Resource Allocation for Cloud Radio Access Networks

Thesis by

Oussama Dhifallah

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science
in Electrical Engineering

King Abdullah University of Science and Technology, Thuwal,

Kingdom of Saudi Arabia

April 2016

The thesis of Oussama Dhifallah is approved by the examination committee

Committee Chairperson: Tareq Y. Al-Naffouri

Committee Member: Mohamed-Slim Alouini

Committee Member: Bernard Ghanem

Committee Member: Basem Shihada

Committee Member: Hayssam Dahrouj

ABSTRACT

Resource Allocation for Cloud Radio Access Networks

Oussama Dhifallah

Cloud-radio access network (CRAN) is expected to be the core network architecture for next generation mobile radio system. In CRANs, joint signal processing is performed at multiple cloud computing centers (clouds) that are connected to several base stations (BSs) via high capacity backhaul links. As a result, large-scale interference management and network power consumption reduction can be effectively achieved.

Unlike recent works on CRANs which consider a single cloud processing and treat inter-cloud interference as background noise, the first part of this thesis focuses on the more practical scenario of the downlink of a multi-cloud radio access network where BSs are connected to each cloud through wireline backhaul links. Assume that each cloud serves a set of pre-known single-antenna mobile users (MUs). This part focuses on minimizing the network total power consumption subject to practical constraints. The problems are solved using sophisticated techniques from optimization theory (e.g. Dual Decomposition-based algorithm and the alternating direction method of multipliers (ADMM)-based algorithm). One highlight of this part is that the proposed solutions can be implemented in a distributed fashion by allowing a reasonable information exchange between the coupled clouds. Additionally, feasible solutions of the considered optimization problems can be estimated locally at each

iteration. Simulation results show that the proposed distributed algorithms converge to the centralized algorithms in a reasonable number of iterations.

To further account of the backhaul congestion due to densification in CRANs, the second part of this thesis considers the downlink of a cache-enabled CRAN where each BS is equipped with a local-cache with limited size used to store the popular files without the need for backhauling. Further, each cache-enabled BS is connected to the cloud via limited capacity backhaul link and can serve a set of pre-known single-antenna MUs. This part assumes that only imperfect channel state information (CSI) is available at the cloud. This part focuses on jointly minimizing the network total power consumption as well as backhaul cost. It then suggests solving this optimization problem using the majorization-minimization (MM) approach. Simulation results show that the proposed algorithm converges in a reasonable number of iterations.

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisor Prof. Hayssam Dahrouj for his continuous guidance, support and advice throughout the course of this work. Without his expertise, the goals of this thesis would have been much more difficult to achieve. His efforts in giving me constant feedback and encouragement are greatly appreciated. Under his supervision, I successfully overcome many research challenges.

Foremost, I would like to thank Prof. Tareq Y. Al-Naffouri and Prof. Mohamed-Slim Alouini for their continuous support and advice. They always showed interest in my research work and never stopped providing me with constant advice not only about my research work, but also about my future career.

I want to express my heartfelt appreciation to my beloved family and friends for their continuous encouragement, their deep moral support and the great interest they express in my work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Motivation

1.1.1 Cloud Radio Access Networks

Cloud-radio access network (CRAN) technology is expected to support the tremendous requirements in mobile data traffic for next generation mobile radio systems (5G) [1][2][3]. In CRANs, base-stations (BSs) from different tiers are connected to the central processor (cloud) via high capacity backhaul links. The cloud then handles all the processing of the baseband signals. Such centralized processing provides a powerful tool to jointly manage the interference, increase network capacity and improve energy efficiency.

In conventional CRAN, there exist three essential components. The first component in CRAN is the central processor (cloud) which is formed by high performance processors where all the baseband signal processing is performed. The second component is the backhaul links which provide connection between the cloud and the BSs and are responsible for transmitting the user signals from the cloud to the BSs. Third part in CRAN is the BS which is responsible for transmitting the received signals from the cloud to the corresponding scheduled users. Since all the scheduling and signal processing is performed in the cloud, the BSs have much less power consumption and complexity which significantly decrease their price of deployment and make the increase of their density possible. Figure 1.1 illustrates an example of the CRAN architecture where the cloud is connected to 4 BSs through high capacity backhaul links and each BS is serving 7 MUs.

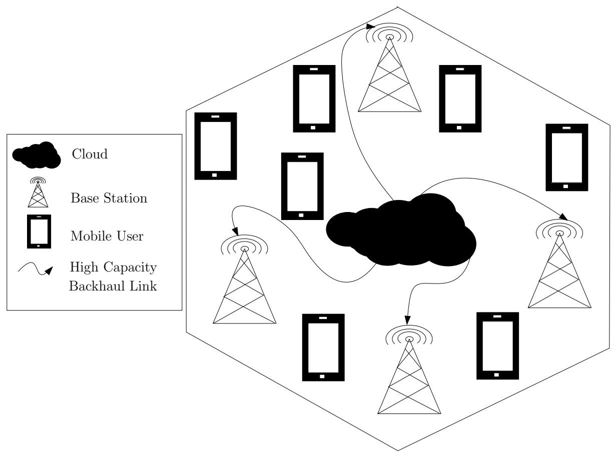

Figure 1.1: CRAN architecture.

Recent works on CRANs consider a single cloud processing and treat inter-cloud interference as background noise. This assumption can be unrealistic in practical scenarios due to the fact that the inter-cloud interference can not be estimated since it depends on the resource allocation strategy adopted by the adjacent clouds. Moreover, connecting many BSs to the same cloud and considering large number of users served by the BSs increase the computation burden on the cloud. The majority of resource allocation optimization problems are computationally expensive especially for large-sized networks. Therefore, distributing the computation across multiple clouds in the network is crucial. Further, centralized resource allocation optimization problems for multi-CRAN lead to joint signal-processing and a high level of cloud-to-cloud backhaul communication. To address this problem, distributed convex optimization approaches are needed so as to solve these resource allocation problems in a distributed fashion across the clouds.

However in dense data networks, the backhaul congestion may become a limiting factor of the CRAN performance. To truly access the advantages harvested by CRAN, the BSs are assumed to be equipped with local-cache with limited size where the popular content requested by the users across the network can be stored. As a result, the BSs directly transmit the data to the corresponding user when the requested data is available at its local-cache. Since many users across the network are likely to request the same content, caching the popular contents at the local memory of the BSs during the off-peak time can significantly reduce the load of the backhaul links.

1.1.2 Research Contribution

The overall setup considered in this thesis is the downlink scenario where the BSs are connected to the cloud through backhaul links and they can serve a set of pre-known single-antenna mobile users. Two scenarios are considered in this thesis :

The chapters 3 and 4 consider the more practical scenario of a CRAN formed by multiple clouds, where each cloud is connected to a cluster of BSs. The chapters focus on minimizing the network total power consumption subject to practical constraints. The main contribution of the chapters is an efficient iterative algorithm (i.e. dual decomposition-based algorithm and the alternating direction method of multipliers (ADMM)-based algorithm) that can be implemented in a distributed fashion across the multiple clouds through a reasonable exchange of information between the coupled clouds. One highlight of the chapters is that the proposed distributed approaches guarantee a feasible solution to the original optimization problem at each iteration.

Unlike the setup considered in the previous two chapters, chapter 5 considers a cache-enabled CRAN where the BSs are equipped with local-cache with limited size. The chapter formulates the problem of minimizing the network total power cost and the backhaul cost subject to per-BS power constraints, the quality of service constraints, the total backhaul capacity constraint and the imperfect CSI constraints in order to design the beamforming vector of each user across the network, the quantization noise covariance matrix of the limited-capacity backhaul links and the BS clustering. The main contribution of the chapter is an efficient and stable algorithm based on the MM approach.

1.2 Thesis Outline

The rest of this thesis is organized as follows. Chapter 2 provides a brief review of convex and non-convex optimization theory. Chapter 3 considers a multi-cloud scenario and focuses on the problem of minimizing the total network power consumption subject to quality-of-service requirements. The main contribution of this chapter is an algorithm based on the dual decomposition approach that can be implemented in a decentralized fashion across the network, with a reasonable amount of information exchange between the clouds. Chapter 4 considers a multi-CRAN where imperfect CSI is available at each cloud and limited capacity backhaul links connecting the BSs to the cloud. The chapter proposes an ADMM-based distributed algorithm to solve the network total power minimization problem. Then, chapter 5 solves the problem of minimizing the network total power cost as well as the backhaul cost in a cache-enabled CRAN using the majorization-minimization (MM) approach. Finally, chapter 6 concludes this thesis.

1.3 Notation

In this thesis, we use the following notations. Lower case letters x, lower case bold letters $\mathbf{x}$ and upper case bold letters $\mathbf{X}$ are used to respectively denote scalars, vectors and matrices. The functions $\mathrm{Tr}(.)$, $||.||_2$ and $||.||_0$ denote respectively the trace of a matrix, the ℓ_2 norm of a vector and the ℓ_0 norm of a vector. The operators $(.)^T$, $(.)^*$ and $(.)^H$ denote respectively the transpose operator, the conjugate operator, the conjugate transpose operator.

Chapter 2

Mathematical Optimization

The resource allocation problems considered in this thesis rely on tools and concepts introduced in optimization theory. This chapter provides a brief review of some convex optimization problems, i.e second-order cone programming (SOCP), semi-definite programming (SDP). Further, this chapter gives an overview of techniques used to construct a distributed algorithm that converges to the global optimal solution in convex optimization problems, i.e. dual decomposition and alternating direction method of multipliers (ADMM). Finally, the chapter presents some techniques used to solve difficult non-convex optimization problems, i.e. the majorization-minimization (MM) algorithm, the SDP rank-1 relaxation and the tightest convex lower bound approximation of non-convex cost functions.

2.1 Introduction

Considerable progress in the field of optimization theory has been achieved, over the past two decades. Optimization is considered as a powerful communications and signal-processing tool and it is used in wide range of applications. Optimization problems considered in communications and signal-processing can be formulated in the following standard form

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ s.t. \quad & g_i(\mathbf{x}) \leq 0, \ \forall \ i \in \{1, 2, .., n\} \\ & h_j(\mathbf{x}) = 0, \ \forall \ j \in \{1, 2, .., m\}, \end{aligned} \tag{2.1}$$

where $\mathbf{x}$ denotes the optimization variable, $\mathbf{x}$ can be real or complex scalar, vector or matrix. f is called the objective function or the cost function. $g_i(\mathbf{x}) \leq 0, \ \forall \ i \in \{1, 2, .., n\}$, are called the inequality constraints and $h_j(\mathbf{x}) = 0, \ \forall \ j \in \{1, 2, .., m\}$, are called the equality constraints.

The optimization problem (2.1) consists of finding the smallest objective value while satisfying the inequality and equality constraints. The feasibility set of the optimization problem is defined as follows

$$\mathcal{F} = \Big\{ \mathbf{x} \in \mathcal{D} \mid g_i(\mathbf{x}) \leq 0, \ \forall \ i \in \{1, 2, .., n\}, h_j(\mathbf{x}) = 0, \ \forall \ j \in \{1, 2, .., m\} \Big\}, \tag{2.2}$$

where $\mathcal{D}$ denotes the domain of the constraint functions g_i and h_j. Specifically, $\mathcal{D}$ is given by

$$\mathcal{D} = \Big\{ \cap_{i=1}^{n} \operatorname{dom}(g_i) \Big\} \cap \Big\{ \cap_{j=1}^{m} \operatorname{dom}(h_j) \Big\}, \tag{2.3}$$

where $\operatorname{dom}(g_i)$ denotes the domain of the function g_i. $\mathbf{x}^*$ is called the optimal solution if it provides the smallest objective value and it satisfies all the equality and inequality constraints, i.e it belongs to the feasibility set $\mathcal{F}$. specifically, $\mathbf{x}^*$ satisfies

$$f(\mathbf{x}^*) = \Big\{ \min_{\mathbf{x}} \quad f(\mathbf{x}), \mathbf{x} \in \mathcal{F} \Big\}. \tag{2.4}$$

2.2 Convex Optimization

Convex optimization is considered as a powerful tool for resource allocation in cloud radio access networks. Convex optimization problems are considered as a special class of mathematical optimization. An optimization problem in the standard form (2.1) is considered as a convex optimization problem when the cost function f is convex and the feasibility set is convex, i.e. the functions $g_i \; \forall \; i \in \{1, 2, .., n\}$ are convex and the functions $h_j, \; \forall \; j \in \{1, 2, .., m\}$ are affine functions. Therefore, convex optimization problems consist of minimizing a convex function while satisfying a set of convex constraints.

Recasting an optimization problem as a convex optimization problem provides interesting advantages. The most important advantage is that convex optimization problem has a single stationary point. Therefore, the global optimal solution can be determined analytically or using reliable and efficient numerical methods (e.g. the interior point method [4]). Another interesting advantage of recasting a problem as a convex optimization problem is the formulation of the associate dual problem. The dual problem may be easier to solve and may offer good directions for distributed solutions.

2.2.1 Second-Order Cone Programming

The second-order cone programming (SOCP) is a convex optimization problem. It consists of minimizing a linear function subject to linear equality constraints and second order cone constraints. The standard form of the SOCP problem can be written as follows

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T\mathbf{x} \\ s.t. \quad & ||\mathbf{A}_i\mathbf{x} + \mathbf{b}_i||_2 \leq \mathbf{u}_i^T\mathbf{x} + v_i, \ \forall \ i \in \{1, 2, .., n\} \\ & \mathbf{D}\mathbf{x} = \mathbf{d}, \end{aligned} \tag{2.5}$$

where the optimization variable is $\mathbf{x} \in \mathbb{R}^m$, and where the optimization parameters are $\mathbf{c} \in \mathbb{R}^m$, $\mathbf{A}_i \in \mathbb{R}^{m_i \times m}$, $\mathbf{b}_i \in \mathbb{R}^{m_i}$, $\mathbf{u}_i \in \mathbb{R}^m$, $v_i \in \mathbb{R}$, and where $\mathbf{D} \in \mathbb{R}^{p \times m}$ and $\mathbf{d} \in \mathbb{R}^p$. The following constraints

$$||\mathbf{A}_i\mathbf{x} + \mathbf{b}_i||_2 \leq \mathbf{u}_i^T\mathbf{x} + v_i, \ \forall \ i \in \{1, 2, .., n\}, \tag{2.6}$$

are called second-order cone (SOC) constraints. The SOC constraints are convex and can represent many classical convex constraints. Specifically, the SOC constraints are equivalent to linear program (LP) constraints when $\mathbf{A}_i = \mathbf{0}$, $\forall \ i \in \{1, 2, .., n\}$ and to quadratic constraints (QC) when $\mathbf{u}_i = \mathbf{0}$, $\forall \ i \in \{1, 2, .., n\}$.

The global optimal solution of the SOCP problem can be found efficiently using reliable numerical algorithms such as the interior point method [4]. The SOCP problem (2.5) has $c = (n + 2p)$ SOC constraints and $v = m$ SOC variables. Therefore, the computational complexity of solving the SOCP problem (2.5) is $\mathcal{O}\left((n + 2p)m^3\right)$ using the interior point method [4]. When n, p and m are large, the computational complexity of solving the SOCP using the interior point method can be huge. To reduce complexity, the SCS solver [5] which is based on the ADMM technique can be used to solve the large-scale SOCP problem (2.5).

2.2.2 Semi-Definite Programming

The semi-definite programming (SDP) is a convex optimization problem. It consists of minimizing a linear function subject to linear equality constraints and positive semi-definite matrix constraints. The standard form of the SDP problem can be formulated as follows

$$\begin{aligned} \min_{\mathbf{X}} \quad & \text{Tr}(\mathbf{C}\mathbf{X}) \\ s.t. \quad & \text{Tr}(\mathbf{A}_i\mathbf{X}) \leq b_i, \ \forall \ i \in \{1, 2, .., n\} \\ & \mathbf{X} \succeq 0, \end{aligned} \tag{2.7}$$

where the optimization variable is $\mathbf{X} \in \mathbb{R}^{m \times m}$, and where the optimization parameters are $\mathbf{C} \in \mathbb{R}^{m \times m}$, $\mathbf{A}_i \in \mathbb{R}^{m \times m}$, $b_i \in \mathbb{R}$. Many optimization problems can be cast as an SDP problem. Specifically, the SOCP problem, quadratically constrained quadratic programming (QCQP) problem and LP problem can be reformulated as an SDP problem.

The global optimal solution of the SDP problem can be found efficiently using reliable numerical algorithms [4]. The SDP problem (2.7) has $c = (n + 1)$ SDP constraints, hence, the worst-case complexity of solving the SDP problem (2.7) is $\mathcal{O}\left((n+1)^4\right)$ using the interior point method. When n is large, the computational complexity of solving the SDP problem using the interior point method can be huge. To reduce complexity, the ADMM technique can be applied to solve the large-scale SDP problem (2.7) [5].

2.3 Distributed Convex Optimization

For large distributed systems, and as it is shown in the next chapters, solving the optimization problems in a centralized fashion using the standard numerical methods may not be physically feasible. Thus, it becomes of central importance to develop distributed algorithms in order to design physically feasible systems as well as reduce the computational complexity of the centralized solutions, particularly in large-scale systems. This section presents an overview of the dual decomposition and the alternating direction method of multipliers which are algorithms that lead to distributed implementation in practical scenario.

2.3.1 Dual Decomposition

Dual decomposition approach [6] can be used when the centralized optimization problem can be implemented in a distributed fashion if a coupling optimization constraint is relaxed. Consider a network formed by p nodes and consider the optimization problem of minimizing a total cost subject to feasibility constraints. Specifically, we consider the optimization problem in the following form

$$\begin{aligned} \min_{\mathbf{x}_i} \quad & \sum_{i=1}^{p} f(\mathbf{x}_i) \\ s.t. \quad & \sum_{j=1}^{p} \mathbf{g}_j(\mathbf{x}_j) \leq \mathbf{c} \\ & \mathbf{x}_i \in \mathcal{X}_i, \ \forall \, i \in \{1, 2, .., p\}, \end{aligned} \tag{2.8}$$

where the optimization variables are $\mathbf{x}_i \in \mathbb{R}^m$, and where the optimization parameters are $\mathbf{c} \in \mathbb{R}^n$, $\mathcal{X}_i$ denotes the feasibility set of the optimization variable $\mathbf{x}_i$ where $\mathcal{X}_i$ is assumed to be convex and the functions $f : \mathbb{R}^m \to \mathbb{R}$ and $\mathbf{g}_j : \mathbb{R}^m \to \mathbb{R}^n$ are convex functions. Clearly, the optimization problem (2.8) can be decoupled into multiple and independent sub-problems when the coupling constraint $\sum_{j=1}^{p} \mathbf{g}_j(\mathbf{x}_j) \leq \mathbf{c}$ is omitted. The first step of the dual decomposition approach consists of formulating the partial Lagrangian over the coupling constraints as follows

$$\mathcal{L}(\mathbf{x}_i, \boldsymbol{\lambda}) = \sum_{i=1}^{p} f(\mathbf{x}_i) + \boldsymbol{\lambda}^T \left(\sum_{j=1}^{p} \mathbf{g}_j(\mathbf{x}_j) - \mathbf{c} \right), \tag{2.9}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^n$ denotes the dual vector associated with the coupling constraint. The dual decomposition approach consists of solving the optimization problem (2.8) iteratively where the iterative algorithm requires to iterate between two levels. At the first level, the partial Lagrangian (2.9) is minimized over the primal variables with fixed dual vector

$$\begin{aligned} \min_{\mathbf{x}_i} \quad & \sum_{i=1}^{p} f(\mathbf{x}_i) + \boldsymbol{\lambda}^T \left(\sum_{j=1}^{p} \mathbf{g}_i(\mathbf{x}_i) - \mathbf{c} \right) \\ s.t. \quad & \mathbf{x}_i \in \mathcal{X}_i, \ \forall \ i \in \{1, 2, .., p\}. \end{aligned} \tag{2.10}$$

Clearly, the above optimization problem is separable and it can be solved locally and independently at each node i. Specifically, the following sub-problem is solved locally at each node

$$\begin{aligned} \min_{\mathbf{x}_i} \quad & f(\mathbf{x}_i) + \boldsymbol{\lambda}^T \mathbf{g}_i(\mathbf{x}_i) \\ s.t. \quad & \mathbf{x}_i \in \mathcal{X}_i. \end{aligned} \tag{2.11}$$

At the second level, the following master problem responsible of updating the dual vector $\boldsymbol{\lambda}$

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \quad & \sum_{i=1}^{p} f_i^*(\boldsymbol{\lambda}) - \boldsymbol{\lambda}^T \mathbf{c} \\ s.t. \quad & \boldsymbol{\lambda} \geq 0, \end{aligned} \tag{2.12}$$

is solved, where $f_i^*(\boldsymbol{\lambda})$ denotes the optimal objective value of the sub-problem (2.11) corresponding to the node i. The master problem (2.12) can be solved using the sub-gradient method. This technique needs the knowledge of a sub-gradient associated with each optimal objective value $f_i^*(\boldsymbol{\lambda}), \ \forall \ i \in \{1, 2, .., p\}$. It can be shown that the sub-gradient update of the dual vector $\boldsymbol{\lambda}$ can be performed locally at each node by allowing a limited information exchange between the nodes [7][8][9].

2.3.2 The Alternating Direction Method of Multipliers

The alternating direction method of multipliers (ADMM) [10] solves optimization problems in the following standard form

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x}) + g(\mathbf{y}) \\ s.t. \quad & \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = \mathbf{c}, \end{aligned} \tag{2.13}$$

where the optimization variables are $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$, and where the optimization parameters are $\mathbf{A} \in \mathbb{R}^{p\times m}$, $\mathbf{B} \in \mathbb{R}^{p\times m}$, $\mathbf{c} \in \mathbb{R}^p$ and the functions $f : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are convex functions. The first step of the ADMM approach consists of formulating the augmented Lagrangian of the optimization problem (2.13) as follows

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}) = f(\mathbf{x}) + g(\mathbf{y}) + \boldsymbol{\lambda}^T (\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{c}) + \frac{\rho}{2}||\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{c}||_2^2, \tag{2.14}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^p$ denotes the dual variable associated with the equality constraint in (2.13) and $\rho > 0$ denotes the penalty parameter. The ADMM approach consists of first minimizing the augmented Lagrangian (2.14) over the primal vector $\mathbf{x}$ while all the other variables are fixed at their current values. Then, the second step of the ADMM algorithm minimizes the augmented Lagrangian (2.14) over the primal vector $\mathbf{y}$ while the remaining variables are fixed. The final step of the ADMM algorithm consists of a simple dual update. Specifically, the ADMM approach consists of the following three updates at the k^{th} iteration

$$\begin{aligned} \mathbf{x}^{k+1} &= \min_{\mathbf{x}} \ \mathcal{L}(\mathbf{x}, \mathbf{y}^k, \boldsymbol{\lambda}^k) \\ \mathbf{y}^{k+1} &= \min_{\mathbf{y}} \ \mathcal{L}(\mathbf{x}^{k+1}, \mathbf{y}, \boldsymbol{\lambda}^k) \\ \boldsymbol{\lambda}^{k+1} &= \boldsymbol{\lambda}^k + \rho \left(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{c}\right). \end{aligned} \tag{2.15}$$

For convex optimization, the ADMM approach converges to the global optimal solution when the following conditions are satisfied [10]

$$\begin{aligned} &\mathbf{A}\mathbf{x}^* + \mathbf{B}\mathbf{y}^* - \mathbf{c} = 0 \\ &0 \in \partial f(\mathbf{x}^*) + \mathbf{A}^T \boldsymbol{\lambda}^*, \end{aligned} \tag{2.16}$$

where $\mathbf{x}^*$, $\mathbf{y}^*$ and $\boldsymbol{\lambda}^*$ denote the optimal primal and dual solution of the optimization problem (2.13), respectively. Since $\mathbf{x}^{k+1}$ is the global optimal solution of the first step of the ADMM algorithm, we can write the following

$$0 \in \partial f(\mathbf{x}^{k+1}) + \mathbf{A}^T \boldsymbol{\lambda}^k + \rho \mathbf{A}^T \left(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^k - \mathbf{c} \right). \tag{2.17}$$

This can be rewritten as follows

$$\rho \mathbf{A}^T \mathbf{B} \left(\mathbf{y}^{k+1} - \mathbf{y}^k \right) \in \partial f(\mathbf{x}^{k+1}) + \mathbf{A}^T \boldsymbol{\lambda}^{k+1}. \tag{2.18}$$

This suggests that a reasonable stopping criterion is that $\mathbf{r}^{k+1} = \mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{c}$ and $\mathbf{s}^{k+1} = \rho \mathbf{A}^T \mathbf{B} \left(\mathbf{y}^{k+1} - \mathbf{y}^k \right)$ must be small where $\mathbf{r}^{k+1}$ and $\mathbf{s}^{k+1}$ are called the primal and dual residuals, respectively. Specifically, the primal and dual residuals should satisfy

$$\left|\left|\mathbf{r}^{k+1}\right|\right|_2 \leq \epsilon_p \text{ and } \left|\left|\mathbf{s}^{k+1}\right|\right|_2 \leq \epsilon_d, \tag{2.19}$$

where ϵ_p and ϵ_d are small positive constants.

The computational complexity of the ADMM algorithm mainly comes from the first two steps at each iteration since the third step is a dual update. The second step of the optimization problems studied in this thesis are convex and an analytical solution can be easily found by setting the gradient of the augmented Lagrangian (2.14) with respect to the primal variable $\mathbf{y}$ to zero. Therefore, the computational complexity of the ADMM approach for the optimization problem studied in the thesis comes mainly from the first step.

2.4 Non-Convex Optimization

Convex optimization is receiving much attention since the global optimal solution can be found analytically or using reliable numerical methods. However, most optimization problems in communication and signal processing applications are non-convex. Non-convex optimization problems are difficult to solve and may have large number of local optima. Many approaches have been proposed to reduce the complexity of solving a non-convex optimization problems. These methods rely mainly on convex relaxations to turn a non-convex optimization problem to a convex optimization problem, then, use the classical methods to find the global optimal solution of the relaxed problem. Further, these methods can only guarantee a local optimal solution or a sub-optimal solution.

