ and $S$ is respectively obtained from (17), and (29) as $S_k^i = E_k^i$. This facilitates an alternating optimization where in each step the corresponding problem is solved to optimality, see Algorithm 2. The defined alternating optimization steps result in a necessary convergence due to the monotonic increase of the objective in each step, and the fact that the eventual system sum rate is bounded from above.

Algorithm 2 AltQCP-WMMSE design for sum rate maximization

1. $\ell \leftarrow 0$: (set iteration number to zero)
2. $V \leftarrow$ right singular matrix initialization [40] Appendix A)
3. repeat
4. $\ell \leftarrow \ell + 1$
5. $U \leftarrow$ solve (17)
6. $V \leftarrow$ solve QCP (30), with fixed $U, S$
7. $S \leftarrow S_k^i = (E_k^i)$
8. until a stable point, or maximum number of $\ell$ reached
9. return $\{V\}$

IV. ROBUST DESIGN WITH IMPERFECT CSI

In many realistic scenarios the CSI matrices can not be estimated or communicated accurately due to the limited channel coherence time as a result of, e.g., reflections from a moving object, or due to dedicating limited resource on the training/feedback process. This issue becomes more significant in a FD system, due to the strong self-interference channel which calls for dedicated silent times for tuning and training process, see [10] Subsection III.A. In particular, the impact of CSI error on the defined MC FD system is three-fold. Firstly, similar to the usual HD scenarios, it results in the erroneous equalization in the receiver, as the communication channels are not accurately known. Secondly, it results in an inaccurate estimation of the received signal from the self-interference path, and thereby degrades the SIC quality. Finally, due to the CSI error, the impact of the distortion signals may not be accurately known, as the statistics of the distortion signals directly depend on the channel situation. In this part we extend the proposed designs in Section III where the aforementioned uncertainties, resulting from CSI error, are also taken into account.

A. Norm-bounded CSI error

In this part we update the defined system model in Section II to the scenario where the CSI is known erroneously. In this respect we follow the so-called deterministic model [42], where the error matrices are not known but located, with a sufficiently high probability, within a known feasible error region. This is expressed as

$$H_{ij}^k = \tilde{H}_{ij}^k + \Delta_{ij}^k, \quad \Delta_{ij}^k \in \Delta_{ij}^k, \quad i, j \in \mathbb{I},$$

(31)

and

$$D_{ij}^k := \{ |D_{ij}^k \Delta_{ij}^k|_F \leq \zeta_{ij}^k \}, \quad \forall i, j \in \mathbb{I}, \quad k \in F_K,$$

(32)

where $\tilde{H}_{ij}^k$ is the estimated channel matrix and $\Delta_{ij}^k$ represents the channel estimation error. Moreover, $D_{ij}^k \succeq 0$ and $\zeta_{ij}^k \geq 0$ jointly define a feasible ellipsoid region for $\Delta_{ij}^k$, which generally depends on the noise and interference statistics, and the used channel estimation method. For further elaboration on the used error model see [42]–[44] and the references therein. The aggregate interference-plus-noise signal at the receiver is hence updated as

$$\nu_i^j = H_{ij}^k e_{ij}^k + H_{ij}^k e_{ij}^k + e_{ij}^k + \Delta_{ij}^k V_{ij}^k s_j^k + n_i^k, \quad j \neq i \in \mathbb{I},$$

(33)

where $\Sigma_j^i$, representing the covariance of $\nu_i^j$, is expressed in (34).

B. Alternating SDP (AltSDP) for worst-case MSE minimization

An optimization problem for minimizing the worst-case MSE under the defined norm-bounded CSI error is written as

$$\min_{\nu, U} \max_{i \in \mathbb{I}} \sum_{k \in F_K} \sum_{j \in \mathbb{I}} \text{tr} \left( S_{ij}^k E_k^i \right),$$

s.t. (16b), $\Delta_{ij}^k \in \Delta_{ij}^k, \forall i, j \in \mathbb{I}, \forall k \in F_K,$

(35)

where $C := \{ \Delta_{ij}^k | \forall i, j \in \mathbb{I}, \forall k \in F_K \}$, and $E_k^i$ is obtained from (15) and (33). Note that the above problem is intractable, due to the inner maximization of quadratic convex objective over $C$, which also invalidates the observed convex QCP structure in (16). In order to formulate the objective into
\[ \Sigma_i^k = \Delta_{ij}^k V_j^k V_j^{H} \Delta_{ij}^{kH} + \sum_{j \in I} H_{ij}^k \Theta_{\kappa,ij} \text{diag} \left( \sum_{l \in \mathbb{F}_K} V_j^l V_j^{H} \right) H_{ij}^k \Theta_{\kappa,ij} \text{diag} \left( \sum_{l \in \mathbb{F}_K} \left( \sigma_{2l}^2 I_M + \sum_{j \in I} H_{ij}^l V_j^l V_j^{H} \right) \right) + \sigma_{\kappa,ij}^2 I_M. \] (34)

