Robust Interference Management via Linear Precoding and Linear/Non-Linear Equalization

Seok-Hwan Park, Ali M. Fouladgar, Tariq Elkourdi, Osvaldo Simeone, Onur Sahin and Shlomo Shamai (Shitz)

Abstract

This work studies the robust design of linear precoding and linear/ non-linear equalization for multi-cell MIMO systems in the presence of imperfect channel state information (CSI). A worst-case design approach is adopted whereby the CSI error is assumed to lie within spherical sets of known radius. First, the optimal robust design of linear precoders is tackled for a MIMO interference broadcast channel (MIMO-IBC) with general unicast/multicast messages in each cell and operating over multiple time-frequency resources. This problem is formulated as the maximization of the worst-case sum-rate assuming optimal detection at the mobile stations (MSs). Then, symbol-by-symbol non-linear equalization at the MSs is considered. In this case, the problem of jointly optimizing linear precoding and decision-feedback (DF) equalization is investigated for a MIMO interference channel (MIMO-IC) with the goal of minimizing the worst-case sum-mean squared error (MSE). Both problems are addressed by proposing iterative algorithms with descent properties. The algorithms are also shown to be implementable in a distributed fashion on processors that have only local and partial CSI by means of the Alternating Direction Method of Multipliers (ADMM). From numerical results, it is shown that the proposed robust solutions significantly improve over conventional non-robust schemes in terms of sum-rate or symbol error rate. Moreover, it is seen that the proposed joint design of linear precoding and DF equalization outperforms existing separate solutions.

Index Terms

S.-H. Park was with the Center for Wireless Communications and Signal Processing Research (CWCSPR), ECE Department, New Jersey Institute of Technology (NJIT), Newark, NJ 07102, when this work was carried out (email: seokhwan81@gmail.com).

O. Simeone. A. M. Fouladgar and T. Elkourdi are with CWCSPR, NJIT (email: osvaldo.simeone@njit.edu).

O. Sahin is with InterDigital Inc., Melville, New York, 11747, USA (email: onur.sahin@interdigital.com).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion, Haifa, 32000, Israel (email: sshlomo@ee.technion.ac.il).

ulti-cell MIMO linear precoding sum-rate maximization robust optimization bounded uncertainty decision-feedback equalization ADMMulti-cell MIMO linear precoding sum-rate maximization robust optimization bounded uncertainty decision-feedback equalization ADMMM

I. INTRODUCTION

The trend towards extreme densification is one of the defining characteristics of 5G systems [1]. As a result, dense heterogeneous networks are expected typical in 5G systems. The study of this class of network architectures is bringing to the fore the need to develop practical and effective interference management techniques. A key constraint on the design of such techniques is the difficulty to collect channel state information (CSI).

Linear precoding at the base stations (BSs) of a multi-cell system provides a practical and effective means to control intra- and inter-cell interference (see, e.g., [2][3]). Linear precoding can be designed assuming optimal decoding at the mobile stations (MSs), and hence invoking the Shannon capacity as the relevant performance metric, see, e.g., [4]; or, alternatively, one can consider suboptimal symbol-by-symbol equalizers at the MSs, for which other performance criteria, such as interference leakage minimization for interference alignment [5] or minimum mean squared error (MSE) [6], are relevant. Moreover, theoretical results demonstrate the potential advantages that can be accrued by performing linear precoding over multiple time-frequency resources (see [7] and references therein).

Linear precoding design requires the availability of CSI at the base stations (BSs). CSI imperfections are inevitable in both time-division and frequency-division duplex systems due to channel estimation errors, CSI outdating and limited feedback (see, e.g., [8][9]). Therefore, a system design that is tailored to the available CSI is bound to incur a potentially severe degradation under the actual unknown channel conditions. In order to prevent unforeseen system failures due to CSI mismatch, a well established approach prescribes the adoption of a *robust optimization* [10] formulation. Accordingly, linear precoding is designed by optimizing the worst-case performance over the set of all plausible CSI conditions given the available CSI at the BSs. In this fashion, a given performance level is guaranteed no matter what the actual CSI is within the given uncertainty set. Alternatively, one can attempt to maximize the average performance with respect to the distribution of the CSI error, see, e.g., [11].

Robust precoder design was studied in [8][9][12] (see also references therein) for multi-cell downlink channels with unicast messages that operate over a single time-frequency resource. Instead, the work [13] considered the robust optimization of single-cell multicast systems in the context of a cognitive-radio network. These works consider optimal decoders and use the sum-rate as the performance metric of interest. The analysis of the degrees of freedom in the presence of imperfect CSI can be found in [14][15].

The robust design of DF equalizers under imperfect CSI is studied in [16] for a single-cell system, e.g., in the absence of interference. The case of multiple cells was considered in [17] in the presence of perfect CSI by considering a separate design of linear precoding and DF equalization.

A. Contributions

This paper has the following main contributions.

1. Linear precoding optimization based on robust sum-rate maximization: We first consider a multi-cell multicast system in which each BS transmits a number of independent messages, each to be decoded by different subset of MSs, over multiple time-frequency resources. The system is referred to as a *MIMO interference broadcast channel* (MIMO-IBC). Unlike the related prior work mentioned above, the model includes the possibility for a BS to multicast the same message to multiple MSs, as studied in [18]-[21] for single-cell systems. Note that multicast is for instance implemented in LTE systems via the evolved multimedia broadcast/multicast service (eMBMS) interface [22]. Robust linear precoding optimization is tackled for this system under the assumptions that the MSs perform optimal decoding in terms of worst-case sum-rate maximization in Sec. III. An iterative algorithm that monotonically increases a lower bound of the objective function is proposed. We then focus on the special case of a MIMO interference channel (MIMO-IC) where each BS communicates to a single MS. For the MIMO-IC, a distributed implementation of the proposed algorithm is derived based on the Alternating Direction Method of Multipliers (ADMM) that only requires local (imperfect) CSI. Numerical results in Sec. V show that the proposed robust solution provides significant performance improvement over conventional non-robust schemes that consider the nominal CSI as being accurate.

2. Linear precoding and DF equalization optimization based on sum-MSE minimization: While the contribution discussed above assumes optimal decoders at the MSs, we consider here the case in which each MS runs a symbol-by-symbol DF equalizer for the MIMO-IC. The DF equalizer is allowed to successively decode any subset of intra-cell and inter-cell streams. Robust joint optimization of linear precoding and DF equalization is tackled for a MIMO-IC in terms of worst-case sum-MSE minimization in Sec. IV. An iterative algorithms is proposed that monotonically decreases an upper bound on the sum-MSE and a distributed implementation is developed based on ADMM. Numerical evidence, reported in Sec. V, shows that the joint optimization considered here is beneficial over the separate optimization considered in the earlier work [17] and that the proposed robust solution outperforms non-robust approaches.

Notations: The set of all $M \times N$ complex matrices is denoted by $\mathbb{C}^{M\times N}$. We use the notation $\mathbf{X} \succeq \mathbf{0}$ to indicate that the matrix $\mathbf{X}$ is positive semidefinite. The operations $(\cdot)^T$ and $(\cdot)^\dagger$ denote transpose and Hermitian transpose of a matrix or vector, and the Frobenius norm and the spectral norm of a matrix $\mathbf{X}$ are denoted by $||\mathbf{X}||_F$ and $||\mathbf{X}||$, respectively. $\mathbb{E}[\cdot]$ represents the expectation operator. We use the notation $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ otherwise.

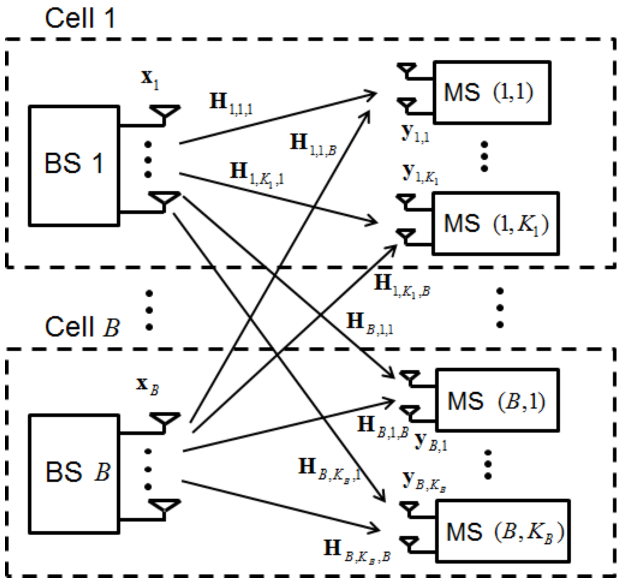

Figure 1. MIMO interference broadcast channel (MIMO-IBC) with B interfering cells.

II. System Model

We study a multi-cell MIMO downlink channel, also known as MIMO-IBC, with B mutually interfering cells that operate over the same time-frequency resources as shown in Fig. 1. We assume that each ith cell has a single multi-antenna BS that wishes to communicate with K_i multi-antenna MS located in the cell. We denote the numbers of antennas at the ith BS and the kth MS in the ith cell as $N_{T,i}$ and $N_{R,(i,k)}$, respectively, and define the sets of all BSs and MSs in the ith cell as $\mathcal{B} \triangleq \{1, \ldots, B\}$ and $\mathcal{K}_i \triangleq \{1, \ldots, K_i\}$, respectively. For notational convenience, let us denote the kth MS in cell i as MS (i, k).

In the most general case, we assume a multicast set-up in which that the ith BS transmits M_i messages, where each mth message $W_{i,m}$ is to be decoded by a set $\mathcal{D}_{i,m} \subseteq \mathcal{K}_i$ of the MSs within the cell. We assume that $\mathcal{D}_{i,m} \neq \emptyset$ and $\mathcal{D}_{i,m} \neq \mathcal{D}_{i,m'}$ for all $m \neq m' \in \mathcal{M}_i \triangleq \{1, \ldots, M_i\}$. Note that the systems with only unicast messages (as in, e.g., [2]) or only multicast messages (as in e.g., [13]) can be captured in this model by choosing $M_i = K_i$ with $\mathcal{D}_{i,m} = \{m\}$ for all $m \in \mathcal{K}_i$ and $M_i = 1$ with $\mathcal{D}_{i,1} = \mathcal{K}_i$, respectively. We also define as $\mathcal{S}_{i,k} \subseteq \mathcal{M}_i$ the set of the indices of messages to be decoded by the (i, k)th MS.

In order to simplify the presentation, we will also consider a simpler system with a single MS per cell, i.e., with $K_i = 1$, as shown in Fig. 2. This set-up amounts to a MIMO-IC with B transmitter-receiver pairs and will be described by using a single index to identify the quantities related to each cell. Specifically, each BS i here will have a single message W_i, i.e., $\mathcal{M}_i \triangleq \{1\}$, which is intended for

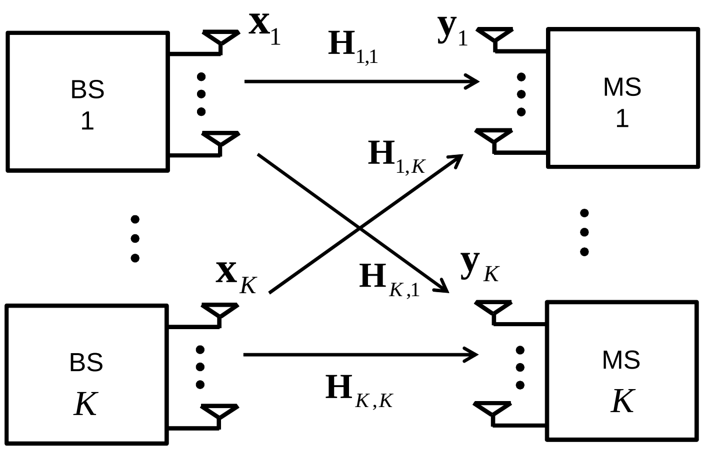

Figure 2. MIMO interference channel (MIMO-IC) with B interfering cells.

the ith MS, i.e., $\mathcal{D}_{i,1} = \mathcal{K}_i = \{1\}$.