An optimization problem in the standard form (2.1) is considered as a non-convex optimization problem when the cost function f is not convex and/or the feasibility set is not convex, i.e. at least one function of the following set of functions $\{g_i \ \forall \ i \in \{1, 2, .., n\}\}$, and/or one function of the following set of function $\{h_j, \ \forall \ j \in \{1, 2, .., m\}\}$ is not linear. This section presents an overview of some convex relaxations largely used in communication and signal processing applications to reduce the complexity of solving non-convex optimization problems.

2.4.1 Majorization-Minimization Approach

The majorization-minimization (MM) procedure [11] consists of transforming a non-convex optimization problem into a tractable optimization problem by iterating between two steps, i.e. a majorization step and a minimization step. The MM algorithm treats optimization problems in the following standard form

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ s.t. \quad & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{2.20}$$

where the optimization variable is $\mathbf{x}$ and the optimization parameters are $\mathcal{X}$ denotes the feasibility set of $\mathbf{x}$ and $f : \mathbb{R}^m \to \mathbb{R}$. We assume that $\mathcal{X}$ is a convex set and f is a non-convex function. The first step of the MM procedure is called the majorization step which consists of finding a surrogate function $g : \mathbb{R}^m \to \mathbb{R}$ that majorizes the objective function in the optimization problem (2.20) at a point $\mathbf{x}^m$. Specifically, the surrogate function g satisfies the following proprieties

$$\begin{aligned} & g(\mathbf{x}|\mathbf{x}^m) \geq f(\mathbf{x}),\ \forall \mathbf{x} \in \mathcal{X} \\ & g(\mathbf{x}^m|\mathbf{x}^m) = f(\mathbf{x}^m). \end{aligned} \tag{2.21}$$

The second step of the MM procedure is called the minimization step which consists of minimizing the constructed surrogate function g over the variable $\mathbf{x}$

$$\begin{aligned} \min_{\mathbf{x}} \quad & g(\mathbf{x}|\mathbf{x}^m) \\ s.t. \quad & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{2.22}$$

where $\mathbf{x}^m$ denotes the current iterate. By iterating between the majorization step and minimization step, the cost function of the optimization problem (2.20) is driven downhill

$$\begin{aligned} f(\mathbf{x}^{m+1}) & \leq g(\mathbf{x}^{m+1}|\mathbf{x}^m) \\ & = \min_{\mathbf{x}} \quad g(\mathbf{x}|\mathbf{x}^m) \ \ s.t. \ \ \mathbf{x} \in \mathcal{X} \\ & \leq g(\mathbf{x}^m|\mathbf{x}^m) \\ & = f(\mathbf{x}^m), \end{aligned} \tag{2.23}$$

where $\mathbf{x}^{m+1}$ denotes the optimal solution of the optimization problem (2.22) and it is the next iterate of the MM procedure. The above property (2.23) follows using (2.21) and (2.22). Given the descent property (2.23), the MM procedure enjoys high numerical stability [11]. Further, the MM procedure converges to a stationary point of the original optimization problem (2.20) [11].

2.4.2 SDP Rank-1 Relaxation

In this subsection, we consider the power minimization problem in a single cloud radio access network scenario with imperfect channel state information (CSI). This optimization can be formulated as follows

$$\begin{aligned}
\min_{\mathbf{W}_u,\lambda_u} \; & \sum_{u\in\mathcal{U}} \mathrm{Tr}\left(\mathbf{W}_u\right) \\
s.t. \; & \sum_{u\in\mathcal{U}} \mathrm{Tr}\left(\mathbf{A}_{bu}\mathbf{W}_u\right) \leq P_b, \;\forall\; b \in \mathcal{B} \\
& \begin{bmatrix} \mathbf{G}_u + \lambda_u \mathbf{E}_u & \mathbf{G}_u \tilde{\mathbf{h}}_u \\ \tilde{\mathbf{h}}_u^H \mathbf{G}_u & \tilde{\mathbf{h}}_u^H \mathbf{G}_u \tilde{\mathbf{h}}_u - \sigma_u^2 - \lambda_u \end{bmatrix} \succeq 0, \;\forall\; u \in \mathcal{U} \\
& \mathbf{G}_u = \frac{1}{\delta_u}\mathbf{W}_u - \sum_{u'\in\mathcal{U}_u} \mathbf{W}_{u'}, \;\lambda_u \geq 0, \;\forall\; u \in \mathcal{U} \\
& \mathbf{W}_u \succeq 0, \;\mathrm{Rank}(\mathbf{W}_u) = 1, \forall\; u \in \mathcal{U},
\end{aligned} \tag{2.24}$$

where the optimization is over the beamforming matrices $\mathbf{W}_u$ and the scalar variables λ_u, and where $\mathcal{B}$ denotes the number of BSs, $\mathcal{U}$ denotes the number of users, $\mathbf{E}_u$ denotes the accuracy of the CSI and $\mathbf{h}_u$ denotes the estimated channel vector of user u. The cost function and all the constraints of the optimization problem (2.24) are convex except the rank-one constraints of the beamforming matrices, i.e. $\mathrm{Rank}(\mathbf{W}_u) = 1$. Therefore, finding the global optimal solution of (2.24) is extremely difficult. To reduce the complexity of solving the above power minimization problem, one can simply drop the rank-one constraints and solve the resulting optimization problem. The rank-one relaxed optimization problem (2.24) is an SDP problem, hence, it can be solved efficiently using reliable numerical methods [4]. In general, the optimal solution of the rank-one relaxed optimization problem does not satisfy the rank-one constraints. In practice, additional approximation methods, such as the Gaussian Randomization [4], are needed to find a rank-one solution whenever needed.

2.4.3 Tightest Convex Lower Bound

One possible convex relaxation of intractable non-convex cost functions is its tightest positively homogeneous convex lower bound [12]. This particularly leads to formulations invariant by rescaling of the data since the convexly relaxed cost function is positively homogeneous. The first step to find the tightest positively homogeneous convex lower bound of the non-convex cost function is to compute its tightest positively homogeneous lower bound but not necessary convex.

Definition 1. *Given a function f, its tightest positively homogeneous lower bound but not necessary convex f_{lb} is given by*

$$f_{lb}(\mathbf{x}) = \inf_{\lambda > 0} \quad \frac{f(\lambda \mathbf{x})}{\lambda}. \tag{2.25}$$

Then, the tightest positively homogeneous convex lower bound can be determined using the following proposition.

Proposition 1. *[12] The tightest positively homogeneous convex lower bound of the real valued function $f : \mathbb{R}^m \to \mathbb{R}$ is well-defined and given by $f_c = f^{**}$, where f^{**} denotes the Fenchel-Legendre bidual of the function f*

This subsection constructs the tightest positively homogeneous convex lower bound of a particular intractable non-convex cost function similar to the cost function considered in chapter 3. Consider the optimization problem in the following form

$$\begin{aligned} \min_{\mathbf{x}} \quad & ||\mathbf{x}||_2^2 + \alpha \mathcal{I}(||\mathbf{x}||_0) \\ s.t. \quad & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{2.26}$$

where $\mathcal{I}(||\mathbf{x}||_0) = 1$ when $||\mathbf{x}||_0 \neq 0$ and $\mathcal{I}(||\mathbf{x}||_0) = 0$ otherwise, and where $\mathcal{X}$ is the feasibility set of the variable $\mathbf{x}$, assumed to be convex and α is a positive constant. Clearly, the cost function of the optimization problem (2.26) is not convex since the function $\mathcal{I}(||\mathbf{x}||_0)$ is not convex. Therefore, finding the global optimal solution of (2.26) is difficult. This subsection computes the tightest positively homogeneous convex lower bound to approximate the cost function. Then, it minimizes the constructed convex function to find a sub-optimal solution of the considered optimization problem.

First, we construct the tightest positively homogeneous lower bound but not necessary convex of the cost function of the optimization problem (2.26), given by

$$C_{lb}(\mathbf{x}) = \inf_{\lambda > 0} \left\{ \lambda ||\mathbf{x}||_2^2 + \frac{\alpha}{\lambda} \mathcal{I}(||\mathbf{x}||_0) \right\}, \tag{2.27}$$

where the optimization variable is λ. Clearly, the above optimization problem is convex. By setting the gradient of the cost function with respect to λ to zero, the optimal solution λ^* is given by

$$\lambda^* = \alpha^{1/2} \mathcal{I}(||\mathbf{x}||_0)^{1/2} ||\mathbf{x}||_2^{-1}. \tag{2.28}$$

Therefore, the tightest positively homogeneous lower bound but not necessary convex of the cost function of the optimization problem (2.26) can be written as follows

$$C_{lb}(\mathbf{x}) = 2||\mathbf{x}||_2 \sqrt{\alpha \mathcal{I}(||\mathbf{x}||_0)}. \tag{2.29}$$

In order to compute the tightest positively homogeneous convex lower bound of the cost function of the optimization problem (2.26), we start by determining the Fenchel conjugate of (2.29), given by

$$\begin{aligned} C_{lb}^*(\mathbf{y}) &= \max_{\mathbf{x}} \left\{ \mathbf{x}^T \mathbf{y} - 2||\mathbf{x}||_2 \sqrt{\alpha \mathcal{I}(||\mathbf{x}||_0)} \right\} \\ &= \max_{\mathcal{A} \subseteq \mathcal{V}} \max_{\mathbf{x}_\mathcal{A} \in \mathbb{R}^A} \left\{ \mathbf{x}_\mathcal{A}^T \mathbf{y}_\mathcal{A} - 2||\mathbf{x}_\mathcal{A}||_2 \sqrt{\alpha F(\mathcal{A})} \right\} \\ &= \max_{\mathcal{A} \subseteq \mathcal{V}} \mathcal{I} \left\{ ||\mathbf{y}_\mathcal{A}||_2 \leq 2 \sqrt{\alpha F(\mathcal{A})} \right\} \\ &= \begin{cases} 0 & \text{if } \Omega(\mathbf{y}) \leq 1 \\ \infty & \text{otherwise} \end{cases}, \end{aligned} \tag{2.30}$$

where $F(\mathcal{A}) = 1$ when $\mathcal{A} \neq \varnothing$ and $F(\mathcal{A}) = 0$ otherwise, and where $\Omega(\mathbf{y}) =$

$\max_{\mathcal{A}\subseteq\mathcal{V},\mathcal{A}\neq\varnothing} \frac{||\mathbf{y}_{\mathcal{A}}||_2}{2\sqrt{\alpha F(\mathcal{A})}}$. Therefore, the tightest positively homogeneous convex lower bound of (2.29) is given by

$$\begin{aligned} C_{lb}^{**}(\mathbf{x}) &= \max_{\mathbf{y}} \ \mathbf{y}^T\mathbf{x} \\ s.t. \quad & \Omega(\mathbf{y}) \leq 1. \end{aligned} \tag{2.31}$$

First, the tightest positively homogeneous convex lower bound of (2.29) can be upper bounded as follows

$$\begin{aligned} C_{lb}^{**}(\mathbf{x}) &\leq \max_{\mathbf{y}} \ ||\mathbf{y}||_2||\mathbf{x}||_2 \\ & s.t. \quad \Omega(\mathbf{y}) \leq 1 \\ &= \max_{\mathbf{y}} \ 2\sqrt{\alpha}||\mathbf{x}||_2\frac{||\mathbf{y}||_2}{2\sqrt{\alpha}} \\ & s.t. \quad \max_{\mathcal{A}\subseteq\mathcal{V},\mathcal{A}\neq\varnothing} \frac{||\mathbf{y}_{\mathcal{A}}||_2}{2\sqrt{\alpha F(\mathcal{A})}} \leq 1 \\ &= 2\sqrt{\alpha}||\mathbf{x}||_2. \end{aligned} \tag{2.32}$$

Further, the tightest positively homogeneous convex lower bound of (2.29) can be lower bounded by setting $\tilde{\mathbf{y}} = 2\sqrt{\alpha}\frac{\mathbf{x}}{||\mathbf{x}||_2}$

$$\begin{aligned} C_{lb}^{**}(\mathbf{x}) &= \max_{\mathbf{y}} \ ||\mathbf{x}||_2||\mathbf{y}||_2 \\ & s.t. \quad \Omega(\mathbf{y}) \leq 1 \\ &\geq ||\mathbf{x}||_2 2\sqrt{\alpha}\frac{||\mathbf{x}||_2}{||\mathbf{x}||_2} \\ &= 2\sqrt{\alpha}||\mathbf{x}||_2. \end{aligned} \tag{2.33}$$

Therefore, the tightest positively homogeneous convex lower bound of (2.29) is given by

$$C_{lb}^{**}(\mathbf{x}) = 2\sqrt{\alpha}||\mathbf{x}||_2. \tag{2.34}$$

The following convex optimization problem is solved in order to construct a sub-optimal solution of the original optimization problem (2.26)

$$\begin{aligned} \min_{\mathbf{x}} \quad & C_{lb}^{**}(\mathbf{x}) = 2\sqrt{\alpha}||\mathbf{x}||_2 \\ s.t. \quad & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{2.35}$$

In order to further improve the performance, the solution of the relaxed optimization problem can be used as a starting point of another iterative algorithm. As a result, the global optimal solution of the original optimization problem can be achieved since the starting point is sub-optimal (i.e close to the global optimal solution).

2.5 Conclusion

This chapter presents an overview of some convex optimization problems and some techniques used to tackle non-convex optimization problems. The concepts introduced in this chapter will be used in the upcoming chapters to provide distributed algorithm for convex resource allocation optimization problems and to tackle non-convex optimization problems arising in resource allocation problems.

Chapter 3

Decentralized Group Sparse Beamforming for Multi-CRAN

Recent studies on CRANs assume the availability of a single processor (cloud) capable of managing the entire network performance; inter-cloud interference is treated as background noise. This chapter[1] considers the more practical scenario of the downlink of a CRAN formed by multiple clouds, where each cloud is connected to a cluster of multiple-antenna BSs via high-capacity wireline backhaul links. The network is composed of several disjoint BSs' clusters, each serving a pre-known set of single-antenna users. To account for both inter-cloud and intra-cloud interference, the chapter considers the problem of minimizing the network total power consumption subject to quality of service constraints, by jointly determining the set of active BSs connected to each cloud and the beamforming vectors of every user across the network. The chapter solves the problem using Lagrangian duality theory through a dual decomposition approach, which decouples the problem into multiple and independent sub-problems, the solution of which depends on the dual optimization problem. The solution then proceeds in updating the dual variables and the active set of BSs at each cloud iteratively. The proposed approach leads to a distributed

[1]A part of this chapter appears in: ©[2015]IEEE O. Dhifallah, H. Dahrouj, T.Y. Al-Naffouri; M.-S. Alouini, "Decentralized Group Sparse Beamforming for Multi-Cloud Radio Access Networks", *in Proc. of IEEE Globecom* , San Diego, USA, Dec 2015.

implementation across the multiple clouds through a reasonable exchange of information between adjacent clouds. The chapter further proposes a centralized solution to the problem. Simulation results suggest that the proposed algorithms significantly outperform the conventional per-cloud update solution, especially at high signal-to-noise-plus-interference-ratio (SINR) target.

3.1 Introduction

Conventional strategies in CRANs assume the existence of a single cloud, and simply treat inter-cloud interference as part of the background noise. This chapter considers the more realistic multi-cloud scenario. A major point in this chapter is that a significant performance gain can be reached by accounting for both inter-cloud and intra-cloud interference, and jointly optimizing the beamforming vectors and the set of active BSs across the network.

The network is formed by a disjoint set of clusters of multi-antenna BSs, and each cloud is connected to one cluster via high capacity wireline backhaul links. Each cluster of BSs communicates with a pre-known set of single-antenna MUs. Unlike the single-cloud model, the performance of the considered system becomes a function of both inter-cloud and intra-cloud interference, and depends on the set of active BSs within each cluster, as well as the beamforming vectors of the users across the network.

Unlike recent literature, this chapter considers a multi-CRAN and focuses on the problem of minimizing the network total power consumption subject to quality-of-service requirements, so as to determine the sparse beamforming solution of each user across the network and the active set of BSs within each cluster of BSs. The main contribution of this chapter is an algorithm that can be implemented in a decentralized fashion across the network, with a reasonable amount of information exchange between the clouds. The solution is based on Lagrangian duality theory using a dual decomposition approach, wherein the original optimization problem is decoupled into multiple and independent sub-problems, each in charge of finding the local dual functions associated with each cloud. By allowing a limited information exchange between the coupled clouds, the proposed approach guarantees a feasible solution to the original problem at each iteration. The chapter further proposes a centralized solution to the problem based on the group sparsity structure of the beamforming vectors. Simulation results show the performance improvement of the proposed solution as compared to the conventional per-cloud update solution.

3.1.1 Related Work

The problem considered in this chapter is related in part to the recent literature on CRAN, especially to reference [2] which addresses the joint BS selection and beamforming design problem in the downlink of a single cloud scenario. The work in [2] considers the problem of minimizing the network total power consumption and solves the problem using techniques from compressive sensing literature. This chapter is also related to reference [13] which addresses the user scheduling and beamforming vectors design problem by maximizing the weighted sum rate. Both references [2] and [13], however, assume a single-cloud scenario and do not account for inter-cloud interference. The problem in this chapter is further related to the multicell beamforming problem studied in conventional cellular networks literature, e.g., [14] [15], which propose decentralized algorithms to solve the problem. The studies in conventional cellular networks [14] [15], however, assume that one user can be served by one BS only, unlike the CRAN architecture which is a user-centric architecture and allows each user to be served by one or more BSs at the same time.

The multi-CRAN setup considered in this chapter is further related to the setup in [16][17]. In [16], an iterative auction approach is used to solve the user-to-cloud association problem which maximizes the network-wide utility, under the general assumption of a pre-known set of active BSs and beamforming vectors. In [17], further, an iterative algorithm based on MM approach is proposed to jointly design the precoding matrix and the correlation matrix of the quantization noise to address the weighted sum-rate maximization problem. The algorithm proposed in [17] is, however, centralized in nature, which is no longer feasible in practice, since otherwise, the clouds would require joint signal-processing and a high level of cloud-to-cloud backhaul communication.

3.1.2 Organization

The rest of this chapter is organized as follows, Section 3.2 presents the system model and formulates the network total power optimization problem. Then, Section 3.3 presents the iterative Decentralized Multi-Cloud Group Sparse Beamforming algorithm. Numerical examples illustrating the performance of the proposed algorithm are given in Section 3.4. Finally, Section 3.5 concludes this chapter.

3.2 System Model and Problem Formulation

3.2.1 System Model

Consider the downlink of a multi-CRAN with C clouds, where each cloud c is connected to B_c BSs via high-capacity wireline backhaul links. Each BS b is equipped with N_b antennas. The network comprises U single-antenna mobile users (MUs). Figure 3.1 illustrates an example of the considered network with 3 clouds, where each cloud is connected to 2 BSs. Let the set $\mathcal{C}$ with size $C = |\mathcal{C}|$ be the set of clouds, and $\mathcal{U}_c$ be the set of users associated with one cloud $c \in \mathcal{C}$. Further, let $\mathcal{A}_c$ denote the set of active BSs connected to cloud c. Let s_u be a complex scalar denoting the data symbol for user u, $\mathbf{w}_{cbu} \in \mathbb{C}^{N_b}$ be the beamforming vector at BS b of cloud c

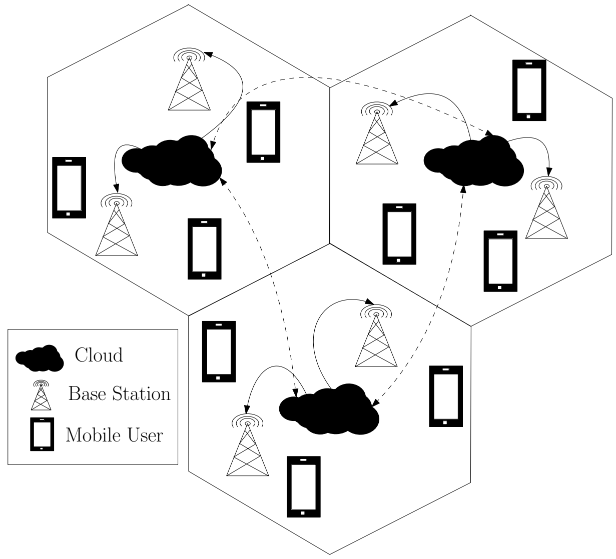

Figure 3.1: The architecture of multi-CRAN.

associated with user u's signal s_u, and $\mathbf{h}_{cbu} \in \mathbb{C}^{N_b}$ be the channel vector from BS b of cloud c to user u. The received signal $y_{cu} \in \mathbb{C}$ of user u in cloud c can be written as

$$\begin{aligned} y_{cu} = & \sum_{b \in \mathcal{A}_c} \mathbf{h}_{cbu}^H \mathbf{w}_{cbu} s_u + \sum_{u' \in \mathcal{U}_{cu}} \sum_{b \in \mathcal{A}_c} \mathbf{h}_{cbu}^H \mathbf{w}_{cbu'} s_{u'} \\ & + \sum_{c' \neq c} \sum_{b' \in \mathcal{A}_{c'}} \sum_{u' \in \mathcal{U}_{c'}} \mathbf{h}_{c'b'u}^H \mathbf{w}_{c'b'u'} s_{u'} + n_u, \end{aligned} \tag{3.1}$$

where $\mathcal{U}_{cu} = \mathcal{U}_c \setminus \{u\}$, and where $n_u \sim \mathcal{CN}(0, \sigma_u^2)$ represents the additive white Gaussian noise which is independent from the transmitted data symbols s_u.

Define the beamforming vector of user $u \in \mathcal{U}_c$ by $\mathbf{w}_{cu} = [\mathbf{w}_{cbu}^T, \quad \forall b \in \mathcal{A}_c]^T \in \mathbb{C}^{N_c}$, which is the stacking of all beamforming vectors due to all active BS $b \in \mathcal{A}_c$ associated with user u's signal s_u, where $N_c = \sum_{b \in \mathcal{A}_c} N_b$. Further, let the channel vector from the set of active BSs $b \in \mathcal{A}_c$ to user u be $\mathbf{h}_{cu} = [\mathbf{h}_{cbu}^T, \quad \forall b \in \mathcal{A}_c]^T \in \mathbb{C}^{N_c}$. Therefore, the received signal at user u in cloud c can be rewritten as

$$y_{cu} = \mathbf{h}_{cu}^H \mathbf{w}_{cu} s_u + \sum_{u' \in \mathcal{U}_{cu}} \mathbf{h}_{cu}^H \mathbf{w}_{cu'} s_{u'} + \sum_{c' \neq c} \sum_{u' \in \mathcal{U}_{c'}} \mathbf{h}_{c'u}^H \mathbf{w}_{c'u'} s_{u'} + n_u. \tag{3.2}$$

3.2.2 Power Model

In this chapter, we consider the following power consumption model of BS b [2]

$$P_{cb} = \begin{cases} \frac{1}{\nu_b} P_{cb,t} + P_{cb,a} & \text{if } P_{cb,t} > 0 \\ P_{cb,s} & \text{if } P_{cb,t} = 0 \end{cases}, \tag{3.3}$$

where $P_{cb,t} = \sum_{u \in \mathcal{U}_c} ||\mathbf{w}_{cbu}||^2$ denotes the total transmit power consumption of BS b connected to cloud c, $\nu_b > 0$ is the drain efficiency of the radio frequency (RF) power amplifier of BS b. Further, $P_{cb,a}$ denotes the power consumed by BS b connected to cloud c in the active mode and $P_{b,s}$ denotes the power consumed by BS b connected to cloud c in the sleep mode. Therefore, the total power consumption can be rewritten as follows

$$\begin{aligned} p(\mathcal{A}, \mathbf{w}) &= \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{A}_c} \sum_{u \in \mathcal{U}_c} \frac{1}{\nu_b} ||\mathbf{w}_{cbu}||^2 + \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{A}_c} P_{cb,a} + \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{Z}_c} P_{cb,s} \\ &= \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{A}_c} \sum_{u \in \mathcal{U}_c} \frac{1}{\nu_b} ||\mathbf{w}_{cbu}||^2 + \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{A}_c} (P_{cb,a} - P_{cb,s}) + \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{B}_c} P_{cb,s} \\ &= \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{A}_c} \sum_{u \in \mathcal{U}_c} \frac{1}{\nu_b} ||\mathbf{w}_{cbu}||^2 + \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{A}_c} P_b^r + \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{B}_c} P_{cb,s}, \end{aligned} \tag{3.4}$$

where $\mathcal{Z}_c$ denotes the inactive set of BS connected to cloud c and $P_b^r = P_{b,a} - P_{b,s}$ denotes the relative wireline backhaul link power consumption which represent the difference between the power consumption of the active BSs and their corresponding transport link in active and sleep modes.

3.2.3 Problem Formulation

Without loss of generality, assume that the average symbol power for user u is unity, i.e. $E(|s_u|^2) = 1$, and that the transmitted data symbols s_u are independent from each other. Therefore, based on (3.2), the SINR of user u served by cloud c can be expressed as

$$\Gamma_u = \frac{\left|\mathbf{h}_{cu}^H \mathbf{w}_{cu}\right|^2}{\sum\limits_{u' \in \mathcal{U}_{cu}} |\mathbf{h}_{cu}^H \mathbf{w}_{cu'}|^2 + \sum\limits_{c' \neq c} \sum\limits_{u' \in \mathcal{U}_{c'}} |\mathbf{h}_{c'u}^H \mathbf{w}_{c'u'}|^2 + \sigma_u^2}. \tag{3.5}$$

One of the constraints of the optimization problem studied in this chapter is the transmit power constraints of every BS b connected to cloud c, given by

$$\sum_{u=1}^{U_c} ||\mathbf{w}_{cbu}||_{\ell_2}^2 \leq P_{cb}, \; \forall \; (b, c) \in (\mathcal{A}_c \times \mathcal{C}), \tag{3.6}$$

where P_{cb} is a given nominal maximum transmit power.

