

Hardware Impairments Aware Transceiver Design for Bidirectional Full-Duplex MIMO OFDM Systems

Omid Taghizadeh *Student Member, IEEE*, Vimal Radhakrishnan, Ali Cagatay Cirik, *Member, IEEE*, Rudolf Mathar, *Senior Member, IEEE*, Lutz Lampe *Senior Member, IEEE*

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 as well as the channel state information error are taken into account. In the first step, we transform the available time-domain characterization of the hardware distortions for FD MIMO transceivers to the frequency domain, via a linear Fourier transformation. As a result, the explicit impact of hardware inaccuracies on the residual self-interference (RSI) and inter-carrier leakage (ICL) is formulated in relation to the intended transmit/received signals. Afterwards, linear precoding and decoding designs are proposed to enhance the system performance following the minimum-mean-squared-error (MMSE) and sum rate maximization strategies, assuming the availability of perfect or erroneous channel state information (CSI). The proposed designs are based on the application of alternating optimization over the system parameters, leading to a necessary convergence. Numerical results indicate that the application of a distortion-aware design is essential for a system with a high hardware distortion, or for a system with a low thermal noise variance.

Keywords—Full-duplex, MIMO, OFDM, hardware impairments, MMSE.

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 (SI) from their own transmitter. Recently, specialized self-interference cancellation (SIC) techniques, e.g., [3]–[6], 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., [7]–[10]. A common idea of such SIC techniques is to subtract the dominant part of the SI signal, e.g., a line-of-sight (LOS) SI 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 CSI in the SI path, due to noise and limited channel coherence time. In this regard, inaccuracy of the analog hardware elements used in subtracting the dominant SI 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 SI path in environments with a small channel coherence time, see [11, 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 SI paths¹.

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 [12], 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 [13], [14] by defining a price/threshold for the SI 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 SI path [15]. Consequently, the consideration of CSI and transceiver error in an FD bidirectional system is further studied in [16], [17] by maximizing the system sum rate, in [18] by minimizing the sum mean-squared-error (MSE), and in [19], [20] 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

O. Taghizadeh, V. Radhakrishnan, and R. Mathar are with the Institute for Theoretical Information Technology, RWTH Aachen University, Aachen, 52074, Germany (email: {taghizadeh, radhakrishnan, mathar}@ti.rwth-aachen.de).

A. C. Cirik and L. Lampe are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (email: {cirik, lampe}@ece.ubc.ca).

Part of this paper has been presented at the 28th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC'17) [1].

¹Since the object is moving rapidly, the reflective paths are more difficult to be accurately estimated.

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 ICL due to the impact of hardware distortions on the strong SI channel². This, in particular, calls for a proper modeling of the ICL as a result of non-linear hardware distortions for FD transceivers. And finally, the channel frequency selectivity shall be opportunistically exploited, by means of a joint design of the linear transmit and receive strategies at all subcarriers, in order to enhance the system performance.

A. Related works on FD MC systems

In the early work by Riihonen *et al.* [21], the performance of a combined analog/digital SIC scheme is evaluated for an FD OFDM transceiver, taking into account the impact of hardware distortions, e.g., limited analog-to-digital convertor (ADC) accuracy. The problem of resource allocation and performance analysis for FD MC communication systems is then addressed in [22]–[28], however, assuming a single antenna transceiver. Specifically, an FD MC system is studied in [22]–[24] in the context of FD relaying, in [26], [27] and [25] in the context of FD cellular systems with non-orthogonal multiple access (NOMA) capability, and in [28] 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 [29], 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 RSI signal is modeled as a purely linear system. As a result, the impacts of the hardware distortions leading to ICL, as observed in [21], 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 ICL are taken into account. Our main contributions, together with the paper organization are summarized as follows:

- In the seminal work by Day *et al.* [12], an FD MIMO transceiver is modeled considering the impacts of hardware distortions in transmit/receiver chains, which is then extensively used for the purpose of FD system design and performance analysis, e.g., [16], [19], [30]–[34]. In the first step, we extend the available time-domain characterization of hardware distortions into an FD MIMO OFDM setup via a linear discrete Fourier transformation. The obtained frequency-domain characterization reveals the statistics of the RSI and ICL, in relation to the intended transmit/receive signals at each subcarrier. Please note that this is in contrast to the available prior works on FD MC systems [22]–[29],

²For instance, a high-power transmission in one of the subcarriers will result in a higher RSI in all of the sub-channels due to, e.g., a higher quantization and power amplifier noise levels.

where ICL is neglected and RSI signal is modeled via a purely linear system.

- Building on the obtained characterization, 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, we extend the studied system to an asymmetric OFDM FD bidirectional setup, where an FD transceiver with a large antenna array simultaneously communicates with multiple single-antenna FD transceivers. The extended scenario is particularly relevant, both due to the recent advances in building FD massive MIMO transceivers [36] as well as the signified impact of hardware distortions due to the lower per-element cost (e.g., low resolution quantization [37]). An algorithm for joint power and subcarrier allocation is then proposed, following the successive inner approximation (SIA) framework [38], with a guaranteed convergence to a solution satisfying Karush–Kuhn–Tucker (KKT) conditions.
- In Section V the proposed design in Section III is extended by also taking into account the impact of CSI error. In particular, 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. Moreover, a methodology to obtain the most destructive CSI error matrices is proposed. This is done by converting the resulting non-convex quadratic problem into a convex program, in order to facilitate worst-case performance analysis under CSI error.

Numerical simulations show that the application of a distortion-aware design is essential, as transceiver accuracy degrades, and ICL becomes a dominant factor.

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 \mathbf{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. $\text{diag}(\cdot)$ returns a diagonal matrix by putting the off-diagonal elements to zero. $[\mathbf{A}_i]_{i=1, \dots, K}$ denotes a tall matrix, obtained by stacking the matrices \mathbf{A}_i , $i = 1, \dots, K$. $\mathcal{R}\{\mathbf{A}\}$ represents the range (column space) of the matrix \mathbf{A} . The set \mathbb{F}_K is defined as $\{1, \dots, K\}$. The set of real, positive real, and complex numbers are respectively denoted as \mathbb{R} , \mathbb{R}^+ , \mathbb{C} .

Table I. USED SYMBOLS

k, i, l	index of a subcarrier, communication direction, and a transmit/receive chain
$\mathbb{I}, \mathbb{V}, \mathbb{U}$	set of comm. directions, precoder and decoder matrices
N_i, M_i, d_i	number of transmit and receive antennas and data streams
$\mathbf{s}_i^k(\tilde{\mathbf{s}}_i^k)$	transmitted (estimated) data symbol
$\mathbf{U}_i^k(\mathbf{V}_i^k)$	linear decoder (precoder) matrix
$\mathbf{y}_i^k(\tilde{\mathbf{y}}_i^k)$	received signal before (after) SI cancellation
$\mathbf{H}_{ij}^k, \hat{\mathbf{H}}_{ij}^k$	the exact, and estimated CSI matrix
$\Delta_{ij}^k, \mathbb{D}_{ij}^k$	CSI error, and the set of feasible CSI errors
$\zeta_{ij}^k, \mathbf{D}_{ij}^k$	radius and shaping matrix for the feasible CSI error region
$\mathbf{e}_{r,i}^k(\mathbf{e}_{t,i}^k)$	receiver (transmitter) distortion over the subcarrier k
$\Theta_{rx,i}(\Theta_{tx,i})$	diagonal matrix of receive (transmit) distortion coefficients
$\mathbf{v}_i^k, \Sigma_i^k$	collective residual SI plus noise signal, and its covariance
$\mathbf{n}_i^k, \sigma_{t,i}^2$	additive thermal noise and its variance
$\mathbf{u}_i^k(\mathbf{v}_i^k)$	undistorted received (transmitted) signal
\mathbf{x}_i^k, P_i	transmit signal, and the maximum transmit power

II. SYSTEM MODEL

A bidirectional OFDM communication between two MIMO FD transceivers is considered. Each communication direction is associated with N_i transmit and M_i 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 $k \in \mathbb{F}_K$ is denoted as $\mathbf{H}_{ii}^k \in \mathbb{C}^{M_i \times N_i}$ where K is the number of subcarriers. The interference channel from i to j -th communication direction is denoted as $\mathbf{H}_{ji}^k \in \mathbb{C}^{M_j \times N_i}$. All channels are quasi-static³, and frequency-flat in each subcarrier.

The transmitted signal in the direction i , subcarrier k is formulated as

$$\mathbf{x}_i^k = \underbrace{\mathbf{V}_i^k \mathbf{s}_i^k}_{=: \mathbf{v}_i^k} + \mathbf{e}_{t,i}^k, \quad \sum_{k \in \mathbb{F}_K} \mathbb{E} \{ \|\mathbf{x}_i^k\|_2^2 \} \leq P_i, \quad (1)$$

where $\mathbf{s}_i^k \in \mathbb{C}^{d_i}$ and $\mathbf{V}_i^k \in \mathbb{C}^{N_i \times d_i}$ respectively represent the vector of the data symbols and the transmit precoding matrix, and $P_i \in \mathbb{R}^+$ imposes the maximum affordable transmit power constraint. The number of the data streams in each subcarrier and in direction i is denoted as d_i , and $\mathbb{E} \{ \mathbf{s}_i^k \mathbf{s}_i^k H \} = \mathbf{I}_{d_i}$. Moreover, $\mathbf{v}_i^k \in \mathbb{C}^{N_i}$ represents the desired signal to be transmitted, where $\mathbf{e}_{t,i}^k$ 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

$$\mathbf{y}_i^k = \underbrace{\mathbf{H}_{ii}^k \mathbf{x}_i^k + \mathbf{H}_{ij}^k \mathbf{x}_j^k}_{=: \mathbf{u}_i^k} + \mathbf{n}_i^k + \mathbf{e}_{r,i}^k, \quad (2)$$

where $\mathbf{n}_i^k \sim \mathcal{CN}(\mathbf{0}_{M_i}, \sigma_{t,i}^2 \mathbf{I}_{M_i})$ is the additive thermal noise. Similar to the transmit signal model, $\mathbf{e}_{r,i}^k$ represents the receiver distortion and models the inaccuracies of the receive chain elements. The *known*, i.e., distortion-free, part of the SI is then subtracted from the received signal, employing an SIC scheme. This is formulated as

$$\tilde{\mathbf{y}}_i^k := \mathbf{y}_i^k - \mathbf{H}_{ij}^k \mathbf{V}_j^k \mathbf{s}_j^k = \mathbf{H}_{ii}^k \mathbf{V}_i^k \mathbf{s}_i^k + \mathbf{v}_i^k, \quad (3)$$

³It indicates that the channel is constant in each communication frame, but may vary from one frame to another frame.

where $\tilde{\mathbf{y}}_i^k$ is the received signal in direction i and subcarrier k , after SIC. Moreover, the aggregate interference-plus-noise term is denoted as $\mathbf{v}_i^k \in \mathbb{C}^{M_i}$, where

$$\mathbf{v}_i^k = \mathbf{H}_{ij}^k \mathbf{e}_{t,j}^k + \mathbf{H}_{ii}^k \mathbf{e}_{t,i}^k + \mathbf{e}_{r,i}^k + \mathbf{n}_i^k, \quad j \neq i. \quad (4)$$