A tractable form, we calculate
\[
\sum_{k \in \mathbb{F}_K} \text{tr} \left( S_k^k E_k^k \right) = \sum_{k \in \mathbb{F}_K} \left( \left\| W_k^H \left( U_k^{kH} H_k^H V_k^k - I_d \right) \right\|^2_F + \left\| W_k^H U_k^{kH} \right\|^2_F 
+ \sum_{j \in I} \sum_{l \in \mathbb{F}_N, m \in \mathbb{F}_K} \left\| W_k^H U_k^{kH} \left( \Theta_{\kappa,ij} \right) \frac{1}{2} \Gamma_{N}^l V_j^l \right\|^2_F 
+ \sum_{j \in I} \sum_{l \in \mathbb{F}_M, m \in \mathbb{F}_K} \left\| W_k^H U_k^{kH} \left( \Theta_{\kappa,ij} \right) \frac{1}{2} \Gamma_{M}^l V_j^l \right\|^2_F 
+ \left\| W_k^H U_k^{kH} \left( \Theta_{\kappa,ij} \right) \frac{1}{2} \right\|^2 \right) 
\geq \sum_{j \in I} \sum_{k \in \mathbb{F}_K} \left\| c_{ij}^k + C_{ij, vec} \left( \Delta_{ij}^k \right) \right\|^2_F, \] (37)

where \( \Gamma_{\lambda}^l = M \times M \) zero matrix except for the \( \lambda \)-th diagonal element equal to 1. In the above expressions \( W_k^i = (S_k^k)^{\frac{1}{2}} \), and
\[
c_{ij}^k := \begin{bmatrix} \delta_{ij} \text{vec} \left( W_k^H \left( U_k^{kH} H_k^H V_k^k - I_d \right) \right) \\
\text{vec} \left( W_k^H U_k^{kH} \left( \Theta_{\kappa,ij} \right) \frac{1}{2} \Gamma_{N}^l V_j^l \right) \\
\text{vec} \left( W_k^H U_k^{kH} \left( \Theta_{\kappa,ij} \right) \frac{1}{2} \Gamma_{M}^l V_j^l \right) \\
\delta_{ij} \text{vec} \left( W_k^H U_k^{kH} \left( \sigma_{2l}^2 I_M + \Theta_{\kappa,ij} \right) \frac{1}{2} \right) \end{bmatrix} \] (38)
\[
C_{ij}^k := \begin{bmatrix} V_j^T \otimes \left( W_k^H U_k^{kH} \right) \\
\left( \left( \Theta_{\kappa,ij} \right) \frac{1}{2} \Gamma_{N}^l V_j^l \right)^T \otimes \left( W_k^H U_k^{kH} \right) \\
\left( \left( \Theta_{\kappa,ij} \right) \frac{1}{2} \Gamma_{M}^l \right)^T \otimes \left( W_k^H U_k^{kH} \right) \\
0_{M_d \times M_d, N_i} \end{bmatrix}, \] (39)

where \( \delta_{ij} \) is the Kronecker delta where \( \delta_{ij} = 1 \) for \( i = j \) and zero otherwise. Moreover we have \( c_{ij}^k \in \mathbb{C}^{d_{ij} \times 1}, C_{ij}^k \in \mathbb{C}^{d_{ij} \times d_{ij}} \).
where $\mathbb{M} := \{\lambda_{ij}^k, \forall i, j \in I, k \in \mathbb{F}_K\}$, and

$$G_i := \left[ \begin{array}{c} \tilde{P}_i \\ \tilde{v}_i \end{array} \right], \quad \tilde{v}_i := \left( I + \Theta_{t_0, i} \right) \frac{1}{2} \mathbf{V}_i^k \right]_{k \in \mathbb{F}_K},$$

and

$$F_{i,j}^k := \left[ \begin{array}{c} \tau_{ij}^k - \lambda_{ij}^k \\ \mathbf{c}_{ij}^k \end{array} \right] : = \left[ \begin{array}{c} \mathbf{I}_{ij} \end{array} \right]_{k \in \mathbb{F}_K} + \mathbf{D}^k_{ij} \mathbf{C}_{ij}^k.$$  \quad (48)

Similar to (49), the obtained problem in (47) is not a jointly, but a separately convex problem over $V$ and $U$, in each case when the other variables are fixed. In particular, the optimization over $V, T, M$ is cast as an SDP, assuming a fixed $U$. Afterwards, the optimization over $U, T, M$ is solved as an SDP, assuming a fixed $V$. The described alternating steps are continued until a stable point is achieved, see Algorithm 3 for a detailed explanation.

Algorithm 3 Alternating SDP (AltSDP) for worst-case MMSE design under CSI error.

1: $\ell \leftarrow 0$; (set iteration number to zero)
2: $V, U \leftarrow$ similar initialization as Algorithm 1
3: repeat
4: $\ell \leftarrow \ell + 1$
5: $V, T, M \leftarrow$ solve SDP (47), with fixed $U$
6: $U, T, M \leftarrow$ solve SDP (47), with fixed $V$
7: until a stable point, or maximum number of $\ell$ reached
8: return $\{U, V\}$

C. WMMSE for sum rate maximization

Under the impact of CSI error, the worst-case rate maximization problem is written as

$$\max \min_{\mathcal{W}} \sum \sum_{k \in \mathbb{F}_K} I_i^k$$

s.t. (16b) $\Delta_{ij}^k \in \mathbb{D}_i^k, \forall i, j \in I, k \in \mathbb{F}_K$. \quad (50b)

