

Distributed Power Control in Downlink Cellular Massive MIMO Systems

Trinh Van Chien*, Emil Björnson*, Erik G. Larsson*, and Tuan Anh Le†

*Department of Electrical Engineering (ISY), Linköping University, SE-581 83 Linköping, Sweden

†Department of Design Engineering & Maths, Middlesex University London, United Kingdom

Email: {trinh.van.chien, emil.bjornson, erik.g.larsson}@liu.se, t.le@mdx.ac.uk

Abstract—This paper compares centralized and distributed methods to solve the power minimization problem with quality-of-service (QoS) constraints in the downlink (DL) of multi-cell Massive multiple-input multiple-output (MIMO) systems. In particular, we study the computational complexity, number of parameters that need to be exchanged between base stations (BSs), and the convergence of iterative implementations. Although a distributed implementation based on dual decomposition (which only requires statistical channel knowledge at each BS) typically converges to the global optimum after a few iterations, many parameters need to be exchanged to reach convergence.

I. INTRODUCTION

Massive MIMO is considered a key technology for 5G networks due to its great improvements in both spectral efficiency (SE) and energy efficiency (EE) over legacy networks [1]. Moreover, resource allocation problems in Massive MIMO are reported to have much lower complexity than that in small-scale systems owing to the fact that the ergodic SE expressions only depend on the large-scale fading coefficients, thanks to the so-called channel hardening property [2]. However, so far, most optimization problems in the Massive MIMO literature have been formulated and solved in a centralized fashion [2], [3]. This requires the network to gather full statistical channel state information (CSI) from all base stations (BSs) at one location in order to centrally allocate resources in every cell and then inform each BS of the decisions. This raises practical questions, especially for dense networks with many BSs and users, about backhaul signaling, scalability, and delays [4]. A classic approach to deal with such issues is distributed optimization [5], which transforms the centralized problem implementations where every BS simultaneously optimizes its local resources based on local information and only parameters that describe the inter-cell interference are exchanged between BSs to iteratively find the globally optimal solution.

A few recent works have applied distributed implementation concepts to Massive MIMO, e.g., [6]–[8]. For the uplink (UL) transmission, a distributed max-min fairness problem for a two-decoding-layers Massive MIMO system is studied in [6] based on the existence of an effective interference

function obeying the rigid conditions [9], [10]. By formulating a non-cooperative game, [7] proposed a distributed EE problem which has a Nash equilibrium. For the downlink (DL) transmission, by utilizing the UL-DL duality, [8] proposed a distributed framework for a total transmit power minimization problem with quality-of-service (QoS) constraints that is also seeking the optimal precoding vectors, which are functions of the small-scale fading realizations. Even though the optimal precoding vectors provides certain gains over heuristic precoding such as zero-forcing (ZF), the algorithm in [8] suffers from the fact that the small-scale fading realizations vary rapidly over both time and frequency. Furthermore, the previous algorithms assumed that the BSs have full access to the channel statistics of the neighboring cells [6] or all the other cells [7], [8], which may require heavy backhaul signaling in the backhaul since users move, new users arrive, and current users disconnect. To the best of our knowledge, no previous work has explicitly investigated what information should be exchanged between BSs in Massive MIMO to solve power minimization problems.

In this paper, we consider the DL transmission of Massive MIMO systems with maximum ratio (MR) or ZF precoding. Each user has a QoS requirement, in terms of an achievable SE, and we study the total transmit power minimization problem. We compare a centralized solution algorithm with two distributed algorithms to answer the following fundamental questions: i) Can a distributed implementation achieve the optimal solution of the centralized problem? ii) Which subset of parameters should be shared between the BSs? iii) How can we limit complexity and backhaul signaling requirements for distributed implementation?

Notations: Upper/lower bold letters are used for matrices/vectors. $(\cdot)^H$ denotes the Hermitian transpose. $\mathbb{E}\{\cdot\}$ denotes the expectation of a random variable, while $\mathcal{CN}(\cdot, \cdot)$ stands for the circularly symmetric Gaussian distribution and $\|\cdot\|$ is the Euclidean norm.

II. SYSTEM MODEL

We consider a Massive MIMO system comprising L cells, each having a BS equipped with M antennas and serving K single-antenna users. The system operates according to a time division duplex (TDD) protocol. The time-frequency resources are divided into coherence intervals of τ_c symbols where the

This paper was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 641985 (5Gwireless). It was also supported by ELLIIT and CENIIT.

channels are assumed to be static and frequency flat. The channel between user k in cell i and BS l is assumed to be uncorrelated Rayleigh fading,

$$\mathbf{h}_{i,k}^l \sim \mathcal{CN}(\mathbf{0}, \beta_{i,k}^l \mathbf{I}_M), \quad (1)$$

where $\beta_{i,k}^l$ is the large-scale fading coefficient.