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

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

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

A. Limited dynamic range in an FD OFDM system

In the seminal work by Day *et al.* [12], a model for the operation of an FD MIMO transceiver is given, relying on the experimental results on the impact of hardware distortions [39]–[42]. In this regard, the inaccuracy of the transmit chain elements, e.g., DAC error, PA and oscillator phase noise, are jointly modeled for each antenna as an additive distortion, and written as $x_l(t) = v_l(t) + e_{t,l}(t)$, see Fig. 1, such that

$$e_{t,l}(t) \sim \mathcal{CN}\left(0, \kappa_l \mathbb{E} \{ |v_l(t)|^2 \} \right), \quad e_{t,l}(t) \perp v_l(t), \quad (6)$$

$$e_{t,l}(t) \perp e_{t,l'}(t), \quad e_{t,l}(t) \perp e_{t,l}(t'), \quad l \neq l' \in \mathbb{L}_T, \quad t \neq t', \quad (7)$$

please see [12, Section II. B,C], [30, Section II. C,D], [16], [19], [31]–[33] for a similar distortion characterization for FD transceivers⁴. In the above arguments, t denotes the instance of time, and v_l , x_l , and $e_{t,l} \in \mathbb{C}$ are respectively the baseband time-domain representation of the intended transmit signal, the actual transmit signal, and the additive transmit distortion at the l -th transmit chain. The set \mathbb{L}_T represents the set of all transmit chains. Moreover, $\kappa_l \in \mathbb{R}^+$ 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 effects of the inaccurate hardware elements, i.e., ADC error, AGC and oscillator phase noise, are presented as additive distortion terms and written as $y_l(t) = u_l(t) + e_{r,l}(t)$ such that

$$e_{r,l}(t) \sim \mathcal{CN}\left(0, \beta_l \mathbb{E} \{ |u_l(t)|^2 \} \right), \quad e_{r,l}(t) \perp u_l(t), \quad (8)$$

$$e_{r,l}(t) \perp e_{r,l'}(t), \quad e_{r,l}(t) \perp e_{r,l}(t'), \quad l \neq l' \in \mathbb{L}_R, \quad t \neq t', \quad (9)$$

where u_l , $e_{r,l}$, and $y_l \in \mathbb{C}$ are respectively the baseband representation of the intended (distortion-free) received signal, additive receive distortion, and the received signal from the l -th receive antenna. The set \mathbb{L}_R represents the set of all receive chains. Similar to the transmit chain characterization, $\beta_l \in \mathbb{R}^+$ is the distortion coefficient for the l -th receive chain, see Fig. 1. The frequency domain representation of the sampled

⁴It is worth mentioning that the accuracy of the above-mentioned modeling varies 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 (iii) and (iv) also hold for an advanced implementation of an 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 RSI via advanced signal processing, see [4, Subsection 3.2]. However, the linear dependence of the remaining distortion signal variance to the desired signal strength varies for different SIC implementations, and should be estimated separately for each transceiver.

time domain signal is obtained as

$$x_l^k = \frac{1}{\sqrt{K}} \sum_{m=0}^{K-1} x_l(mT_s) e^{-j\frac{2\pi mk}{K}} = \frac{1}{\sqrt{K}} \sum_{m=0}^{K-1} v_l(mT_s) e^{-j\frac{2\pi mk}{K}} + \frac{1}{\sqrt{K}} \sum_{m=0}^{K-1} e_{t,l}(mT_s) e^{-j\frac{2\pi mk}{K}}, \quad (10)$$

$\underbrace{\hspace{10em}}_{=:v_l^k} \qquad \underbrace{\hspace{10em}}_{=:e_{t,l}^k}$

$$y_l^k = \frac{1}{\sqrt{K}} \sum_{m=0}^{K-1} y_l(mT_s) e^{-j\frac{2\pi mk}{K}} = \frac{1}{\sqrt{K}} \sum_{m=0}^{K-1} u_l(mT_s) e^{-j\frac{2\pi mk}{K}} + \frac{1}{\sqrt{K}} \sum_{m=0}^{K-1} e_{r,l}(mT_s) e^{-j\frac{2\pi mk}{K}}, \quad (11)$$

$\underbrace{\hspace{10em}}_{=:u_l^k} \qquad \underbrace{\hspace{10em}}_{=:e_{r,l}^k}$

where T_s is the sampling time, and KT_s is the OFDM block duration prior to the cyclic extension, see [43] for a detailed discussion on OFDM technology.

Lemma II.1. *The impact of hardware distortions in the frequency domain is characterized as*

$$e_{t,l}^k \sim \mathcal{CN}\left(0, \frac{\kappa_l}{K} \sum_{m \in \mathbb{F}_K} \mathbb{E}\{|v_l^m|^2\}\right), \quad e_{r,l}^k \perp v_l^k, \quad e_{r,l}^k \perp e_{t,l}^k, \quad (12)$$

$$e_{r,l}^k \sim \mathcal{CN}\left(0, \frac{\beta_l}{K} \sum_{m \in \mathbb{F}_K} \mathbb{E}\{|u_l^m|^2\}\right), \quad e_{r,l}^k \perp u_l^k, \quad e_{r,l}^k \perp e_{t,l}^k, \quad (13)$$

transforming the statistical independence, as well as the proportional variance properties from the time domain.

Proof: See the Appendix. \blacksquare

The above lemma indicates that the distortion signal variance at each subcarrier, relates to the total distortion power at the corresponding chain, indicating the impact of ICL. This can be interpreted as a variance-dependent thermal noise, where the temporal independence of signal samples results in a flat power spectral density over the active communication bandwidth. In this part 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 relevant in practice since it enables the design algorithms to reduce communication task, e.g., transmit power, on the chains with noisier elements. Following Lemma II.1, the statistics of the distortion terms, introduced in (1), (2) can be inferred as

$$e_{t,i}^k \sim \mathcal{CN}\left(\mathbf{0}_{N_i}, \Theta_{\text{tx},i} \sum_{k \in \mathbb{F}_K} \text{diag}\left(\mathbb{E}\{\mathbf{v}_i^k \mathbf{v}_i^{kH}\}\right)\right), \quad (14)$$

$$e_{r,i}^k \sim \mathcal{CN}\left(\mathbf{0}_{M_i}, \Theta_{\text{rx},i} \sum_{k \in \mathbb{F}_K} \text{diag}\left(\mathbb{E}\{\mathbf{u}_i^k \mathbf{u}_i^{kH}\}\right)\right), \quad (15)$$

where $\Theta_{\text{tx},i} \in \mathbb{R}^{N_i \times N_i}$ ($\Theta_{\text{rx},i} \in \mathbb{R}^{M_i \times M_i}$) is a diagonal matrix including distortion coefficients κ_l/K (β_l/K) for the

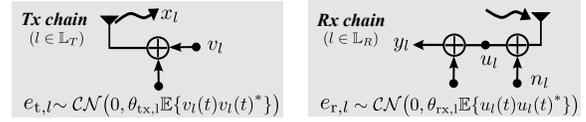


Figure 1. Limited dynamic range is modeled by injecting additive distortion terms at each transmit or receive chain. $e_{t,l}$ and $e_{r,l}$ denote the distortion terms, and n_l represents the additive thermal noise.

corresponding chains⁵.

Via the application of (14)-(15) on (4), the covariance of the received collective interference-plus-noise signal is obtained as

$$\begin{aligned} \Sigma_i^k &:= \mathbb{E}\{\mathbf{v}_i^k \mathbf{v}_i^{kH}\} \\ &\approx \sum_{j \in \mathbb{I}} \mathbf{H}_{ij}^k \Theta_{\text{tx},j} \text{diag}\left(\sum_{l \in \mathbb{F}_K} \mathbf{V}_j^l \mathbf{V}_j^{lH}\right) \mathbf{H}_{ij}^{kH} + \sigma_{i,k}^2 \mathbf{I}_{M_i} \\ &\quad + \Theta_{\text{rx},i} \text{diag}\left(\sum_{l \in \mathbb{F}_K} \left(\sigma_{i,l}^2 \mathbf{I}_{M_i} + \sum_{j \in \mathbb{I}} \mathbf{H}_{ij}^l \mathbf{V}_j^l \mathbf{V}_j^{lH} \mathbf{H}_{ij}^{lH}\right)\right), \quad (16) \end{aligned}$$

where $\Sigma_i^k \in \mathbb{C}^{M_i \times M_i}$ is obtained considering $0 \leq \beta_l \ll 1$, $0 \leq \kappa_l \ll 1$, and hence ignoring the terms containing higher orders of the distortion coefficients in (16).

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 [44], where an adequately long training sequence can be applied, see [12, Subsection III.A]. The impact of the CSI inaccuracy is later addressed in Section V.
- As expected, the role of the distortion signals on the RSI, including the resulting ICL, is evident from (16). It is the main goal of the remaining parts of this chapter to incorporate and evaluate this impact on the design of the defined MC system⁶.

III. LINEAR TRANSCIEVER 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 MSE matrix of the defined system is calculated as

$$\begin{aligned} \mathbf{E}_i^k &:= \mathbb{E}\left\{\left(\hat{\mathbf{s}}_i^k - \mathbf{s}_i^k\right)\left(\hat{\mathbf{s}}_i^k - \mathbf{s}_i^k\right)^H\right\} \\ &= \left(\mathbf{U}_i^k H \mathbf{H}_{ii}^k \mathbf{V}_i^k - \mathbf{I}_{d_i}\right)\left(\mathbf{U}_i^k H \mathbf{H}_{ii}^k \mathbf{V}_i^k - \mathbf{I}_{d_i}\right)^H \\ &\quad + \mathbf{U}_i^k H \Sigma_i^k \mathbf{U}_i^k, \quad (17) \end{aligned}$$

⁵A simpler mathematical presentation can be obtained by assuming the same transceiver accuracy over all antennas, similar to [12], [30]. In such a case, the defined diagonal matrices can be replaced by a scalar.

⁶Please note that the obtained modeling and design framework for the OFDM systems can be similarly generalized to any MC technology with orthogonal waveforms, i.e., with zero intrinsic interference. For a number of the recently developed waveforms, e.g., the filter banked multi-carrier (FBMC) scheme where the adjacent waveforms are not orthogonal [61], the main insight of this work still holds, i.e., the non-linear hardware distortions lead to RSI and ICL, which is detrimental to the performance of FD systems. However, the exact characterization and design of the transmit (receive) strategies require non-trivial extensions.

where Σ_i^k is given in (16). 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

$$\begin{aligned} \min_{\mathbb{V}, \mathbb{U}} \quad & \sum_{i \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} \text{tr}(\mathbf{S}_i^k \mathbf{E}_i^k) \\ \text{s.t.} \quad & \text{tr} \left((\mathbf{I}_{N_i} + K \Theta_{\text{rx},i}) \sum_{l \in \mathbb{F}_K} \mathbf{V}_i^l \mathbf{V}_i^{lH} \right) \leq P_i, \quad \forall i \in \mathbb{I}, \end{aligned} \quad (18a)$$

$$(18b)$$

where $\mathbb{X} := \{\mathbf{X}_i^k, \forall i \in \mathbb{I}, \forall k \in \mathbb{F}_K\}$, with $\mathbb{X} \in \{\mathbb{U}, \mathbb{V}\}$, and (18b) 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 (18) is not a jointly convex problem. Nevertheless, it holds a QCP structure separately over the sets \mathbb{V} and \mathbb{U} , in each case when other variables are fixed. In this regard, the objective (18a) can be decomposed over \mathbb{U} for different communication directions, and for different subcarriers. The optimal MMSE receive filter can be hence calculated in closed form as

$$\mathbf{U}_{i,\text{mmse}}^k = \left(\Sigma_i^k + \mathbf{H}_{ii}^k \mathbf{V}_i^k \mathbf{V}_i^{kH} \mathbf{H}_{ii}^{kH} \right)^{-1} \mathbf{H}_{ii}^k \mathbf{V}_i^k. \quad (19)$$

