Network Optimization of Evolving Mobile Systems with Presence of Interference Coupling

LEI YOU
Abstract

The rapid development from 4G to 5G of mobile communications poses significant challenges in providing high rate and capacity, making it more crucial for efficient utilization of time-frequency resource via optimally configuring the network. Mathematical optimization serves as a powerful tool for addressing this type of problems. However, gauging its potential in large-scale cellular networks is non-trivial due to the inherent coupling relation of interference among cells. To address this issue, the dissertation adopts a so-called load-coupling system that mathematically formulates the mutual influence caused by radio resource allocation among cells. The model defines the time-frequency resource consumption in each cell as the cell load. The load of one cell governs the interference that the cell generates to the others, since the cell transmits more frequently with higher load. The model enables joint optimization of resource allocation in multiple cells with respect to the dynamics of resource occupancy of cells. Under the load coupling model, the dissertation applies mathematical optimization to resolve resource management problems with respect to a number of evolving technologies, such as coordinated multipoint (CoMP) transmission, wireless relays, cloud radio access networks (C-RAN), and non-orthogonal multiple access (NOMA). Six research papers are included in the dissertation. Paper I addresses the question of how network planning and coordination may increase the efficiency of spectrum usage, by jointly optimizing user association and resource allocation with CoMP. Paper II investigates the potential of relay cooperation for energy saving. As an extension of Paper I, Paper III studies the capacity maximization for a target group of users, while keeping the quality-of-service (QoS) of other users being strictly met. Paper IV provides a general framework and a series of theoretical analysis for algorithmically enabling resource optimization in multi-cell NOMA with load coupling, where users are allowed to group together for sharing time-frequency resource by successive interference cancellation (SIC). Under this framework, Paper V explores the potential of NOMA networks. For a restricted setup of NOMA, the paper achieves globally optimal resource usage efficiency, in terms of power allocation, user pair selection, and time-frequency resource allocation. Finally, Paper VI, serving as a complementary note, overcomes a key obstacle in analyzing convergence of applying load coupling in NOMA networks.

Keywords: Resource optimization, load coupling, OFDMA, NOMA

Lei You, Department of Information Technology, Computing Science, Box 337, Uppsala University, SE-75105 Uppsala, Sweden.

© Lei You 2019

ISSN 1651-6214
ISBN 978-91-513-0726-8
urn:nbn:se:uu:diva-391133 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-391133)
“Nothing is true, everything is permitted.”
—The Creed
List of Papers

The following papers are included in the dissertation.


VI **L. You** and D. Yuan “A Note on Decoding Order in Optimizing Multi-Cell NOMA,” under review by *IEEE Transactions on Wireless Communications*.

Reprints were made with permission from the publishers.
The author also contributed to the following publications which are not included in this dissertation:


Acknowledgement

First and foremost, I would like to express my sincere gratitude to my main supervisor Prof. Di Yuan for the continuous support of my Ph.D study and research. It was you who guided me on mathematical optimization that is becoming the way of my future career. I appreciate your patience for helping me overcome sloppiness, stubbornness, and arrogance along the path towards a mature researcher. I feel grateful for the time you took for breaking things down for me to understand at the early stage of my study, and the countless effort you paid throughout all my Ph.D life for training me to be professional. *I just want to thank you for being the best supervisor in the world to me.* I have enjoyed working with you so much over the years.

I would also like to thank my co-supervisor, Prof. Pierre Flener. Thank you for always having put efforts on my study. Your meticulous attitude towards everything in work and life has been more than impressive. Thank you so much for your kindly support on my studies, teaching, research work, and writing.

It has been a great joy working with my co-authors Dr. Lei Lei, Dr. Nikolaos Pappas, Prof. Peter Värbrand, Dr. Sumei Sun, Dr. Symeon Chatzinotas, and Prof. Björn Ottersten. Thank you for your collaboration and interesting ideas. A special thank you to Dr. Qi Liao, who has been extremely helpful with discussions and relevant techniques.

I am grateful to my colleagues Ghafour Ahani, Dr. Pawel Wiatr, Dr. Tao Deng, Dr. Yunyun Zhu, Gaoyang Dai, Yuan Gao, Albert Mingkun Yang, and Xiuming Liu. It has been full of fun for studying and discussing research with you together. Thanks for your kindly help, support and the friendly environment you brought to me, as well as all the fun we have had in the last few years. Also, I would like to thank my friends and former colleagues Dr. Vangelis Angelakis and Dr. Scott Fowler, for the help they provided on my early stage of Ph.D study. Great thanks to Dr. Marian Codreanu. It was a pleasant time of discussing the knowledge of physical layer communications with you. Thanks for your tough questions which incepted me to widen my research from various perspectives. Many thanks to Dr. Yixin Zhao, who generously lent me help to studying and campus life upon my early arrival to Sweden. Special thanks to Ioannis Avgouleas, who had shared office with me for a year at Norrköping, where we had a lot of fun with coffee and mathematics.