Via the application of Lemma III.1 and (27) the rate maximization problem is equivalently written as

$$\max \min_{\mathcal{W}} \sum \sum_{k \in \mathbb{F}_K} B \left( \log \left| W_i^k W_i^H \right| \right) + d_i - \text{tr} \left( W_i^k E_i^k W_i^k \right)$$

s.t. (50b), \quad (51a)

where $\mathcal{W} := \{W_i^k, \forall i \in I, k \in \mathbb{F}_K\}$. The above problem is not tractable in the current form, due to the inner min-max structure. Following the max-min exchange introduced in [41], Section III, and undertaking similar steps as in (37)-(46) the problem (51) is turned into

$$\max \sum \sum_{k \in \mathbb{F}_K} B \left( 2 \log |W_i^k| + d_i - \sum_{j \in I} \sum_{k \in \mathbb{F}_K} a_{ij}^k \right)$$

s.t. $F_{i,j}^k \succeq 0, G_i \succeq 0, \forall i, j \in I, k \in \mathbb{F}_K$, \quad (52a)

where $F_{i,j}^k, G_i$ are defined in (47). It is observable that the transformed problem holds a separately, but not a jointly, convex structure over the optimization variable sets. In particular, the optimization over $V, T, M$ and $U, T, M$ are cast as SDP in each case when other variables are fixed. Moreover, the optimization over $W$ can be efficiently implemented using the MAX-DET algorithm [43], see Algorithm 4. Similar to Algorithm 3 due to the monotonic increase of the objective in each optimization iteration the algorithm converges to a stationary point. See [41, Section III] for arguments regarding convergence and optimization steps for a problem with a similar variable separation.

Algorithm 4 AltSDP-WMMSE algorithm for worst-case rate maximization under CSI error

1: $\ell \leftarrow 0$; (set iteration number to zero)
2: $V, U \leftarrow$ similar initialization as Algorithm 1
3: $W \leftarrow$ identity matrix initialization
4: repeat
5: $\ell \leftarrow \ell + 1$
6: $V, T, M \leftarrow$ solve SDP (47), with fixed $U, W$
7: $U, T, M \leftarrow$ solve SDP (47), with fixed $V, W$
8: until a stable point, or maximum number of $\ell$ reached
9: return $\{U, V\}$

V. DISCUSSIONS

In this section we provide useful insights regarding the proposed designs in Section III and IV from the aspects of the required computational complexity, as well as the worst-case CSI error matrices.

A. Worst case CSI error

It is beneficial to obtain the least favorable CSI error matrices, as they provide guidelines for the future channel estimation strategies. For instance, this helps us to choose a channel training sequence that reduces the radius of the CSI error feasible regions in the most destructive directions. Moreover, such knowledge is a necessary step for cutting-set-based methods [49] which aim to reduce the design complexity by iteratively identifying the most destructive error matrices and explicitly incorporating them into the future design steps. In the current setup, the worst-case channel error matrices are identified by maximizing the weighted MSE objective in (35) within their defined feasible region. This is expressed as

$$\max_{\mathcal{W}} \sum \sum_{k \in \mathbb{F}_K} \text{tr} \left( W_i^k E_i^k W_i^k \right),$$

s.t. (50b), \quad (53a)

Due to the uncoupled nature of the error feasible set, and the value of the objective function over $\Delta_{ij}^k$, following (37), the above problem is decomposed as

$$\min_{b_{ij}^k} \left\| c_{ij}^k D_{ij}^k b_{ij}^k \right\|_2^2 - 2 \text{Re} \left\{ b_{ij}^k H D_{ij}^k C_{ij}^k H c_{ij}^k \right\} - c_{ij}^k H E_{ij}^k$$

s.t. $b_{ij}^k H b_{ij}^k \leq \zeta_{ij}^k$, \quad (54b)
where \( \text{Re}\{\cdot\} \) represents the real part of a complex value. Note that the objective in (54a) is a non convex function and cannot be minimized using the usual numerical solvers in the current form. Following the zero duality gap results for the non-convex quadratic programs \([50],[51]\), we focus on the dual function of (54). The corresponding Lagrangian function that the objective in (54a) is a non convex function and cannot be obtained dual variable \( \rho \) and the optimal value of \( b \) that the optimality and the worst case bound on computational complexity are related to different dimensions in the problem structure. Nevertheless, the actual computational load may vary in practice, due to the structure simplifications and depending on the used numerical solver. Furthermore, the overall algorithm complexity also depends on the number of optimization iterations required for convergence. See Subsection VI-A for a study on the convergence behavior, as well as a numerical evaluation of the algorithm computational complexity.

VI. SIMULATION RESULTS

In this section we evaluate the behavior of the studied FD MC system via numerical simulations. In particular, we evaluate the proposed designs in Sections III and IV for various system situations, and under the impact of transceiver

If one of the aforementioned conditions is not satisfied, an infinitely large value of \( b_{ij} \) can be chosen in the negative direction of \( A_{ij} \), if \( A_{ij} \) is not positive semi-definite, or in the direction \( D_{ij}^k H C_{ij}^k c_{ij}^k \) within the null-space of \( A_{ij} \).