Nevertheless, the defined problem is coupled over \mathbf{V}_i^k , due to the impact of ICL, as well as the power constraint (18b). The Lagrangian function, corresponding to the optimization (18) over \mathbb{V} is expressed as

$$\mathcal{L}(\mathbb{V}, \boldsymbol{\nu}) := \sum_{i \in \mathbb{I}} \left(\nu_i P_i(\mathbb{V}) + \sum_{k \in \mathbb{F}_K} \text{tr}(\mathbf{S}_i^k \mathbf{E}_i^k) \right), \quad (20)$$

$$P_i(\mathbb{V}) := -P_i + \text{tr} \left((\mathbf{I}_{N_i} + K \Theta_{\text{rx},i}) \sum_{l \in \mathbb{F}_K} \mathbf{V}_i^l \mathbf{V}_i^{lH} \right), \quad (21)$$

where $\boldsymbol{\nu} := \{\nu_i, i \in \mathbb{I}\}$ is the set of dual variables. The dual function, corresponding to the above Lagrangian is defined as

$$\mathcal{F}(\boldsymbol{\nu}) := \min_{\mathbb{V}} \mathcal{L}(\mathbb{V}, \boldsymbol{\nu}) \quad (22)$$

where the optimal \mathbf{V}_i^k is obtained as

$$\begin{aligned} \mathbf{V}_i^{k*} = & \left(\mathbf{J}_i^k + \nu_i (\mathbf{I}_{N_i} + K \Theta_{\text{rx},i}) + \mathbf{H}_{ii}^{kH} \mathbf{U}_i^k \mathbf{S}_i^k \mathbf{U}_i^{kH} \mathbf{H}_{ii}^k \right)^{-1} \\ & \times \mathbf{H}_{ii}^{kH} \mathbf{U}_i^k \mathbf{S}_i^k, \end{aligned} \quad (23)$$

and

$$\begin{aligned} \mathbf{J}_i^k := & \sum_{l \in \mathbb{F}_K} \sum_{j \in \mathbb{I}} \left(\mathbf{H}_{ji}^{kH} \text{diag} \left(\mathbf{U}_j^l \mathbf{S}_j^l \mathbf{U}_j^{lH} \Theta_{\text{rx},j} \right) \mathbf{H}_{ji}^k \right. \\ & \left. + \text{diag} \left(\mathbf{H}_{ji}^{lH} \mathbf{U}_j^l \mathbf{S}_j^l \mathbf{U}_j^{lH} \mathbf{H}_{ji}^l \Theta_{\text{rx},i} \right) \right). \end{aligned} \quad (24)$$

Due to the convexity of the original problem (18) over \mathbb{V} , the defined dual problem is a concave function over $\boldsymbol{\nu}$, with $\mathcal{P}_i(\mathbb{V})$ as a subgradient, see [45, Eq. (6.1)]. As a result, the optimal $\boldsymbol{\nu}$ is obtained from the maximization

$$\boldsymbol{\nu}^* = \underset{\boldsymbol{\nu} \geq 0}{\text{argmax}} \mathcal{F}(\boldsymbol{\nu}), \quad (25)$$

following a standard subgradient update, [45, Subsection 6.3.1].

Utilizing the proposed optimization framework, the alternating optimization over \mathbb{V} and \mathbb{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 (18a) 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: $\mathbb{V} \leftarrow$ right singular matrix initialization, see [46, Appendix A]
 - 3: $\mathbb{U} \leftarrow$ solve (19)
 - 4: **repeat**
 - 5: $\ell \leftarrow \ell + 1$
 - 6: $\mathbb{V} \leftarrow$ solve (23) or QCP (18), with fixed \mathbb{U}
 - 7: $\mathbb{U} \leftarrow$ solve (19) or QCP (18) with fixed \mathbb{V}
 - 8: **until** a stable point, or maximum number of ℓ reached
 - 9: **return** $\{\mathbb{U}, \mathbb{V}\}$
-

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

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

$$I_i^k = \log_2 \left| \mathbf{I}_{d_i} + \mathbf{V}_i^{kH} \mathbf{H}_{ii}^{kH} (\Sigma_i^k)^{-1} \mathbf{H}_{ii}^k \mathbf{V}_i^k \right|, \quad (26)$$

where Σ_i^k is defined in (16). The sum rate maximization problem can be hence presented as

$$\max_{\mathbb{V}} \sum_{i \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} \omega_i I_i^k, \quad \text{s.t. (18b)}. \quad (27)$$

where $\omega_i \in \mathbb{R}^+$ is the weight associated with the communication direction i . The optimization problem (27) 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 (19), the resulting MSE matrix is obtained as

$$\mathbf{E}_{i,\text{mmse}}^k = \left(\mathbf{I}_{d_i} + \mathbf{V}_i^{kH} \mathbf{H}_{ii}^{kH} (\Sigma_i^k)^{-1} \mathbf{H}_{ii}^k \mathbf{V}_i^k \right)^{-1}. \quad (28)$$

By recalling (26), and upon utilization of $\mathbf{U}_{i,\text{mmse}}^k$, we observe the following useful connection to the rate function

$$I_i^k = -\log_2 \left| \mathbf{E}_{i,\text{mmse}}^k \right|, \quad (29)$$

which facilitates the decomposition of rate function via the following lemma, see also [35, Eq. (9)].

Lemma III.1. *Let $\mathbf{E} \in \mathbb{C}^{d \times d}$ be a positive definite matrix. The maximization of the term $-\log |\mathbf{E}|$ is equivalent to the maximization*

$$\max_{\mathbf{E}, \mathbf{S}} -\text{tr}(\mathbf{S}\mathbf{E}) + \log |\mathbf{S}| + d, \quad (30)$$

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

$$\mathbf{S} = \mathbf{E}^{-1}, \quad (31)$$

at the optimality.

Proof: See [47, Lemma 2]. \blacksquare

By recalling (29), and utilizing Lemma III.1, the original optimization problem over \mathbb{V} can be equivalently formulated as

$$\max_{\mathbb{V}, \mathbb{U}, \mathbb{S}} \sum_{i \in \mathbb{I}} \omega_i \sum_{k \in \mathbb{F}_K} \left(\log |\mathbf{S}_i^k| + d_i - \text{tr}(\mathbf{S}_i^k \mathbf{E}_i^k) \right) \quad \text{s.t. (18b),} \quad (32)$$

where $\mathbb{S} := \{\mathbf{S}_i^k \succ 0, \forall i \in \mathbb{I}, \forall k \in \mathbb{F}_K\}$. The obtained optimization problem (32) is not a jointly convex problem. Nevertheless, it is a QCP over \mathbb{V} when other variables are fixed, and can be obtained with a similar structure as for (18). Moreover, the optimization over \mathbb{U} and \mathbb{S} is respectively obtained from (19), and (31) as $\mathbf{S}_i^k = \mathbf{E}_i^k{}^{-1}$. This facilitates an alternating optimization where in each step the corresponding problem is solved to optimality, see Algorithm 2. The defined alternating optimization steps results 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: Algorithm 1, Steps 1-2 (initialization)
 - 2: **repeat**
 - 3: Algorithm 1, Steps 5-7
 - 4: $\mathbb{S} \leftarrow \mathbf{S}_i^k = (\mathbf{E}_i^k)^{-1}$
 - 5: **until** a stable point, or maximum number of ℓ reached
 - 6: **return** $\{\mathbb{V}\}$
-

IV. BIDIRECTIONAL FD MASSIVE MIMO SYSTEMS: JOINT POWER AND SUBCARRIER ALLOCATION

In this part, we extend the studied system into an asymmetric setup, where an FD transceiver equipped with a large antenna array (e.g., a basestation) performs a bidirectional communication with multiple FD single-antenna nodes (e.g., users). Thanks to the FD capability and multi-user beamforming, the communication at different directions can flexibly coexist on shared subcarriers, improving the spectral efficiency, or can be accommodated on different subcarriers in order to control the interference. Please note that the impact of hardware impairments is known to be significant for a system with a large antenna array, due to the lower per-element cost, e.g., low resolution ADC and DAC [37]. This signifies the role of the characterization in Lemma II.1 regarding the impact of hardware impairments for an FD MIMO OFDM system. In order to extend the defined setup to an asymmetric one, we denote the set of communication directions from (to) the users to (from) the massive MIMO transceiver as \mathbb{I}_{UL} (\mathbb{I}_{DL}), such that $\mathbb{I} = \mathbb{I}_{UL} \cup \mathbb{I}_{DL}$. Moreover, the lower-case notations⁷ ($\tilde{\mathbf{f}}_i^k$)

⁷The channel dimensions are accordingly obtained as $\mathbf{H}_{ij}^k \in \mathbb{C}^1$ when $i \in \mathbb{I}_{UL}$, $j \in \mathbb{I}_{DL}$, $\mathbf{h}_{ii}^k \in \mathbb{C}^{\tilde{M}} (\mathbb{C}^{1 \times \tilde{N}})$ when $i \in \mathbb{I}_{UL}(\mathbb{I}_{DL})$, and $\mathbf{H}_{ji}^k \in \mathbb{C}^{\tilde{M} \times \tilde{N}}$ when $i \in \mathbb{I}_{DL}$, $j \in \mathbb{I}_{UL}$. \tilde{N} (\tilde{M}) represents the number of transmit (receive) antennas at the massive MIMO transceiver.

\mathbf{f}_i^k and \mathbf{u}_i^k are used to represent the (normalized) transmit and receive linear filters⁸. Moreover, we have $\mathbf{f}_i^k = \tilde{\mathbf{f}}_i^k \sqrt{p_{i,k}}$ where $p_{i,k}$ denotes the transmit power. In this part, we perform a joint subcarrier and power allocation with the goal of maximizing the system sum rate. An upper bound on the achievable information rate is obtained as

$$R_{i,k}^{\text{UB}} = \gamma_0 \log_2 \left(1 + \frac{\left| (\mathbf{u}_i^k)^H \mathbf{h}_{ii}^k \tilde{\mathbf{f}}_i^m \right|_2^2 p_{i,k}}{\sigma_{i,k}^2 + \sum_{j \in \mathbb{I}} \sum_{m \in \mathbb{F}_K} \gamma_{ij}^{km} p_{j,m}} \right) \quad (33)$$

where $0 < \gamma_0 < 1$ indicates the portion of the frame duration dedicated to data communication, and

$$\begin{aligned} \gamma_{ij}^{km} := & \underbrace{\delta_{ij}^{km} \left| (\mathbf{u}_i^k)^H \mathbf{H}_{ij}^m \tilde{\mathbf{f}}_j^m \right|_2^2}_{\text{co-channel interference}} + \\ & + \underbrace{\left(\mathbf{u}_i^k \right)^H \mathbf{H}_{ij}^k \Theta_{\text{tx},j} \text{diag} \left(\tilde{\mathbf{f}}_j^m \left(\tilde{\mathbf{f}}_j^m \right)^H \right) \left(\mathbf{H}_{ij}^k \right)^H \mathbf{u}_i^k}_{\text{transmitter distortion}} \\ & + \underbrace{\left(\mathbf{u}_i^k \right)^H \Theta_{\text{rx},i} \text{diag} \left(\mathbf{H}_{ij}^m \tilde{\mathbf{f}}_j^m \left(\tilde{\mathbf{f}}_j^m \right)^H \left(\mathbf{H}_{ij}^m \right)^H \right) \mathbf{u}_i^k}_{\text{receiver distortion}}, \end{aligned} \quad (34)$$

where $\delta_{ij}^{km} = 0$ if $k \neq m$ or $j \in \mathbb{I}_{DL}$, $i \in \mathbb{I}_{UL}$ and otherwise $\delta_{ij}^{km} = 1$. Please note that the given upper bound in (33) is obtained similar to [49] assuming an accurate CSI, please see Section V for the consideration of CSI error. It is worth mentioning the impact of hardware distortions on the RSI, as well as the ICL is evident from (34)⁹. The optimization problem for maximizing the system sum rate is formulated as

$$\max_{p_{i,k} \geq 0} \sum_{i \in \mathbb{I}} \omega_i \sum_{k \in \mathbb{F}_K} R_{i,k}^{\text{UB}}, \quad \text{s.t.} \quad \sum_{k \in \mathbb{F}_K} p_{i,k} \leq P_i, \quad i \in \mathbb{I}_{UL}, \quad (35a)$$

$$\sum_{i \in \mathbb{I}_{DL}} \sum_{k \in \mathbb{F}_K} p_{i,k} \leq P_i, \quad i \in \mathbb{I}_{DL}. \quad (35b)$$