A. Signal Model

The BSs perform joint precoding over L time or frequency slots with arbitrarily varying channel gains[1]. In the MIMO-IBC of Fig. 1, the signal $\mathbf{y}_{i,k} \triangleq [\mathbf{y}_{i,k}(1); \ldots; \mathbf{y}_{i,k}(L)] \in \mathbb{C}^{LN_{R,(i,k)} \times 1}$ received by MS (i,k) over the slots, with $\mathbf{y}_{i,k}(l)$ representing the received signal at slot $l \in \mathcal{L} \triangleq \{1, \ldots, L\}$, can be written as

$$\mathbf{y}_{i,k} = \sum_{j \in \mathcal{B}} \mathbf{H}_{i,k,j} \mathbf{x}_j + \mathbf{z}_{i,k}, \tag{1}$$

where $\mathbf{H}_{i,k,j} \triangleq \mathrm{diag}(\mathbf{H}_{i,k,j}(1), \ldots, \mathbf{H}_{i,k,j}(L)) \in \mathbb{C}^{LN_{R,(i,k)} \times LN_{T,j}}$ denotes the channel response matrix from the jth BS to the MS (i,k), with $\mathbf{H}_{i,k,j}(l) \in \mathbb{C}^{N_{R,(i,k)} \times N_{T,j}}$ representing the matrix at slot $l \in \mathcal{L}$; $\mathbf{x}_j \triangleq [\mathbf{x}_j(1); \ldots; \mathbf{x}_j(L)] \in \mathbb{C}^{LN_{T,j} \times 1}$ is the signal transmitted by the jth BS with $\mathbf{x}_j(l) \in \mathbb{C}^{N_{T,j} \times 1}$ being the signal transmitted at slot l; $\mathbf{z}_{i,k} \triangleq [\mathbf{z}_{i,k}(1); \ldots; \mathbf{z}_{i,k}(L)] \in \mathbb{C}^{LN_{R,(i,k)} \times 1}$ denotes the additive noise distributed as $\mathbf{z}_{i,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{i,k})$ with $\mathbf{\Sigma}_{i,k} \triangleq \mathrm{diag}(\mathbf{\Sigma}_{i,k}(1), \ldots, \mathbf{\Sigma}_{i,k}(L))$, where $\mathbf{\Sigma}_{i,k}(l) \succeq \mathbf{0}$ denotes the covariance matrix of the noise $\mathbf{z}_{i,k}(l)$ at slot l. We assume that each BS i is subject to the transmit power constraint

$$\frac{1}{L} \mathbb{E}\left[\|\mathbf{x}_i\|^2 \right] \leq P_i. \tag{2}$$

The notation is simplified for the MIMO-IC in Fig. 2, so that we can write

$$\mathbf{y}_i = \sum_{j \in \mathcal{B}} \mathbf{H}_{i,j} \mathbf{x}_j + \mathbf{z}_i \tag{3}$$

[1]We recall that one way to obtain varying channels in the frequency domain is to have flat-fading channels but with asynchronous interference [23].

for the signal received by MS i, where $\mathbf{H}_{i,j} \in \mathbb{C}^{LN_{R,i} \times LN_{T,j}}$ denotes the channel response matrix from the jth BS to the ith MS.

B. Linear Precoding

Each BS i in the MIMO-IBC transmits the linearly precoded signal

$$\mathbf{x}_i = \sum_{m \in \mathcal{M}_i} \mathbf{V}_{i,m} \mathbf{s}_{i,m}, \tag{4}$$

where $\mathbf{s}_{i,m} \in \mathbb{C}^{d_{i,m} \times 1}$ is the baseband signal encoding the message $W_{i,m}$ intended for MSs (i,k) with $k \in \mathcal{D}_{i,m}$; and the matrix $\mathbf{V}_{i,m} \in \mathbb{C}^{LN_{T,i} \times d_{i,m}}$ is the associated precoding matrix. The dimension $d_{i,m}$ of the encoded signal $\mathbf{s}_{i,m}$ cannot exceed the maximum number of signal dimensions on the channels between the ith BS and the MSs in set $\mathcal{D}_{i,m}$, namely $L \cdot \min\{N_{T,i}, N_{R,(i,k)}\}$ for $k \in \mathcal{D}_{i,m}$. Note that, for the MIMO-IC, the transmitted signal is simplified to

$$\mathbf{x}_i = \mathbf{V}_i \mathbf{s}_i, \tag{5}$$

where $\mathbf{V}_i \in \mathbb{C}^{LN_{T,i} \times d_i}$ is the precoding matrix for the signal $\mathbf{s}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ intended for the ith MS.

In the MIMO-IBC with the linear precoding model (4) described above, the received signal (1) can be rewritten as

$$\mathbf{y}_{i,k} = \sum_{m \in \mathcal{S}_{i,k}} \mathbf{H}_{i,k,i} \mathbf{V}_{i,m} \mathbf{s}_{i,m} + \sum_{q \in \mathcal{M}_i \setminus \mathcal{S}_{i,k}} \mathbf{H}_{i,k,i} \mathbf{V}_{i,q} \mathbf{s}_{i,q} + \sum_{j \in \mathcal{B} \setminus \{i\}} \sum_{q \in \mathcal{M}_j} \mathbf{H}_{i,k,j} \mathbf{V}_{j,q} \mathbf{s}_{j,q} + \mathbf{z}_{i,k}, \tag{6}$$

where the first term is the desired signals to be decoded by the receiving MS (i,k) while the second and third terms stand for the intra-cell and inter-cell interference signals, respectively. For the MIMO-IC, using (5), the received signal (3) reduces to

$$\mathbf{y}_i = \mathbf{H}_{i,i} \mathbf{V}_i \mathbf{s}_i + \sum_{j \in \mathcal{B} \setminus \{i\}} \mathbf{H}_{i,j} \mathbf{V}_j \mathbf{s}_j + \mathbf{z}_i, \tag{7}$$

showing only the presence of inter-cell interference.

C. Channel State Information

With the exceptions of Sec. III-D and Sec. IV-E, we assume that the optimization of the system is performed at a central unit that receives the CSI about the channel matrices $\{\mathbf{H}_{i,k,j}\}_{i \in \mathcal{B}, k \in \mathcal{K}_i}$ from the jth BS for all $j \in \mathcal{B}$. These channel matrices can be estimated by each BS either via uplink training for Time Division Duplex systems or fed back by the MSs in a Frequency Division Duplex system. As a result, the channel matrices $\hat{\mathbf{H}}_{i,k,j}$ available at the central unit are different from the actual channel matrix $\mathbf{H}_{i,k,j}$ due to, e.g., *i)* channel estimation errors at the BSs; *ii)* channel outdating; *iii)* capacity limitation on the backhaul links connecting the BSs to the central unit.

In order to address the CSI inaccuracy, we adopt a standard deterministic uncertainty model, whereby the actual channel $\mathbf{H}_{i,k,j}$ is related to the CSI $\hat{\mathbf{H}}_{i,k,j}$ as

$$\mathbf{H}_{i,k,j} = \hat{\mathbf{H}}_{i,k,j} + \mathbf{\Delta}_{i,k,j}, \tag{8}$$

where the error matrix $\mathbf{\Delta}_{i,k,j}$, defined as $\mathbf{\Delta}_{i,k,j} \triangleq \mathrm{diag}(\mathbf{\Delta}_{i,k,j}(1), \ldots, \mathbf{\Delta}_{i,k,j}(L))$, is known to lie in the bounded spherical uncertainty region [8][9][12]

$$\mathcal{U}_{i,k,j}(l) \triangleq \left\{ \mathbf{\Delta}_{i,k,j}(l) \middle| \, \|\mathbf{\Delta}_{i,k,j}(l)\|_F \leq \varepsilon_{i,k,j}(l) \right\}. \tag{9}$$

Therefore, the central unit performs the system design based on the knowledge of the nominal CSIs $\{\hat{\mathbf{H}}_{i,k,j}\}_{i,j\in\mathcal{B},k\in\mathcal{K}_i}$ and the parameters $\{\varepsilon_{i,k,j}(l)\}_{l\in\mathcal{L},i,j\in\mathcal{B},k\in\mathcal{K}_i}$ that measure the inaccuracy of the reported CSI as per (9). Note that, following the discussion above, the notation simplifies in a natural fashion for the MIMO-IC. Finally, we assume that the MSs have all necessary local CSI, which can be obtained via downlink training.

A distributed implementation of the proposed algorithms that does not require a central unit will be discussed in Sec. III-D and Sec. IV-E.

III. Robust Sum-Rate Maximization over Linear Precoding

In this section, we investigate the optimal design of the precoding matrices $\mathbf{V} \triangleq \{\mathbf{V}_{i,m}\}_{i\in\mathcal{B},m\in\mathcal{M}_i}$ for the MIMO-IBC with the aim of maximizing the sum-rate. A key underlying assumption here is that the BSs use (point-to-point) capacity-achieving codes and that the MSs perform optimal decoding, so that the Shannon capacity can be used as the performance criterion. We start by deriving an achievable sum-rate for given precoding matrices $\mathbf{V} \triangleq \{\mathbf{V}_{i,m}\}_{i\in\mathcal{B},m\in\mathcal{M}_i}$ in Sec. III-A. The design problem is then formulated in Sec. III-B and tackled in Sec. III-C. A distributed situation based on ADMM is proposed in Sec. III-D.

A. Sum-Rate

In order to decode the intended message set $\mathcal{S}_{i,k}$, the MS (i,k) performs successive interference cancellation with a predetermined order $\pi_{i,k} : \{1, \ldots, |\mathcal{S}_{i,k}|\} \to \mathcal{S}_{i,k}$. More specifically, MS (i,k) first decodes the signal $\mathbf{s}_{i,\pi_{i,k}(1)}$ based on the received signal $\mathbf{y}_{i,k}^{(1)} \triangleq \mathbf{y}_{i,k}$ by treating all other signals as noise. Then, MS (i,k) cancels the decoded signal $\mathbf{s}_{i,\pi_{i,k}(1)}$ on the received signal $\mathbf{y}_{i,k}$ obtaining $\mathbf{y}_{i,k}^{(2)} \triangleq \mathbf{y}_{i,k} - \mathbf{H}_{i,k,i}\mathbf{V}_{i,\pi_{i,k}(1)}\mathbf{s}_{i,\pi_{i,k}(1)}$. The next signal $\mathbf{s}_{i,\pi_{i,k}(2)}$ is hence decoded based on $\mathbf{y}_{i,k}^{(2)}$ by treating all remaining signals as noise, and the process is repeated until all the signals $\{\mathbf{s}_{i,m}\}_{m\in\mathcal{S}_{i,k}}$ are decoded. We emphasize that in prior works [8][9][12] the intended message set includes a single message, hence not requiring successive interference cancellation and significantly simplifying the analysis.

Following, e.g., [24][25] (see also references therein), we find it useful to express the achievable sum-rate in terms of the minimum MSE covariance matrices. The minimum MSE matrix $\mathbf{E}_{i,m,k}^{*} \triangleq$

$\mathbb{E}[\mathbf{e}_{i,m,k}\mathbf{e}_{i,m,k}^{\dagger}]$ is the covariance of the error vector $\mathbf{e}_{i,m,k} \triangleq \hat{\mathbf{s}}_{i,m,k}^{*} - \mathbf{s}_{i,m}$ that measures the error between the desired signal $\mathbf{s}_{i,m}$ and the minimum MSE estimate $\hat{\mathbf{s}}_{i,m,k}^{*}$ of $\mathbf{s}_{i,m}$ at the MS (i,k). We recall that MS (i,k) decodes $\mathbf{s}_{i,m}$ based on the signal $\mathbf{y}_{i,k}^{(\pi_{i,k}^{-1}(m))}$ obtained via the successive interference cancellation procedure discussed above. The following relationship is well known to exist between the maximum achievable rate $R_{i,m,k}^{*}$ that can be decoded at MS (i,k) for message m and the minimum MSE covariance matrix:

$$R_{i,m,k}^{*}(\mathbf{V},\mathbf{H}) = \frac{1}{L}\log\det\left(\mathbf{E}_{i,m,k}^{*-1}(\mathbf{V},\mathbf{H})\right), \tag{10}$$

where we have made explicit the dependence of the rate $R_{i,m,k}^{*}$ and the minimum MSE covariance matrix $\mathbf{E}_{i,m,k}^{*}$ on the precoding matrices $\mathbf{V}$ and the channel matrices $\mathbf{H} \triangleq \{\mathbf{H}_{i,k,j}\}_{i,j\in\mathcal{B},k\in\mathcal{K}_i}$.

We now derive the minimum MSE covariance matrix $\mathbf{E}_{i,m,k}^{*}(\mathbf{V},\mathbf{H})$. To this end, we introduce the matrix $\mathbf{U}_{i,m,k}^{*} \in \mathbb{C}^{LN_{R,(i,k)}\times d_{i,m}}$, which describes the minimum MSE equalizer applied by MS (i,k) on $\mathbf{y}_{i,k}^{(\pi_{i,k}^{-1}(m))}$ to estimate the signal $\mathbf{s}_{i,m}$ for $n \in \{1,\ldots,|\mathcal{S}_{i,k}|\}$, as $\hat{\mathbf{s}}_{i,m,k}^{*} = \mathbf{U}_{i,m,k}^{*\dagger}\mathbf{y}_{i,k}^{(\pi_{i,k}^{-1}(m))}$, namely

$$\mathbf{U}_{i,m,k}^{*} = \left(\boldsymbol{\Omega}_{i,m,k} + \mathbf{H}_{i,k,i}\mathbf{V}_{i,m}\mathbf{V}_{i,m}^{\dagger}\mathbf{H}_{i,k,i}^{\dagger}\right)^{-1}\mathbf{H}_{i,k,i}\mathbf{V}_{i,m}. \tag{11}$$

In (11), we have defined

$$\begin{aligned}\boldsymbol{\Omega}_{i,m,k} \triangleq \boldsymbol{\Sigma}_{i,k} + &\sum_{q\in\mathcal{M}_i\setminus\{m,\pi_{i,k}(1),\ldots,\pi_{i,k}(\pi_{i,k}^{-1}(m)-1)\}} \mathbf{H}_{i,k,i}\mathbf{V}_{i,q}\mathbf{V}_{i,q}^{\dagger}\mathbf{H}_{i,k,i}^{\dagger} \\ &+ \sum_{j\in\mathcal{B}\setminus\{i\}}\sum_{q\in\mathcal{M}_j}\mathbf{H}_{i,k,j}\mathbf{V}_{j,q}\mathbf{V}_{j,q}^{\dagger}\mathbf{H}_{i,k,j}^{\dagger},\end{aligned} \tag{12}$$

which denotes the covariance matrix of the interference-plus-noise signals when MS (i,k) decodes the signal $\mathbf{s}_{i,m}$.