I express my gratitude to Dr. Ying Huang and Fredrik Larsson for driving me to the hospital when my arm was injured at that winter. I am particularly grateful to Bohui Jiang Axelsson and Peter Axelsson for the joyful time we had spent together in Uppland Väsby every year. Many thanks to Xiaoming
Lin, for all the valuable and professional advices on my career development, as well as being there whenever I need wisdom. I also wish to thank all my dear friends in China, Sweden, United Kingdom, Greece, Singapore, and Estonia, for all the memorable moments we have experienced together.

Many thanks to Prof. Di Yuan and his daughter Ms. Emilie Yuan for helping me with the Swedish summary. Special thanks to Dr. Yunyun Zhu for designing and drawing the cover of this dissertation. Thank you so much for being such creative and ingenious.

Last but not least, great thanks to my family. I would like to thank my wife Lejuan, and express my deepest gratitude to our parents, for all of their encouragement, support, and love ♥.

The dissertation work has been partially supported by Swedish Research Council and the European Union.

Stockholm, August 2019
Lei You
Sammanfattning på Svenska


Avhandlingen tillämpar matematisk optimering för att uppnå effektiv resursanvändning i storskaliga mobila nät. Forskningen har bestått av att utveckla


# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4G</td>
<td>The fourth generation</td>
</tr>
<tr>
<td>5G</td>
<td>The fifth generation</td>
</tr>
<tr>
<td>RRM</td>
<td>Radio resource management</td>
</tr>
<tr>
<td>BS</td>
<td>Base station</td>
</tr>
<tr>
<td>CoMP</td>
<td>Coordinated multipoint</td>
</tr>
<tr>
<td>C-RAN</td>
<td>Cloud radio access networks</td>
</tr>
<tr>
<td>OFDMA</td>
<td>Orthogonal frequency division multiple access</td>
</tr>
<tr>
<td>OMA</td>
<td>Orthogonal multiple access</td>
</tr>
<tr>
<td>NOMA</td>
<td>Non-orthogonal multiple access</td>
</tr>
<tr>
<td>RU</td>
<td>Resource unit</td>
</tr>
<tr>
<td>RB</td>
<td>Resource block</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality-of-service</td>
</tr>
<tr>
<td>SIC</td>
<td>Successive interference cancellation</td>
</tr>
<tr>
<td>IC</td>
<td>Interference cancellation</td>
</tr>
<tr>
<td>JT</td>
<td>Joint transmission</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal-to-interference-and-noise ratio</td>
</tr>
<tr>
<td>RRH</td>
<td>Remote radio head</td>
</tr>
<tr>
<td>SIF</td>
<td>Standard interference function</td>
</tr>
<tr>
<td>BBU</td>
<td>Baseband units</td>
</tr>
<tr>
<td>PPP</td>
<td>Poisson point process</td>
</tr>
<tr>
<td>LTE</td>
<td>Long-term evolution</td>
</tr>
</tbody>
</table>
## Contents

Part I: Introduction and Overview ................................................................. 15

1 Introduction .............................................................................................. 17
   1.1 Motivation .................................................................................... 17
   1.2 Dissertation Outline and Organization ........................................... 18

2 Multi-Cell Resource Optimization .......................................................... 19
   2.1 Resource Allocation in Multi-Cell OFDMA .................................... 19
   2.2 Resource Allocation in Multi-Cell NOMA ..................................... 20
   2.3 Interference Modeling ................................................................... 22

3 Resource Allocation with Cell Load Coupling ........................................ 24
   3.1 Load Coupling from OMA to NOMA ......................................... 24
      3.1.1 Load Coupling of OMA .................................................. 24
      3.1.2 Load Coupling of NOMA ............................................... 25
   3.2 Existing Solutions and Challenges ................................................. 27

4 Scope and Contributions ......................................................................... 29

References .................................................................................................. 33

Part II: Research Papers ............................................................................. 39
Part I:
Introduction and Overview
1. Introduction

1.1 Motivation

In wireless communications networks, radio resource management (RRM) concerns the efficient usage of all kinds of resources for radio communications. RRM has been an important topic in the past decades. From the fourth generation (4G) of broadband cellular network technology to 5G, the rapid growth of the number of devices participating in communications and their fast increasing demand, pose significant challenges to RRM and makes it a more and more crucial topic [1–4]. In cellular systems, multiple access technologies enable a set of users to share the common resource of a network, such that all of them can utilize the communication service. In general, RRM in cellular networks can be classified into two scenarios: a cellular network consists of only one cell, and one consists of multiple cells, referred to as single-cell and multi-cell scenarios, respectively. Here, a cell is a geographic area served by a base station (BS). The multi-cell scenarios are much more complicated than the single-cell ones due to interference coupling among cells. For RRM optimization in multi-cell scenarios, modeling the interference among users and cells is essential. In this dissertation, the terms “cell” and “BS” are used interchangeably.