It can be observed that (35) falls into the class of smooth difference-of-convex (DC) optimization problems. In this regard, we propose an iterative optimization, following the SIA framework [38] which is proven to converge to a point satisfying KKT optimality conditions. Let $p_{i,k,0}$ be a feasible transmit power value. Then, employing the first order Taylor's approximation on the concave terms, $R_{i,k}^{\text{UB}}$ is lower-bounded as

$$\begin{aligned} R_{i,k}^{\text{UB}} \geq & \gamma_0 \log_2 \left(\left\| \mathbf{h}_{ii}^k \right\|_2^2 p_{i,k} + \sum_{j \in \mathbb{I}} \sum_{m \in \mathbb{F}_K} \gamma_{ij}^{km} p_{j,m} + \sigma_{i,k}^2 \right) \\ & - \gamma_0 \log_2 \left(\sum_{j \in \mathbb{I}} \sum_{m \in \mathbb{F}_K} \gamma_{ij}^{km} p_{j,m,0} + \sigma_{i,k}^2 \right) \\ & - \frac{\gamma_0 \sum_{j \in \mathbb{I}} \sum_{m \in \mathbb{F}_K} \gamma_{ij}^{km} (p_{j,m} - p_{j,m,0})}{\log(2) \sum_{j \in \mathbb{I}} \sum_{m \in \mathbb{F}_K} \gamma_{ij}^{km} p_{j,m,0} + \sigma_{i,k}^2} =: \overline{R_{i,k}^{\text{UB}}}, \end{aligned} \quad (36)$$

⁸Due to the properties of the large antenna arrays, the transmit precoder and receive filters are usually chosen via a maximum ratio transmission/combining (MRT/MRC) strategy [37], a projection to the null-space of the SI channel [36] or via a joint user and SI spatial zero-forcing [48], [49], resulting in a different performance-complexity tradeoff.

⁹In particular to a massive MIMO transceiver, where low-resolution quantization is used, the distortion coefficient κ_l in $\Theta_{\text{tx},i}$ (and similarly β_l in $\Theta_{\text{rx},i}$) is obtained as κ_l [dB] = $-6.02b_l$, where b_l is the number of quantization bits at the chain l .

where $\overline{R_{i,k}^{UB}}$ is a jointly concave function over $p_{i,k}$, facilitating an iterative update where in each iteration the convex problem $\max_{p_{i,k} \geq 0} \sum_{i \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} \overline{R_{i,k}^{UB}}$ s.t. (35b) is solved to the optimality. The proposed iterative update is continued until a stable solution is obtained. It can be observed that $\overline{R_{i,k}^{UB}}$ represents a tight and global lower bound to $R_{i,k}^{UB}$, with a shared slope at the point of approximation $p_{j,m,0}$ ¹⁰. As a result, the proposed iterative update follows the requirements set in [38, Theorem 1], with a proven convergence to a solution satisfying the KKT conditions.

V. 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 an FD system, due to the strong SI channel which calls for dedicated silent times for tuning and training process, see [12, 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 SI 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 [51], where the error matrices are not known but located, with a sufficiently high probability, within a known feasible error region¹¹. This is expressed as

$$\mathbf{H}_{ij}^k = \tilde{\mathbf{H}}_{ij}^k + \Delta_{ij}^k, \quad \Delta_{ij}^k \in \mathbb{D}_{ij}^k, \quad i, j \in \mathbb{I}, \quad (37)$$

and

$$\mathbb{D}_{ij}^k := \left\{ \Delta_{ij}^k \mid \|\mathbf{D}_{ij}^k \Delta_{ij}^k\|_F \leq \zeta_{ij}^k \right\}, \quad \forall i, j \in \mathbb{I}, \quad k \in \mathbb{F}_K, \quad (38)$$

where $\tilde{\mathbf{H}}_{ij}^k$ is the estimated channel matrix and Δ_{ij}^k represents the channel estimation error. Moreover, $\mathbf{D}_{ij}^k \succeq 0$ and $\zeta_{ij}^k \geq 0$ jointly define a feasible ellipsoid region for Δ_{ij}^k which generally depends on the noise and interference statistics, and the used channel estimation method. For further elaboration on the

¹⁰This is directly concluded for a first-order Taylor's approximation on any smooth convex function [50].

¹¹The feasible error region can be obtained from the statistical distribution of the true CSI values, as a minimum radius ball or ellipsoid containing the true CSI values with a desired confidence probability, or via the knowledge of the CSI quantization strategy, in case the CSI error is dominated by feedback quantization.

used error model see [51] and the references therein.

The aggregate interference-plus-noise signal at the receiver is hence updated as

$$\mathbf{v}_i^k = \mathbf{H}_{ij}^k \mathbf{e}_{t,j}^k + \mathbf{H}_{ii}^k \mathbf{e}_{t,i}^k + \mathbf{e}_{r,i}^k + \Delta_{ij}^k \mathbf{V}_j^k \mathbf{s}_j^k + \mathbf{n}_i^k, \quad j \neq i \in \mathbb{I}, \quad (39)$$

where Σ_i^k , representing the covariance of \mathbf{v}_i^k , is expressed in (40).

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

$$\begin{aligned} \min_{\mathbb{V}, \mathbb{U}} \max_{\mathbb{Y}} \sum_{i \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} \text{tr}(\mathbf{S}_i^k \mathbf{E}_i^k), \\ \text{s.t. (18b), } \Delta_{ij}^k \in \mathbb{D}_{ij}^k, \quad \forall i, j \in \mathbb{I}, \quad k \in \mathbb{F}_K, \end{aligned} \quad (41)$$

where $\mathbb{Y} := \{\Delta_{ij}^k, \forall i, j \in \mathbb{I}, \forall k \in \mathbb{F}_K\}$, and \mathbf{E}_i^k is obtained from (17) and (40). Note that the above problem is intractable, due to the inner maximization of quadratic convex objective over \mathbb{Y} , which also invalidates the observed convex QCP structure in (18). In order to formulate the objective into a tractable form, we calculate

$$\begin{aligned} \sum_{k \in \mathbb{F}_K} \text{tr}(\mathbf{S}_i^k \mathbf{E}_i^k) &= \sum_{k \in \mathbb{F}_K} \left(\left\| \mathbf{W}_i^k \mathbf{H}_i^k \left(\mathbf{U}_i^k \mathbf{H}_{ii}^k \mathbf{V}_i^k - \mathbf{I}_{d_i} \right) \right\|_F^2 \right. \\ &\quad + \left\| \mathbf{W}_i^k \mathbf{H}_i^k \mathbf{U}_i^k \mathbf{H}_{i3-i}^k \mathbf{V}_{3-i}^k \right\|_F^2 + \sigma_{i,k}^2 \left\| \mathbf{W}_i^k \mathbf{H}_i^k \mathbf{U}_i^k \right\|_F^2 \\ &\quad + \sum_{j \in \mathbb{I}} \sum_{l \in \mathbb{F}_{N_j}} \sum_{m \in \mathbb{F}_K} \left\| \mathbf{W}_i^k \mathbf{H}_i^k \mathbf{U}_i^k \mathbf{H}_{ij}^k \left(\Theta_{\text{rx},j} \right)^{\frac{1}{2}} \Gamma_{N_j}^l \mathbf{V}_j^m \right\|_F^2 \\ &\quad + \sum_{j \in \mathbb{I}} \sum_{l \in \mathbb{F}_{M_i}} \sum_{m \in \mathbb{F}_K} \left\| \mathbf{W}_i^k \mathbf{H}_i^k \mathbf{U}_i^k \left(\Theta_{\text{rx},i} \right)^{\frac{1}{2}} \Gamma_{M_i}^l \mathbf{H}_{ij}^m \mathbf{V}_j^m \right\|_F^2 \\ &\quad \left. + \left\| \mathbf{W}_i^k \mathbf{H}_i^k \mathbf{U}_i^k \mathbf{H}_i^k \left(\Theta_{\text{rx},i} \sum_{q \in \mathbb{F}_K} \sigma_{i,q}^2 \right)^{\frac{1}{2}} \right\|_F^2 \right) \end{aligned} \quad (42)$$

$$= \sum_{j \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} \left\| \mathbf{c}_{ij}^k + \mathbf{C}_{ij}^k \text{vec}(\Delta_{ij}^k) \right\|_2^2, \quad (43)$$

where Γ_M^l is an $M \times M$ zero matrix except for the l -th diagonal element equal to 1. In the above expressions $\mathbf{W}_i^k = (\mathbf{S}_i^k)^{\frac{1}{2}}$, and

$$\mathbf{c}_{ij}^k := \begin{bmatrix} \delta_{ij} \text{vec} \left(\mathbf{W}_i^k \mathbf{H}_i^k \left(\mathbf{U}_i^k \mathbf{H}_{ii}^k \mathbf{V}_i^k - \mathbf{I}_{d_i} \right) \right) \\ \left[\text{vec} \left(\mathbf{W}_i^k \mathbf{H}_i^k \mathbf{U}_i^k \mathbf{H}_{i3-i}^k \mathbf{V}_{3-i}^k \right) \right]_{l \in \mathbb{F}_{N_j}, m \in \mathbb{F}_K} \\ \left[\text{vec} \left(\mathbf{W}_i^k \mathbf{H}_i^k \mathbf{U}_i^k \mathbf{H}_{ij}^k \left(\Theta_{\text{rx},j} \right)^{\frac{1}{2}} \Gamma_{N_j}^l \mathbf{V}_j^m \right) \right]_{l \in \mathbb{F}_{N_j}, m \in \mathbb{F}_K} \\ \left[\text{vec} \left(\mathbf{W}_i^k \mathbf{H}_i^k \mathbf{U}_i^k \mathbf{H}_{ij}^k \left(\Theta_{\text{rx},i} \right)^{\frac{1}{2}} \Gamma_{M_i}^l \mathbf{H}_{ij}^m \mathbf{V}_j^m \right) \right]_{l \in \mathbb{F}_{M_i}, m \in \mathbb{F}_K} \\ \delta_{ij} \text{vec} \left(\mathbf{W}_i^k \mathbf{H}_i^k \mathbf{U}_i^k \mathbf{H}_i^k \left(\sigma_{i,k}^2 \mathbf{I}_{M_i} + \Theta_{\text{rx},i} \sum_{m \in \mathbb{F}_K} \sigma_{i,m}^2 \right)^{\frac{1}{2}} \right) \end{bmatrix}, \quad (44)$$

