Device-to-Device Communications for Future Cellular Networks: Challenges, Trade-Offs, and Coexistence

SERVEH SHALMASHI

Abstract

The steep growth in mobile data traffic has gained a lot of attention in recent years. With current infrastructure deployments and radio resources, operators will not be able to cope with the upcoming demands. Consequently, discussions of the next generation of mobile networks, referred to as the fifth generation (5G), have started in both academia and industry. In addition to more capacity, stringent requirements for improving energy efficiency, decreasing delays, and increasing reliability have been envisioned in 5G. Many solutions have been put forward, one of them being device-to-device (D2D) communications where users in close proximity can transmit directly to one another bypassing the base station (BS).

In this thesis, we identify trade-offs and challenges of integrating D2D communications into cellular networks and propose potential solutions. To maximize gains from such integration, resource allocation and interference management are key factors. We start by introducing cooperation between D2D and cellular users in order to minimize any interference between the two user types and identifying the scenarios where this cooperation can be beneficial. It is shown that an increase in the number of cellular users within the coverage area and in the size of the cell is associated with a higher probability of cooperation. With this cooperation, we can potentially increase the number of connected devices, reduce the delay, increase the cell sum rate, and offload an overloaded cell.

Next, we consider D2D communications underlaying the uplink of cellular networks. In such a scenario, any potential gain from resource sharing (time, frequency, or space) is determined by how the interference is managed. The quality and performance of the interference management techniques depend on the availability of the channel state information (CSI) and the location of nodes as well as the frequency of updates regarding such information. The more information is required, the more signaling is needed, which results in higher power consumption by the users. We investigate the trade-off between the availability of full CSI, which necessitates instantaneous information, and that of limited CSI, which requires infrequent updates. Our results show that with limited CSI, a good performance (in terms of the sum rate of both user types) can be achieved if a small performance loss is tolerated by cellular users. In addition, we propose a novel approach for interference management which only requires the information on the number of D2D users without any knowledge about their CSI. This blind approach can achieve a small outage probability with very low computational complexity when the number of scheduled D2D users is small.

We then study the problem of mode selection, i.e., if a user should transmit in the D2D mode or in the conventional cellular mode. We identify the decision criteria for both overlay and underlay scenarios with two different objectives. We find out

that the D2D communication is beneficial in macro cells or at cell boundaries. The area in which D2D mode is optimal varies with the objective of the network, transmit power, required quality-of-service, and the number of BS antennas.

In the second part of this thesis, we study the effects of integration and coexistence of underlay D2D communications with another promising technology proposed for 5G, namely massive multiple-input-multiple-output (MIMO). Potential benefits of both technologies are known individually, but the possibility of and performance gains from their coexistence are not adequately addressed. We evaluate the performance of this hybrid network in terms of energy efficiency and the average sum rate. Comprehensive analysis reveals that the performance highly depends on the D2D user density. We conclude that underlay D2D communications can only coexist with massive MIMO systems in the regime of low D2D user density. By introducing a high number of D2D users, gains from the massive MIMO technology degrade rapidly, and therefore in this case, the D2D communications should use the overlay approach rather than the underlay, or the network should only allow a subset of D2D transmissions to be active at a time.

Sammanfattning

Den stora ökningen i mobildatatrafik de senaste åren har tilldragit sig mycket intresse. Med nuvarande infrastruktur och radioresurser kommer inte mobiloperatörerna att kunna hantera de kommande kraven. Därför har diskussioner kring den femte generationens (5G) mobila nätverk startat inom både akademin och industrin. Utöver högre kapacitet så kommer strikta krav på ökad energieffektivitet, minskad fördröjning samt ökad tillförlitlighet att planeras för 5G. En av många lösningar som har föreslagits är enhet-till-enhetskommunikation (device-to-device communications, D2D, på engelska), vilket innebär att närliggande mobilanvändare kan sända direkt till varandra utan att gå genom basstationen.

I denna avhandling identifierar vi kompromisser och problem kring, samt föreslår lösningar för, integrering av D2D-kommunikation i cellulära nätverk. Viktiga faktorer för att maximera vinsten av sådan integrering är resursallokering och störningshantering. Avhandlingen börjar med att beskriva samarbetet mellan D2D- och cellulära användare för att minska störningen mellan de två användartyperna, samt för att identifiera scenarier där denna typ av samarbete kan vara fördelaktigt. Vi visar att samarbetssannolikheten ökar med antalet cellulära användare i täckningsområdet, samt när cellstorleken ökar. Denna typ av samarbete kan användas för att öka antalet anslutna enheter, minska fördröjningen, öka cellsummadatatakten eller avlasta överlastade celler.

Härnäst studerar vi D2D-kommunikation underliggande upplänken i cellulära nätverk. I ett sådant scenario bestäms eventuell vinst från resursdelning (t.ex. i tid, frekvens eller rymd) av hur störningen hanteras. Kvaliteten och prestandan hos störningshanteringen beror på tillgängligheten av kanalkännedom och information om nodernas position, samt uppdateringsfrekvensen för dessa. Ju mer information som behövs, desto mer signalering krävs, vilket leder till högre effektförbrukning hos användarna. Vi undersöker kompromissen mellan fullt tillgänglig kanalkännedom, vilket kräver momentan information, och ett scenario där kanalkännedomen är begränsad, vilket enbart kräver uppdatering med låg frekvens. Våra resultat visar att god summadatatakt kan uppnås när enbart begränsad kanalkännedom är tillgänglig, om en liten prestandaförlust tillåts för cellulära användare. Vi föreslår dessutom en ny metod för störningshantering som enbart kräver information om antalet D2D-användare, utan vetskap om deras kanalkännedom. Denna blinda metod kan uppnå hög täckningssannolikhet med låg beräkningskomplexitet när antalet schemalagda D2D-användare är lågt.

Vi studerar även lägesvalsproblemet, dvs. om en användare ska sända i D2D-läge eller i konventionellt cellulärt läge. Vi karaktäriserar beslutskriterierna för både överliggande och underliggande scenarier med två olika objektivfunktioner och visar att D2D-kommunikation är fördelaktig i makroceller samt vid cellkan-

terna. Området för D2D-optimalitet varierar med objektivfunktionen för nätverket, sändareffekten, servicekvalitetskraven och antalet basstationsantenner.

I den andra delen av avhandlingen så studerar vi effekter kring integrering och samexistens av underliggande D2D-kommunikation med en annan lovande teknologi för 5G, nämligen massiv multiple input-multiple output (massiv MIMO). De individuella fördelarna för de två teknologierna är välkända, men eventuella prestandavinster när teknologierna samexisterar har inte studerats tillräckligt. Vi undersöker prestanda i detta hybridnätverk i termer av energieffektivitet och genomsnittlig summadatatakt. En noggrann analys visar att prestandan beror på tätheten av D2D-användare. Vi drar slutsatsen att underliggande D2D-kommunikation bara kan samexistera med massiv MIMO när tätheten av D2D-användare är låg. När det existerar många D2D-användare minskas prestandavinsten från massiv MIMO snabbt och därför bör D2D-kommunikationen ske i överliggande läge istället för underliggande läge. Alternativt kan nätverket tillåta att enbart en delmängd av D2D-sändningar är aktiva samtidigt.

Acknowledgments

It is almost the end of an amazing and at the same time challenging and rewarding four-year journey in my life. There were many moments that I could not see a light at the end of the tunnel, but I am glad my perseverance as well as encouragements, inspirations, educations, and supports that I received along the way have brought me to this point. I want to take this opportunity to acknowledge those who contributed.

First, I am grateful to my main advisor Dr. Ki Won Sung for being a supportive person and providing me with many opportunities to learn from. I am thankful to my co-adviser Prof. Jens Zander for being an inspiring and a fair mentor. I always counted on his wisdom when I needed any advice. Furthermore, I would like to thank Assoc. Prof. Ben Slimane for facilitating my start at KTH and for our collaborations during my study.

I am thankful to my co-authors for our discussions and collaborations. I am particularly grateful to Asst. Prof. Emil Björnsson for our fruitful discussions, for his friendly advice, and for being a source of inspiration for high-quality work. His insights and passion to teach, especially in the area of MIMO, have contributed a lot to my thesis. Emil generously shared his time and knowledge with me as if I am his own student; I really appreciate his sense of responsibility. I would like to extend my gratitude to Assoc. Prof. Marios Kountouris for his openness in sharing and discussing ideas. I have learned a lot from his in-depth knowledge and experience in the area of stochastic geometry. I am grateful to Prof. Mérouane Debbah for hosting me in his group at Supélec and our collaborations. I would also like to thank everyone in the Large Networks and Systems Group at Supélec who made my stay in Paris enjoyable. Ericsson Research Foundation is acknowledged for partially funding my visit to Supélec.

I would like to further acknowledge Assoc. Prof. Guowang Miao and Assoc. Prof. Zhu Han for our collaborations, and Göran Andersson for helping me with Mathematica. I would like to appreciate the help of Hossein Shokri-Ghadikolaei and Farshad Naghibi for proofreading part of my thesis. Their positivity and selflessness helped me finally finish the thesis! I would also like to thank Rasmus Brandt for the Swedish translation of the abstract and for being amazingly quick and a welcoming person. I extend my gratitude to Dr. Hamed Farhadi for his openness to discussion and generously willing to answer any questions. I am very grateful for all his help. Many thanks to the administrative team, in particular Sara Winther and Csaba Goszleth who make our working environment pleasant and easy.

I am grateful to Dr. Magnus Frodigh for reviewing my PhD proposal and thesis. His valuable comments have greatly improved the thesis. I would like to acknowledge the efforts by Assoc. Prof. Jose F. Monserrat for acting as faculty opponent, and by

the grading committee formed by Assoc. Prof. James Gross, Assoc. Prof. Elisabeth Uhlemann, and Dr. Claes Tidestav.

I have enjoyed the friendly research environment in the Radio Systems Lab—thanks to my former and present colleagues: Dr. Sibel Tombaz, Dr. Evanny Obregon, Dr. Du Ho Kang, Dr. Lei Shi, Dr. Ali Özyagci, Amirhossein Ghanbari, Andrés Laya, Ashraf Widaa, Tatjana Apanasevic, Luis Martinez, Amin Azari, Haris Celik, Yanpeng Yang, Assoc. Prof. Jan Markendahl , Prof. Claes Beckman, and Mats Nilson.

Lunch time with Mozhgan and Forough has always been a pleasant break from work. Thanks for always being there for me. Outside of my academic life, I am blessed to have many friends who have shared many happy moments with me. Thank you for your encouragements and supports. I am sincerely grateful to my yoga family who listened to me, uplifted me, and made me smile during the last year.

I owe everything I have to endless love of my parents Manijeh and Esmael, my brother Souran and my lovely sisters Sorour and Sonia. I am very grateful to have you all. Special thanks to my extended family in Sweden, Jahan, Maliheh and Atin. Finally, I want to thank a precious person in my life, Farshad, for his love that warms me up and his support during uncountable weekends and late nights that he spent with me in my office to keep me going.

Serveh Shalmashi
Stockholm, May 2015

Contents

Part I

Thesis Overview

Chapter 1

Introduction

During the past decade, the volume of mobile data traffic has increased at a rapid pace and quantitative studies predict that the exponential growth will continue in the future as illustrated in Figure 1.1. The growth is mainly due to emerging popular multimedia applications that are supported by new types of devices such as smartphones and tablets [Eri13, Eri12]. Moreover, multiple devices may be used by the same user to connect to the Internet through the existing cellular infrastructure, which contributes to increased data traffic [Rea10]. Consequently, the total mobile data traffic generated is predicted to have a 1000-fold increase by the year 2020 [HSS13]. This is extremely demanding in terms of network resources and link capacity.

Besides the issue of large data volume in the upcoming decade, user experience is also an important challenge. Current networks may offer good quality-of-service (QoS) in isolated areas, but they cannot meet the extreme capacity demands on future wireless systems in areas where they have to handle situations where users are located in close proximity to one another, such as shopping malls, festivals, stadiums, and even office buildings [PBM+13]. Users want to be connected anytime, anywhere. Increasing capacity and connectivity will translate into higher energy consumption and costs, which in turn are not economical or sustainable from operational perspective.

During the years, mobile broadband technologies have evolved. Long-term evolution (LTE) and LTE-Advanced systems, which have embodied the fourth generation (4G) networks, have reached a certain level of maturity. Now, we are on the verge of a transition into a state of fully connected society where high capacity is needed, but incremental changes in the current systems and technologies are not enough to make this transition [ABC+14]—fundamental changes are needed to handle future non-homogeneous networks as well as new trends in user behavior and applications such as high quality video streaming and augmented reality.

Therefore, discussions of a new standard have taken place in academia and industry in order to envision the needs and requirements of, and possible use cases for future networks, referred to as the fifth generation (5G). The exact definition of
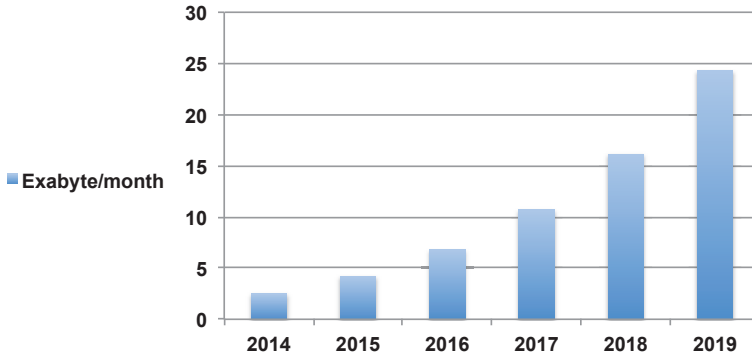
Figure 1.1: Mobile data traffic growth prediction [Cis15].

5G is not yet clear, but it needs to take into account a wider rang of use cases and characteristics. Therefore, stringent key performance indicators (KPIs) and tight requirements have been proposed in order to handle higher mobile data volumes, reduce latency, and increase the number of connected devices, while at the same time increasing energy efficiency (EE) and reducing costs [OBB+14,Eri13,BJDO14].

5G networks are supposed to support the existing and evolving technologies and simultaneously integrate new solutions which have been proposed to meet the new requirements [ABC+14, DMP+14]. In order to increase network capacity, one option is to improve the efficiency of available radio resources; another option is to increase resources such as the amount of available spectrum, the number of antennas and the number of base stations (BSs). However, adding radio resources is not necessarily cost and energy efficient, and it may sometimes take a long time for them to be put into practice. There are many new concepts, design criteria, and scenarios that have been proposed for 5G; some of them, if implemented, will bring fundamental changes at the architectural and node level. One example of such proposed technologies is device-to-device (D2D) communications which will change the nature of conventional network design.