To evaluate the minimum MSE covariance matrix $\mathbf{E}_{i,m,k}^{*}(\mathbf{V},\mathbf{H})$, it is then useful to define the MSE matrix $\mathbf{E}_{i,m,k}(\mathbf{V},\mathbf{U},\mathbf{H}) \triangleq \mathbb{E}[(\mathbf{U}_{i,m,k}^{\dagger}\mathbf{y}_{i,k}^{(\pi_{i,k}^{-1}(m))} - \mathbf{s}_{i,m})(\mathbf{U}_{i,m,k}^{*}\mathbf{y}_{i,k}^{(\pi_{i,k}^{-1}(m))} - \mathbf{s}_{i,m})^{\dagger}]$ obtained for a generic matrix $\mathbf{U}_{i,m,k}$, which can be written as (see, e.g., [25])

$$\begin{aligned}\mathbf{E}_{i,m,k}(\mathbf{V},\mathbf{U},\mathbf{H}) = &\sum_{j\in\mathcal{B}}\sum_{q\in\mathcal{M}_{i,m,k,j}}\left(\mathbf{U}_{i,m,k}^{\dagger}\mathbf{H}_{i,k,j}\mathbf{V}_{j,q} - \delta_{(i,m),(j,q)}\mathbf{I}\right) \\ &\cdot\left(\mathbf{V}_{j,q}^{\dagger}\mathbf{H}_{i,k,j}^{\dagger}\mathbf{U}_{i,m,k} - \delta_{(i,m),(j,q)}\mathbf{I}\right) + \mathbf{U}_{i,m,k}^{\dagger}\boldsymbol{\Sigma}_{i,k}\mathbf{U}_{i,m,k},\end{aligned} \tag{13}$$

where we have defined the variable $\mathbf{U} \triangleq \{\mathbf{U}_{i,m,k}\}_{i\in\mathcal{B},m\in\mathcal{M}_i,k\in\mathcal{D}_{i,m}}$, and the set $\mathcal{M}_{i,m,k,j} = \mathcal{M}_i \setminus \{\pi_{i,k}(1),\ldots,\pi_{i,k}(\pi_{i,k}^{-1}(m)-1)\}$ if $i=j$ and $\mathcal{M}_{i,m,k,j} = \mathcal{M}_j$ otherwise. The MSE matrix $\mathbf{E}_{i,m,k}^{*}(\mathbf{V},\mathbf{H})$ is then obtained as $\mathbf{E}_{i,m,k}^{*}(\mathbf{V},\mathbf{H}) = \mathbf{E}_{i,m,k}(\mathbf{V},\mathbf{U}^{*},\mathbf{H})$, using the optimal decoding matrices $\mathbf{U}_{i,m,k} = \mathbf{U}_{i,m,k}^{*}$ in (11).

B. Problem Formulation

In this section, we formulate the problem of maximizing the worst-case sum-rate, where the worst case is evaluated over all possible CSI error matrices $\{\mathbf{\Delta}_{i,k,j}(l) \in \mathcal{U}_{i,k,j}(l)\}_{l\in\mathcal{L},i,j\in\mathcal{B},k\in\mathcal{K}_i}$ within the uncertainty sets (9). The optimization is done over the precoding matrices $\mathbf{V}$ for fixed decoding orders $\{\pi_{i,k}\}_{i\in\mathcal{B},k\in\mathcal{K}_i}$ at the MSs. The problem is stated as

$$\underset{\mathbf{V},R}{\text{maximize}} \sum_{i\in\mathcal{B}} \sum_{m\in\mathcal{M}_i} R_{i,m} \tag{14a}$$

$$\text{s.t. } R_{i,m} \leq \min_{\{\mathbf{\Delta}_{i,k,j}(l)\in\mathcal{U}_{i,k,j}(l)\}_{l\in\mathcal{L},j\in\mathcal{B}}} R^{*}_{i,m,k}(\mathbf{V}, \hat{\mathbf{H}} + \mathbf{\Delta}) \tag{14b}$$

$$\text{for all } i \in \mathcal{B},\ m \in \mathcal{M}_i,\ k \in \mathcal{D}_{i,m},$$

$$\sum_{m\in\mathcal{M}_i} \text{tr}\left(\mathbf{V}_{i,m}\mathbf{V}^{\dagger}_{i,m}\right) \leq LP_i, \text{ for all } i \in \mathcal{B}, \tag{14c}$$

where we have defined the variable $R \triangleq \{R_{i,m}\}_{i\in\mathcal{B},m\in\mathcal{M}_i}$. The constraint (14b) imposes that each signal $\mathbf{s}_{i,m}$ is decodable by the destination MSs $k \in \mathcal{D}_{i,m}$ for all possible error matrices $\{\mathbf{\Delta}_{i,k,j}(l) \in \mathcal{U}_{i,k,j}(l)\}_{l\in\mathcal{L},i,j\in\mathcal{B},k\in\mathcal{K}_i}$; and the condition (14c) corresponds to the transmit power constraint (14c). Note that solving problem (14) is difficult due to its non-convexity even in the presence of perfect CSI, i.e., with $\{\varepsilon_{i,k,j}(l) = 0\}_{l\in\mathcal{L},i,j\in\mathcal{B},k\in\mathcal{K}_i}$.

Remark 1. Some special cases of the problem (14) were studied in prior works. Specifically, the unicast system, i.e., $\mathcal{D}_{i,m} = \{m\}$ for $m \in \mathcal{K}_i$, with single-antenna MSs, i.e., $N_R = 1$, and no time-frequency extension, i.e., $L = 1$, was treated in [8], while the broadcast set-up with a single user per cell, i.e., $K_i = 1$, and $L = 1$ was studied in [12], and the single-cell broadcast system, i.e., $B = 1$, with $L = 1$ was investigated in [9]. Moreover, the single-cell set-up, i.e., $B = 1$, with a multicast message, i.e., $\mathcal{D}_{i,1} = \mathcal{K}_i$, and $L = 1$ was studied in [18][20] assuming that perfect CSI is available at BS, i.e., $\varepsilon_{i,k,j}(l) = 0$, and the study was extended in [21] to the multi-cell scenario. The algorithm proposed here hence unifies and generalizes [8][9][12][18][20][21].

C. Robust Sum-Rate Optimization

In this subsection, we present the proposed iterative algorithm to tackle the problem (14). We first apply Fenchel duality, as summarized in Lemma 3 in Appendix **??**, to the constraint (14b), so as to restate it in the equivalent form

$$R_{i,m} \leq \min_{\{\mathbf{\Delta}_{i,k,j}(l)\in\mathcal{U}_{i,k,j}(l)\}_{l\in\mathcal{L},j\in\mathcal{B}}} \max_{\mathbf{S}_{i,m,k}\succeq\mathbf{0},\mathbf{U}_{i,m,k}} \left\{ \begin{array}{c} -\text{tr}\Big(\mathbf{S}_{i,m,k}\mathbf{E}_{i,m,k}(\mathbf{V},\mathbf{U},\hat{\mathbf{H}}+\mathbf{\Delta})\Big) \\ +\log\det\left(\mathbf{S}_{i,m,k}\right) + d_{i,m} \end{array} \right\}. \tag{15}$$

Then, in order to make the problem more tractable, we obtain a lower bound on the optimal value of the problem (14) by exchanging the order of the min and the max operations in the constraint (15), as done in [12, Sec. III], leading to the problem

$$\underset{\mathbf{V},R,\mathbf{U},\mathbf{S}\succeq\mathbf{0}}{\text{maximize}} \sum_{i\in\mathcal{B}}\sum_{m\in\mathcal{M}_i} R_{i,m} \tag{16a}$$

$$\text{s.t. } R_{i,m} \leq \min_{\{\boldsymbol{\Delta}_{i,k,j}(l)\in\mathcal{U}_{i,k,j}(l)\}_{l\in\mathcal{L},j\in\mathcal{B}}} \left\{ \begin{array}{c} -\mathrm{tr}\Big(\mathbf{S}_{i,m,k}\mathbf{E}_{i,m,k}(\mathbf{V},\mathbf{U},\hat{\mathbf{H}}+\boldsymbol{\Delta})\Big) \\ +\log\det\left(\mathbf{S}_{i,m,k}\right)+d_{i,m} \end{array} \right\},$$

$$\text{for all } i\in\mathcal{B},\ m\in\mathcal{M}_i,\ k\in\mathcal{D}_{i,m}, \tag{16b}$$

$$\sum_{m\in\mathcal{M}_i} \mathrm{tr}\left(\mathbf{V}_{i,m}\mathbf{V}_{i,m}^{\dagger}\right) \leq LP_i,\ \text{for all } i\in\mathcal{B}, \tag{16c}$$

where we have defined the variable $\mathbf{S} \triangleq \{\mathbf{S}_{i,m,k}\}_{i\in\mathcal{B},m\in\mathcal{M}_i,k\in\mathcal{D}_{i,m}}$. To see that the optimal value of (16) lower-bounds that of (14), with (15) in lieu of (14b), observe that, unlike (14), in (16) the variables $\mathbf{U}$ and $\mathbf{S}$ cannot be chosen as a function of the error matrices $\boldsymbol{\Delta}_{i,k,j}(l)$.

Problem (16) is made complicated by the nested minimization with respect to the uncertainty variables $\boldsymbol{\Delta}_{i,k,j}(l)$. Since this problem appears to be prohibitively complex, the following lemma provides an alternative problem which does not include the matrices $\boldsymbol{\Delta}_{i,k,j}(l)$ and whose optimal solution lower-bounds that of (16).

Lemma 1. *The optimal value of the problem*

$$\underset{\substack{\mathbf{V},\mathbf{U},\tilde{\mathbf{S}}\succeq\mathbf{0};\\ (R,\gamma,\tau,\mu)\geq 0}}{\text{maximize}} \quad \sum_{i\in\mathcal{B}}\sum_{m\in\mathcal{M}_i} R_{i,m} \tag{17a}$$

$$\text{s.t. } LR_{i,m} \leq -\gamma_{i,m,k} + 2\log\det\left(\tilde{\mathbf{S}}_{i,m,k}\right) + d_{i,m},$$

$$\text{for all } i\in\mathcal{B},\ m\in\mathcal{M}_i,\ k\in\mathcal{D}_{i,m}, \tag{17b}$$

$$\gamma_{i,m,k} \geq \sum_{j\in\mathcal{B}} \tau_{i,m,k,j} + \left\|\boldsymbol{\Sigma}_{i,k}^{1/2}\mathbf{U}_{i,m,k}\tilde{\mathbf{S}}_{i,m,k}\right\|_F^2,$$

$$\text{for all } i\in\mathcal{B},\ m\in\mathcal{M}_i,\ k\in\mathcal{D}_{i,m}, \tag{17c}$$

$$\begin{bmatrix} \tau_{i,m,k,j} - \sum_{l\in\mathcal{L}}\mu_{i,m,k,j}(l) & \mathbf{c}_{i,m,k,j}^{\dagger} & \mathbf{0} \\ \mathbf{c}_{i,m,k,j} & \mathbf{I} & -\mathbf{C}_{i,m,k,j} \\ \mathbf{0} & -\mathbf{C}_{i,m,k,j}^{\dagger} & \mathrm{diag}(\{\mu_{i,m,k,j}(l)\}_{l\in\mathcal{L}})\otimes\mathbf{I} \end{bmatrix} \succeq \mathbf{0},$$

$$\text{for all } i,j\in\mathcal{B},\ m\in\mathcal{M}_i,\ k\in\mathcal{D}_{i,m}, \tag{17d}$$

$$\sum_{m\in\mathcal{M}_i} \mathrm{tr}\left(\mathbf{V}_{i,m}\mathbf{V}_{i,m}^{\dagger}\right) \leq LP_i,\ \text{for all } i\in\mathcal{B}, \tag{17e}$$

provides a lower bound on the optimal value of problem (16), where we have defined the variables $\tilde{\mathbf{S}} \triangleq \{\tilde{\mathbf{S}}_{i,m,k}\}_{i\in\mathcal{B},m\in\mathcal{M}_i,k\in\mathcal{D}_{i,m}}$, $\gamma \triangleq \{\gamma_{i,m,k}\}_{i\in\mathcal{B},m\in\mathcal{M}_i,k\in\mathcal{D}_{i,m}}$, $\tau \triangleq \{\tau_{i,m,k,j}\}_{i,j\in\mathcal{B},m\in\mathcal{M}_i,k\in\mathcal{D}_{i,m}}$,