In the past years, an interference modeling approach based on the use of cell load, has become a frequently used method in the analysis of large-scale multi-cell scenarios [5–27]. The modeling approach is also known as “cell load coupling” [8], which defines the cell load to be the proportion of allocated resource in the time-frequency domain of each cell. Consider a 2-D resource grid of time and frequency. The grid can be reused by multiple cells. The proportion of time-frequency allocation of a cell is defined to be the load of the cell, which serves as a scaling parameter for the interference originating from this cell. The model captures the influence of resource allocation on the interference that a cell generates to the others, and is shown to possess good accuracy in performance evaluation of large-scale multi-cell networks. Furthermore, it has been proved that some classic results for tackling the power control problem [28], also apply to the load coupling system in specific scenarios [5–27]; this eases the analysis and problem solving to some extent.

This dissertation addresses radio resource optimization under the load coupling system, taking into consideration several evolving technologies such as coordinated multipoint (CoMP) transmission, wireless relay techniques, cloud radio access networks (C-RAN), and non-orthogonal multiple access
The main objective of this dissertation is to investigate the potential of mathematical optimization in RRM with respect to power and spectrum resource efficiency. The dissertation aims at providing effective and scalable algorithmic solutions for resource allocation in large-scale networks. What's more, all the solution methodologies derived by this dissertation can be implemented in a distributed manner.

1.2 Dissertation Outline and Organization

The dissertation has two parts. Part I gives the background of multi-cell resource optimization, discusses inter-cell interference modeling, as well as presents the cell load coupling system. Part II consists of six papers. Paper I addresses the question of how cell coordination may increase the efficiency of spectrum usage, by jointly optimizing user association and resource allocation with CoMP transmission. Paper II investigates the potential of relay cooperation for energy saving. As an extension of Paper I, Paper III studies capacity maximization for a target group of users, while keeping the quality-of-service (QoS) of other users being strictly met. Paper IV provides a general framework and a series of theoretical analysis for algorithmically enabling resource optimization in multi-cell NOMA with load coupling, where multiple users can use the same time-frequency resource by successive interference cancellation (SIC). With this framework, Paper V explores the potential of NOMA networks, by achieving globally optimal resource efficiency, in terms of power allocation, user pair selection, and time-frequency resource allocation. Finally, Paper VI, serving as a complementary note, overcomes a key obstacle in analyzing convergence of applying load coupling in NOMA networks.
2. Multi-Cell Resource Optimization

In this chapter, we introduce the background of radio resource allocation in multi-cell scenarios. In Section 2.1, we give a discussion of resource optimization of orthogonal frequency division multiple access (OFDMA) in multi-cell scenarios. Section 2.2 discusses resource allocation under NOMA with SIC. Section 2.3 outlines common methodologies for multi-cell interference modeling.

2.1 Resource Allocation in Multi-Cell OFDMA

OFDMA is widely used in 4G. The OFDMA technique provides a flexible mechanism for resource allocation [29, 30]. Namely, OFDMA operates on a time-frequency resource grid, composing of multiple resource units (RUs). An RU consists of one or multiple resource blocks (RBs) and is the minimum unit for resource allocation. In OFDMA, each RU can be allocated to up to one user in a cell, such that there is no time-frequency resource reuse among users within the cell, see Figure 2.1 for an illustration, where a cell allocates four RUs to two users. The fact that the RUs allocated to users in a cell are non-overlapping, is referred to as orthogonal resource allocation. In addition to RU allocation, the transmit power allocated on each RU needs to be determined as well. Power and time-frequency resource allocations are two major aspects in OFDMA resource management [29].

![Figure 2.1](image)

*Figure 2.1.* Time-frequency resource allocation in OFDMA, where user 1 and user 2 are served by one cell. The RUs marked red are allocated. The number on each RU indicates which user the RU is allocated to.
In multi-cell scenarios, resource allocation needs to take into account the inter-cell interference, as the data rate delivered to a user on an RU is determined by the power of transmission, the interference, and the noise. In general, a user that receives more interference needs to be allocated more power and time-frequency resource, for the same QoS requirement. On the other hand, the more resource a cell allocates to its users, the more interference it generates to others. Due to the interference among cells, resource allocation in one cell may affect the performance of the others, which in turn affect the performance of the cell. This chain of effects makes resource allocation in multi-cell scenarios much more complicated than the single-cell case. Therefore, inter-cell interference modeling is crucial in resource allocation of multi-cell OFDMA.

Two techniques can be applied in OFDMA for further improving resource efficiency. One promising technique is joint transmission (JT) with CoMP, as illustrated in Figure 2.2. With CoMP, multiple cells can simultaneously serve a user using the same RUs, without interfering each other [31].