In early generations of mobile systems, the network-centric design was introduced, based on the notion of cell, uplink, and downlink communications. At that time, the application of mobile networks was mainly for voice communication and there was an implicit assumption that users are not in close proximity to one another. However, this assumption is not tenable anymore as the main current trends are content (file) sharing and interest sharing (e.g., online-gaming and social networks, where users in close proximity happen to interact more). Hence, it is important to consider proximity awareness as a design parameter [BHL+14]. To this end, one of the broad visions of 5G is its emphasis on device-centric solutions and the need for smarter devices. D2D communication appears to be an enabling technology for this vision, which allows users in close proximity to communicate directly with each other, bypassing the base station (BS).

D2D communications can bring many benefits, It can potentially save some resources such as transmit power in the BS or mobile devices because of the direct short-range communication, specially if the user is located at the cell edge [FDM$^+$12]. If spectrum sharing between cellular and D2D users is allowed, it can improve the spectrum usage, resulting in a larger number of connected devices even in a highly utilized network. At the same time, user data rates and capacity per area unit would increase and the latency would be reduced [DRW$^+$09].

D2D communications which allow for the local management of short-range communications, can ease down the load of the backhaul and core network. They can also be a good enabler for creating caching or local data sharing zones, which again would lead to an increased number of connected devices in the network as well as higher cost efficiency [AWM14, LMU$^+$14]. Consequently, D2D communications can pave the way for massive machine-to-machine communications. Moreover, D2D communications can help extend the coverage where a mobile device relays the information of another out-of-range user to its destination. Therefore, a greater degree of reliability and availability can be achieved in the network. D2D communications are also envisioned to be an enabler for another technology, referred to as vehicle-to-vehicle (V2V) communications [LMU$^+$14].

In this thesis, we will focus on the integration of D2D communications in future wireless networks.

1.1 D2D Communications

In the conventional cellular transmission mode, the user equipment (UE) first transmits its data to the BS using uplink resources; then the BS forwards the data to the corresponding receiver using downlink resources. However, if the transmitting UE and the receiving UE are in close proximity to each other, the BS can allow the users to directly communicate with each other. This direct transmission mode is referred to as the *D2D mode* [FDM$^+$12, SBSD14].

D2D communications can be integrated into cellular networks in different ways. In terms of spectrum resources, they are divided into two categories: [MHR14b, AWM14]:

- *Inband communications*, in which D2D users can use the same licensed spectrum as cellular user equipments (CUEs). This category is further divided into *overlay* and *underlay* transmissions. That is, depending on the intended application, D2D communications can use dedicated resources (time/frequency), i.e., the overlay approach, or reuse the resources of other CUEs in the cell, i.e., the underlay approach. The allocation of dedicated resources is important for applications such as multi-casting and public safety, whereas resource sharing can improve efficiency of the available resources.

- *Outband communications*, in which D2D users use the unlicensed spectrum, such as the industrial, scientific, and medical (ISM) bands, for their transmis-
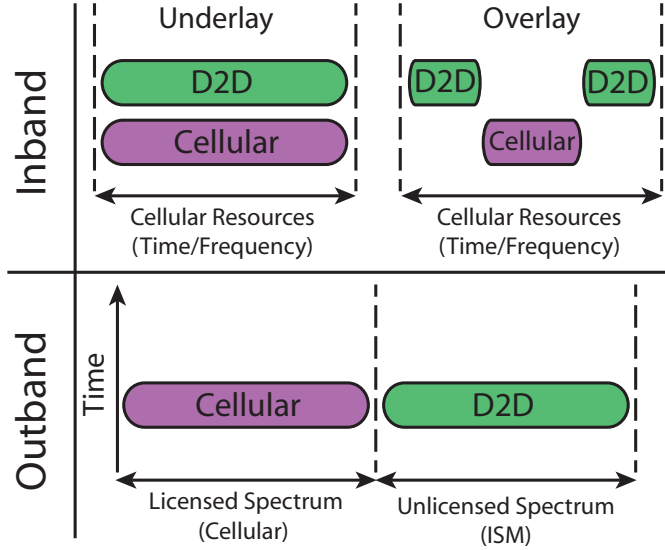
Figure 1.2: Spectrum resource for D2D communications [AWM14].

sions. This, on the one hand, results in the elimination of interference to and
from CUEs and, on the other hand, decreases the network control over D2D
communications. In addition, D2D communications need to adapt to other
technologies transmitting in the same unlicensed band.

A schematic view of how D2D users can access the spectrum of cellular users is
illustrated in Figure 1.2. In terms of network control, D2D communications are
divided into two categories:

- *Network-assisted communications*, in which the infrastructure node (i.e., the
  BS) assists with radio resource management, device discovery, establishing
  D2D connections, mobility management, and security issues. In this thesis,
  we will focus on network-assisted D2D communications.

- *Autonomous communications*, in which, as in the Ad-hoc networks, the BS
  has no control over the D2D communications. The autonomous D2D commu-
  nications can be used in case of network failure or when there is no coverage.

Many previous studies have proposed integration of short-range communications
into the infrastructure network. Two examples of these can be found in the con-
text of mobile Ad-hoc networks (MANET) and cognitive radio networks (CRN),
for which both pros and cons are well-studied, e.g,. see [HS04, ZdV05, GJMS09].
Furthermore, the concept of mobile relays to forward information to other mobiles
was already studied in [WCDT01,KLA13]. Although short-range communication is
not a new concept, D2D communication has only recently gained momentum since

it was proposed in the 3rd generation partnership project (3GPP) LTE standard meetings for public safety in case of network failure. It is now considered to be one of the system concepts of the future 5G networks [HRTA14, LMU$^+$14]. From an architectural perspective, the D2D communication is similar to MANET and CRN. However, unlike MANET or CRN, the D2D communications envisioned for 3GPP and 5G networks is inband network-assisted D2D communication where the network plays a major role. In MANET and CRN, the cellular network's resources are shared between two different systems, but in the D2D communication, they are shared between two different user types. Moreover, cellular networks are not oblivious to the D2D users, who are managed by the BS through the control plane; that is to say, the BS initiates, synchronizes, optimizes, and manages the resources for CUEs and between CUEs and D2D users.

Note that when MANET and CRN are used, lack of coordination makes it difficult for both systems to gain from resource sharing. For example, in CRN, spectrum sensing is very challenging and consumes a lot of power from cognitive radios (secondary users). In addition, the same lack of coordination in underlay CRN makes it a very difficult task to manage interference. Similarly, in MANET, collision avoidance and synchronization are important issues. The coordination provided by the BS for the inband network-assisted D2D communication makes it easier to handle these problems and results in a technology that is more appealing from a technical and business perspective. Other technologies that identify the need for close proximity communications are Bluetooth, Zigbee, and WiFi direct [AWM14]. However, they operate in the unlicensed band to which the users have no exclusive rights and where there is also no coordination among them. Then, synchronization and device discovery drain the batteries of devices quickly. Systems operating in unlicensed bands should use limited power and follow certain rules in order to manage interference. The range of operation is limited to a few meters especially for Bluetooth and Zigbee. However, these problems can be solved by the BS's control in network-assisted D2D communication. There are lots of scenarios where D2D communications can be beneficial in driving the 5G networks. It can be used to enable very critical applications like V2V communications and even can play a major role in integrating sensor networks into cellular networks as an instance of machine-type communications [LMU$^+$14].

Many studies in the literature so far have investigated the inband network-assisted D2D communications with an emphasis on the underlay scenario. Studies such as [JYD$^+$09, YTDR09, DYRJ10, DRW$^+$09, FDM$^+$12] considered the feasibility of D2D communications as an underlay to a cellular network and showed the potential gains in spectrum efficiency. Other potential benefits include saving energy at both the network and the user level as well as omitting one extra hop (i.e., the BS) in the transmission, thus reducing communication delays. However, questions remain to be answered about how this technology can be integrated in 5G networks and what changes are required to guarantee these gains.

In this thesis, we mainly consider network-assisted underlay D2D communications except in one section where we also study the overlay scenario.

1.2   Thesis Focus and Research Questions

Introducing D2D communications in well-planned cellular networks brings out new technical challenges, including new decision-making criteria for scheduling problems, radio resource management, interference management, coexistence with other techniques and technologies; all these issues should be tackled in order to guarantee the integration of D2D communications in cellular infrastructure. In addition, there are other challenges that are not within the scope of this thesis, such as changes in mobile device hardware for direct communication and device discovery.

In D2D overlaid networks, in particular, some issues needs to be addressed: how to partition the channels with existing cellular users, what the best operational mode is for each user, how many D2D users can transmit per channel, and how to avoid intra-interference between D2Ds. While in the underlay approach, the important challenges include finding new methods to deal with the extra source of interference, i.e., D2D transmissions, seeking new decision-making algorithms for scheduling and user pairing as well as determining the best mode of operation. Although interference management in D2D underlaid networks may not be easy, such networks allow for efficient reuse of the spectrum in spatial and temporal domains owing to the close proximity of users. Consequently, they can lead to increased potential gains in terms of capacity. This category of networks has been the focus of most of the studies in the literature.

Initial studies in this area deal with simple scenarios with one D2D user and one CUE in a single cell, and the results show the feasibility of underlay D2D communications. However, if multiple devices reuse the spectrum band of one CUE, the effect of signaling overhead and aggregate interference becomes more important in the scheduling decisions. Such effects have not been treated properly in those previous studies. Similarly, regarding mode selection problems, the criteria for decision-making are not made clear in the literature, and they are often based on very simple scenarios. Finally, the effects of integration and coexistence of D2D communications with other techniques and technologies that are commonly used in cellular networks have not been well investigated so far.

In what follows, we will consider two most important problems in D2D communications; we will use the existing approaches in more realistic scenarios in an attempt to answer a set of research questions relevant to each problem. The first part, mainly addresses the state of the art in current networks, whereas the second part tries to incorporate a broadened vision and studies the impacts of D2D communications on currently available solutions for 5G. What follows is the first high-level research question (HQ) that we try to answer:

- **HQ1**: What are the important trade-offs in radio resource management for integrating D2D communications into cellular networks and how should they be treated?

There are different techniques which handle radio resource scheduling and interference management in order to guarantee gains from spectrum sharing between

D2D and cellular users. These techniques are based on power control [YTDR09, JYD$^+$09, LLAH14], opportunistic medium access control [CK14], and developing guard zones [MLPH11] for any of devices that should be protected. Effective interference management and scheduling techniques depend on the information available in the nodes, such as channel state information (CSI) in the network and information that may be overheard by users in close proximity.

By introducing intelligence in devices, users can acquire information from other nodes and potentially reuse the extra information in an opportunist manner. By overhearing other users close-by in a crowded area, blind spots and coverage can be improved. In a crowded area where the resources are limited, cooperation between the cellular and D2D users can introduce an extra degree of freedom in order to manage interference and refrain users from transmitting at the same time. Therefore, we first consider a crowded-communication scenario and try to answer the following research question (RQ):

– **RQ1-1**: In which scenarios is cooperation beneficial and how much resources from D2D users should be allocated for cooperation in order to obtain the required gain (trade-off between cooperation and no cooperation)?

Many studies on underlay D2D communications consider interference control on the assumption that full CSI knowledge of all nodes in the network and users' location information are available at the BS, e.g., see [MLPH11, YDRT11]. However, such an assumption is not always practical, depending very much on signaling overhead and transmit power of the nodes which provide the BS with the CSI. Even if it is possible to acquire all information at the BS, the computational complexity to handle the optimization problems for decision-making may be too time consuming to be manageable. As a result, the optimization problems may not be scalable. In this regard, we investigate the following questions:

– **RQ1-2**: What is the trade-off between system performance and signaling overhead (trade-off between full CSI and limited CSI)?

– **RQ1-3**: Is it possible to omit the need for CSI in the scheduling process in order to minimize signaling overhead in the network while still gaining from D2D communications (trade-off between full CSI and no CSI)?

The finial aspect regarding HQ1 that we study is related to the operation mode of the user, i.e., cellular or D2D mode. The choice of an operation mode is closely connected with proper user pairing and scheduling, which in turn contribute greatly to the gains achieved from this type of communications. So far, in the literature, the operation mode has often been decided based on the distance between the user and the BS as well as between different users [DYRJ10], or it is considered as part of the resource allocation and scheduling process where the emphasis is on developing a joint radio resource management algorithm [WZZY13]. However, we investigate the effects of other parameters including the network's objective, the users' expected QoS, and the available resources at the BS such as transmit power and the number

of antennas, which have been neglected before. In particular, we try to answer the
following question:

 – **RQ1-4**: When is the D2D mode preferable to the cellular mode for the user
   in the network? What are the crucial parameters in this strategic decision
   (trade-off between D2D mode and cellular mode)?

As we mentioned earlier, 5G networks are supposed to integrate newly pro-
posed solutions and techniques to meet the stringent requirements envisioned for
5G [BJDO14]. In addition to D2D communications, other new concepts for 5G net-
works include massive MIMO (densification in terms of the number of antennas),
ultra dense networks (densification in terms of the number of BSs), and a huge
number of connected devices known as machine-type communications (densifica-
tion in terms of the number of devices). Potential benefits of these solutions are
known individually but not in combination. Especially, the possibility of these solu-
tions coexisting with one another is yet to be made clear. The evolution of cellular
networks has mainly aimed at achieving higher data rates. Now other objectives
are being considered for 5G networks, including improved coverage, reliability, scal-
ability and energy efficiency. In this study, we take into account energy efficiency,
which is the objective of the second high-level research question:

 • **HQ2**: How do the extra resources and degrees of freedom in the BS resulting
   from a large number of antennas impact the energy (EE) and average sum
   rate (ASR) in underlay D2D networks?

The insights obtained from investigating RQ1-4 motivate us to look further
into this matter, and therefore, we first study the coexistence of two technologies,
namely massive MIMO and D2D communications. The extra degrees of freedom
offered by having multiple antennas at BSs are highly desirable in the design of
future mobile networks, because many users can then be multiplexed and the inter-
user interference can be controlled. Furthermore, the performance of cell-edge users
can be greatly improved owing to the higher SNR [BDF+13, BJ13, GKH+07]. There
are some studies that focus on EE for D2D communications; however, they are lim-
ited to single-antenna BSs. These studies include [YK12, MHR+14a, WXS+13]. The
study [YK12] focuses on a coalition formation method, [MHR+14a] designs a re-
source allocation scheme that is energy efficient, and [WXS+13] aims at optimizing
the battery life of user devices. From RQ1-4, we know that different conclusions
are achieved, based on different network objectives. Thus, we study both EE and
ASR taking into account the number of BS antennas, the number of cellular users
and the density of D2D users within a given coverage area.

1.3   Thesis Outline and Contributions

This section provides an outline of the thesis and summarizes the main contri-
butions. This thesis consists of two parts: the first part comprises Chapters 1–5;

Chapters 2–5 will be described below. The second part includes the corresponding papers.

Chapter 2

This chapter gives an overview of the different models used and introduces the common assumptions on which the following chapters are based.

Chapter 3

This chapter considers trade-offs for integrating D2D communications into cellular networks, with a focus on radio resource management.