During the UL channel estimation phase, each coherence interval dedicates τ_p symbols for pilot transmission. We assume that the same set of $\tau_p = K$ orthonormal pilot signals $\{\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_K\}$, with $\|\boldsymbol{\psi}_k\|^2 = \tau_p, \forall k$, is reused in each cell wherein user k in cell l allocates the power $\hat{p}_{l,k} > 0$ to its pilot signal. The received training signal at BS l is

$$\mathbf{Y}_l = \sum_{i=1}^L \sum_{t=1}^K \sqrt{\hat{p}_{i,t}} \mathbf{h}_{i,t}^l \boldsymbol{\psi}_t^H + \mathbf{N}_l, \quad (2)$$

where \mathbf{N}_l is additive noise with each element independently distributed as $\mathcal{CN}(0, \sigma_{\text{UL}}^2)$, where σ_{UL}^2 is the variance. The channel between user k in cell l and BS l is estimated by multiplying the received signal in (2) with $\boldsymbol{\psi}_k$ as

$$y_{l,k} = \mathbf{Y}_l \boldsymbol{\psi}_k = \sum_{i=1}^L \sum_{t=1}^K \sqrt{\hat{p}_{i,t}} \mathbf{h}_{i,t}^l \boldsymbol{\psi}_t^H \boldsymbol{\psi}_k + \mathbf{N}_l \boldsymbol{\psi}_k. \quad (3)$$

Using minimum mean square error (MMSE) estimation [11], [12], the channel estimate is distributed as

$$\hat{\mathbf{h}}_{i,k}^l \sim \mathcal{CN}(\mathbf{0}, \gamma_{i,k}^l \mathbf{I}_M), \quad (4)$$

where the variance $\gamma_{i,k}^l$ is

$$\gamma_{i,k}^l = \frac{\hat{p}_{i,k} \tau_p (\beta_{i,k}^l)^2}{\sum_{i'=1}^L \hat{p}_{i',k} \tau_p \beta_{i',k}^l + \sigma_{\text{UL}}^2}. \quad (5)$$

In this paper, the channel estimates are used to construct linear precoding vectors for the DL data transmission.

A. Downlink Data Transmission

In the DL data transmission, BS l transmits a Gaussian signal $s_{l,k}$ to its user k with $\mathbb{E}\{|s_{l,k}|^2\} = 1$. The received baseband signal at user k in cell l is

$$y_{l,k} = \sum_{i=1}^L \sum_{t=1}^K \sqrt{\rho_{i,t}} (\mathbf{h}_{i,t}^l)^H \mathbf{w}_{i,t} s_{i,t} + n_{l,k}, \quad (6)$$

where the additive noise is $n_{l,k} \sim \mathcal{CN}(0, \sigma_{\text{DL}}^2)$, $\rho_{i,t}$ is the power allocated to user t in cell i for transmission of the data symbol $s_{i,t}$ and $\mathbf{w}_{i,t}$ is the corresponding normalized linear precoding vector:

$$\mathbf{w}_{l,k} = \begin{cases} \frac{\hat{\mathbf{h}}_{l,k}^l}{\sqrt{\mathbb{E}\{\|\hat{\mathbf{h}}_{l,k}^l\|^2\}}}, & \text{for MR,} \\ \frac{\hat{\mathbf{H}}_l^H \mathbf{r}_{l,k}}{\sqrt{\mathbb{E}\{\|\hat{\mathbf{H}}_l^H \mathbf{r}_{l,k}\|^2\}}}, & \text{for ZF,} \end{cases} \quad (7)$$

where $\hat{\mathbf{H}}_l^l = [\hat{\mathbf{h}}_{l,1}^l, \dots, \hat{\mathbf{h}}_{l,K}^l] \in \mathbb{C}^{M \times K}$ is the estimated channel matrix of the K users in cell l , $\mathbf{r}_{l,k}$ is the k th column of $(\hat{\mathbf{H}}_l^H \hat{\mathbf{H}}_l^l)^{-1}$. Using standard techniques, a closed-form

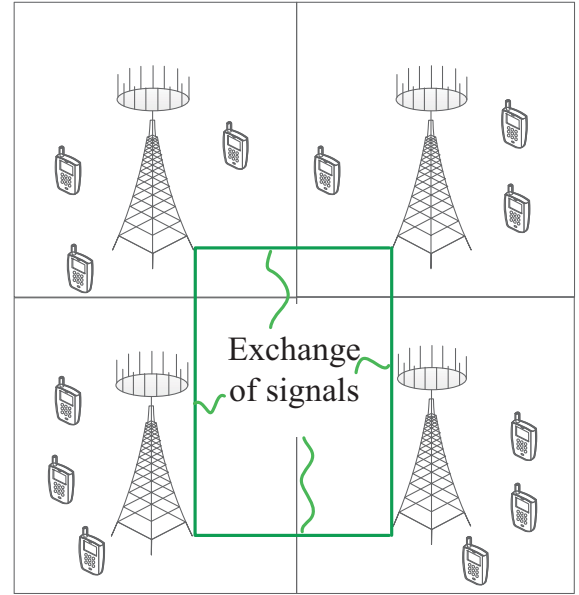


Fig. 1: A multi-cell Massive MIMO system where the problem (11) is optimized in distributed manner.

lower bound on the DL ergodic capacity with MR or ZF precoding is obtained.

Lemma 1. [13, Corollary 3] In the DL, the closed-form expression for the ergodic SE of user k in cell l is

$$R_{l,k} = \left(1 - \frac{\tau_p}{\tau_c}\right) \log_2(1 + \text{SINR}_{l,k}), \quad (8)$$

where the effective signal-to-interference-plus-noise ratio (SINR), denoted by $\text{SINR}_{l,k}$, is

$$\text{SINR}_{l,k} = \frac{G \rho_{l,k} \gamma_{l,k}^l}{G \sum_{\substack{i=1 \\ i \neq l}}^L \rho_{i,k} \gamma_{l,k}^i + \sum_{i=1}^L \sum_{t=1}^K \rho_{i,t} z_{l,k}^i + \sigma_{\text{DL}}^2}. \quad (9)$$

The parameters G and $z_{l,k}^i$ are specified by the precoding scheme. MR precoding gives $G = M$ and $z_{l,k}^i = \beta_{l,k}^i$, while ZF precoding gives $G = M - K$ and $z_{l,k}^i = \beta_{l,k}^i - \gamma_{l,k}^i$.