$$\boldsymbol{\Sigma}_i^k = \boldsymbol{\Delta}_{ij}^k \mathbf{V}_j^k \mathbf{V}_j^{kH} \boldsymbol{\Delta}_{ij}^{kH} + \sum_{j \in \mathbb{I}} \mathbf{H}_{ij}^k \boldsymbol{\Theta}_{\text{tx},j} \text{diag} \left(\sum_{l \in \mathbb{F}_K} \mathbf{V}_j^l \mathbf{V}_j^{lH} \right) \mathbf{H}_{ij}^{kH} + \boldsymbol{\Theta}_{\text{rx},i} \text{diag} \left(\sum_{l \in \mathbb{F}_K} \left(\sigma_{i,l}^2 \mathbf{I}_{M_i} + \sum_{j \in \mathbb{I}} \mathbf{H}_{ij}^l \mathbf{V}_j^l \mathbf{V}_j^{lH} \mathbf{H}_{ij}^{lH} \right) \right) + \sigma_{i,k}^2 \mathbf{I}_{M_i}. \quad (40)$$

$$\mathbf{C}_{ij}^k := \begin{bmatrix} \mathbf{V}_j^{kT} \otimes (\mathbf{W}_i^{kH} \mathbf{U}_i^{kH}) \\ \left[\left((\boldsymbol{\Theta}_{\text{tx},j})^{\frac{1}{2}} \boldsymbol{\Gamma}_{N_j}^l \mathbf{V}_j^m \right)^T \otimes (\mathbf{W}_i^{kH} \mathbf{U}_i^{kH}) \right]_{l \in \mathbb{F}_{N_j}, m \in \mathbb{F}_K} \\ \left[\mathbf{V}_j^{kT} \otimes (\mathbf{W}_i^{mH} \mathbf{U}_i^{mH} (\boldsymbol{\Theta}_{\text{rx},i})^{\frac{1}{2}} \boldsymbol{\Gamma}_{M_i}^l) \right]_{l \in \mathbb{F}_{M_i}, m \in \mathbb{F}_K} \\ \mathbf{0}_{M_i d_i \times M_i N_j} \end{bmatrix}, \quad (45)$$

where δ_{ij} is the Kronecker delta where $\delta_{ij} = 1$ for $i = j$ and zero otherwise. Moreover we have $\mathbf{c}_{ij}^k \in \mathbb{C}^{\tilde{d}_{ij} \times 1}$, $\mathbf{C}_{ij}^k \in \mathbb{C}^{\tilde{d}_{ij} \times M_i N_j}$ such that

$$\tilde{d}_{ij} := d_i d_j (1 + K(N_j + M_i)) + d_i M_i. \quad (46)$$

Please note that (42) is obtained by recalling (17) and (40) and the known matrix equality [52, Eq. (516)], and (44)-(45) are calculated via the application of [52, Eq. (496), (497)].

By applying the Schur's complement lemma on the epigraph form of the quadratic norm (43), i.e., $\|\mathbf{c}_{ij}^k + \mathbf{C}_{ij}^k \text{vec}(\boldsymbol{\Delta}_{ij}^k)\|_2^2 \leq \tau_{ij}^k$, the optimization problem (41) is equivalently written as

$$\min_{\mathbb{V}, \mathbb{U}, \mathbb{T}} \max_{\mathbb{Y}} \sum_{i \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} \tau_{ij}^k, \quad \text{s.t. (18b), } \|\mathbf{b}_{ij}^k\|_F \leq \zeta_{ij}^k, \quad (47a)$$

$$\begin{bmatrix} 0 & \mathbf{b}_{ij}^{kH} \tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \\ \mathbf{C}_{ij}^k \tilde{\mathbf{D}}_{ij}^k \mathbf{b}_{ij}^k & \mathbf{0}_{\tilde{d}_{ij} \times \tilde{d}_{ij}} \end{bmatrix} + \begin{bmatrix} \tau_{ij}^k & \mathbf{c}_{ij}^{kH} \\ \mathbf{c}_{ij}^k & \mathbf{I}_{\tilde{d}_{ij}} \end{bmatrix} \succeq 0, \quad (47b)$$

where $\mathbb{T} := \{\tau_{ij}^k, \forall i, j \in \mathbb{I}, \forall k \in \mathbb{F}_K\}$ and

$$\tilde{\mathbf{D}}_{ij}^k := \mathbf{I}_{N_j} \otimes (\mathbf{D}_{ij}^k)^{-1}, \quad (48)$$

$$\tilde{\boldsymbol{\Delta}}_{ij}^k := \mathbf{D}_{ij}^k \boldsymbol{\Delta}_{ij}^k, \quad \mathbf{b}_{ij}^k := \text{vec}(\tilde{\boldsymbol{\Delta}}_{ij}^k), \quad (49)$$

are defined for notational simplicity. The problem (47) is still intractable, due to the inner maximization. The following lemma converts this structure into a tractable form.

Lemma V.1. *Generalized Petersen's sign-definiteness lemma: Let $\mathbf{Y} = \mathbf{Y}^H$, and $\mathbf{X}, \mathbf{P}, \mathbf{Q}$ are arbitrary matrices with complex valued elements. Then we have*

$$\mathbf{Y} \succeq \mathbf{P}^H \mathbf{X} \mathbf{Q} + \mathbf{Q}^H \mathbf{X}^H \mathbf{P}, \quad \forall \mathbf{X} : \|\mathbf{X}\|_F \leq \zeta, \quad (50)$$

if and only if

$$\exists \lambda \geq 0, \quad \begin{bmatrix} \mathbf{Y} - \lambda \mathbf{Q}^H \mathbf{Q} & -\zeta \mathbf{P}^H \\ -\zeta \mathbf{P} & \lambda \mathbf{I} \end{bmatrix} \succeq 0. \quad (51)$$

Proof: See [53, Proposition 2], [54]. \blacksquare

By choosing the matrices in Lemma V.1 such that $\mathbf{X} = \mathbf{b}_{ij}^k$,

$$\mathbf{Q} = \begin{bmatrix} -1, & \mathbf{0}_{1 \times \tilde{d}_{ij}} \end{bmatrix} \text{ and}$$

$$\mathbf{Y} = \begin{bmatrix} \tau_{ij}^k & \mathbf{c}_{ij}^{kH} \\ \mathbf{c}_{ij}^k & \mathbf{I}_{\tilde{d}_{ij}} \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \mathbf{0}_{M_i N_j \times 1}, & \tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \end{bmatrix},$$

the optimization problem in (47) is equivalently written as

$$\min_{\mathbb{V}, \mathbb{U}, \mathbb{T}, \mathbb{M}} \sum_{i,j \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} \tau_{ij}^k \quad (52a)$$

$$\text{s.t. } \mathbf{F}_{i,j}^k \succeq 0, \quad \mathbf{G}_i \succeq 0, \quad \forall i, j \in \mathbb{I}, k \in \mathbb{F}_K, \quad (52b)$$

where $\mathbb{M} := \{\lambda_{ij}^k, \forall i, j \in \mathbb{I}, k \in \mathbb{F}_K\}$, and

$$\mathbf{G}_i := \begin{bmatrix} P_i & \tilde{\mathbf{v}}_i^H \\ \tilde{\mathbf{v}}_i & \mathbf{I} \end{bmatrix}, \quad \tilde{\mathbf{v}}_i := \left[\text{vec} \left((\mathbf{I} + K \boldsymbol{\Theta}_{\text{tx},i})^{\frac{1}{2}} \mathbf{V}_i^k \right) \right]_{k \in \mathbb{F}_K},$$

$$\mathbf{F}_{i,j}^k := \begin{bmatrix} \tau_{ij}^k - \lambda_{ij}^k & \mathbf{c}_{ij}^{kH} & \mathbf{0}_{1 \times M_i N_j} \\ \mathbf{c}_{ij}^k & \mathbf{I}_{\tilde{d}_{ij}} & -\zeta_{ij}^k \mathbf{C}_{ij}^k \tilde{\mathbf{D}}_{ij}^k \\ \mathbf{0}_{M_i N_j \times 1} & -\zeta_{ij}^k \tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} & \lambda_{ij}^k \mathbf{I}_{M_i N_j} \end{bmatrix}.$$

Similar to (32), the obtained problem in (52) is not a jointly, but a separately convex problem over \mathbb{V} and \mathbb{U} , in each case when the other variables are fixed. In particular, the optimization over $\mathbb{V}, \mathbb{T}, \mathbb{M}$ is cast as an SDP, assuming a fixed \mathbb{U} . Afterwards, the optimization over $\mathbb{U}, \mathbb{T}, \mathbb{M}$ is solved as an SDP, assuming a fixed \mathbb{V} . The described alternating steps are continued until a stable point is achieved, see Algorithm 3 for the detailed procedure.

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

- 1: $\ell \leftarrow 0$; (set iteration number to zero)
- 2: $\mathbb{V}, \mathbb{U} \leftarrow$ similar initialization as Algorithm 1
- 3: **repeat**
- 4: $\ell \leftarrow \ell + 1$
- 5: $\mathbb{V}, \mathbb{T}, \mathbb{M} \leftarrow$ solve SDP (52), with fixed \mathbb{U}
- 6: $\mathbb{U}, \mathbb{T}, \mathbb{M} \leftarrow$ solve SDP (52), with fixed \mathbb{V}
- 7: **until** a stable point, or maximum number of ℓ reached
- 8: **return** $\{\mathbb{U}, \mathbb{V}\}$

C. WMMSE for sum rate maximization

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

$$\max_{\mathbb{V}} \min_{\mathbb{Y}} \sum_{i \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} I_i^k \quad (53a)$$

$$\text{s.t. (18b), } \boldsymbol{\Delta}_{ij}^k \in \mathbb{D}_{ij}^k, \quad \forall i, j \in \mathbb{I}, k \in \mathbb{F}_K. \quad (53b)$$

Via the application of Lemma III.1, and (29) the rate maximization problem is equivalently written as

$$\max_{\mathbb{V}} \min_{\mathbb{Y}} \max_{\mathbb{U}, \mathbb{W}} \sum_{i \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} \left(\log \left| \mathbf{W}_i^k \mathbf{W}_i^{kH} \right| + d_i - \text{tr} \left(\mathbf{W}_i^k \mathbf{E}_i^k \mathbf{W}_i^k \right) \right) \quad (54a)$$

$$\text{s.t. (53b),} \quad (54b)$$

where $\mathbb{W} := \{\mathbf{W}_i^k, \forall i \in \mathbb{I}, \forall k \in \mathbb{F}_K\}$. The above problem is not tractable in the current form, due to the inner min-max structure. Following the max-min exchange introduced in [47,

Section III], and undertaking similar steps as in (43)-(52a) the problem (54) is turned into

$$\max_{\mathbb{V}, \mathbb{U}, \mathbb{W}, \mathbb{T}, \mathbb{M}} \sum_{i \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} \left(2 \log |\mathbf{W}_i^k| + d_i - \sum_{j \in \mathbb{I}} \tau_{ij}^k \right) \quad (55a)$$

$$\text{s.t.} \quad \mathbf{F}_{i,j}^k \succeq 0, \mathbf{G}_i \succeq 0, \forall i, j \in \mathbb{I}, k \in \mathbb{F}_K, \quad (55b)$$

where $\mathbf{F}_{i,j}^k, \mathbf{G}_i$ are defined in (52). 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 $\mathbb{V}, \mathbb{T}, \mathbb{M}$ and $\mathbb{U}, \mathbb{T}, \mathbb{M}$ are cast as SDP in each case when other variables are fixed. Moreover, the optimization over \mathbb{W} can be efficiently implemented using the MAX-DET algorithm [55], 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 [47, 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: Algorithm 1, Steps 1-3 (initialization)
- 2: **repeat**
- 3: $\mathbb{W}, \mathbb{T}, \mathbb{M} \leftarrow$ solve MAT-DET (52), with fixed \mathbb{V}, \mathbb{U}
- 4: Algorithm 3, Steps 4-6
- 5: **until** a stable point, or maximum number of ℓ reached
- 6: **return** $\{\mathbb{U}, \mathbb{V}\}$

