Robust Transceiver Optimization for Downlink Coordinated Base Station Systems: Distributed Algorithm

Tadilo Endeshaw Bogale, Student Member, IEEE, Luc Vandendorpe, Fellow, IEEE and Batu Krishna Chalise, Member, IEEE

Abstract—This paper considers the joint transceiver design for downlink multiuser multi-input single-output (MISO) systems with coordinated base stations (BSs) where imperfect channel state information (CSI) is available at the BSs and mobile stations (MSs). By incorporating antenna correlation at the BSs and taking channel estimation errors into account, we solve two robust design problems: 1) minimizing the weighted sum of mean-square-error (MSE) with per BS antenna power constraint, and 2) minimizing the total power of all BSs with per user MSE target and per BS antenna power constraints. These problems are solved as follows. First, for fixed receivers, we propose centralized and novel computationally efficient distributed algorithms to jointly optimize the precoders of all users. Our centralized algorithms employ the second-order-cone programming (SOCP) approach, whereas, our novel distributed algorithms use the Lagrangian dual decomposition, modified matrix fractional minimization and an iterative method. Second, for fixed BS precoders, the receivers are updated by the minimum mean-square-error (MMSE) criterion. These two steps are repeated until convergence is achieved. In all of our simulation results, we have observed that the proposed distributed algorithms achieve the same performance as that of the centralized algorithms. Moreover, computer simulations verify the robustness of the proposed robust designs compared to the non-robust/naive designs.

Index Terms—Robust Transceiver, multiuser MIMO, distributed optimization and convex optimization.

I. INTRODUCTION

The next generation multimedia communications are expected to support high data rates. To meet this demand, multi-antenna systems are recommended as they significantly increase the spectral efficiency of wireless channels [1], [2]. The performance enhancement is achieved by exploiting the transmit and receive diversity. In [3], a fundamental relation between mutual information and minimum mean-square-error (MMSE) has been established for multiple-input multiple-output (MIMO) Gaussian channels. Furthermore, it has been shown that different transceiver optimization problems are equivalently reformulated as a function of MMSE matrix, for instance, minimizing bit error rate, maximizing capacity etc [4]–[6]. For these reasons, mean-square-error (MSE)-based design problems are commonly examined in multiuser networks.

In general, the uplink channel MSE-based problems are better understood than that of the downlink channel problems. For this reason, most of the research works examine MSE-based problems for the downlink multiuser systems [5], [7]–[12]. However, all of the these papers examine their problems for conventional downlink networks. In these networks, base stations (BSs) from different cells communicate with their respective remote terminals independently. Hence, in the latter network, inter-cell interference is obliged to be considered as a background noise. Recently, it has been shown that BS coordination communication is a promising technique to significantly improve the capacity of wireless channels by mitigating (or possibly canceling) inter-cell interference [13]–[15]. The BS coordination can be performed by two approaches. In the first approach, BSs are coordinated at the beamforming (precoder) level [14], whereas in the second approach, coordination takes place both at the signal and beamforming (precoder) levels [13], [15]. It is well known that the latter coordination approach has better performance gain compared to the former one [15], [16]. This performance improvement, however, requires additional signal coordination. In the current paper, we focus on the second BS coordination approach (the approach of [13] and [15]). In [17], four MSE-based linear transceiver optimization problems have been considered for coordinated BS MIMO systems. These problems are examined by assuming that the total power of each BS or the individual power of each BS antenna is constrained. The optimization problems of [17] are solved as follows. First, by keeping the receivers constant, the precoders of all users are jointly optimized using a second-order-cone programming (SOCP) approach (SOCP problems are convex and can be solved using interior point (IP) methods [18]). Second, for the given BS precoders, the receiver of each user is optimized by the MMSE method. The first and second steps are repeated in an iterative manner to jointly optimize the transmitters and receivers. Thus, in [17], the receiver of each user can be optimized independently and distributively. However, the joint optimization of the precoders has been carried out by
a centralized algorithm. When the number of users and/or BSs increase, the computational cost of the joint precoder design also increases [19]. Consequently, solving the precoder optimization problem in a centralized manner, especially for large-scale coordinated networks, is not a computationally efficient approach. This motivates us to develop distributed algorithms to solve MSE-based problems for coordinated BS systems with per BS antenna power constraints in [20]. This paper solves its optimization problems distributively by applying the Lagrangian dual decomposition, modified matrix fractional minimization and an iterative technique.

In the current paper, we extend the work of [20] to robust case. The goal of this work is to jointly optimize the transmitters and receivers of all users when imperfect channel state information (CSI) is available both at the BSs and mobile stations (MSs), and with antenna correlation at the BSs. It is known in [21]–[23] that transmit antenna correlation matrices depend on array parameters (such as array geometry, antenna spacing), and the average angle of arrival (AOA) of scattered signals from user and the corresponding angular spread. This means that the transmit antenna correlation matrices (which capture spatial variation) vary at a rate much slower than the fast fading component (that captures temporal variations) of downlink channel. Thus, errors caused from the estimation of fast fading part of the channel can significantly outnumber the errors caused from the estimation of slowly varying antenna correlation matrices. Based upon these discussions, it is clear that the transmit correlation matrices can be obtained from long term channel statistics with a reasonable accuracy [24]. Due to this reason, perfect transmit antenna correlation matrices are assumed to be available at the BSs and MSs. We also assume that each BS is equipped with multiple antennas and each MS is equipped with single antenna. For our robust transceiver designs, a stochastic approach has been utilized. We design the transmitters and receivers by considering the practically relevant scenario where spatial correlation matrices as seen from each BS are different for different MSs and the variance of the estimation errors corresponding to the estimates of the channels are different. For this CSI model, we examine the following MSE-based robust design problems.

1) The robust minimization of the weighted sum MSE with per BS antenna power constraint ($P_1$).
2) The robust minimization of the total sum power of all BSs with per BS antenna power and per user MSE target constraints ($P_2$).

To the best of our knowledge, problems $P_1$ and $P_2$ are non-convex. Hence, convex optimization tools can not be used to solve them. Each of these problems are solved iteratively as follows. First, for given receivers, we propose centralized and novel computationally efficient distributed algorithms to design the optimal precoders of all users. Our centralized algorithm designs the precoders of all users using SOCP approach and our novel distributed algorithm designs the precoders of all users by employing the Lagrangian dual decomposition, modified matrix fractional minimization and an iterative approach. Second, like in [17] and [20], the receiver of each user is optimized independently using minimum average mean-square-error (MAMSE) approach. These steps are repeated until convergence is achieved. The centralized and distributed algorithms require the complete channel estimates of all MSs. The centralized precoder design algorithms are developed by extending the approach of [17] to the case where imperfect CSI is available at the BSs and MSs. However, this extension does not change the fact that the robust problem can be reformulated into a SOCP problem as in the case of non-robust problem [17]. Thus, the novelty of our current paper relies mainly on the proposed distributed algorithms where we have used modified matrix fractional minimization techniques to solve the precoder design problems of $P_1$ and $P_2$ distributively. As a result, our new distributed algorithms are able to solve the transceiver design problems of $P_1$ and $P_2$ with less computational cost than that of the centralized algorithms. Furthermore, in all of our simulation results, we have observed that our novel distributed algorithms achieve the same performance as that of the centralized algorithms. The main contributions of this paper is thus summarized as follows.

1) We propose novel computationally efficient distributed algorithms to jointly optimize the precoders of all users for problems $P_1$ and $P_2$. As will be clear later, the proposed distributed algorithms can be extended straightforwardly to solve $P_1$ and $P_2$ for MIMO coordinated BS systems with per BS antenna (groups of BS antennas) power constraints.
2) We have demonstrated that the proposed distributed algorithms for $P_1$ and $P_2$ achieve the same performance as that of the centralized algorithms proposed for $P_1$ and $P_2$.
3) We examine the joint effect of channel estimation errors and antenna correlations on the performance of $P_1$ and $P_2$.

The remaining part of this paper is organized as follows. We present the coordinated multi-antenna BS system model in Section II. In Section III, the multiuser channel model under imperfect CSI is presented. Section IV discusses the robust design problems $P_1$ - $P_2$, and the proposed centralized and distributed algorithms. The extensions of our centralized and distributed algorithms for $P_1$ and $P_2$ in a MIMO coordinated BS system is discussed in Section V. In Section VI, computer simulations are used to compare the performance of the centralized and distributed algorithms, and the robust and non-robust/novel designs. Finally, conclusions are drawn in Section VII.