![Figure 2.2. Two cells serve one user by reusing RUs via CoMP.](image)

Another technique is wireless relay, which provides a flexible way for coverage extension and fading alleviation of wireless channels, as illustrated in Figure 2.3. The cell transmits data to the relay via a wireless backhaul, and the user accesses the data from the relay by an access link. It is shown that wireless backhaul technologies have advantages over the fiber-based solution [32]. To avoid the loop interference [33], the backhaul and access links should operate on orthogonal resources. Namely, the RUs utilized by the two types of links do not overlap.

### 2.2 Resource Allocation in Multi-Cell NOMA

NOMA is a promising technique in 5G [34–45]. SIC, serving as the key technique of NOMA, enables multiple users of one cell to reuse same RUs [29]. With superposition coding, the signal can be separated at the receiver side. With interference cancellation (IC), a user decodes a strong interfering signal, as if it is the intended receiver. The decoded signal is then removed from the composite signal, before the user’s own signal is decoded. Due to the non-orthogonality resulted by RU sharing, users may receive intra-cell interference due to signals to other users using the same RUs. In SIC, the IC process is performed successively for users. Namely, all users receiving data on an
RU follow a decoding order that is determined based on the channels’ conditions, inter-cell interference, and noise. In general, in SIC, a user $a$ can decode another user $b$’s signal, if, for $b$’s signal, its signal-to-interference-and-noise ratio (SINR) at $a$ is no worse than that of $b$ [29]. The signals of the remaining users form intra-cell interference. In single-cell scenarios, the signal strength is mainly determined by the channel conditions and noise power. Assuming uniform noise, then the user being closest to the base station decodes all the others’ signals on the RU within the same cell, and receives no intra-cell interference (see Figure 2.4 for an illustration). In multi-cell scenarios, the signal strength is also affected by the inter-cell interference. Therefore, the decoding order of one cell is influenced by the resource allocation of other cells.

Figure 2.3. Illustration of wireless relay for cell coverage extension. The cell allocates six RUs for serving user 1, and two RUs for transmitting data to the wireless relay. The wireless relay receives the date on these two RUs, and allocates one RU for the transmission to user 2.

Figure 2.4. NOMA with SIC. Two users are allocated the same set of RUs.

In this dissertation, for an RU, the users that share the RU are referred to as the user group of the RU, and the selection of users for sharing RUs is re-
ferred to as user grouping. For the users in a group of an RU, the power on this RU is split among them. Compared to OFDMA, both user grouping and power split are new dimensions for resource optimization. Similar to traditional OFDMA, power split, user grouping, and time-frequency resource allocation cannot be optimized independently due to interference coupling. What’s more, all of them influence the decoding orders of user groups. The decoding orders are not known a priori in multi-cell NOMA due to interference coupling, posing challenges to system optimization.

2.3 Interference Modeling

Interference modeling has been widely investigated for multi-cell OFDMA. There are several types of approaches of interference modeling: stochastic, exact, worst-case, and approximate approaches.

- The stochastic approach (e.g. [20, 46–48]) assumes specific distributions of cells and users in cellular networks. Namely, to model the random behavior of the interference, the locations of cells or users are represented as stochastic processes, usually Poisson point process (PPP) and its variants. This type of modeling approaches is applicable to analyzing large-scale cellular networks. On the other hand, it does not apply to a network with specific given topology.

- Deterministic approaches are commonly used (e.g. [49, 50]) in formulating RRM as mathematical optimization problems. Binary variables are introduced to indicate whether a specific RU (or a specific channel) is allocated for transmission. In some cases, the binary variable is replaced by power allocation. The interference on each RU is then derived using the values of these variables. Compared to stochastic geometry, this type of approaches generally applies for analysis of networks with any given topology. However, it usually leads to high computational complexity in analysis and optimization, and sometimes intractability in problem solving.

- The worst-case approach (e.g. [51, 52]) simply considers the case that every cell generates the maximum interference to others, regardless of the time-frequency resource occupancy in the cell. The approach effectively simplifies modeling of multi-cell scenarios, though having issues due to underestimating the system performance. Interference modeling based on this method may significantly lose accuracy when the time-frequency resource occupancy is low; in this case the interference would be severely overestimated.

- Approximate approaches [5–9, 9–15, 15–23, 23–27, 53–55] strike a balance between exactness and tractability. In the past decade, a modeling approach that is named “load coupling” has been proposed and is becoming widely adopted for multi-cell resource optimization [5–27, 54, 55].
The model addresses a significant aspect in multi-cell resource optimization, namely, the dependency of interference on resource allocation. In this model, a cell’s load is defined to be the proportion of allocated time-frequency resource in a cell. The cell load is used for approximating the interference that the cell generates to others.