$\mu \triangleq \{\mu_{i,m,k,j}(l)\}_{l\in\mathcal{L},i\in\mathcal{B},m\in\mathcal{M}_i,k\in\mathcal{D}_{i,m}}$ *and the notations*

$$\tilde{\mathbf{S}}_{i,m,k} \triangleq \mathbf{S}_{i,m,k}^{1/2}, \tag{18}$$

$$\mathbf{c}_{i,m,k,j,q} \triangleq \mathrm{vec}\left(\tilde{\mathbf{S}}_{i,m,k}^{\dagger}(\mathbf{U}_{i,m,k}^{\dagger}\hat{\mathbf{H}}_{i,k,j}\mathbf{V}_{j,q} - \delta_{(i,m),(j,q)}\mathbf{I})\right), \tag{19}$$

$$\mathbf{c}_{i,m,k,j} \triangleq \left[\mathbf{c}_{i,m,k,j,\mathcal{M}_{i,m,k,j}(1)}; \ldots; \mathbf{c}_{i,m,k,j,\mathcal{M}_{i,m,k,j}(|\mathcal{M}_{i,m,k,j}|)}\right], \tag{20}$$

$$\mathbf{C}_{i,m,k,j,q}(l) \triangleq \mathbf{V}_{j,q}^{T}(l) \otimes (\tilde{\mathbf{S}}_{i,m,k}^{\dagger}\mathbf{U}_{i,m,k}^{\dagger}(l)), \tag{21}$$

$$\mathbf{C}_{i,m,k,j}(l) \triangleq \left[\mathbf{C}_{i,m,k,j,\mathcal{M}_{i,m,k,j}(1)}(l); \ldots; \mathbf{C}_{i,m,k,j,\mathcal{M}_{i,m,k,j}(|\mathcal{M}_{i,m,k,j}|)}(l)\right], \tag{22}$$

$$\text{and } \mathbf{C}_{i,m,k,j} \triangleq -\left[\varepsilon_{i,k,j}(1)\mathbf{C}_{i,m,k,j}(1), \ldots, \varepsilon_{i,k,j}(L)\mathbf{C}_{i,m,k,j}(L)\right]. \tag{23}$$

In (21), the matrices $\mathbf{U}_{i,m,k}(l)$ and $\mathbf{V}_{i,m}(l)$ represent the submatrices of $\mathbf{U}_{i,m,k}$ and $\mathbf{V}_{i,m}$, respectively, corresponding to the lth slot. Moreover, the problem (17) is equivalent to the problem (16) when the precoding is performed over a single time-frequency slot, i.e., $L = 1$.

Proof. See Appendix **??**. □ □

Problem (17) is still not convex with respect to the optimization variables. However, the problem of optimizing one of three sets of variables $\mathbf{V}$, $\mathbf{U}$ and $\tilde{\mathbf{S}}$ along with the remaining variables R, γ, τ and μ when fixing the other two sets can be seen to be convex using standard arguments [29]. Based on this observation, we propose an alternating optimization algorithm as summarized in Table Algorithm 1. Note that the subproblems in Step 2-4 can be solved using standard convex optimization tools (e.g., [29]).

Remark 2. The proposed algorithm, being based on alternating optimization, is guaranteed to provide a sequence of feasible solutions with non-decreasing objective function, hence guaranteeing the convergence of Algorithm 1. The effectiveness of the approach, which is based on optimizing the problem (17) in lieu of (14) will be validated via numerical results.

Remark 3. The complexity of solving convex problems is polynomial in the size of the unknowns and hence the same complexity order is inherited by each iteration of Algorithm 1. The speed of convergence will be discussed via numerical results in Sec. V.

D. Distributed Implementation

The discussion above assumes that a single control unit is available to perform the optimization. In the special case of the MIMO-IC, we will now discuss how Algorithm 1 can be also performed in a distributed fashion. The extension to the general MIMO-IBC is left as future work. Similar to, e.g., [25], this decentralization is done by assigning different subtasks to distinct processors that are allowed to communicate with one another. We focus on the case $L = 1$ in order to simplify the presentation, although the extension to any L is straightforward. Unlike [25] and references therein, which consider

Algorithm 1 Iterative Algorithm for problem (17)

Step 1. Initialize the matrices $\mathbf{V}^{(1)}$ and $\tilde{\mathbf{S}}^{(1)}$ to arbitrary feasible matrices for problem (17) and set $t = 1$.

Step 2. Solve the problem (17) with respect to the variables $\mathbf{U}$, R, γ, τ and μ for fixed variables $\mathbf{V} = \mathbf{V}^{(t)}$ and $\tilde{\mathbf{S}} = \tilde{\mathbf{S}}^{(t)}$, and update the matrices $\mathbf{U}^{(t+1)}$ as a solution of this problem.

Step 3. Solve the problem (17) with respect to the variables $\tilde{\mathbf{S}}$, R, γ, τ and μ for fixed variables $\mathbf{U} = \mathbf{U}^{(t+1)}$ and $\mathbf{V} = \mathbf{V}^{(t)}$, and update the matrices $\tilde{\mathbf{S}}^{(t+1)}$ as a solution of this problem.

Step 4. Solve the problem (17) with respect to the variables $\mathbf{V}$, R, γ, τ and μ for fixed variables $\mathbf{U} = \mathbf{U}^{(t+1)}$ and $\tilde{\mathbf{S}} = \tilde{\mathbf{S}}^{(t+1)}$, and update the matrices $\mathbf{V}^{(t+1)}$ as a solution of this problem.

Step 5. Stop if a convergence criterion on the objective function is satisfied. Otherwise, set $t \leftarrow t + 1$ and go back to Step 2.

perfect CSI, the robust problem (17) requires some additional steps in order to be made distributed. For this purpose, as discussed below, we will resort to ADMM [27][30]. Throughout this section, we use the simplified notation introduced for the MIMO-IC in Sec. II. Accordingly, the problem (17) reduces to

$$\underset{\substack{\{\mathbf{U}_i, \mathbf{V}_i, \tilde{\mathbf{S}}_i \succeq \mathbf{0},\\ (R_i, \gamma_i, \tau_{j,i}, \mu_{j,i}) \geq 0\}_{i,j\in\mathcal{B}}}}{\text{maximize}} \quad \sum_{i\in\mathcal{B}} R_i \tag{24a}$$

$$\text{s.t. } R_i \leq -\gamma_i + 2\log\det\left(\tilde{\mathbf{S}}_i\right) + d_i, \text{for all } i \in \mathcal{B}, \tag{24b}$$

$$\gamma_j \geq \sum_{k\in\mathcal{B}} \tau_{j,k} + \left\|\mathbf{\Sigma}_j^{1/2}\mathbf{U}_j\tilde{\mathbf{S}}_j\right\|_F^2, \text{ for all } j \in \mathcal{B}, \tag{24c}$$

$$\begin{bmatrix} \tau_{j,i} - \mu_{j,i} & \mathbf{c}_{j,i}^\dagger & \mathbf{0} \\ \mathbf{c}_{j,i} & \mathbf{I} & -\mathbf{C}_{j,i} \\ \mathbf{0} & -\mathbf{C}_{j,i}^\dagger & \mu_{j,i}\mathbf{I} \end{bmatrix} \succeq \mathbf{0}, \text{ for all } i, j \in \mathcal{B},$$

$$\text{tr}\left(\mathbf{V}_i\mathbf{V}_i^\dagger\right) \leq P_i \text{ for all } i \in \mathcal{B}; \tag{24d}$$

where $\mathbf{C}_{j,i} \triangleq \mathbf{V}_i^\dagger \otimes \left(\tilde{\mathbf{S}}_j\mathbf{U}_j^\dagger\right)$ and $\mathbf{c}_{j,i} \triangleq \text{vec}\left(\tilde{\mathbf{S}}_j\left(\mathbf{U}_j^\dagger\hat{\mathbf{H}}_{j,i}\mathbf{V}_i - \delta_{j,i}\mathbf{I}\right)\right)$.

In the distributed implementation under study, there are $2B$ processors, one for each BS and one for each MS, where each processor needs only be informed about the CSI concerning the associated BS or MS. For instance, the processor associated with BS i should only be informed about the local CSI $\hat{\mathbf{H}}_{j,i}$ for all $j \in \mathcal{B}$. Moreover, the processor associated with BS i produces the precoding matrix $\mathbf{V}_i$, while the processor corresponding to MS i yields the matrix $\mathbf{U}_i$. This division of the optimization task into distinct processors enables, on the one hand, a parallel and modular implementation at the control unit and, on the other hand, a distributed implementation across BSs and MSs. As mentioned, the processors operate by exchanging messages in an iterative fashion until convergence. In the following, we refer to the processor associated to a BS or an MS simply with the index of the BS or MS. The schedule of the subtasks and of the exchanged messages, with reference to Table Algorithm 1, is as follows.

1) Step 2 and Step 3: After initialization (Step 1), Step 2 in Table Algorithm 1 is carried out in parallel by all MSs. Specifically, each MS $i \in \mathcal{B}$ optimizes over the variables $\mathbf{U}_i$, R_i, γ_i, $\tau_{i,j}$ and $\mu_{i,j}$ for all $j \in \mathcal{B}$. It can be seen that this parallelization comes at no loss of optimality, since the variables pertaining to each MS i appear in different constraints in problem (24). Moreover, it is easily verified that the terms in (24) that are relevant for MS i only depend on the local CSI $\hat{\mathbf{H}}_{i,j}$ for all $i \in \mathcal{B}$. Similarly, Step 3) can be carried out in parallel by each MS by optimizing over variables $\tilde{\mathbf{S}}_i$, R_i, γ_i, $\tau_{i,j}$ and $\mu_{i,j}$ for all $j \in \mathcal{B}$. At the end of this step, each MS i sends the obtained matrices $\mathbf{U}_i$ and $\tilde{\mathbf{S}}_i$ to all BSs.

2) Step 4: Following the approach above, we would like to parallelize Step 4 across the BSs. Specifically, each BS i should optimize over the variables $\mathbf{V}_i$, R_i, γ_i, $\boldsymbol{\tau}_i$ =$[\boldsymbol{\tau}_{1,i} ... \boldsymbol{\tau}_{K,i}]^T$ and $\mu_{j,i}$ for all $j \in \mathcal{B}$. However, this turns out to be not directly possible due to the constraints (24c), which couple the variables $\boldsymbol{\tau}_i$ of different BSs i. To solve this problem, first observation is that, at the optimum solution for Step 4, the constraints (24c) should be satisfied with equality[2]. Therefore, the coupling constraint (24c) can be written as $\sum_{i \in \mathcal{B}} \boldsymbol{\tau}_i = \boldsymbol{\gamma} - \mathbf{u}$, where we have defined $\boldsymbol{\gamma} = [\gamma_1 ... \gamma_K]^T$ and $\mathbf{u} = [||\boldsymbol{\Sigma}_1^{1/2} \mathbf{U}_1 \tilde{\mathbf{S}}_1||_F^2 ... ||\boldsymbol{\Sigma}_K^{1/2} \mathbf{U}_K \tilde{\mathbf{S}}_K||_F^2]^T$. Linear equality constraints such as this one can be handled in a distributed fashion by the ADMM algorithm, which is known to convergence to the optimal solution of the original problem [27][30][3]. We review the ADMM algorithm for the problem at hand in the rest of this section.

Define the augmented Lagrangian function for BS i

$$L_i(R_i, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\lambda}) = R_i + \boldsymbol{\lambda}^T (\boldsymbol{\tau}_i - \boldsymbol{\gamma}) + \frac{\rho}{2} \left\| \sum_{j \in \mathcal{B}} \boldsymbol{\tau}_j - \boldsymbol{\gamma} + \mathbf{u} \right\|^2, \tag{25}$$

where $\rho \geq 0$ is a parameter of the algorithm. We will use the notation $\boldsymbol{\tau}_{\sim i}$ to denote all vectors $\boldsymbol{\tau}_j$ with $j \neq i$ and similarly for $\boldsymbol{\gamma}_{\sim i}$. Following ADMM, the distributed implementation of step 4 of Algorithm

[2]If the constraints were not satisfied with equality for some optimal solution, it would be possible to decrease at least one γ_i for some i, without violating the constraints. This would in turn allow a larger value of the corresponding rate R_i to be selected, which would result in a contradiction.

[3]For a review of the convergence properties of the ADMM algorithm, we refer to [27] for the case $K = 2$ and to [30] for the general case.

Algorithm 2 ADMM implementation of step 4 of Algorithm 1

Step 1. For each BS $i = 1, ..., K$, solve the convex problem (26), where $\mathbf{U} = \mathbf{U}^{(t+1)}$ and $\tilde{\mathbf{S}} = \tilde{\mathbf{S}}^{(t+1)}$, and update $\boldsymbol{\gamma}^{[m-1]} \leftarrow [\gamma_i^*, \boldsymbol{\gamma}_{\sim i}^{[m-1]}]$ and $\boldsymbol{\tau}^{[m-1]} \leftarrow [\boldsymbol{\tau}_i^*, \boldsymbol{\tau}_{\sim i}^{[m-1]}]$, with γ_i^* and $\boldsymbol{\tau}_i^*$ equal to the calculated optimal values.