Notations: The following notations are used throughout this paper. Upper/lower case boldface letters denote matrices/column vectors. $\text{vec}(X)$, $\text{tr}(X)$, $X^*$, $X^H$ and $E(X)$ denote vectorization, trace, optimal, transpose, conjugate trans-
pose and expected value of $X$, respectively. $I_n(1)$ is an identity matrix of size $n \times n$ (appropriate size) and $\mathbb{C}^{M \times M}$ represent spaces of $M \times M$ matrices with complex entries. The diagonal and block-diagonal matrices are represented by $\text{diag}(\cdot)$ and $\text{blkdiag}(\cdot)$ respectively. Subject to is denoted by s.t. $\|x\|_n$ is the the $n$th norm of a vector $x$.

system [17], [19]. It is assumed that $n_k$ is a zero-mean circularly symmetric complex Gaussian (ZMCSCG) random variable with the variance $\sigma^2_k$, i.e., $n_k \sim \mathcal{CN}(0, \sigma^2_k)$. We also assume that the symbol $d_k$ is a ZMCSCG random variable with unit variance and is independent of $\{d_{k,l}\}_{l=1,l\neq k}^L$ and noise $n_k$, i.e., $E\{d_k d_{k,l}^H\} = 1, E\{d_k d_{k,l}^H\} = 0, \forall l \neq k$ and $E\{d_k n_k^H\} = 0$.

For this system model, when perfect CSI is available at the BSs and MSs, the MSE of the $k$th user can be expressed as

$$
\xi_k = E_{d_k,n_k} \{(d_k - e_k)(d_k - e_k)^H\} = w_k^H \left( \sum_{l=1}^L h_{lk}^H B_l \sum_{l=1}^L h_{lk}^H B_l \right)^H + \sigma^2_k w_k - w_k^H \sum_{l=1}^L h_{lk}^H b_{lk} - \sum_{l=1}^L h_{lk}^H h_{lk} w_k + 1.
$$

### III. CHANNEL MODEL

Considering antenna correlation at the BSs, we model the Rayleigh fading MISO channel between the $l$th BS and the $k$th MS as $h_{lk}^H = h_{ulk}^H \tilde{R}_{blk}^{1/2}$, where the elements of $h_{ulk}^H$ are independent and identically distributed (i.i.d) ZMCSG random variables all with unit variance and $\tilde{R}_{blk} \in \mathbb{C}^{N \times N}$ is the antenna correlation matrix as seen from the $l$th BS [26], [24]. The channel estimation of the $k$th MS ($h_{lk}^H$) can be performed on $h_{ulk}, \forall l$, using an orthogonal training method [27]. Upon doing so, the true channel between the $l$th BS and $k$th MS $h_{lk}^H$ is given by [12]

$$
h_{lk}^H = (\hat{h}_{ulk}^H + e_{ulk}^H) \tilde{R}_{blk}^{1/2} = \hat{h}_{lk}^H + e_{ulk}^H \tilde{R}_{blk}^{1/2} \tag{2}
$$

where $\hat{h}_{ulk}^H$ is the MMSE estimate of $h_{ulk}^H$,$ \hat{h}_{lk}^H = \hat{h}_{ulk}^H \tilde{R}_{blk}^{1/2}$ and $e_{ulk}^H$ is the estimation error for which its entries are i.i.d with $\mathcal{CN}(0, \sigma^2_{e,ulk})$. In the case of linearly dependent channel estimation errors, $e_{ulk}^H$ and $\hat{h}_{lk}^H$ can be expressed as $e_{ulk}^H = \tilde{e}_{ulk}^H Z_{lk}$ and $h_{lk}^H = \hat{h}_{lk}^H + \tilde{e}_{ulk}^H \tilde{R}_{blk}$, where $Z_{lk} \in \mathbb{C}^{N_{lk} \times 2}$, $\tilde{e}_{ulk}^H \in \mathbb{C}^{1 \times N_{lk}} = \mathcal{CN}(0, \tilde{\sigma}^2_{e,ulk})$ and $\tilde{R}_{blk} = \tilde{Z}_{ulk} \tilde{R}_{blk}^{1/2}$. For simplicity, the current paper examines $\mathcal{P}1$ and $\mathcal{P}2$ for the channel model given in (2). As will be clear later, the approaches of this paper can also be applied to solve $\mathcal{P}1$ and $\mathcal{P}2$ with linearly dependent channel estimation error models.

The main idea of the robust design is that $\{e_{ulk}^H\}_{k=1}^K, \forall l$ are unknown but $(\hat{h}_{ulk}^H, \tilde{R}_{blk}$ and $\sigma^2_{e,ulk})_{k=1}^K, \forall l$ are available. We assume that the $k$th MS estimates its channel (i.e., $\hat{h}_{ulk}, \forall l$) and feeds $\{\hat{h}_{ulk}^H, \sigma^2_{e,ulk}\}, \forall l$ back to the BSs without any error and delay. Since the $l$th BS has the channel estimates of all MSs, it can compute $\tilde{R}_{blk}$ locally from the long term channel statistics of $\hat{h}_{ulk}^H$ [24]. Thus, both the BSs and MSs have the same channel imperfections. The average mean-square-error
(AMSE) of the $k$th MS ($\bar{\xi}_k$) can be expressed as

$$
\bar{\xi}_k = E_{e_k} E_{p_k} \left( \zeta_b \right) = w_k^H \left( \sum_{i=1}^L \tilde{h}_{ik}^H b_i \right)
$$

$$
= w_k^H \left( \sum_{i=1}^L \tilde{h}_{ik}^H b_i \right) + \sum_{i=1}^L \sigma_{e_k}^2 \text{tr} \left( R_{ik}^{1/2} b_i b_i^H R_{ik}^{1/2} \right) + \sigma_k^2 \right) w_k - w_k^H \sum_{i=1}^L \tilde{h}_{ik}^H b_i - \sum_{i=1}^L b_i \tilde{h}_{ik} w_k + 1.$$

(3)

where $\tilde{h}_{ik} = \left[ \tilde{h}_{ik}^1, \cdots, \tilde{h}_{ik}^L \right]^T$, $R_{ik} = \text{blkdiag}(\sigma_{e_k}^2 \tilde{R}_{1ik}, \cdots, \sigma_{e_k}^2 \tilde{R}_{Lik})$, and in the second equality we use the fact $E_{e_k} [e_k^H \Phi e_k] = \sigma_e^2 \text{tr} \left( \Phi \right)$, if the entries of $e$ are i.i.d with $CN(0, \sigma_e^2)$ and $\Phi$ is a matrix given in [12], [28]. Note that one can extend the precoder/decoder design problems of $P1$ and $P2$ by incorporating channel estimation and feedback errors with time delay. In this case, the expression of $\bar{\xi}_k$ will be different from (3). Hence, solving $P1$ and $P2$ with erroneous feedback and non-zero delay is an open research topic.

IV. PROBLEM FORMULATIONS AND PROPOSED SOLUTIONS

In this section, we examine problems $P1$ and $P2$. For each problem, we use the following general optimization framework. First, for fixed receivers, the precoders of all users are optimized. second, using the latter precoders, the optimal receiver of each user is designed by the MAMSE receiver method. Since designing the receivers by MAMSE approach is optimal for any robust MSE-based problem, $P1$ and $P2$ utilize the same MAMSE expression to design their receivers. Finally, these two steps are repeated until convergence is achieved. The MAMSE receiver of the $k$th user is given by [20]

$$
w_k = \frac{\tilde{h}_{ik}^H b_i}{\tilde{h}_{ik}^H \sum_{i=1}^K b_i \eta_i + \text{tr} \left( R_{ik} \sum_{i=1}^K b_i b_i^H \right) + \sigma_k^2.}
$$

In the following, for fixed receivers $\{w_k\}_{k=1}^K$, we propose centralized and computationally efficient distributed precoder design algorithms for problems $P1$ and $P2$.

A. Robust weighted sum MSE minimization problem ($P1$)

The robust weighted sum MSE minimization with per BS antenna power constraint problem can be formulated as

$$
\min_{\{w_k, b_k\}_{k=1}^K} \sum_{k=1}^K \eta_k \bar{\xi}_k, \quad \text{s.t.} \quad \sum_{k=1}^K b_k b_k^H \leq p_n, \quad \forall n
$$

(5)

where $\eta_k$ is the AMSE weighting factor of the $k$th user and $p_n$ is the maximum available power at the $n$th BS antenna. The antenna numbers are assigned from the first antenna of BS$_1$ (which corresponds to antenna 1) to the last antenna of BS$_L$ (which corresponds to antenna $N$).