The above modeling approaches have also been used in multi-cell NOMA interference modeling [34,41,45,47], though compared to traditional OFDMA, multi-cell NOMA poses much more research challenges due to the interplay between SIC and inter-cell interference [35, 38, 41, 54]. Namely, the decoding order for performing SIC in one cell depends on the interference from the other cells, and hence also the resource allocations in other cells. Interference modeling in multi-cell NOMA is still an up-to-date topic.
3. Resource Allocation with Cell Load Coupling

This chapter introduces the load coupling model with respect to two scenarios, i.e., orthogonal time-frequency access (OMA) and NOMA. The former is used for modeling traditional OFDMA scenarios. The latter is a generalized case of the former and takes into account SIC that allows multiple users to use one RU simultaneously. Recall that the term “orthogonality” refers to the non-overlapping RU allocation in time-frequency resource grid.

3.1 Load Coupling from OMA to NOMA

Denote by $\mathcal{I}$ the set of cells, and $\mathcal{J}$ the set of users. Denote by $p_i$ the transmit power on one RU of cell $i$ ($i \in \mathcal{I}$), and $g_{ij}$ the gain of user $j$ ($j \in \mathcal{J}$) with respect to cell $i$. Denote by $\sigma^2$ the noise power. We use $i$ to refer to the serving cell of user $j$ unless stated otherwise.

3.1.1 Load Coupling of OMA

In OMA, the SINR of user $j$ is given below.

$$\text{SINR}_j = \frac{p_i g_{ij}}{\sum_{k \in \mathcal{I} \setminus \{i\}} p_k g_{kj} \rho_k + \sigma^2}$$

where $\rho_k$ in the denominator is the resource allocation of cell $k$ and serves as a scaling parameter of the inter-cell interference from cell $k$. In the load coupling model, $\rho_i$ ($i \in \mathcal{I}$) is also referred to as the “cell load” of $i$, defined in (3.2) below. We use $\mathcal{J}_i$ to represent the users served by cell $i$.

$$\rho_i = \sum_{j \in \mathcal{J}_i} x_j$$

where $x_j$ is the proportion of RUs allocated to user $j$ in cell $i$. For example, in Figure 2.1, $x_1$ and $x_2$ equal 12.5% and the cell load is 25%. Denote by $B$ and $M$ the bandwidth of each RU, and the total number of RUs in one cell for the time and total bandwidth under consideration, respectively. On each RU, the achievable amount of bits for user $j$ is $\log_2(1 + \text{SINR}_j)$. Then, $x_j MB \log_2(1 + \text{SINR}_j)$ computes the total amount of bits delivered to user.
Denote by \( d_j \) the bits demand of user \( j \). In order to satisfy this demand, we should have \( x_j MB \log_2(1 + \text{SINR}_j) \geq d_j \). Namely, the minimum required resource allocation for the demand requirement is as follows:

\[
x_j = \frac{d_j}{MB \log_2(1 + \text{SINR}_j)}
\]

In this dissertation, we use normalized \( d_j \) such that the two constants \( M \) and \( B \) are not necessary in our presentation.

Equations (3.1)-(3.3) are the load coupling model of OMA systems. If \( \rho_k = 0 \), it means that there is no time-frequency resource in use by cell \( k \). In this case, \( k \) does not generate interference to others. On the other hand, if \( \rho_k = 1 \), then all the RBs in cell \( k \) are used for transmission and cell \( k \) constantly interferes with others, i.e., \( p_{k}g_{kj} \) for all \( j \in \mathcal{J}_i \). Intuitively, a cell’s load can be viewed as the likelihood that the cell interferes with others. Generally, \( \rho_k (k \in \mathcal{I}) \) captures the coupling relationship between resource allocation and interference. One can observe that the load coupling equations quantify how the change of resource allocation influences the interference that the cell generates to others. Combining (3.1)-(3.3), the load coupling system in OMA reads (\( i \in \mathcal{I} \)):

\[
\rho_i = \frac{d_j}{\log_2 \left( 1 + \sum_{k \in \mathcal{I} \setminus \{i\}} p_k g_{kj} \rho_k + \sigma^2 \right)}
\]

The equations of (3.4) of all cells \( i \in \mathcal{I} \) form a non-linear equation system. The system can be addressed based on Perron-Frobenius theorem [5–27].

### 3.1.2 Load Coupling of NOMA

Consider the NOMA scenario, namely, one RU can be allocated to multiple users with SIC. We refer to the users sharing the same RUs as a group\(^1\). We use \( u \) to refer to a generic group. For any user \( j \in u \), denote by \( q_{ju} \) the portion of power used for user \( j \) on each RU allocated to group \( u \). For cell \( i (i \in \mathcal{I}) \), denote by \( \mathcal{U}_i \) the set of all groups of users in \( \mathcal{J}_i \). In analogy with this, we use \( \mathcal{U}_j \) to refer to the set of all groups that user \( j \) belongs to. In the most general case, for any cell \( i \) there can be \( 2|\mathcal{J}_i| - 1 \) groups in \( \mathcal{U}_i \). And for any user \( j \) served by cell \( i \), \( \mathcal{U}_j \) is a subset of \( \mathcal{U}_i \).