Step 2. Set $\boldsymbol{\tau}^{[m]} \leftarrow \boldsymbol{\tau}^{[m-1]}$ and $\boldsymbol{\gamma}^{[m]} \leftarrow \boldsymbol{\gamma}^{[m-1]}$, and update the Lagrangian vector as

$$\boldsymbol{\lambda}^{(m)} = \boldsymbol{\lambda}^{(m-1)} + \rho \left(\sum_{j \in \mathcal{B}} \boldsymbol{\tau}_j^{[m]} - \boldsymbol{\gamma}^{[m]} + \mathbf{u} \right). \tag{27}$$

1 is obtained as summarized in Table Algorithm 2, where we have defined the problem

$$\underset{\mathbf{V}_i, (R_i, \gamma_i, \boldsymbol{\tau}_i, \{\mu_{j,i}\}_{j \in \mathcal{B}}) \geq 0}{\text{maximize}} \quad L_i(R_i, [\boldsymbol{\tau}_i, \boldsymbol{\tau}_{\sim i}^{[m-1]}], [\gamma_i, \boldsymbol{\gamma}_{\sim i}^{[m-1]}], \boldsymbol{\lambda}^{[m-1]}) \tag{26a}$$

$$\text{s.t. } R_i \leq -\gamma_i + 2 \log \det \left(\tilde{\mathbf{S}}_i \right) + d_i, \tag{26b}$$

$$\begin{bmatrix} \tau_{j,i} - \mu_{j,i} & \mathbf{c}_{j,i}^\dagger & \mathbf{0} \\ \mathbf{c}_{j,i} & \mathbf{I} & -\mathbf{C}_{j,i} \\ \mathbf{0} & -\mathbf{C}_{j,i}^\dagger & \mu_{j,i} \mathbf{I} \end{bmatrix} \succeq \mathbf{0}, \text{ for all } j \in \mathcal{B}, \tag{26c}$$

$$\text{tr} \left(\mathbf{V}_i \mathbf{V}_i^\dagger \right) \leq P_i. \tag{26d}$$

Note that the problem (26) can be solved at BS i based only on local CSI. Moreover, in Algorithm 2, the BSs only need to exchange information about the solutions γ_i^* and $\boldsymbol{\tau}_i^*$ obtained at each step, since this information is required to solve problem (26) and to update the Lagrangian vector according to (27).

IV. Robust MSE Optimization of Linear Precoding and DF Equalization

In the previous section, an underlying assumption was that each MS performs information-theoretically optimum decoding of all the streams intended for it. In this section, instead, we assume that each MS implements a simpler symbol-by-symbol (i.e., per-channel use) DF equalizer. For generality, the DF equalizer is allowed to decode and cancel not only the streams intended for the given MS but also an arbitrary subset of interfering signals. A key differentiating feature of the proposed approach with respect to existing art is that we tackle the *joint* optimization of linear precoding and DF equalization. We will focus here on the MIMO-IC and we will set $L = 1$. This will be done in order to simplify the discussion but the extension to the more general MIMO-IBC could be done by generalizing in a relatively straightforward way the analysis covered here. We start by reviewing DF equalization in the

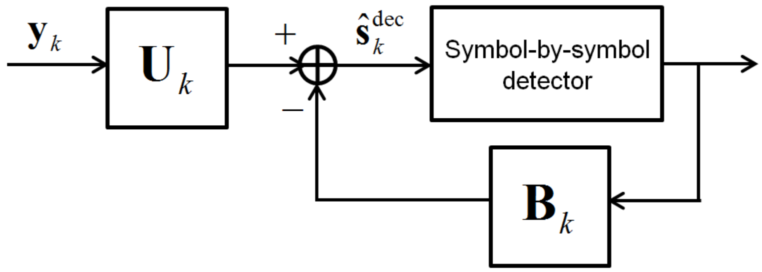

Figure 3. Illustration of the DF equalizer at receiver k.

context of the MIMO-IC in the next subsection. Then, we formulate the problem in Sec. IV-C, propose a solution in Sec. IV-D and discuss a distributed implementation in Sec. IV-E.

A. DF Equalization

As shown in Fig. 3, each MS k performs symbol-by-symbol DF equalization with a feedforward linear equalization matrix $\mathbf{U}_k \in \mathbb{C}^{d_k^{\mathrm{dec}} \times N_{R,k}}$ and a feedback matrix $\mathbf{B}_k \in \mathcal{L}^{d_k^{\mathrm{dec}}}$, where $\mathcal{L}^N$ is the set of strictly lower triangular matrices in $\mathbb{C}^{N \times N}$ and d_k^{dec} is the total number of data streams decoded by receiver k. Specifically, receiver k decodes the vector of signals $\mathbf{s}_k^{\mathrm{dec}} \triangleq [\mathbf{s}_{q_1^k}; \ldots; \mathbf{s}_{q^k_{|\mathcal{I}_k^{\mathrm{dec}}|}}; \mathbf{s}_k] \in \mathbb{C}^{d_k^{\mathrm{dec}} \times 1}$, where we have defined the set $\mathcal{I}_k^{\mathrm{dec}} = \{q_1^k, \ldots, q^k_{|\mathcal{I}_k^{\mathrm{dec}}|}\} \subseteq \mathcal{I}_k \triangleq \mathcal{K} \setminus \{k\}$ of the indices of the interfering signals to be decoded and canceled prior to the desired signal $\mathbf{s}_k$ and $d_k^{\mathrm{dec}} \triangleq d_k + \sum_{j \in \mathcal{I}_k^{\mathrm{dec}}} d_j$. According to the structure of the matrix $\mathbf{B}_k$, cancellation is performed stream-by-stream starting from the top of vector $\mathbf{s}_k^{\mathrm{dec}}$ and hence ending with the desired signal $\mathbf{s}_k$, see, e.g., [28].

Similar to [28] (see also references therein), assuming that decoding of all the previously decoded streams is successful, the signals $\hat{\mathbf{s}}_k^{\mathrm{dec}}$ at the input of the decision device can be written as

$$\hat{\mathbf{s}}_k^{\mathrm{dec}} = \left(\mathbf{U}_k \bar{\mathbf{H}}_k - \mathbf{B}_k\right) \mathbf{s}_k^{\mathrm{dec}} + \mathbf{U}_k \sum_{j \in \mathcal{I}_k \setminus \mathcal{I}_k^{\mathrm{dec}}} \mathbf{H}_{k,j} \mathbf{V}_j \mathbf{s}_j + \mathbf{U}_k \mathbf{z}_k, \tag{28}$$

where the matrix $\bar{\mathbf{H}}_k$ is defined as

$$\bar{\mathbf{H}}_k \triangleq \left[(\mathbf{H}_{k,q_1^k} \mathbf{V}_{q_1^k}) \ldots (\mathbf{H}_{k,q^k_{|\mathcal{I}_k^{\mathrm{dec}}|}} \mathbf{V}_{q^k_{|\mathcal{I}_{i,k}^{\mathrm{dec}}|}}) \, (\mathbf{H}_{k,k} \mathbf{V}_k) \right]. \tag{29}$$

B. Sum-MSE

As done in [28] and related references therein, we consider the MSE at the output of the DF equalizer as the performance metric used to optimize the precoding matrices $\mathbf{V}$, the feedback equalization matrices $\mathbf{U}$ and the feedback matrices $\mathbf{B} \triangleq \{\mathbf{B}_k\}_{k \in \mathcal{K}}$. The MSE matrix is given as $\mathbf{E}_k \triangleq \mathbb{E}[\mathbf{e}_k \mathbf{e}_k^\dagger]$ with the error signal being defined as $\mathbf{e}_k \triangleq \hat{\mathbf{s}}_k^{\mathrm{dec}} - \mathbf{s}_k^{\mathrm{dec}}$. With the estimate (28) and for given matrices $(\mathbf{V}, \mathbf{U}, \mathbf{C}, \mathbf{H})$, with $\mathbf{C} \triangleq \{\mathbf{C}_k = \mathbf{I} + \mathbf{B}_k\}_{k\in\mathcal{K}}$, the MSE matrix $\mathbf{E}_k^{\mathrm{DFE}}(\mathbf{V},\mathbf{U},\mathbf{C},\mathbf{H})$, similar to [28], can be seen to be equal to

$$\begin{aligned}\mathbf{E}_k^{\mathrm{DFE}}(\mathbf{V},\mathbf{U},\mathbf{C},\mathbf{H}) = \left(\mathbf{U}_k\bar{\mathbf{H}}_k - \mathbf{C}_k\right)\left(\bar{\mathbf{H}}_k^{\dagger}\mathbf{U}_k^{\dagger} - \mathbf{C}_k^{\dagger}\right) \\ + \sum_{j\in\mathcal{I}_k\setminus\mathcal{I}_k^{\mathrm{dec}}} \mathbf{U}_k\mathbf{H}_{k,j}\mathbf{V}_j\mathbf{V}_j^{\dagger}\mathbf{H}_{k,j}^{\dagger}\mathbf{U}_k^{\dagger} + \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{U}_k^{\dagger}.\end{aligned} \tag{30}$$

We remark that the MSE function $\mathbf{E}_k^{\mathrm{DFE}}(\mathbf{V},\mathbf{U},\mathbf{C},\mathbf{H})$ is convex with respect to one of two variables $\mathbf{V}$ and $\mathbf{U}$, along with the remaining variable $\mathbf{C}$, if we fix the other one. This property will be used to derive an iterative alternating algorithm below.

C. Problem Formulation

In this section, we tackle the problem of minimizing the worst-case sum of the MSEs over all possible error matrices $\{\mathbf{\Delta}_{k,j} \in \mathcal{U}_{k,j}\}_{k,j\in\mathcal{K}}$ within the uncertainty sets over the linear precoding and DF equalization matrices. The problem is stated as

$$\underset{\mathbf{V}}{\text{minimize}} \sum_{k\in\mathcal{K}} \max_{\{\mathbf{\Delta}_{k,j}\in\mathcal{U}_{k,j}\}_{j\in\mathcal{K}}} \min_{\mathbf{U},\mathbf{C}} \mathrm{tr}\left(\mathbf{E}_k^{\mathrm{DFE}}(\mathbf{V},\mathbf{U},\mathbf{C},\hat{\mathbf{H}}+\mathbf{\Delta})\right) \tag{31a}$$

$$\text{s.t. } \mathrm{tr}\left(\mathbf{V}_j\mathbf{V}_j^{\dagger}\right) \le P_j, \text{ for all } j\in\mathcal{K}, \tag{31b}$$

$$\mathbf{C}_k \in \mathcal{L}^{d_k^{\mathrm{dec}}}, \ \mathrm{diag}(\mathbf{C}_k) = \mathbf{1}, \text{ for all } k\in\mathcal{K}. \tag{31c}$$

The constraint (31b) imposes that the precoding matrices $\mathbf{V}$ satisfy the power constraint in (2). Note that solving problem (31) is difficult due to its non-convexity even in the presence of perfect CSI, i.e., with $\{\varepsilon_{k,j} = 0\}_{k,j\in\mathcal{K}}$.

Remark 4. Some special cases of the problem (31) were studied in prior works. Specifically, the case with a single-user, i.e., $K = 1$, and hence no interference, was studied in [28] and [16] assuming the availability of perfect CSI and imperfect CSI, respectively. The case of multiple users, i.e., $K \ge 2$, was studied in [17], but the optimization was tackled only for the design of the variables $\mathbf{U}$ and $\mathbf{C}$ under the assumption of perfect CSI.

D. Robust Sum-MSE Optimization

In this subsection, we present an iterative algorithm to tackle the problem (31). In order to make the problem more tractable, we obtain a feasible solution for (31) by formulating an alternative problem whose solution provides an upper bound on the optimal MSE of the original problem (31). This is done by exchanging the order of the min and the max operations in the cost function (31a). To see that the optimal value of the resulting problem upper-bounds that of (31), observe that, unlike (31), the variables $\mathbf{U}$ and $\mathbf{C}$ cannot be chosen as a function of the error matrices $\mathbf{\Delta}_{k,j}$. This reduces the scope of the optimization domain and generally leads to suboptimal, but computationally efficient, solutions to problem (31).

The problem at hand is made complicated by the nested maximization with respect to the uncertainty variables $\mathbf{\Delta}_{k,j}$. The following lemma provides an equivalent problem which does not include the matrices $\mathbf{\Delta}_{k,j}$.