The first problem deals with a crowded communication environment where the number of users is higher than what the BS can actually support. Due to interference, spectrum sharing leads to a performance gain usually when users are well separated spatially. We propose to use the idea of spectrum sharing in the downlink of cellular networks, even when users are packed tightly together. This is possible if the D2D user cooperates with the BS in order to transmit the cellular user's data along with its own data. Due to the device's power limits, the objective is to minimize the D2D user's transmit power allocated for collaboration on the condition that the cellular user's performance does not degrade. We show the feasibility of such cooperation as well as the scenarios and parameters that lead to high performance gains. The first part of the chapter is based on the following paper:

- **Paper A:** S. Shalmashi and S. B. Slimane, "Cooperative Device-to-Device Communications in the Downlink of Cellular Networks," in *Proceedings of IEEE Wireless Communication and Networking Conference (WCNC)*, Istanbul, Turkey, April 2014.

The second problem deals with interference management in a scenario with multiple D2D communications reusing the uplink resource of a cellular user, based on the underlay paradigm. Therefore, the effect of aggregated interference of D2D users at the BS should be considered carefully in order to guarantee the QoS for the cellular user and the gains from D2D communications. However, the quality of D2D user scheduling depends on the available CSI at the BS. We study this problem in two scenarios. First, we formulate this problem with two constraints: (i) the instantaneous signal-to-interference-plus-noise ratio (SINR) constraint where full (instantaneous) CSI is available, and (ii) the SINR outage constraint where limited CSI is available. We show that there is a trade-off between the signaling overhead for acquiring CSI at the BS and the system performance. Second, in order to protect the BS, an aggregated interference constraint is considered and we formulate the problem with the objective that the number of D2D links is maximized. We study the problem with two different interference constraints: (i) the peak interference constraint (PIC) and (ii) the average interference constraint (AIC). In the former, the assumption is that the full CSI is available in all nodes, while in the latter, we assume that the BS has no knowledge of D2D users' locations and their CSI.

The solution of the first formulation is used as the baseline for comparison with performance results obtained in the second formulation. With the AIC, we derive an upper bound on the number of D2D users that can be admitted for simultaneous transmission. The performance results are then compared with those of the optimal solution obtained from the PIC. The bound in the AIC is very practical and cost-effective in terms of signaling overhead and transmit powers, and at the same time, it is computationally efficient, especially in a scenario where the BS receives a request for admission from a new D2D user. This problem is studied in the following papers:

- **Paper B:** S. Shalmashi, G. Miao, and S. B. Slimane, "Interference Management for Multiple Device-to-Device Communications Underlaying Cellular Networks," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, London, UK, September 2013.

- **Paper C:** S. Shalmashi, G. Miao, Z. Han, and S. B. Slimane, "Interference Constrained Device-to-Device Communications," in *Proceedings of IEEE International Conference on Communications (ICC)*, Sydney, Australia, June 2014.

The next problem deals with mode selection in a network where multiple antennas are deployed at the BS. We consider two scenarios regarding the resources for D2D communications. In the first scenario, resources for D2D communications are dedicated to the user, whereas in the second scenario resources are shared with the cellular user. Given the type of resources, dedicated or shared, we decide on the preferable mode of communication, i.e., the direct communication (D2D mode) or the communication through the BS (cellular mode). In addition, we formulate the optimization problem with two different objectives for each type of resources. The optimization problems are (i) maximizing the QoS given a constant transmit power, and (ii) minimizing the transmit power given a fixed QoS. In both cases, the optimal decision criteria are derived. In the case of dedicated resources, we find the area where the receiver should be located in order for the D2D mode to be optimal. Besides, we show that the size of this area is affected by the transmit power in problem (i), and the number of antennas and the pre-defined QoS in problem (ii). In the case of spectrum sharing, we derive an upper bound of interference that can be tolerated by the D2D receiver. We show that these two problems have different behaviors in terms of D2D optimality. This part of the chapter is based on the following paper:

- **Paper D:** S. Shalmashi, E. Björnson, S. B. Slimane, and M. Debbah "Closed-Form Optimality Characterization of Network-Assisted Device-to-Device Communications," in *Proceedings of IEEE Wireless Communication and Networking Conference (WCNC)*, Istanbul, Turkey, April 2014.

Chapter 4

This chapter considers D2D communications underlaying cellular networks as an enabling technology. We address the issue of coexistence between D2D communication and one of the other solutions proposed for 5G, namely massive MIMO. We assume that D2D users reuse the downlink resources of cellular networks in an underlay fashion. In addition, multiple antennas at the BS are used in order to simultaneously support multiple cellular users. The network model involves a number of cellular users who are randomly distributed in the cell and a number of D2D users who are distributed according to a homogeneous Poisson point process (PPP). Two metrics are considered, namely the average sum rate (ASR) and energy efficiency (EE). We derive tractable and directly computable expressions and study the trade-offs between the ASR and EE taking into account the number of BS antennas, and the density of D2D users within a given coverage area in two scenarios regarding the number of cellular users: (i) when the number of cellular users is fixed, (ii) when the number of cellular users is a function of the number of antennas. This chapter is based on the following papers:

- **Paper E:** S. Shalmashi, E. Björnson, M. Kountouris, K. W. Sung, and M. Debbah, "Energy Efficiency and Sum Rate when Massive MIMO meets Device-to-Device Communication," in *Proceedings of IEEE International Conference on Communications (ICC) Workshop on Device-to-Device Communication for Cellular and Wireless Networks*, London, UK, June 2015.

- **Paper F:** S. Shalmashi, E. Björnson, M. Kountouris, K. W. Sung, and M. Debbah, "Energy Efficiency and Sum Rate Tradeoffs for Massive MIMO Systems with Underlaid Device-to-Device Communications," submitted to *IEEE Transactions on Communications*, May 2015.

Chapter 5

This chapter concludes the thesis and discusses possible directions for future research.

1.3.1   Contributions Outside the Scope of this Thesis

The author of this thesis has also contributed to the following publications which are outside the scope of this thesis.

- **[TSM13]:** A. Thanos, S. Shalmashi, and G. Miao, "Network-Assisted Discovery for Device-to-Device Communications," in *Proceedings of IEEE Global Communications Conference (GLOBECOM) Workshop - Device-to-Device (D2D) Communication With and Without Infrastructure*, Atlanta, GA, December 2013.

- **[SS13]:** S. Shalmashi and S. B. Slimane, "On Secondary User Transmission Schemes in Relay-Assisted Cognitive Radio Networks," in *Proceedings of IEEE Vehicular Technology Conference (VTC Spring)*, Dresden, Germany, June 2013.

- **[SS12]:** S. Shalmashi and S. B. Slimane, "Performance Analysis of Relay-Assisted Cognitive Radio Systems with Superposition Coding," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sydney, Australia, September 2012.

Modeling and Methodology

In order to tackle our high-level research questions, as described in Section 1.2, we start by highlighting the commonalities in modeling approaches and methodologies that are used in Chapter 3 and Chapter 4.

2.1 Channel Model

Communications in wireless networks are limited by several factors, such as propagation environment, interference, and noise. There are a number of causes of signal attenuation in the wireless medium, including distance (known as path loss), large obstacles (known as shadowing), and the reception of multiple copies of the same signal which has been attenuated and phase shifted (known as multi-path fading). Furthermore, the environment varies as the users' positions are changed. The interference is caused by signals received from unintended transmitters, and the effect of the thermal noise stems from the receiver's electronics and is usually modeled as additive white Gaussian noise (AWGN). All radio resource management techniques depend on the amount of information available about channel impairments, the users' locations, and the frequency of updates (i.e., when there is a change, how fast the updated information is provided).

The channel model used in this thesis takes into account the effects of path loss, the multi-path fading, and in one scenario, the shadowing. The path loss model for D2D communications has not been standardized yet and we follow the model described in [XH10], which is based on the International Telecommunication Union's (ITU) recommendations for micro urban environments [IR09]. The path loss model is defined as

$$\text{PL} = A + 10\alpha \log_{10}(r), \tag{2.1}$$

where r is the distance between the transmitter and the receiver measured in meter. A and α are path loss coefficient and path loss exponent, respectively. As shown in [XH10], A is a function of the carrier frequency (f_c). The values of A and α are given in Table 2.1 for both line-of-sight (LoS) and non-line-of-sight (NLoS)

Table 2.1: Path loss parameters.

Device	type of PL	α	A
BS - UE	PL_{LoS}	2.2	34.04
BS - UE	PL_{NLoS}	3.67	30.55
UE - UE	PL_{LoS}	1.69	38.84
UE - UE	PL_{NLoS}	4	28.03

scenarios. The average path loss is calculated as [XH10]

$$\overline{PL} = \beta PL_{LoS} + (1 - \beta)PL_{NLoS}, \tag{2.2}$$

where β is the probability of line-of-sight, which for outdoor users between the BS and a device is defined as

$$\beta = \min\left(\frac{18}{r}, 1\right)\left(1 - \exp\left(\frac{-r}{36}\right)\right) + \exp\left(\frac{-r}{36}\right), \tag{2.3}$$

and between two devices as

$$\beta = \begin{cases} 1, & r \le 4, \\ \exp\left(-(r-4)/3\right), & 4 < r < 60, \\ 0, & r \ge 60. \end{cases} \tag{2.4}$$

In our results, either the two-slope model in [XH10] is used, or only the NLoS part is considered due to the complexity of calculations.

Log-normal shadowing is generated using a correlated model described in [ZK01, ZCS10]. The multi-path fading component is distributed according to the circularly symmetric complex Gaussian distribution with $\mathcal{CN}(0,1)$. We assume a Rayleigh block fading channel in which the channel remains constant during one time slot and varies over different time slots.

2.2 Uplink or Downlink Resources for D2D Communications?

In mobile broadband services, the spectrum allocated for cellular networks is licensed. One of the benefits of licensed spectrum is the network's ability to control interference and guarantee a certain level of quality-of-service for users. Cellular users are well scheduled in time, frequency, and space in order to minimize the inter- and intra-tier interferences. Communications in cellular networks takes place in two directions, namely the *uplink* (UL) and the *downlink* (DL). In the UL direction, the cellular user equipment (CUE) sends its data to the base station (BS) whereas in the DL, the BS forwards the data to the intended receivers. The DL and

Table 2.2: Scenarios for D2D communications considered in the thesis.

Scenario	Chapter 3				Chapter 4	
	Paper A	Paper B	Paper C	Paper D	Paper E	Paper F
Uplink resources		✓	✓	✓		
Downlink resources	✓				✓	✓

UL transmissions are separated in time or frequency in order to avoid any impact on each other. The techniques that are used for interference and radio resource management in cellular networks regarding both the DL and UL have evolved through the years.

In this thesis, we introduce inband underlay device-to-device (D2D) communications in cellular networks, which will change the characteristics of interference in such networks. We also consider overlay D2D communications in the first part of RQ1-4, where the D2D user is assigned a dedicated time frame for transmission. When D2D communications utilize the resources of cellular users in an underlay fashion, a natural question arises: "which resources should be reused for D2D transmissions, the UL resources of the cellular network or the DL resources?"

Practically, D2D communications can be enabled in either the UL or DL of a cellular network, or in both [DYRJ10]. But the characteristics and victims of the interference are different in each case. For instance, in the UL, the BS suffers from the interference of D2D users, but with the use of advanced signal processing techniques at the BS, better protection can be provided. At the same time, the transmission from CUEs can cause severe interference to the D2D users if they are located closely to each other. In the case of the DL, the transmission from the BS has a high impact on the D2D users, and at the same time, D2D communications interfere with the CUEs' reception. Most of the studies in the literature consider the D2D communications in the UL since there are some regulatory restrictions in certain regions for reusing the DL resources [FR11].

Analytically, however, it is often easier to handle the DL scenario than the UL due to the dependencies between the interference distribution and user locations [NDA13]. In this thesis, we study D2D communications in both the UL and DL scenarios, as shown in Table 2.2. Furthermore, we assume that the traffic model for both cellular and D2D users is full-buffer, i.e., they always have packets to transmit.

2.3 Network Modeling Approaches

The common models in the literature for underlay D2D communications can be categorized into two groups, based on their approach to deal with radio resource allocation and management problems: *instantaneous* approach and *statistical* approach. The former deals with the instantaneous system information such as chan-

nel gains and link distances in the problem formulations and assumes that all
the necessary information on the nodes is available at the BS for the decision-
making process. Based on this assumption, instantaneous optimal decisions are
made regarding power and channel allocations, criteria for mode selection, ad-
mission control, or any other scheduling issues. Examples of such models can be
found in [JYD$^+$09, YDRT11, DRW$^+$09, JKR$^+$09, DYRJ10, BFA11, FR11, MLPH11].
The latter approach, i.e., the statistical approach, exploits the system's statisti-
cal information, such as the distribution of locations of users and BSs as well as
the distribution of channel gains. Such information is assumed to be valid for a
longer time span than is the instantaneous information. The statistical information
is used to model the network, and accordingly statistically optimal decisions are
reached [LPXW12, LAG, YTDR09].

Each of the above two approaches has its own pros and cons and can be used for
specific purposes. The instantaneous approach is very useful in order to understand
the fundamental limits and potential gains of the network and its performance. It
can also be used in feasibility studies. However, finding the instantaneous optimal
decisions may involve high signaling overhead to exchange network information as
well as high computational complexity, and the approach suffers from scalability
issues. Therefore, suboptimal heuristic algorithms or distributed decision making
with local information are often used in practice. Results obtained from the instan-
taneous approach can be used as a basis for comparison, and for developing these
suboptimal algorithms.

In this thesis, in order to answer RQ1-1 and RQ1-4 in Chapter 3 (Paper A and
Paper D, respectively), we assume that the instantaneous channel information is
available at the BS, CUEs, and D2D users. RQ1-1 deals with opportunistic D2D
communications in congested-communication scenarios whereas RQ1-4 studies the
mode selection problem. In Chapter 4 (Papers E–F), which addresses the second
high-level research question, we assume that only the instantaneous channel infor-
mation of the CUEs is available at the BS. The rest of the thesis is based on the
statistical approach. To answer RQ1-2 and RQ1-3 in Chapter 3 (Paper B–C), we
study the trade-off between the availability of CSI and network performance, and
develop a blind scheduling algorithm that requires no information of the D2D users,
but that of the number of existing users in the cell.

Furthermore, in Chapter 4, we take advantage of a class of mathematical tools
from *stochastic geometry* in order to characterize the average performance of ran-
dom networks. In other words, the stochastic geometry tools can provide us with
information about the average performance of a network over all random topolo-
gies seen from a generic node weighted by their probability of occurrence. In certain
cases, a tractable analysis is possible in order to characterize the network perfor-
mance over random topologies, as opposed to the fixed hexagonal topology where
we have to deal with heavy Monte-Carlo simulations. In this type of modeling,
the locations of nodes or BSs are assumed to be points of a point process (PP)
and random. One common PP is the Poisson point process (PPP) [Hae13]. A PP,
$\Pi = \{x_i; i = 1, 2, 3, \dots\} \subset \mathbb{R}^d$ is a PPP if and only if the number of points inside

any compact set $\mathcal{B} \subset \mathbb{R}^d$ is a Poisson random variable, and the number of points in disjoint sets are independent. We use stochastic geometry in our modeling in order to study the second high-level research question (HQ2).