1) Centralized precoder design for $P1$: The objective function of the above problem can be expressed as

$$
\min_{k=1}^K \eta_k \bar{\xi}_k = \left\{ \begin{array}{l}
\eta_k \bar{\xi}_k = \frac{\tilde{h}_{ik}^H b_i}{\tilde{h}_{ik}^H \sum_{i=1}^K b_i \eta_i + \text{tr} \left( R_{ik} \sum_{i=1}^K b_i b_i^H \right) + \sigma_k^2} \end{array} \right.
$$

$$
= \text{tr} \left( \frac{\eta_k \tilde{h}_{ik}^H b_i}{\sigma_k^2 \eta_k w_k^H R_{ik}^H b_i} - \eta_k \tilde{h}_{ik}^H b_i - \eta_k \tilde{h}_{ik} b_i + \sigma_k^2 \eta_k w_k^H + \eta_k \right)
$$

$$
= \text{tr} \left( \frac{\eta_k \tilde{h}_{ik}^H b_i}{\sigma_k^2 \eta_k w_k^H R_{ik}^H b_i} - \eta_k \tilde{h}_{ik}^H b_i - \eta_k \tilde{h}_{ik} b_i + \sigma_k^2 \eta_k w_k^H + \eta_k \right)
$$

$$
+ \text{tr} \left( B^H \Psi B \right) + \text{tr} \left( \frac{\sigma_k^2 \eta_k W^H W}{} \right)
$$

(6)

where $\eta = \text{diag}(\eta_1, \cdots, \eta_K)$, $W = \text{diag}(w_1, \cdots, w_K)$, $H = [h_1, \cdots, h_K]$, $\sigma = \text{diag}(\sigma_1, \cdots, \sigma_K)$ and $\Psi = \sum_{k=1}^K \eta_k w_k^H R_{ik}$. For fixed receivers $\{w_k\}_{k=1}^K$ and using (6), the global optimal $\{b_k\}_{k=1}^K$ of (5) can be obtained by solving the following problem [17]

$$
\min_{\chi, b_k} \chi \text{s.t.} ||\mu||_2 \leq \chi, \quad ||b_n||_2 \leq \sqrt{p_n}, \quad \forall n
$$

(7)

where $\mu = \text{vec} \left( \sqrt{\eta} W^H H^H B - \sqrt{\eta} \right) : \text{vec} \left( \sqrt{\Psi} B \right)$ and $\tilde{h}_{ik}^H$ is the $b$th row of $H^H$.

As we can see, (7) is a SOCP problem for which the global optimal solution is obtained by existing convex optimization tools [18]. This problem has $2NK + 1$ real optimization variables, $N$ second-order-cone (SOC) constraints where each of them consists of $2K$ real dimensions and one SOC constraint with $2K(K + N)$ real dimensions. According to [29] (see page 196), the computational complexity of the latter problem in terms of number of iterations is upper bounded by $O(\sqrt{N} + 1)$ where the complexity of each iteration is within the order of $O((2K^2 + 4K^2)N)$. Thus, the total worst-case computational complexity of (7) is given by $O((2K^2 + 4K^2)N)$. This shows that for large networks, the centralized precoder design scheme appears to be impractical. This motivates us to design the precoders of each user distributively with less computational cost than that of the centralized precoder design approach.

2) Distributed precoder design for $P1$: For fixed $\{w_k\}_{k=1}^K$, to solve the precoders of (5) distributively, we propose the Lagrangian dual decomposition technique 2. To this end, we first express the Lagrangian function associated with

\footnote{Since the precoder design problem of (5) is convex, there is a zero duality gap between the primal and dual problems.}
\( L(\lambda, B) = \sum_{k=1}^{K} \eta_k \tilde{b}_k + \sum_{n=1}^{N} \lambda_n \left( \sum_{i=1}^{K} b_i \lambda_n H_{i,n,n} - p_n \right) \)

\[
\begin{align*}
&= \sum_{k=1}^{K} \left\{ b_k^H \left( \sum_{i=1}^{K} \eta_i \tilde{h}_i w_i H_i + \eta_i w_i H_i R_{bi} \right) b_k \\
&\quad - \eta_k w_k H_k b_k - \eta_k b_k^H H_k w_k + \sigma_k^2 \eta_k w_k H_k w_k + \eta_k \right\} \\
&+ \sum_{n=1}^{N} \lambda_n \left( \sum_{i=1}^{K} b_i \lambda_n H_{i,n,n} - p_n \right) \\
&= \sum_{k=1}^{K} \left\{ b_k^H A b_k - \eta_k w_k H_k b_k - \eta_k b_k^H H_k w_k + \sigma_k^2 \eta_k w_k H_k w_k + \eta_k \right\} - \sum_{n=1}^{N} \lambda_n p_n \\
&= g(\lambda) = \min_{\{b_k\}_{k=1}^{K}} L(\lambda, B)
\end{align*}
\]
Thus, $O_{arises from matrix inversion. According to \[30\], the major computational load of the method are the same.

Algorithm I: Iterative algorithm to solve (12)

1) Initialization: Set $\{\lambda_n = 1\}_{n=1}^N$.

Repeat

2) With the current $\lambda$, compute $\{g_n\}_{n=1}^N$ using (17) and update $\{\lambda_n\}_{n=1}^N$ with (18).

3) Share the above $\{\lambda_n\}_{n=1}^N$ among all processors.

4) Calculate the objective function of (12).

Until convergence.

As we can see, Algorithm I is developed to get the minimum value of the objective function of (12). Thus, this algorithm should stop iteration when the objective function of (12) is not decreasing significantly [5]. One simple approach of doing this is to stop Algorithm I from iteration when $\phi_i - \phi_{i+1} < \delta$, where $\phi_i$ is the objective function of (12) at the $i$th iteration and $\delta$ is the desired accuracy. For our simulation results, we have used the latter approach to declare convergence of Algorithm I with $\delta = 10^{-12}$.

Convergence: Since (13) is jointly convex in $\{g_n, t_n\}_{n=1}^N$ and $\{\lambda_n\}_{n=1}^N$, at each step of Algorithm I, the objective function of (13) is non-increasing. This implies that at each iteration of this algorithm, the objective function of (12) is also non-increasing. Moreover, it is clearly seen that the objective function of (12) is lower bounded by 0. These two facts show that Algorithm I is always convergent\(^1\). Although we are not able to prove the global optimality of Algorithm I analytically, in all of our simulation results, we have observed that the optimal $\lambda$ of (12) obtained by Algorithm I and the SDP method are the same.

Computational complexity: The major computational load of Algorithm I arises from matrix inversion. According to [30], matrix inversion has a complexity on the order of $O(N^2.376)$.

Thus, Algorithm I requires $O(N^2.376)$ per iteration. As will be shown later in Section VI, in all of our simulations, Algorithm I converges to an optimal solution in less than 10 iterations. This shows that the proposed distributed algorithm significantly reduces the computational load of the precoder design for $P1$.

Note: When the $n$th power constraint of (5) is inactive, at optimality, the corresponding Lagrangian multiplier should be zero. However, when we reformulate (12) into (13), each of $\{\lambda_n\}_{n=1}^N$ is not allowed to be zero. This shows that the development of distributed algorithm for (12) with $\{\lambda_n \geq 0\}_{n=1}^N$ is an open problem.

Using $\{\lambda_n\}_{n=1}^N$ of Algorithm I, the optimal $\{b_k\}_{k=1}^K, \forall k$ of (5) can be computed by (10), and with these $\{b_k\}_{k=1}^K, \forall k$, the decoder of each user can be computed using the MAMSE receiver approach (4). In summary, we solve $P1$ (5) distributively as shown in Algorithm II.

Algorithm II: Distributed algorithm for problem $P1$ (5)

Initialization: Set $B = H$ and normalize the rows of $B$ such that the power constraint of each antenna is satisfied with equality. Then, initialize $\{w_k\}_{k=1}^K$ by MAMSE receiver (4). Set the maximum number of iterations $i_{max}$.

Repeat

1) Compute the optimal $\{\lambda_n\}_{n=1}^N$ with Algorithm I.

2) Solve for $\{b_k\}_{k=1}^K$ using (10).