The SINR of user \( j \) in group \( u \) is as follows,

\[
\text{SINR}_{ju} = \frac{q_{ju}g_{ij}}{\sum_{h \in u} q_{hu}g_{ij} \theta_{hj} + \sum_{k \in \mathcal{I} \setminus \{i\}} p_k g_{kj} \rho_k + \sigma^2}
\]

\(^1\)We also use the term “user pair” or “pair” when RUs are shared by two users.
where $\theta_{hj}$ is a binary indicator: $\theta_{hj} = 1$ if and only if user $j$ can decode the signal intended for user $h$. Note that whether the decoding can be successfully performed or not depends on the channel condition, inter-cell interference, and noise. To be specific, for two users $j$ and $h$ served by cell $i$, if

$$\sum_{k \in \mathcal{I}\setminus\{i\}} P_k g_{kj} \rho_k + \sigma^2 \leq \sum_{k \in \mathcal{I}\setminus\{i\}} P_k g_{kh} \rho_k + \sigma^2,$$

then user $j$ can decode user $h$. Define

$$w_j = \left( \sum_{k \in \mathcal{I}\setminus\{i\}} P_k g_{kj} \rho_k + \sigma^2 \right) / g_{ij}$$

(3.6)

Then $\theta_{hj} = 1$ if and only if

$$w_h \geq w_j.$$  

To satisfy the demand of user $j$, we have

$$\sum_{u \in \mathcal{U}_j} \log_2 \left( 1 + \frac{q_{ju} g_{ij}}{\sum_{h \in \mathcal{U}_u} q_{hu} g_{ij} \theta_{hj} + \sum_{k \in \mathcal{I}\setminus\{i\}} P_k g_{kj} \rho_k + \sigma^2} \right) x_u \geq d_j$$

(3.7)

where the resource allocation in cell $i$ is computed by

$$\rho_i = \sum_{u \in \mathcal{U}_i} x_u$$

(3.8)

and the sum of powers of all users that reuse RUs in group $u$ does not exceed the total power on the RU, namely,

$$\sum_{j \in u} q_{ju} \leq P_i$$

(3.9)

The decoding order $\theta$ for any $h$ and $j$ is determined by $w_h$ and $w_j$, i.e.,

$$\theta_{hj} = \begin{cases} 1 & \text{if } w_h \geq w_j \\ 0 & \text{otherwise} \end{cases}$$

(3.10)

Mathematically, (3.6)-(3.10) form the load coupling system of NOMA. Unlike OMA, the load coupling system of NOMA is no longer an equation system of the cell load levels, rather it forms a performance region, expressed in terms of power split $q_{ju}$, time-frequency resource allocation $x_u$, and decoding order $\theta_{hj}$. From another perspective, finding the minimum required cell load for satisfying user demands amounts to optimizing these variables. We remark that these new dimensions for optimization do not exist in the case of OMA. Indeed, the load coupling system of OMA is a special case of that of NOMA: Without SIC, one RU can be allocated to at most one user, such that $x_u$ can be replaced by $x_j$. Also, in the SINR formula, the intra-cell interference term disappears, and $q_{ju}$ can be replaced by $p_i$.

\footnote{Strictly speaking, if $w_h = w_j$, then either $\theta_{hj} = 1$ or $\theta_{jh} = 1$ holds.}
3.2 Existing Solutions and Challenges

The load coupling system is a model of multi-cell scenarios for characterizing the dependency of resource allocation and inter-cell interference. In this dissertation, the model serves as a basis for joint optimization of power and time-frequency resource with user bits demand requirement. One of the difficulties in solving the load-coupling system is its high non-linearity. Previous literature, based on Perron-Frobenius Theorem, have shown that the load-coupling system of OMA can be solved by fixed-point iterations with geometric convergence rate [56]. Let \( \rho \) be the load vector of all cells, and \( \rho_{-i} \) the vector without cell \( i \). Denote by \( f_i(\rho_{-i}) \) the right-hand side of (3.4), i.e.

\[
f_i(\rho_{-i}) = \frac{d_j}{\log_2 \left( 1 + \frac{p_i g_{ij}}{\sum_{k \in \mathcal{I} \setminus \{i\}} p_k g_{kj} \rho_k + \sigma^2} \right)}
\]

(3.11)

Denote by \( f \) the vector form of \( f_i \) (\( i \in \mathcal{I} \)). The solution of the non-linear equation system \( \rho^* = f(\rho^*) \), if exists, is proved to be unique, and can be solved by fixed point iterations. Namely, the iterative process \( \rho_i^{(k+1)} = f(\rho_i^{(k)}) \) (\( k = 0, 1, 2, \ldots \infty \)) converges to the unique fixed point, with arbitrary non-negative starting point \( \rho_i^{(0)} \) [8].

The dissertation investigates two relatively new transmission techniques with the load coupling model of OMA.