Lemma 2. *The problem at hand is equivalent to the problem*

$$\underset{\mathbf{V},\mathbf{U},\mathbf{C},\gamma,\tau,\mu\geq 0}{\text{minimize}} \sum_{k\in\mathcal{K}} \gamma_k \tag{32a}$$

$$\text{s.t. } \operatorname{tr}\left(\mathbf{V}_j\mathbf{V}_j^{\dagger}\right) \leq P_j, \text{ for all } j\in\mathcal{K}, \tag{32b}$$

$$\mathbf{C}_k \in \mathcal{L}^{d_k^{\text{dec}}}, \ \operatorname{diag}(\mathbf{C}_k) = \mathbf{1}, \text{ for all } k\in\mathcal{K}, \tag{32c}$$

$$\gamma_k \geq \sum_{j\in\mathcal{K}} \tau_{k,j} + \left\|\mathbf{U}_k\mathbf{\Sigma}_k^{1/2}\right\|_F^2 \text{ for all } k\in\mathcal{K}, \tag{32d}$$

$$\begin{bmatrix} \tau_{k,j} & \mathbf{w}_{k,j}^{\text{DFE}\dagger} & \mathbf{0} \\ \mathbf{w}_{k,j}^{\text{DFE}} & \mathbf{I} & -\varepsilon_{k,j}\mathbf{W}_{k,j}^{\text{DFE}} \\ \mathbf{0} & -\varepsilon_{k,j}\mathbf{W}_{k,j}^{\text{DFE}\dagger} & \mu_{k,j}\mathbf{I} \end{bmatrix} \succeq \mathbf{0} \text{ for all } k,j\in\mathcal{K}, \tag{32e}$$

where we have defined the notations $\gamma \triangleq \{\gamma_k\}_{k\in\mathcal{K}}$, $\tau \triangleq \{\tau_{k,j}\}_{k,j\in\mathcal{K}}$, $\mu \triangleq \{\mu_{k,j}\}_{k,j\in\mathcal{K}}$, $\mathbf{W}_{k,j}^{\text{DFE}} \triangleq \mathbf{V}_j^T \otimes \mathbf{U}_k$, $\mathbf{C}_k = [\mathbf{C}_{k,i_1^k} \dots \mathbf{C}_{k,i^k_{|\mathcal{I}_k^{\text{dec}}|}} \mathbf{C}_{k,k}]$ *and*

$$\mathbf{w}_{k,j}^{\text{DFE}} \triangleq \begin{cases} \operatorname{vec}(\mathbf{G}_k\hat{\mathbf{H}}_{k,j}\mathbf{V}_j - \mathbf{C}_j), & j\in\{k\}\cup\mathcal{I}_k^{\text{dec}} \\ \operatorname{vec}(\mathbf{G}_k\hat{\mathbf{H}}_{k,j}\mathbf{V}_j), & \text{otherwise} \end{cases}. \tag{33}$$

Proof. Lemma 2 follows from the same arguments used in Appendix **??** to prove Lemma 1 and is not detailed here. □ □

Problem (32) is still not convex with respect to the optimization variables. However, the problem of optimizing one of two variables $\mathbf{V}$ and $\mathbf{U}$, along with the remaining variables $\mathbf{C}$, γ, τ and μ, when fixing the other one can be seen to be convex. Based on this observation, we propose an alternating optimization algorithm as summarized in Table Algorithm 3. Note that the subproblems in Step 2-3 can be solved using standard convex optimization tools (e.g., [29]). The same consideration in Remark 2 and Remark 3 apply here.

E. Distributed Implementation

While the algorithm discussed above for the solution of (32) requires a centralized implementation, it is also possible to implement it using a distributed approach, following the same steps detailed in Sec. III-D, as long as one sets $\mathcal{I}_k^{\text{dec}} = \emptyset$, i.e., if each MS only performs DF equalization on the stream intended for itself. The general case is more challenging and is left as interesting open problem. The

Algorithm 3 Iterative Algorithm for problem (32)

Step 1. Initialize the matrices $\mathbf{V}^{(1)}$ to arbitrary feasible matrices for problem (32) and set $t = 1$.

Step 2. Solve the problem (32) with respect to the variables $\mathbf{U}$, $\mathbf{C}$ γ, τ and μ for fixed variables $\mathbf{V} = \mathbf{V}^{(t)}$, and update the matrices $\mathbf{U}^{(t+1)}$ as a solution of this problem.

Step 3. Solve the problem (32) with respect to the variables $\mathbf{V}$, $\mathbf{C}$ γ, τ and μ for fixed variables $\mathbf{U} = \mathbf{U}^{(t+1)}$, and update the matrices $\mathbf{V}^{(t+1)}$ as a solution of this problem.

Step 4. Stop if a convergence criterion on the objective function is satisfied. Otherwise, set $t \leftarrow t + 1$ and go back to Step 2.

distributed implementation works as follows. After initialization (Step 1), Step 2 in Table Algorithm 3 can be carried out with no loss of optimality in parallel by all MSs with only local CSI, whereby each MS $i \in \mathcal{B}$ optimizes over the variables $\mathbf{U}_i$, $\mathbf{C}_i$ γ_i, $\tau_{i,j}$ and $\mu_{i,j}$ for all $j \in \mathcal{B}$. Then, Step 3 can be solved in a distributed way across the BSs, whereby each BS i optimizes over the variables $\mathbf{V}_i$, $\mathbf{C}_i$, γ_i, $\tau_{j,i}$ and $\mu_{j,i}$ for all $j \in \mathcal{B}$, by means of the ADMM algorithm. The details follow in a straightforward way from the discussion in Sec. III-D and will not be detailed here.

V. Numerical Results

In this section, we present numerical results to illustrate the advantage of the proposed robust designs. We start with the sum-rate performance in the presence of optimal decoding at the MSs and then evaluate the Symbol Error Rate (SER) performance gains of the proposed DF-based design.

A. Sum-Rate Maximization over Linear Precoding

Here, we evaluate the performance of the robust strategy proposed in Sec. III. The following two baseline strategies will be considered for reference:

1) Non-robust design: The precoding matrices $\mathbf{V}$ are designed assuming that there is no uncertainty on the reported channel matrices so that $\hat{\mathbf{H}} = \mathbf{H}$. We refer to this approach as "*non-robust*". This scheme can be implemented via Algorithm 1 by setting $\{\varepsilon_{i,k,j}(l) = 0\}_{l\in\mathcal{L},i,j\in\mathcal{B},k\in\mathcal{K}_i}$.

2) MMSE design: Instead of focusing on the weighted sum-rate maximization, one may be interested in minimizing the worst-case sum of MSEs over all MS, which can be written as

$$\sum_{i\in\mathcal{B}}\sum_{m\in\mathcal{M}_i}\sum_{k\in\mathcal{D}_{i,m}}\max_{\{\mathbf{\Delta}_{i,k,j}(l)\in\mathcal{U}_{i,k,j}(l)\}_{l\in\mathcal{L},j\in\mathcal{B}}}\mathrm{tr}\Big(\mathbf{S}_{i,m,k}\mathbf{E}_{i,m,k}(\mathbf{V},\mathbf{U},\hat{\mathbf{H}}+\mathbf{\Delta})\Big), \tag{34}$$

as studied in [9, Sec. II-B] for the case of a single-cell, i.e., $B = 1$, without symbol extension, i.e., $L = 1$. This approach will be referred to as "*MMSE*" scheme and can be implemented via Algorithm 1 by fixing the weight matrices to $\{\mathbf{S}_{i,m,k} = \mathbf{I}\}_{i\in\mathcal{B},m\in\mathcal{M}_i,k\in\mathcal{D}_{i,m}}$ and removing Step 4.

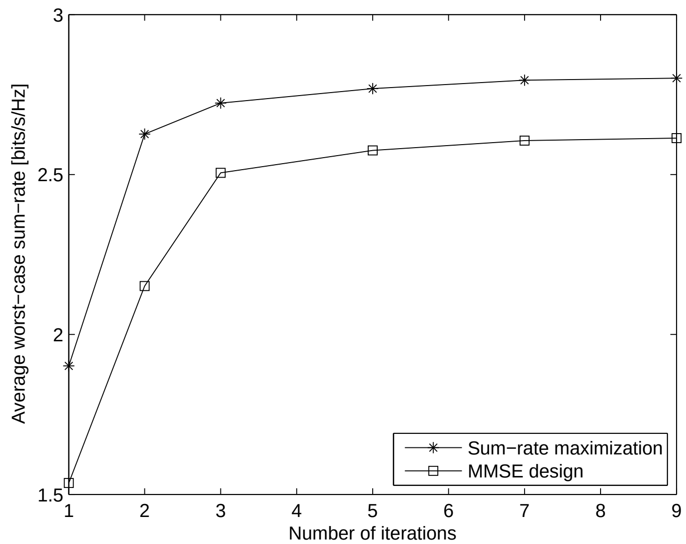

Figure 4. Average worst-case sum-rate versus the number of iterations for the multi-cell downlink system with $B = 3$, $K_i = 1$, $N_{T,i} = N_{R,(i,k)} = 1$, $P_i = 10$ dB,$L = 2$, $\varepsilon = 0.5$ and unicast messages $\mathcal{D}_{i,m} = \{m\}$ for $m \in \mathcal{K}_i$.

We set the CSI uncertainty levels to be all equal as $\varepsilon_{i,k,j}(l) = \varepsilon$ for all $i, j \in \mathcal{B}$, $k \in \mathcal{K}_i$ and $l \in \mathcal{L}$. We assume that the elements of the nominal channel matrices $\hat{\mathbf{H}}_{i,k,j}(l)$ are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit variance, and are also independent across the indices i, k, j and l. For each realization of the nominal CSIs $\{\hat{\mathbf{H}}_{i,k,j}\}_{i,j\in\mathcal{B},k\in\mathcal{K}_i}$ and given precoding matrices $\mathbf{V}$, we measure the worst-case sum-rate by randomly generating N_{worst} samples of the uncertainty terms $\mathbf{\Delta}_{i,k,j}(l) \in \mathcal{U}_{i,k,j}(l)$ for all $i, j \in \mathcal{B}$, $k \in \mathcal{K}_i$ and $l \in \mathcal{L}$, and taking the minimum of the measured sum-rates using (10). Specifically, we obtain each sample $\mathbf{\Delta}_{i,k,j}(l)$ as $\mathbf{\Delta}_{i,k,j}(l) = \mathbf{W}_{i,k,j}(l)/||\mathbf{W}_{i,k,j}(l)||_F \cdot \varepsilon_{i,k,j}(l)$ where the matrix $\mathbf{W}_{i,k,j}(l)$ has i.i.d. complex Gaussian random elements with zero mean and unit variance. Thus, the measured worst-case sum-rate is expected to approach the actual worst-case sum-rate as the number N_{worst} of samples increases. In our simulations, we used $N_{\text{worst}} = 10^4$ samples.

In Fig. 4, we show an example of the convergence rates of the proposed schemes summarized in Sec. III to a locally optimal point. Specifically, Fig. 4 plots the average worst-case sum-rate obtained with the sum-rate maximization and MMSE designs versus the number of iterations for the multi-cell downlink system with $B = 3$, $K_i = 1$, $N_{T,i} = N_{R,(i,k)} = 1$, $P_i = 10$ dB, $L = 2$, $\varepsilon = 0.5$ and unicast message sets $\mathcal{D}_{i,m} = \{m\}$ for $m \in \mathcal{K}_i$. It is observed that a few, around 7, iterations are sufficient for the convergence of the schemes.

Fig. 5 plots the average worst-case sum-rate versus the uncertainty level ε for the multi-cell downlink

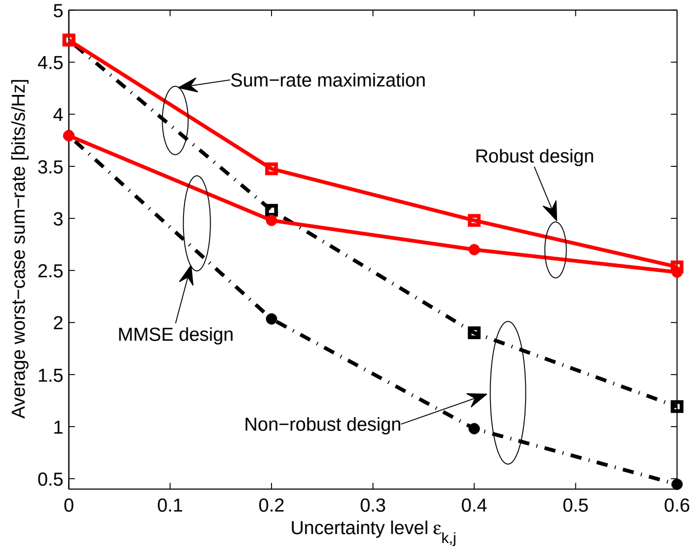

Figure 5. Average worst-case sum-rate versus the uncertainty level ε for the multi-cell downlink system with $B = 3$, $K_i = 1$, $N_{T,i} = N_{R,(i,k)} = 1$, $P_i = 10$ dB,$L = 2$ and unicast messages $\mathcal{D}_{i,m} = \{m\}$ for $m \in \mathcal{K}_i$.

system with $B = 3$, $K_i = 1$, $N_{T,i} = N_{R,(i,k)} = 1$, $P_i = 10$ dB, $L = 2$ and unicast message sets $\mathcal{D}_{i,m} = \{m\}$ for $m \in \mathcal{K}_i$. The performance gain of the proposed robust scheme over the non-robust design is seen to increase rapidly with the CSI uncertainty level ε for both sum-rate maximization and MMSE design. For instance, for $\varepsilon = 0.1$, $\varepsilon = 0.2$ and $\varepsilon = 0.3$, the gains in terms of sum-rate are 5%, 13% and 28% for sum-rate maximization design and 17%, 50% and 67% for MMSE design, respectively. Moreover, the performance loss of the robust MMSE scheme compared to the robust sum-rate maximization scheme vanishes when the CSI is sufficiently inaccurate. This shows that a simpler MMSE design may be sufficient when the CSI is not accurate enough.

Fig. 6 plots the average sum-rate obtained with the MMSE design versus the number L of (time-frequency) slots for the multi-cell downlink system with $B = 3$, $K_i = 1$, $N_{T,i} = N_{R,(i,k)} = 1$, $\varepsilon = 0.15$ and unicast message sets $\mathcal{D}_{i,m} = \{m\}$ for $m \in \mathcal{K}_i$. From the figure, we observe that joint precoding across multiple slots provides a significant gain compared to precoding over a single slot, due to the possibility to perform power allocation and interference alignment precoding across the slots. Note that this gain will be decreased in the presence of correlated channel matrices due to the lower diversity. Moreover, the gain of the robust scheme compared to the non-robust scheme is uniform over L slots and more significant at high SNR, where the channel estimation error is dominant over the channel noise power.