Note that the semi-definite presentation in (58b) automatically satisfies \( A_{ij} \geq 0 \), and \( D_{ij}^k H C_{ij}^k c_{ij}^k \in \mathbb{R}\{A_{ij}^k\} \).

thereby complicate the structure of the resulting optimization problem. In this part, we analyze the arithmetic complexity associated with the Algorithm \([52]\). Note that Algorithm \([52]\) is considered as a general framework, containing Algorithm III as a special case, since it takes into account the impacts of hardware distortion jointly with CSI error.

The optimization over \( V, U \) are separately cast as SDP. A general SDP problem is defined as

\[
\min_{z} \quad p^T z, \quad \text{s.t.} \quad z \in \mathbb{R}^n, \quad Y_0 + \sum_{i=1}^{n} z_i Y_i \succeq 0, ||z||_2 \leq q,
\]

(60)

where the fixed matrices \( Y_i \) are symmetric block-diagonal, with \( M \) diagonal blocks of the sizes \( l_m \times l_m \), \( m \in \mathbb{F}_M \), and define the specific problem structure, see \([52]\) Subsection 4.6.3. The arithmetic complexity of obtaining an \( \epsilon \)-solution to the defined problem, i.e., the convergence to the \( \epsilon \)-distance vicinity of the optimum is upper-bounded by

\[
O(1) \left( 1 + \sum_{m=1}^{M} l_m \right)^\frac{3}{2} \left( n^3 + n^2 \sum_{m=1}^{M} l_m^2 + n \sum_{m=1}^{M} l_m^3 \right) \text{digit}(\epsilon),
\]

(61)

where \( \text{digit}(\epsilon) \) is obtained from \([52]\) Subsection 4.1.2, and affected by the required solution precision. The required computation of each step is hence determined by size of the variable space and the corresponding block diagonal matrix structure, which is obtained in the following:

1) Optimization over \( V, T, M \): The size of the variable space is given as \( n = 2K(4 + \sum_{i\in I} d_i N_i) \). Moreover, the block sizes are calculated as \( l_m = 2 + 2K d_i N_i \), \( \forall i \in I \), corresponding to the semi-definite constraint on \( G_i \), and as \( l_m = 2 + 2d_i + 2M_i N_j \), \( \forall i, j \in I \), \( k \in \mathbb{F}_K \), corresponding to the semidefinite constraint on \( F_{i,j}^k \) from (47). The overall number of the blocks is calculated as \( M = 2 + 4K \).

2) Optimization over \( U, T, M \): The size of the variable space is given as \( n = 2K (4 + \sum_{i\in I} d_i M_i) \). The block sizes are calculated as \( l_m = 2 + 2d_i + 2M_i N_j \), \( \forall i, j \in I \), \( k \in \mathbb{F}_K \), corresponding to the semidefinite constraint on \( F_{i,j}^k \) from (47). The overall number of the blocks is calculated as \( M = 4K \).

3) Remarks: The above analysis intends to show how the bounds on computational complexity are related to different dimensions in the problem structure. Nevertheless, the actual computational load may vary in practice, due to the structure simplifications and depending on the used numerical solver. Furthermore, the overall algorithm complexity also depends on the number of optimization iterations required for convergence. See Subsection VI-A for a study on the convergence behavior, as well as a numerical evaluation of the algorithm computational complexity.
inaccuracy and CSI error. Communication channels $H_{ij}$ follow an uncorrelated Rayleigh flat fading model with variance $\rho$. For the self-interference channel we follow the characterization reported in [37]. In this respect we have $H_{ij} \sim \mathcal{CN}(\sqrt{\frac{\rho_{4} K_{R}}{M_{i} + T + K_{R}}}, \sigma_{\text{si}})$ where $\rho_{4}$ represents the self-interference channel strength, $H_{0}$ is a deterministic term, and $K_{R}$ is the Rician coefficient. For each channel realization, the resulting performance is evaluated by employing different design strategies and for various system parameters. The overall system performance is then averaged over 100 channel realizations. Unless otherwise is stated, the following values are used to define our default setup: $K = 4$, $K_{R} = 10$, $M := M_{i} = N_{i} = 2$, $\rho = -20$ dB, $\rho_{4} = 1$, $\sigma_{\text{si}}^{2} := \sigma_{i,k}^{2} = -90$ dB, $P_{\text{max}} := P_{i} = 1$, $d_{i} = 1$, $\kappa = -30$ dB where $\Theta_{i,k} = \kappa I_{M_{i}}$, and $\Theta_{k,i} = \kappa I_{N_{i}}$, and $\zeta_{ij} = -15$ dB, $\forall i, j \in I$, $k \in \mathbb{F}_{K}$.

### A. Algorithm analysis

Due to the alternating structure, the convergence behavior of the proposed algorithms is of interest, both as a verification for algorithm operation as well as an indication of the algorithm efficiency in terms of the required computational effort. In this part, the performance of AltQCP and AltSDP algorithms are studied in terms of the average convergence behavior and computational complexity. Moreover, the impact of the choice of the algorithm initialization is evaluated.