Trade-Offs for Integrating D2D Communications in Cellular Networks

Introducing device-to-device (D2D) communications in well-planned cellular networks can potentially improve spectrum utilization, boost energy efficiency and capacity, and reduce communication delays.[1] However, this integration requires careful radio resource management in order to guarantee the aforementioned gains resulting from such hybrid networks.

This chapter provides some key trade-offs in the design of radio resource management algorithms. The term trade-off refers to some interdependencies in the solution space where improving one direction can degrade the other. In particular, the performance of any radio resource management algorithm is heavily dependent on gathering and processing information, which can increase delays and energy consumption and should be kept to a minimum in mobile devices. In general, to induce better performance, a centralized approach at the base station (BS) with all the required information is employed to find globally optimal results. Note that better performance is tied to a higher load for exchanging and storing information. The corresponding solution may require lots of resources that might not be available, or it may be very costly and complex to be implemented. Therefore, the optimal decisions always have high dependencies on the overhead and scalability.

Integrating D2D communications in cellular networks further increases the amount of information exchange among different entities in the network. Therefore, key trade-offs in such networks should be investigated. Appropriate solutions could be put forward taking into account these trade-offs as well as the applications and dynamics of the considered scenarios.

To this end, we study four scenarios in this chapter assuming that the traffic model for both cellular and D2D users is full-buffer: First, we consider a crowded-communication scenario, as can be found in places such as shopping malls and

---

[1]Part of the material presented in this chapter is based on our work in [SMS13] (© 2013 IEEE), [SBSD14], [SS14], and [SMHS14] (© 2014 IEEE), which have been published, and on the author's Licentiate thesis [Sha14]. The material is reused with permission.

stadiums, where cellular user equipments (CUEs) may be in close proximity to one another and D2D users reuse the same resource as do CUEs. In order to reduce the complexity of interference management and avoid gathering lots of local information, we introduce cooperation between D2D and cellular users and investigate the trade-off between cooperation and no cooperation. Then, in order to address issues involved in interference management techniques for acquiring channel state information (CSI), such as complexity, scalability, and efficient power management, we move on to study trade-offs between the availability of full CSI, which requires instantaneous updates, and that of limited CSI, which requires infrequent information updates. Next, we investigate a centralized approach with access to full CSI and develop a novel algorithm which does not require any CSI at the BS. Finally, we study the trade-off between transmission in the D2D mode and that in the cellular mode, and we highlight key parameters that affect this decision-making process.

In what follows, we provide a summary of the modeling approaches used in this thesis and the key results regarding the first research question (HQ1) posed in Section 1.2. The details of the analysis and more results can be found in Papers A–D.

3.1   Cooperative D2D Communications

One of the challenging communication scenarios involves the crowded environments where there are many users in close proximity to one another in a small area. In such environments, the cellular network can easily become congested due to the high number of connections. Therefore, it is desirable to increase the area spectral efficiency and the number of connected devices per shared resources (time/frequency). In order to minimize interference among users, interactions and cooperation between users are beneficial. Cooperation requires that the cooperative entities follow the same protocol and have some common knowledge, or that they are willing to sacrifice some of their resources, such as power and time, in order to improve each other's performance.

Cooperative D2D communications can be a solution to the above scenario by allowing spectrum sharing between cellular links and direct D2D links. Cooperative D2D communications make use of the broadcast nature of the wireless channel in which users in close proximity can overhear each other's broadcasted information. In this scenario, we assume that the D2D transmitter can act as a relay to assist CUE's transmission, while at the same time having the opportunity to transmit to its own receiver.

If cooperation between a CUE and a D2D user is allowed, the downlink transmission is divided into two phases (time slots). In the first phase, the BS transmits (broadcasts) while both the CUE and D2D transmitter listen. If such cooperation is beneficial, then in the second phase, the D2D transmitter employs the superposition coding scheme to transmit to the intended CUE and its own receiver. In the superposition coding scheme, the D2D transmitter sends a linear combination of
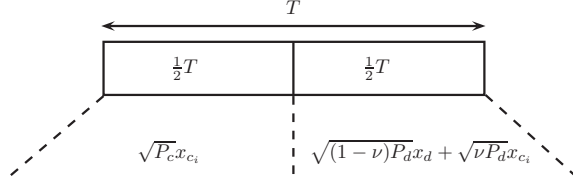
Figure 3.1: Frame structure with two equal-sized time slot.

its own signal and the intended CUE's signal, as shown in Figure 3.1. We assume that there is only one D2D pair which seeks an opportunistic cooperation with one of the M CUEs. Now, there are several questions that need to be addressed:

- How should the D2D user and CUE cooperate so that both systems benefit from such cooperation?

- Which CUE should be selected for cooperation?

- How much of the D2D user's transmit power (ν) should be assigned to the CUE's signal, which is the main concern for such cooperation?

To answer these questions, let R_{c_i} be the achievable rate for the ith CUE with cooperation and R_{dir} be the rate when there is no cooperation and only the BS transmits. Then, such cooperation would be beneficial for the CUE if

$$R_{c_i} \geq R_{\mathrm{dir}}. \tag{3.1}$$

Moreover, in order for the D2D user to be able to cooperate with a CUE, it should be able to decode the signal of that CUE in the first phase.

The D2D user can benefit more from such cooperation if it can spend less power on the CUE's signal, and consequently more power on its own signal. Therefore, the objective of the D2D user is

$$\underset{i \in \mathcal{A}}{\text{minimize}} \quad \nu_i, \tag{3.2}$$

where $\mathcal{A}$ is the set of CUEs that the D2D user can cooperate with, i.e., those CUEs that satisfy the condition in (3.1). For each cooperating CUE in the set $\mathcal{A}$, the smallest power fraction for relaying is the solution to the optimization problem (3.2), denoted by ν_i^*, for $i \in \mathcal{A}$. Among the set of CUEs, the one which needs the minimum relay power is chosen for cooperation, i.e.,

$$r = \underset{i \in \mathcal{A}}{\arg\min} \quad \nu_i^*, \tag{3.3}$$

where r is the index of the selected CUE. Since, the D2D transmitter and receiver are located in close proximity to each other, if the D2D transmitter can decode the
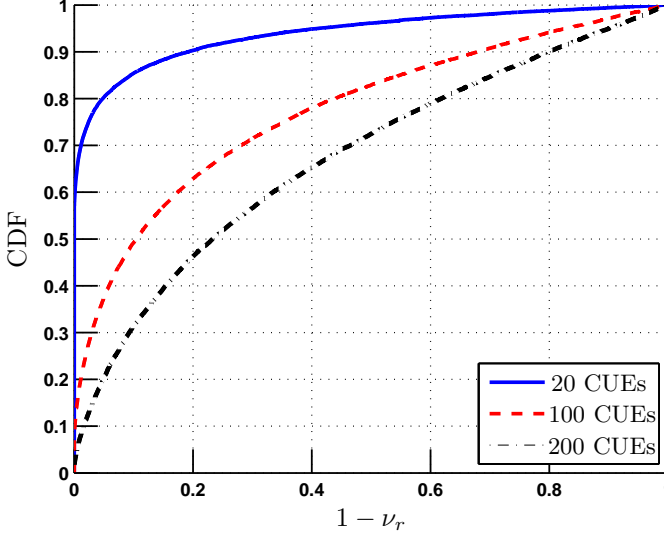
Figure 3.2: CDF of $1-\nu_r$ when $R = 200$ m. (© 2014 IEEE. Reused with permission.)

CUE's signal in the first phase, then it is highly probable that the D2D receiver can also decode it. If this is the case, the D2D receiver can cancel the effect of CUE's signal from the superimposed signal received in the second transmission phase, which in turn will improve the gain for the D2D user.

We evaluate the feasibility of our model with Monte-Carlo simulations. First, we study the amount of power that is used for the D2D link's communication, i.e., $1 - \nu_r$, with different numbers of available CUEs $M \in \{20, 100, 200\}$ and two different cell sizes. $1 - \nu_r = 0$ corresponds to the case where cooperation is not possible (beneficial) since the D2D user should allocate all its transmit power for relaying ($\nu_r = 1$). Figure 3.2 depicts the cumulative distribution function (CDF) of $1 - \nu_r$ when the cell radius is $R = 200$ m. As the results indicate, the probability of cooperation is a function of the number of CUEs that are available in the area. For instance, when the density of CUEs is small (e.g., $M = 20$), in almost 60% of realizations, beneficial cooperation is not possible. However, by increasing the density of CUEs, to, for instance, $M = 200$, successful cooperation is achieved in almost 98% of instances. Note that when the density of CUEs is low, the probability is small that a CUE could be found which has lower requirements on its achievable rate.

Changing the cell radius to $R = 500$ m in Figure 3.3, we observe increased opportunities for cooperation, which is the result of a lower direct-link rate requirement in (3.1). For instance, when $M = 20$, in almost 90% of realizations a beneficial cooperation is possible. Note that we do not use any power control scheme for the BS's transmit power; therefore, the direct-link data rate is higher in a smaller cell.
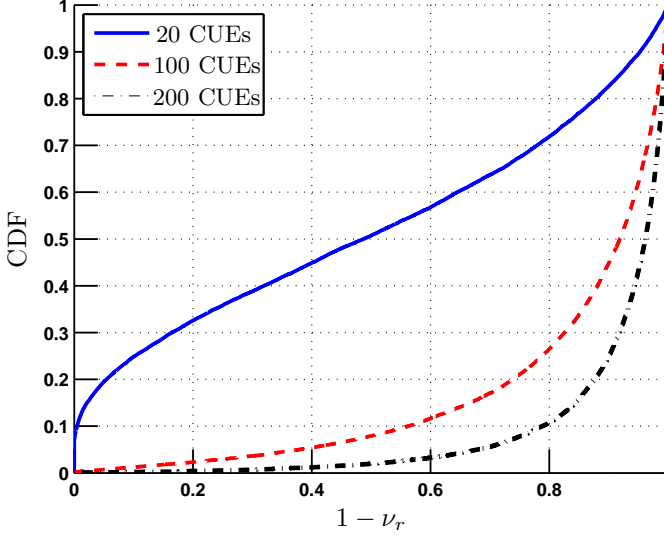
Figure 3.3: CDF of $1-\nu_r$ when $R = 500$ m. (© 2014 IEEE. Reused with permission.)

The scenario with the smaller cell radius is equivalent to the one with a requirement to raise expected gains from the CUE, which results in reduced cooperation and an increased amount of power required for cooperation.

To study the effect of throughput improvements, we consider the scenario with $R = 500$ m and $M = 20$. In Figure 3.4, the data rates for the D2D user and the CUE with and without cooperation are studied. We observe that the cooperative CUE achieves the direct-link data rate which is the minimum requirement for cooperation while the D2D link can achieve higher data rates if interference cancellation is possible.

Therefore, cooperative D2D communications provide not only opportunities for transmission in high-density areas, but also a high data rate for the D2D user, leading to a higher cell sum rate. Such cooperation results in a higher number of connected devices and reduced delays, and can be used to offload an overloaded cell or to extend the coverage area.

3.2 Interference Management for Multiple D2D Communications

Interference management is one important problem in integrating D2D communications underlaying cellular networks. The network performance depends on the availability of global or local information and the frequency with which this information is updated and provided for the decision-making process. In this section, we address these two problems. We model a set up that comprises K potential D2D
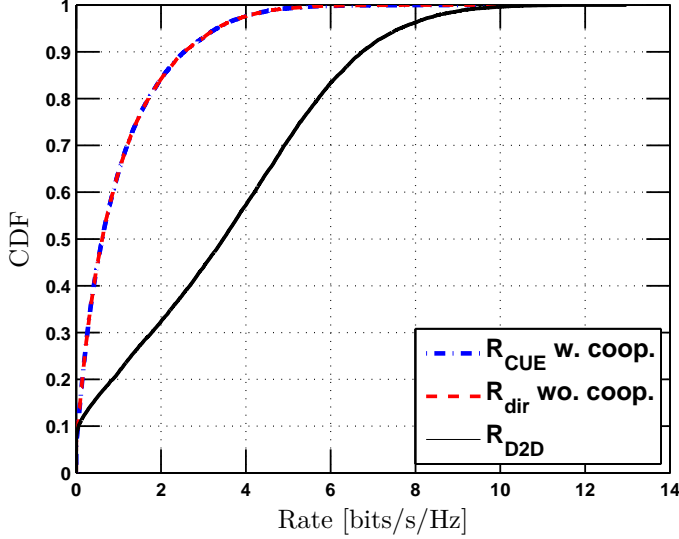
Figure 3.4: CDF of data rates when $R = 500$ m and $M = 20$. (© 2014 IEEE. Reused with permission.)

links in addition to one CUE that communicates through the BS in a single cell. The D2D communications reuse the resources of the CUE in the uplink. Since more than one user is allowed to be scheduled, the aggregated interference resulting from D2D communications becomes very important. We assume that the CUE uses its maximum transmit power and optimize the power of active D2D links such that the quality-of-service (QoS) of the CUE is not degraded. To this end, we formulate the interference management problem in the following two sections under different assumptions regarding the availability of CSI. To model active D2D users, we define a binary random variable $x_k \in \{0, 1\}$, $k \in \{1, \ldots, K\}$, where $x_k = 1$ corresponds to the event that the kth D2D user is active, and $x_k = 0$, if otherwise.

3.2.1 Full CSI versus Limited CSI

We first consider the effect of the time scale for updating channel information, from which we study the trade-off between the availability of full CSI, which requires instantaneous updates, and that of limited CSI, which requires only path loss information. In order to manage the interference from multiple D2D transmissions, the BS controls the transmit power of active D2D users. One simple power control method which has also been used in DS-CDMA systems is to assume that the interference power received at the BS is the same for all D2D nodes regardless of their positions. This power control method enables users who are far away from the BS to communicate directly with each other in the D2D mode while the ones

who are close to the BS may prefer to be scheduled in the cellular mode.

Let $\bar{p}$ be the interference power received from all D2D users, and γ_0 and γ_k denote the CUE's and the kth D2D user's instantaneous signal-to-noise-plus-interference ratio (SINR), respectively. Then, the corresponding achievable data rates are $R_0 = \log_2(1+\gamma_0)$ and $R_k = \log_2(1+\gamma_k)$, measured in bps/Hz. Our objective is to maximize the overall spectral efficiency of the network, i.e.,

$$\underset{p_k,x_k}{\text{maximize}} \quad R_0 + \sum_{k=1}^{K} x_k R_k. \tag{3.4}$$

In order to protect the BS, we formulate the problem with two different constraints. In the first case, we use a fast adaptation approach, adapting the D2D users' transmit power to the instantaneous channel gains. In this type of solution, the channel gains of all D2D links are required by the BS, which results in high signaling overhead among the network entities. One way to reduce this overhead is to consider using only the average channel gains, which will lead to the use of a slow adaption approach. In this case, power adaptations of the D2D transmitters are not instantaneous; that is to say, the power allocation can only compensate for slow fading. In this case, a certain threshold for the outage of the CUE should be allowed.