The objective function of the optimization problem studied in this chapter is a generalization of the network total power consumption function studied in [2] in a single cloud setup. This chapter considers the network total power consumption across the multiple clouds (3.4), which can be written as follows

$$p(\mathcal{A}, \mathbf{w}) = \sum_{c=1}^{C} \sum_{b \in \mathcal{A}_c} \sum_{u \in \mathcal{U}_c} \frac{1}{\nu_b} ||\mathbf{w}_{cbu}||_{\ell_2}^2 + \sum_{c=1}^{C} \sum_{b \in \mathcal{A}_c} P_b^r, \tag{3.7}$$

where the constant term in (3.4) is omitted, $\mathcal{A} = \bigcup\limits_{c=1}^{C} \mathcal{A}_c$ denotes the set of active BSs in the entire network. The considered network power consumption model consists of the transmit power consumption of the active BSs and the relative wireline backhaul link power consumption.

The optimization problem studied in this chapter considers the minimization of the multi-CRAN network total power consumption subject to SINR and transmit power constraints, which can be formulated as follows

$$\begin{aligned} \min_{\mathbf{w},\mathcal{A}} \quad & \sum_{c=1}^{C} \sum_{b\in\mathcal{A}_c} \sum_{u\in\mathcal{U}_c} \frac{1}{\nu_b} ||\mathbf{w}_{cbu}||_{\ell_2}^2 + \sum_{c=1}^{C} \sum_{b\in\mathcal{A}_c} P_b^r \\ s.t. \quad & \sum_{u\in\mathcal{U}_c} ||\mathbf{w}_{cbu}||_{\ell_2}^2 \ \leq \ P_{cb}, \ \forall \ (b,c) \in (\mathcal{A}_c \times \mathcal{C}) \\ & \Gamma_u \geq \delta_u, \ \forall \ (u,c) \in (\mathcal{U}_c \times \mathcal{C}), \end{aligned} \tag{3.8}$$

where the optimization is over the beamforming vectors $\mathbf{w} = [\mathbf{w}_{cbu}^T, \quad \forall (b,u,c) \in (\mathcal{A}_c \times \mathcal{U}_c \times \mathcal{C})]^T \in \mathbb{C}^N$, where $N = \sum_{c=1}^{C} U_c N_c$, and over the set of active BSs $\mathcal{A}$, and where δ_u denotes the SINR target for user u. The above problem (3.8) is a mixed discrete and continuous optimization problem. Therefore, finding the global optimal solution for such optimization problem would likely require a combinatorial search, that is clearly infeasible for any sized network. Further, centralized solutions to the problem are no longer feasible, since otherwise, the clouds would require joint signal-processing and a high level of cloud-to-cloud backhaul communication. Thus, instead of looking for global optimal solutions, the following section proposes a sub-optimal decentralized algorithm to solve the problem. Simulation results suggest that the proposed decentralized algorithm converges to the centralized solution in a finite number of iterations. Furthermore, the proposed solution significantly outperforms the conventional per-cloud update solution.

3.3 Decentralized Multi-Cloud Group Sparse Beamforming Algorithm

3.3.1 Proposed Decentralized Solution

This section proposes a decentralization solution to the network total power minimization problem (3.8), using a dual decomposition approach, first proposed in the context of multi-cell systems [15]. Specifically, we first start by introducing inter-cloud interference terms, and additional auxiliary variables so as to decouple the SINR constraints. The obtained optimization problem can then be solved locally at each cloud with the help of a dual decomposition approach [8][6], thereby allowing the exchange of the local inter-cloud interference terms between the coupled clouds. Note that this approach decouples the original optimization problem into multiple and independent sub-problems in charge of finding the local dual functions, controlled by a master problem responsible for updating the dual variables.

Problem Relaxation

First, define the inter-cloud interference terms $\xi^2_{c'u}$ from an interfering cloud c' to user u served by any cloud (different than c') as follows

$$\xi^2_{c'u} = \sum_{u' \in \mathcal{U}_{c'}} \left| \mathbf{h}^H_{c'u} \mathbf{w}_{c'u'} \right|^2 . \tag{3.9}$$

By relaxing the equality in the inter-cloud interference term $\xi^2_{c'u}$ in (3.9) into an inequality, the network total power minimization problem (3.8) can be reformulated as

$$\begin{aligned} \min_{\mathbf{w},\mathcal{A},\boldsymbol{\xi}} \quad & \sum_{c=1}^{C} \sum_{b \in \mathcal{A}_c} \sum_{u \in \mathcal{U}_c} \frac{1}{\nu_b} ||\mathbf{w}_{cbu}||^2_{\ell_2} + \sum_{c=1}^{C} \sum_{b \in \mathcal{A}_c} P^r_b \\ s.t. \quad & \sum_{u \in \mathcal{U}_c} ||\mathbf{w}_{cbu}||^2_{\ell_2} \ \leq \ P_{cb}, \ \forall \ (b,c) \in (\mathcal{A}_c \times \mathcal{C}) \\ & \Gamma_u = \frac{|\mathbf{h}^H_{cu} \mathbf{w}_{cu}|^2}{\sum_{u' \in \mathcal{U}_{cu}} |\mathbf{h}^H_{cu} \mathbf{w}_{cu'}|^2 + \sum_{c' \neq c} \xi^2_{c'u} + \sigma^2_u} \geq \delta_u, \ \forall \ (u,c) \in (\mathcal{U}_c \times \mathcal{C}) \\ & \xi^2_{c'u} \geq \sum_{u' \in \mathcal{U}_{c'}} \left| \sum_{b' \in \mathcal{A}_{c'}} \mathbf{h}^H_{c'b'u} \mathbf{w}_{c'b'u'} \right|^2, \ \forall u \notin \mathcal{U}_{c'}, \ \forall c' \in \mathcal{C}, \end{aligned} \tag{3.10}$$

where the optimization is over the beamforming vectors, the set of active BSs in the entire network $\mathcal{A}$, and the inter-cloud interference vector $\boldsymbol{\xi} = [\xi_{c'u}, \quad \forall u \notin \mathcal{U}_{c'} \ \forall c' \in$

$\mathcal{C}]^T \in \mathbb{R}_+^{U'}$, where $U' = \sum_{c=1}^{C} \sum_{c' \neq c} U_{c'}$. Relaxing the inter-cloud interference constraints with inequality in the reformulated problem (3.10) is, in general, sub-optimal as compared to (3.8) since the achieved SINR is lower (i.e., more power is required to achieve the SINR target). However, the inequality turns out to be always tight.

Proposition 2. *The original optimization problem (3.8) and the relaxed optimization problem (3.10) have the same optimal solution.*

Proof. First, we prove by contradiction that the third constraints in (3.10) are achieved with equality at the optimal solution. To this end, let $\mathbf{w}^*$, $\mathcal{A}^*$ and $\boldsymbol{\xi}^*$ be the optimal solution of the optimization problem (3.10). In addition, we assume by contradiction that there exist at least $u \notin \mathcal{U}_{c'}$ and $c' \in \mathcal{C}$ such that

$$\xi^{*2}_{c'u} \geq \sum_{u' \in \mathcal{U}_{c'}} \left| \sum_{b' \in \mathcal{A}^*_{c'}} \mathbf{h}^H_{c'b'u} \mathbf{w}^*_{c'b'u'} \right|^2 . \tag{3.11}$$

Therefore, we can find $\tilde{\xi}^2_{c'u} < {\xi^*_{c'u}}^2$ which still satisfies the third constraints in (3.10). Such a $\tilde{\xi}^2_{c'u}$ further leads to a higher SINR value of user u. Therefore, the power constraint can be decreased by reducing the power transmitted from the cloud c to user u while satisfying the SINR constraints. This is a contradiction of the optimality assumption of ${\xi^*_{c'u}}^2$, which proves our proposition. □

Note that the SINR constraints in problem (3.10) are still coupled across the different clouds, because of the inter-cloud interference terms, which still prevents the derivation of a decentralized strategy. However, using a similar approach to the decoupling strategy, first proposed in the multi-cell problem [15], (3.10) can be decoupled by introducing a set of auxiliary local variables at each cloud and a set of corresponding equality constraints. Problem (3.10) can then be reformulated as

$$
\begin{aligned}
\min_{\mathbf{w},\mathcal{A},\tilde{\boldsymbol{\xi}}} \quad & \sum_{c=1}^{C} \sum_{b\in\mathcal{A}_c} \sum_{u\in\mathcal{U}_c} \frac{1}{\nu_b} ||\mathbf{w}_{cbu}||_{\ell_2}^2 + \sum_{c=1}^{C} \sum_{b\in\mathcal{A}_c} P_b^r \\
s.t. \quad & \sum_{u\in\mathcal{U}_c} ||\mathbf{w}_{cbu}||_{\ell_2}^2 \ \leq \ P_{cb}, \ \forall \ (b,c) \in (\mathcal{A}_c \times \mathcal{C}) \\
& \Gamma_u^{(c)} = \frac{|\mathbf{h}_{cu}^H \mathbf{w}_{cu}|^2}{\sum\limits_{u'\in\mathcal{U}_{c_u}} |\mathbf{h}_{cu}^H \mathbf{w}_{cu'}|^2 + \sum\limits_{c'\neq c} \xi_{c'u}^{(c)^2} + \sigma_u^2} \geq \delta_u, \ \forall \ (u,c) \in (\mathcal{U}_c \times \mathcal{C}) \\
& \xi_{c'u}^{(c')^2} \geq \sum_{u'\in\mathcal{U}_{c'}} \left| \sum_{b'\in\mathcal{A}_{c'}} \mathbf{h}_{c'b'u}^H \mathbf{w}_{c'b'u'} \right|^2, \ \forall u \notin \mathcal{U}_{c'}, \ \forall c' \in \mathcal{C} \\
& \xi_{c'u}^{(c')} = \xi_{c'u}^{(c_u)}, \ \forall \ (u,c') \in (\mathcal{U} \times \bar{\mathcal{C}}_u),
\end{aligned}
\tag{3.12}
$$

where c_u denotes the cloud serving user u, $\bar{\mathcal{C}}_u = \{c|c \neq c_u\}$, and where the optimization is over the beamforming vectors, the set of active BSs $\mathcal{A}$, and the inter-cloud interference vector $\tilde{\boldsymbol{\xi}} = [\xi_{c'u}^{(c)} \ \forall c \in \{c', c_u\} \ \forall u \notin \mathcal{U}_{c'} \ \forall c' \in \mathcal{C}]^T \in \mathbb{R}_+^{2U'}$. Note that by introducing the consistency constraints $\xi_{c'u}^{(c')} = \xi_{c'u}^{(c_u)}$ in (3.12), the SINR constraints become separable.

Subproblems Formulation

The distributed algorithm to solve problem (3.12) is obtained by adopting a dual decomposition approach. Specifically, we start by forming the partial Lagrangian over the consistency constraints as (i.e., by first accounting to the consistency constraints only)

$$
\begin{aligned}
\mathcal{L}(\mathbf{w},\mathcal{A},\boldsymbol{\xi},\boldsymbol{\lambda}) = & \sum_{c=1}^{C} \sum_{b\in\mathcal{A}_c} \sum_{u\in\mathcal{U}_c} \frac{1}{\nu_b} ||\mathbf{w}_{cbu}||_{\ell_2}^2 + \sum_{c=1}^{C} \sum_{b\in\mathcal{A}_c} P_b^r \\
& + \sum_{u=1}^{U} \sum_{c'\neq c_u} \lambda_{c'u} \left(\xi_{c'u}^{(c')} - \xi_{c'u}^{(c_u)} \right),
\end{aligned}
\tag{3.13}
$$

where $\lambda_{c'u}$ denotes the Lagrange multiplier associated with the consistency constraint $\xi_{c'u}^{(c')} = \xi_{c'u}^{(c_u)}$, and $\boldsymbol{\lambda} = [\lambda_{c'u}, \quad \forall u, \quad \forall c' \in \bar{\mathcal{C}}_u]^T \in \mathbb{R}^{U'}$.

Lemma 1. *The right hand side term in the partial Lagrangian (3.13) can be rewritten as follows*

$$J = \sum_{u=1}^{U} \sum_{c' \neq c_u} \lambda_{c'u} \left(\xi_{c'u}^{(c')} - \xi_{c'u}^{(c_u)} \right) = \sum_{c=1}^{C} \boldsymbol{\lambda}_c^T \boldsymbol{\xi}^{(c)}, \tag{3.14}$$

where the local dual variable $\boldsymbol{\lambda}_c = [\lambda_{cu'}\ \forall u' \notin \mathcal{U}_c,\ \ -\lambda_{c'u}\ \forall c' \neq c\ \forall u \in \mathcal{U}_c]^T$ *and the local inter-cloud interference vector* $\boldsymbol{\xi}^{(c)} = [\xi_{cu'}^{(c)}\ \forall u' \notin \mathcal{U}_c,\ \xi_{c'u}^{(c)}\ \forall c' \neq c\ \forall u \in \mathcal{U}_c]^T$*.*

Proof. Define the set $S_1 = \{\ (c, u') \mid c \in \mathcal{C}, u' \notin \mathcal{U}_c\ \}$,and the set $S_2 = \{\ (c, u) \mid c \in \mathcal{C}, u \in \mathcal{U}_c\ \}$. S_1 and S_2 are respectively equivalent to the following sets $S_1' = \{\ (c, u) \mid u \in \mathcal{U}, c \neq c_u\ \}$ and $S_2' = \{\ (c, u) \mid u \in \mathcal{U}, c = c_u\ \}$. By exploiting this equivalence, J can be written as follows

$$\begin{aligned} J &= \sum_{u=1}^{U} \sum_{c' \neq c_u} \lambda_{c'u} \left(\xi_{c'u}^{(c')} - \xi_{c'u}^{(c_u)} \right) \\ &= \sum_{u=1}^{U} \sum_{c \neq c_u} \lambda_{cu} \xi_{cu}^{(c)} - \sum_{u=1}^{U} \sum_{c \neq c_u} \lambda_{cu} \xi_{cu}^{(c_u)} \\ &= \sum_{c=1}^{C} \left\{ \sum_{u' \notin \mathcal{U}_c} \lambda_{cu'} \xi_{cu'}^{(c)} - \sum_{u \in \mathcal{U}_c} \sum_{c' \neq c} \lambda_{c'u} \xi_{c'u}^{(c)} \right\} \\ &= \sum_{c=1}^{C} \boldsymbol{\lambda}_c^T \boldsymbol{\xi}^{(c)}. \quad \square \end{aligned} \tag{3.15}$$

Using the above Lemma, (3.12) can be transformed into a separable optimization problem, wherein the local dual functions, called $g_c(\boldsymbol{\lambda}_c)$, can be found separately at each cloud as follows

$$\begin{aligned} g_c(\boldsymbol{\lambda}_c) = \min_{\mathbf{w}, \mathcal{A}, \boldsymbol{\xi}^{(c)}} & \sum_{b \in \mathcal{A}_c} \sum_{u \in \mathcal{U}_c} \frac{1}{\nu_b} ||\mathbf{w}_{cbu}||_{\ell_2}^2 + \sum_{b \in \mathcal{A}_c} P_b^r + \boldsymbol{\lambda}_c^T \boldsymbol{\xi}^{(c)} \\ s.t. \quad & \sum_{u \in \mathcal{U}_c} ||\mathbf{w}_{cbu}||_{\ell_2}^2 \ \leq \ P_{cb},\ \forall\ b \in \mathcal{A}_c \\ & \Gamma_u^{(c)} = \frac{|\mathbf{h}_{cu}^H \mathbf{w}_{cu}|^2}{\sum_{u' \in \mathcal{U}_{cu}} |\mathbf{h}_{cu}^H \mathbf{w}_{cu'}|^2 + \sum_{c' \neq c} \xi_{c'u}^{(c)^2} + \sigma_u^2} \geq \delta_u \\ & \xi_{cu'}^{(c)^2} \geq \sum_{u \in \mathcal{U}_c} \left| \sum_{b \in \mathcal{A}_c} \mathbf{h}_{cbu'}^H \mathbf{w}_{cbu} \right|^2,\ \forall\ u' \notin \mathcal{U}_c. \end{aligned} \tag{3.16}$$

For a fixed active set of BSs $\mathcal{A}$, the optimization problem (3.16) can be easily recast as a SOCP problem, which can then be solved using efficient numerical algorithms [4]. Therefore, the global optimal solution of the above optimization problem can be obtained by searching over all the possible active sets of BSs and selecting the set that provides the minimum cost value. Such approach to solve the optimization problem (3.16) grows exponentially with the number of BSs, which makes it unpractical. Since the minimization problem in formulation (3.16) is decoupled for every cloud, this chapter rather proposes solving (3.16) using an iterative group sparse beamforming (GSBF) algorithm, first proposed in a single cloud setup [2]. Such approach provides a sub-optimal solution to (3.16), but has a maximum number of iterations that is linear in the number of BSs. The GSBF approach relies on finding a tightest convex positively homogeneous lower bound [12] of the cost function, and then uses the MM algorithm [11] in order to induce group sparsity.

Proposition 3. *The tightest convex positively homogeneous lower bound of the sparse representation of the cost function in (3.16) is given by*

$$\Omega(\mathbf{w}_c) = 2\sum_{b=1}^{B_c} \sqrt{\frac{P_b^r}{\nu_b}}||\mathbf{w}_{cb}||_{\ell_2} + \boldsymbol{\lambda}_c^T \boldsymbol{\xi}^{(c)}. \tag{3.17}$$

The proof of the above proposition uses steps similar to the ones used in [2] and used in chapter 2, and hinges upon the fact that the second term in (3.17) is convex. Steps for determining the MM algorithm and the sparse representation of the cost function in (3.16) are also omitted in this chapter, as they mirror the steps used in [2].

Master Problem

Solving (3.16) for every local dual function initially at every cloud c requires the knowledge of the local dual variable $\boldsymbol{\lambda}_c$. Such variables are found by solving the master problem of the dual decomposition, which can be written as follows

$$\max_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda}) = \sum_{c=1}^{C} g_c(\boldsymbol{\lambda}_c). \tag{3.18}$$

This dual problem can be solved iteratively and independently at each cloud using the sub-gradient method with the following update rule

$$\lambda_{c'u}(i+1) = \lambda_{c'u}(i) + \mu \left(\xi_{c'u}^{(c')} - \xi_{c'u}^{(c_u)} \right), \ \forall\ c',\ \forall\ u, \tag{3.19}$$

where $\xi_{c'u}^{(c')} - \xi_{c'u}^{(c_u)}$ is a sub-gradient of the dual function g at $\lambda_{c'u}$ [8], i represents the iteration index and μ denotes the step size. In this work, we only focus on a diminishing step size. As the simulations section suggest, the decentralized algorithm converges to the centralized solution in a finite number of iterations. Note that at each iteration of the sub-gradient algorithm, the coupled clouds need to exchange their current local inter-cloud interference terms, which is a reasonable level of information exchange.

Feasible Primal Solution

The dual decomposition approach consists of iterating between the optimization problem (3.16) and the update rule (3.19). At each iteration of this algorithm, the introduced auxiliary variables may not lead to a feasible solution for the original network total power minimization problem (3.8), i.e., $\xi_{c'u}^{(c')} \neq \xi_{c'u}^{(c_u)}$ during the intermediate iteration. To reach a set of feasible solution at each iteration, we propose using fixed average inter-cloud interference terms defined as follows

$$\xi_{c'u} = \frac{1}{2} \left(\xi_{c'u}^{(c')} + \xi_{c'u}^{(c_u)} \right), \ \forall\ u \notin \mathcal{U}_{c'},\ \forall\ c' \in \mathcal{C}. \tag{3.20}$$

Then, resolve the optimization problem (3.16) for fixed average inter-cloud interference terms. Note that the obtained solution remains sub-optimal, yet a feasible, solution to the original problem (3.8).

Iterative Decentralized Algorithm

The proposed decentralized solution requires to iterate between two levels. At the first level, multiple and independent sub-problems are solved locally. At the second level, there is a master problem responsible for updating the dual variables. At each iteration, a solution of feasible beamforming vectors and active set of BSs is estimated locally. The proposed iterative Decentralized Multi-Cloud Group Sparse Beamforming (I-DMCGSB) algorithm is summarized in Table (1).

Algorithm 1 The iterative Decentralized Multi-Cloud Group Sparse Beamforming (I-DMCGSB)

Initialization : Initialize the dual variables $\boldsymbol{\lambda}_c(0)$ and set the iteration index to $i = 0$.

Repeat :

1: Solve the optimization problem (3.16). Then, broadcast the local inter-cloud interference terms to the adjacent clouds. If it is infeasible, go to **End**;

2: Update the dual variables as in (3.19);

3: *Optional: Compute the average inter-cloud interference terms as in (3.20). Then, solve the optimization problem (3.16) using the fixed average inter-cloud interference terms to estimate a feasible primal solution.*

4: Update the iteration index $i = i + 1$;

Until : Desired level of convergence.

5: Find the optimal beamforming vectors $\mathbf{w}^*_{cbu}$, $\forall\ b \notin \mathcal{B}_c$, $\forall\ u \in \mathcal{U}_c$, the active set of BSs $\mathcal{A}^*_c$ and the optimal power consumption $p^*_c(\mathcal{A}^*_c, \mathbf{w}^*_{cu})$.

End

3.3.2 Centralized Solution

In this section, we derive a centralized solution for the optimization problem (3.8). To this end, we assume that all the processing is performed in a shared processor connecting all the clouds. The total power minimization problem (3.8) can be reformulated as

$$\begin{aligned}
\min_{\mathbf{w}',\mathcal{A}} \quad & \sum_{b\in\mathcal{A}} \sum_{u\in\mathcal{U}} \frac{1}{\nu_b} ||\mathbf{w}_{bu}||_{\ell_2}^2 + \sum_{b\in\mathcal{A}} P_b^r \\
s.t. \quad & \sum_{u\in\mathcal{U}} ||\mathbf{w}_{bu}||_{\ell_2}^2 \ \leq \ P_b, \ \forall\ b \in \mathcal{A} \\
& \Gamma'_u = \frac{|\mathbf{h}'^H_u \mathbf{w}'_u|^2}{\sum_{u'\neq u} |\mathbf{h}'^H_u \mathbf{w}'_{u'}|^2 + \sigma_u^2} \geq \delta_u, \ \forall\ u \in \mathcal{U} \\
& ||\mathbf{w}_{bu}||_{\ell_2} \ \leq \ 0, \ \forall\ (u,b) \in (\mathcal{U} \times \bar{\mathcal{A}}_u),
\end{aligned} \tag{3.21}$$

where $\mathbf{w}'_u$ denotes the beamforming vector of user u which is defined as the stacking of all beamforming vectors used by the set of active BSs $b \in \mathcal{A}$ to serve user u, and $\mathbf{h}'_u$ represents the channel vector from the entire set of active BSs $b \in \mathcal{A}$ to user u, $\mathbf{w}_{bu}$ denotes the beamforming vector at BS b for user u, $N' = U \sum_{b\in\mathcal{A}} N_b$, $\bar{\mathcal{A}}_u$ denotes the active set of BSs that can not serve user u and P_b denotes the maximum transmit power of BS b. In order to guarantee the equivalence between the centralized formulation and the original network total power minimization problem (3.8), we introduce additional inequality constraints $||\mathbf{w}_{bu}||_{\ell_2} \ \leq \ 0, \ \forall\ (u,b) \in (\mathcal{U} \times \bar{\mathcal{A}}_u)$, so as to impose that the beamforming vectors form BS b to user u connected to different cloud is equal to zero. Thus, solving the optimization problem (3.21) is equivalent to solving the original problem (3.8).

The introduced constraints in (3.21) are SOC constraints. The above optimization problem can, therefore, be recast as a SOCP problem for a fixed active set of BSs $\mathcal{A}$, which can then be solved efficiently [4]. In order to reduce the complexity, the iterative GSBF algorithm can be used which provides a sub-optimal solution to (3.21) with low computational complexity.

3.3.3 Per-Cloud Update Solution

Conventional CRANs operate on a per-cloud basis; inter-cloud interference is treated as background noise. The baseline approach to solve (3.8) then consists of iteratively solving the network total power consumption minimization problem on a per-cloud basis, where in each iteration, the cloud under consideration solves the problem locally while treating inter-cloud interference terms as part of the background noise. Specifically, in iteration i, cloud c determines its set of beamformers and active set of BSs for fixed inter-cloud interference. The resulting solution then affects the solution at cloud c' in the following iteration, as the inter-cloud interference terms need to be updated. Such solution, called per-cloud update solution, iteratively proceeds in updating the solution of each cloud, in view of the other clouds' solutions.

3.3.4 Complexity Analysis

The implementation of the Iterative Decentralized Multi-Cloud Group Sparse Beamforming algorithm requires to solve the optimization problem (3.16) using the GSBF algorithm and then it updates the dual variables at each iteration. Therefore, the computation complexity of the I-DMCGSB algorithm comes mainly from solving the optimization problem (3.16) at each cloud c. Further, the computational complexity of the GSPF algorithm comes mainly from solving an SOCP problem with computational complexity proportional to $\mathcal{O}(B_c^{3.5})$ for cloud c using the interior-point method [4]. The centralized solution directly solves the optimization problem (3.21) using the GSBF algorithm. Therefore, the computational complexity of the centralized solution at each iteration is proportional to $\mathcal{O}(B^{3.5})$ using the interior point method. Further, the per-cloud update solution requires to iteratively solve the optimization problem (3.12) using the GSBF algorithm with fixed inter-cloud interference terms. Therefore, the computational complexity of the per-cloud update solution is proportional to $\mathcal{O}(B_c^{3.5})$ for cloud c at each iteration using the interior point method.