Fig. 7 plots the average sum-rate obtained with the MMSE design versus the number $K_i = K$ of

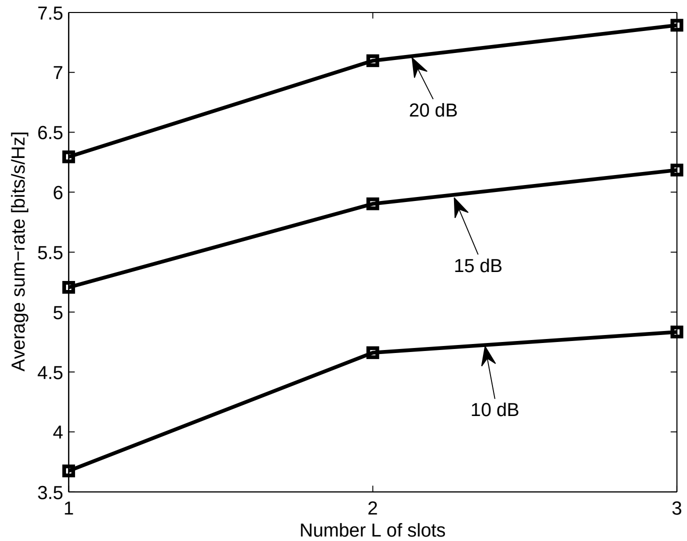

Figure 6. Average sum-rate versus the number L of slots for the multi-cell downlink system with $B = 3$, $K_i = 1$, $N_{T,i} = N_{R,(i,k)} = 1$, $\varepsilon = 0.15$ and unicast message sets $\mathcal{D}_{i,m} = \{m\}$ for $m \in \mathcal{K}_i$.

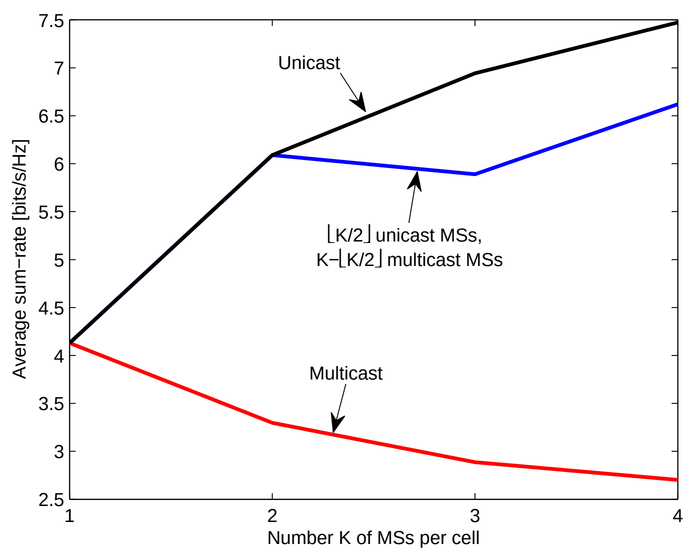

Figure 7. Average sum-rate versus the number K_i of MSs for the multi-cell downlink system with $B = 2$, $N_{T,i} = 2$, $N_{R,(i,k)} = 1$, $L = 1$, $P_i = 15$ dB and $\varepsilon = 0.15$.

MSs for the multi-cell downlink system with $B = 2$, $N_{T,i} = 2$, $N_{R,(i,k)} = 1$, $L = 1$, $P_i = 15$ dB and $\varepsilon = 0.15$. We compare the performance obtained with three different message assignments: *i*) Each BS transmits K unicast messages to the MSs in the cell; *ii*) Each BS transmits a single multicast message to all the MSs in the cell; and *iii*) Each BS transmits $\lfloor K/2 \rfloor$ unicast messages to $\lfloor K/2 \rfloor$ MSs and a single multicast message to the remaining $K - \lfloor K/2 \rfloor$ MSs. It is observed that increasing the number of unicast messages results in improved sum-rate performance due to the added flexibility in allocating the available transmission resources. In contrast, the sum-rate is degraded by increasing the number of MSs requesting multicast messages because of the increased number of rate constraints imposed on the multicast data streams (see (14b)). Finally, it is noted that, as the number of MSs increases, the performance gain of the robust unicast scheme grows larger with respect to the non-robust strategy: in this regime, the less effective control of inter-MS interference afforded by imperfect CSI leads to a more significant performance degradation. This trend is not present for the multicast case, in which having more MSs leads to smaller gains achievable by precoding based on CSI.

B. Sum-MSE Minimization over Linear Precoding and DF Equalization

We now turn to the evaluation of the performance of linear precoding and DF equalization in terms of SER. We consider uncoded transmission and focus on a 4-QAM constellation, as done, e.g., in [28].

We start by focusing on the case with perfect CSI, i.e., $\mathbf{H}_{i,j} = \hat{\mathbf{H}}_{i,j}$ for all $i, j \in \mathcal{B}$. The purpose is to demonstrate the gains of the proposed joint optimization of linear precoding and DF equalization over a separate optimization. According to the latter approach, adopted, e.g., in [17], the precoding is first optimized according to some criterion, here the MSE, and then the DF equalizers are optimized for the given precoding matrices. Another benchmark is provided by the MMSE solution with linear precoding and equalization discussed above. DF equalization is designed with $d_k^{\mathrm{dec}} = d_i$, i.e., each MS only decodes and cancels the data streams intended for itself. Fig. 8 shows the SER for $K = 2$, $d_i = 2$ and $N_{T,i} = N_{R,i} = 4$ for all $i \in \mathcal{B}$ versus the signal-to-noise ratio (SNR) $P_i = P$ for all $i \in \mathcal{B}$ for the three mentioned techniques. The gains of DF equalization versus linear equalization are apparent, and the proposed joint optimization leads to gains of around 2.5 dB at SER equal to 10^{-6}.

Turning to the case with imperfect CSI, we assume that the CSI errors $\mathbf{\Delta}_{i,j}$ are generated randomly and we evaluate the worst-case SER using the same procedure discussed above in the context of the sum-rate. Fig. 9 shows the SER versus SNR P for the proposed robust solution as compared to a non-robust solution that is designed assuming perfect CSI, i.e., that $\mathbf{H}_{i,j} = \hat{\mathbf{H}}_{i,j}$ for all $i, j \in \mathcal{B}$. We set $K = 4$, $d_i = 2$ and $N_{T,i} = N_{R,i} = 4$ for all $i \in \mathcal{B}$ and ε is selected to be 10% of the average norm of the channel matrices $(E[||\hat{\mathbf{H}}_{i,j}||^2])^{1/2}$. The robust scheme is shown to provide significant performance gains, namely around 3 dB at 6×10^{-3}. It should also be mentioned that both schemes attain their error floors at $P = 15$ dB (not shown in the figure).

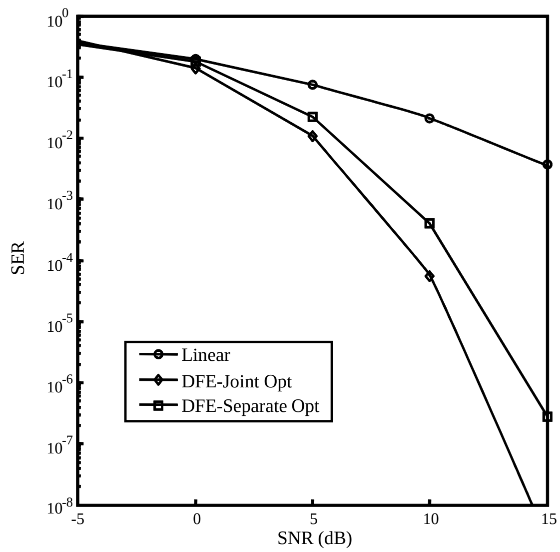

Figure 8. SER versus SNR for linear equalization and DF equalization assuming separate or joint optimization ($K = 2$, $d_i = 2$, $N_{T,i} = N_{R,i} = 4$, $P_i = P$ for all $i \in \mathcal{B}$).

Finally, we further elaborate on the distributed solution discussed in Sec. IV-E. Specifically, we focus on the convergence properties of the proposed ADMM approach. Fig. 9 shows the SER for the centralized solution discussed as far as compared with the decentralized solution with two or five ADMM iteration. It is seen that five iterations are sufficient to obtain the same performance as the centralized scheme.

VI. Concluding Remarks

The design of linear and non-linear transceivers for multi-cell MIMO systems has been studied under the requirement of robustness with respect to the uncertainty on the CSI. Adopting a deterministic worst-case design approach, we have tackled the sum-rate optimal robust design of linear precoders for a MIMO interference broadcast channel (MIMO-IBC) with general unicast/multicast messages in each cell and operating over multiple time-frequency resources. We have then considered the optimal robust joint design of linear precoding and symbol-by-symbol DF equalization at the MSs using the MSE as the performance criterion. Both problems have been addressed by proposing iterative algorithms that are also shown to be implementable in a distributed fashion via the ADMM algorithm on processors that have only local and partial CSI. Numerical results have demonstrated that the proposed robust solutions improves significantly over conventional non-robust schemes. The gain in sum-rate is seen to be particularly relevant for unicast message assignments with a larger number of users and for moderate-to-large SNR. Moreover, the proposed joint design of linear precoding and DF equalization is shown to outperform existing separate design approaches.

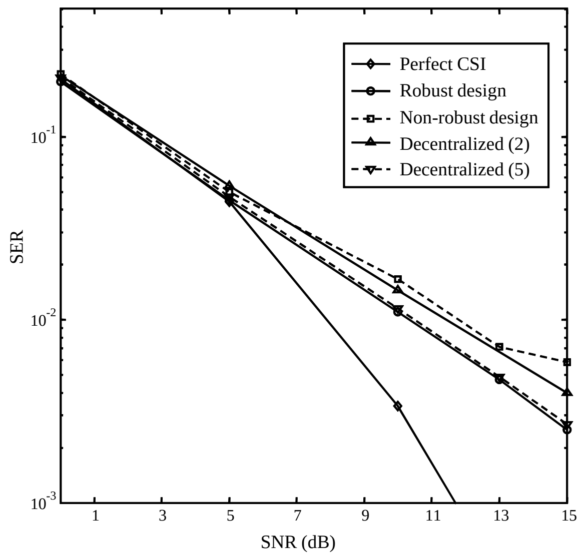

Figure 9. SER versus SNR for the centralized robust and non-robust designs of linear precoding and DF equalization and for the decentralized implementation with two and five ADMM inner iterations ($K = 4$, $d_i = 2$, $N_{T,i} = N_{R,i} = 4$ for all $i \in \mathcal{B}$ and $\varepsilon = 0.1(E[||\hat{\mathbf{H}}_{i,j}||^2])^{1/2}$).

APPENDIX

In this section, we review some useful lemmas that are used in the derivations presented in the text.

Lemma 3. (Fenchel Conjugate Function [26]) *Consider a matrix $\mathbf{E} \in \mathbb{C}^{d\times d}$ with $\mathbf{E} \succ \mathbf{0}$. Then, we have the equality*

$$\log\det\left(\mathbf{E}^{-1}\right) = \max_{\mathbf{S}\succeq\mathbf{0}} \left\{-\mathrm{tr}\left(\mathbf{SE}\right) + \log\det\left(\mathbf{S}\right) + d\right\}. \tag{35}$$

Lemma 4. (Sign Definiteness Lemma [9]) *Let $\mathbf{A}$, $\{\mathbf{P}_i, \mathbf{Q}_i\}_{i=1}^N$ be given matrices with appropriate sizes and $\mathbf{A} = \mathbf{A}^\dagger$. Then, the condition*

$$\mathbf{A} \succeq \sum_{i=1}^{N} \left(\mathbf{P}_i^\dagger \mathbf{X}_i \mathbf{Q}_i + \mathbf{Q}_i^\dagger \mathbf{X}_i^\dagger \mathbf{P}_i\right) \tag{36}$$

holds for all $\mathbf{X}_i$ satisfying $||\mathbf{X}_i|| \leq \varepsilon_i$, $i \in \{1, \ldots, N\}$ if there exist real nonnegative numbers $\mu_1, \ldots, \mu_N \geq 0$ that satisfy the condition

$$\begin{bmatrix} \mathbf{A} - \sum_{i=1}^N \mu_i \mathbf{Q}_i^\dagger \mathbf{Q}_i & -\varepsilon_1 \mathbf{P}_1^\dagger & \cdots & -\varepsilon_N \mathbf{P}_N^\dagger \\ -\varepsilon_1 \mathbf{P}_1 & \mu_1 \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\varepsilon_N \mathbf{P}_N & \mathbf{0} & \cdots & \mu_N \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \tag{37}$$

The converse is also true when $N = 1$ [10, Sec. IV][31, Sec. 2.6.3].