In Fig. 2 the average convergence behavior is depicted for different values of $\kappa$ [dB]. In particular, "Min" and "Avg" curves respectively represent the minimum, and the average value of the algorithm objective at the corresponding optimization step over the choice of 20 random initializations. Moreover, "RSM" represents the right-singular matrix initialization proposed in [40]. Appendix A. It is observed that the algorithms converge, within $10 - 30$ optimization iterations, specially as $\kappa$ is small. Although the global optimality of the final solution can not be verified due to the possibility of local solutions, the numerical experiments suggest that the applied RSM initialization shows a better convergence behavior compared to a random initialization. Moreover, it is observed that a higher transceiver inaccuracy results in a slower convergence and a gap with optimality. This is expected, as larger $\kappa$ leads to a more complex problem structure. Note that the algorithm AltQCP shows a smaller value of objective compared to that of AltSDP for any value of $\kappa$, since the impact of CSI error is not considered in the algorithm objective.

In addition to the algorithm convergence behavior, the required computational complexity is affected by the problem dimension, and the required per-iteration complexity, see Subsection V-B. In Fig. 3 the required computation time (CT) is depicted for different number of antennas, as well as different number of subcarriers. It is observed that the AltSDP results in a significantly higher CT, compared to AltQCP. This is expected as the consideration of CSI error in AltSDP results in a larger problem dimension, and hence higher complexity. Moreover, the obtained closed-form solution expressions in AltQCP result in a more efficient implementation. Nevertheless, the required CT for AltQCP is still higher than the threshold-based low-complexity approaches, see Subsection VI-B1 due to the expanded problem dimension associated with the impact of residual self-interference and inter-carrier leakage.

### B. Performance comparison

In this part we evaluate the performance of the proposed AltSDP and AltQCP algorithms in terms of the resulting worst-case MSE, see Subsection V-A under various system conditions.

1) **Comparison benchmarks:** In order to facilitate a meaningful comparison, we consider popular approaches for the design of FD single-carrier bidirectional systems, or the available designs for other MC systems with simplified assumptions, see Subsection I-A. The following approaches are hence implemented as our evaluation framework:

- **AltSDP:** The AltSDP algorithm proposed in Section IV.
- **AltQCP:** In this algorithm, the impact of the hardware distortions leading to inter-carrier leakage, as well as CSI error are jointly taken into account.

\footnote{The reported CT is obtained using an Intel Core i5-3320M processor with the clock rate of 2.6 GHz and 8 GB of random-access memory (RAM). As our software platform we have used MATLAB 2013a, on a 64-bit operating system.}
• **AltQCP**: The AltQCP algorithm proposed in Section III. The algorithm operates on the simplified assumption that the CSI error does not exist, i.e., $\zeta = 0$, and hence focuses on the impact of hardware distortions.

• **HD**: The AltSDP algorithm is used on an equivalent HD setup, where the communication directions are separated via a time division duplexing (TDD) scheme. The equivalent number of data streams to the FD case is transmitted, and evaluated in terms of the resulting worst-case MSE.

• $\kappa = 0$: The algorithm follows a similar approach as the literature instances where the impact of CSI error is taken into account, e.g., [13], [31]. Nevertheless the impact of hardware distortion, leading to inter-carrier leakage, is ignored.

• **SC**: The optimal single carrier design applied to the defined MC system. Following a similar approach as in [10], [16], [17], [20] for the design of single carrier systems, the impact of CSI error and hardware distortions are taken into account.

Other than the approaches that directly deal with the impact of residual self-interference, e.g., [10], [16], [17], [20], a low complexity design framework is proposed in [11] - [13], by introducing an interference power threshold, denoted as $P_{th}$. In this approach, it is assumed that the self-interference signal can be perfectly subtracted, given the self-interference power is kept below $P_{th}$. In this regard, we evaluate the extended version of [13] on the defined MC setup for three values of $P_{th}$:

- $P_{th} = \infty$: representing a design by assuming a perfect hardware and CSI assumption, and with no limit on the self-interference power. This case corresponds to a linear precoder/decoder design with no consideration on the impact of self-interference.

- $P_{th} = \text{High}$: similar to the previous case, but setting $P_{th} = P_i$. This represents a system with a relatively high dynamic range, i.e., a high tolerable self-interference power. A perfect hardware and CSI is assumed.

- $P_{th} = \text{Low}$: similar to the previous case, but setting $P_{th} = P_i/10$. This represents a system with a relatively low dynamic range, i.e., a small tolerable self-interference power. A perfect hardware and CSI is assumed.

2) **Visualization**: In Figs. 4-8 the average performance of the defined benchmark algorithms in terms of the worst-case (WC) MSE are depicted. The average sum rate behavior of the system is depicted in Fig. 10 (a)-(c).

In Fig. 4 the impact of transceiver inaccuracy is depicted on the resulting WC-MSE. It is observed that the estimation accuracy is degraded as $\kappa$ increases. For the low-complexity algorithms, where the impact of hardware distortion is not considered, the resulting MSE goes to infinity as $\kappa$ increases. Nevertheless, the resulting MSE reaches a saturation point for the distortion-aware algorithms, i.e., AltSDP and AltQCP. This is since for the data streams affected with a large distortion intensity, the decoder matrices are set to zero which limits the resulting MSE to the magnitude of the data symbols. Moreover, the AltSDP method outperforms the other performance benchmarks for all values of $\kappa$. It is worth mentioning that the significant gain of an FD system with low $\kappa$ over the HD counterpart, disappears for a larger levels of hardware distortion where AltSDP and HD result in a close performance.