The decentralized algorithm solves iteratively the considered optimization problem. Therefore, it may lead to higher computational complexity compared to the centralized solution since the SOCP problems solved at each iteration of the distributed and centralized algorithms have comparable complexity and the distributed algorithm needs more iterations to converge. However, centralized solution is not practical, since the clouds would require joint signal-processing and a high level of cloud-to-cloud communication. In large-scale network, the decentralized algorithm can lead to lower computational complexity compared to the centralized algorithm since it solves the centralized problem in a distributed fashion which reduces the complexity of the SOCP problems solved at each cloud. The per-cloud update solution and the decentralized solution have comparable computational complexity. However, simulation results show that the decentralized solution perform better than the per-cloud update solution.

3.3.5 Convergence Properties

This chapter uses the dual decomposition to construct a distributed algorithm to the network total power minimization problem in a multi-CRAN scenario. The first step of the dual decomposition approach consists of solving the optimization problem (3.16). Since the optimization problem (3.16) is a mixed discrete and continuous optimization problem, the computational complexity of directly solving it grows exponentially with the number of BSs, which makes it unpractical. The iterative GSBF algorithm [2] is used to construct a sub-optimal solution to (3.16) with reasonable computational complexity. However, the dual decomposition approach needs the global optimal solution of (3.16) to guarantee the convergence to the decentralized solution. The gap between the sub-optimal solution achieved by the GSBF algorithm and the global optimal solution of the optimization problem (3.16) is relatively small especially for small sized network. When the decentralized solution does not converge to the centralized one, it is guaranteed that the solution obtained using the dual decomposition based algorithm (1) is feasible for the original problem.

3.4 Simulation Results

This section investigates the performance of the proposed algorithms. Consider a two-cloud CRAN, where each cloud is connected to two BSs, i.e., $B_1 = B_2 = 2$. Each cluster of BSs serves two single-antenna MUs. Each BS is equipped with five antennas. We assume that the BSs and MUs in the first cloud are uniformily and independently distributed in the square region $[0\ 2000]\times[0\ 2000]$ meters. In the second cloud, the BSs and MUs are uniformily and independently distributed in the square region $[2000\ 4000]\times[0\ 2000]$ meters. The channel vectors from the BSs to the MUs are generated assuming a distance-dependent path loss $L(d_{bu}) = 128.1 + 37.6\log_{10}(d_{bu})$, and Rayleigh fading component, where d_{bu} denotes the distance between BS b and user u in kilometers. The noise power spectral density is $\sigma_u^2 = -96$ dBm $\forall u$. The transmit antenna power gain is set to 9 dBi. We set the initial dual variable to $\lambda_{cu}(0) = 0.01,\ \forall\ u \in \mathcal{U},\ \forall\ c \in \bar{\mathcal{C}}_u$. The drain efficiency of the radio frequency power amplifier is set to $\nu_b = 0.25,\ \forall\ b$. The relative wireline backhaul link power consumption is set to $P_b^r = 2 + l_b,\ \forall\ b$, where l_b denotes the number of BS b in cloud c. We set the maximum transmit power of BS b in cloud c to $P_{cb} = 1W$.

First, we set the SINR target to 5 dB and the step size $\mu = \frac{2}{\sqrt{i}}$. Figure 3.2 plots the network power consumption (in Watts) versus the number of iterations so as to illustrate the convergence behavior of the proposed I-DMCGSB algorithm and the conventional per-cloud update algorithm. The figure shows that the estimated feasible primal solution is a tight upper bound on the solution of the original optimization problem. It can be noticed that our proposed decentralized solution I-DMCGSB algorithm converges to the centralized solution in less than 100 iterations. Figure 3.2 particularly shows that the proposed joint optimization solution (centralized and

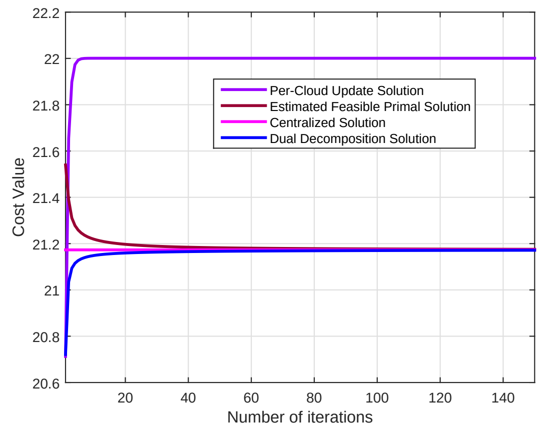

Figure 3.2: Convergence behavior of the I-DMCGSB and the conventional per-cloud update algorithms with SINR target=5 dB and $K = 2$.

decentralized) significantly outperforms the conventional per-cloud update solution.

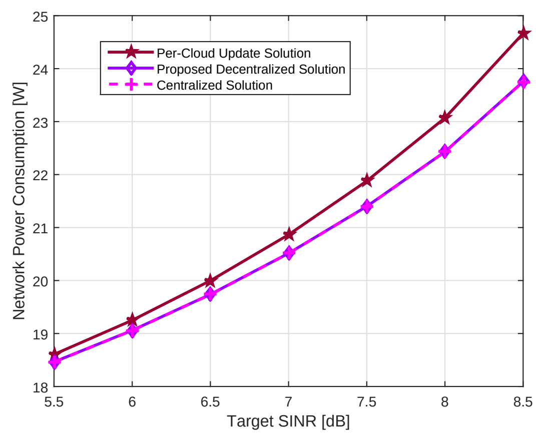

Figure 3.3: Network total power consumption versus SINR target with $K = 2$.

Then, we compare the energy efficiency of the proposed and the conventional per-cloud update decentralized algorithms. Figure 3.3 plots the network total power consumption as a function of SINR target. The figure shows that the proposed decentralized solution has the same performance as the centralized solution for all values of SINR target. Furthermore, the proposed solution significantly outperforms the per-cloud update solution, especially at high SINR target.

Finally, to investigate the impact of changing the step size on the convergence speed of the algorithm, Figure 3.4 shows $P_{\text{decentralized}} - P_{\text{centralized}}$ as a function of the iteration index for different step sizes under a SINR target of 4 dB. The figure shows that the convergence speed of the proposed decentralized I-DMCGSB algorithm is sensitive to the used step size. Figure 3.4 especially shows how a well chosen step size may lead to a significant improvement in the convergence speed of the algorithm.

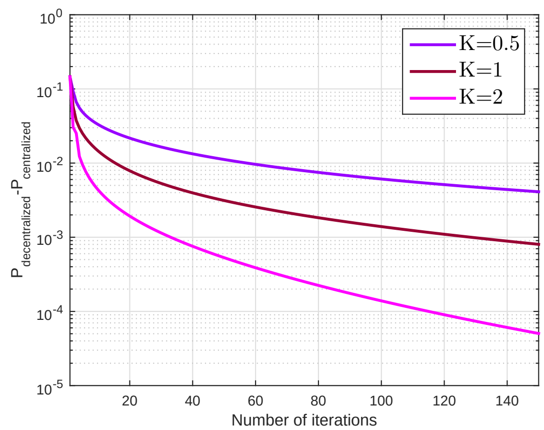

Figure 3.4: Suboptimality versus iteration number for dual decomposition with different step-sizes and fixed SINR target=4 dB.

3.5 Conclusion

In this chapter, a downlink multi-cloud radio access network is considered, where each cloud is connected to several multiple-antenna BSs via high capacity backhaul links. The chapter proposes a decentralized algorithm to solve the network total power minimization problem using a dual decomposition approach. A feasible active set of BSs and beamforming vectors can be obtained locally at each iteration by allowing a limited information exchange between the adjacent clouds. Simulation results show that the proposed decentralized algorithm converges to the centralized solution in a finite number of iterations, and the proposed algorithms significantly outperform the conventional per-cloud update solution, especially at high signal-to-noise-plus-interference- ration (SINR) target.

Chapter 4

Distributed Robust Power Minimization for the Downlink of Multi-CRAN

The algorithm design and performance analysis in the previous chapter relies on important assumptions, such as perfect CSI and unlimited capacity backhaul links, which may not always be feasible. This chapter[1] considers the downlink of a multi-CRAN where each cloud is connected to several BSs through limited-capacity wireline backhaul links. The performance of the system becomes therefore a function of not only inter-cloud and intra-cloud interference but also the compression schemes of the limited capacity backhaul links. The chapter assumes independent compression scheme and imperfect CSI where the CSI errors belong to an ellipsoidal bounded region. The problem of interest becomes the one of minimizing the network total transmit power subject to BS power and quality of service constraints, as well as backhaul capacity and CSI error constraints. The chapter suggests solving the problem using the ADMM approach. One of the highlight of the chapter is that the proposed ADMM-based algorithm can be implemented in a distributed fashion

[1]A part of this chapter will appear in: O. Dhifallah, H. Dahrouj, T. Y.Al-Naffouri, M-S. Alouini, "Distributed Robust Power Minimization for the Downlink of Multi-Cloud Radio Access Networks", *submitted to Globecom 2016* .

across the multi-cloud network by allowing a limited amount of information exchange between the coupled clouds. Simulation results show that the proposed distributed algorithm provides a similar performance to the centralized algorithm in a reasonable number of iterations.

4.1 Introduction

The existing strategies in CRAN omit the inter-cloud interference and assume that one single cloud is connected to BSs through high capacity backhaul links. It is mostly also assumed that perfect CSI is available at the cloud, which is not often practical. This chapter considers the more realistic scenario of a multi-CRAN where each cloud is connected to several BSs through limited capacity backhaul links, and only imperfect CSI is available at each cloud. Further, each cluster of BSs serves a pre-known set of single-antenna MUs. The chapter assumes that imperfect CSI is available at the cloud, where the CSI errors belong to a known ellipsoidal region. The network performance becomes therefore a direct function of the quantization noise covariance matrix (assumed to be diagonal for mathematical tractability, i.e., independent compression schemes), the beamforming vector of each user across the network, and the BS to user association strategy.

Unlike recent literature, this chapter considers a multi-CRAN and formulates the problem of minimizing the network total power consumption subject to per-BS power and quality of service constraints, as well as the backhaul capacity limits and CSI error constraints, so as to determine the beamforming vector of each user across the network and the optimal quantization noise covariance matrix associated with each cloud. Using the SDP procedure [18] and the S-Procedure method [4], the considered non-convex optimization problem can be reformulated as a convex SDP minimization problem. Since centralized solutions require joint signal processing and high signaling overhead, the chapter proposes an ADMM-based distributed algorithm to solve the formulated problem (Another interesting application of the ADMM algorithm is provided in Appendix B). Specifically, each cloud computes the beamforming vectors of its users locally where a limited information exchange between the adjacent clouds is allowed. Then, it performs a joint precoding/beamforming of user signals which are sent to the BSs with nonzero beamforming coefficients through the limited-capacity backhaul links. Simulation results show that the proposed distributed algorithm provides a similar performance to the centralized algorithm in a reasonable number of iterations.

4.1.1 Related Work

The optimization problem considered in this chapter is partly related to reference [19] which solves the weighed sum rate optimization problem for the data-sharing and the compression-based strategies in the downlink of a single cloud. The chapter is further related to reference [20] which designs the quantization noise covariance matrix of the limited capacity wireline backhaul links, the transmit power of the wireless backhaul links and the beamforming vector of each user across the network by minimizing the network total power consumption. However, references [19] [20], consider a single cloud scenario and assume that perfect channel state information is available at the cloud. This chapter is also related to reference [7] which considers a muli-CRAN scenario, and provides a distributed solution to the sparse beamforming problem based on a dual decomposition approach; however, reference [7] assumes perfect CSI at the cloud, and infinite capacity backhaul links, unlike the current chapter. Additionally, the optimization problem considered in this chapter is related to the multi-cell beamforming problem studied in [21][22]. The works in [21][22] solve the robust beamforming optimization problem by providing a distributed algorithm based on ADMM and primal decomposition respectively. However, references [21][22] assume that each user can be served by one BS only, unlike the CRAN scenario. The performance in [21][22] is, further, not optimized over the quantization noise covariance matrix.

4.1.2 Organization

The rest of this chapter is organized as follows, Section 4.2 presents the system model and formulates the total network power optimization problem. Then, Section 4.3 presents the proposed iterative decentralized algorithm. Numerical examples illustrating the performance of the proposed distributed algorithm are given in Section 4.4. Finally, Section 4.5 concludes this chapter.

4.2 System Model and Problem Formulation

4.2.1 System Model

Consider the downlink of a multi-CRAN with C clouds where each cloud is connected to B BSs through capacity-limited backhaul links. The network comprises several single-antenna MUs. For simplicity of analysis, the chapter assumes that both BSs and MUs have single-antenna each. Figure 4.1 illustrates a typical example of the considered architecture with 3 clouds, where each cloud is connected to 3 BSs and is serving 6 MUs.

Let $\mathcal{U}_{cb}$ denotes the set of users associated with BS b connected to cloud c. Let $\mathbf{w}_{cu} = [w_{c1u},\ w_{c2u}, .., \ w_{cB_cu}]^T \in \mathbb{C}^{B_c}$ be the beamforming vector from cloud c to user u. We assume that if BS b, connected to cloud c, is not serving user u, i.e. $u \notin \mathcal{U}_{cb}$, then the corresponding beamforming coefficient w_{cbu} is set to zero. Let $\mathbf{h}_{cu} = [h_{c1u},\ h_{c2u}, .., \ h_{cB_cu}]^T \in \mathbb{C}^{B_c}$ be the channel vector from cloud c to user u. The

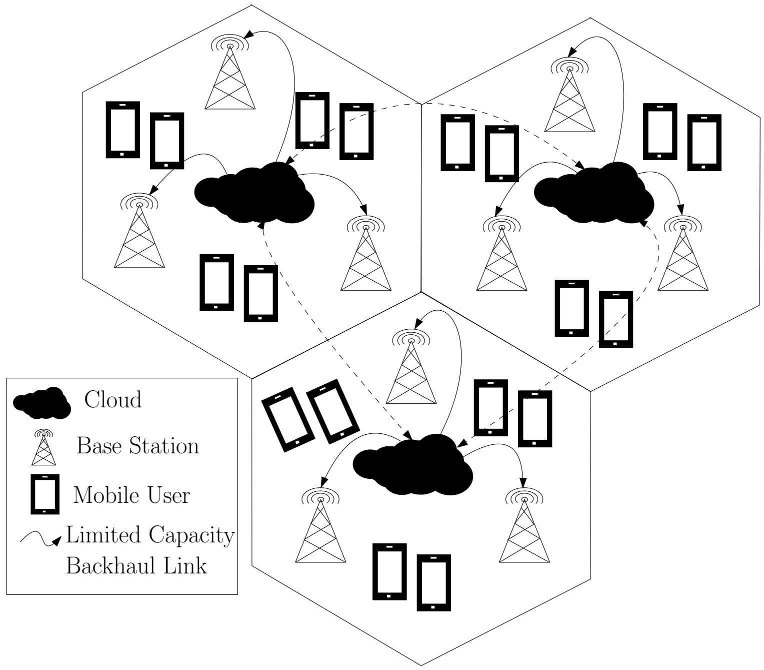

Figure 4.1: An example of Multi-CRAN, in which, the BSs are connected to a cloud through wireline capacity-limited backhaul links.

received signal $y_u \in \mathbb{C}$ at user u can be written as

$$y_u = \mathbf{h}_{cu}^H \mathbf{w}_{cu} s_u + \sum_{u' \in \mathcal{U}_{cu}} \mathbf{h}_{cu}^H \mathbf{w}_{cu'} s_{u'} + \sum_{c \in \mathcal{C}} \sum_{b \in \mathcal{B}_c} h_{cbu}^* v_{cb} + \sum_{c' \neq c} \sum_{u' \in \mathcal{U}_{c'}} \mathbf{h}_{c'u}^H \mathbf{w}_{c'u'} s_{u'} + n_u, \tag{4.1}$$

where v_{cb} denotes the quantization noise scalar assumed to be independent from the transmitted data symbols s_u, and where the additive white Gaussian noise n_u is assumed to be independent from the quantization noise v_{cb}. Let $\mathbf{v}_c = [v_{c1}, v_{c2}, \cdots, v_{cB_c}]$ be the quantization noise vector of cloud c which is assumed to be non-uniform white Gaussian with diagonal covariance matrix $\mathbf{Q}_c \in \mathbb{C}^{B_c \times B_c}$ where the diagonal entries q_{cb} satisfy

$$q_{cb} \geq 0, \; \forall \, b \in \mathcal{B}_c. \tag{4.2}$$

4.2.2 Problem Formulation

This chapter considers the problem of minimizing the network total power consumption while satisfying the per-BS power and quality of service constraints, as well as the backhaul-capacity limits and CSI error constraints. Specifically, each cloud designs the beamforming vectors of its users locally where only a limited information exchange between the adjacent clouds is allowed. After designing the optimal beamforming vectors, each cloud c transmits the local users message s_u and the beamforming coefficients $\mathbf{w}_{cu}$ to the BSs with nonzero value in $\mathbf{w}_{cu}$. In this chapter, the CSI errors are assumed to be bounded by an ellipsoidal region. Therefore, the true channel vector from cloud c to user u can be expressed as follows

$$\mathbf{h}_{cu} = \tilde{\mathbf{h}}_{cu} + \mathbf{e}_{cu},\ \forall\ (u, c) \in (\mathcal{U}_c \times \mathcal{C}), \tag{4.3}$$

where $\tilde{\mathbf{h}}_{cu}$ and $\mathbf{e}_{cu}$ denote respectively the estimated channel vector and the CSI error vector. Further, the CSI error vector satisfies the following elliptical constraints

$$\mathbf{e}_{cu}^H \mathbf{E}_{cu} \mathbf{e}_{cu} < 1,\ \forall\ (u, c) \in (\mathcal{U}_c \times \mathcal{C}), \tag{4.4}$$

where $\mathbf{E}_{cu}$ is a positive definite matrix specifying the accuracy of the CSI. Without loss of generality, the data symbols are assumed to have unit power, i.e. $E(|s_u|^2) = 1$, $\forall u$, and independent from each other. In this case, the SINR of user u can be expressed as

$$\Gamma_u = \frac{\left|\mathbf{h}_{cu}^H \mathbf{w}_{cu}\right|^2}{\sum\limits_{u' \in \mathcal{U}_{cu}} |\mathbf{h}_{cu}^H \mathbf{w}_{cu'}|^2 + \mathbf{h}_{cu}^H \mathbf{Q}_c \mathbf{h}_{cu} + \sum\limits_{c' \neq c} \psi_{c'u}^2 + \sigma_u^2}, \tag{4.5}$$

where $\psi_{c'u}^2 = \sum\limits_{u' \in \mathcal{U}_{c'}} \left|\mathbf{h}_{c'u}^H \mathbf{w}_{c'u'}\right|^2 + \mathbf{h}_{c'u}^H \mathbf{Q}_{c'} \mathbf{h}_{c'u}$ denotes the inter-cloud interference terms from an interfering cloud c' to user u served by cloud c, for $c \neq c'$. The transmit power constraints of each BS b served by cloud c is one of the constraints of the optimization problem studied in this chapter and is given by

$$\sum_{u=1}^{U_{cb}} |w_{cbu}|^2 + q_{cb} \leq P_{cb},\ \forall\ (b,c) \in (\mathcal{B}_c \times \mathcal{C}). \tag{4.6}$$

Assuming that the backhaul links have limited capacity, the cloud needs to perform joint precoding and compression of user signals before forwarding them to the corresponding BSs. Such compression introduces quantization noise to the user signal. As a result, the beamforming vector associated with user u in cloud c, the quantization noise level q_{cb} and the backhaul capacity C_{cb} are related as follows

$$\log_2\left(1 + \frac{\sum\limits_{u=1}^{U_{cb}} |w_{cbu}|^2}{q_{cb}}\right) \leq C_{cb},\ \forall\ (b,c) \in (\mathcal{B}_c \times \mathcal{C}), \tag{4.7}$$

where we assume an ideal vector quantizer and independent quantization at each BS. Define δ_u to be the target SINR of user u. In order to successfully decode the received message, the information theoretical achievable data rate of each user u should be greater than the fixed rate transmission. This constraint can be translated as follows

$$\log_2\left(1 + \Gamma_u\right) \geq R_u,\ \forall\ (u,c) \in (\mathcal{U}_c \times \mathcal{C}), \tag{4.8}$$

where $R_u = \log_2\left(1 + \delta_u\right)$ denotes the fixed rate transmission of user u.
The optimization problem studied in this chapter is the minimization of the network total power consumption which can be formulated as follows

$$\begin{aligned} \min_{\mathbf{w},\mathbf{q}} \quad & \sum_{c=1}^{C} \sum_{b\in\mathcal{B}_c} \sum_{u\in\mathcal{U}_{cb}} |w_{cbu}|^2 + \sum_{c=1}^{C} \sum_{b\in\mathcal{B}_c} q_{cb} \\ s.t. \quad & \text{Constraints (4.2), (4.3), (4.4), (4.5), (4.6), (4.7) and (4.8)}, \end{aligned} \tag{4.9}$$

where the optimization is over the beamforming vectors $\mathbf{w} = [\mathbf{w}_{cu}, \quad \forall (u,c) \in (\mathcal{U}_{bc} \times \mathcal{C})]^T \in \mathbb{C}^{U_{bc}B_cC}$, and the quantization noise vector $\mathbf{q} = [q_{cb}, \quad \forall (b,c) \in (\mathcal{B}_c \times \mathcal{C})]^T \in \mathbb{R}^{B_cC}$, and where δ_u denotes the SINR target for user u. The above optimization problem is of high complexity due to the non-convexity of the SINR constraints and to the fact that the CSI errors belong to an infinitely uncountable set (4.4). To efficiently derive a decentralized solution to the optimization problem (4.9), we adopt an ADMM strategy which is first proposed in optimization theory [10]. Since centralized solution requires joint signal-processing and high signaling overhead, the ADMM approach permits a distributed implementation of the algorithm by allowing a limited information exchange between the coupled clouds. Simulation results demonstrate that the proposed decentralized solution converges to the centralized solution in a reasonable number of iterations.

4.3 Robust Multi-Cloud Algorithm

The main objective of this section is to use the ADMM approach to provide a distributed algorithm to solve the network total power minimization problem (4.9). First, we start by relaxing the inter-cloud interference terms and introducing local auxiliary variables. However, the relaxed optimization problem is intractable due to the infinite set of CSI errors. In this respect, the SDP procedure [18] and the S-Procedure method [4] can be used to find an equivalent convex SDP optimization problem. Then, the ADMM algorithm [10] is applied to derive a distributed algorithm to solve the problem.

4.3.1 Problem Relaxation

By relaxing the equality in the inter-cloud interference terms expression into inequality, the network total power minimization problem can be reformulated as follows

$$\min_{\mathbf{w},\mathbf{q},\boldsymbol{\psi}} \quad \sum_{c=1}^{C} \sum_{b\in\mathcal{B}_c} \sum_{u\in\mathcal{U}_{cb}} |w_{cbu}|^2 + \sum_{c=1}^{C} \sum_{b\in\mathcal{B}_c} q_{cb}$$
$$s.t. \quad \text{Constraints (4.2), (4.3), (4.4), (4.5), (4.6), (4.7) and (4.8)} \tag{4.10}$$
$$\psi_{c'u}^2 \geq \sum_{u'\in\mathcal{U}_{c'}} \left|\mathbf{h}_{c'u}^H \mathbf{w}_{c'u'}\right|^2 + \mathbf{h}_{c'u}^H \mathbf{Q}_{c'} \mathbf{h}_{c'u},$$

where the optimization is over the beamforming vector $\mathbf{w}$, the quantization noise vector $\mathbf{q}$, and the inter-cloud interference vector $\boldsymbol{\psi} = [\psi_{c'u}, \quad \forall u \notin \mathcal{U}_{c'} \, \forall c' \in \mathcal{C}]^T \in \mathbb{R}_+^{U'}$ where $U' = \sum_{c=1}^{C} \sum_{c' \neq c} U_{c'}$.

Proposition 4. *The original optimization problem (4.9) and the relaxed optimization problem (4.10) have the same optimal solution.*

Proof. The proof is similar to the proof of proposition 2. □

4.3.2 Distributed Solution via ADMM

The optimization problem (4.10) is still of high complexity for the same reasons mentioned earlier. In order to reduce the complexity, we utilize an approach similar to the one used in [22]. Specifically, we start by defining the optimization variables $\mathbf{W}_{cu} = \mathbf{w}_{cu}\mathbf{w}_{cu}^H$ where $\text{rank}(\mathbf{W}_{cu}) = 1$. Then, the S-Procedure method [4] and a rank-one SDR approach [18] are used to obtain an equivalent SDP minimization problem.