3) Update the MAMSE receivers $\{w_k\}_{k=1}^K$ with (4).

4) Calculate the objective function of (5).

Until convergence.

In our simulation, we declare the convergence of this algorithm when $\bar{\xi} - \bar{\xi}_{i+1} < 10^{-6}$, where $\bar{\xi}_i$ is the achieved weighted sum AMSE at the $i$th iteration of Algorithm II.

Convergence: It can be shown that at each iteration of Algorithm II, the objective function of (5) is non-increasing. Since we are interested to get any local optimal $\{b_k, w_k\}_{k=1}^K$ that yields the local minimum $\bar{\xi}$, the convergence analysis of Algorithm II with respect to the optimization variables $\{b_k, w_k\}_{k=1}^K$ is not required.

Implementation of Algorithm II: For the implementation of this distributed algorithm, for simplicity, it is assumed that $L = N = K$, and $P1$ is solved in a central controller which has $K$ parallel processors. We can implement Algorithm II distributively as follows.

Initialization: Each processor sets $\{w_k\}_{k=1}^K$ as in Algorithm II and $\{\lambda_n = 1\}_{n=1}^N$.

1) With the current $\{\lambda_n\}_{n=1}^N$ and $\{w_k\}_{k=1}^K$, the $n$th processor computes $g_n$, using (17) and updates its $\lambda_n$ by (18), $\forall n$. Then, $\{\lambda_n\}_{n=1}^N$ are shared among all processors. These two steps are repeated until $\{\lambda_n\}_{n=1}^N$ are found to be optimal.

2) Using $\{\lambda_n\}_{n=1}^N$ of step 1, the $k$th processor computes the optimal $b_k$ by (10), $\forall k$ and $\{b_k\}_{k=1}^K$ are shared among all processors. Again, using these precoders, the $k$th processor computes $w_k$ with (4), $\forall k$ and $\{w_k\}_{k=1}^K$ are shared to processors.

3) Steps (1) and (2) are repeated until Algorithm II is convergent.

4) The controller finally sends the optimal precoders and decoders to the corresponding BSs and MSs, respectively.

Note that in some scenario we might be obliged to design the precoders and decoders of all users without a central controller. In this case, one can apply the above implementation approach just by replacing the role of processors with that of BSs.

\(^1\)Note that since the aim of problem (12) is to get any $\{\lambda_n\}_{n=1}^N$, which achieves the smallest objective function of (12), we believe that the convergence analysis of Algorithm I with respect to the optimization variables $\{\lambda_n\}_{n=1}^N$ is not required.
B. Robust power minimization problem (P2)

The robust power minimization constrained with the MSE target of each user and the power of each BS antenna problem is formulated as

\[
\min_{\{b_k, w_k\}_{k=1}^K} \sum_{k=1}^K b_k^H b_k,
\]

s.t. \( \sum_{k=1}^K b_k b_k^H n_n \leq p_n, \quad \sum_{k=1}^K b_k b_k^H n_n \leq \eta_k, 0 < \eta_k < 1, \forall k \) (19)

where \( \eta_k \) is the \( k \)th user MSE target.

1) Centralized precoder design for P2: For fixed receivers \( \{w_k\}_{k=1}^K \), using (3) and applying the same technique as in (7), the above problem can be equivalently expressed as [17]

\[
\min_{B, \bar{\chi}} \bar{\chi}
\]

s.t. \( \|\text{vec}(B)\|_2 \leq \bar{\chi}, \|b_n\|_2 \leq \sqrt{p_n}, \forall n \)

\[
\left\| \left( H_k^H b_k - \theta_k \right) ; \text{vec} \left( \sqrt{R_{kk}} B w_k \right) \right\|_2 \leq \sqrt{\gamma_k - \sigma_k w_k^H w_k}, \forall k
\]

where \( \theta_k \) is a column vector of size \( K \) with the \( k \)th element equal to 1 and all the other elements equal to 0. The above problem is a SOCP for which the global optimal solution can be obtained with \( O(\sqrt{(K+N+1)(2KN+1)^2(2K^2 + 2KN^2 + 4KN)}) \) computational load [29].

2) Distributed precoder design for P2: For fixed \( \{w_k\}_{k=1}^K \), like in P1, we utilize the Lagrangian dual decomposition method to solve P2 distributively. The Lagrangian function of (19) is given as

\[
L(\lambda, \bar{\nu}, B) = \sum_{k=1}^K \left\{ b_k^H (I_N + \sum_{i=1}^K \nu_i \hat{h}_i w_i w_i^H \hat{h}_i^H + \nu_i w_i w_i^H R_{bi}) + \lambda b_k - 2 \Re \{ \nu_k w_k^H \hat{h}_k^H b_k \} + \sigma_k^2 \nu_k w_k^H w_k + \nu_k (1 - \eta_k) \right\} - \sum_{n=1}^N \lambda_n p_n
\]

where \( \lambda \) and \( \bar{\nu} = \text{diag}(\nu_1, \cdots, \nu_K) \) are the Lagrangian multipliers for the first and second constraint sets of (19), respectively. The dual function of (19) is computed by

\[
g(\lambda, \bar{\nu}) = \min_{\{b_k\}_{k=1}^K} L(\lambda, \bar{\nu}, B)
\]

\[
= \sum_{k=1}^K \nu_k \alpha_k - \sum_{k=1}^K \nu_k w_k^H \hat{h}_k^H \bar{A}^{-1} \hat{h}_k w_k - \sum_{n=1}^N \lambda_n p_n
\]

where \( \alpha_k = 1 + \sigma_k^2 w_k^H w_k - \eta_k, \bar{A} = I_N + \sum_{i=1}^K \sqrt{\nu_i} \hat{h}_i w_i w_i^H \hat{h}_i^H + \nu_i w_i w_i^H R_{bi} + \lambda \) and the second equality is obtained after substituting the optimum \( b_k \) which is given by

\[
b_{k}^* = \nu_k \bar{A}^{-1} \hat{h}_k w_k, \forall k. \quad (23)
\]

From (4) and (23) it can be clearly seen that if (19) is feasible, \( \{v_k > 0\}_{k=1}^K \) must be satisfied. Furthermore, for given \( \{\lambda_n\}_{n=1}^N \) and \( \{\eta_k\}_{k=1}^K \), the precoder of each user can be optimized distributively by using (23). The optimal \( \{\lambda_n\}_{n=1}^N \) and \( \{\nu_k\}_{k=1}^K \) of (21) can be obtained by solving the dual problem of (19) which can be expressed as

\[
\begin{align*}
\max_{\{\lambda_n \geq 0\}_{n=1}^N, \{\nu_k > 0\}_{k=1}^K} & \quad g(\lambda, \bar{\nu}) = \sum_{k=1}^K \nu_k \alpha_k - w_k^H \hat{h}_k^H \bar{A}^{-1} \hat{h}_k w_k \nu_k^2 \\
\text{s.t.} & \quad \max_{\{\lambda_n \geq 0\}_{n=1}^N, \{\nu_k > 0\}_{k=1}^K} \sum_{k=1}^K \nu_k \alpha_k - w_k^H \hat{h}_k^H \bar{A}^{-1} \hat{h}_k w_k \nu_k^2 \\
& \quad - \sum_{n=1}^N \lambda_n p_n. \quad (24)
\end{align*}
\]