In the first scenario with fast adaption, a pre-determined SINR threshold, γ_0^{th} is defined to protect the CUE. Then, the instantaneous SINR constraint is

$$\gamma_0 \geq \gamma_0^{\text{th}}. \tag{3.5}$$

Denoting the number of active D2D links by L, we have

$$\frac{p_0 G_{00}}{\bar{p}\sum_{k=1}^{K} x_k + N} = \frac{p_0 G_{00}}{\bar{p}L + N} \geq \gamma_0^{\text{th}}, \tag{3.6}$$

where p_0 is the transmit power of the CUE, G_{00} is the instantaneous channel gain from the CUE to the BS, and N is the receiver's noise power. Thus, the relationship between the maximum tolerable received interference power ($\bar{p}$) from each D2D transmitter at the BS and the maximum number of supported D2D users are

$$\bar{p} \leq \frac{1}{L}\Big(\frac{p_0 G_{00}}{\gamma_0^{\text{th}}} - N\Big). \tag{3.7}$$

Using an iterative heuristic approach, we can find the value of L and the corresponding transmit powers that maximize the objective (3.4). The results obtained from the heuristic approach (Algorithm 1 in Paper B) is also compared with a brute-force enumeration approach in performance evaluation.

In the second scenario with slow adaptation, we assume that the power updates in the D2D transmitters are based only on the path loss information and do not

follow the fast fading. Therefore, a small outage probability is allowed at the BS which is given as

$$P_{\text{out}} = 1 - e^{-\frac{\gamma_0^{\text{th}}}{\bar{\gamma}_0}} \left(\frac{\frac{\bar{\gamma}_0}{\gamma_0^{\text{th}}}}{\bar{\gamma}_d + \frac{\bar{\gamma}_0}{\gamma_0^{\text{th}}}} \right)^L, \tag{3.8}$$

where $\bar{\gamma}_0$ and $\bar{\gamma}_d$ are the average SNR of the CUE and D2D users, respectively. The outage constraint becomes

$$P_{\text{out}} \leq p_{\text{out}}^{\text{th}}, \tag{3.9}$$

where $p_{\text{out}}^{\text{th}} > 0$ is the tolerable outage probability at the BS. Using (3.8)–(3.9), we can calculate an upper bound on $\bar{p}$, the allowable interference power which is received at the BS from each active D2D user, as

$$\bar{p} \leq \frac{p_0 \bar{g}_{00}}{\gamma_0^{\text{th}}} \left[\left(\frac{e^{-\frac{\gamma_0^{\text{th}}}{\bar{\gamma}_0}}}{1 - p_{\text{out}}^{\text{th}}} \right)^{\frac{1}{L}} - 1 \right]^+, \tag{3.10}$$

where $[z]^+ = \max\{z, 0\}$. To find the optimal solution for the problem (3.4) based on the availability of average channel gains with the constraint (3.9), we employ an enumeration approach.

For performance evaluation, we assume two pre-defined thresholds for the CUE $\gamma_0^{\text{th}} = 2, 6$ dB. Figure 3.5 shows the average number of active D2D users versus different numbers of available D2D users in the cell for both adaptation approaches. As the results indicate, a higher number of D2D links are scheduled with the fast adaption approach than with the slow adaption method. In slow adaption, due to the lack of instantaneous channel knowledge, the transmit power of D2D users is overestimated, which in turn results in fewer active D2D links. Figure 3.6 shows a similar trend regarding the average cell sum rate. Furthermore, the difference between fast and slow adaption is more noticeable when a higher number of users participate in the scheduling decision. Thus, the slow adaptation approach is likely to require less frequent information exchanges, but might not achieve the same gain as does the fast adaption approach due to over provisioning. Since the performance of the slow adaptation approach is not too far from that of the fast adaptation, it may be preferable based on the dynamics of the network and the application scenario.

3.2.2   Full CSI versus No CSI

Next, we study the trade-off between the availability of full CSI and no CSI at all. The goal is to maximize the number of active D2D users in the cell while ensuring that the aggregated interference to both cellular and D2D users meets the users' QoS requirements. The problem is formulated under two distinctive constraints, namely peak interference constraint (PIC) and average interference constraint (AIC).
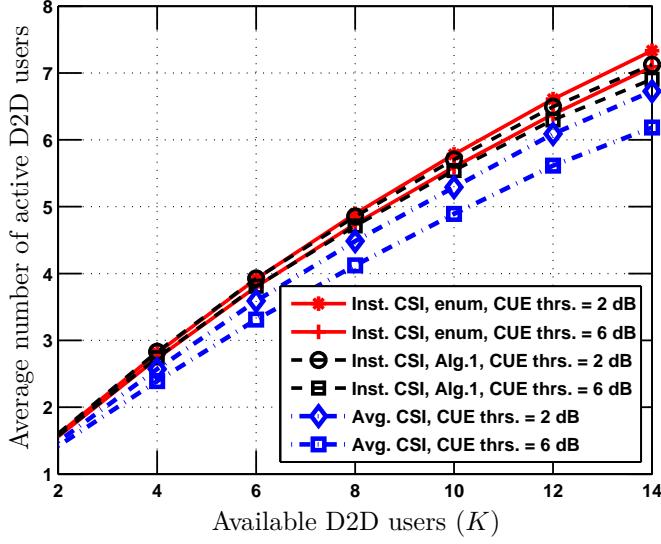
Figure 3.5: Average nr. of active D2D users vs. nr. of D2Ds in the cell for $\gamma_0^{\text{th}} = 2$ dB and $p_{\text{out}}^{\text{th}} = 10\%$. (© 2013 IEEE. Reused with permission.)
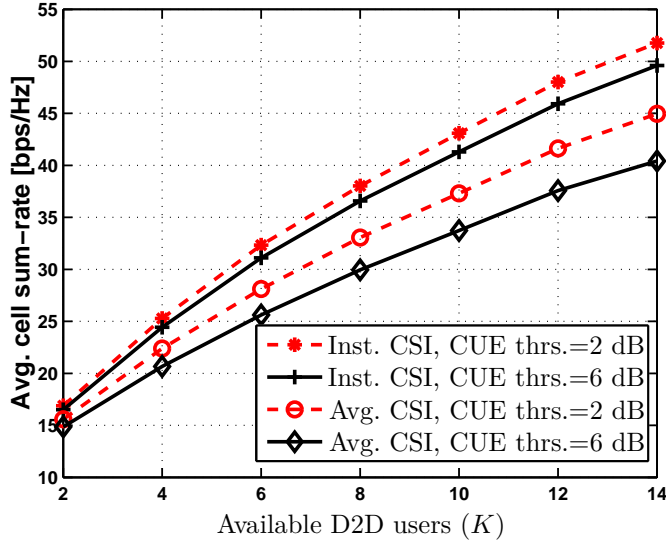

Figure 3.6: Average cell sum rate based on average and instantaneous CSI for $\gamma_0^{\text{th}} = 0, 6$ dB and $p_{\text{out}}^{\text{th}} = 10\%$. (© 2013 IEEE. Reused with permission.)

We formulate the first optimization problem under PIC as

$$\max_{\left\{\substack{p_k \in \mathbb{R}_+ \\ x_k \in \{0,1\}}\right\}_{\forall k}} \sum_{k=1}^{K} x_k \tag{3.11}$$

subject to

$$\gamma_k \geq \gamma_d^{\text{th}}, \qquad\qquad \forall k \in \{1, \ldots, K\}, \tag{3.12a}$$

$$\sum_{k=1}^{K} x_k p_k G_{k0} \leq I_{\text{th}}, \tag{3.12b}$$

$$x_k p_k - P_{\max}^d \leq 0, \qquad\qquad \forall k \in \{1, \ldots, K\}, \tag{3.12c}$$

where p_k is the transmit power of the kth D2D transmitter and G_{k0} is the instantaneous channel gain of the kth D2D transmitter to the BS. Constraint (3.12a) accounts for the QoS of D2D users, constraint (3.12b) assures that the interference at the BS is under a certain limit I_{th}, and the last constraint corresponds to the maximum allowable transmit power of each active device $P_{\max}^d$.

In the second problem formulation under AIC, we assume that D2D user locations are random and unknown to the BS, which only has the statistics of the D2D users' CSI. Therefore, the interference constraint in (3.12b) is changed to

$$\mathbb{E}_{p,G}\left[\sum_{k=1}^{K} x_k p_k G_{k0}\right] \leq I_{\text{th}}. \tag{3.13}$$

The constraint (3.12a) is omitted in this formulation since the BS does not have any information about the D2D users' CSI and cannot guarantee their QoS. We consider channel inversion as the power control policy and derive an upper bound on the number of D2D users that can simultaneously be active as

$$L \leq \frac{c_d}{c_0} \frac{I_{\text{th}}}{p_r \beta(R, \alpha_0, \alpha_d)}. \tag{3.14}$$

In the above upper bound, I_{th} is determined by the CUE's QoS requirement, c_d, c_0 are the path loss coefficients for the D2D users and the CUE, respectively. $\beta(R, \alpha_0, \alpha_d)$, which is a function of the cell radius R and the path loss exponents α_0 and α_d, is calculated from (3.13) based on the channel distribution and the power control policy.

In this problem formulation, the only available information at the BS is the total number of users who have the D2D capability. Therefore, the BS decides on the admission of D2D links with minimal information and a very low complexity. This approach is useful, for example, when the BS should decide if a new D2D user can join the current band.

We validate the performance of the above two methods with Monte-Carlo simulations. We consider a cell size of $R = 350$ m. First, we compare the cell throughput
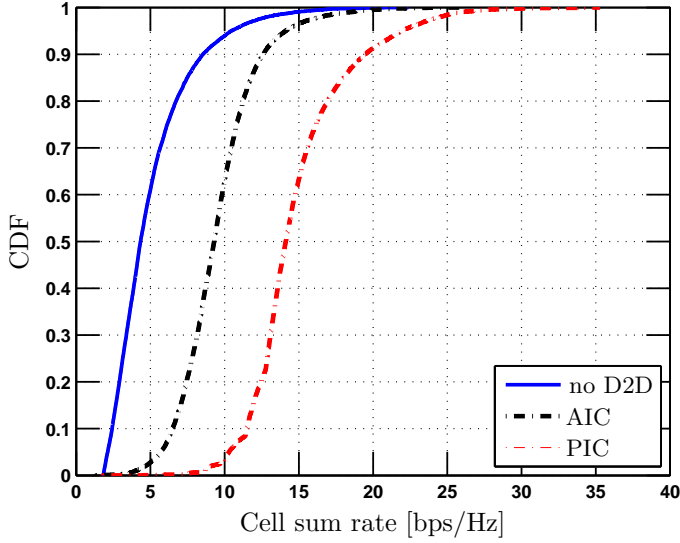
Figure 3.7: CDF of the cell sum rate for different scenarios with $K = 8$, $\gamma_0^{\text{th}} = 4$ dB, and $\gamma_d^{\text{th}} = 2$ dB. (© 2014 IEEE. Reused with permission.)
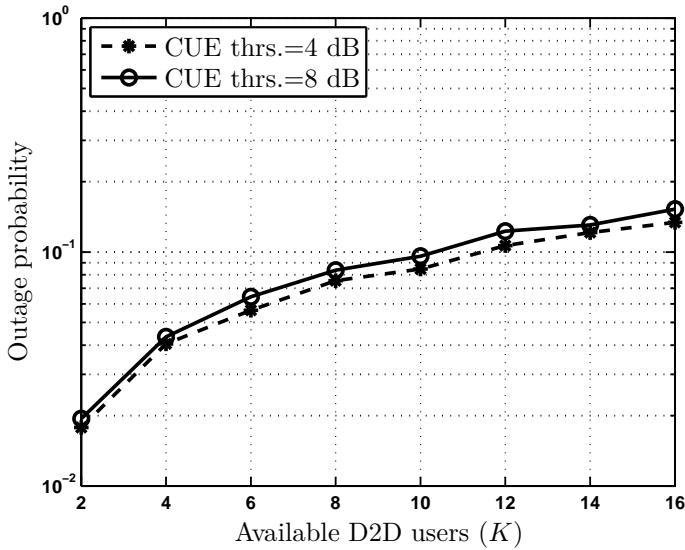

Figure 3.8: CUE's outage probability vs. available D2D users (K), with $\gamma_0^{\text{th}} = 0, 4$ dB and $\gamma_d^{\text{th}} = 2$ dB. (© 2014 IEEE. Reused with permission.)

achieved in three scenarios: conventional cellular transmission with no D2D communications, cellular transmission coexisting with multiple D2D users under PIC where full CSI is known at the BS, and D2D communications under AIC in which no CSI is available. As shown in Figure 3.7, the results indicate that our proposed method can improve performance even though no information about channel state of D2D users is available. In the results presented here, we removed the solutions regarding $I_{\text{th}} < 0$ as in such cases without transmission of D2D users, the CUE is already in outage. Clearly, there is a direct relation between the performance and the amount of available information. However, the proposed approach reduces the complexity for faster decision making in resource allocation and still can improve the performance compared with the conventional system with no D2D communications.

As in the AIC approach, by which only the average interference constraint can be guaranteed but not the instantaneous QoS for the CUE, we also study the CUE's outage probability, which is given by

$$P_{\text{out}} = \Pr\left(\sum_{k=1}^{K} x_k p_k G_{k0} > I_{\text{th}}\right), \tag{3.15}$$

in our simulations. Figure 3.8 shows the outage probability of the CUE at the BS caused by D2D transmissions under AIC. It is observed that even though there is no available information about D2D users, the outage probability of the CUE is quite low especially when the number of D2D users is low.

3.3   Mode Selection in D2D Communications

In this section, we consider the scenario depicted in Figure 3.9 where UE_1 would like to communicate with UE_2. Both users are located in the same cell and equipped with a single antenna whereas the BS is equipped with N antenna arrays. The main question that we try to answer is:

- When is the D2D mode (direct transmission) preferable to the cellular mode?

To be able to compare these two modes, we assume that the length of transmission is the same for both. That is, in the cellular mode both uplink and downlink are used for communications while in the D2D mode, two uplink resources are used, as shown in Figure 3.10.