SDP Reformulation

Based on the introduced quadratic variables $\mathbf{W}_{cu}$, the cost function of the optimization problem (4.10), the constraints (4.6) and (4.7) can respectively be rewritten as follows

$$\sum_{c=1}^{C} \sum_{u\in\mathcal{U}_c} \text{Tr}\left(\mathbf{W}_{cu}\right) + \sum_{c=1}^{C} \text{Tr}\left(\mathbf{Q}_c\right) \tag{4.11}$$

$$\sum_{u \in \mathcal{U}_c} \operatorname{Tr}\left(\mathbf{A}_{cbu}\mathbf{W}_{cu}\right) + \operatorname{Tr}\left(\mathbf{U}_{cb}\mathbf{Q}_c\right) \leq P_{cb} \tag{4.12}$$

$$\sum_{u \in \mathcal{U}_c} \operatorname{Tr}\left(\mathbf{A}_{cbu}\mathbf{W}_{cu}\right) \leq \left(2^{C_{cb}} - 1\right) \operatorname{Tr}\left(\mathbf{U}_{cb}\mathbf{Q}_c\right), \tag{4.13}$$

where $\mathbf{A}_{cbu}$ and $\mathbf{U}_{cbu}$ denote the diagonal matrix with 1 at the main diagonal entry b and zeros otherwise. After introducing the local inter-cloud interference terms $\xi_{c'u}^{(c)}$ and additional equality constraints, the rank-one relaxation of the network total power minimization problem (4.10) can be formulated as follows

$$\begin{aligned}
&\min_{\substack{\mathbf{W}_{cu},\mathbf{Q}_c,\boldsymbol{\psi} \\ \tilde{\boldsymbol{\xi}},\lambda_{cu},\beta_{c'u}}} \sum_{c=1}^{C} \sum_{u \in \mathcal{U}_c} \operatorname{Tr}\left(\mathbf{W}_{cu}\right) + \sum_{c=1}^{C} \operatorname{Tr}\left(\mathbf{Q}_c\right) \\
&s.t. \;\; \text{Constraints (4.12) and (4.13)} \\
&\qquad \boldsymbol{\Delta}_{cu} \succeq 0,\; \lambda_{cu} \geq 0,\; \forall\, (u,c) \in (\mathcal{U}_c \times \mathcal{C}) \\
&\qquad \boldsymbol{\Theta}_{c'u} \succeq 0,\; \beta_{c'u} \geq 0,\; \forall u \notin \mathcal{U}_{c'},\; \forall c' \in \mathcal{C} \\
&\qquad \xi_{c'u}^{(c')} = \psi_{c'u},\; \forall\, (u,c') \in (\mathcal{U} \times \bar{\mathcal{C}}_u) \\
&\qquad \xi_{c'u}^{(c_u)} = \psi_{c'u},\; \forall\, (u,c') \in (\mathcal{U} \times \bar{\mathcal{C}}_u) \\
&\qquad \mathbf{W}_{cu} \succeq 0, \mathbf{Q}_c \succeq 0, \mathbf{Q}_c \text{ is diagonal}, \forall\, (c,u) \in (\mathcal{C} \times \mathcal{U}_c),
\end{aligned} \tag{4.14}$$

where the optimization is over the beamforming matrices $\mathbf{W}_{cu}$, the quantization noise covariance matrices $\mathbf{Q}_c$, the global inter-cloud interference vector $\boldsymbol{\psi}$, the local inter-cloud interference vector $\tilde{\boldsymbol{\xi}} = [\xi_{c'u}^{(c)}\ \forall c \in \{c', c_u\}\ \forall u \notin \mathcal{U}_{c'}\ \forall c' \in \mathcal{C}]^T \in \mathbb{R}_+^{2U'}$, and the introduced variables λ_{cu} and $\beta_{c'u}$, and where matrices $\boldsymbol{\Delta}_{cu}$ and $\boldsymbol{\Theta}_{c'u}$ are denoted by

$$\boldsymbol{\Delta}_{cu} = \begin{bmatrix} \mathbf{G}_{cu}^{1} + \lambda_{cu}\mathbf{E}_{cu} & \mathbf{G}_{cu}^{1}\tilde{\mathbf{h}}_{cu} \\ \tilde{\mathbf{h}}_{cu}^{H}\mathbf{G}_{cu}^{1} & \tilde{\mathbf{h}}_{cu}^{H}\mathbf{G}_{cu}^{1}\tilde{\mathbf{h}}_{cu} - \sum_{c' \neq c} \xi_{c'u}^{(c)} - \sigma_u^2 - \lambda_{cu} \end{bmatrix}, \tag{4.15}$$

$$\mathcal{F} = \Bigg\{ \mathbf{W}_{cu}, \mathbf{Q}_c, \tilde{\boldsymbol{\xi}}, \boldsymbol{\lambda}, \boldsymbol{\beta} \Bigg| \sum_{u' \in \mathcal{U}_c} \mathrm{Tr}\left(\mathbf{A}_{cbu'} \mathbf{W}_{cu'}\right) + \mathrm{Tr}\left(\mathbf{U}_{cb} \mathbf{Q}_c\right) \leq P_{cb}, \sum_{u' \in \mathcal{U}_c} \mathrm{Tr}\left(\mathbf{A}_{cbu'} \mathbf{W}_{cu'}\right) - (2^{C_{cb}} - 1)\mathrm{Tr}\left(\mathbf{U}_{cb} \mathbf{Q}_c\right) \leq 0, \boldsymbol{\Delta}_{cu} \succeq 0, \lambda_{cu} \geq 0; \mathbf{W}_{cu} \succeq 0, \mathbf{Q}_c \succeq 0, \mathbf{Q}_c \text{ is diagonal}, \forall (u, b, c) \in (\mathcal{U}_{cb} \times \mathcal{B}_c \times \mathcal{C}); \boldsymbol{\Theta}_{c'u} \succeq 0, \beta_{c'u} \geq 0, \forall u \notin \mathcal{U}_{c'},\ \forall c' \in \mathcal{C} \Bigg\}. \tag{4.18}$$

---

and

$$\boldsymbol{\Theta}_{c'u} = \begin{bmatrix} -\mathbf{G}^2_{c'u} + \beta_{c'u}\mathbf{E}_{c'u} & -\mathbf{G}^2_{c'u}\tilde{\mathbf{h}}_{c'u} \\ -\tilde{\mathbf{h}}^H_{c'u}\mathbf{G}^2_{c'u} & -\tilde{\mathbf{h}}^H_{c'u}\mathbf{G}^2_{c'u}\tilde{\mathbf{h}}_{c'u} + \xi^{(c')}_{c'u} - \beta_{c'u} \end{bmatrix}, \tag{4.16}$$

and where

$$\begin{cases} \mathbf{G}^1_{cu} = \frac{1}{\delta_u}\mathbf{W}_{cu} - \sum\limits_{u' \in \mathcal{U}_{cu}} \mathbf{W}_{cu'} - \mathbf{Q}_c \\ \mathbf{G}^2_{c'u} = \sum\limits_{u' \in \mathcal{U}_{c'}} \mathbf{W}_{c'u} + \mathbf{Q}_{c'} \end{cases}, \tag{4.17}$$

and where $\delta_u = 2^{R_u} - 1$. This chapter focuses on finding a distributed solution to problem (4.14) which is a rank-one relaxed problem to the optimization problem (4.10). In practice, additional approximation methods are needed to find a rank-one solution whenever needed. Further, the known approximation methods, such as the Gaussian randomization [4], can not be implemented in a distributed fashion. However, as shown in the simulations, the solution reached from solving (4.14) is rank-one in most of the cases.

ADMM-Based Distributed Solution

To derive a distributed solution to the relaxed problem (4.14) using the ADMM approach, we start by defining the following indicator function

$$I(\mathbf{W}_{cu}, \mathbf{Q}_c, \tilde{\boldsymbol{\xi}}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \begin{cases} 0 & \text{if } \{\mathbf{W}_{cu}, \mathbf{Q}_c, \tilde{\boldsymbol{\xi}}, \boldsymbol{\lambda}, \boldsymbol{\beta}\} \in \mathcal{F} \\ +\infty & \text{Otherwise} \end{cases}, \tag{4.19}$$

where $\mathcal{F}$ denotes the feasibility set given in (4.18), $\boldsymbol{\lambda} = [\lambda_{cu},\ \forall\ (u,c) \in (\mathcal{U}_c \times \mathcal{C})]^T \in \mathbb{R}^{U_c C}$ and where $\boldsymbol{\beta} = [\beta_{c'u},\ \forall u \notin \mathcal{U}_{c'},\ \forall c' \in \mathcal{C}]^T \in \mathbb{R}^{U'}$. Therefore, the augmented Lagrangian of the optimization problem (4.14) can be written as follows

$$\begin{aligned} \mathcal{L}_\rho(\mathbf{W}_{cu}, \mathbf{Q}_c, \tilde{\boldsymbol{\xi}}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\nu}) = \sum_{c=1}^{C} \Bigg\{ \sum_{u \in \mathcal{U}_c} \mathrm{Tr}\left(\mathbf{W}_{cu}\right) + \mathrm{Tr}\left(\mathbf{Q}_c\right) \\ + \frac{\rho}{2} \left|\left| \boldsymbol{\xi}^{(c)} - \boldsymbol{\psi}^{(c)} + \frac{1}{\rho}\boldsymbol{\nu}^{(c)} \right|\right|_{\ell_2}^2 \Bigg\} + I(\mathbf{W}_{cu}, \mathbf{Q}_c, \tilde{\boldsymbol{\xi}}, \boldsymbol{\lambda}, \boldsymbol{\beta}), \end{aligned} \tag{4.20}$$

where the vectors $\boldsymbol{\nu}^{(c)} = [\lambda^{(c)}_{cu'};\ \forall u' \notin \mathcal{U}_c,\ \lambda^{(c)}_{c'u};\ \forall c' \neq c\, \forall u \in \mathcal{U}_c]^T$ and $\boldsymbol{\xi}^{(c)} = [\xi^{(c)}_{cu'};\ \forall u' \notin \mathcal{U}_c,\ \xi^{(c)}_{c'u};\ \forall c' \neq c\, \forall u \in \mathcal{U}_c]^T$ denote the local dual vector and the local vector associated with the inter-cloud interference vector $\boldsymbol{\psi}^{(c)} = [\psi_{cu'};\ \forall u' \notin \mathcal{U}_c,\ \psi_{c'u};\ \forall c' \neq c\, \forall u \in \mathcal{U}_b]^T$.

The ADMM algorithm provides a distributed solution to the network total power minimization problem by performing two primal minimization steps and a dual update in each iteration. Specifically, the first step of the ADMM algorithm fixes the vectors $\boldsymbol{\psi}$ and $\boldsymbol{\nu}$ and minimizes the augmented Lagrangian (4.20) over the primal variables $\mathbf{W}_{cu}$, $\mathbf{Q}_c$, $\tilde{\boldsymbol{\xi}}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$. Then, the second step of the ADMM algorithm minimizes the augmented Lagrangian (4.20) over the primal vector $\boldsymbol{\psi}$ while the remaining variables are fixed. The final step of the ADMM algorithm consists of a simple dual update. Next, we show that the first step is fully separable and it can be implemented in a distributed fashion. However, the second and third steps can be implemented in a distributed fashion only by allowing a limited information exchange between the coupled clouds.

First Step Optimization: In the first step, the ADMM approach minimizes the augmented Lagrangian over the primal variables $\mathbf{W}_{cu}$, $\mathbf{Q}_c$,$\tilde{\boldsymbol{\xi}}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$. Specifically, the first step can be formulated as follows

$$\begin{aligned} \min_{\substack{\mathbf{W}_{cu},\mathbf{Q}_c,\tilde{\boldsymbol{\xi}} \\ \boldsymbol{\lambda},\boldsymbol{\beta}}} \quad & \sum_{c=1}^{C} \left\{ \sum_{u\in\mathcal{U}_c} \operatorname{Tr}\left(\mathbf{W}_{cu}\right) + \operatorname{Tr}\left(\mathbf{Q}_c\right) + \frac{\rho}{2} \left|\left| \boldsymbol{\xi}^{(c)} - \boldsymbol{\psi}^{(c)^k} + \frac{1}{\rho}\boldsymbol{\nu}^{(c)^k} \right|\right|_{\ell_2}^2 \right\} \\ s.t. \quad & \{\mathbf{W}_{cu}, \mathbf{Q}_c, \tilde{\boldsymbol{\xi}}, \boldsymbol{\lambda}, \boldsymbol{\beta}\} \in \mathcal{F}. \end{aligned} \tag{4.21}$$

It can be noticed that the above optimization problem is completely separable. Specifically, the following optimization problem is solved independently at each cloud c

$$\begin{aligned} \min_{\substack{\mathbf{W}_{cu},\mathbf{Q}_c,\boldsymbol{\xi}^{(c)} \\ \boldsymbol{\lambda}^{(c)},\boldsymbol{\beta}^{(c)}}} \quad & \left\{ \sum_{u\in\mathcal{U}_c} \operatorname{Tr}\left(\mathbf{W}_{cu}\right) + \operatorname{Tr}\left(\mathbf{Q}_c\right) + \frac{\rho}{2} \left|\left| \boldsymbol{\xi}^{(c)} - \boldsymbol{\psi}^{(c)^k} + \frac{1}{\rho}\boldsymbol{\nu}^{(c)^k} \right|\right|_{\ell_2}^2 \right\} \\ s.t. \quad & \sum_{u\in\mathcal{U}_c} \operatorname{Tr}\left(\mathbf{A}_{cbu}\mathbf{W}_{cu}\right) + \operatorname{Tr}\left(\mathbf{U}_{cb}\mathbf{Q}_c\right) \leq P_{cb} \\ & \sum_{u\in\mathcal{U}_c} \operatorname{Tr}\left(\mathbf{A}_{cbu}\mathbf{W}_{cu}\right) \leq \left(2^{C_{cb}} - 1\right) \operatorname{Tr}\left(\mathbf{U}_{cb}\mathbf{Q}_c\right) \\ & \boldsymbol{\Delta}_{cu} \succeq 0,\ \lambda_{cu} \geq 0,\ \forall\ u \in \mathcal{U}_c \\ & \boldsymbol{\Theta}_{cu'} \succeq 0,\ \beta_{cu'} \geq 0,\ \forall u' \notin \mathcal{U}_c \\ & \mathbf{W}_{cu} \succeq 0, \mathbf{Q}_c \succeq 0, \mathbf{Q}_c \text{ is diagonal}, \forall\ u \in \mathcal{U}_c, \end{aligned} \tag{4.22}$$

where the optimization is over the local beamforming matrices $\mathbf{W}_{cu}$, and the local quantization noise covariance matrix $\mathbf{Q}_c$, the local inter-cloud interference vector $\boldsymbol{\xi}^{(c)}$, and over the local vectors $\boldsymbol{\beta}^{(c)} = [\beta_{cu'},\ \forall\ u' \notin \mathcal{U}_c]^T$ and $\boldsymbol{\lambda}^{(c)} = [\lambda_{cu},\ \forall\ u \in \mathcal{U}_c]^T$. The above optimization problem is an SDP problem. Thus, it can be solved using efficient numerical algorithms [4].

Second Step Optimization: In the second step, the ADMM approach minimizes the augmented Lagrangian over the primal variables $\boldsymbol{\psi}$. Specifically, the second step can be formulated as follows

$$\psi_{cu}{}^{k+1} = \underset{\boldsymbol{\psi}}{\operatorname{argmin}} \sum_{c=1}^{C} \sum_{u \notin \mathcal{U}_c} \left(\psi_{cu} - \xi_{cu}^{(c)}{}^{k+1} + \frac{1}{\rho} \nu_{cu}^{(c)}{}^{k} \right)^2 + \left(\psi_{cu} - \xi_{cu}^{(c_u)}{}^{k+1} + \frac{1}{\rho} \nu_{cu}^{(c_u)}{}^{k} \right)^2. \quad (4.23)$$

The optimization problem (4.23) is convex and separable on the variables ψ_{cu}. Setting the gradient of the cost function in (4.23) with respect to $\boldsymbol{\psi}$ to zero, the inter-cloud interference terms ψ_{cu} $\forall c \in \mathcal{C}$ $\forall u \notin \mathcal{U}_c$ are respectively updated as follows

$$\psi_{cu}{}^{k+1} = \frac{1}{2} \left(\xi_{cu}^{(c)}{}^{k+1} + \xi_{cu}^{(c_u)}{}^{k+1} \right) + \frac{1}{2\rho} \left(\nu_{cu}^{(c)}{}^{k} + \nu_{cu}^{(c_u)}{}^{k} \right). \quad (4.24)$$

The above update rule can be preformed in a distributed fashion by allowing the exchange of the local inter-cloud interference terms between the coupled clouds.

Dual variable update: The last step of the ADMM approach consists of respectively updating the dual variables $\nu_{cu}^{(c')}$ $\forall c' \in (c, c_u)$ $\forall u \notin \mathcal{U}_c$ $\forall c \in \mathcal{C}$ as follows

$$\begin{cases} \nu_{cu}^{(c)}{}^{k+1} = \nu_{cu}^{(c)}{}^{k} + \rho \left(\xi_{cu}^{(c)}{}^{k+1} - \psi_{cu}{}^{k+1} \right), \ \forall u \notin \mathcal{U}_c, \\ \nu_{c'u}^{(c)}{}^{k+1} = \nu_{c'u}^{(c)}{}^{k} + \rho \left(\xi_{c'u}^{(c)}{}^{k+1} - \psi_{c'u}{}^{k+1} \right), \ \forall c' \neq c, u \in \mathcal{U}_c \end{cases}, \quad (4.25)$$

It can be noticed that the above update rule can be preformed in a distributed fashion.

Iterative Distributed Algorithm: The proposed distributed solution consists of iterating between the first and second step optimization and the dual variable update. Note that the solution reached through our algorithm also determines the users scheduling by first assuming that all the users can be served by each BS, then, only selects the users with nonzero beamforming vector. Further, at each iteration of the ADMM-based distributed algorithm, the introduced local inter-cloud interference terms may not lead to a feasible solution since $\xi_{c'u}^{(c')} \neq \xi_{c'u}^{(c_u)}$. However, a feasible solution can be estimated by fixing the inter-cloud interference terms and solving the optimization problem (4.22) over the beamforming matrices $\mathbf{W}_{cu}$, the quantization noise covariance matrices $\mathbf{Q}_c$, and the vectors $\boldsymbol{\lambda}^{(c)}$ and $\boldsymbol{\beta}^{(c)}$. The proposed iterative Distributed Robust Multi-Cloud Beamforming algorithm is summarized in Table (2).

Algorithm 2 The Iterative Distributed Robust Multi-Cloud Beamforming algorithm

Initialization : Initialize the dual variables $\boldsymbol{\nu}_c(0)$ and set the iteration index to $i = 0$.

Repeat :

1: Solve the minimization problem (4.22). Then, transmit the local inter-cloud interference terms to the adjacent clouds. If it is infeasible, go to **End**;

2: Update the inter-cloud interference terms as in (4.24);

3: Update the dual variables as in (4.25);

4: *Optional: fix the inter-cloud interference terms, then, solve the optimization problem (4.22) to estimate a feasible primal solution. If it is infeasible, then another iteration is needed.*

5: Update the iteration index $i = i + 1$;

Until : Desired level of convergence.

6: If the optimal solution is rank-one: Find the optimal beamforming vectors $\mathbf{w}^*_{cbu}$, $\forall\, b \notin \mathcal{B}_c$, $\forall\, u \in \mathcal{U}_c$, the users scheduling and the optimal power consumption. Otherwise: use the randomization techniques to find rank-one solution.

End

4.3.3 Centralized Solution

The centralized approach consists of reformulating the optimization problem (4.9) as an SDP problem using the S-Procedure method and a rank-one semi-definite relaxation. Specifically, the centralized optimization problem can be formulated as follows

$$
\begin{aligned}
\min_{\substack{\mathbf{W}_{cu},\mathbf{Q}_c,\boldsymbol{\psi}\\ \lambda_{cu},\beta_{c'u}}} \quad & \sum_{c=1}^{C} \sum_{u\in\mathcal{U}_c} \operatorname{Tr}\left(\mathbf{W}_{cu}\right) + \sum_{c=1}^{C} \operatorname{Tr}\left(\mathbf{Q}_c\right) \\
s.t. \quad & \text{Constraints (4.12) and (4.13)} \\
& \boldsymbol{\Delta}_{cu} \succeq 0,\ \lambda_{cu} \geq 0,\ \forall\ (u,c) \in (\mathcal{U}_c \times \mathcal{C}) \\
& \boldsymbol{\Theta}_{c'u} \succeq 0,\ \beta_{c'u} \geq 0,\ \forall u \notin \mathcal{U}_{c'},\ \forall c' \in \mathcal{C} \\
& \mathbf{W}_{cu} \succeq 0, \mathbf{Q}_c \succeq 0, \mathbf{Q}_c \text{ is diagonal}, \forall\ (c,u) \in (\mathcal{C} \times \mathcal{U}_c),
\end{aligned}
\tag{4.26}
$$

where the matrices $\boldsymbol{\Delta}_{cu}$ and $\boldsymbol{\Theta}_{c'u}$ are given by

$$\boldsymbol{\Delta}_{cu} = \begin{bmatrix} \mathbf{G}_{cu}^{1} + \lambda_{cu}\mathbf{E}_{cu} & \mathbf{G}_{cu}^{1}\tilde{\mathbf{h}}_{cu} \\ \tilde{\mathbf{h}}_{cu}^{H}\mathbf{G}_{cu}^{1} & \tilde{\mathbf{h}}_{cu}^{H}\mathbf{G}_{cu}^{1}\tilde{\mathbf{h}}_{cu} - \sum_{c' \neq c} \psi_{c'u} - \sigma_u^2 - \lambda_{cu} \end{bmatrix}, \tag{4.27}$$

and

$$\boldsymbol{\Theta}_{c'u} = \begin{bmatrix} -\mathbf{G}_{c'u}^{2} + \beta_{c'u}\mathbf{E}_{c'u} & -\mathbf{G}_{c'u}^{2}\tilde{\mathbf{h}}_{c'u} \\ -\tilde{\mathbf{h}}_{c'u}^{H}\mathbf{G}_{c'u}^{2} & -\tilde{\mathbf{h}}_{c'u}^{H}\mathbf{G}_{c'u}^{2}\tilde{\mathbf{h}}_{c'u} + \psi_{c'u} - \beta_{c'u} \end{bmatrix}. \tag{4.28}$$

This optimization problem is an SDP problem, hence, it can be solved using the interior point method [4]. The optimal solution of the optimization problem does not always satisfy the rank-one constraints. In practice, additional approximation methods, such as the Gaussian randomization, are needed to find a rank-one solution whenever needed.

4.3.4 Complexity Analysis

The implementation of the proposed ADMM-based distributed algorithm requires first to solve the convex optimization problem (4.22) at each cloud with $u = (2B_c + 2U + U_c + 2)$ constraints and $v = (2U + (C-1)U_c + 1)$ variables and then it updates the inter-cloud interference terms and the dual variables at each iteration. Therefore, the computation complexity of the ADMM-based distributed algorithm comes mainly from solving the optimization problem (4.22). On the other hand, the centralized solution directly solves the SDP problem with $(2B + U + 2C + 2CU)$ SDP constraints and $v = (2CU + C)$ SDP variables. Therefore, the ADMM-based distributed algorithm leads to larger complexity compared to the centralized solution in small-sized network since the convex problems solved at each iteration of the distributed algorithm and the SDP problem solved by the centralized algorithm have comparable complexity. However, centralized solution is not practical, since the clouds would require joint signal-processing and a high level of cloud-to-cloud communication. In large-scale network, the ADMM-based distributed algorithm may lead to lower computational complexity as compared to the centralized algorithm especially for large number of coupled clouds C since it solves the centralized problem in a distributed fashion which significantly reduces the complexity of the SDP problems solved at each cloud.

4.4 Simulation Results

This section illustrates the performance of the proposed distributed solution. To this end, consider a multi-CRAN scenario formed by two clouds, where each cloud is connected to six single-antenna BSs. Further, each cloud can serve four single-antenna MUs. Assume that the BSs and MUs in both clouds are uniformily and independently distributed in the square region [2000 4000] $\times$ [2000 4000] meters and [0 2000] $\times$ [0 2000] meters, respectively. Further, assume that the estimated channel vectors from the BSs to the MUs are generated using a distance-dependent path loss $L(d_{bu}) = 128.1 + 37.6\log_{10}(d_{bu})$, and Rayleigh fading component. The noise power spectral density is set to $\sigma_u^2 = -98$ dBm $\forall u$. The initial dual variable is set to $\nu_{cu}(0) = 0.01, \ \forall \ u \in \mathcal{U}, \ \forall \ c \in \bar{\mathcal{C}}_u$. We set the maximum transmit power of BS b in cloud c to $P_{cb} = 1W$ and the backhaul capacity limit between BS b and cloud c to $C_{cb} = 40$ Mbps. For illustration purposes, we also set $\mathbf{E}_{cu} = \frac{1}{a}\mathbf{I}_{B_c}$ where $a > 0$. In all below figures the solutions were rank-one.

In the first example, we consider different SINR targets and we set $a = 0.01$ and $\rho = 4$. Figure 4.2 shows the convergence behavior of the proposed distributed algorithm. It can be noticed that the ADMM-based distributed algorithm converges to the centralized algorithm in a reasonable number of iterations for all the considered SINR values.

Then, we compare the convergence behavior of the proposed distributed algorithm

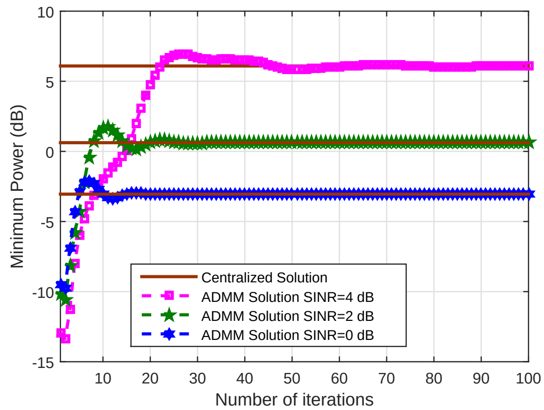

Figure 4.2: Convergence behavior of the ADMM-based and the conventional centralized algorithms with SINR target=5 dB.

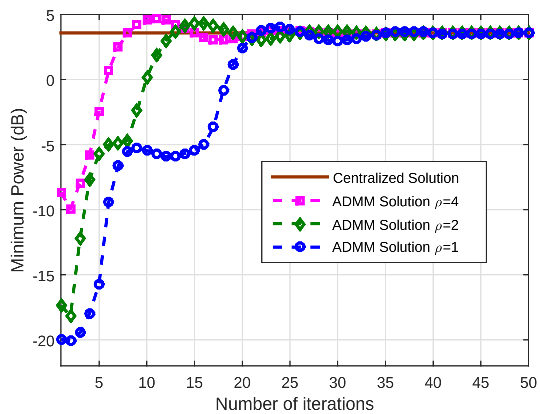

Figure 4.3: Minimum power versus the iteration number for different penalty parameters with SINR target=6 dB.

for different penalty parameter ρ and for target SINR= 6 dB. Figure 4.3 plots the optimal minimum power consumption versus the number of iterations. It can be noticed that the convergence speed of the ADMM-based distributed algorithm strongly depends on the selected penalty parameter. Therefore, a properly chosen penalty parameter leads to faster convergence speed of the proposed distributed algorithm.