As the closed-form expression of the ergodic SE in (8) is independent of the small-scale fading, it can be used to solve resource allocation problems whose solutions are applicable over a long time period [6]–[8]. In this paper, we introduce a distributed implementation for the total transmit power minimization problem.

B. Total Transmit Power Minimization Problem

The total transmit power at BS l is $\sum_{k=1}^K \rho_{l,k}, \forall l$. Suppose user k in cell l has the QoS requirement $R_{l,k} \geq \xi_{l,k}$, where $R_{l,k}$ is given by (8). The total transmit power minimization

problem of the L BSs subject to these QoS requirements is formulated as

$$\begin{aligned} & \underset{\{\rho_{l,k} \geq 0\}}{\text{minimize}} && \sum_{l=1}^L \sum_{k=1}^K \rho_{l,k} \\ & \text{subject to} && R_{l,k} \geq \xi_{l,k}, \forall l, k, \\ & && \sum_{t=1}^K \rho_{l,t} \leq P_{\max,l}, \forall l, \end{aligned} \quad (10)$$

where $P_{\max,l}$ is the maximum transmit power at BS l . Converting from the SE requirements to the corresponding SINR values, i.e., set $\hat{\xi}_{l,k} = 2^{\frac{\tau_c \xi_{l,k}}{\tau_c - \tau_p}} - 1$, and then problem (10) is reformulated as

$$\begin{aligned} & \underset{\{\rho_{l,k} \geq 0\}}{\text{minimize}} && \sum_{l=1}^L \sum_{k=1}^K \rho_{l,k} \\ & \text{subject to} && \text{SINR}_{l,k} \geq \hat{\xi}_{l,k}, \forall l, k, \\ & && \sum_{t=1}^K \rho_{l,t} \leq P_{\max,l}, \forall l. \end{aligned} \quad (11)$$

Since problem (11) is a linear program [1], its optimal solution can be obtained in polynomial time by an interior-point algorithm or a simplex method, e.g., using CVX [14]. A centralized implementation requires the KL^2 large-scale fading coefficients of all channels (i.e., $\beta_{i,k}^l, \forall i, k, l$), the KL^2 variances of all the channel estimates (i.e., $\gamma_{i,k}^l, \forall i, k, l$), and the KL QoS requirements of the users to be gathered at one location in the network, which can then solve (11) to obtain the optimal power control (i.e., $\rho_{l,k}, \forall l, k$). The optimal power solution then needs to be fed back to the BSs by sending KL additional parameters over the backhaul.

C. Basic Form of Distributed Implementation

In a distributed implementation of problem (11), each BS performs data power allocation for its own users and only iteratively exchange signals with the other BSs as shown in Fig. 1. A basic form of (11) is that every BS acquires the $2K(L-1)^2 + K(L-1)$ parameters, which the centralized implementation requires, and then solves (11) locally. This removes the need for sending the solution over the backhaul.

III. DISTRIBUTED IMPLEMENTATION BY DUAL DECOMPOSITION

In this section, a distributed implementation of (11) is studied based on dual decomposition. Its convergence and the computational complexity are also investigated.

A. Assumptions for Distributed Implementation

The proposed distributed implementation for problem (11) is based on two levels of optimization: a master level and the L sublevels are schematically illustrated in Fig. 2. At Sublevel l , BS l will locally optimize the transmit powers to its K users utilizing only partial information. The following assumptions are made to allocate power for the DL transmission:

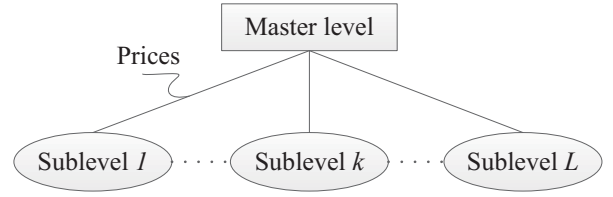


Fig. 2: Distributed implementation of problem (11) based on the dual decomposition.

- 1) BS l possesses statistical information including the large-scale fading coefficients $\beta_{l,k}^l$ and the channel estimate variance $\gamma_{i,k}^l$ of the K local users.
- 2) BS l has the statistical information of the channels to the $K(L-1)$ interfering users, i.e., $\beta_{i,k}^l, \gamma_{i,k}^l, \forall k, i \neq l$. Those can also be measured locally.
- 3) BS l jointly optimizes the DL powers allocated to its K local users based on the above prior information.
- 4) The inter-cell interference is considered as consistency parameters in a dual-decomposition approach. These are the only parameters needed to exchange between BSs and will be defined hereafter.