The mode selection problem is closely related to the way the optimization problem is modeled for the system. The resource allocation problems are modeled using two different approaches. The first one is to maximize the spectral efficiency or QoS for a given transmit power $p_{\text{UE}} = p_{\text{UE}}^*$, which is written as

$$\begin{aligned} \underset{R}{\text{maximize}} \quad & R \\ \text{subject to} \quad & \max\left(R_{\text{cell}}(p_{\text{UE}}^*), R_{\text{D2D}}(p_{\text{UE}}^*)\right) \geq R. \end{aligned} \tag{P1}$$
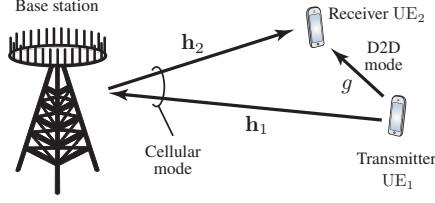
Figure 3.9: Illustration of the system model where UE_1 communicates with UE_2, either via the BS (cellular mode) or by direct transmission (D2D mode). (© 2014 IEEE. Reused with permission.)
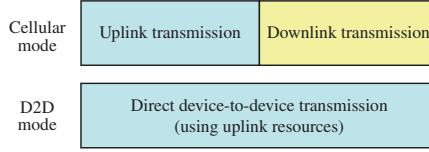

Figure 3.10: By cutting out the middleman (the BS), D2D mode can effectively use twice the amount of resources for data transmission than cellular mode. (© 2014 IEEE. Reused with permission.)

The other approach is to minimize the transmit power required to maintain a given QoS level R^*, which is written as

$$
\begin{aligned}
&\underset{p_{\text{UE}}}{\text{minimize}} && p_{\text{UE}} \\
&\text{subject to} && \max\Big(R_{\text{cell}}(p_{\text{UE}}), R_{\text{D2D}}(p_{\text{UE}})\Big) \geq R^*.
\end{aligned}
\tag{P2}
$$

We study these two formulations for D2D communications with dedicated and shared resources and show how they may behave differently with respect to the optimal mode of operation.

3.3.1 D2D Mode Optimality with Dedicated Resources

If the D2D resource is dedicated, it can be proved that the solution to (P1) for a given transmit power $p_{\text{UE}}^* > 0$ is achieved by the D2D mode if

$$
|g|^2 \geq \sqrt{\frac{\|\mathbf{h}_1\|^2 (\sigma_{\text{UE}}^2)^2 \kappa}{\sigma_{\text{BS}}^2 p_{\text{UE}}^*}},
\tag{3.16}
$$

where $g \in \mathbb{C}$ denotes the direct link channel between the UEs, and $\mathbf{h}_1 \in \mathbb{C}^{N \times 1}$ denotes the channel between UE_1 and the BS. The parameter κ decides whether the UE and BS can double the energy per channel use in the cellular mode ($\kappa = 2$), since they only transmit half of the time, and whether the energy is fixed ($\kappa = 1$). The additive circularly-symmetric complex Gaussian noise has variance σ_i^2, $i \in \{\text{UE}, \text{BS}\}$.

For a given QoS $R^* > 0$, the solution to (P2) is achieved by the D2D mode if and only if

$$|g|^2 \geq \frac{1}{2^{R^*} + 1} \frac{\sigma_{\text{UE}}^2}{\sigma_{\text{BS}}^2} \|\mathbf{h}_1\|^2 \kappa. \tag{3.17}$$

3.3.2 D2D Mode Optimality with Shared Resources

In the case of shared resources, a simple sufficient condition for D2D mode optimality can be proved to be

$$|g|^2 \geq \sqrt{\frac{|\mathbf{h}_1^H \mathbf{w}_1|^2 I_{\text{D2D}}^2 \kappa}{p_{\text{UE}}^* I_{\text{ul}}}}, \tag{3.18}$$

where $\mathbf{w}_1 \in \mathbb{C}^{N \times 1}$ is the unit-norm receiver beamforming vector. I_{ul} and I_{D2D} denote the interference-plus-noise powers at UE_2 and the BS in the uplink, respectively.

The solution to (P2) for a given R^* is achieved by the D2D mode if and only if

$$|g|^2 \geq \frac{1}{2^{R^*} + 1} \frac{I_{\text{D2D}}}{I_{\text{ul}}} |\mathbf{h}_1^H \mathbf{w}_1|^2 \kappa. \tag{3.19}$$

3.3.3 Geometrical Insights

In order to gain geometrical insights into D2D mode optimality, we consider a simple path loss model:

$$|g|^2 = c_g d_g^{-b_g}, \tag{3.20}$$

$$\|\mathbf{h}_1\|^2 = N c_h d_h^{-b_h}, \tag{3.21}$$

where d_g and d_h refer to the distance between UE_1 and UE_2 and that between UE_1 and the BS, respectively. Furthermore, $c_g, c_h, b_g, b_h > 0$ are some arbitrary path loss parameters.

Using this model, e.g., in (3.16), with dedicated resources leads to

$$\frac{d_g^{-b_g}}{d_h^{-\frac{b_h}{2}}} \geq \sqrt{\frac{N}{\sigma_{\text{BS}}^2 p_{\text{UE}}^*}} \frac{\sigma_{\text{UE}}^2 \sqrt{\kappa c_h}}{c_g}. \tag{3.22}$$

Given a fixed distance d_h between UE_1 and the BS, we can compute the circular area A around UE_1 where UE_2 (or all receivers in multi-casting) should be located to enable the D2D mode. From (3.22), the optimality condition becomes

$$A = \pi d_g^2 \leq \pi d_h^{\frac{b_h}{b_g}} \left(\frac{p_{\text{UE}}^*}{N} \frac{\sigma_{\text{BS}}^2 c_g^2}{(\sigma_{\text{UE}}^2)^2 \kappa c_h} \right)^{\frac{1}{b_g}}. \tag{3.23}$$
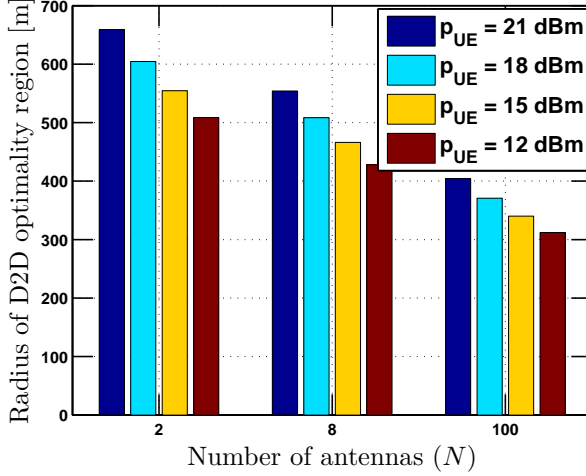
Figure 3.11: Radius of the circular D2D optimality region vs. number of antennas for (P1) with dedicated resources. (© 2014 IEEE. Reused with permission.)

It is observed that the area of D2D mode optimality increases with the distance from the BS. Therefore, the D2D mode is more probable in large macro cells and/or when UE_1 is located at the cell border. Moreover, the area grows with the transmit power as $(p_{UE}^*)^{1/b_g}$ and decreases as $1/N^{1/b_g}$ with the number of antennas. Similar conditions can be derived for (P2).

In the case of shared spectrum, in order to gain geometrical insights, in addition to the path loss model, we assume zero-forcing (ZF) beamforming at the BS to cancel the interference. This assumption causes the average SNR loss $|\mathbf{h}_1^H \mathbf{w}_1|^2 = \frac{N-M}{N}\|\mathbf{h}_1\|^2 = (N-M)c_h d_h^{-b_h}$, where M $(M < N)$ is the number of interferers. The interference experienced by UE_2 and its distance from UE_1 depend on its coordinates (x_r, y_r). Then, we have the D2D optimality condition derived from (3.18) as

$$I_{\text{D2D}}(x_r, y_r) \leq \sqrt{\frac{p_{UE}^* \sigma_{BS}^2 c_g^2}{(N-M)\kappa c_h} \frac{d_h^{b_h}}{d_g^{2b_g}(x_r, y_r)}}. \tag{3.24}$$

A similar approach can be applied for (P2).

For performance evaluation, we consider a single circular cell of radius R with the BS in the center. The distance from the D2D transmitter UE_1 to the BS is fixed to $R/2$. The scenario where dedicated resources are allocated to UE_1 is considered in Figure 3.11 and Figure 3.12. The results in Figure 3.11 show the radius of the D2D optimality area for (P1) versus the number of BS antennas and different transmit power levels. As proved in the analytical part, the area of optimality becomes larger as the power increases. However, the area is reduced as the number of BS antennas is increased. In Figure 3.12, for (P2), the D2D optimality region also becomes small
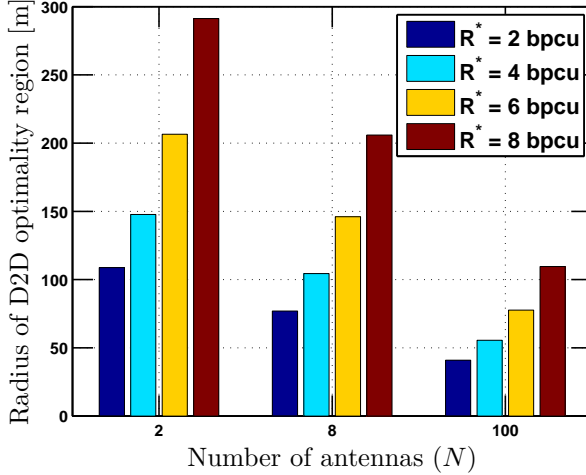
Figure 3.12: Radius of the circular D2D optimality region vs. number of antennas for (P2) with dedicated resources. (© 2014 IEEE. Reused with permission.)

if the QoS constraint is small and the number of BS antennas is large.

Figure 3.11 and Figure 3.12 show that these two problem formulations behave differently. In the case of spectrum sharing, we consider a scenario in which M interferers are placed on a circle of radius $R/2$ at equal distances apart. We assume a grid of possible positions for D2D receivers separated by 5 m in the cell area. Figures 3.13–3.14 show the probability of D2D mode optimality for each receiver position based on the bounds derived for fading channels. The D2D optimality region in (P1) is larger than the region derived in (P2). The reason is that the cost of combating the interference is using higher power, and therefore the probability of D2D mode optimality is lower in (P2) as depicted in Figure 3.14.

In this chapter, we studied four different trade-offs in resource management for D2D communications, which depend on the network scenario. For instance, cooperation is beneficial in crowded-communication environments where the network is overloaded and strict interference management is required. To increase the number of connected devices or cell sum rate, underlay D2D transmissions can be employed. Although enabling D2D communications increases the complexity of radio resource management, if a small performance loss is acceptable by the CUEs, limited CSI can be used. If fast decisions are required or if best-effort services suffice for the D2D communications, blind scheduling algorithms, which do not need any CSI knowledge, can be employed. Finally, in the last trade-off, whether a user transmits in the cellular mode or the D2D mode depends on the objective of the network, the required QoS, the user's location, and resources available at the BS.

Figure 3.13: Probability of D2D mode optimality in the case of shared spectrum with fading for (P1) with $N=8$ and $M=7$. (© 2014 IEEE. Reused with permission.)
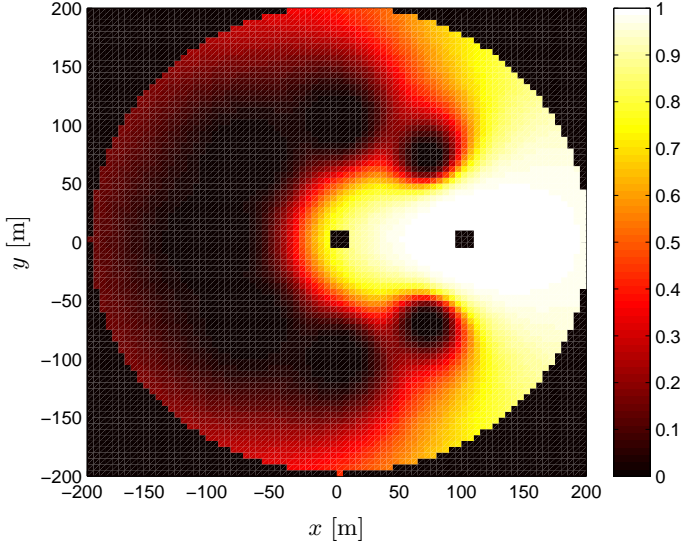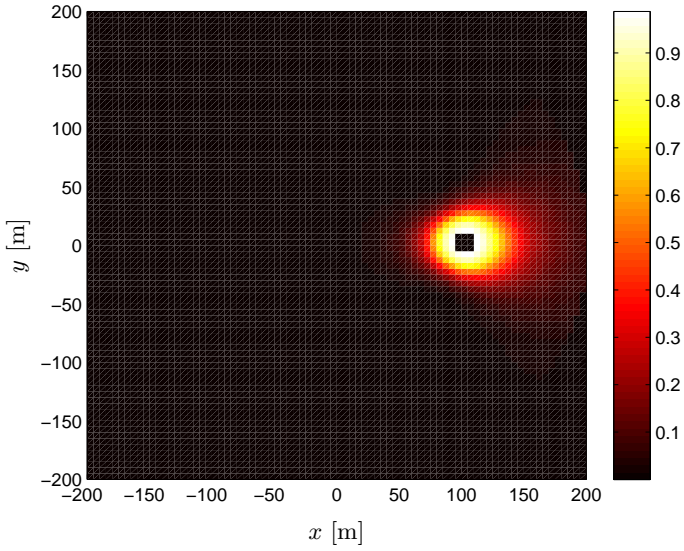

Figure 3.14: Probability of D2D mode optimality in the case of shared spectrum with fading for (P2) with $N=8$ and $M=7$. (© 2014 IEEE. Reused with permission.)

Coexistence between D2D Communications and Massive MIMO

Early work on device-to-device (D2D) communications has focused on single antenna systems.[1] However, moving towards multi-antenna systems is unavoidable—especially considering the focus of the research community on a recent technology trend for 5G networks, referred to as massive (or large-scale) MIMO. Massive MIMO is a type of multi-user MIMO (MU-MIMO) technology where the base station (BS) uses an array of hundreds of active antennas to simultaneously serve tens of users through the use of coherent transmission processing [RPB$^+$13]. Massive MIMO is known to be a very spectral and energy efficient way to obtain uniform coverage over a given area.

In spite of the known benefits of this technology, it is not clear how the overall performance of networks would be affected, if it is combined with another technology proposed for 5G, namely D2D communications. In this chapter, we provide a summary of the modeling approaches and key results of the thesis with regard to the second research question (HQ2) posed in Section 1.2. The details of the analysis and more results can be found in Papers E–F.