**Figure 4.** WC MSE vs. transceiver inaccuracy ($\kappa$). As transceiver accuracy degrades, the resulting MSE increases.

**Figure 5.** WC MSE vs. feasible CSI error radius ($\zeta$). The resulting MSE increases as CSI accuracy degrades.

**Figure 6.** WC MSE vs. thermal noise variance ($\sigma^2_n$). Low noise regime results in a low MSE for the proposed algorithms, but degrades the performance of the algorithms with perfect hardware assumption.
In Fig. 7 the impact of the communication channel strength (\( \rho \)) is observed. The resulting MSE decreases as the channel strength increases, but saturates due to the impact of hardware distortion.

In Fig. 8 the impact of the number of antennas (\( M \)) is observed. The resulting MSE decreases as the number of antennas increase.

In Fig. 9 the impact of the number of independent subcarriers (\( K \)) is observed. The resulting MSE increases as \( K \) increases. The performance of SC scheme is optimal for small \( K \), but degrades rapidly as \( K \) increases.

In Fig. 5 the impact of the CSI error is depicted. It is observed that the estimation MSE increases for a larger value of \( \zeta \). For the low-complexity algorithms where the impact of CSI error is not considered, the resulting MSE goes to infinity, as \( \zeta \) increases. Nevertheless, the performance of the AltSDP method saturates by choosing zero decoder matrices, following a similar concept as for Fig. 4. It is observed that the performance of the AltSDP and AltQCP methods deviate as \( \zeta \) increases, however, they obtain a similar performance for a small \( \zeta \). Similar to Fig. 4 a significant gain is observed in comparison to the HD and SC cases, for a system with accurate CSI.

In Fig. 6 the impact of the thermal noise variance is depicted. It is observed that the resulting performance degrades for the distortion-aware algorithms, as the noise variance increases. Nevertheless, we observe a significant performance degradation for the threshold-based algorithms, particularly \( P_{th} = \text{Low} \), in the low noise regime. This is since the imposed interference power threshold tends to reduce the transmit power, which results in a larger decoder matrices in a low-noise regime. This, in turn, results in an increased impact of distortion. Nevertheless, as the noise variance increases, the algorithm chooses decoding matrices with a smaller norm in order to reduce the impact of noise. This also reduces the impact of hardware distortions. Similar to Fig. 4, the proposed AltSDP method outperforms the other comparison benchmarks. It is observed that the performance degradation caused by ignoring the CSI error in AltQCP, or by applying a simplified single carrier design, is significant particularly for a system with a small noise variance.

In Fig. 7 the impact of the communication channel strength is observed on the resulting system performance. It is observed that the MSE decreases in most parts as the communication channel becomes stronger. Nevertheless, the system performance saturates, due to the impact of hardware distortion which increases proportional to the transmit/receive power at each chain. Moreover, the performance of the methods with a perfect hardware/CSI assumption saturates at a higher MSE, due to the impact of the ignored effect. Moreover, the algorithms AltQCP and AltSDP result in an approximately similar performance for a system with a high channel strength. This is since for a high \( \rho \) regime, the impact of thermal noise and CSI error become less significant. As a result the system performance is dominated by the impact of distortion which is amplified due to the higher channel strength.

In Fig. 8 the impact of the number of antennas is observed. As expected, a higher number of antennas results in an increased performance for all of the performance benchmarks. In particular, a higher number of antennas enables the system to better overcome the CSI error, for a fixed \( \zeta \), and also to direct the transmit power in the desired channel and not in the self-interference path.

In Fig. 9 the impact of the number of subcarriers is observed on the resulting MSE. It is observed that a higher number of subcarriers result in a higher error for all benchmark methods. This is expected as a higher number of subcarriers enables a higher number of communication streams, resulting in a lower available per-stream power. The performance of the SC design reaches optimality of a single carrier system, as expected. Nevertheless the performance of the SC scheme deviates from optimality as \( K \) increases, and results in the highest MSE in comparison to the evaluated benchmarks, for \( K \geq 5 \). This is
expected, as higher independent subcarriers represent a channel with a higher frequency selectivity which calls for a specialized MC design.

In Fig. 10 the average sum rate behavior of the system depicted in Fig. 10(a), the impact of hardware inaccuracy is depicted. It is observed that a higher $\kappa$ results in a smaller sum rate. Moreover, the obtained gains via the application of the defined MC design in comparison to the designs with frequency-flat assumption, and via the application of FD setup in comparison to HD setup, are evident for a system with accurate hardware conditions. Conversely, it is observed that a design with consideration of hardware impairments is essential as $\kappa$ increases. In Figs. 10(b) and (c), the opposite impacts of noise level, and the maximum transmit power are observed on the system sum rate. It is observed that the system sum rate increases as noise level decreases, or as the maximum transmit power increases. In both cases, the gain of AltQCP method, in comparison to the methods which ignore the impact of hardware distortions are observed for a high SNR conditions, i.e., for a system with a high transmit power or a low noise level.

**VII. Conclusion**

The application of bi-directional FD communication presents a potential for improving the spectral efficiency. Nevertheless, such systems are limited due to the impact of residual self-interference. This issue becomes more crucial in a multi-carrier system, where the residual self-interference spreads over multiple carriers, due to the impact of hardware distortion. In this work we have presented a modeling and design framework for a FD MIMO OFDM system, taking into account the impact of hardware distortions leading to inter-carrier leakage, as well as the impact of CSI error.