B. Details of the Distributed Implementation

Since the cost function in the optimization problem (11) is not strictly convex, a standard dual decomposition implementation is not guaranteed to converge [15].¹ Thus, in order to ensure the convergence of the distributed implementation, we will convert (11) into a convex problem which involves a strictly convex cost function by introducing a new variable $\tilde{\rho}_{l,k} \geq 0$ such as $\rho_{l,k} = \tilde{\rho}_{l,k}^2, \forall l, k$. The effective SINR value of user k in cell l is reformulated as

$$\begin{aligned} \text{SINR}_{l,k} &= \frac{G \tilde{\rho}_{l,k}^2 \gamma_{l,k}^l}{G \sum_{\substack{i=1 \\ i \neq l}}^L \tilde{\rho}_{i,k}^2 \gamma_{l,k}^i + \sum_{i=1}^L \sum_{t=1}^K \tilde{\rho}_{i,t}^2 z_{l,k}^i + \sigma_{\text{DL}}^2} \\ &= \frac{G \tilde{\rho}_{l,k}^2 \gamma_{l,k}^l}{\sum_{t=1}^K \tilde{\rho}_{l,t}^2 z_{l,k}^t + \sum_{\substack{i=1 \\ i \neq l}}^L \left(G \tilde{\rho}_{i,k}^2 \gamma_{l,k}^i + \sum_{t=1}^K \tilde{\rho}_{i,t}^2 z_{l,k}^i \right) + \sigma_{\text{DL}}^2}. \end{aligned} \quad (12)$$

In the last equation of (12), the first term in the denominator represents the mutual non-coherent interference of the K local users served by BS l . Hence, BS l can locally evaluate this term. The second term includes the coherent interference caused by the users utilizing nonorthogonal pilot signals and the non-coherent interference from all users in the other cells. If BS l wishes to evaluate the SE of each local user, it will need to obtain this information from the other $L-1$ BSs (i.e., the prices in Fig. 2) to compute the

¹A function f is strictly convex if for two variables $x_1 \neq x_2$ in the feasible domain of f and a scalar $\alpha \in (0, 1)$ that gives $\alpha x_1 + (1-\alpha)x_2$ also in the feasible domain, then $f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha)f(x_2)$.

second part. In order to reduce the amount of exchanged information among BSs, we introduce so-called consistency variables $\theta_{l,k}^i$ and $\tilde{\theta}_{l,k}^i$ which represent the exact and believed value of $G\tilde{\rho}_{l,k}^2\gamma_{l,k}^i + \sum_{t=1}^K \tilde{\rho}_{l,t}^2 z_{l,k}^i$, respectively. Thus, (11) is reformulated as

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^L \sum_{t=1}^K \tilde{\rho}_{l,t}^2 && (13a) \\ & \{\tilde{\rho}_{l,k} \geq 0, \theta_{l,k}^i \geq 0, \tilde{\theta}_{l,k}^i \geq 0\} \end{aligned}$$

$$\begin{aligned} & \text{subject to} && \frac{G\tilde{\rho}_{l,k}^2\gamma_{l,k}^i}{\sum_{t=1}^K \tilde{\rho}_{l,t}^2 z_{l,k}^i + \sum_{\substack{i=1 \\ i \neq l}}^L (\tilde{\theta}_{l,k}^i)^2 + \sigma_{\text{DL}}^2} \geq \hat{\xi}_{l,k}, \forall l, k, && (13b) \end{aligned}$$

$$G\tilde{\rho}_{l,k}^2\gamma_{l,k}^i + \sum_{t=1}^K \tilde{\rho}_{l,t}^2 z_{l,k}^i \leq (\theta_{l,k}^i)^2, \forall l, i, k, i \neq l, \quad (13c)$$

$$\theta_{l,k}^i \leq \tilde{\theta}_{l,k}^i, \forall l, i, k, i \neq l, \quad (13d)$$

$$\sum_{k=1}^K \tilde{\rho}_{l,k}^2 \leq P_{\max,l}, \forall l. \quad (13e)$$

The constraints (13d) are called consistency constraints. There are in total $2L(L-1)K$ consistency variables $\theta_{l,k}^i$ and $\tilde{\theta}_{l,k}^i, \forall k, i, l$, thus (13) involves $LK(2LK-1)$ optimization variables. We stress that (11) and (13) are equivalent since $\theta_{l,k}^i = \tilde{\theta}_{l,k}^i$ at the optimum. The dual decomposition approach is used to decompose problem (13) into L subproblems that each can be solved locally at a BS. To that end, a partial Lagrangian function, which is related to the differences between the consistency variables, is formed as

$$\mathcal{L}(\{\tilde{\rho}_{l,k}, \lambda_{l,k}^i\}) = \sum_{l=1}^L \sum_{k=1}^K \left(\tilde{\rho}_{l,k}^2 + \sum_{\substack{i=1 \\ i \neq l}}^L \lambda_{l,k}^i (\theta_{l,k}^i - \tilde{\theta}_{l,k}^i) \right), \quad (14)$$

where $\lambda_{l,k}^i \geq 0$ is the Lagrange multiplier associated with the constraint $\theta_{l,k}^i \leq \tilde{\theta}_{l,k}^i$. The dual function of (14) is computed as the superposition of the L local dual functions

$$g_l(\{\lambda_{l,k}^i\}) = \sum_{i=1}^L g_l(\{\lambda_{l,k}^i\}), \quad (15)$$

where the local dual function of BS l , denoted by $g_l(\{\lambda_{l,k}^i\})$, is formulated as

$$\begin{aligned} & g_l(\{\lambda_{l,k}^i\}) = \\ & \inf_{\{\tilde{\rho}_{l,k} \geq 0, \theta_{l,k}^i \geq 0, \tilde{\theta}_{l,k}^i \geq 0\}} \sum_{k=1}^K \tilde{\rho}_{l,k}^2 + \sum_{k=1}^K \sum_{\substack{i=1 \\ i \neq l}}^L (\lambda_{l,k}^i \theta_{l,k}^i - \lambda_{l,k}^i \tilde{\theta}_{l,k}^i). \end{aligned} \quad (16)$$