Table 4.1: Probability that the optimal solution is rank-one.

a	0.001	0.005	0.01	0.05	0.1
$\delta_u = 0$ dB	0.993	0.973	0.951	0.797	0.635
$\delta_u = 4$ dB	0.962	0.940	0.919	0.797	0.679

Table 4.1 presents the probability that the optimal solution found using the centralized and distributed solution is rank-one for 1000 channel realizations and for

various SINR target and scalar a. It can be noticed that reducing the scalar a improves the probability that the optimal solution is rank-one. It is greater than 0.9 when $a \leq 0.01$. In the case when the rank is not 1, additional randomization techniques are needed in order to retrieve the rank-1 solution. But this comes at the expense of the distributed nature of our proposed algorithm. Therefore, the CSI errors should be small in order to guarantee a rank-one solution to the considered optimization problem in a distributed fashion.

4.5 Conclusion

Sophisticated resource allocation optimization techniques are expected to be at the heart of research for next generation mobile radio systems. This chapter considers the downlink of a multi-CRAN, where each cloud serves a pre-known set of single antenna mobile users, each served by several BSs. The chapter assumes that each BS is connected to one cloud via limited-capacity backhaul link and that only imperfect CSI is available at each cloud. The chapter formulates the network total power minimization problem subject to the above practical constraints. It solves the problem using an ADMM-based algorithm that can be implemented in a distributed fashion across the network by allowing a limited information exchange between the coupled clouds. Simulation results show that the ADMM-based algorithm provides the same performance as the centralized algorithm in a reasonable number of iterations.

Chapter 5

Robust Power Minimization in Cache-Enabled CRAN

In dense data networks, the backhaul congestion may become a limiting factor of the CRAN performance. To truly access the advantages harvested by CRAN, this chapter[1] considers the downlink of a cache-enabled CRAN where each cache-enabled BS is connected to the cloud via limited capacity backhaul link and it can serve a set of pre-known single-antenna MUs. This chapter focuses on minimizing the network total power consumption and the backhaul cost. The chapter then suggests solving this optimization problem using the MM approach. Simulation results show that the proposed algorithm converges in a reasonable number of iterations.

5.1 Introduction

This chapter considers a cache-enabled cloud where the popular content requested by the users across the network is stored in the local limited-size memory of the BSs. Further, this chapter considers the more realistic scenario where the cloud is connected to the BSs via limited-capacity backhaul links and only imperfect CSI is

[1] A part of this chapter will appear in: O. Dhifallah, H. Dahrouj, T. Y.Al-Naffouri, M-S. Alouini, "Robust Beamforming in Cache-Enabled Cloud Radio Access Networks", *IEEE wireless Communications Letters* , (Under Preparation).

available at the cloud. Each BS across the network can serve a set of pre-known single-antenna MUs. Further, this chapter assumes that independent compression procedure is applied to the user signals. Therefore, the performance of the considered system becomes a function of both the compression scheme, the beamforming vector of each user across the network.

If the content requested by a MU is available in the local-cache of the BS, it directly transmits the data without backhaul cost. However, if the requested data is not available in the local-cache, the BS fetches the data from the cloud which performs joint precoding and compression of the requested signal before forwarding it to the corresponding BS. This operation introduces quantization noise to the user signal. Further, the requested data is not quantized when it is available in the local-cache of the BS. The chapter formulates the problem of minimizing the network total power cost and the backhaul cost subject to per-BS power constraint, the quality of service constraints, the total backhaul capacity constraint and the imperfect CSI constraints in order to design the beamforming vector of each user across the network, the quantization noise covariance matrix of the limited-capacity backhaul links and the BS clustering. This chapter uses the SDP procedure [18], the S-Procedure method [4], and the MM technique [11] to solve the considered optimization problem.

5.1.1 Related Work

The optimization problem considered in this chapter is partly related to reference [23] where the authors assume a multicast cache-enabled CRAN and formulates the total network cost minimization problem in order to design the multicast beamforming vectors and BS clustering. However, reference [23] assumes perfect CSI available at the cloud and neglects the effect of the compression of user signals. The optimization problem formulated in this chapter is related to the work in [24] which assumes a conventional CRAN and focuses on solving the network total power minimization problem for both the data-sharing strategy and compression strategy. This chapter is also related to the work in [13] which focuses on solving the utility maximization problem for both the dynamic and static BS clustering. Additionally, the chapter is related to the work in [20] where the authors assume hybrid connections between the cloud and the BSs and focuses on minimizing the network total power minimization to design the beamforming vector of each user across the network, the quantization noise covariance matrix of the wireline backhaul links, the transmit power of the wireless backhaul links and. However, the works in [24], [13] and [20] assume a conventional CRAN where all the BSs fetch the requested data from the cloud and perfect CSI is available at the cloud.

5.1.2 Organization

The rest of this chapter is organized as follows, Section 5.2 presents the system model and formulates the total network cost minimization problem. Then, Section 5.3 presents the proposed iterative algorithm. Numerical examples illustrating the performance of the proposed algorithm are given in Section 5.4. Finally, Section 5.5 concludes this chapter.

5.2 System Model and Problem Formulation

5.2.1 System Model

Consider the downlink of a cache-enabled CRAN, where the cloud is connected to B BSs through wireline limited-capacity backhaul links. Each BS can serve U single-antenna MUs. Further, we assume that the cloud has access to the whole data files and each BS is equipped with a local-cache with finite storage. For the simplicity of analysis, we assume that each BS is equipped with single-antenna, all the files have the same size and the knowledge of cache and request status are known for the cloud.

Figure 5.1 illustrates an example of the considered network with 3 cache-enabled BSs and 6 MUs.

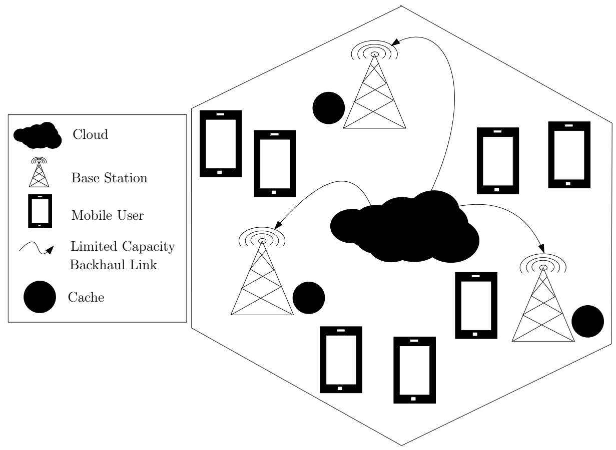

Figure 5.1: The considered CRAN, in which, each cache-enable BS is connected to the cloud through wireline capacity-limited backhaul link.

Let $\mathcal{B} = \{1, .., B\}$ be the set of BSs connected to the cloud through wireline limited-capacity backhaul links. Further, let $\mathcal{U} = \{1, .., U\}$ be the set of users across the network. Let the beamforming vector from all the BSs to user $u \in \mathcal{U}$ be $\mathbf{w}_u = [w_{1u},\ w_{2u}, ..,\ w_{Bu}]^T \in \mathbb{C}^B$ and let the channel vector from all the BSs to user u be $\mathbf{h}_u = [h_{1u},\ h_{2u}, ..,\ h_{Bu}]^T \in \mathbb{C}^B$. Further, let the quantization noise vector be $\mathbf{v} = [v_1,\ v_2, \cdots,\ v_B] \in \mathbb{C}^B$. We assume that $\mathbf{v}$ is non-uniform white Gaussian with diagonal covariance matrix $\mathbf{Q} \in \mathbb{C}^{B\times B}$ where $q_b \geq 0,\ \forall b \in \mathcal{B}$ denotes the diagonal entries. We assume that each cache-enabled BS has a local storage with size F_b. Define the cache placement matrix $\mathbf{P} \in \mathbb{R}^{B\times F}$, where $p_{bf} = 1$ when the content f is cached in BS b and $p_{bf} = 0$ otherwise,i.e. $\sum_{f=1}^{F} p_{bf} < F_b,\ \forall b \in \mathcal{B}$. Therefore, the received signal $y_u \in \mathbb{C}$ at user u can be written as

$$y_u = \mathbf{h}_u^H \mathbf{w}_u s_u + \sum_{u' \in \mathcal{U}_u} \mathbf{h}_u^H \mathbf{w}_{u'} s_{u'} + \sum_{b\in\mathcal{B}} \alpha_b h_{bu}^* v_b + n_u, \tag{5.1}$$

where $\mathcal{U}_u = \mathcal{U}\backslash\{u\}$, s_u denotes the transmitted data symbol for user u, $n_u \sim \mathcal{CN}(0, \sigma_u^2)$ denotes the additive white Gaussian noise which is independent from the transmitted data symbols s_u and the quantization noise v_b, and where $\alpha_b = 0$ when the contents requested from BS b by all the users is cached in BS b and $\alpha_b = 1$ otherwise.

5.2.2 Backhaul Cost and Power Model

When the content f_u requested by user u from BS b is available in the local cache of BS b, the BS transmit the data directly without backhaul cost. However, when the content f_u is not in the cache of BS b, it needs to retrieve it from the cloud. In this case, the backhaul cost is the transmission rate of user u. Therefore, the backhaul cost of BS b serving user u can be expressed as follows

$$B_{bu} = \begin{cases} \log_2(1+\delta_u) & \text{if } p_{bf_u} = 0 \text{ and } |w_{bu}|^2 > 0 \\ 0 & \text{if } p_{bf_u} = 1 \text{ and } |w_{bu}|^2 = 0 \end{cases}. \tag{5.2}$$

In this chapter, we consider the following power consumption model of BS b

$$P_b = \begin{cases} \frac{1}{\nu_b} P_{b,t} + P_{b,a} + p_b P_{b,c} + (U - p_b) P_{b,cl} & \text{if } P_{b,t} > 0 \\ P_{b,s} & \text{if } P_{b,t} = 0 \end{cases}, \tag{5.3}$$

where $P_{b,t} = \left(\sum_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b\right)$ denotes the total transmit power consumption, ν_b denotes the power amplifier efficiency and $p_b = \sum_{u\in\mathcal{U}} p_{bf_u}$. Further, $P_{b,a}$ denotes the power consumed by BS b in the active mode, $P_{b,c}$ denotes the power consumed by BS b in the sleep mode, $P_{b,c}$ denotes the maximum power needed by BS b to retrieve the requested data from its local cache and $P_{b,c}$ denotes the maximum power needed by BS b to retrieve the requested data from the cloud. Therefore, the total power consumption can be rewritten as follows

$$P_t = \sum_{b\in\mathcal{B}} \left\{ \frac{1}{\nu_b} \left(\sum_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b \right) + \left|\left| P_{b,t} \right|\right|_0 P_{rb} + P_{b,s} \right\}, \tag{5.4}$$

where $P_{rb} = P_{b,t} + P_{b,a} + p_b P_{b,c} + (U - p_b) P_{b,cl} - P_{b,s}$ denotes the relative power consumption.

5.2.3 Problem Formulation

This chapter focuses on minimizing the total network cost, which consists of the total power consumption and the backhaul consumption, while satisfying the per-BS power consumption, the total backhaul capacity constraint and quality of service constraints, as well as CSI error constraints. In this chapter, we assume that imperfect channel state information is available at the cloud. Further, the channel errors are assumed to be bounded by an elliptical region. Therefore, the true channel vector of user u can be written as follows

$$\mathbf{h}_u = \tilde{\mathbf{h}}_u + \mathbf{e}_u, \ \forall \ u \in \mathcal{U}, \tag{5.5}$$

where $\mathbf{e}_u$ denotes the CSI error vector of user u and it satisfies the following elliptical constraints

$$\mathbf{e}_u^H \mathbf{E}_u \mathbf{e}_u < 1, \tag{5.6}$$

where $\mathbf{E}_u$ is a known positive definite matrix and it specifies the accuracy of the CSI. We assume that the user symbols have unit power, i.e. $E(|s_u|^2) = 1$, $\forall \ u \in \mathcal{U}$, and independent from each other, from the quantization noise and from the additive noise. In this case, the SINR of user u can be written as follows

$$\Gamma_u = \frac{\left|\mathbf{h}_u^H \mathbf{w}_u\right|^2}{\sum\limits_{u' \in \mathcal{U}_u} |\mathbf{h}_u^H \mathbf{w}_{u'}|^2 + \sum\limits_{b \in \mathcal{B}} \alpha_b |h_{bu}|^2 q_b + \sigma_u^2}. \tag{5.7}$$

This chapter considers the problem of minimizing the total network cost which consists of the total power consumption (5.4) and the backhaul cost (5.2), i.e.

$$\begin{aligned} C_N = \sum_{b\in\mathcal{B}} \Bigg\{ \frac{1}{\nu_b}\left(\sum_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b\right) + \left\| \sum_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b \right\|_0 P_{rb} \\ + \sum_{u\in\mathcal{U}} \left\| |w_{bu}|^2 \right\|_0 R_u \left(1 - p_{bf_u}\right) \Bigg\}, \end{aligned} \tag{5.8}$$

where $R_u = \log_2(1+\delta_u)$. One of the constraints studied in this chapter is the per-BS power constraint which is given by

$$\sum_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b \leq P_b, \ \forall\, b \in \mathcal{B}. \tag{5.9}$$

If the requested data from a MU is not available in the local-cache of the BS, it fetches the data from the cloud through the limited-capacity backhaul link. Since the backhaul link has limited capacity, the cloud needs to compress the requested data before forwarding it to the corresponding BS. In this chapter, it is assumed that user signals are compressed independently. Further, if the requested data from all the users is available in the base station local-cache, we assume that the transmitted user signals are not quantized. Therefore, the beamforming vector associated with user u, the quantization noise level q_b and the total backhaul capacity C are related as follows

$$\sum_{b\in\mathcal{B}} \left\| \sum_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b \right\|_0 \tilde{R}_b \leq C, \tag{5.10}$$

where $\tilde{R}_b, \ \forall\, b \in \mathcal{B}$ is given by

$$\tilde{R}_b = \log_2\left(1 + \frac{\sum\limits_{u\in\mathcal{U}} (1 - p_{bf_u})|w_{bu}|^2}{q_b}\right). \tag{5.11}$$

The optimization problem studied in this chapter is the minimization of the network total cost which can be formulated as follows

$$\begin{aligned} &\min_{\mathbf{w},\mathbf{q}} \quad C_N \\ &s.t. \;\; \text{Constraints (5.5), (5.6), (5.9), and (5.10)} \\ &\tilde{\Gamma}_u = \frac{\left|\mathbf{h}_u^H \mathbf{w}_u\right|^2}{\sum\limits_{u' \in \mathcal{U}_u} \left|\mathbf{h}_u^H \mathbf{w}_{u'}\right|^2 + \mathbf{h}_u^H \mathbf{Q} \mathbf{h}_u + \sigma_u^2} \geq \delta_u, \; \forall \, u \in \mathcal{U} \\ &q_b \geq 0, \; \forall b \in \mathcal{B}, \end{aligned} \tag{5.12}$$

where the optimization is over the beamforming vectors $\mathbf{w} = [\mathbf{w}_u, \; \forall u \in \mathcal{U}]^T \in \mathbb{C}^U$, and the quantization noise vector $\mathbf{q} = [q_b, \; \forall b \in \mathcal{B}]^T \in \mathbb{R}^B$.

Remark 1. *If the contents requested from BS b by all the users across the network are available in the local cache of BS b, the corresponding quantization noise q_b is only in the SINR expression of the optimization problem (5.12). Therefore, it is set to zero in order to increase the SINR value and satisfy the SINR constraints. In this case, the SINR expressions $\tilde{\Gamma}_u$ in (5.12) are equivalent to the expressions in (5.7).*

The optimization problem (5.12) is not convex due to the ℓ_0-norm in the cost function (5.12) and the total backhaul capacity constraint (5.10). Further, the CSI errors belong to an infinite set (5.6). Thus, the considered optimization problem (5.12) is NP-hard. To efficiently derive an approximate solution to the optimization problem (5.12), we start by relaxing the total backhaul capacity constraint (5.10). Then, we approximate the ℓ_0-norm in the cost function of the optimization problem (5.12). Finally, we use the MM algorithm to guarantee a stationary point to the obtained optimization problem.

5.3 Proposed Solution

5.3.1 Problem Relaxation

The optimization problem (5.12) is difficult to deal with due to the non-convexity of the cost function, the total backhaul capacity constraint (5.10) and the infinite set of possible CSI errors. In this subsection, we relax the total backhaul capacity constraint to reduce the complexity. Specifically, we start by approximating the ℓ_0-norm using the following fact, for $x > 0$

$$||x||_0 = I(x > 0) = \lim_{\epsilon \to 0} \frac{\log\left(1 + \epsilon^{-1} x\right)}{\log(1 + \epsilon^{-1})}. \tag{5.13}$$

Therefore, the total backhaul capacity constraint (5.10) can be approximated as follows

$$\lambda_\epsilon \sum_{b \in \mathcal{B}} \log\left(1 + \epsilon^{-1} \left\{ \sum_{u \in \mathcal{U}} |w_{bu}|^2 + \alpha_b q_b \right\}\right) \tilde{R}_b \leq C, \tag{5.14}$$

where $\lambda_\epsilon = \frac{1}{\log(1+\epsilon^{-1})}$ and $\epsilon > 0$ is a small constant. The approximated constraint is still non-convex. This chapter replaces the approximated left hand side of the constraint (5.14) by its upper-bound to convexifiy the original constraint (5.10).

Theorem 1. *An upper-bound of the function in the left hand side of the constraint (5.14) is given by*

$$C_M^m = \lambda_\epsilon \sum_{b \in \mathcal{B}} \tilde{R}_b \left\{ \frac{\sum\limits_{u \in \mathcal{U}} |w_{bu}|^2 + \alpha_b q_b}{\sum\limits_{u \in \mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1} + \epsilon} + a_b^{m-1} \right\}, \tag{5.15}$$

where $a_b^{m-1} = \log(1 + \epsilon^{-1}\{\sum_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1}\}) - \{\sum_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1}\}$ $/\{\sum_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1} + \epsilon\}$, *and where* w_{bu}^{m-1} *and* q_b^{m-1} *denote the beamforming vector from BS* b *to user* u *and quantization noise of BS* b *of the previous iteration.*

Proof. For $\epsilon > 0$, the function $x \to \log(1 + \epsilon^{-1} x)$ is a concave function on the interval

$[0 \ \ +\infty[$. Based on the first order propriety of concave functions, the following inequality holds for any $x \geq 0$ and $x^{m-1} \geq 0$

$$\log(1+\epsilon^{-1}x) \leq \frac{x}{\epsilon + x^{m-1}} + \log(1+\epsilon^{-1}x^{m-1}) - \frac{x^{m-1}}{\epsilon + x^{m-1}}. \tag{5.16}$$

Let w_{bu}^{m-1} and q_b^{m-1} denote the beamforming vector from BS b to user u and quantization noise of BS b of the previous iteration of the MM algorithm and let $x = \sum_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b$ and $x^{m-1} = \sum_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1}$. Substituting x and x^{m-1} in (5.16), the right hand side of (5.16) can be rewritten as follows

$$\begin{aligned} T_b = & \frac{\sum\limits_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b}{\epsilon + \sum\limits_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1}} + \log(1+\epsilon^{-1}\{\sum_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1}\}) \\ & - \frac{\sum\limits_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1}}{\epsilon + \sum\limits_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1}} = \frac{\sum\limits_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b}{\epsilon + \sum\limits_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1}} + a_b^{m-1}. \end{aligned} \tag{5.17}$$

The following inequality holds for any w_{bu} and $q_b \geq 0$

$$\begin{aligned} & \lambda_\epsilon \sum_{b\in\mathcal{B}} \log\left(1+\epsilon^{-1}\left\{\sum_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b\right\}\right) \tilde{R}_b \leq \\ & \lambda_\epsilon \sum_{b\in\mathcal{B}} \left\{\frac{\sum\limits_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b}{\epsilon + \sum\limits_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1}} + a_b^{m-1}\right\} \tilde{R}_b. \end{aligned} \tag{5.18}$$

Further, we can readily see that the above inequality is achieved with equality when $w_{bu} = w_{bu}^{m-1}$ and $q_b = q_b^{m-1}$, $\forall b \in \mathcal{B},\ u \in \mathcal{U}$. Therefore, the majorized version of the function in the left hand side of the constraint (5.10) is the right hand side of the inequality (5.18). This completes the proof of theorem 1. □

After replacing the left hand side of the constraint (5.14) by (5.15), the constraint (5.14) is still non-convex due to the presence of $\tilde{R}_b$. To further reduce the complexity, we propose to solve the optimization problem (5.12) iteratively with fixed $\tilde{R}_b$ obtained from the previous iteration. Specifically, at the m^{th} iteration, we solve the following relaxed optimization problem

$$\begin{aligned}
&\min_{\mathbf{w},\mathbf{q}} \qquad C_N \\
&s.t. \quad \text{Constraints (5.5), (5.6) and (5.9)} \\
&\lambda_\epsilon \sum_{b\in\mathcal{B}} \gamma_b^{m-1} \tilde{R}_b^{m-1} \left(\sum_{u\in\mathcal{U}} |w_{bu}|^2 + \alpha_b q_b + \frac{a_b^{m-1}}{\gamma_b^{m-1}} \right) \leq C \\
&\tilde{\Gamma}_u = \frac{\left|\mathbf{h}_u^H \mathbf{w}_u\right|^2}{\sum\limits_{u'\in\mathcal{U}_u} |\mathbf{h}_u^H \mathbf{w}_{u'}|^2 + \mathbf{h}_u^H \mathbf{Q} \mathbf{h}_u + \sigma_u^2} \geq \delta_u, \ \forall\ u \in \mathcal{U} \\
&\gamma_b^{m-1} = \frac{1}{\sum\limits_{u\in\mathcal{U}} |w_{bu}^{m-1}|^2 + \alpha_b q_b^{m-1} + \epsilon}, \ \forall\ b \in \mathcal{B} \\
&\quad q_b \geq 0, \ \forall b \in \mathcal{B}, .
\end{aligned} \tag{5.19}$$

The optimization problem (5.19) is still of high complexity due to the non-convexity of the cost function and the infinite set of possible CSI errors. In order to further reduce the complexity, we reformulate the optimization problem (5.19) as an SDP problem using the SDP relaxation. Then, the MM algorithm is used to guarantee a stationary point to the relaxed optimization problem.

5.3.2 SDP Reformulation

In this section, we define the SDP optimization variables $\mathbf{W}_u = \mathbf{w}_u \mathbf{w}_u^H$ where $\text{rank}(\mathbf{W}_u) = 1, \ \forall\ u \in \mathcal{U}$. Then, the S-Procedure method [4] and a rank-one semi-definite relaxation (SDR) approach [18] are used to formulate an equivalent SDP minimization problem. Based on the introduced quadratic variables $\mathbf{W}_u$ and after dropping the non-convex rank-one constraints, the relaxed minimization problem

$$\hat{C}_N = \sum_{b\in\mathcal{B}} \Bigg\{ \frac{1}{\nu_b} \left(\sum_{u\in\mathcal{U}} \text{Tr}\left(\mathbf{A}_b\mathbf{W}_u\right) + \text{Tr}\left(\mathbf{A}_b\mathbf{Q}\right) \right) + \left\| \sum_{u\in\mathcal{U}} \text{Tr}\left(\mathbf{A}_b\mathbf{W}_u\right) + \text{Tr}\left(\mathbf{A}_b\mathbf{Q}\right) \right\|_0 P_b^r$$
$$+ \sum_{u\in\mathcal{U}} \left\| \text{Tr}\left(\mathbf{A}_b\mathbf{W}_u\right) \right\|_0 R_u \left(1 - p_{bf_u}\right) \Bigg\}. \tag{5.21}$$

(5.19) can be reformulated as follows

$$\begin{aligned} \min_{\mathbf{W}_u,\mathbf{Q},\lambda_u} \quad & \hat{C}_N \\ s.t. \sum_{u\in\mathcal{U}} \quad & \text{Tr}\left(\mathbf{A}_b\mathbf{W}_u\right) + \alpha_b \text{Tr}\left(\mathbf{A}_b\mathbf{Q}\right) \le P_b \\ & \sum_{b\in\mathcal{B}} \hat{R}_b^{m-1} \left(\sum_{u\in\mathcal{U}} \text{Tr}\left(\mathbf{A}_b\mathbf{W}_u\right) + \alpha_b \text{Tr}\left(\mathbf{A}_b\mathbf{Q}\right) + \hat{a}_b^{m-1} \right) \le C \\ & \mathbf{\Delta}_u \succeq 0, \;\; \lambda_u \ge 0, \;\; \forall\, u \in \mathcal{U} \\ & \mathbf{W}_u \succeq 0, \mathbf{Q} \succeq 0, \mathbf{Q} \text{ is diagonal}, \forall\, u \in \mathcal{U}, \end{aligned} \tag{5.20}$$

where $\hat{C}_N$ is given in (5.21), $\hat{R}_b^{m-1} = \gamma_b^{m-1}\tilde{R}_b^{m-1}$, $\hat{a}_b^{m-1} = \frac{a_b^{m-1}}{\gamma_b^{m-1}}$, $\mathbf{A}_b$ denotes the diagonal matrix with 1 at the main diagonal entry b and zeros otherwise, and where the optimization is over the beamforming matrices $\mathbf{W}_u$, the quantization noise covariance matrix $\mathbf{Q}$, and the introduced variables λ_u, and where the matrix $\mathbf{\Delta}_u$ are denoted by

$$\mathbf{\Delta}_u = \begin{bmatrix} \mathbf{G}_u + \lambda_u \mathbf{E}_u & \mathbf{G}_u \tilde{\mathbf{h}}_u \\ \tilde{\mathbf{h}}_u^H \mathbf{G}_u & \tilde{\mathbf{h}}_u^H \mathbf{G}_u \tilde{\mathbf{h}}_u - \sigma_u^2 - \lambda_u \end{bmatrix}, \tag{5.22}$$

and where

$$\mathbf{G}_u = \frac{1}{\delta_u}\mathbf{W}_u - \sum_{u'\in\mathcal{U}_u} \mathbf{W}_{u'} - \mathbf{Q}. \tag{5.23}$$

Although the feasibility set of the optimization problem (5.20) is convex, it is still of high complexity due to the non-convexity of the cost function (5.21). Next, we focus

$$\eta_b^{m-1} = \frac{1}{\nu_b} + \frac{\lambda_\epsilon P_b^r}{\epsilon + \sum\limits_{u\in\mathcal{U}_b} \mathrm{Tr}\left(\mathbf{A}_b\mathbf{W}_u^{m-1}\right) + \mathrm{Tr}\left(\mathbf{A}_b\mathbf{Q}^{m-1}\right)},$$
$$\beta_{bu}^{m-1} = \frac{R_u(1-p_{bf_u})\lambda_\epsilon}{\epsilon + \mathrm{Tr}\left(\mathbf{A}_b\mathbf{W}_u^{m-1}\right)}. \tag{5.25}$$

---

on determining an approximate solution to the relaxed optimization problem (5.20) by approximating the ℓ_0-norm, then, using the MM algorithm.