It can be shown that the above problem can be formulated as SDP [18]. Thus, (24) can be solved by using convex optimization tools with complexity on the order of \( O(\sqrt{K(N + 1)}(N + 2K)^2 K(N + 1)^2) \). However, our interest is to obtain the optimal values of \( \{\lambda_n\}_{n=1}^N \) and \( \{\nu_k\}_{k=1}^K \) of the above problem distributively. The above problem can be rewritten as

\[
\min_{\{\lambda_n \geq 0\}_{n=1}^N, \{\nu_k > 0\}_{k=1}^K} \sum_{k=1}^K \nu_k w_k^H \hat{h}_k^H (\mathbf{W} \mathbf{Y} \mathbf{W}^H + \mathbf{I})^{-1} \hat{h}_k w_k
\]

\[
\quad + \sum_{n=1}^N \lambda_n + \sum_{k=1}^K \nu_k \alpha_k p_n
\]

\[
\quad + \sum_{n=1}^N \lambda_n + \sum_{k=1}^K \nu_k \alpha_k p_n \quad (25)
\]

where \( \mathbf{Y} = \text{blkdiag}(\nu, \lambda), \mathbf{W} = [\mathbf{W}_1, \cdots, \mathbf{W}_K] \), \( \nu = \text{blkdiag}(\nu_1, \cdots, \nu_K) \), \( \mathbf{W}_k = (w_k \hat{h}_k \hat{h}_k^H + \mathbf{R}_{bk})^{1/2} \). Now, by applying matrix fractional minimization of [18] on the first sum terms of the above problem, we can reformulate (25) as (see page 198 of [18])

\[
\min_{\{g_k, \bar{t}_k, \nu_k\}_{k=1}^K, \{\lambda_n\}_{n=1}^N} \sum_{k=1}^K g_k^H \mathbf{Y}^{-1} g_k + \bar{t}_k^H \bar{t}_k + \sum_{n=1}^N \lambda_n p_n
\]

\[
\quad - \sum_{k=1}^K \nu_k \bar{t}_k, \quad \text{s.t.} \quad \bar{t}_k + \mathbf{W} g_k = \nu_k \hat{h}_k w_k, \forall k. \quad (26)
\]

From the equality constraint of (26), we get \( \bar{t}_k = \nu_k \hat{h}_k w_k - \mathbf{W} g_k \). By substituting this \( \bar{t}_k \) into the objective function of the above problem, (26) can be rewritten as

\[
\min_{\{g_k, \nu_k\}_{k=1}^K, \{\lambda_n\}_{n=1}^N} \sum_{k=1}^K g_k^H \mathbf{Y}^{-1} g_k + \nu_k \hat{h}_k w_k - \mathbf{W} g_k^H
\]

\[
\quad + \sum_{k=1}^N \lambda_n p_n - \sum_{k=1}^K \nu_k \alpha_k \nu_k \quad (27)
\]

For fixed \( \{\nu_k\}_{k=1}^K \) and \( \{\lambda_n\}_{n=1}^N \), the optimal \( g_k \) of problem (27) is given by

\[
g_k = \nu_k \mathbf{Y}^{-1} + \mathbf{W} \mathbf{W}^H + \hat{h}_k w_k \nu_k
\]

\[
= \nu_k \mathbf{Y} \mathbf{W}^{H, \mathbf{A}^{-1}} \hat{h}_k w_k, \forall k \quad (28)
\]

where the second equality is obtained by employing matrix inversion Lemma [28]. To develop distributed algorithm for the above problem, we introduce the following variables: \( \mathbf{G}_k \in \mathcal{C}^{N \times K} \) as the \( \mathbf{G}_n^{(N(k-1)+1:K)} \) submatrix of \( \mathbf{G}^* = [\mathbf{g}_1, \cdots, \mathbf{g}_K] \) and \( \{\nu_k\}_{k=1}^K \) as the \( K+N \) row
of $\tilde{G}^*$. For the given $\{v_k\}_{k=1}^K$ and $\{\lambda_n\}_{n=1}^N$, $\tilde{G}_k^*$ and $(u_n^*)^H$ can also be computed as

$$\tilde{G}_k^* = v_k \tilde{W}_k^H \tilde{\Gamma}_k, \quad \forall k, \quad u_n^* = \lambda_n \tilde{\Gamma}_n, \quad \forall n \quad (29)$$

where $\tilde{\Gamma} = \tilde{A}^{-1} \tilde{W} \tilde{\nu}$ and $\tilde{\Gamma}_n$ is the $n$th row of $\tilde{\Gamma}$.

Now, we solve problem (27) distributively as follows. First, for fixed $\{v_k\}_{k=1}^K$ and $\{\lambda_n\}_{n=1}^N$, the optimal $\tilde{g}_k^*$ of (27) and the introduced variables $(\tilde{G}_k^*, \tilde{u}_n^*)$ are computed using (28) and (29), respectively. Then, using these $\tilde{g}_k^*, \tilde{G}_k^*$ and $\tilde{u}_n^*$, $v_k$ and $\lambda_n$ are updated independently and distributively by

$$v_k^* = \min_{\nu_k \geq 0} \nu_k^2 \rho_{k1} - \nu_k \rho_{k2} + \frac{\rho_{k3}}{\nu_k} \quad (30)$$

$$\lambda_n^* = \sqrt{\frac{\rho_{nn}}{\nu_n}}, \quad \forall n \quad (31)$$

where $\rho_{k1} = \tilde{h}_k^H \tilde{w}_k \tilde{h}_k$, $\rho_{k2} = 2 \Re \{v_k^* \tilde{h}_k^H \tilde{W} \tilde{g}_k^*\} + \alpha_k$, $\rho_{k3} = \text{tr}(\tilde{G}_k^* (\tilde{G}_k^*)^H)$ and $\rho_{nn} = \langle \tilde{u}_n^* \rangle^H \tilde{u}_n^*$. If $\rho_{k1} \neq 0$, by applying first order derivative, it can be shown that (30) has exactly one real solution which is given by [31]

$$v_k^* = \frac{1}{6 \rho_{k1}} \left[ \rho_{k2} + \mu_1 + \mu_2 \right], \quad \forall k \quad (32)$$

where $\mu_1 = \frac{3}{4} \left[ 2 \rho_{k2} + c_k - \zeta_k \right], \mu_2 = \frac{1}{4} \left[ 2 \rho_{k2} + c_k + \zeta_k \right]$, $c_k = 108 \rho_{k1} \rho_{k3}$ and $\zeta_k = \sqrt{c_k (4 + 4 \rho_{k2})}$. One can easily see that $\rho_{k1} \geq 0, \mu_1 \geq 0, \mu_2 \geq 0$ and $\rho_{k2} > 0$. Moreover, when $\rho_{k1} > 0$, $v_k^*$ of (32) is always positive. To summarize, (25) can be solved distributively in an iterative manner as in Algorithm III.

**Algorithm III**: Distributed algorithm to solve (25)

1) Initialization: Set $\{\lambda_n = 1\}_{n=1}^N$ and $\{v_k = 1\}_{k=1}^K$.

2) Repeat

3) With the current $\{\lambda_n\}_{n=1}^N$ and $\{v_k\}_{k=1}^K$, compute $\tilde{g}_k^*, \tilde{G}_k^*$ and $\tilde{u}_n^*$ using (28) and (29), $\forall n, \forall k$. Then, update $\lambda_n$ and $v_k$ by (31) and (32), respectively, $\forall n, \forall k$.

3) Share the above $\{\lambda_n\}_{n=1}^N$ and $\{v_k\}_{k=1}^K$ among all processors.

4) Calculate the objective function of (25).

Until convergence.

**Feasibility study for $\mathcal{P}_2$**: The problem (19) is infeasible, if there exists at least one MS with either $\rho_{k1} = 0$ or $v_k^* \gg 0$. This can be justified as follows. For the former case (i.e., $\exists k$ such that $\rho_{k1} = 0$), one can easily verify that (19) is infeasible. For the latter case (i.e., when $\{v_k^* > 0\}_k^K$ and $\exists k$ such that $v_k^* \gg 0$), although $v_k^*$ are not permitted to be $\infty$, $v_k^*$ can be arbitrarily very large number. And when $v_k^*$ is large, one can use (23) to show that the $k$th MS needs additional power at least in one of the BS antennas to satisfy its AMSE target. This case corresponds to the scenario where (25) is an unbounded problem. During the iterative stages of Algorithm III, when (19) is infeasible, we have observed from the simulation results that $(v_k)_{k=1}^K$ of (32) increases rapidly for at least one MS and the latter algorithm never converges to a point.

**Convergence**: If $\mathcal{P}_2$ is feasible, it can be shown that Algorithm III is guaranteed to converge. However, we are not able to show the global optimality of Algorithm III analytically. Nonetheless, in all simulation results we observe that the optimal $\lambda$ and $\nu$ of (25) obtained by the latter algorithm and SDP method are the same.

**Computational complexity**: As can be seen from (28) and (29), in the proposed distributed algorithm, the main computational load comes from the computation of $\tilde{A}^{-1}$ which can also be computed efficiently with $O(N^2.376)$ [30]. Moreover, in all of our simulation results, we have observed that Algorithm III converges to an optimal solution within few iterations.

Once we get the optimal $\{\lambda_n\}_{n=1}^N$ and $\{v_k\}_{k=1}^K$, like in $\mathcal{P}_1$, the precoders and decoders of each user can be optimized using (23) and (4), respectively. It follows that $\mathcal{P}_2$ (19) can be solved distributively like in Algorithm II of $\mathcal{P}_1$.