From (16), problem (13) can be decomposed into the L subproblems and the l th subproblem is

$$\begin{aligned} & \text{minimize} && \sum_{k=1}^K \tilde{\rho}_{l,k}^2 + \sum_{k=1}^K \sum_{\substack{i=1 \\ i \neq l}}^L (\lambda_{l,k}^i \theta_{l,k}^i - \lambda_{l,k}^i \tilde{\theta}_{l,k}^i) \\ & \{\tilde{\rho}_{l,k} \geq 0, \theta_{l,k}^i \geq 0, \tilde{\theta}_{l,k}^i \geq 0\} && \\ & \text{subject to} && \frac{G\tilde{\rho}_{l,k}^2\gamma_{l,k}^i}{\sum_{t=1}^K \tilde{\rho}_{l,t}^2 z_{l,k}^i + \sum_{\substack{i=1 \\ i \neq l}}^L (\tilde{\theta}_{l,k}^i)^2 + \sigma_{\text{DL}}^2} \geq \hat{\xi}_{l,k}, \forall k, \\ & && G\tilde{\rho}_{l,k}^2\gamma_{l,k}^i + \sum_{t=1}^K \tilde{\rho}_{l,t}^2 z_{l,k}^i \leq (\theta_{l,k}^i)^2, \forall i, k, i \neq l, \\ & && \sum_{k=1}^K \tilde{\rho}_{l,k}^2 \leq P_{\max,l}. \end{aligned} \quad (17)$$

Notice that BS l uses $K(2L-1)$ optimization variables to solve the l th subproblem, while all the L subproblems can be processed in parallel by the L BSs for given values of all the Lagrange multipliers $\lambda_{l,k}^i$. The globally optimal solution to (17) is obtained due to its convexity as stated in Theorem 1.

Theorem 1. *The optimization problem (17) is equivalent to the following second-order cone (SOC) program*

$$\begin{aligned} & \text{minimize} && s_l \\ & \{\tilde{\rho}_{l,k} \geq 0, s_l, \{\theta_{l,k}^i \geq 0, \tilde{\theta}_{l,k}^i \geq 0\}\} && \\ & \text{subject to} && \|\hat{\mathbf{u}}_{l,k}\| \leq \tilde{\rho}_{l,k} \sqrt{\frac{G\gamma_{l,k}^i}{\hat{\xi}_{l,k}}}, \forall k, \\ & && \|\mathbf{u}_l\| \leq \frac{1}{2} \left(1 + s_l + \sum_{k=1}^K \sum_{\substack{i=1 \\ i \neq l}}^L (\lambda_{l,k}^i \tilde{\theta}_{l,k}^i - \lambda_{l,k}^i \theta_{l,k}^i) \right), \\ & && \|\tilde{\mathbf{u}}_{l,k}\| \leq \theta_{l,k}^i, \forall i, k, i \neq l, \\ & && \|\check{\mathbf{u}}_l\| \leq \sqrt{P_{\max,l}}, \end{aligned} \quad (18)$$

where $\hat{\mathbf{u}}_{l,k} \in \mathbb{C}^{K+L}$, $\mathbf{u}_l \in \mathbb{C}^{K+1}$, $\tilde{\mathbf{u}}_{l,k} \in \mathbb{C}^{K+1}$, and $\check{\mathbf{u}}_l \in \mathbb{C}^K$ are respectively defined as

$$\begin{aligned} & \hat{\mathbf{u}}_{l,k} = \\ & \left[\sqrt{z_{l,k}^i} (\tilde{\rho}_{l,1}, \dots, \tilde{\rho}_{l,K}), \tilde{\theta}_{l,k}^1, \dots, \tilde{\theta}_{l,k}^{l-1}, \tilde{\theta}_{l,k}^{l+1}, \dots, \tilde{\theta}_{l,k}^L, \sigma_{\text{DL}} \right]^T, \\ & \mathbf{u}_l = \\ & \left[\tilde{\rho}_{l,1}, \dots, \tilde{\rho}_{l,K}, \frac{1}{2} \left(1 - s_l + \sum_{k=1}^K \sum_{\substack{i=1 \\ i \neq l}}^L (\lambda_{l,k}^i \theta_{l,k}^i - \lambda_{l,k}^i \tilde{\theta}_{l,k}^i) \right) \right]^T, \\ & \tilde{\mathbf{u}}_{l,k} = \left[\tilde{\rho}_{l,k} \sqrt{G\gamma_{l,k}^i}, \sqrt{z_{l,k}^i} (\tilde{\rho}_{l,1}, \dots, \tilde{\rho}_{l,K}) \right]^T, \\ & \check{\mathbf{u}}_l = [\tilde{\rho}_{l,1}, \dots, \tilde{\rho}_{l,K}]^T. \end{aligned}$$

Proof. Utilizing the epi-graph representation [16], the objective function of (17) turns to the following constraint by introducing a new optimization variable s_l ,

$$\sum_{k=1}^K \tilde{\rho}_{l,k}^2 + \sum_{k=1}^K \sum_{\substack{i=1 \\ i \neq l}}^L (\lambda_{l,k}^i \theta_{l,k}^i - \lambda_{l,k}^i \tilde{\theta}_{l,k}^i) \leq s_l, \quad (19)$$

and it is equivalent to

$$\sum_{k=1}^K \tilde{\rho}_{l,k}^2 \leq s_l - \sum_{k=1}^K \sum_{i=1, i \neq l}^L (\lambda_{i,k}^l \theta_{i,k}^l - \lambda_{l,k}^i \tilde{\theta}_{l,k}^i). \quad (20)$$