It is observed that the application of a distortion-aware design is essential, as transceiver accuracy degrades, and inter-carrier leakage becomes a dominant factor. Moreover, a significant gain is observed compared to the usual single-carrier approaches, for a channel with frequency selectivity. However, the aforementioned improvements are obtained at the expense of a higher design computational complexity.

![Figure 10](image)

**REFERENCES**

[1] O. Taghizadeh, V. Radhakrishnan, A. C. Cirik, S. Shojae, R. Mathar, and L. Lampe, “Linear precoder and decoder design for bidirectional Full-Duplex MIMO OFDM systems,” in *PIMRC’17 (Submitted)*, Montreal, Canada.

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

[3] 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.

[4] Y. Hua, P. Liang, Y. Ma, A. Cirk, and G. Qian, “A method for broadband full-duplex MIMO radio,” *IEEE Signal Processing Letters*, vol. 19, Dec. 2011.

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

[6] M. Jain, J. I. Choi, T. Kim, D. Bharadia, K. Srinivasan, S. Seth, P. Levis, S. Katti, and P. Sinha, “Practical, real-time, full duplex wireless,” in *Proceedings of 17th Annual International Conference on Mobile Computing and Networking (MobiCom)*, Las Vegas, NV, Sep. 2011.

[7] J.-H. Lee, I. Sohn, and Y.-H. Kim, “Analog cancelation schemes for full-duplex mimo ofdm systems,” *AEU-International Journal of Electronics and Communications*, vol. 70, no. 3, pp. 272–277, 2016.

[8] D. Kim, H. Lee, and D. Hong, “A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers,” *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2017–2046, Fourthquarter 2015.

[9] A. Sabharwal, P. Schniter, D. Guo, D. Bliss, S. Rangarajan, and R. Wichman, “In-band full-duplex wireless: Challenges and opportunities,” *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 9, pp. 1637–1652, Sept 2014.

[10] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, “Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range,” *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3702–3713, July 2012.

[11] S. Huberman and T. Le-Ngoc, “Self-interference-threshold-based MIMO full-duplex precoding,” *IEEE Transactions on Vehicular Technology*, vol. 64, no. 8, pp. 3803–3807, Aug 2015.

[12] ———, “Self-interference pricing-based MIMO full-duplex precoding,” *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 549–552, Dec 2014.

[13] J. Zhang, O. Taghizadeh, J. Luo, and M. Haardt, “Full duplex wireless communications with partial interference cancellation,” *Proceedings of
J. Zhang, O. Taghizadeh, and M. Haardt, “Robust transmit beamforming design for full-duplex point-to-point MIMO systems,” Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS), Aug. 2013.

A. C. Cirik, R. Wang, and Y. Hua, “Weighted-sum-rate maximization for bi-directional full-duplex MIMO systems,” in 2013 Asilomar Conference on Signals, Systems and Computers, Nov 2013, pp. 1632–1636.

A. C. Cirik, J. Zhang, M. Haardt, and Y. Hua, “Sum-rate maximization for bi-directional full-duplex MIMO systems under multiple linear constraints,” in Proc. 15th Int. Workshop Signal Processing Advances in Wireless Communications (SPAWC 2014), Toronto, Canada, Jun. 2014.

A. Cirik, R. Wang, Y. Hua, and M. Latva-aho, “Weighted sum-rate maximization for full-duplex MIMO interference channels,” Communications, IEEE Transactions on, vol. 63, March 2015.

O. Taghizadeh and R. Mathar, “Worst-Case robust sum rate maximization for Full-Duplex Bi-Directional MIMO systems under channel knowledge uncertainty,” in IEEE ICC 2017 Signal Processing for Communications Symposium (ICC’17 SP’C), Paris, France, May 2017, pp. 5950–5956.

A. C. Cirik, R. Wang, Y. Rong, and Y. Hua, “MSE based transceiver designs for bi-directional full-duplex MIMO systems,” in 2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), June 2014, pp. 384–388.

A. C. Cirik, O. Taghizadeh, L. Lampe, R. Mathar, and Y. Hua, “Sum-power minimization under rate constraints in full-duplex MIMO systems,” VTC 2016, Montreal, Canada, 2016.

A. Cirik, S. Biswas, S. Vuppala, and T. Ratnarajah, “Beamforming design for full-duplex MIMO interference channels: QoS and energy-efficiency considerations,” IEEE Transactions on Communications, vol. PP, no. 99, pp. 1–1, 2016.

O. Taghizadeh, A. C. Cirik, R. Mathar, and L. Lampe, “Sum power minimization for TDD-Enabled Full-Duplex Bi-Directional MIMO systems under channel uncertainty,” in European Wireless 2017 (EW2017), Dresden, Germany, May 2017, pp. 412–417.

T. Riihonen and R. Wichman, “Analog and digital self-interference cancellation in full-duplex MIMO-OFDM transceivers with limited resolution in A/D conversion,” in 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Nov 2012, pp. 45–49.