5.3.3 Majorization-Minimization Approach

The optimization problem (5.20) is still of high complexity due to the non-convex cost function. This section focuses on finding an approximate solution to (5.20) by first approximating the ℓ_0-norm and then using the MM technique. Specifically, by removing the limit from (5.13) and replacing ϵ by a small value, the cost function in the optimization problem (5.20) can be approximated as follows

$$\begin{aligned}\tilde{C}_N = \sum_{b\in\mathcal{B}} \Bigg\{ &\frac{1}{\nu_b}\left(\sum_{u\in\mathcal{U}} \mathrm{Tr}\left(\mathbf{A}_b\mathbf{W}_u\right) + \mathrm{Tr}\left(\mathbf{A}_b\mathbf{Q}\right)\right) \\ &+ \lambda_\epsilon \log\left(1 + \epsilon^{-1}\left\{\sum_{u\in\mathcal{U}} \mathrm{Tr}\left(\mathbf{A}_b\mathbf{W}_u\right) + \mathrm{Tr}\left(\mathbf{A}_b\mathbf{Q}\right)\right\}\right) P_b^r \\ &+ \lambda_\epsilon \sum_{u\in\mathcal{U}} \log\left(1 + \epsilon^{-1}\mathrm{Tr}\left(\mathbf{A}_b\mathbf{W}_u\right)\right) R_u\left(1 - p_{bf_u}\right)\Bigg\}.\end{aligned} \tag{5.24}$$

Clearly, the above cost function is not convex. Next, the MM algorithm is used to find a stationary point to the obtained optimization problem. The MM approach consists of first finding a surrogate function that majorizes the cost function. Then, it minimizes the obtained function until a local optimal solution of the optimization problem with cost function (5.24) is reached.

Theorem 2. *The surrogate function that majorizes the objective function (5.24) is* *given by*

$$C_N^m(\mathbf{W}_u, \mathbf{Q}) = \sum_{b \in \mathcal{B}} \eta_b^{m-1} \left(\sum_{u \in \mathcal{U}} \mathrm{Tr}\left(\mathbf{A}_b \mathbf{W}_u\right) + \mathrm{Tr}\left(\mathbf{A}_b \mathbf{Q}\right) \right) + \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} \beta_{bu}^{m-1} \mathrm{Tr}\left(\mathbf{A}_b \mathbf{W}_u\right) + c(\mathbf{W}_u^{m-1}, \mathbf{Q}^{m-1}), \tag{5.26}$$

where η_b^{m-1}, β_{bu}^{m-1} *are given in (5.25) and* $c(\mathbf{W}_u^{m-1}, \mathbf{Q}^{m-1})$ *is a constant depending only on the beamforming matrices* $\mathbf{W}_u^{m-1}$ *and the quantization noise covariance matrix* $\mathbf{Q}^{m-1}$ *of the previous iteration.*

Proof. Based on theorem 1, the proof follows. □

Using the above theorem, the MM approach solves the following optimization problem at the m^{th} iteration

$$\begin{aligned} \min_{\mathbf{W}_u, \mathbf{Q}, \lambda_u} \quad & C_N^m(\mathbf{W}_u, \mathbf{Q}) \\ s.t. \sum_{u \in \mathcal{U}} \quad & \mathrm{Tr}\left(\mathbf{A}_b \mathbf{W}_u\right) + \alpha_b \mathrm{Tr}\left(\mathbf{A}_b \mathbf{Q}\right) \leq P_b, \ \forall \, b \in \mathcal{B} \\ & \sum_{b \in \mathcal{B}} \hat{R}_b^{m-1} \left(\sum_{u \in \mathcal{U}} \mathrm{Tr}\left(\mathbf{A}_b \mathbf{W}_u\right) + \alpha_b \mathrm{Tr}\left(\mathbf{A}_b \mathbf{Q}\right) + \hat{a}_b^{m-1} \right) \leq C \\ & \boldsymbol{\Delta}_u \succeq 0, \ \lambda_u \geq 0, \ \forall \, u \in \mathcal{U} \\ & \mathbf{W}_u \succeq 0, \mathbf{Q} \succeq 0, \mathbf{Q} \text{ is diagonal}, \forall \, u \in \mathcal{U}. \end{aligned} \tag{5.27}$$

The above optimization problem is an SDP problem. Therefore, it can be solved using efficient numerical algorithms [4].

5.3.4 Proposed Iterative Algorithm

To solve the optimization problem, one can use a two loop algorithm where the outer loop is responsible of updating $\hat{R}_b^{m-1}$ and $\hat{a}_b^{m-1}$ and the inner loop is responsible of solving the SDP problem (5.27) and updating η_b^{m-1} and β_b^{m-1}. Such algorithm can significantly increase of computational complexity. To reduce the complexity, the proposed iterative algorithm combines the two loops and iterates between two levels. At the first level, it solves the optimization problem (5.27). Then, it updates $\tilde{R}_b^m$, γ_b^m, η_b^m and β_b^m. We summarize the proposed relaxed MM algorithm to find an approximate solution to (5.12) in Table 3.

Algorithm 3 The Iterative Relaxed MM Algorithm

Initialization : Initialize $\tilde{R}_b^0$, $\hat{a}_b^0$, η_b^0 and β_b^0 and the iteration index to $m = 1$.
Repeat :
1: Solve the optimization problem (5.27). If it is infeasible, go to **End**;
2: Update $\hat{R}_b^m$, $\hat{a}_b^m$, η_b^m and β_b^m;
3: Update the iteration index $m = m + 1$;
Until : Desired level of convergence.
4: Find the optimal beamforming vectors $\mathbf{w}_u^*$, $\forall\, u \in \mathcal{U}$ and the optimal quantization noise covariance matrix $\mathbf{Q}^*$.
End

The following theorem shows the convergence properties of the proposed algorithm to solve the optimization problem (5.12).

Theorem 3. *The Iterative Relaxed MM Algorithm is guaranteed to converge to a feasible solution of the original optimization problem as ϵ tends to 0.*

Proof. Please refer to Appendix A for details. □

The proposed iterative algorithm does not always lead to a rank-one solution. However, a rank-one solution can be guaranteed using the randomization techniques [4].

5.3.5 Complexity Analysis

The implementation of the Iterative Relaxed MM Algorithm requires to solve the semi-definite program (5.27) with $c = (B + 3U + 3)$ SDP constraints and $v = (2U + 1)$ SDP variables at each iteration. Further, the second step of the proposed algorithm consists of updating $\hat{R}_b^{m+1}$, $\hat{a}_b^{m+1}$, η_b^{m+1} and β_b^{m+1}. Therefore, the computation complexity of the proposed algorithm comes mainly from solving the SDP problem. If the obtained solution does not satisfy the relaxed rank-one constraints. The gaussian randomization technique [4] is applied to estimate an approximate rank-one solution. This step may significantly increase the computational complexity of the proposed algorithm since to guarantee a good solution the gaussian randomization technique may need a very large number of samples.

To solve the SDP problems generated at each iteration of the proposed algorithm, one may use the interior point method which is implemented in most advanced solvers, e.g. SDPT3 and SeDuMi. In large-scale network (i.e, B and U are large), the computation complexity of the proposed algorithm using the interior point method can be huge due to solving a large-scale SDP problem at each iteration. To reduce complexity, the splitting conic solver (SCS) [5], which is based on the ADMM algorithm [10], can be applied to solve the large-scale SDP problem (5.27). The SCS solves the large-scale SDP problem by performing parallel cone projection and subspace projection which significantly reduces the computational complexity as compared to the interior point based solvers.

5.4 Simulation Results

This section provides simulation examples to show the convergence behavior of the proposed Iterative Relaxed MM Algorithm. To this end, consider a cache-enabled CRAN scenario formed by $B = 16$ single-antenna cache-enabled BS, where each bs is connected to $U = 5$ single-antenna MUs. Assume that the BSs and MUs are uniformily and independently distributed in the square region [0 1000] $\times$ [0 1000] meters. Further, the estimated channel vectors are generated using Rayleigh fading component and a distance-dependent path loss, modeled as $L(d_{bu}) = 128.1 + 37.6\log_{10}(d_{bu})$. Each user across the network randomly requests one content for the BSs according to the content popularity, modeled as Zipf distribution with skewness parameter 1. The noise power spectral density is set to $\sigma_u^2 = -98$ dBm $\forall u$. We set the maximum transmit power of BS b to $P_b = 1$ Watts, the total backhaul capacity limit to $C = 560$ Mbps, the relative power consumption to $P_{rb} = 38$ Watts, $\nu_b = 2.5$ and the accuracy matrix $\mathbf{E}_u = \frac{1}{a}\mathbf{I}_B$ where $a > 0$. Additionally, $\tilde{R}_b^0$, $\hat{a}_b^0$, η_b^0 and β_b^0 $\forall b \in \mathcal{B}$ are set to 1.

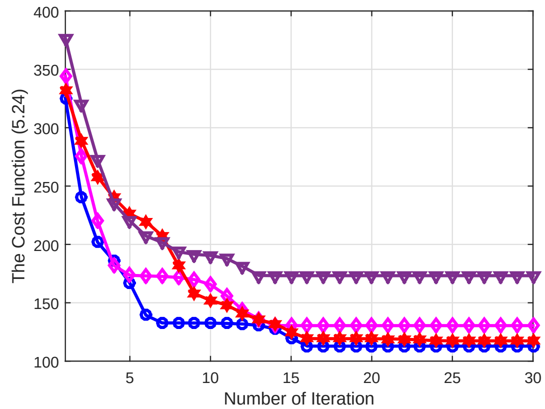

Figure 5.2: Convergence behavior of the Iterative Relaxed MM Algorithm.

We set the SINR target to $\delta_u = 10$dB $\forall u \in \mathcal{U}$ and the positive constant a to 0.01 and $\epsilon = 10^{-6}$. Figure 5.2 shows the convergence behavior of the proposed Iterative Relaxed MM Algorithm for different channel realizations. It can be noticed that the proposed algorithm converges for all the considered channel realizations. Figure 5.2 further shows that the proposed algorithm has a reasonable convergence speed where the maximum number of iteration is 20 for the considered channel realizations. It is especially remarkable that the cost function (5.24) is driven downhill. This results validates the statement shown in theorem 3.

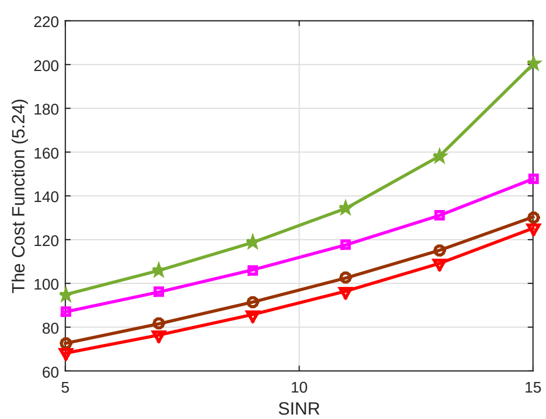

Figure 5.3: The cost function (5.24) versus SINR target.

Figure 5.3 illustrates the value of the cost function (5.24) as function of the target SINR. It can be noticed that increasing the quality of service requirements increases the total power consumption and the backhaul consumption.

5.5 Conclusion

This chapter considers the downlink of a cache-enabled CRAN, where each cache-enabled single-antenna BS serves a pre-known set of single antenna mobile users. The chapter assumes that only imperfect CSI is available at the cloud and each BS is connected to the cloud through limited-capacity backhaul link. The chapter provides an iterative algorithm to solve the network total cost problem. Simulation results show that the proposed algorithm convergence in a reasonable number of iterations.

Chapter 6

Concluding Remarks

6.1 Summary

Sophisticated resource allocation optimization techniques are expected to be at the heart of research for next generation mobile radio systems. This thesis considers the downlink of CRAN architecture where several BSs serve a pre-know set of single antenna mobile users. After providing an overview of useful convex and non-convex optimization algorithm, this thesis mainly focuses on two scenarios: multi-CRAN and cache-enabled CRAN.

The thesis first considers the more practical scenario of a muli-CRAN architecture where each cloud is connected to several BSs. The network total power minimization problem subject to practical constraints is formulated. An iterative algorithm based on distributed convex approaches (i.e. dual decomposition and ADMM approaches) is presented. A key feature of the algorithm is that it can be implemented in a distributed fashion by allowing a limited information exchange between the coupled clouds. Further, the proposed approach guarantees a feasible solution to the original problem at each iteration.

Then, a cache-enabled CRAN scenario is considered where the cloud is connected to several single-antenna cache-enabled BSs through limited capacity backhaul links. The problem of minimizing the network total power cost and the backhaul cost subject to per-BS power constraint, the quality of service constraints, the total backhaul capacity constraint and the imperfect CSI constraints is considered in order to design the beamforming vector of each user across the network, the quantization noise covariance matrix of the limited-capacity backhaul links and the BS clustering. An iterative algorithm based on the MM approach is presented. One of the highlight of the proposed iterative algorithm is that it converges in a reasonable number of iteration.

6.2 Future Work

To meet the increasing demands in mobile data traffic and high energy efficiency, a tremendous increase in the network density is expected which leads to large-scale optimization problems. As a result, the distributed algorithms presented in this thesis will have high computational complexity due to solving large-scale SOCP problem or large-scale SDP problem at each iteration. Most of the existing solvers (e.g. SDPT3) relies on the interior point method which is infeasible for large-scale problem due to its computational complexity. The new solver SCS [5] based on the first order method ADMM can be used to solve the large-scale SOCP problem or large-scale SDP problem in a distributed fashion. However, the accuracy of the SCS solver is inferior as compared to the interior point based solvers. Potential question that arise here is whether the application of the ADMM technique to the large-scale SOCP problem or large-scale SDP problem can preserve the convergence of the distributed algorithm with reasonable computational complexity.

REFERENCES

[1] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] Y. Shi, J. Zhang, and K. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[3] H. Dahrouj, A. Douik, O. Dhifallah, T. Y. Al-Naffouri, and M. S. Alouini, "Resource allocation in heterogeneous cloud radio access networks: advances and challenges," *IEEE, Wireless Communications*, vol. 22, no. 3, pp. 66–73, June 2015.

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

[5] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *Journal of Optimization Theory and Applications*, pp. 1–27, 2016.

[6] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, Aug 2006.

[7] O. Dhifallah, H. Dahrouj, T. Y. Al-Naffouri, and M. S. Alouini, "Decentralized group sparse beamforming for multi-cloud radio access networks," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[8] S. Boyd, L. Xiao, A. Mutapcic, and J. Mattingley, "Notes on decomposition methods: course reader for convex optimization II, Stanford," 2007, available

online: http://www.stanford.edu/class/ee364b/.

[9] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Select. Areas Commun.*, vol. 24, p. 14391451, 2006.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[11] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer.Statistician*, pp. 30–37, 2004.

[12] G. Obozinski and F. Bach, "Convex Relaxation for Combinatorial Penalties," *ArXiv e-prints*, May 2012.

[13] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[14] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[15] A. Tolli, H. Pennanen, and P. Komulainen, "Decentralized minimum power multi-cell beamforming with limited backhaul signaling," *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 570–580, February 2011.

[16] H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Distributed cloud association in downlink multicloud radio access networks," *CoRR*, vol. abs/1503.00267, 2015.

[17] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Inter-cluster design of precoding and fronthaul compression for cloud radio access networks," *IEEE Wireless Communications Letters*, vol. 3, no. 4, pp. 369–372, Aug 2014.

[18] G. Zheng, K. kit Wong, and T. sang Ng, "Robust linear MIMO in the downlink: A worst-case optimization with ellipsoidal uncertainty regions," *EURASIP J. Adv. Signal Process.*, June 2008.

[19] P. Patil, B. Dai, and W. Yu, "Performance comparison of data-sharing and compression strategies for cloud radio access networks," *European Signal Processing Conference (EUSIPCO)*, Sept 2015.

[20] O. Dhifallah, H. Dahrouj, T. Y. Al-Naffouri, and M. Alouini, "Joint hybrid backhaul and access links design in cloud-radio access networks," in *IEEE 82nd Vehicular Technology Conference, VTC Fall 2015, Boston, MA, USA, September 6-9, 2015*, 2015, pp. 1–5.

[21] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2988–3003, June 2012.

[22] H. Pennanen, A. Tolli, and M. Latva-aho, "Decentralized robust beamforming for coordinated multi-cell MISO networks," *IEEE Signal Processing Letters*, vol. 21, no. 3, pp. 334–338, March 2014.

[23] H. Zhou, M. Tao, E. Chen, and W. Yu, "Content-centric multicast beamforming in cache-enabled cloud radio access networks," *CoRR*, vol. abs/1504.05663, 2015.

[24] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *CoRR*, vol. abs/1601.01070, 2016.

[25] A. Tolli, H. Pennanen, and P. Komulainen, "SINR balancing with coordinated multi-cell transmission," in *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, April 2009, pp. 1–6.

[26] S. Joshi, M. Codreanu, and M. Latva-aho, "Distributed resource allocation for MISO downlink systems via the alternating direction method of multipliers," *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 1, 2014.

APPENDICES

A Detailed Proof of Theorem 3

In this appendix, we will present a detailed proof of Theorem 3. We start by proving that the Iterative Relaxed MM Algorithm is guaranteed to converge. Specifically, we will first prove that the objective function (5.24) is driven downhill, i.e.

$$\tilde{C}_N(\mathbf{W}_u^{m+1}, \mathbf{Q}^{m+1}) \leq \tilde{C}_N(\mathbf{W}_u^{m}, \mathbf{Q}^{m}) \ \forall \ m, \tag{A.1}$$

where m denotes the iteration index of the optimization problem (5.27). To this end, let $\mathbf{W}_u^m$, $\mathbf{Q}^m$, $\mathbf{W}_u^{m+1}$ and $\mathbf{Q}^{m+1}$ denote the optimal solution of the optimization problem (5.27) at iteration m and $m+1$ respectively. We consider two particular cases in terms of whether $\mathbf{W}_u^m$ and $\mathbf{Q}^m$ are feasible for the optimization problem (5.27) at the iteration $m+1$.

Case one: When $\mathbf{W}_u^m$ and $\mathbf{Q}^m$ are feasible for the optimization problem (5.27) at the iteration $m+1$. In this case,we have

$$\begin{aligned} \tilde{C}_N(\mathbf{W}_u^{m+1}, \mathbf{Q}^{m+1}) &\leq C_N^m(\mathbf{W}_u^{m+1}, \mathbf{Q}^{m+1}) \\ &= \min_{\mathbf{W}_u, \mathbf{Q}, \lambda_u} C_N^m(\mathbf{W}_u, \mathbf{Q}) \\ &\quad\ s.t. \text{ Constraints in (5.27)} \\ &\leq C_N^m(\mathbf{W}_u^m, \mathbf{Q}^m) \\ &= \tilde{C}_N(\mathbf{W}_u^m, \mathbf{Q}^m), \end{aligned} \tag{A.2}$$

where (A.2) follows because $\mathbf{W}_u^m$ and $\mathbf{Q}^m$ are feasible for the optimization problem (5.27) at the iteration $m+1$.

Case two: When $\mathbf{W}_u^m$ and $\mathbf{Q}^m$ are not feasible for the optimization problem (5.27) at the iteration $m+1$. In this case, $\mathbf{W}_u^m$ and $\mathbf{Q}^m$ only violate the following constraint

$$\sum_{b\in\mathcal{B}} \hat{R}_b^m \left(\sum_{u\in\mathcal{U}} \mathrm{Tr}\left(\mathbf{A}_b \mathbf{W}_u^m\right) + \alpha_b \mathrm{Tr}\left(\mathbf{A}_b \mathbf{Q}^m\right) + \hat{a}_b^m \right) \leq C. \tag{A.3}$$

Given the optimization problem considered in chapter 5, the per-BS power constraints and the above constraint (A.3) are only feasibility constraints and the SINR constraints are both feasibility and optimality constraints. This means that the SINR constraints are the only constraints that determine the optimal solution of the optimization problem (5.27). We know that $\mathbf{W}_u^m$ and $\mathbf{Q}^m$ verifies the SINR constraints and the per-BS power constraints of the optimization problem (5.27) because $\mathbf{W}_u^m$ and $\mathbf{Q}^m$ are the optimal solution (5.27) of at the iteration m. Further, we know that $\mathbf{W}_u^m$ and $\mathbf{Q}^m$ are the optimal solution of (5.27) at the iteration $m+1$. Since violating the constraints (A.3) does not influence the optimality of (5.27), we have

$$\begin{aligned} C_N^m(\mathbf{W}_u^{m+1}, \mathbf{Q}^{m+1}) = & \min_{\mathbf{W}_u, \mathbf{Q}, \lambda_u} C_N^m(\mathbf{W}_u, \mathbf{Q}) \\ & s.t. \text{ Constraints in (5.27)} \\ \leq & \; C_N^m(\mathbf{W}_u^m, \mathbf{Q}^m). \end{aligned} \tag{A.4}$$

Therefore, the Iterative Relaxed MM Algorithm is guaranteed to converge. The relaxed constraint (A.3) is an upper bound of the approximated total backhaul capacity constraint (5.14). Thus, the solution of the Iterative Relaxed MM Algorithm is a feasible solution of the original optimization problem as ϵ tends to 0.

B Other Application of ADMM

This appendix[1] provides another interesting application of the ADMM approach in cognitive radio (CR) network. The downlink of a CR network formed by multiple primary and secondary transmitters is considered, where each multi-antenna transmitter serves a pre-known set of single-antenna users. This appendix assumes that the secondary and primary transmitters can transmit simultaneously their data over the same frequency bands, so as to achieve a high system spectrum efficiency. The appendix considers the balancing problem of maximizing the minimum SINR subject to both total power constraint of the secondary transmitters, and maximum interference constraint at each primary user due to secondary transmissions.

System Model and Problem Formulation

System Model

Consider the downlink of a spectrum sharing-based CR network with B_p primary transmitters and B_s secondary transmitters. Assume that each transmitter b is equipped with N_b antennas. Further, assume that the network comprises U_p single-antenna primary receivers and U_s single-antenna secondary receivers where each primary transmitter b serves U_{pb} primary users and each secondary transmitter b serves U_{sb} secondary users. The channel state information is assumed to be perfectly known by the secondary transmitters and the user association is assumed to be predefined. Let $\mathbf{w}_{bu} \in \mathbb{C}^{N_b}$ be the beamforming vector at transmitter b associated with user u, and

[1] A part of this appendix will appear in: SO. Dhifallah, H. Dahrouj, T. Y.Al-Naffouri, M-S. Alouini, "SINR Balancing with Decentralized Coordination in Cognitive Radio Networks", *IEEE Transactions on Vehicular Technology*, (Second Round Revision).

let $\mathbf{h}_{bu} \in \mathbb{C}^{N_b}$ be the channel vector from transmitter b to user u. Let b_u denotes the transmitter serving user u and $\bar{\mathcal{B}}_{su}$ denotes the set of all the secondary transmitters except for the secondary transmitter serving user u. The received signal $y_u \in \mathbb{C}$ at the secondary user u served by the secondary transmitter b can be written as follows

$$\begin{aligned} y_u = \mathbf{h}_{bu}^H \mathbf{w}_{bu} q_u + \sum_{u' \in \mathcal{U}_{sbu}} \mathbf{h}_{bu}^H \mathbf{w}_{bu'} q_{u'} + \sum_{b' \neq b} \sum_{u' \in \mathcal{U}_{sb'}} \mathbf{h}_{b'u}^H \mathbf{w}_{b'u'} q_{u'} \\ + \sum_{b' \in \mathcal{B}_p} \sum_{u' \in \mathcal{U}_{pb'}} \mathbf{h}_{b'u}^H \mathbf{w}_{b'u'} q_{u'} + n_u, \end{aligned} \tag{B.1}$$

where $\mathcal{U}_{sbu} = \mathcal{U}_{sb} \setminus \{u\}$ and $\mathcal{B}_p$ denotes the set of primary transmitters, and where q_u is a complex scalar denoting the data symbol for the secondary user u and $n_u \sim \mathcal{CN}(0, \sigma_u^2)$ represents the additive white Gaussian noise which is assumed to be independent from the transmitted data symbols q_u.