By applying the identity $x - y = \frac{1}{4}(1+x-y)^2 - \frac{1}{4}(1-x+y)^2$, i.e., with $x = s_l$ and $y = \sum_{k=1}^K \sum_{i=1, i \neq l}^L (\lambda_{i,k}^l \theta_{i,k}^l - \lambda_{l,k}^i \tilde{\theta}_{l,k}^i)$, for the right-hand side of (20), this constraint can be converted to a SOC constraint. The other constraints of (17) can be easily converted to the corresponding SOC constraints as in (18). \square

At iteration $n + 1$, after obtaining the optimal solutions to the L subproblems in (18), every BS will update the Lagrange multipliers at the so-called master level by considering the following master dual problem:

$$\begin{aligned} & \text{maximize} && \sum_{l=1}^L \sum_{k=1}^K \sum_{\substack{i=1 \\ i \neq l}}^L \lambda_{l,k}^{i,(n+1)} \left(\theta_{l,k}^{i,*,(n)} - \tilde{\theta}_{l,k}^{i,*,(n)} \right), \\ & \{\lambda_{l,k}^{i,(n+1)} \geq 0\} \end{aligned} \quad (21)$$

where $\theta_{l,k}^{i,*,(n)}$ and $\tilde{\theta}_{l,k}^{i,*,(n)}$ are the global optimums in the n th iteration to the L subproblems in (18). The subgradient projection method [5] can be adopted to update the Lagrangian multipliers as

$$\lambda_{l,k}^{i,(n+1)} = \left[\lambda_{l,k}^{i,(n)} - \zeta^{(n)} \left(\tilde{\theta}_{l,k}^{i,*,(n)} - \theta_{l,k}^{i,*,(n)} \right) \right]_+, \forall l, i, k, \quad (22)$$

where $\zeta^{(n)}$ is a positive step-size at the n th iteration and $[\cdot]_+$ is the projection onto the nonnegative orthant. If the master problem is solved at an arbitrary BS, it acquires the $2K(L-1)^2$ consistency parameters from the $L-1$ remaining BSs. The $2K(L-1)$ updated Lagrange multiplier $\lambda_{i,k}^{l,(n)}$ and $\lambda_{l,k}^{i,(n)}$, $\forall i, l, k, i \neq L$ should be sent back to BS l in every iteration. In total, the number of exchanged parameters for each iteration is $4K(L-1)^2$. The proposed distributed implementation of problem (11) is presented in Algorithm 1.

Furthermore, the convergence property of Algorithm 1 is established in the following theorem.

Theorem 2. *If the step-size $\zeta^{(n)}$ is small enough, the distributed implementation in Algorithm 1 yields the optimal solution to (11).*

Proof. The constraint functions of problem (18) are strictly convex, so the solution to the master dual problem (21) converges to the global optimum with $\theta_{l,k}^i - \tilde{\theta}_{l,k}^i \rightarrow 0, \forall l, i, k$, [4], [17]. Consequently, the centralized and distributed implementation have the same optimal solution. \square

C. Complexity Analysis of (18)

Since problem (18) contains SOC constraints, a standard interior-point method (IPM) [18], [19] can be used to find its optimal solution. The worst-case runtime of the IPM is computed as follows.

Definition 1: For a given tolerance $\epsilon > 0$, the set of $\tilde{\rho}_{l,k}^\epsilon, s_l^\epsilon, \theta_{i,k}^{l,\epsilon}, \tilde{\theta}_{l,k}^{i,\epsilon}$ is called an ϵ -solution to problem (18) if

$$s_l^\epsilon \leq s_l^* + \epsilon, \quad (23)$$

Algorithm 1 Distributed implementation for problem (11)

Input: $P_{\max,l}, \forall l; \xi_{l,k}, \forall l, k; \beta_{l,k}^i, \forall l, i, k$; Initial Lagrange multipliers $\lambda_{l,k}^{i,(1)}, \forall l, i, k$; Set up $n = 1$ and initialize $\zeta^{(1)}$.

1. *Iteration n:*

- 1.1. BS l solves problem (18) utilizing $\lambda_{l,k}^{i,(n)}$ from the master level.
- 1.2. BS l store the currently optimal powers: $\rho_{l,k}^{*,(n)} = \tilde{\rho}_{l,k}^{*,(n)}, \forall l, k$.
- 1.2. BS l send the optimal values $\tilde{\theta}_{l,k}^{i,*,(n)}$ and $\theta_{l,k}^{i,*,(n)}, \forall l, i, k$, to the master level
- 1.3. Update $\lambda_{l,k}^{i,(n+1)}$ by using (22) and then send back to BS l .

2. If *Stopping criterion is satisfied* \rightarrow *Stop*. Otherwise, go to Step 3.

3. Set $n = n + 1$ and update $\zeta^{(n+1)}$, then go to Step 1.

Output: The optimal solutions: $\rho_{l,k}^* = \rho_{l,k}^{*,(n)}, \forall l, k$.

where s_l^* is the globally optimal solution to the optimization problem (18).

The number of decision variables of problem (18) is on the order of $m = \mathcal{O}(K(2L-1) + 1)$ where $\mathcal{O}(\cdot)$ represents the big-O notation, thus we obtain Lemma 2.