T. Riihonen, K. Haneda, S. Werner, and R. Wichman, “SINR analysis of full-duplex OFDM repeaters,” in 2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, Sept 2009, pp. 3169–3173.

M. Mokhtar, N. Al-Dhahir, and R. Hamila, “OFDM full-duplex DF relaying under I/Q imbalance and loopback self-interference,” IEEE Transactions on Vehicular Technology, vol. 65, no. 8, pp. 6737–6741, Aug 2016.

N. Li, Y. Li, M. Peng, and W. Wang, “Resource allocation in multi-carrier full-duplex amplify-and-forward relaying networks,” in 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), May 2016, pp. 1–5.

Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, “Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems,” IEEE Transactions on Communications, vol. PP, no. 99, pp. 1–1, 2017.

M. Al-Imari, M. Ghoraiishi, and P. Xiao, “Radio resource allocation for full-duplex multicarrier wireless systems,” in 2015 International Symposium on Wireless Communication Systems (ISWCS), Aug 2015, pp. 571–575.

A. C. Cirik, K. Rikkinen, Y. Rong, and T. Ratnarajah, “A subcarrier and power allocation algorithm for ofdma full-duplex systems,” in 2015 European Conference on Networks and Communications (EuCNC), June 2015, pp. 11–15.

W. Li, J. Lilleberg, and K. Rikkinen, “On rate region analysis of half- and full-duplex OFDM communication links,” IEEE Journal on Selected Areas in Communications, vol. 32, no. 9, pp. 1688–1698, Sept 2014.

D. W. K. Ng, E. S. Lo, and R. Schober, “Dynamic resource allocation in MIMO-OFDMA systems with full-duplex and hybrid relaying,” IEEE Transactions on Communications, vol. 60, no. 5, pp. 1291–1304, 2012.

G. Santella and F. Mazzenga, “A hybrid analytical-simulation procedure for performance evaluation in M-QAM-OFDM schemes in presence of nonlinear distortions,” IEEE Trans. Veh. Technol., vol. 47, pp. 142–151, 1998.

W. Namgoong, “Modeling and analysis of nonlinearities and mismatches in AC-coupled direct-conversion receiver,” IEEE Trans. Wireless Commun., vol. 47, pp. 163–173, Jan. 2005.

H. Suzuki, T. V. A. Tran, I. B. Collings, G. Daniels, and M. Hedley, “Transmitter noise effect on the performance of a MIMO-OFDM hardware implementation achieving improved coverage,” IEEE J. Sel. Areas in Communication, vol. 26, pp. 867–876, Aug. 2008.

S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, “Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design,” IEEE Transactions on Wireless Communications, vol. 7, no. 12, pp. 4792–4799, December 2008.

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.

M. Duarte, C. Dick, and A. Sabharwal, “Experiment-driven characterization of full-duplex wireless systems,” IEEE Transactions on Wireless Communications, December 2012.

U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, “Wireless backhauling of 5G small cells: challenges and solution approaches,” IEEE Wireless Communications, vol. 22, no. 5, pp. 22–31, October 2015.

D. Bertesekas, “Nonlinear programming. athena scientific,” Belmont, Massachusetts, 1999.

H. Shen, B. Li, M. Tao, and X. Wang, “MSE-based transceiver designs for the MIMO interference channel,” IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3480–3489, November 2010.

J. Jose, N. Prasad, M. Khojastepour, and S. Rangarajan, “On robust weighted-sum rate maximization in MIMO interference networks,” in 2011 IEEE International Conference on Communications (ICC), June 2011, pp. 1–6.

J. Wang and D. P. Palomar, “Worst-case robust MIMO transmission with imperfect channel knowledge,” IEEE Transactions on Signal Processing, vol. 57, no. 8, pp. 3086–3100, 2009.

R. Lorenz and S. P. Boyd, “Robust minimum variance beamforming,” IEEE Transactions on Signal Processing, vol. 53, May 2005.

A. Pascual, D. P. Palomar, A. I. Neira, and M. A. Lagunas, “A robust maximin approach for MIMO communications with partial channel state information based on convex optimization,” IEEE Transactions on Signal Processing, vol. 54, Jan. 2006.

S. Petersen and M. Pedersen, Matrix Cookbook, 2008.

Y. C. Eldar and N. Merhav, “A competitive minimax approach to robust estimation of random parameters,” IEEE Transactions on Signal Processing, vol. 52, no. 7, pp. 1931–1946, July 2004.

M. V. Khebennikov and P. S. Scherbakov, “Petersens lemmata on matrix uncertainty and its generalizations,” Automation and Remote Control, vol. 69, no. 11, pp. 1932–1945, 2008.

L. Vendenderhge, S. Boyd, and S.-P. Wu, “Determinant maximization with linear matrix inequality constraints,” SIAM Journal on matrix analysis and applications, vol. 19, no. 2, pp. 499–533, 1998.

A. Mutapcic and S. Boyd, “Cutting-set methods for robust convex optimization with pessimization oracles,” Optimization Methods & Software, vol. 24, no. 3, pp. 381–406, 2009.

X. Zheng, X. Sun, D. Li, and Y. Xu, “On zero duality gap in nonconvex
quadratic programming problems,” *Journal of Global Optimization*, vol. 52, no. 2, pp. 229–242, 2012.

[51] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[52] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Siam, 2001, vol. 2.