D. 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 [56] 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 (41) within their defined feasible region. This is expressed as

$$\max_{\mathbb{Y}} \sum_{i \in \mathbb{I}} \sum_{k \in \mathbb{F}_K} \text{tr}(\mathbf{W}_i^{kH} \mathbf{E}_i^k \mathbf{W}_i^k), \quad (56a)$$

$$\text{s.t.} \quad \|\mathbf{D}_{ij}^k \Delta_{ij}^k\|_F \leq \zeta_{ij}^k, \forall i, j \in \mathbb{I}, k \in \mathbb{F}_K. \quad (56b)$$

Due to the uncoupled nature of the error feasible set, and the value of the objective function over Δ_{ij}^k , following (43), the above problem is decomposed as

$$\min_{\mathbf{b}_{ij}^k} - \left\| \mathbf{C}_{ij}^k \tilde{\mathbf{D}}_{ij}^k \mathbf{b}_{ij}^k \right\|_2^2 - 2 \text{Re} \left\{ \mathbf{b}_{ij}^{kH} \tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{c}_{ij}^k \right\} - \mathbf{c}_{ij}^{kH} \mathbf{c}_{ij}^k \quad (57a)$$

$$\text{s.t.} \quad \mathbf{b}_{ij}^{kH} \mathbf{b}_{ij}^k \leq \zeta_{ij}^{k2}, \quad (57b)$$

where $\text{Re}\{\cdot\}$ represents the real part of a complex value. Note that the objective in (57a) is a non convex function and can not be minimized using the usual numerical solvers in the current form. Following the zero duality gap results for the non-convex

quadratic problems [57], we focus on the dual function of (57). The corresponding Lagrangian function to (57) is constructed as

$$\mathcal{L}(\mathbf{b}_{ij}^k, \rho_{ij}^k) = \mathbf{b}_{ij}^{kH} \mathbf{A}_{ij}^k \mathbf{b}_{ij}^k - 2 \text{Re} \left\{ \mathbf{b}_{ij}^{kH} \tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{c}_{ij}^k \right\} - \mathbf{c}_{ij}^{kH} \mathbf{c}_{ij}^k - \rho_{ij}^k \zeta_{ij}^{k2}, \quad (58)$$

where ρ_{ij}^k is the dual variable and

$$\mathbf{A}_{ij}^k := \rho_{ij}^k \mathbf{I}_{N_j M_i} - \tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{C}_{ij}^k \tilde{\mathbf{D}}_{ij}^k. \quad (59)$$

Consequently, the value of the dual function is obtained as

$$\mathbf{g}(\rho_{ij}^k) = - \mathbf{c}_{ij}^{kH} \mathbf{C}_{ij}^k \tilde{\mathbf{D}}_{ij}^k (\mathbf{A}_{ij}^k)^{-1} \tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{c}_{ij}^k - \mathbf{c}_{ij}^{kH} \mathbf{c}_{ij}^k - \rho_{ij}^k \zeta_{ij}^{k2},$$

if $\mathbf{A}_{ij}^k \succeq 0$, and $\tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{c}_{ij}^k \in \mathcal{R}\{\mathbf{A}_{ij}^k\}$, and otherwise is unbounded from below¹². By applying the Schur complement lemma, the maximization of the dual function is written using the epigraph form as

$$\max_{\rho_{ij}^k \geq 0, \phi_{ij}^k} - \phi_{ij}^k \quad (60a)$$

$$\text{s.t.} \quad \begin{bmatrix} \phi_{ij}^k - \mathbf{c}_{ij}^{kH} \mathbf{c}_{ij}^k - \rho_{ij}^k \zeta_{ij}^{k2} & \mathbf{c}_{ij}^{kH} \mathbf{C}_{ij}^k \tilde{\mathbf{D}}_{ij}^k \\ \tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{c}_{ij}^k & \mathbf{A}_{ij}^k \end{bmatrix} \succeq 0, \quad (60b)$$

where $\phi_{ij}^k \in \mathbb{R}$ is an auxiliary variable¹³. By plugging the obtained dual variable ρ_{ij}^k into (58), and considering the fact that $-\tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{C}_{ij}^k \tilde{\mathbf{D}}_{ij}^k + \rho_{ij}^k \mathbf{I}_{N_j M_i} \succeq 0$ as a result of (60), the optimal value of \mathbf{b}_{ij}^k is obtained from (58) as

$$\mathbf{b}_{ij}^{k*} = \left(-\tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{C}_{ij}^k \tilde{\mathbf{D}}_{ij}^k + \rho_{ij}^k \mathbf{I}_{N_j M_i} \right)^{-1} \tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{c}_{ij}^k,$$

where $(\cdot)^*$ represents the optimality and the worst case Δ_{ij}^k is consequently calculated via $\text{vec}(\Delta_{ij}^k) = \tilde{\mathbf{D}}_{ij}^k \mathbf{b}_{ij}^{k*}$.

E. Computational complexity

The proposed designs in Section III and V are based on the alternative design of the optimization variables. Furthermore, it is observed that the consideration of non-linear hardware distortions, leading to ICL, as well as the impact of CSI error, result in a higher problem dimension and thereby complicate the structure of the resulting optimization problem. In this part, we analyze the arithmetic complexity associated with the Algorithm 3. Note that Algorithm 3 is considered as a general framework, containing Algorithm 1 as a special case, since it takes into account the impacts of hardware distortion jointly with CSI error.

The optimization over \mathbb{V}, \mathbb{U} are separately cast as SDP. A

¹²If one of the aforementioned conditions is not satisfied, an infinitely large value of \mathbf{b}_{ij}^k can be chosen in the negative direction of \mathbf{A}_{ij}^k , if \mathbf{A}_{ij}^k is not positive semi-definite, or in the direction $\tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{c}_{ij}^k$ within the null-space of \mathbf{A}_{ij}^k .

¹³Note that the semi-definite presentation in (60b) automatically satisfies $\mathbf{A}_{ij}^k \succeq 0$, and $\tilde{\mathbf{D}}_{ij}^{kH} \mathbf{C}_{ij}^{kH} \mathbf{c}_{ij}^k \in \mathcal{R}\{\mathbf{A}_{ij}^k\}$.

general SDP problem is defined as

$$\min_{\mathbf{z}} \mathbf{p}^T \mathbf{z}, \text{ s.t. } \mathbf{z} \in \mathbb{R}^n, \mathbf{Y}_0 + \sum_{i=1}^n z_i \mathbf{Y}_i \succeq 0, \|\mathbf{z}\|_2 \leq q,$$

where the fixed matrices \mathbf{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 [58, Subsection 4.6.3]. The arithmetic complexity of obtaining an ϵ -solution to the defined problem, i.e., the convergence to the ϵ -distance vicinity of the optimum is upper-bounded by

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

where $\mathcal{O}(1)$ is a positive constant and invariant to the problem dimensions [58], and $\text{digit}(\epsilon)$ is obtained from [58, 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 $\mathbb{V}, \mathbb{T}, \mathbb{M}$* : The size of the variable space is given as $n = 2K(4 + \sum_{i \in \mathbb{I}} d_i N_i)$. Moreover, the block sizes are calculated as $l_m = 2 + 2Kd_i N_i$, $\forall i \in \mathbb{I}$, corresponding to the semi-definite constraint on \mathbf{G}_i , and as $l_m = 2 + 2d_{ij} + 2M_i N_j$, $\forall i, j \in \mathbb{I}, k \in \mathbb{F}_K$, corresponding to the semidefinite constraint on $\mathbf{F}_{i,j}^k$ from (52). The overall number of the blocks is calculated as $M = 2 + 4K$.

2) *Optimization over $\mathbb{U}, \mathbb{T}, \mathbb{M}$* : The size of the variable space is given as $n = 2K(4 + \sum_{i \in \mathbb{I}} d_i M_i)$. The block sizes are calculated as $l_m = 2 + 2d_{ij} + 2M_i N_j$, $\forall i, j \in \mathbb{I}, k \in \mathbb{F}_K$, corresponding to the semidefinite constraint on $\mathbf{F}_{i,j}^k$ from (52). 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.

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 V for various system situations, and under the impact of transceiver inaccuracy and CSI error. Communication channels \mathbf{H}_{ij}^k follow an uncorrelated Rayleigh flat fading model with variance ρ . For the SI channel we follow the characterization reported in [42], indicating a Rician distribution for the SI channel. In this respect we have $\mathbf{H}_{ij} \sim \mathcal{CN}\left(\sqrt{\frac{\rho_{si} K_R}{1+K_R}} \mathbf{H}_0, \frac{\rho_{si}}{1+K_R} \mathbf{I}_{M_i} \otimes \mathbf{I}_{N_j}\right)$ where ρ_{si} represents the SI channel strength, \mathbf{H}_0 is a deterministic term,¹⁴ and K_R is the Rician coefficient. For each channel realization, the resulting performance is evaluated by

¹⁴For simplicity, we choose \mathbf{H}_0 as a matrix of all-1 elements.

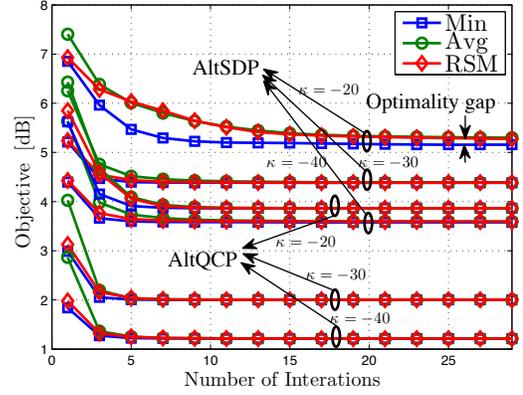


Figure 2. Average convergence behavior for AltQCP and AltSDP algorithms. AltQCP converges with fewer steps, and leads to a smaller optimality gap compared to AltSDP. Both algorithms converge in 10-30 iterations.

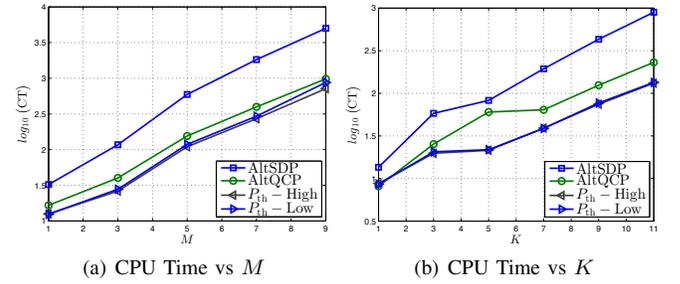


Figure 3. Comparison of the algorithm computational complexities, in terms of the required CPU time (CT), for different system dimensions, i.e., different K and M . κ represents the hardware inaccuracy, i.e., $\Theta_{rx,i} = \kappa \mathbf{I}_{M_i}$, $\Theta_{tx,i} = \kappa \mathbf{I}_{N_i}$.

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: $K = 4$, $K_R = 10$, $M := M_i = N_j = 2$, $\rho = -20$ dB, $\rho_{si} = 1$, $\sigma_n^2 := \sigma_{i,k}^2 = -30$ dB, $P_{\max} := P_i = 1$, $d_i = 1$, $\kappa = -30$ dB where $\Theta_{rx,i} = \kappa \mathbf{I}_{M_i}$ and $\Theta_{tx,i} = \kappa \mathbf{I}_{N_i}$, and $\zeta_{ij}^k = -15$ dB, $\omega_i = 1$, $\forall i, j \in \mathbb{I}$, $k \in \mathbb{F}_K$.