Lemma 2. *The computational complexity to obtain ϵ -solution to problem (18) is*

$$\delta (LK^3 + 6LK^2 + KL^2 + 6LK + 3K + 5 + m^2) m, \quad (24)$$

where $\delta = \ln(\epsilon^{-1})\sqrt{2LK+4}$.

Proof. First, problem (18) has K SOC constraints of dimension $(K+L+1)$, $(L-1)K+1$ SOC constraints of dimension $(K+2)$, and 1 SOC constraints of dimension $(K+1)$. Based on these observations, one can follow the same steps as in [20, Section V-A] to arrive at (24). Note that the term δ in (24) is the order of the number of iteration required to reach ϵ -solution to problem (18) while the remaining terms represent the per-iteration computation costs [20], [21]. \square

D. Numerical Results

Next, we study the convergence of the dual decomposition approach. We consider a wrapped-around cellular network with $L = 4, M = 500, \tau_c = 200, P_{\max,l} = 40$ W, $\hat{p}_{l,k} = 200$ mW, and $\xi_{l,k} = 0.5$ b/s/Hz, $\forall l, k$. The KL users are randomly distributed over the coverage area with uniform distribution. The distance of user k in cell l to BS l is denoted as $d_{l,k}^l$ km and $d_{l,k}^l \geq 0.035$ km. The system bandwidth and noise figure are 20 MHz and -96 dBm, respectively. The large-scale fading coefficient is modeled as

$$\beta_{l,k}^l [dB] = -148.1 - 37.6 \log_{10}(d_{l,k}^l) + z_{l,k}^l, \quad (25)$$

where the shadow fading $z_{l,k}^l$ follows a log-normal Gaussian distribution with standard variation 7 dB. The Lagrange multipliers are initially selected as $\lambda_{l,k}^{i,(1)} = 0, \forall l, i, k$, while the

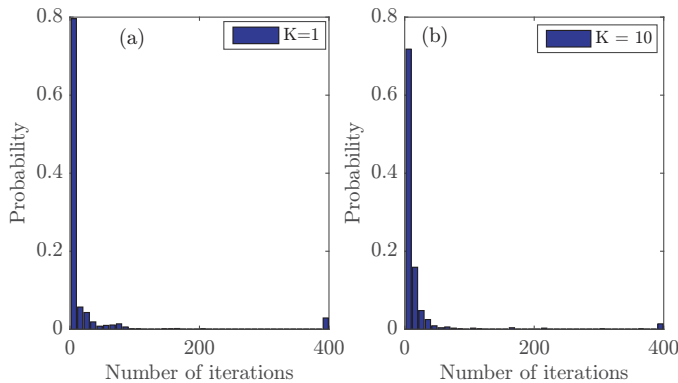


Fig. 3: The distribution of the number of iterations required for Algorithm 1 to obtain 95% of the global optimum: (a) $K = 1$; (b) $K = 10$.

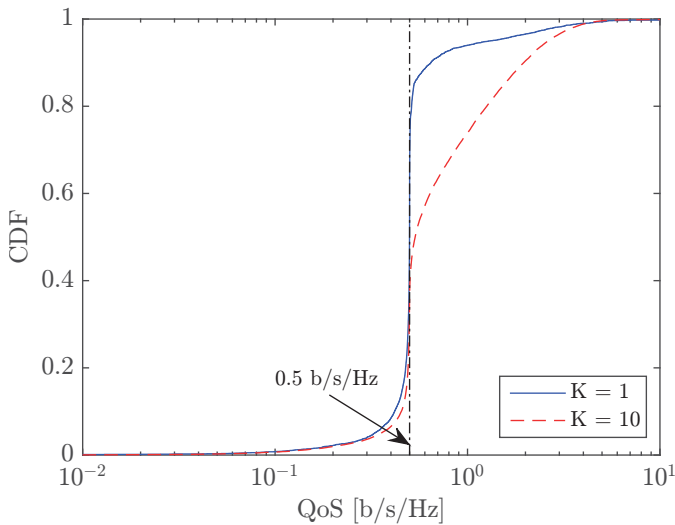


Fig. 4: Actual QoS [b/s/Hz] of each user at the 95% of the global optimum.

step size in (22) is defined as $\zeta^{(n)} = 0.01$. ZF precoding is used for simulations. Monte-Carlo results are obtained over 1000 different realizations of user locations.

Fig. 3 shows the probability of the number of iterations which Algorithm 1 needs to obtain the 95% of the global optimum for (a) $K = 1$ and (b) $K = 10$. Algorithm 1 converges to the desired objective value after a few iterations. For example, with $K = 10$, 12.4% of the realizations of user locations only need one iterations to attain 95% of the global optimum. If we set 400 iterations as a maximum level to terminate the algorithm for the realizations of users which are in trouble to obtain 95% of the global optimum, there are 1.4% of the user realizations facing with this issue. Meanwhile, by comparing Fig. 3a and Fig. 3b, we observe that Algorithm 1 requires more iterations to converge to the global optimum when increasing the number of users in the network.

Fig. 4 gives the cumulative distribution probability (CDF) of the actual QoS per user when the algorithm has reached 95%

TABLE I: The total number of optimization variables and exchanged variables for the considered approaches

Method	Execution Place	Total Number of Opt. variables	Total Number of Ex. Parameters
Centralized Imp.	Core Network	KL	$2KL^2 + 2KL$
Basic form of Distributed Imp.	Base Stations	KL^2	$2K(L-1)^2L + K(L-1)L$
Distributed Imp. by Dual Decomp.	Base Stations	$2KL^2 - KL + L$	$4K(L-1)^2N + 2K(L-1)L$

of the global optimum. The results confirm that the proposed distributed implementation satisfies the QoS requirements of most of the users, while some get somewhat lower QoS and some get higher. However, the performance gap between the proposed algorithm and its centralized counterpart is wider as the number of users increases, but mainly in the sense that users get higher QoS than required.