- One is JT with CoMP, and its potential in C-RAN. JT enables one user to be served simultaneously by multiple cells, and in C-RAN scenarios, multiple remote radio heads (RRHs). With JT, the signal in the numerator of the SINR in (3.1) is enhanced. Besides, for any user \( j \), it does not receive interference from its serving cells, and hence the inter-cell interference in the denominator reads \( \sum_{k \in \mathcal{I} \setminus \mathcal{I}_j} p_k g_{kj} \rho_k \), where \( \mathcal{I}_j \) is the set of cells serving user \( j \) via JT. The load levels of users are more tightly coupled in this case, since a user can be tied to more than one cell.

- The other is relay technique with wireless backhaul. In this case, the time-frequency resource occupancy in a cell is due to not only serving users, but also for backhaul transmission. The resource allocation variable is link-wise instead of being user-wise; the latter is represented in (3.3).

More details of the two cases above can be found in Papers I–III in Part II. The fixed-point iteration based method for solving the load coupling model of OMA serves as the keystone in addressing scenarios with JT and wireless relay.

Analysis of the load coupling system of NOMA is more challenging, because \( \rho_i \) does not submit to a closed form formula of \( \rho_{-i} \). In addition, from an optimization’s perspective, the load coupling system of NOMA is much more complicated than OMA, since the decoding order of each group is dependent on the inter-cell interference and hence the resource allocation of other cells. In any algorithmic framework that is based on iterating on variable values, any
change of resource allocation may alter the decoding order. Therefore, the resulting algorithm is hard to analyze in terms of optimality and convergence.

The fixed-point iterations based framework does not apply directly to NOMA with SIC. The dissertation proposes a new optimization framework for load coupling in multi-cell NOMA. Namely, define the function $f_i(\rho_{-i})$ as follows.

$$f_i(\rho_{-i}) = \min_{q,x \geq 0, \theta \in \{0,1\}} \rho_i \text{ s.t. (3.6)-(3.10)} \quad (3.12)$$

This dissertation reveals that, $f_i$ in (3.12) possesses the same property as that in (3.11). The dissertation investigates multi-cell NOMA resource optimization, with evaluation and demonstration of the performance of multi-cell NOMA with respect to resource efficiency and load balancing. On top of that, the dissertation reveals that a class of multi-cell NOMA optimization problems are tractable as long as their single-cell version is.

We remark that solving $\rho = f(\rho)$ for both (3.11) and (3.12) is indeed finding the minimum time-frequency resource allocation for satisfying the demands of users. It naturally serves as the keystone for addressing all resource optimization problems in this dissertation. To the best of our knowledge, these results remained unknown before the dissertation.
4. Scope and Contributions

This dissertation aims at investigating the potential of mathematical optimization in addressing resource optimization in multi-cell scenarios with several evolving technologies. Cell load coupling serves as a unified interference modeling approach for all the papers included in this dissertation. Papers I–III focus on multi-cell OMA scenarios. The papers have mainly addressed the joint optimization of user association and power/time-frequency resource allocation. Papers IV–VI address multi-cell resource optimization of NOMA. Specifically, Paper I addresses the question of how network planning and coordination may increase the efficiency of spectrum usage, by jointly optimizing user association and resource allocation with CoMP transmission. Paper II investigates the potential of relay cooperation for energy saving. As an extension of Paper I, Paper III studies the capacity maximization for any given group of users, while keeping the bits demand of other users being strictly met. Paper IV proposes an algorithmic framework for optimizing multi-cell NOMA. Under this framework, Paper V explores the potential of NOMA networks, and proposes an exact algorithm for a class of joint optimization problems of power allocation, user pair selection, and time-frequency resource allocation. Paper VI, serving as a complementary note, overcomes a key obstacle for modeling the dependency of decoding order and inter-cell interference in multi-cell NOMA. Paper VI proves that a restriction in Paper IV and Paper V can be dropped without affecting the main conclusions.

In all of these papers, the main idea, the core concept of system modeling and problem formulation, as well as the fundamentals of theoretical results, are derived by the discussions among all the authors. The author of this dissertation serves as the first and the corresponding authors of all the six papers, who has mainly taken the responsibility of the development of both theoretical and experimental results, including but not limited to algorithm design, lemma/theorem proofs, simulations/benchmarking, as well as writing. The papers and the main scientific contributions are summarized as follows.

**Paper I:** Load Optimization with User Association in Cooperative and Load-Coupled LTE Networks

The paper addresses the problem of optimizing user association for time-frequency resource efficiency and load balancing in OFDMA cellular networks. First, the paper explores the potential of using JT techniques for delivering users’ bits demands more efficiently. Second, the paper validates the
effectiveness of JT for two optimization problems, sum load minimization and maximum load minimization. The paper proves that both problems possess \( \mathcal{NP} \)-hardness. Third, a mixed integer linear programming based scheme is proposed for problem solving, which is essentially based on a linearization of the OMA load coupling system. This approach also enables a bounding scheme for performance benchmarking. Finally, the paper derives a set of partial optimality conditions, the fulfillment of which guarantees solution improvements for the two optimization problems. Experimental results show the effectiveness of the proposed solution methods. The paper suggests that, under certain circumstances, optimization problems with non-linear load coupling constraints can be effectively addressed with linear approximation with near-optimal performance. The optimality conditions also serve as one of the foundations for association optimization in Paper II and Paper III.