V. EXTENSION TO MIMO COORDINATED BASE STATION SYSTEMS

For multiuser MIMO coordinated BS systems, the solution approaches of Section IV can be applied to solve the following problems. (1) The robust minimization of symbol (user) wise weighted sum MSE with per BS antenna power constraint problem. (2) The robust minimization of the total sum power of all BSs with per BS antenna power and per symbol (user) MSE target constraints. In this section, we examine the robust symbol wise weighted sum MSE minimization with a per BS antenna power constraint problem for the multiuser MIMO coordinated BS systems ($\mathcal{P}_1$) only. For the MIMO coordinated BS systems, the channel estimation technique of Section III can be utilized. Upon doing so, the true channel between the $l$th BS and the $k$th user $H_{lk}^H$ and its MMSE estimate $\hat{H}_{lk}^H$ are related by [11]

$$H_{lk}^H = \hat{H}_{lk}^H + E_{blk}^H R_{lk}^{1/2} = \hat{H}_{lk}^H + B_{lk}^H \quad (33)$$

where $B_{lk}^H$ is $C_{Mk \times N}$ and $M_k$ are the estimation error matrix and number of antennas of the $k$th MS, respectively, and the entries of $E_{blk}^H$ are i.i.d with $CN(0, \sigma^2_{blk})$. Like in (3), the AMSE of the $k$th MS $l$th symbol ($\xi_{kl}$) can be expressed as

$$\xi_{kl} = 1 + w_{kl}^H (\hat{H}_{lk}^H \sum_{j=1}^K S_j b_{jm} b_{jm}^H \hat{H}_{lk}^H + \text{tr}(R_{lk} \sum_{j=1}^K S_j b_{jm} b_{jm}^H)) + \sigma^2_{blk} w_{kl} \tilde{h}_{lk}^H b_{kl}^H \quad (34)$$

where $\tilde{h}_{lk}^H$ is $C_{N \times 1}$ and $w_{kl}^H \in C_{Mk \times N}$ are the precoder and decoder vectors of the $k$th MS $l$th symbol, respectively, and $\hat{H}_{lk}^H = \{\hat{H}_{lk}^H, \cdots, \hat{H}_{lk}^H\} \in C_{Mk \times N}$. Using (33) and (34), we can formulate $\mathcal{P}_1$ as

$$\min_{(w_{kl}, b_{kl})} \sum_{k=1}^K \sum_{l=1}^L \eta_{kl} \xi_{kl},$$

s.t. $\sum_{k=1}^K \sum_{l=1}^L b_{kl} b_{kl}^H \leq p_n, \quad \forall n \quad (35)$

Note that all the other problems of this section can be examined like in $\mathcal{P}_1$. 

8
where \( \eta_{ki} \) is the AMSE weighting factor of the \( k \)th MS \( i \)th symbol. For given precoder vectors \( \{b_{ki}, \forall i\}_{k=1}^{K} \), the receivers \( \{w_{ki}, \forall i\}_{k=1}^{K} \) of the above problem can be optimized by the following MAM approach

\[
\begin{align*}
\mathbf{w}_{ki} = & \left( \mathbf{H}_{ki}^{H} \sum_{j=1}^{K} \sum_{m=1}^{S_j} b_{jm} b_{jm}^{H} \mathbf{H}_{ki}^{H} \right)^{-1} \mathbf{H}_{ki}^{H} \mathbf{b}_{ki}, \forall k, i.
\end{align*}
\]

Next, for fixed \( \{b_{ki}, \forall i\}_{k=1}^{K} \) we summarize the centralized and distributed precoder design algorithms of \( \mathcal{P}1 \).

### A. Centralized precoder design of \( \mathcal{P}1 \)

By employing (34), \( \sum_{k=1}^{K} \sum_{i=1}^{S_k} \xi_{ki} \) can be written as a quadratic expression like that of (6). This shows that the precoder design problem of (35) can be formulated as SOCP for which the global optimal solution can be obtained using convex optimization tools (see also [17]).

### B. Distributed precoder design of \( \mathcal{P}1 \)

Here like in \( \mathcal{P}1 \), the precoder design problem of \( \mathcal{P}1 \) can be solved distributively by applying the Lagrangian dual decomposition and modified matrix fractional minimization approaches. After some straightforward steps, the Lagrangian function associated with \( \mathcal{P}1 \) can be expressed as

\[
L(\lambda, \{b_{ki}, \forall i\}_{k=1}^{K}) = \sum_{k=1}^{K} \sum_{i=1}^{S_k} \left\{ b_{ki}^{H} \mathbf{A} b_{ki} - \eta_{ki} w_{ki}^{H} \mathbf{H}_{ki}^{H} \mathbf{b}_{ki} - \eta_{ki} w_{ki}^{H} \mathbf{H}_{ki}^{H} \mathbf{b}_{ki} + \eta_{ki} w_{ki}^{H} \mathbf{H}_{ki}^{H} \mathbf{b}_{ki} + \sigma_{ki}^{2} \eta_{ki} w_{ki}^{H} \mathbf{H}_{ki}^{H} w_{ki} + \eta_{ki} \right\} - \sum_{n=1}^{N} \lambda_{n} p_{n} \tag{37}
\]

where \( \lambda = \text{diag}(\lambda_1, \cdots, \lambda_N) \) are the Lagrangian multipliers corresponding to the constraint sets of (35) and \( \mathbf{A} = \sum_{j=1}^{K} \sum_{m=1}^{S_j} \eta_{jm} \mathbf{H}_{ji} w_{jm}^{H} \mathbf{H}_{ji} + \text{tr}\{w_{jm}^{H} \mathbf{w}_{jm}\} \mathbf{R}_{jk} + \lambda \).

By employing the above expression and after some mathematical manipulations, the dual problem of (35) can be formulated as

\[
\begin{align*}
\max_{\{\lambda_{n}\geq0\}_{n=1}^{N}} \min_{\{b_{ki}, \forall i\}_{k=1}^{K}} & L(\lambda, \{b_{ki}, \forall i\}_{k=1}^{K}) = \\
\max_{\{\lambda_{n}\geq0\}_{n=1}^{N}} \sum_{k=1}^{K} \sum_{i=1}^{S_k} & \left\{ \eta_{ki}(\sigma_{ki}^{2} w_{ki}^{H} w_{ki} + 1) - \sigma_{ki}^{2} w_{ki}^{H} \mathbf{H}_{ki}^{H} \mathbf{H}_{ki} w_{ki}^{H} \mathbf{H}_{ki}^{H} \mathbf{b}_{ki} + \eta_{ki} \right\} - \sum_{n=1}^{N} \lambda_{n} p_{n}. \tag{38}
\end{align*}
\]

This problem has exactly the same structure at that of (11). Thus, with the help of Lemma 1, we can develop distributed algorithm to solve the above problem. Consequently, \( \mathcal{P}1 \) can be solved distributively like that of \( \mathcal{P}1 \).

### VI. Simulation Results

In this section, we present the simulation results for problems \( \mathcal{P}1 \) and \( \mathcal{P}2 \). The spacial antenna correlation matrix between the \( t \)th BS and the \( k \)th user \( \mathbf{R}_{tk} \) is taken from a widely used exponential correlation model as \( \mathbf{R}_{tk} = \rho_{tk}^{\|j-i\|} \mathbf{I}_{K}, \forall l \), where \( 0 \leq \rho_{tk} < 1 \) and \( i \neq j \) \( \in \mathcal{N} \), and \( \{\sigma_{e1k} = \sigma_{c1k} = \cdots = \sigma_{e2k} = \sigma_{c2k} = \sigma_{e1k} = \sigma_{c1k} \}_{k=1}^{K} \). We have used exponential correlation model because of the following two reasons. First, exponential correlation model is physically reasonable in a way that the correlation between two transmit antennas decreases as the distance between them increases [32]. Second, this model is a widely used antenna correlation model for an urban area communications [26].

#### A. Simulation results for \( \mathcal{P}1 \)

In this subsection, we consider a system with \( L = 2 \) BSs where each BS has \( 2 \) antennas and \( K = 4 \) MSs. We use \( \rho_{b11} = 0.25, \rho_{b12} = 0.5, \rho_{b13} = 0.2, \rho_{b14} = 0.4, \rho_{b21} = 0.6, \rho_{b22} = 0.1, \rho_{b23} = 0.8 \) and \( \rho_{b24} = 0.15 \).