Problem Formulation

Based on the introduced signal model (B.1), the SINR of the secondary user u served by the secondary transmitter b can be expressed as

$$\Gamma_u = \frac{\left|\mathbf{h}_{bu}^H \mathbf{w}_{bu}\right|^2}{\sum\limits_{u' \in \mathcal{U}_{sbu}} \left|\mathbf{h}_{bu}^H \mathbf{w}_{bu'}\right|^2 + \sum\limits_{b' \neq b} \sum\limits_{u' \in \mathcal{U}_{sb'}} \left|\mathbf{h}_{b'u}^H \mathbf{w}_{b'u'}\right|^2 + \theta_u^2 + \sigma_u^2}, \tag{B.2}$$

where we assume that the transmitted data symbols q_u for the user u have unit power, i.e. $E(|q_u|^2) = 1$, and independent from each other. Besides, $\theta_u^2 = \sum_{b' \in \mathcal{B}_p} \sum_{u' \in \mathcal{U}_{pb'}} |\mathbf{h}_{b'u}^H \mathbf{w}_{b'u'}|^2$ denotes the interference seen by the secondary user u due to primary transmission which is assumed to be known by the secondary transmitters.

The optimization problem studied in this section considers the maximization of the minimum SINR subject to the total power constraint of the secondary transmitters and the maximum interference seen by each primary user due to secondary transmissions. Specifically, the considered optimization problem can be formulated as follows

$$\begin{aligned} \max_{\mathbf{w}} \quad & \min_{u} \ \Gamma_u \\ s.t. \quad & \sum_{b \in \mathcal{B}_s} \sum_{u \in \mathcal{U}_{sb}} ||\mathbf{w}_{bu}||_{\ell_2}^2 \ \leq \ P \\ & \sum_{b \in \mathcal{B}_s} \sum_{u \in \mathcal{U}_{sb}} |\mathbf{h}_{bu'}^H \mathbf{w}_{bu}|^2 \ \leq \ \beta_{u'}, \ \forall \ u' \in \mathcal{U}_p, \end{aligned} \tag{B.3}$$

where the optimization is over the beamforming vectors $\mathbf{w} = [\mathbf{w}_{bu}^T; \ \forall (b,u) \in (\mathcal{B}_s \times \mathcal{U}_{sb})]^T \in \mathbb{C}^N$, where $N = \sum_{b=1}^{B_s} U_{sb} N_b$, $\mathcal{B}_s$ denotes the set of secondary transmitters and $\beta_{u'}$ denotes the maximum interference that can be tolerated by the primary user u'. The above problem (B.3) can be easily solved using the bisection method [25]. However, centralized solutions to solve the SINR balancing optimization problem are impractical, since otherwise, the secondary transmitters would require joint signal-processing and high signaling overhead. Thus, a distributed algorithm between the secondary transmitters is proposed using ADMM approach and by allowing a limited information exchange between the secondary transmitters.

Distributed Max–Min SINR Beamforming

Problem Relaxation

First, define the secondary inter-cell interference terms ${\psi_{b'u}}^2$ from an interfering secondary transmitter b' to the secondary user u served by any secondary transmitter different than b' as ${\psi_{b'u}}^2 = \sum_{u' \in \mathcal{U}_{sb'}} \left| \mathbf{h}_{b'u}^H \mathbf{w}_{b'u'} \right|^2$. Second, define the primary inter-cell interference terms ${\chi_{b'u'}}^2$ from an interfering secondary transmitter b' to the primary user u' served by any primary transmitter as ${\chi_{b'u'}}^2 = \sum_{u \in \mathcal{U}_{sb'}} \left| \mathbf{h}_{b'u'}^H \mathbf{w}_{b'u} \right|^2$. Finally, define the local power consumption terms ${\kappa_{b'}}^2$ of the secondary transmitter b' as ${\kappa_{b'}}^2 = \sum_{u' \in \mathcal{U}_{sb'}} ||\mathbf{w}_{b'u'}||_{\ell_2}^2$. By relaxing the equalities in the secondary inter-cell interference terms ${\psi_{b'u}}^2$, the primary inter-cell interference terms ${\chi_{b'u'}}^2$ and the local power consumption terms ${\kappa_{b'}}^2$ into inequalities, the SINR balancing problem (B.3) can be reformulated as

$$\begin{aligned}
&\max_{\mathbf{w},\boldsymbol{\psi},\boldsymbol{\chi},\boldsymbol{\kappa}} \min_{u} \Gamma_u = \frac{|\mathbf{h}_{bu}^H \mathbf{w}_{bu}|^2}{\sum\limits_{u'\in\mathcal{U}_{sbu}} |\mathbf{h}_{bu}^H \mathbf{w}_{bu'}|^2 + \sum\limits_{b'\neq b} {\psi_{b'u}}^2 + \theta_u^2 + \sigma_u^2} \\
&s.t. \quad \sum_{u\in\mathcal{U}_{sb}} ||\mathbf{w}_{bu}||_{\ell_2}^2 + \sum_{b'\neq b} {\kappa_{b'}}^2 \ \leq \ P \\
&\qquad \sum_{u\in\mathcal{U}_{sb}} |\mathbf{h}_{bu'}^H \mathbf{w}_{bu}|_{\ell_2}^2 + \sum_{b'\neq b} {\chi_{b'u'}}^2 \ \leq \ \beta_{u'}, \ \forall \ u' \in \mathcal{U}_p \\
&\qquad {\psi_{b'u}}^2 \geq \sum_{u'\in\mathcal{U}_{sb'}} \left|\mathbf{h}_{b'u}^H \mathbf{w}_{b'u'}\right|^2, \ \forall u \notin \mathcal{U}_{b'}, \ \forall b' \in \mathcal{B}_s \\
&\qquad {\chi_{b'u'}}^2 \geq \sum_{u\in\mathcal{U}_{sb'}} \left|\mathbf{h}_{b'u'}^H \mathbf{w}_{b'u}\right|^2, \ \forall \ u' \in \mathcal{U}_p, \ \forall b' \in \mathcal{B}_s \\
&\qquad {\kappa_{b'}}^2 \geq \sum_{u'\in\mathcal{U}_{sb'}} ||\mathbf{w}_{b'u'}||_{\ell_2}^2, \ \forall b' \in \mathcal{B}_s,
\end{aligned} \tag{B.4}$$

where the optimization is over the beamforming vectors, the secondary inter-cell interference vector $\boldsymbol{\psi} = [\psi_{b'u};\ \forall u \notin \mathcal{U}_{sb'}, \forall b' \in \mathcal{B}_s]^T \in \mathbb{R}^{U'}$, the primary inter-cell interference vector $\boldsymbol{\chi} = [\chi_{b'u'};\ \forall u' \in \mathcal{U}_p, \forall b' \in \mathcal{B}_s]^T \in \mathbb{R}^{U_p B_s}$, the local power consumption vector $\boldsymbol{\kappa} = [\kappa_{b'};\ \forall b' \in \mathcal{B}_s]^T \in \mathbb{R}^{B_s}$, where $U' = \sum\limits_{b=1}^{B_s} \sum\limits_{b'\neq b} U_{b'}$. Relaxing the secondary inter-cell interference, the primary inter-cell interference and the local power consumption constraints with inequality in the reformulated problem (B.4) is, in general, sub-optimal as compared to (B.3). However, the inequality turns out to be always tight at the optimal solution.

Proposition 5. *The original optimization problem (B.3) and the relaxed optimization problem (B.4) have the same optimal solution.*

Proof. Let $\mathbf{w}^*$, $\boldsymbol{\psi}^*$, $\boldsymbol{\chi}^*$ and $\boldsymbol{\kappa}^*$ denote the optimal solution of the optimization problem (B.4) and $\Gamma_{u^*}^*$ denotes the optimal objective value of the optimization problem (B.4). Further, let $\tilde{\mathbf{w}}$ and $\tilde{\Gamma}_{\tilde{u}}$ denotes the optimal solution and the optimal objective value of the optimization problem (B.3), respectively. First, we can readily see that the optimal solution of (B.3) are feasible for the optimization problem (B.4) since $\tilde{\mathbf{w}}$ satisfies the third, fourth and fifth constraints of (B.4) with equality. Therefore, the optimal solution of the optimization problem (B.3) is upper bounded by the optimal solution of the optimization problem (B.4), i.e. $\tilde{\Gamma}_{\tilde{u}} \leq \Gamma^*_{u^*}$. Next, we will show that $\tilde{\Gamma}_{\tilde{u}} = \Gamma^*_{u^*}$. To this end, we start by showing that the third constraints of the optimization problem (B.4) are tight for the user with the minimum SINR u^*. We assume by contradiction that there exist at least $b' \in \mathcal{B}_{su^*}$, where $\mathcal{B}_{su^*}$ denotes all the secondary transmitters not serving user u^*, such that

$$\psi^*_{b'u^*}{}^2 > \sum_{u' \in \mathcal{U}_{sb'}} \left|\mathbf{h}^H_{b'u^*}\mathbf{w}^*_{b'u'}\right|^2 . \tag{B.5}$$

Therefore, we can find $\tilde{\psi}^2_{b'u^*} < \psi^*_{b'u^*}{}^2$ which satisfies the third constraint of the optimization problem (B.4). Such a $\tilde{\psi}^2_{b'u^*}$ further leads to a higher minimum SINR value of user u^* and it does not influence the minimum SINR value of all the other users. This contradicts with the optimality assumption stated in the beginning of this proof. Therefore, the third constraints of the optimization problem (B.4) are tight for the user with the minimum SINR u^*. Further, we can readily see that the optimal beamforming vector $\mathbf{w}^*$ of the optimization problem (B.4) satisfies the constraints of the original optimization problem (B.3). We know that the cost function of the optimization problem (B.4) is lower bounded by the cost function of the optimization problem (B.3). Given that the third constraints of the optimization problem (B.4) are tight for the user with the minimum SINR u^*, i.e. $\psi^*_{b'u^*}{}^2 = \sum_{u' \in \mathcal{U}_{sb'}} \left|\mathbf{h}^H_{b'u^*}\mathbf{w}^*_{b'u'}\right|^2$, $\forall b' \in \mathcal{B}_{su^*}$ and $\psi^{*2}_{b'u^-} \geq \sum_{u' \in \mathcal{U}_{sb'}} \left|\mathbf{h}^H_{b'u^-}\mathbf{w}^*_{b'u'}\right|^2$ for all the users u^- different than u^* and $\forall b' \in \mathcal{B}_{su^-}$. Therefore, replacing $\psi^{*2}_{b'u}$ by $\sum_{u' \in \mathcal{U}_{sb'}} \left|\mathbf{h}^H_{b'u}\mathbf{w}^*_{b'u'}\right|^2$, $\forall u \notin \mathcal{U}_{b'}$, $\forall b' \in \mathcal{B}_s$ in the cost function of the optimization problem (B.4) leads to the same SINR value of the user u^* and increases the SINR value of all the other users. Further, the beamforming vector $\mathbf{w}^*$ satisfies the constraints of the optimization problem (B.3).

Therefore, the beamforming vector $\mathbf{w}^*$ solves the optimization problem (B.3) and also we have $\tilde{\Gamma}_{\tilde{u}} = \Gamma^*_{u^*}$. This completes the proof of proposition 5. $\square$

Decentralized Solution via the Alternating Direction Method of Multipliers

This section proposes a decentralization algorithm to solve the SINR balancing optimization problem (B.3), using ADMM approach, first introduced in the context of multi-cell systems [26]. To this end, we first start by introducing local auxiliary variables and equality constraints in order to decouple the SINR constraints, the total power constraint and the primary interference constraints. Then, the ADMM algorithm [10] is used to achieve a distributed solution to the considered optimization problem. Specifically, problem (B.4) can be reformulated as

$$\begin{aligned} \max_{\mathbf{w},\boldsymbol{\psi},\boldsymbol{\chi},\boldsymbol{\kappa},\boldsymbol{\xi},\boldsymbol{\phi},\boldsymbol{\varrho}} \quad & \frac{1}{B_s}\sum_{b\in\mathcal{B}_s}\delta^{(b)} \\ s.t. \quad & \left\{\mathbf{w}_b, \boldsymbol{\xi}^{(b)}, \boldsymbol{\phi}^{(b)}, \boldsymbol{\varrho}^{(b)}, \boldsymbol{\delta}^{(b)}\right\} \in \mathcal{F}_b,\ \forall\ b \in \mathcal{B}_s \\ & \xi^{(b')}_{b'u} = \psi_{b'u}, \xi^{(b_u)}_{b'u} = \psi_{b'u},\ \forall\ (u,b') \in (\mathcal{U}_s \times \bar{\mathcal{B}}_{su}) \\ & \phi^{(b)}_{b'u'} = \chi_{b'u'},\ \forall\ (u',b,b') \in (\mathcal{U}_p \times \mathcal{B}_s \times \mathcal{B}_s) \\ & \varrho^{(b)}_{b'} = \kappa_{b'},\ \forall\ (b,b') \in (\mathcal{B}_s \times \mathcal{B}_s); \delta^{(b)} = \gamma,\ \forall\ b \in \mathcal{B}_s, \end{aligned} \tag{B.6}$$

where the vector $\boldsymbol{\xi} = [\xi^{(b)}_{b'u};\ \forall b \in \{b', b_u\}\ \forall u \notin \mathcal{U}_{b'}\ \forall b' \in \mathcal{B}_s]^T \in \mathbb{R}^{2U'}$, vector $\boldsymbol{\phi} = [\phi^{(b)}_{b'u'};\ \forall b \in \mathcal{B}_s\ \forall u \in \mathcal{U}_p\ \forall b' \in \mathcal{B}_s]^T \in \mathbb{R}^{{B_s}^2 U_p}$ and vector $\boldsymbol{\varrho} = [\varrho^{(b)}_{b'};\ \forall b \in \mathcal{B}_s\ \forall b' \in \mathcal{B}_s]^T \in \mathbb{R}^{B^{s2}}$ and where $\mathcal{F}^b$ denotes the feasibility set defined by the inequality constraints in (B.4) associated with the secondary transmitter b. Note that introducing the equality constraints in (B.6) decouples the SINR constraints, the total power constraint and the maximum interference caused by the secondary transmission constraints. To derive a distributed solution to the optimization problem (B.6),

$$g_b\left(\boldsymbol{\xi}^{(b)}, \boldsymbol{\phi}^{(b)}, \boldsymbol{\varrho}^{(b)}\right) = \left\|\boldsymbol{\xi}^{(b)} - \boldsymbol{\psi}^{(b)^k} + \frac{1}{\rho}\boldsymbol{\lambda}^{(b)^k}\right\|_{\ell_2}^2 + \left\|\boldsymbol{\phi}^{(b)} - \boldsymbol{\chi}^{(b)^k} + \frac{1}{\rho}\boldsymbol{\mu}^{(b)^k}\right\|_{\ell_2}^2 + \left\|\boldsymbol{\varrho}^{(b)} - \boldsymbol{\kappa}^{(b)^k} + \frac{1}{\rho}\boldsymbol{\vartheta}^{(b)^k}\right\|_{\ell_2}^2. \tag{B.8}$$

---

we start by formulating the augmented Lagrangian of the optimization problem (B.6) as in chapter 4. The ADMM algorithm solves the relaxed optimization problem (B.6) iteratively by performing two primal minimization steps and a dual variable update at each iteration.

First Step Minimization: The first step of the ADMM approach consists of minimizing the augmented Lagrangian over the variables $\mathbf{w}$, $\boldsymbol{\xi}$, $\boldsymbol{\phi}$, $\boldsymbol{\varrho}$ and $\boldsymbol{\delta}$ while all the other variables are fixed at their current values. It can be shown that this optimization problem is fully separable between the secondary transmitters. Specifically, each secondary transmitter b can update its corresponding beamforming vector and the introduced local variables independently by solving the following optimization problem

$$\begin{aligned} \min_{\substack{\mathbf{w}_b, \boldsymbol{\xi}^{(b)}, \delta^{(b)} \\ \boldsymbol{\phi}^{(b)}, \boldsymbol{\varrho}^{(b)}}} \; & -\frac{\delta^{(b)}}{B_s} + \frac{\rho}{2}\, g_b\left(\boldsymbol{\xi}^{(b)}, \boldsymbol{\phi}^{(b)}, \boldsymbol{\varrho}^{(b)}\right) + \frac{\rho}{2}\left|\delta^{(b)} - \gamma^k + \frac{1}{\rho}\nu^{(b)^k}\right|^2 \\ s.t. \; & \left\{\mathbf{w}_b, \boldsymbol{\xi}^{(b)}, \boldsymbol{\phi}^{(b)}, \boldsymbol{\varrho}^{(b)}, \boldsymbol{\delta}^{(b)}\right\} \in \mathcal{F}_b, \end{aligned} \tag{B.7}$$

where the function $g_b\left(\boldsymbol{\xi}^{(b)}, \boldsymbol{\phi}^{(b)}, \boldsymbol{\varrho}^{(b)}\right)$ is given in (B.8) and where the optimization is over the local beamforming vector $\mathbf{w}_b$ and the local variables $\boldsymbol{\xi}^{(b)}$, $\boldsymbol{\phi}^{(b)}$, $\boldsymbol{\varrho}^{(b)}$ and $\delta^{(b)}$ of the secondary transmitter b. The above optimization problem is not convex due to the SINR constraints. However, for a fixed $\delta^{(b)}$, the optimization problem (B.7) can be easily reformulated as a Second-Order Cone Programming (SOCP). Thus, it can be solved using efficient numerical algorithms [4].

Proposition 6. *The following function is uni-modal on an interval* $[0\ \delta^{(b)}_{max}]$, $\forall b \in \mathcal{B}_s$

$$h\left(\delta^{(b)}\right) = -\frac{\delta^{(b)}}{B_s} + \frac{\rho}{2}\left|\delta^{(b)} - \gamma^k + \frac{1}{\rho}{\nu^{(b)}}^k\right|^2 + \frac{\rho}{2}g_b^*, \tag{B.9}$$

where g_b^* *is the optimal objective value of (B.7) for a fixed* $\delta^{(b)}$.

The proof of the above proposition is omitted in this appendix, as it mirror the proof used in [26]. Using the above proposition, the optimization problem (B.7) can be easily solved using the golden search method with any predefined accuracy $\epsilon > 0$.

Second Step Minimization: The second step of the ADMM approach consists of minimizing the augmented Lagrangian over the variables $\boldsymbol{\psi}$, $\boldsymbol{\chi}$, $\boldsymbol{\kappa}$ and γ while all the other variables are fixed at their current values. It can be shown that this optimization problem is convex and separable on the variables $\boldsymbol{\psi}$, $\boldsymbol{\chi}$, $\boldsymbol{\kappa}$ and γ. Specifically, the optimal secondary inter-cell interference terms, the optimal primary inter-cell interference terms , the optimal local power consumption terms and γ are respectively updated as follows

$$\begin{cases} {\psi_{bu}}^{k+1} = \frac{1}{2}\left({\xi^{(b)}_{bu}}^{k+1} + {\xi^{(b_u)}_{bu}}^{k+1}\right) + \frac{1}{2\rho}\left({\lambda^{(b)}_{bu}}^k + {\lambda^{(b_u)}_{bu}}^k\right) \\ {\chi_{bu'}}^{k+1} = \frac{1}{B_s}\sum\limits_{b'\in\mathcal{B}_s}\left({\phi^{(b')}_{bu'}}^{k+1} + \frac{1}{\rho}{\mu^{(b')}_{bu'}}^k\right) \\ {\kappa_b}^{k+1} = \frac{1}{B_s}\sum\limits_{b'\in\mathcal{B}_s}\left({\varrho^{(b')}_b}^{k+1} + \frac{1}{\rho}{\vartheta^{(b')}_b}^k\right) \\ \gamma^{k+1} = \frac{1}{B_s}\sum\limits_{b'\in\mathcal{B}_s}\left({\delta^{(b')}}^{k+1} + \frac{1}{\rho}{\nu^{(b')}}^k\right) \end{cases}, \tag{B.10}$$

Note that the above updates can be performed independently at each secondary transmitter by allowing the exchange of the introduced local variables between the coupled secondary transmitters.

Dual Variable Update: The last step of the ADMM algorithm consists of updating the dual variables associated with the equality constraints in (B.6). The dual variables $\lambda_{bu}^{(b')}$ $\forall b' \in (b, b_u)$ $\forall u \notin \mathcal{U}_{sb}$ $\forall b \in \mathcal{B}_s$; $\mu_{bu'}^{(b')}$ $\forall u \in \mathcal{U}_p$ $\forall b' \in \mathcal{B}_s$ $\forall b \in \mathcal{B}_s$; $\vartheta_b^{(b')}$ $\forall b' \in \mathcal{B}_s$ $\forall b \in \mathcal{B}_s$ and $\nu^{(b)}$ $\forall$ $b \in \mathcal{B}_s$ are respectively updated as follows

$$\begin{cases} {\lambda_{bu}^{(b')}}^{k+1} = {\lambda_{bu}^{(b')}}^{k} + \rho \left({\xi_{bu}^{(b')}}^{k+1} - {\psi_{bu}}^{k+1} \right) \\ {\mu_{bu'}^{(b')}}^{k+1} = {\mu_{bu'}^{(b')}}^{k} + \rho \left({\phi_{bu'}^{(b')}}^{k+1} - {\chi_{bu'}}^{k+1} \right) \\ {\vartheta_{b}^{(b')}}^{k+1} = {\vartheta_{b}^{(b')}}^{k} + \rho \left({\varrho_{b}^{(b')}}^{k+1} - {\kappa_{b}}^{k+1} \right) \\ {\nu^{(b)}}^{k+1} = {\nu^{(b)}}^{k} + \rho \left({\delta^{(b)}}^{k+1} - \gamma^{k+1} \right) \end{cases}, \tag{B.11}$$

It can be noticed that the dual variable update rules can be performed independently at each secondary transmitter.

Simulation Example

This subsection evaluates the performance of the proposed decentralized algorithm. Consider two secondary transmitters and one primary transmitter, i.e. $\mathcal{B}_s = 2$ and $\mathcal{B}_p = 1$. Further, assume that each transmitter serves two single-antenna receivers and each transmitter is equipped with three antennas. In the first secondary cell, the receivers are assumed to be uniformly and independently distributed in the square region [0 500] $\times$ [0 500] meters and the transmitter is assumed to be located at $(250, 250)$ meters. In the second secondary cell, the receivers are assumed to be uniformly and independently distributed in the square region [500 1000] $\times$ [0 500] meters and the transmitter is located at $(750, 250)$ meters. In the primary cell, the receivers are assumed to be uniformly and independently distributed in the square region [1000 1500] $\times$ [0 500] meters. The channel model is assumed to be formed by a distance-dependent path loss $L(d_{bu}) = 128.1 + 37.6\log_{10}(d_{bu})$, and Rayleigh fading component. We set the initial secondary inter-cell interference, the primary inter-cell interference, the local power consumption and all the dual variables to 0.01. Besides, we assume that $\beta_{u'} = -90$ dBm/Hz $\forall u' \in \mathcal{U}_p$.

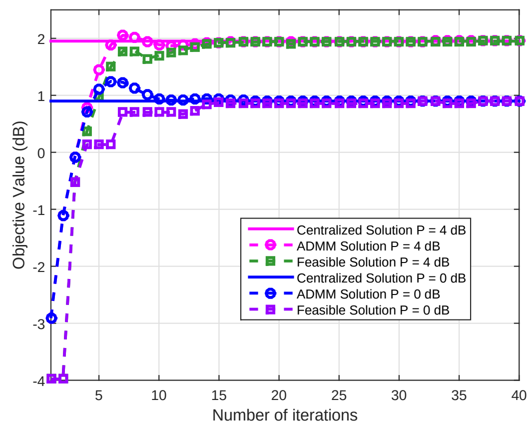

Figure B.1: Convergence behavior of the proposed algorithm for different maximum total transmit power.

We consider different maximum total transmit power values and we set the accuracy $\epsilon = 10^{-3}$. Figure B.1 illustrates the convergence behavior of the proposed algorithm. It can be noticed that our proposed decentralized solution converges to the centralized solution [25] in a reasonable number of iterations for different maximum total transmit power.

C Papers Submitted and Under Preparation

- [B1] A. Douik, H. Dahrouj, O. Dhifallah, T.Y. Al-Naffouri, M.-S. Alouini, "Coordinated Scheduling in Cloud Radio Access Networks", *submitted to Cambridge University Press.*

- [J3] O. Dhifallah, H. Dahrouj, T. Y.Al-Naffouri, M-S. Alouini, "Robust Beamforming in Cache-Enabled Cloud Radio Access Networks", *IEEE wireless Communications Letters* , (Under Preparation).

- [J2] O. Dhifallah, H. Dahrouj, T. Y.Al-Naffouri, M-S. Alouini, "SINR Balancing with Decentralized Coordination in Cognitive Radio Networks", *IEEE Transactions on Vehicular Technology*, (Second Round Revision).

- [J1] H. Dahrouj, A. Douik, O. Dhifallah, T.Y. Al-Naffouri, M.-S. Alouini, "Resource Allocation in Heterogeneous Cloud Radio Access Networks: Advances and Challenges", *IEEE Wireless Communications*, vol.22, no.3, pp.66,73, June 2015.

- [C3] O. Dhifallah, H. Dahrouj, T. Y.Al-Naffouri, M-S. Alouini, "Distributed Robust Power Minimization for the Downlink of Multi-Cloud Radio Access Networks", *submitted to Globecom 2016* .

- [C2] O. Dhifallah, H. Dahrouj, T.Y. Al-Naffouri; M.-S. Alouini, "Decentralized Group Sparse Beamforming for Multi-Cloud Radio Access Networks", *in Proc. of IEEE Globecom* , San Diego, USA, Dec 2015.

- [C1] O. Dhifallah, H. Dahrouj, T.Y. Al-Naffouri; M.-S. Alouini, "Joint Hybrid Backhaul and Access Links Design in Cloud-Radio Access Networks", *in Proc. of IEEE VTC* , Boston, USA, Sept 2015.