4.1 System Model

In order to answer the HQ2, we consider a network as shown in Figure 4.1: a single cell scenario where the BS is located in the origin with a circular coverage area of radius R. The BS serves multiple single-antenna cellular user equipments (CUEs) which are uniformly distributed over the coverage area. These CUEs are simultaneously served in the downlink direction, using an array of T_c antennas located at the BS. Furthermore, it is assumed that the number of CUEs is $1 \leq U_c \leq T_c$, since this is the interval where multi-antenna transmissions can control

---

[1]Part of the material presented in this chapter is based on our work, which has been published in [SBK$^+$15b] (© 2015 IEEE) and [SBK$^+$15a], which has been submitted for publication. The material is reused with permission.
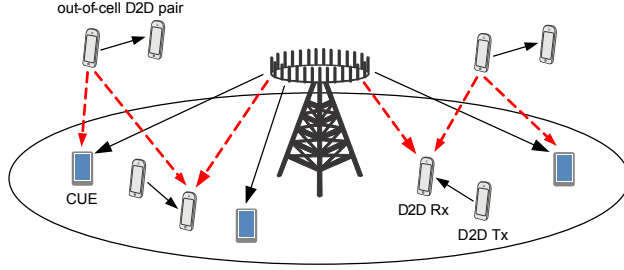
Figure 4.1: System model with a large number of antennas at the BS and underlay D2D communications. (© 2015 IEEE. Reused with permission.)

interference [BBO14]. A data precoding technique, known as zero-forcing (ZF), is used in the BS, as it can mitigate interference between the CUEs. In addition to the CUEs, there are other single-antenna users who communicate directly with each other, operating in the D2D communication mode and therefore bypassing the BS. The locations of D2D transmitters (D2D Tx) are given by the homogeneous Poisson point process (PPP) Φ with density λ_d in $\mathbb{R}^2$. The parameter λ_d is the average number of D2D Tx per unit area. The D2D receiver (D2D Rx) corresponding to any D2D Tx is randomly generated in an isotropic direction with a fixed distance away from its corresponding D2D Tx—a model that is similar to the one considered in [LLAH14].

We assume equal power allocation for both CUEs and D2D users. P_c denotes the total transmit power of the BS, and the transmit power per CUE is $\frac{P_c}{U_c}$. The transmit power of the D2D Tx is denoted by P_d.

4.2   Performance Metrics

In this section, we first introduce the performance metrics that are considered. Traditional metrics for network design include high peak rates and average rates. In the evolution of cellular networks, however, energy efficiency (EE) of networks is becoming more important and gaining more attention. Therefore, one of the performance metrics considered in this part is EE. Conventional approaches for network design are:

- maximizing rates with a fixed power budget;

- minimizing transmit power for fixed rates.

However, a new problem is how to balance rate and power consumption. In this regard, it becomes important to take into account overhead signaling and circuit power, and to use more detailed models that deal with power consumption.

We define EE as the ratio between the average sum rate (ASR) and the total consumed power, i.e.,

$$EE = \frac{ASR}{Total\ power}. \tag{4.1}$$

The ASR is obtained from the total rates of both D2D users and CUEs as

$$ASR = U_c \bar{R}_c + \pi R^2 \lambda_d \bar{R}_d, \tag{4.2}$$

where $\pi R^2 \lambda_d$ is the average number of D2D users in the cell and $\bar{R}_t$ with $t \in \{c, d\}$ denotes the average rates of the CUEs and D2D users, respectively. We compute $\bar{R}_t$ as the successful transmission rate by

$$\bar{R}_t = \sup_{\beta_t \geq 0} B_w \log_2(1 + \beta_t) P_{cov}^t(\beta_t), \tag{4.3}$$

where

$$P_{cov}^t(\beta_t) = \Pr\{SINR_t \geq \beta_t\} \tag{4.4}$$

is the coverage probability when the received SINR is higher than a specified threshold β_t needed for successful reception. Note that $SINR_t$ contains random channel fading and random user locations. Finding the supremum guarantees the best rates for the D2D users and the CUEs. If we know the coverage probability $(P_{cov}^t(\beta_t))$, the expression in (4.3) can easily be computed by using line search for each user type independently.

For the total power consumption, we consider a detailed model described in [BSHD15] as

$$Total\ power = \frac{1}{\eta}\big(P_c + \lambda_d \pi R^2 P_d\big) + C_0 + T_c C_1 + \big(U_c + 2\lambda_d \pi R^2\big)C_2, \tag{4.5}$$

where $P_c + \lambda_d \pi R^2 P_d$ is the total transmission power, η is amplifier efficiency ($0 < \eta \leq 1$), C_0 is the load-independent power consumption at the BS, C_1 is the power consumption per BS antenna, C_2 is the power consumption per user device, and $U_c + 2\lambda_d \pi R^2$ is the average number of active users.

In order to calculate EE and the ASR, we first calculate the theoretical coverage probability $P_{cov}^t(\beta_t)$ for both cellular and D2D users using tools from stochastic geometry. We then compute the ASR and EE by substituting $P_{cov}^t(\beta_t)$ in (4.3). Finally, we use the theoretical results to characterize the impact of different parameters on both the ASR and EE.

4.3 Numerical Results

Using the results from the theoretical derivations of coverage probabilities given in [SBK+15b, SBK+15a] and a detailed model of power consumption [BSHD14], we assess the performance of the setup in Figure 4.1 in terms of the ASR and EE. There are three important parameters which impact the performance of both metrics: the density of D2D users, the number of BS antennas, and the number of CUEs. We
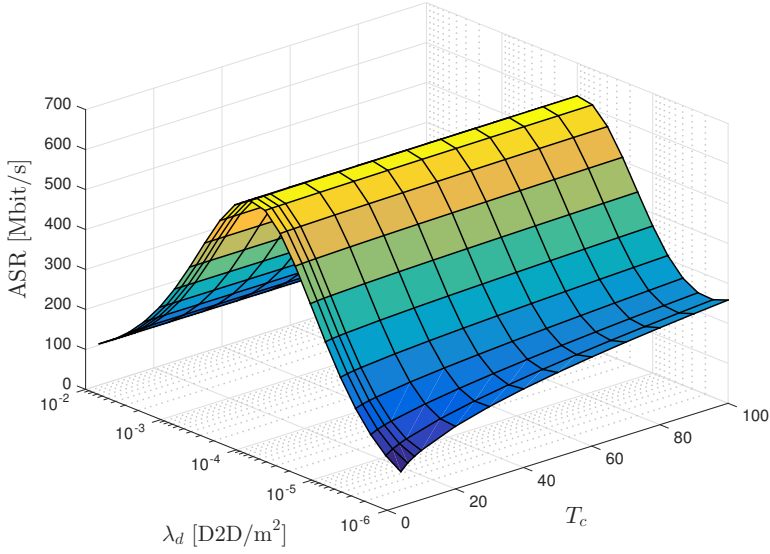
Figure 4.2: ASR [Mbit/s] as a function of the D2D user density λ_d and the BS antennas T_c with $U_c = 4$.
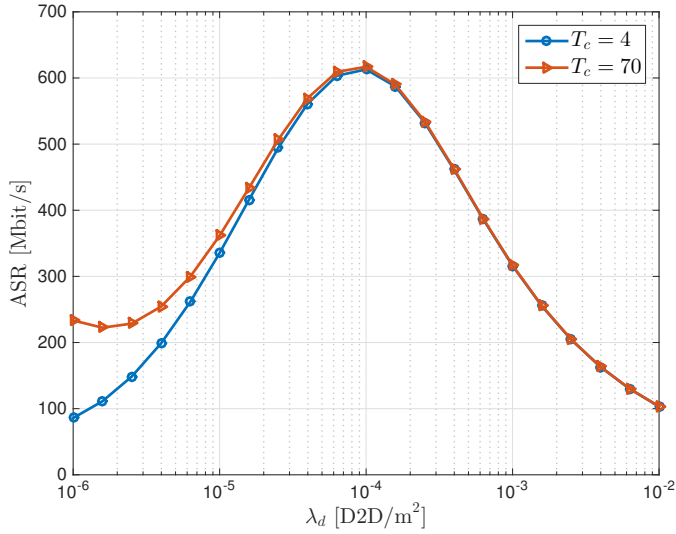

Figure 4.3: ASR [Mbit/s] as a function of the D2D user density λ_d for $T_c \in \{4, 70\}$ with $U_c = 4$.
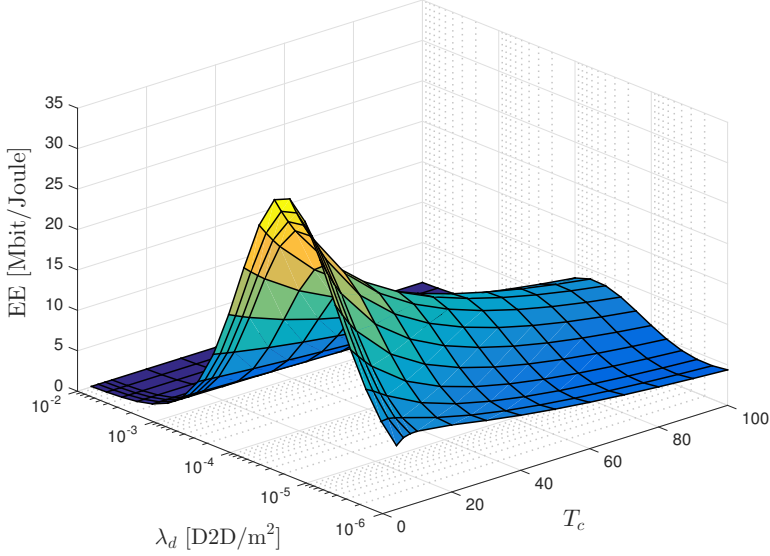
Figure 4.4: EE [Mbit/Joule] as a function of the D2D user density λ_d and the number of BS antennas T_c with $U_c = 4$.

consider two scenarios for our simulations in terms of the number of CUEs. First, we set a fixed number of CUEs ($U_c = 4$), and then, we move on to the scenario where the number of CUEs is a function of the number of BS antennas with a fixed ratio $\frac{T_c}{U_c} = 5$.

4.3.1 Fixed Number of CUEs

In Figure 4.2, the behavior of the ASR is shown as a function of different numbers of BS antennas T_c and the density of D2D users λ_d for $U_c = 4$. It is observed that increasing the number of BS antennas contributes to the slow growth of the ASR. Besides, there is an optimal value of the D2D user density λ_d which results in the maximum ASR for each number of BS antennas. However, there is a difference in the shape of the ASR between lower T_c values and higher T_c values, which can be seen in the 2-D plot in Figure 4.3 for $T_c = \{4, 70\}$. For $T_c = 4$, the rate contributed by the CUEs to the sum rate is low; therefore adding D2D users to the network (i.e., increasing λ_d), which may cause interference, will nevertheless leads to an increase in the ASR. This increase in the ASR continues until reaching a point where the interference among D2D users themselves limits the data rate per link and results in decreasing the ASR.

By increasing the number of antennas to $T_c = 70$, the rates of the CUEs become larger. In contrast, by introducing a small number of D2D users, the effect of the initial interference from D2D users becomes visible; that is, the decrease in rates of the CUEs' is not compensated by what D2D users contribute to the ASR. If we

further increase the number of D2D users, even though the rate per link decreases for both CUEs and D2D users, the resulting aggregate D2D rate becomes higher and the ASR starts to increase. The same reasoning as in $T_c = 4$ applies for the second turning point; that is, in areas with high D2D densities, the interference from D2D users is the limiting factor for the ASR. Thus, in the case of low D2D densities, increasing the number of BS antennas is beneficial in terms of the ASR. However, in the interference-limited regime (high λ_d), increasing the number of BS antennas does not impact overall network performance. Simply put, in high D2D density areas, the gain that can be achieved from massive MIMO is degraded by the interference from D2D users.

Figure 4.4 shows network performance in terms of EE as a function of the parameters λ_d and T_c. Similar to the ASR, EE behaves differently with a different number of BS antennas. With a high number of BS antennas, it decreases because the total circuit power becomes dominant and the increase in the ASR is not sufficient enough to compensate for the decrease in EE. In Figure 4.4, if we consider the EE behavior versus T_c, we see a different behavior between scenarios of low and high D2D user densities. It is observed that the low-density scenario initially benefits from more BS antennas until the sum of the circuit power consumption of all antennas at the BS dominates performance and leads to a gradual decrease of EE. As the figure implies, there is an optimal number of BS antennas for achieving maximal EE in the low-density scenario. However, in the high-density scenario, which is the interference-limited regime, EE decreases quite rapidly with T_c. Increasing the number of BS antennas in this case cannot improve the ASR; at the same time, the circuit power consumption increases as a result of the higher number of BS antennas, which in turn leads to poor network EE.

4.3.2 Number of CUEs as a Function of the Number of BS Antennas

In this section, we evaluate network performance when the ratio between the number of CUEs and the number of BS antennas is fixed by $\frac{T_c}{U_c} = 5$. The general trend of network performance is the same as in the case of a fixed number of CUEs, as discussed in the previous section. However, there are some differences, as highlighted in Figure 4.5 and Figure 4.7, regarding the ASR and EE, respectively. As it is shown in Figure 4.5, in the low D2D density regime, the ASR increases linearly with the number of CUEs (equivalently with the number of BS antennas) as the main massive MIMO gains come from multiplexing rather than just having many antennas. But, in the case of a high D2D density, the ASR is almost flat.

The above behavior can be easily explained by Figure 4.6 where the ASR resulting from the CUEs is plotted against the ASR that is contributed by D2D users. In the scenario where we have $T_c = 70$ and $U_c = 14$, the cellular network contributes more to the total ASR in the low D2D density regime (e.g., $\lambda_d = 10^{-6}$) due to a high number of CUEs and BS antennas. In this region, the network gains from massive MIMO. By increasing the number of D2D users, however, the gains are un-
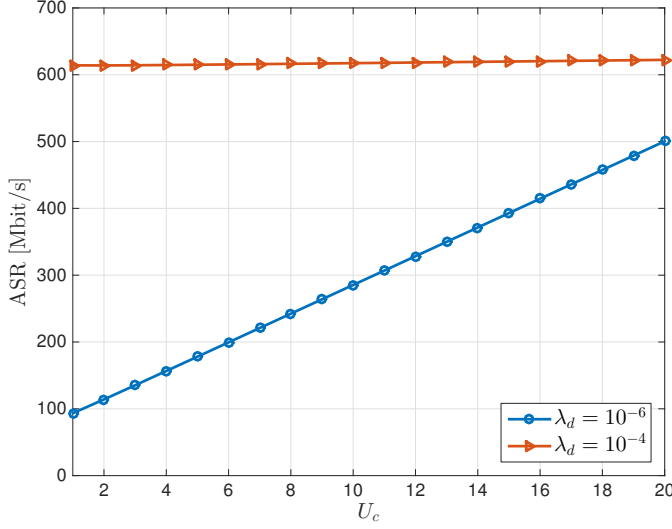
Figure 4.5: ASR [Mbit/s] as a function of the number of CUEs U_c with D2D user density $\lambda_d \in \{10^{-6}, 10^{-4}\}$ for a fixed ratio $\frac{T_c}{U_c} = 5$.

done because the interference caused by the D2D users has established dominance and consequently degrades network performance that was achieved by interference cancellation between CUEs. Therefore, with a medium D2D user density, if there is a fixed-rate constraint upon CUEs, the network can still benefit from underlay D2D communications. But in the-high density regime (e.g., $\lambda_d = 10^{-4}$), the cellular ASR is too small.

Finally, Figure 4.7 illustrates that in the low D2D density regime, even though the ASR increases linearly, EE almost stays unchanged despite the increased number of CUEs, and correspondingly of BS antennas. This is again due to the increase of the total circuit power consumption with T_c which compensates for the higher ASR. However, EE performance decreases with the number of CUEs in the high-density of D2D users. This is due to the fact that the sum rate contributed by the CUEs is small because of the interference from high number of D2D users, and additionally increasing U_c (and accordingly T_c) increases the circuit power consumption without any gain in the ASR.