![Fig. 2. The average power utilized by the first antenna of BS 2 for the robust centralized, robust distributed, non-robust centralized [17] and non-robust distributed [20] designs when \( \rho_{b11} = 0.25, \rho_{b12} = 0.5, \rho_{b13} = 0.2, \rho_{b14} = 0.4, \rho_{b21} = 0.6, \rho_{b22} = 0.1, \rho_{b23} = 0.8 \) and \( \rho_{b24} = 0.15 \).](image)

- **Robust Centralized Algorithm**
- **Robust Distributed Algorithm**
- **Non-Robust Centralized Algorithm**
- **Non-Robust Distributed Algorithm**

This behavior has also been observed in [17] and [20] where the sum MSE minimization with per BS antenna power constraint problem is examined.

- **Simulation results for \( \mathcal{P}2 \)**

In this subsection, we consider a system with \( L = 2 \) BSs where each BS has \( 2 \) antennas and \( K = 4 \) MSs. We use \( \rho_{b11} = 0.25, \rho_{b12} = 0.5, \rho_{b13} = 0.2, \rho_{b14} = 0.4, \rho_{b21} = 0.6, \rho_{b22} = 0.1, \rho_{b23} = 0.8 \) and \( \rho_{b24} = 0.15 \).
of BS$_2$ for both of these designs are not necessarily the same. Although the robust and non-robust designs do not use the same power at each antenna, we have noticed at all SNR values that the average total sum power of all antennas of the latter designs are almost the same. In the sequel, we compare the performance of the robust centralized, robust distributed, non-robust centralized and non-robust distributed algorithms in terms of sum AMSE. For this purpose, we define the signal-to-noise ratio (SNR) as $P_{\text{sum}}/\sigma^2$, where $P_{\text{sum}}$ is the total sum power utilized by all BS antennas of the robust distributed algorithm and $\sigma^2$ is the noise variance. The SNR is controlled by varying $\sigma^2$.

We first compare the performance of the robust centralized and robust distributed algorithms in terms of sum AMSE. Fig. 3 shows that the robust centralized and distributed algorithms achieve the same sum MSE. Next, we compare the performance of the robust and non-robust algorithms with the non-robust design of [20]. As can be seen from Fig. 3, the proposed robust designs have better performance than that of the non-robust design in [20] and this improvement is better at high SNR regions.

To see the effect of antenna correlation matrix on the achievable sum AMSE of robust and non-robust designs, we change the latter $\{\rho_{blk}\}_{k=1}^{K}$ to $\rho_{011} = 0.35, \rho_{012} = 0.5, \rho_{013} = 0.3, \rho_{014} = 0.4, \rho_{021} = 0.6, \rho_{022} = 0.2, \rho_{023} = 0.8$ and $\rho_{024} = 0.25$. For this setting, we plot the total sum AMSE of our robust design and the non-robust design of [20] in Fig. 4. By examining this figure and Fig. 3, one can see that the sum AMSE increases as the antenna correlation coefficient increases for both the robust and non-robust designs. This is because when $\{\rho_{blk}, l, k\}$ increases, the number of symbols with low channel gain increases (this can be easily seen from the eigenvalue decomposition of $R_{blk}$). Consequently, for a given SNR value, the total sum AMSE also increases.

The above scenario gracefully fits to that of [12] where the robust weighted sum MSE minimization with a total BS power constraint problem is examined for the conventional downlink MIMO systems.

### B. Simulation results for $P2$

For the simulation result of $P2$, we also consider a system with $L = 2$ BSs where each BS has 2 antennas and $K = 4$ MSs. We use $\rho_{011} = 0.25, \rho_{012} = 0.5, \rho_{013} = 0.2, \rho_{014} = 0.4, \rho_{021} = 0.6, \rho_{022} = 0.1, \rho_{023} = 0.8$ and $\rho_{024} = 0.15$, and $\sigma^2_{e1} = 0.01, \sigma^2_{e2} = 0.02, \sigma^2_{e3} = 0.03$ and $\sigma^2_{e4} = 0.04$. It is assumed that $\{\sigma^2_k = \sigma^2\}_{k=1}^{K}, \{\theta_n = 15\}_{n=1}^{4}$ and the AMSE target of each user is set to $\{\xi_k = 0.2\}_{k=1}^{K}$. For better explanation of the simulation results of this subsection, we use the following channel estimates obtained from the above settings. The rows of $H_{ijkl}$ represent the channel estimate between all BSs and the $k$th MS. We first compare the performance of the robust centralized and robust distributed algorithms in terms of the total sum power of all BSs. Fig. 5 shows that the robust centralized and distributed algorithms utilize the same total power. Then, we compare the performance of the proposed robust design and the non-robust design of [20]$\dagger$ in terms of total sum power of all BSs which is also plotted.

$\dagger$Note that the feasible initial $\{w_k\}_{k=1}^{K}$ for problem $P2$ can be obtained from the solution of $P1$.

$\dagger$For $P2$, we have also shown in the latter paper that [17] and [20] have the same performance.
in Fig. 5. This figure shows that the robust design utilizes more power than that of the non-robust design for all noise variances. Now, for the total power given in the latter figure, the AMSE of each user for both designs are plotted in Fig. 6. This figure shows that for all noise variances, the non-robust design does not satisfy the AMSE requirement but the proposed robust design ensures the AMSE requirement efficiently. To ensure the latter AMSE, however, the robust design utilizes more power than that of the non-robust design. This shows that Fig. 6 does not reveal the actual performance of the proposed robust design. Thus, for a fair comparison, we tune \( \{\sigma^2_{ek}\}_{k=1}^K \) such that the total power utilized by the robust and non-robust designs are the same (or very close to each other)\(^8\) and then we compare the performance of these two designs by their achieved AMSEs of each user. For the power requirement of Fig. 7, the AMSE of each user is plotted in Fig. 8. The latter figure shows that for the same total power, the robust design still outperforms the non-robust design.

For \( \mathcal{P}2 \), we have noticed that for each channel estimate different error variance (after numerical tuning) is required to get the same total sum powers for the robust and non-robust designs. However, the performance behavior exhibited in all channel estimates fits to that of \( \hat{H}_{P2}^H \). The detail simulation results of this problem for the other channel estimates are omitted to reduce redundancy. Moreover, to see the effect of antenna correlation factor for \( \mathcal{P}2 \), we use the previously mentioned AMSE targets (i.e., \( \{\varepsilon_k = 0.2\}_{k=1}^K \)). With these AMSE targets, when we increase the antenna correlation factor, we observe that the total power requirement of the whole network also increases. The simulation results which show this fact has not been included for conciseness.

**C. Convergence characteristics of Algorithm I**

As we have mentioned in Section IV-A.1, the centralized algorithm to solve (7) has limited practical interest when the number of BSs and/or MSs are large. Moreover, in Section IV-A.2, the computational complexity of our distributed algorithm to solve (12) (once the complexity of (12) with Algorithm I is studied, the complexity of (7) with this algorithm is

\[
\hat{H}_{P2}^H = \begin{bmatrix}
0.2328 - 0.0868i & 0.1344 + 0.3848i & 0.2407 - 0.3118i & -0.2276 - 1.3829i \\
1.6717 + 0.4976i & 0.5254 - 0.9034i & 0.0206 + 0.1318i & 1.1292 - 0.8314i \\
-0.0426 + 1.7262i & -0.6380 - 0.5663i & -0.2765 - 0.4716i & -0.2760 - 0.1349i \\
0.3266 + 0.0882i & -0.3655 + 0.8997i & -0.5944 - 1.1556i & -0.2692 - 0.6244i
\end{bmatrix}
\]

\( (39) \)
networks. However, we are not able to compute analytically.

Due to this, we examine the convergence characteristics of Algorithm I for a 19-cell hexagonal structure coordinated BS system as in [33]. Each BS is located at the center of its cell, whereas each MS is located randomly inside these 19-cells with uniform distribution. The propagation model between each BS and MS contains two components. One is the path loss component decaying with distance, and the other one is the Rayleigh fading random component which has a zero mean and unit variance. For this simulation, we use \( \{ \eta_k = 1, \sigma^2_{elk} = 0.02, \rho_{blk} = 0.25, \forall l \} \) and all the other parameter settings are summarized as shown in Table I. For the channel realizations of these parameters, we examine the convergence characteristics of Algorithm I at different iterative stages of Algorithm II (i.e., with different \( \{ w_k \}^{K}_{k=1} \)) as shown in Fig. 9. As can be seen from this figure, Algorithm I converges to an optimal solution in less than 10 iterations.