In this appendix, we show that the optimal solution of the problem (16) is lower-bounded by that of the problem (17) in Lemma 1. We first write an epigraph form of the problem (16) as

$$\underset{\mathbf{V},R,\mathbf{U},\mathbf{S}\succeq\mathbf{0},\gamma}{\text{maximize}} \sum_{i\in\mathcal{B}}\sum_{m\in\mathcal{M}_i} R_{i,m} \tag{38a}$$

$$\text{s.t. } R_{i,m} \leq -\gamma_{i,m,k} + \log\det\left(\mathbf{S}_{i,m,k}\right) + d_{i,m}, \text{ for all } i\in\mathcal{B},\, m\in\mathcal{M}_i,\, k\in\mathcal{D}_{i,m}, \tag{38b}$$

$$\gamma_{i,m,k} \geq \text{tr}\left(\mathbf{S}_{i,m,k}\mathbf{E}_{i,m,k}(\mathbf{V},\mathbf{U},\hat{\mathbf{H}}+\boldsymbol{\Delta})\right), \tag{38c}$$

$$\text{for all } \{\boldsymbol{\Delta}_{i,k,j}(l)\in\mathcal{U}_{i,k,j}(l)\}_{l\in\mathcal{L},j\in\mathcal{B}},\, i\in\mathcal{B},\, m\in\mathcal{M}_i,\, k\in\mathcal{D}_{i,m},$$

$$\sum_{m\in\mathcal{M}_i}\text{tr}\left(\mathbf{V}_{i,m}\mathbf{V}_{i,m}^{\dagger}\right) \leq LP_i, \text{ for all } i\in\mathcal{B}, \tag{38d}$$

where we have defined the variables $\gamma \triangleq \{\gamma_{i,m,k}\}_{i\in\mathcal{B},m\in\mathcal{M}_i,k\in\mathcal{D}_{i,m}}$ and $\tilde{\mathbf{S}} \triangleq \{\tilde{\mathbf{S}}_{i,m,k}\}_{i\in\mathcal{B},m\in\mathcal{M}_i,k\in\mathcal{D}_{i,m}}$ with $\tilde{\mathbf{S}}_{i,m,k} \triangleq \mathbf{S}_{i,m,k}^{1/2}$.

From the MSE expression in (13), we can see that the constraints (38c) are equivalent to the conditions

$$\gamma_{i,m,k} \geq \sum_{j\in\mathcal{B}}\tau_{i,m,k,j} + \left\|\boldsymbol{\Sigma}_{i,k}^{1/2}\mathbf{U}_{i,m,k}\tilde{\mathbf{S}}_{i,m,k}\right\|_F^2, \text{ for all } i\in\mathcal{B},\, m\in\mathcal{M}_i,\, k\in\mathcal{D}_{i,m}, \tag{39}$$

$$\text{and } \tau_{i,m,k,j} \geq \sum_{q\in\mathcal{M}_{i,m,k,j}}\left\|\tilde{\mathbf{S}}_{i,m,k}^{\dagger}\left(\mathbf{U}_{i,m,k}^{\dagger}(\hat{\mathbf{H}}_{i,k,j}+\boldsymbol{\Delta}_{i,k,j})\mathbf{V}_{j,q} - \delta_{(i,m),(j,q)}\mathbf{I}\right)\right\|_F^2, \tag{40}$$

$$\text{for all } \{\boldsymbol{\Delta}_{i,k,j}(l)\in\mathcal{U}_{i,k,j}(l)\}_{l\in\mathcal{L}},\, i,j\in\mathcal{B},\, m\in\mathcal{M}_i,\, k\in\mathcal{D}_{i,m},$$

where we have introduced auxiliary variables $\tau_{im,,k,j}$ to simplify the expression and the sets $\mathcal{M}_{i,m,k,j}$ are defined in (13). After some manipulations, the constraint (40) can be rewritten as

$$\tau_{i,m,k,j} \geq \left\|\mathbf{c}_{i,m,k,j} + \sum_{l\in\mathcal{L}}\mathbf{C}_{i,m,k,j}(l)\mathbf{d}_{i,k,j}(l)\right\|^2, \tag{41}$$

$$\text{for all } \{\|\mathbf{d}_{i,,k,j}(l)\| \leq \varepsilon_{i,k,j}(l)\}_{l\in\mathcal{L}},\, i,j\in\mathcal{B},\, m\in\mathcal{M}_i,\, k\in\mathcal{D}_{i,m},$$

where we have defined the vector $\mathbf{d}_{i,k,j}(l) \triangleq \text{vec}(\boldsymbol{\Delta}_{i,k,j}(l))$ and the notations $\mathbf{c}_{i,m,k,j}$ and $\mathbf{C}_{i,m,k,j}(l)$ are defined in (19)-(22).

Applying the Schur complement Lemma [29, Appendix C] to the constraint (41), we obtain the following equivalent linear matrix inequality.

$$\begin{bmatrix}\tau_{i,m,k,j} & \mathbf{c}_{i,m,k,j}^{\dagger}\\ \mathbf{c}_{i,m,k,j} & \mathbf{I}\end{bmatrix} + \sum_{l\in\mathcal{L}}\begin{bmatrix}0 & \mathbf{d}_{i,k,j}^{\dagger}(l)\mathbf{C}_{i,m,k,j}^{\dagger}(l)\\ \mathbf{C}_{i,m,k,j}(l)\mathbf{d}_{i,k,j}(l) & \mathbf{0}\end{bmatrix} \succeq \mathbf{0}, \tag{42}$$

$$\text{for all } \{\|\mathbf{d}_{i,k,j}(l)\| \leq \varepsilon_{i,k,j}(l)\}_{l\in\mathcal{L}},\, i,j\in\mathcal{B},\, m\in\mathcal{M}_i,\, k\in\mathcal{D}_{i,m}.$$

From Lemma 4, we can see that the constraint (42) holds if the condition

$$\begin{bmatrix} \tau_{i,m,k,j} - \sum_{l\in\mathcal{L}} \mu_{i,m,k,j}(l) & \mathbf{c}_{i,m,k,j}^{\dagger} & \mathbf{0} \\ \mathbf{c}_{i,m,k,j} & \mathbf{I} & -\mathbf{C}_{i,m,k,j} \\ \mathbf{0} & -\mathbf{C}_{i,m,k,j}^{\dagger} & \mathrm{diag}\left(\{\mu_{i,m,k,j}(l)\}_{l\in\mathcal{L}}\right) \otimes \mathbf{I} \end{bmatrix} \succeq \mathbf{0}, \tag{43}$$

is satisfied for some $\{\mu_{i,m,k,j}(l) \geq 0\}_{l\in\mathcal{L}}$ and for all $i, j \in \mathcal{B}$, $m \in \mathcal{M}_i$ and $k \in \mathcal{D}_{i,m}$ with the notation $\mathbf{C}_{i,m,k,j} \triangleq -[\varepsilon_{i,k,j}(1)\mathbf{C}_{i,m,k,j}(1), \ldots, \varepsilon_{i,k,j}(L)\mathbf{C}_{i,m,k,j}(L)]$. Also, the converse is true when $L = 1$ (see, e.g., [10, Sec. IV][31, Sec. 2.6.3]). Note that the condition (43) implies (42) but is not equivalent unless $L = 1$. Therefore, replacing the condition (40) with (43) leads to a problem whose solution lower-bounds that of the problem (38). Substituting the conditions (39), (40) and (43) into the problem (38) results in the problem (17) in Lemma 1, which concludes the proof.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi; S. V. Hanly, A. Lozano, A. C. K. Soong and J. C. Zhang, "What Will 5G Be?," *IEEE Journ. Sel. Areas Comm.*, vol. 32, no. 6, pp. 1065-1082, Jun. 2014.

[2] H. Huh, H. C. Papadopoulos and G. Caire, "Multiuser MISO transmitter optimization for intercell interference mitigation," *IEEE Trans. Sig. Processing*, vol. 58, no. 8, pp. 4272-4285, Aug. 2010.

[3] D. Stotz and H. Bolcskei, "Degrees of freedom in vector interference channels," arXiv:1210.2259.

[4] P. C. Weeraddana, M. Codreanu, M. Latva-aho, A. Ephremides, and C. Fischione, "Weighted sum-rate maximization in wireless networks: A review," *Foundations and Trends in Networking*, vol. 6, no. 1-2, pp. 1-163, 2012.

[5] K. Gomadam, V. R. Cadambe and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309-3322, Jun. 2011.

[6] M. Hong, R.-Y. Sun, H. Baligh and Z.-Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Comm.*, vol. 31, no. 2, pp. 226-240, Feb. 2013.

[7] S. A. Jafar, "Interference Alignment: A New Look at Signal Dimensions in a Communication Network," *Foundations and Trends in Communications and Information Theory*, vol. 7, no. 1, pp. 1-136, 2011.

[8] A. Tajer, N. Prasad and X. Wang, "Robust linear precoder design for multi-cell downlink transmission," *IEEE Trans. Sig. Processing*, vol. 59, no. 1, pp. 235-251, Jan. 2011.

[9] E. A. Gharavol and E. G. Larsson, "The sign-definiteness lemma and its applications to robust transceiver optimization for multiuser MIMO systems," *IEEE Trans. Sig. Processing*, vol. 61, no. 2, pp. 238-252, Jan. 2013.

[10] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Sig. Processing*, vol. 52, no. 7, pp. 1931-1946, Jul. 2004.

[11] P. Patcharamaneepakorn, A. Doufexi and S. Armour, "Leakage-based transceiver designs for MIMO interfering broadcast channels," in *Proc. IEEE Wireless Comm. Sig. Processing (WCSP 2013)*, Hangzhou, China, Oct. 2013.

[12] J. Jose, N. Prasad, M. Khojastepour and S. Rangarajan, “On robust weighted-sum rate maximization in MIMO interference networks,” in *Proc. IEEE Intern. Conf. Comm. (ICC 2011)*, Kyoto, Japan, Jun. 2011.

[13] Y. Huang, Q. Li, W.-K. Ma and S. Zhang, “Robust multicast beamforming for spectrum sharing-based cognitive radios,” *IEEE Trans. Sig. Processing*, vol. 60, no. 1, pp. 527-533, Jan. 2012.

[14] H. Bolcskei and I. J. Thukral, “Interference alignment with limited feedback,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT 2009)*, Seoul, Korea, Jul. 2009.

[15] R. T. Krishnamachari and M. K. Varanasi, “Interference alignment under limited feedback for MIMO interference channels,” *IEEE Trans. Sig. Processing*, vol. 61, no. 15, pp. 3908-3917, Aug. 2013.

[16] M. B. Shenouda and T. N. Davidson, “Tomlinson-Harashima precoding for broadcast channels with uncertainty,” *IEEE Journ. Sel. Areas Comm.,* vol. 25, no. 7, pp. 1380-1389, Sep. 2007.

[17] V. Ntranos and G. Caire, “A comparison of decoding techniques for interference alignment,” in *Proc. IEEE Int. Symp. Comm. Control Sig. Processing (ISCCSP 2012)*, Rome, Italy, May 2012.

[18] N. D. Sidiropoulos, T. N. Davidson and Z.-Q. Luo, “Transmit beamforming for physical-layer multicasting,” *IEEE Trans. Sig. Processing*, vol. 54, no. 6, pp. 2239-2251, Jun. 2006.

[19] M. A. Khojastepour, A. Salehi-Golsefidi and S. Rangarajan, “Towards an optimal beamforming algorithm for physical layer multicasting,” in *Proc. IEEE Inf. Theory Workshop (ITW 2011)*, Paraty, Brazil, Oct. 2011.

[20] H. Zhu, N. Prasad and S. Rangarajan, “Precoder design for physical layer multicasting,” *IEEE Trans. Sig. Processing*, vol. 60, no. 11, pp. 5932-5947, Nov. 2012.

[21] Z. Xiang, M. Tao and X. Wang, “Coordinated multicast beamforming in multicell networks,” *IEEE Trans. Wireless Comm.*, vol. 12, no. 1, pp. 12-21, Jan. 2013.

[22] S. Sesia, I. Toufik and M. Baker, *LTE - The UMTS long term evolution: from theory to practice*, Wiley, 2011.

[23] M. Torbatian, H. Najafi, and M. O. Damen, “Asynchronous interference alignment,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3148-3157, Sept. 2012.

[24] S. S. Christensen, R. Agarwal, E. Carvalho and J. Cioffi, “Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design,” *IEEE Trans. Wireless Comm.*, vol. 7, no. 12, pp. 4792-4799, Dec. 2008.

[25] Q. Shi, M. Razaviyayn, Z.-Q. Luo and C. He, “An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” *IEEE Trans. Sig. Processing*, vol. 59, no. 9, pp. 4331-4340, Sep. 2011.

[26] J. Borwein and A. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples,* Springer Verlag, 2006.

[27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the Alternating Direction Method of Multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, 2011.

[28] M. B. Shenouda and T. N. Davidson, “A framework for designing MIMO systems with decision feedback equalization or Tomlinson-Harashima precoding,” *IEEE Journ. Sel. Areas Comm.,* vol. 26, no. 2, pp. 401-411, Feb. 2008.

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

[30] M. Hong and Z.-Q. Luo, “On the linear convergence of the alternating direction method of multipliers,” arXiv preprint arXiv:1208.3922 (2012).

[31] S. Boyd, L. E. Ghaoui, E. Feron and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory,* Philadelphia, PA: SIAM Studies in Applied Mathematics, 1994.