The paper has been published in *IEEE Transactions on Wireless Communications*.

**Paper II:** Energy-Aware Wireless Relay Selection in Load-Coupled OFDMA Cellular Networks

The paper investigates network energy minimization via optimizing wireless relay selection in OFDMA networks. The paper accounts for the impact of load levels of both cells and relays on transmission energy. The paper proves the hardness in terms of the tractability of the energy-aware relay selection problem. To tackle the computational complexity, the paper derives theoretical results that provide insights in designing an effective and efficient algorithm. Numerically, the proposed method exhibits good performance in energy saving.

The paper has been published in *IEEE Communications Letters*.

**Paper III:** User-Centric Performance Optimization with Remote Radio Head Cooperation in C-RAN

The paper focuses on the scenarios of C-RAN, where distributed remote radio heads (RRHs) are coordinated by baseband units (BBUs) in the cloud. The centralization of signal processing provides flexibility for CoMP of RRHs to cooperatively serve users. The paper aims at enhancing capacity performance by jointly optimizing RRH-user association and time-frequency resource allocation. The paper analyzes the computational complexity of the problem. As an extension of Paper I, the paper derives a method towards optimizing the user-centric performance with user differentiation, such that the capacity of a target group of users can be scaled up, while the other users’ bits demands are met exactly. The paper proves that, under fixed RRH-user association,
optimally scaling up the capacity for any target group amounts to solving a so-called iterated function. Based on this result, an algorithm for joint RRH-user association and time-frequency resource allocation is proposed.

The paper is in minor revision for *IEEE Transactions on Wireless Communications*.

**Paper IV: A Framework for Optimizing Multi-Cell NOMA: Delivering Demand with Less Resource**

The paper provides a general framework for user grouping and power allocation, taking into account inter-cell interference, for optimizing resource allocation of NOMA in multi-cell networks. To be specific, in order to deliver the user demand with most efficient time-frequency resource usage, optimization has to be performed along several dimensions: user grouping for RU reuse, power split among the users in a group, and the portions of RUs allocated to each user group. Under a restricted setup of user grouping and power split, the paper provides a series of theoretical analysis, to algorithmically enable optimization approaches. The resulting algorithmic notion supports a general conclusion. Namely, for any performance metric that is monotonically increasing in the cell load, the convergence towards global optimum is guaranteed via the load coupling system. Numerical results demonstrate the gain of NOMA in achieving significantly higher efficiency in delivering demands to users.

The paper has been published in *Proceedings of IEEE GLOBECOM 2017*.

**Paper V: Resource Optimization with Load Coupling in Multi-Cell NOMA**

The paper explores the potential of NOMA networks in achieving optimal resource utilization, without the restriction of power split imposed by Paper IV. Towards this goal, the paper investigates a broad class of problems consisting in optimizing power allocation and user pairing for RU reuse for any cost function that is monotonically increasing in cell load. The paper proposes an algorithm that is guaranteed to converge to global optimality for this problem class. Numerically, the paper evaluates and demonstrates the performance of multi-cell NOMA in terms of resource efficiency and load balancing. The results show that sub-optimality due to optimizing power allocation and user pairing separately may affect the performance of NOMA, whereas significant improvements are obtained by optimizing both jointly.

The paper has been published in *IEEE Transactions on Wireless Communications*.
Paper VI: A Note on Decoding Order in Optimizing Multi-Cell NOMA

This technical note presents a theoretical result for resource optimization with NOMA. We remark that both Paper IV and Paper V impose a restriction in user grouping such that the decoding order for any group in the optimization problem is independent of inter-cell interference. The algorithmic framework in both papers relies on the use of fixed-point iterations across cells. One difficulty here is that the order of decoding for SIC in NOMA is generally not known a priori. This is because the decoding order in one cell depends on interference, which, in turn, is governed by resource allocation in other cells, and vice versa. The restriction of user grouping is a workaround to deal with this difficulty. The technical note derives and proves the following result: The convergence is guaranteed, even if the order changes over the iterations. The result not only waives the need of the previous workaround, but also implies that a wide class of resource optimization problems for multi-cell NOMA is tractable, as long as that for single cell is.

The paper is submitted to IEEE Transactions on Wireless Communications.
References


[41] Y. Fu, Y. Chen, and C. W. Sung, “Distributed power control for the down-
link of multi-cell NOMA systems,” *IEEE Transactions on Wireless Communications*, to appear.


[54] L. You, L. Lei, D. Yuan, S. Sun, S. Chatzinotas, and B. Ottersten, “A


Part II: Research Papers
Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1843

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)