### Table I

<table>
<thead>
<tr>
<th>Simulation parameters for convergence of Algorithm I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of BSs</td>
</tr>
<tr>
<td>Number of antennas at each BS</td>
</tr>
<tr>
<td>Transmit power of each BS antenna</td>
</tr>
<tr>
<td>Radius of each cell</td>
</tr>
<tr>
<td>Reference distance (d0)</td>
</tr>
<tr>
<td>Path loss exponent</td>
</tr>
<tr>
<td>Mean path loss at d0</td>
</tr>
<tr>
<td>Channel bandwidth</td>
</tr>
<tr>
<td>Receiver noise figure</td>
</tr>
<tr>
<td>Receiver vertical antenna gain</td>
</tr>
<tr>
<td>Receiver temperature</td>
</tr>
<tr>
<td>SNR</td>
</tr>
</tbody>
</table>

**Fig. 8.** Comparison of the AMSE achieved in each user for the robust design and non-robust design of [20] after tuning the error variance.

**Fig. 9.** Convergence characteristics of Algorithm I at different iterative stages of Algorithm II.

### D. Overall computational complexity to solve (7)

In Section IV-A.1, we have presented the worst-case computational cost of IP methods to solve (7) centrally. However, in most practical problems, IP methods require less computational cost than that of their worst-case complexities. To the best of our knowledge, computing the exact computational complexity of IP methods for this problem requires immense effort and time. Hence, we believe that such a task is beyond the scope of our current work. However, we have carried out extensive simulations to compare the computational time of our proposed distributed algorithm with that of the centralized algorithm which uses IP method. In the following, we describe the simulation platform and methodology we have used, and discuss the results.

According to [34], MOSEC is a computationally efficient optimization package which uses IP methods to solve large-scale optimization problems. Moreover, for SOCP problems, MOSEC requires less computational time than that of SeDuMi, LOQO, SDPT3 and CPLEX [35], [36]. This motivates us to compare the computational time of Algorithm I with that of MOSEC to solve (7). Our Matlab codes were run on a personal computer with 1.6 GHz, 2GB dual core processor under Windows XP. For comparison between these two algorithms, we have used a coordinated BS system with \( L = N/2, N = K, \{ \rho_{blk} = 0.25, \sigma^2_{elk} = 0.02, \sigma^2_k = 0.1, \}^{K}_{k=1} \) and \( \{ pm = 2 \}^{N}_{n=1} \). It is assumed that problem (7) has been solved by a central controller with \( K \) processors and all other parameters are taken as mentioned in the first paragraph of Section VI. Table II shows the amount of time required to solve (7) by Algorithm I and MOSEC at different iterative stages of Algorithm II (i.e., for different \( \{ w_k \}^{K}_{k=1} \)). As can be seen from Table II, our proposed distributed algorithm requires less computational time than that of MOSEC. From this table we can notice that our distributed algorithm has practical interest especially when

---

To get the computational time of Algorithm I per processor, first we get the computational time of Algorithm I by assuming one processor (i.e., personal computer), then we divide the latter computational time by \( K \).
The convergence characteristics of Algorithm III and the overall computational complexity of (20) can be studied like in Sections VI-C and VI-D, respectively.

VII. CONCLUSIONS

This paper considers the joint transceiver design for multi-user MISO systems with coordinated BSs where imperfect CSI is available at the BSs and MSs. By incorporating antenna correlation at the BSs and taking channel estimation errors into account, we solve two robust design problems. The problems are solved as follows. First, for fixed receivers, we propose centralized and novel computationally efficient distributed algorithms to jointly optimize the precoders of all users. The centralized algorithms employ the SOCP approach, whereas the distributed algorithms use the Lagrangian dual decomposition, modified matrix fractional minimization and an iterative method. Second, for fixed BS precoders, the receivers are updated by the MAMSE criterion. These two steps are repeated until convergence is achieved. Computer simulations demonstrate that our proposed distributed algorithms achieve the same performance as that of the centralized algorithms. Simulation results also verify the superior performance of the stochastic robust designs compared to that of the non-robust/naive designs.

REFERENCES

Tadilo Endeshaw Bogale (S’09) was born in Gondar, Ethiopia. He received his B.Sc and M.Sc degree in Electrical Engineering from Jimma University, Jimma, Ethiopia and Karlstad University, Karlstad, Sweden in 2004 and 2008, respectively. From 2004-2007, he was working in Ethiopian Telecommunications Corporation (ETC) in mobile project department. Since 2009 he has been working towards his PhD degree and as an assistant researcher at the ICTEAM institute, University Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium. His research interests include robust (non-robust) transceiver design for multiuser MIMO systems, centralized and distributed algorithms, and convex optimization techniques for multiuser systems.

Batu Krishna Chalise was born in Kathmandu, Nepal. He received the B. E. degree in electronics engineering from Tribhuvan University, Kathmandu, in 1998 and the M. S. and Ph.D. degrees in electrical engineering from the University of Duisburg-Essen, Duisburg, Germany, in 2001 and 2006, respectively. In 1999, he was a lecturer with the Institute of Engineering, Kathmandu. For a short period in 2001, he was a Researcher with the Fraunhofer-Institute of Microelectronic Circuits and Systems (IMS), Duisburg. From 2002-2006, he was a Research Assistant with the Department of Communication Systems, University of Duisburg-Essen, where his Ph.D. research was supported by a grant from the Ministry of Education and Science of North Rhein-Westphalia (NRW), Germany. He was a Postdoctoral Researcher with the Communication and Remote Sensing Laboratory, Université catholique de Louvain, Louvain La Neuve, Belgium from December 2006 till June 2010. Currently, he is the Postdoctoral Research Fellow with the Center for Advanced Communications, Villanova University, Villanova, USA. His research interests include cooperative and opportunistic wireless communications, robust algorithms for multi-antenna systems and convex optimization.

Luc Vandendorpe (M’93-SM’99-F’06) was born in Mouscron, Belgium, in 1962. He received the Electrical Engineering degree (summa cum laude) and the Ph.D. degree from the Universit Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1985 and 1991, respectively. Since 1985, he has been with the Communications and Remote Sensing Laboratory of UCL, where he first worked in the field of bit rate reduction techniques for video coding. In 1992, he was a Visiting Scientist and Research Fellow at the Telecommunications and Traffic Control Systems Group of the Delft Technical University, The Netherlands, where he worked on spread spectrum techniques for personal communications systems. From October 1992 to August 1997, he was Senior Research Associate of the Belgian NSF at UCL, and invited Assistant Professor. He is currently a Professor and head of the Institute for Information and Communication Technologies, Electronics and Applied Mathematics.

His current interest is in digital communication systems and more precisely resource allocation for OFDM(A)-based multicell systems, MIMO and distributed MIMO, sensor networks, turbo-based communications systems, physical layer security and UWB based positioning.

Dr. Vandendorpe was co-recipient of the 1990 Biennial Alcatel-Bell Award from the Belgian NSF for a contribution in the field of image coding. In 2000, he was co-recipient (with J. Louveaux and P. Deryck) of the Biennial Siemens Award from the Belgian NSF for a contribution about filter-bank-based multicarrier transmission. In 2004, he was co-winner (with J. Czyz) of the Face Authentication Competition, FAC 2004. He is or has been TPC member for numerous IEEE conferences (VTC, Globecom, SPAWC, ICC, PIMRC, WCNC) and for the Turbo Symposium. He was Co-Technical Chair (with P. Duhamel) for the IEEE ICASSP 2006. He was an Editor for Synchronization and Equalization of the IEEE TRANSACTIONS ON COMMUNICATIONS between 2000 and 2002, Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS between 2003 and 2005, and Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING between 2004 and 2006. He was Chair of the IEEE Benelux joint chapter on Communications and Vehicular Technology between 1999 and 2003. He was an elected member of the Signal Processing for Communications committee between 2000 and 2005, and an elected member of the Sensor Array and Multichannel Signal Processing committee of the Signal Processing Society between 2006 and 2008. Currently, he is an elected member of the Signal Processing for Communications committee. He is the Editor-in-Chief for the EURASIP Journal on Wireless Communications and Networking. L. Vandendorpe is a Fellow of the IEEE.