IV. COMPARISON OF CENTRALIZED AND DISTRIBUTED IMPLEMENTATION APPROACHES

Table I summarizes the total number of the optimization variables and the exchanged parameters for the centralized and the two distributed implementations. In this table, N represents the number of iterations used in the dual decomposition algorithm to reach convergence. Even though the distributed approach utilizing dual decomposition divides (11) into the L subproblems where every BS locally optimizes the power control coefficients, it also introduces new optimization variables leading to a large number of optimization variables, both in total and when comparing each subproblem with the centralized problem. Note that the global problem is a linear program, while the subproblems are SOC programs, which are more complex to solve even when the number of optimization variables are equal.

The total number of exchanged parameters with the dual decomposition approach may be less than in the basic form of distributed implementation, because N is usually small as shown in Section III-D. However, in situations where the number of iterations needed for the dual decomposition approach increases, the number of exchanged parameters may surpass that of the two other implementations.

Unlike in small-scale networks, where the small-scale fading coefficient from every antenna to every user needs to be exchanged to solve the centralized problem, in Massive MIMO only a few statistical parameters need to be exchanged per user. Hence, in Massive MIMO, the distributed implementation by dual decomposition does not bring any substantial reductions in backhaul signaling compared to the centralized counterpart; this is basically a consequence of the channel hardening property [1].

V. CONCLUSION

This paper compared distributed and centralized implementations of the total transmit power minimization problem with

QoS requirements in multi-cell Massive MIMO systems. In the distributed implementation based on dual decomposition, every BS can independently and simultaneously minimize its transmit powers, while controlling inter-cell interference based on the Lagrange multipliers $\{\lambda_{l,k}^i\}$ provided by a central entity, while a subset of nominal parameters representing the strength of mutual interference $\{\theta_{l,k}^i, \tilde{\theta}_{l,k}^i\}$ should be sent to the central entity to update these Lagrange multipliers. This iterative method converges to the global optimum after a few iterations (in the most realizations of the user locations) and its convergence is theoretically guaranteed. In small-scale MIMO networks, this approach can substantially reduce the computational complexity and backhaul signaling. However, in Massive MIMO systems, this distributed implementation does not bring any significant reductions in the amount of exchanged information, since each channel is only described by a few parameters. Therefore, for resource allocation problems that only involve large-scale fading coefficients, a centralized implementation is preferable in terms of both backhaul signaling and complexity. In practice, a combination of the two approaches might be preferable, for example, where the resource allocation is not optimized on every BS or on a single central entity, but at multiple central entities that are responsible for larger coverage areas.

REFERENCES

- [1] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
- [2] T. V. Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and user association optimization for Massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6384 – 6399, 2016.
- [3] H. Q. Ngo, A. E. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: Uniformly great service for everyone," in *Proc. IEEE SPAWC*, 2015.
- [4] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 2-3, pp. 113–381, 2013.
- [5] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, 2006.
- [6] A. Adhikary, A. Ashikhmin, and T. L. Marzetta, "Uplink interference reduction in large-scale antenna systems," *IEEE Trans. Wireless Commun.*, vol. 65, no. 5, pp. 2194–2206, 2017.
- [7] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, "Distributed energy-efficient UL power control in Massive MIMO with hardware impairments and imperfect CSI," in *Proc. IEEE ISWCS*, 2015, pp. 311–315.
- [8] H. Asgharimoghaddam, A. Tolli, and N. Rajatheva, "Decentralizing the optimal multi-cell beamforming via large system analysis," in *Proc. IEEE ICC*, 2014, pp. 5125–5130.
- [9] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, 1995.
- [10] T. A. Le and M. R. Nakhai, "Downlink optimization with interference pricing and statistical CSI," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2339–2349, 2013.
- [11] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.
- [12] T. V. Chien, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell Massive MIMO systems," *IEEE Trans. Wireless Commun.*, 2018, accepted for publication.
- [13] T. V. Chien and E. Björnson, *Massive MIMO Communications*. Springer International Publishing, 2017, pp. 77–116.
- [14] CVX Research Inc., "CVX: Matlab software for disciplined convex programming, academic users," <http://cvxr.com/cvx>, 2015.
- [15] S. Boyd, L. Xiao, A. Mutapic, and J. Mattingley, "Primal and dual decomposition: Notes on decomposition methods." [Online]. Available: <https://stanford.edu/class/ee364b/lectures.html>
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [17] H. Pennanen, A. Tölli, and M. Latva-aho, "Decentralized coordinated downlink beamforming via primal decomposition," *IEEE Signal Process. Lett.*, vol. 18, no. 11, pp. 647–650, 2011.
- [18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [19] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*. Society for Industrial and Applied Mathematics, 2001.
- [20] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690–5705, Nov. 2014.
- [21] T. A. Le, Q.-T. Vien, H. X. Nguyen, D. W. K. Ng, and R. Schober, "Robust chance-constrained optimization for power-efficient and secure SWIPT systems," *IEEE Trans. Green Comm. and Networking*, vol. 1, no. 3, pp. 333–346, 2017.