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 κ [dB]. In particular, "Min" and "Avg" curves respectively represent the minimum, and the average value of the algorithm objective at the corresponding optimization step

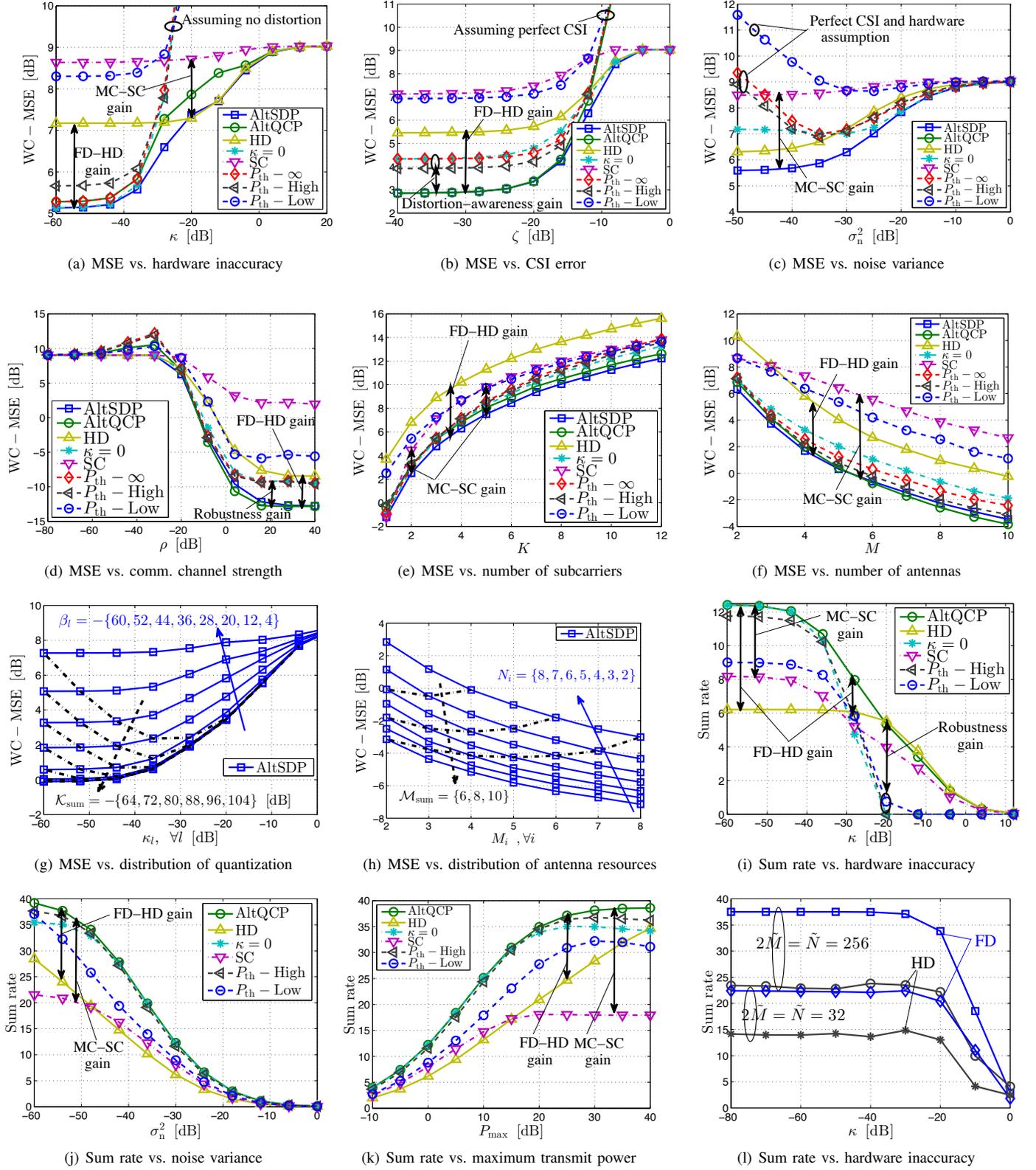


Figure 4. System performance under various specifications. The application of a distortion-aware design is essential for a system with erroneous hardware or with a high signal-to-noise ratio (SNR). The default parameter set is $\kappa = -50$ dB, $\sigma_n^2 := \sigma_{i,k}^2 = -30$ dB for (i)-(l).

over the choice of 20 random initializations. Moreover, "RSM" represents the right-singular matrix initialization proposed in [46, Appendix A]. It is observed that the algorithms converge, within 10 – 30 optimization iterations, specially as κ 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 κ 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 κ , 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-E. In Fig. 3, the required computation time (CT) is depicted for different number of antennas, as well as different number of subcarriers¹⁵. 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 RSI and ICL.

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-D, 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 V. The impact of the hardware distortions leading to ICL, as well as CSI error are taken into account.
- *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.

¹⁵Please note that the obtained CT may change depending on the used software or hardware platforms. Hence, they are reported merely as a numerical comparison measure [59], [60]. 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.

- $\kappa = 0$: The impact of CSI error is taken into account similar to, e.g., [15], [29]. Nevertheless the impact of hardware distortion, leading to ICL, is ignored.
- *SC*: The optimal single carrier design applied to the defined MC system, following a similar approach as in [12], [16]. The impact of CSI error and hardware distortions are taken into account.

Other than the approaches that directly deal with the impact of RSI, e.g., [12], [16], a low complexity design framework is proposed in [13], [14], by introducing an interference power threshold, denoted as P_{th} . In this approach, it is assumed that the SI signal can be perfectly subtracted, given the SI power is kept below P_{th} . In this regard, we evaluate the extended version of [14] on the defined MC setup for three values of P_{th} :

- $P_{th} - \{\infty, High, Low\}$: representing a design by respectively choosing $P_{th} = \infty, P_i, P_i/10$, representing a system with perfect, high, and low dynamic range conditions.

2) *Visualization*: In Figs. 4 (a)-(h) 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. 4 (i)-(l).

In Fig. 4 (a) the impact of transceiver inaccuracy is depicted on the resulting WC-MSE. It is observed that the estimation accuracy is degraded as κ increases. For the low-complexity algorithms, where the impact of hardware distortion is not considered, the resulting MSE goes to infinity as κ 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 κ . It is worth mentioning that the significant gain of an FD system with low κ over the HD counterpart, disappears for a larger levels of hardware distortion where AltSDP and HD result in a close performance.

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

In Fig. 4 (c) 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 $P_{th} - Low$, 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 (a), 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. 4 (d) 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 increase 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 ρ 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. 4 (e) 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. 4 (f) 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 ζ , and also to direct the transmit power in the desired channel and not in the SI path.

In Fig. 4 (g) the impact of the accuracy of transmit and receiver chains are studied, where $\kappa[\text{dB}] + \beta[\text{dB}] = \mathcal{K}_{\text{sum}}$, i.e., the sum-accuracy (in dB scale) is fixed. For instance, for an FD transceiver with massive antenna arrays where the utilization of analog cancelers is not feasible, and also the quantization bits are considered as costly resources, the value of \mathcal{K}_{sum} is related to the total number of quantization bits. The similar evaluation regarding the number of transmit/receive antennas is performed in Fig. 4 (h), where $M_t + M_r = \mathcal{M}_{\text{sum}}$. It is observed that different available resources, i.e., \mathcal{K}_{sum} , \mathcal{M}_{sum} , result in different optimal allocations. However, as a general insight, it is observed that the performance is degraded when resources

are concentrated only on transmit or receive side. Please note that similar approach can be used for evaluating different cost models for accuracy and antenna elements, regarding different system setups, or SIC specifications.

In Fig. 4 (i)-(l) the average sum rate behavior of the system depicted. In Fig. 4 (i), the impact of hardware inaccuracy is depicted. It is observed that a higher κ 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 κ increases. In Fig. 4 (j) and (k), 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 increases¹⁶. 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. In Fig 4 (l) the performance of the asymmetric setup, studied in Section IV, is depicted, assuming $|\mathbb{I}| = 5$, and $d_i = 1$. It is observed that the gain of FD system (over the HD counterpart) vanishes rapidly as the hardware distortions increase.

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 RSI. This issue becomes more crucial in a MC system, where the RSI spreads over multiple carriers, due to the impact of hardware distortion. In this work we have presented a modeling and design framework for an FD MIMO OFDM system, taking into account the impact of hardware distortions leading to ICL, 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 ICL 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.

APPENDIX

We start the proof with the characterization of the impact of distortion on the transmit chains. The proof to the receiver characterization is obtained similarly. The statistical independence properties at the frequency domain directly follows from the time domain statistical independence $e_{t,l}(t) \perp v_l(t)$, and $e_{t,l}(t) \perp e_{t,l'}(t)$, and the linear nature of the transformation in (10). The Gaussian and zero-mean properties similarly follow for $c_{t,l}^k$ as a linearly weighted sum of the zero-mean

¹⁶For the algorithms with a zero-distortion assumption, the maximum allowed transmit power is utilized to reduce the impact of thermal noise. However, this result in an amplified distortion effect, and a reduced performance as SNR increases, also see the MSE peaks in Fig. 4 (c)-(d) for the same algorithms.

Gaussian values $e_{t,l}(mT_s)$. The variance of $e_{t,l}^k$ can be hence obtained as

$$\mathbb{E} \left\{ \left| e_{t,l}^k \right|^2 \right\} = \mathbb{E} \left\{ \frac{1}{K} \left(\sum_{m=0}^{K-1} e_{t,l}(mT_s) e^{-\frac{j2\pi mk}{K}} \right) \left(\sum_{n=0}^{K-1} e_{t,l}^*(nT_s) e^{\frac{j2\pi nk}{K}} \right) \right\} \quad (61)$$

$$= \frac{1}{K} \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} \mathbb{E} \{ e_{t,l}(mT_s) e_{t,l}^*(nT_s) \} e^{-\frac{j2\pi(m-n)k}{K}} \quad (62)$$

$$= \kappa_l \mathbb{E} \{ |v_l(t)|^2 \} \quad (63)$$

$$= \frac{\kappa_l}{K} \sum_{m=1}^K \mathbb{E} \{ |v_l^m|^2 \} \quad (64)$$

where (61) is obtained via direct application of (10), and (63) is obtained from (6), and the statistical independence of $e_{t,l}$ at the subsequent time samples from (7). The identity (64) follows from the Parseval's theorem on the energy conversation over orthonormal Fourier basis.

REFERENCES

- [1] O. Taghizadeh, V. Radhakrishnan, A. C. Cirik, S. Shojaei, R. Mathar, and L. Lampe, "Linear precoder and decoder design for bidirectional Full-Duplex MIMO OFDM systems," in *IEEE PIMRC'17*, Montreal, Canada, 2017.
- [2] S. Hong, J. Brand, J. I. Choi, M. Jain, J. Mehlman, S. Katti, and P. Levis, "Applications of self-interference cancellation in 5G and beyond," *IEEE Comm. Magazine*, vol. 52, no. 2, pp. 114–121, Feb 2014.
- [3] D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proceedings of the 11th USENIX Conf. on Networked Systems Design and Implementation*, ser. NSDI'14, Berkeley, CA, USA, 2014, pp. 359–372.
- [4] D. Bharadia, E. McMillin, and S. Katti, "Full duplex radios," in *Proceedings of the ACM SIGCOMM 2013 Conf. on SIGCOMM*, ser. SIGCOMM '13. New York, NY, USA: ACM, 2013, pp. 375–386. [Online]. Available: <http://doi.acm.org/10.1145/2486001.2486033>
- [5] J.-H. Lee, I. Sohn, and Y.-H. Kim, "Analog cancellation schemes for full-duplex MIMO OFDM systems," *AEU-International J. of Electronics and Comm.*, vol. 70, no. 3, pp. 272–277, 2016.
- [6] J. H. Lee, "Self-interference cancelation using phase rotation in full-duplex wireless," *IEEE Trans. on Veh. Tech.*, vol. 62, no. 9, pp. 4421–4429, Nov 2013.
- [7] D. Kim, S. Park, H. Ju, and D. Hong, "Transmission capacity of full-duplex-based two-way ad hoc networks with arq protocol," *IEEE Trans. on Veh. Tech.*, vol. 63, no. 7, pp. 3167–3183, Sept 2014.
- [8] G. Zhang, K. Yang, P. Liu, and J. Wei, "Power allocation for full-duplex relaying-based d2d communication underlying cellular networks," *IEEE Trans. on Veh. Tech.*, vol. 64, no. 10, pp. 4911–4916, Oct 2015.
- [9] M. Duarte, A. Sabharwal, V. Aggarwal, R. Jana, K. K. Ramakrishnan, C. W. Rice, and N. K. Shankaranarayanan, "Design and characterization of a full-duplex multi-antenna system for wifi networks," *IEEE Trans. on Vehicular Technology*, vol. 63, no. 3, pp. 1160–1177, March 2014.
- [10] R. Feng, Q. Li, Q. Zhang, and J. Qin, "Robust secure beamforming in MISO full-duplex two-way secure communications," *IEEE Trans. on Veh. Tech.*, vol. 65, no. 1, pp. 408–414, Jan 2016.
- [11] 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 Conf. on Mobile Computing and Networking (MobiCom)*, Las Vegas, NV, Sep. 2011.
- [12] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range," *IEEE Trans. on Sig. Proc.*, vol. 60, no. 7, pp. 3702–3713, July 2012.
- [13] S. Huberman and T. Le-Ngoc, "Self-interference pricing-based MIMO full-duplex precoding," *IEEE Wireless Comm. Letters*, vol. 3, no. 6, pp. 549–552, Dec 2014.
- [14] J. Zhang, O. Taghizadeh, J. Luo, and M. Haardt, "Full duplex wireless communications with partial interference cancellation," *Proceedings of the 46th Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA*, Nov. 2012.
- [15] 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.
- [16] A. Cirik, R. Wang, Y. Hua, and M. Latva-aho, "Weighted sum-rate maximization for full-duplex MIMO interference channels," *Comm., IEEE Trans. on*, vol. 63, March 2015.
- [17] 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'17*, Paris, France, May 2017, pp. 5950–5956.
- [18] A. C. Cirik, R. Wang, Y. Rong, and Y. Hua, "MSE based transceiver designs for bi-directional full-duplex MIMO systems," in *IEEE SPAWC'14*, June 2014, pp. 384–388.
- [19] A. Cirik, S. Biswas, S. Vuppala, and T. Ratnarajah, "Beamforming design for full-duplex MIMO interference channels: QoS and energy-efficiency considerations," *IEEE Trans. on Comm.*, vol. PP, no. 99, pp. 1–1, 2016.
- [20] 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.
- [21] 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 Conf. Record of the Forty Sixth Asilomar Conf. on Signals, Systems and Computers (ASILOMAR)*, Nov 2012, pp. 45–49.
- [22] 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 Comm.*, Sep 2009, pp. 3169–3173.
- [23] M. Mokhtar, N. Al-Dhahir, and R. Hamila, "OFDM full-duplex DF relaying under IQ imbalance and loopback self-interference," *IEEE Trans. on Veh. Tech.*, vol. 65, no. 8, pp. 6737–6741, Aug 2016.
- [24] N. Li, Y. Li, M. Peng, and W. Wang, "Resource allocation in multi-carrier full-duplex amplify-and-forward relaying networks," in *IEEE VTC Spring*, May 2016, pp. 1–5.
- [25] 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 Trans. on Comm.*, vol. PP, no. 99, pp. 1–1, 2017.
- [26] M. Al-Imari, M. Ghorashi, and P. Xiao, "Radio resource allocation for full-duplex multicarrier wireless systems," in *ISWCS'15*, Aug 2015, pp. 571–575.
- [27] A. C. Cirik, K. Rikkinen, Y. Rong, and T. Ratnarajah, "A subcarrier and power allocation algorithm for OFDMA full-duplex systems," in *EuCNC'15*, June 2015, pp. 11–15.
- [28] W. Li, J. Lilleberg, and K. Rikkinen, "On rate region analysis of half- and full-duplex OFDM communication links," *IEEE J. on Selected Areas in Comm.*, vol. 32, no. 9, pp. 1688–1698, Sep 2014.
- [29] D. W. K. Ng, E. S. Lo, and R. Schober, "Dynamic resource allocation in MIMO-OFDMA systems with full-duplex and hybrid relaying," *IEEE Trans. on Comm.*, vol. 60, no. 5, pp. 1291–1304, 2012.
- [30] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex MIMO relaying: Achievable rates under limited dynamic range," *IEEE J. on Selected Areas in Comm.*, vol. 30, no. 8, pp. 1541–1553, Sep 2012.
- [31] W. Xie, X. Xia, Y. Xu, K. Xu, and Y. Wang, "Massive MIMO

- full-duplex relaying with hardware impairments,” *J. of Comm. and Networks*, vol. 19, no. 4, pp. 351–362, August 2017.
- [32] X. Xia, D. Zhang, K. Xu, W. Ma, and Y. Xu, “Hardware impairments aware transceiver for full-duplex massive MIMO relaying,” *IEEE Trans. on Sig. Proc.*, vol. 63, no. 24, pp. 6565–6580, Dec 2015.
- [33] S. Jia and B. Aazhang, “Signaling design of two-way MIMO full-duplex channel: Optimality under imperfect transmit front-end chain,” *IEEE Trans. on Wireless Comm.*, vol. 16, no. 3, pp. 1619–1632, March 2017.
- [34] O. Taghizadeh, A. Cirik, and R. Mathar, “Hardware impairments aware transceiver design for full-duplex amplify-and-forward MIMO relaying,” *IEEE Trans. on Wireless Comm.*, vol. PP, no. 99, pp. 1–1, 2017.
- [35] 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 Trans. on Wireless Comm.*, vol. 7, no. 12, pp. 4792–4799, Dec 2008.
- [36] E. Everett, C. Shepard, L. Zhong, and A. Sabharwal, “Softnull: Many-antenna full-duplex wireless via digital beamforming,” *IEEE Trans. on Wireless Comm.*, vol. 15, no. 12, pp. 8077–8092, Dec 2016.
- [37] C. Kong, C. Zhong, S. Jin, S. Yang, H. Lin, and Z. Zhang, “Full-duplex massive MIMO relaying systems with low-resolution ADCs,” *IEEE Trans. on Wireless Comm.*, vol. 16, no. 8, pp. 5033–5047, Aug 2017.
- [38] B. R. Marks and G. P. Wright, “Technical note—a general inner approximation algorithm for nonconvex mathematical programs,” *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.
- [39] W. Namgoong, “Modeling and analysis of nonlinearities and mismatches in ac-coupled direct-conversion receiver,” *IEEE Trans. on Wireless Comm.*, vol. 4, no. 1, pp. 163–173, Jan 2005.
- [40] G. Santella and F. Mazzino, “A hybrid analytical-simulation procedure for performance evaluation in M-QAM-OFDM schemes in presence of nonlinear distortions,” *IEEE Trans. on Veh. Tech.*, vol. 47, no. 1, pp. 142–151, Feb 1998.
- [41] 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. on Selected Areas in Comm.*, vol. 26, no. 6, pp. 867–876, Aug 2008.
- [42] M. Duarte, C. Dick, and A. Sabharwal, “Experiment-driven characterization of full-duplex wireless systems,” *IEEE Trans. on Wireless Comm.*, vol. 11, no. 12, pp. 4296–4307, Dec 2012.
- [43] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM wireless communications with MATLAB*. John Wiley & Sons, 2010.
- [44] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, “Wireless backhauling of 5G small cells: challenges and solution approaches,” *IEEE Wireless Comm.*, vol. 22, no. 5, pp. 22–31, Oct 2015.
- [45] D. Bertsekas, “Nonlinear programming. athena scientific,” *Belmont, Massachusetts*, 1999.
- [46] H. Shen, B. Li, M. Tao, and X. Wang, “MSE-based transceiver designs for the MIMO interference channel,” *IEEE Trans. on Wireless Comm.*, vol. 9, no. 11, pp. 3480–3489, Nov 2010.
- [47] J. Jose, N. Prasad, M. Khojastepour, and S. Rangarajan, “On robust weighted-sum rate maximization in MIMO interference networks,” in *IEEE ICC’11*, June 2011, pp. 1–6.
- [48] T. Parfait, Y. Kuang, and K. Jerry, “Performance analysis and comparison of ZF and MRT based downlink massive MIMO systems,” in *2014 Sixth International Conf. on Ubiquitous and Future Networks (ICUFN)*, July 2014, pp. 383–388.
- [49] X. Xia, D. Zhang, K. Xu, W. Ma, and Y. Xu, “Hardware impairments aware transceiver for full-duplex massive MIMO relaying,” *IEEE Transactions on Sig. Proc.*, vol. 63, no. 24, pp. 6565–6580, Dec 2015.
- [50] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.
- [51] J. Wang and D. P. Palomar, “Worst-case robust MIMO transmission with imperfect channel knowledge,” *IEEE Trans. on Sig. Proc.*, vol. 57, no. 8, pp. 3086–3100, 2009.
- [52] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Technical University of Denmark, nov 2012, version 20121115. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?3274>
- [53] Y. C. Eldar and N. Merhav, “A competitive minimax approach to robust estimation of random parameters,” *IEEE Trans. on Signal Processing*, vol. 52, no. 7, pp. 1931–1946, July 2004.
- [54] M. V. Khlebnikov and P. S. Shcherbakov, “Petersens lemma on matrix uncertainty and its generalizations,” *Automation and Remote Control*, vol. 69, no. 11, pp. 1932–1945, 2008.
- [55] L. Vandenberghe, S. Boyd, and S.-P. Wu, “Determinant maximization with linear matrix inequality constraints,” *SIAM J. on matrix analysis and applications*, vol. 19, no. 2, pp. 499–533, 1998.
- [56] A. Mutapcic and S. Boyd, “Cutting-set methods for robust convex optimization with pessimizing oracles,” *Optimization Methods & Software*, vol. 24, no. 3, pp. 381–406, 2009.
- [57] X. Zheng, X. Sun, D. Li, and Y. Xu, “On zero duality gap in nonconvex quadratic programming problems,” *J. of Global Optimization*, vol. 52, no. 2, pp. 229–242, 2012.
- [58] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Siam, 2001, vol. 2.
- [59] B. Li, C. Z. Wu, H. H. Dam, A. Cantoni and K. L. Teo, “A parallel low complexity zero-forcing beamformer design for multiuser MIMO systems via a regularized dual decomposition method,” *IEEE Trans. on Sig. Proc.*, vol. 63, no. 16, pp. 4179–4190, Aug.15, 2015.
- [60] L. M. Cortés-Peña, J. R. Barry and D. M. Blough, “Jointly optimizing stream allocation, beamforming and combining weights for the MIMO interference channel,” *IEEE Trans. Wireless Comm.*, vol. 14, no. 4, pp. 2245–2256, April 2015.
- [61] R. Zakaria and D. Le Ruyet, “A novel filter-bank multicarrier scheme to mitigate the intrinsic interference: application to MIMO systems,” in *IEEE Trans. Wireless Comm.*, vol. 11, no. 3, pp. 1112–1123, March 2012.