The conclusion is that the D2D user density has a very high impact on a network which uses massive MIMO technology. In the downlink, these two technologies can only coexist when there is a low density of D2D users and with careful interference coordination. The number of CUEs should be a function of the number of BS antennas in order for the network to benefit from a high number of BS antennas in terms of the ASR and EE. Otherwise, in cases where the density of D2D users is high, the D2D communication should use the overlay approach rather than underlay, or the network should allow a subset of D2D users to transmit.
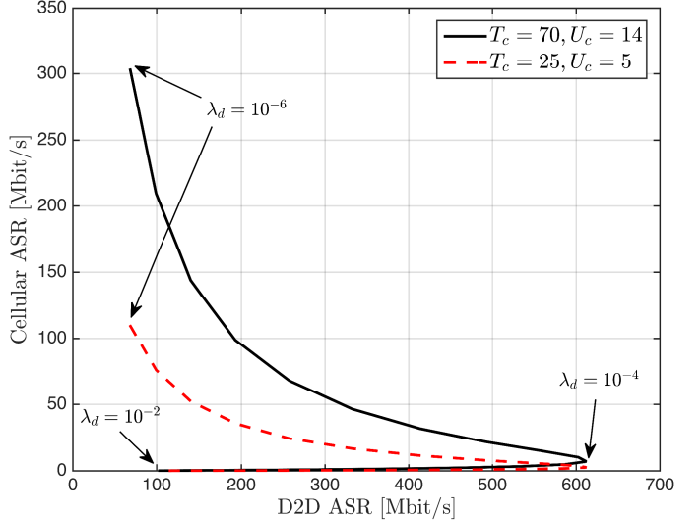
Figure 4.6: Cellular ASR vs. D2D ASR [Mbit/s] for a fixed ratio $\frac{T_c}{U_c} = 5$. The curves are obtained by varying the value of λ_d from 10^{-6} to 10^{-2}.
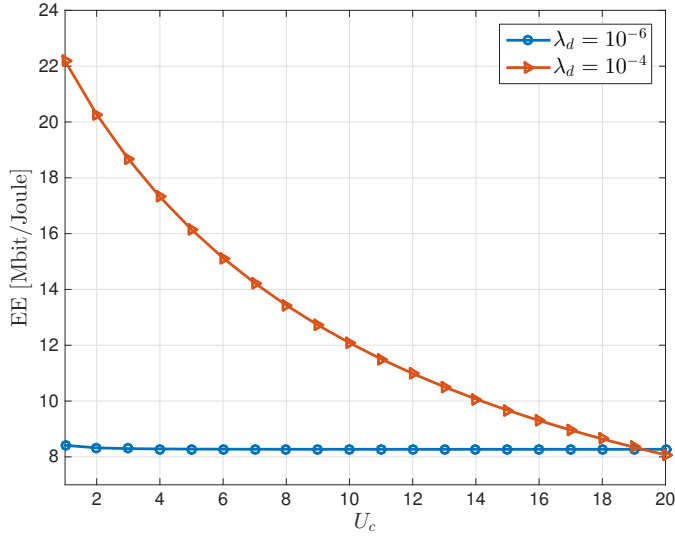

Figure 4.7: EE [Mbit/Joule] as a function of the number of CUEs U_c with the D2D user density $\lambda_d \in \{10^{-6}, 10^{-4}\}$ for a fixed ratio $\frac{T_c}{U_c} = 5$.

Conclusions and Future Research Directions

The explosion of data traffic volume poses a significant challenge for current cellular networks and is one of main drivers for the next generation of mobile networks, referred to as the fifth generation (5G). There are many solutions proposed for 5G that either try to increase the efficiency of the available resources or aim at providing new radio resources or infrastructures. Some of these solutions, if implemented, require fundamental changes at the node and architectural levels. Device-to-device (D2D) communication is a good example of such proposed solutions, by which a user communicates directly to its receiver bypassing the base station (BS). There are different ways to integrate D2D communications in networks. In this thesis, we devoted our attention to the inband network-assisted D2D communications where the network plays a major role in initiating, coordinating, and optimizing D2D communications. From a resource (time/frequency) allocation perspective, D2D users either share the same resources with cellular users (the underlay approach) or are assigned a dedicated portion of the cellular resources (the overlay approach). The scenarios considered in this thesis are mostly focused on the underlay D2D communications which can improve the spectral efficiency even in highly loaded networks.

In this chapter, we provide our concluding remarks with regard to the research questions posed in Section 1.2 and discuss some potential directions for future research.

5.1   Concluding Remarks

The first part of the thesis' contributions are related to the first high-level research question:

- **HQ1**: What are the important trade-offs of radio resource management in integrating D2D communications in cellular networks and how should they be treated?

To answer this question, we investigated four different scenarios for D2D communications as follows.

In the first scenario, we investigated the feasibility of cooperation between cellular and D2D users, especially in crowded-communication environments. We formulated an optimization problem to minimize the total power allocated for such cooperation while ensuring no performance loss for cellular users. Our results show that the possibilities of cooperation and overall improvement in cell throughput are increased with the number of cellular users within the cell and with the cell size. Such cooperation leads to a higher number of connected devices, reduces delays, increases the cell sum rate, and can be used to offload an overloaded cell or to extend the coverage area.

Next, we considered the coexistence of multiple D2D users with a cellular user. We formulated the problem of sum rate maximization with a signal-to-interference-plus-noise ratio (SINR) constraint upon the cellular user in two scenarios: (i) full CSI (instantaneous updates) and (ii) limited CSI (infrequent updates). We studied the effects of the amount of available information on the performance gains from spectrum sharing in underlay D2D communications. The results indicate that multiple D2D users can share the spectrum with a cellular link without any performance loss when full CSI is available. When there is only limited CSI is available, the network can still gain from integrating D2D communications in the same spectrum band if a small performance loss can be tolerated. Limited CSI can also decrease the computational complexity and signaling overhead (due to infrequent CSI updates), which in turn leads to less power consumption by the users. In the next step, we proposed a novel interference management scheme that requires no CSI of the D2D users, resulting in a very low signaling overhead. Specifically, the BS can schedule as many D2D users as possible in the same frequency band as shared by the cellular user. The results show that the network performance achieved by using this scheme varies depending on the cell range, the path-loss component, and the power control policy. At the same time, with our proposed scheme, it is still possible to have performance gains even when there is no CSI of the D2D users available at the BS.

In the last scenario, we studied the problem of mode selection in a network with a multi-antenna BS. We identified the decision criteria concerning the user's choice of an operation mode with two different objectives. The first objective is to maximize the quality-of-service (QoS) given a fixed transmit power, and the second objective is to minimize the transmit power given a specific QoS requirement. In both cases, we found the optimal conditions for determining when the D2D mode is preferable to the cellular mode in the scenario with dedicated resources. In addition, we derived an upper bound on the tolerable interference, by which the user can decide to operate in the D2D mode or the cellular one. Our results show that the two problem formulations behave differently and that their performance depends on the transmit power, the required QoS, and the number of BS antennas.

In summary, in this part we identified and studied four different trade-offs in resource management for D2D communications, namely cooperation (with cellu-

lar users) versus no cooperation, full CSI versus limited CSI, full CSI versus no CSI, and D2D mode versus cellular mode. These trade-offs depend on the network scenario. For instance, cooperation is beneficial in crowded-communication environments such as shopping malls and airports, where the network is overloaded and strict interference management is required. In scenarios where the goal is to increase the number of connected devices or cell throughput, simultaneous D2D transmissions can be allowed. Although enabling D2D communications increases the complexity of decision making in terms of radio resource management, if a small performance loss is acceptable by the cellular users, limited CSI can be used for scalability and for reducing the complexity of decision-making process. If very fast decisions are required or if best-effort services suffice for the D2D communications, blind scheduling algorithms, which do not need any CSI knowledge, can be employed. In the last trade-off, whether a user transmits in the cellular mode or the D2D mode depends on the objective of the network, the user's location, and resources available at the BS. Finally, in all the above scenarios, D2D communications are more beneficial at cell borders and in larger cells.

In the second part of this thesis, we shifted our focus towards 5G networks. There are many technologies and techniques proposed for 5G, of which the potential benefits are known individually but not in combination. Specially, the possibility of these solutions coexisting with one another has not been made clear so far. With this motivation, we studied the coexistence of the D2D communications with another promising enabler of 5G, namely, massive MIMO technology. In this part, we addressed the second high-level research question:

- **HQ2**: How do the extra resources and degrees of freedom in the BS resulting from a large number of antennas impact energy efficiency (EE) and the average sum rate (ASR) in underlay D2D networks?

To answer this question, we characterized the relation between the ASR and EE in terms of the number of BS antennas and D2D user density in two scenarios: first, one with a fixed number of cellular users, and the other in which the number of cellular users scales with the number of BS antennas within a given coverage area. We derived tractable expressions for the ASR and EE in both scenarios, and studied the trade-offs between important parameters. Our results show that both ASR and EE have different behaviors in scenarios with a different (low or high) density of D2D users. In the case of the fixed number of cellular users, increasing the number of BS antennas in the low D2D user density regime marginally improves the ASR; however, the increase is linear when the number of cellular users is scaled with the number of BS antennas. This is the regime where the integration of underlay D2D communications into the cellular networks is beneficial in terms of the ASR. The scenario where the number of cellular users is scaled with the number of BS antennas is more favorable than the first scenario with a fixed number of cellular users, as the additional degree of freedom is exploited to serve additional cellular users.

From an EE perspective, in the scenario with a low density of D2D users, when the number of cellular users is fixed, EE is increased with the number of BS antennas until the circuit power consumption from many BS antennas becomes dominant. However, if the number of cellular users is scaled with the number of BS antennas, the ASR increases at nearly the same rate as the circuit power consumption, and therefore EE remains almost constant. As in the case of the ASR, this is the regime where the integration of underlay D2D communications into the cellular networks is beneficial in terms of EE.

In conclusion, in the high D2D user density regime, there is a small gain in terms of the ASR from adding a great number of BS antennas. Whereas the EE degrades significantly. In fact, in this regime the interference from D2D users drastically degrades the gain from having many BS antennas. Therefore, in this regime, the D2D communication should use the overlay approach, or the network should only allow a subset of the D2D transmissions to be active at a time.

5.2   Future Work

In this section, based on the assumptions, results, and observations discussed in this thesis, we provide several suggestions for future research as listed below:

- In this thesis, we considered a single-cell scenario where we assumed the effect of inter-cell interference is negligible. Only in Paper E and Paper F, the effects of out-of-cell D2D users have been considered. One straightforward extension is to investigate how the proposed solutions may behave in the context of multi-cell scenarios.

- Except in Paper A, we neglected the effect of shadowing on our channel models for simplicity; however, more gains in D2D communications can be obtained given the shadowing effect. The effect of a detailed channel model is missing in the study of D2D communications. It is possible that many D2D users have a line-of-sight channel to their receivers due to close proximity to each other. This can potentially contribute to improved performance for D2D users, who may even be able to communicate using much less power.

- In Paper B and the second part of Paper C, only the cellular users are protected from aggregate interference created by scheduling multiple D2D users. The effects of interference between D2D users are neglected. The lack of coordination between independent D2D pairs creates high interference when multiple D2D users are operating simultaneously in crowded places such as stadiums and shopping malls. It is necessary to find a distributed algorithm which considers the effects of interference on both D2D and cellular users with local channel information.

- As we studied in Papers D–F which deal with the context of massive MIMO systems, providing multiple BS antennas is beneficial for the D2D communi-

cations in some cases. As an interesting extension, we can study the case of distributed antenna systems (DAS) with a perfect backhaul assumption. DAS is considered to be a low-cost infrastructure which can potentially increase capacity and coverage (as opposed to densification in terms of the number of BSs). Moreover, deploying antenna elements (AE) is much easier than deploying a BS. DAS can be seen as another form of massive MIMO but with different interference characteristics. It is a very relevant question to study the trade-off between the D2D mode and the cellular mode in the DAS case.

- Finally, investigating the interplay between D2D communications and massive MIMO in other higher frequency bands, such as millimeter wave (mmWave), will be an interesting future direction. In mmWave systems we need directivity gain to compensate for severe channel attenuation. This directionality, however, promises a significant gain in D2D communications due to a substantially lower amount of multiuser interference in mmWave networks.

Part II

Included Papers

Bibliography

[ABC⁺14]   J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[AGD⁺11]   G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[ARZ95]   M. Andersin, Z. Rosberg, and J. Zander, "Gradual removals in cellular PCS with constrained power control and noise," in *Proc. IEEE Int. Symp. on Personal, Indoor, Mobile Radio Commun. (PIMRC)*, Toronto, ON, Canada, Sep. 1995, pp. 56–60.

[AWM14]   A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, Fourth quarter 2014.

[BBO14]   E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.

[BDF⁺13]   R. Baldemair *et al.*, "Evolving wireless communications: Addressing the challenges and expectations of the future," *IEEE Veh. Technol. Mag.*, vol. 8, no. 1, pp. 24–30, Mar. 2013.

[BFA11]   M. Belleschi, G. Fodor, and A. Abrardo, "Performance analysis of a distributed resource allocation scheme for D2D communications," in *Proc. IEEE Global Telecommun. Conf. Workshops*, Dec. 2011, pp. 358–362.

[BFP⁺12]   M. Belleschi, G. Fodor, D. D. Penda, M. Johansson, and A. Abrardo, "A joint power control and resource allocation algorithm for D2D communications," KTH, School of Electrical Engineering (EES), Automatic Control, Tech. Rep., 2012.

[BHL⁺14]   F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[BJ13]   E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 2-3, pp. 113–381, 2013.

[BJDO14]   E. Björnson, E. Jorswieck, M. Debbah, and B. Ottersten, "Multi-objective signal processing optimization: The way to balance conflicting metrics in 5G systems," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 14–23, Nov. 2014.

[BKD13]   E. Björnson, M. Kountouris, and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," in *Proc. IEEE Int. Conf. on Telecommun. (ICT)*, May 2013.

[BKRL06]   A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

[BLD14]   E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *submitted to IEEE Trans. Wireless Commun.*, 2014. [Online]. Available: http://arxiv.org/abs/1412.7102

[BSHD14]   E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?" in *Proc. IEEE Wireless Commun. Network. Conf. (WCNC)*, Istanbul, Turkey, Apr. 2014.

[BSHD15]   E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, 2015, to appear. [Online]. Available: http://arxiv.org/abs/1403.6150

[BV04]   S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[Cis15]   "Cisco visual networking index: Global mobile data traffic forecast update, 2014–2019," White paper, Cisco, Feb. 2015.

[CK14]   Z. Chen and M. Kountouris, "Distributed SIR-aware opportunistic access control for D2D underlaid cellular networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Austin, TX, Dec. 2014.

[DKA13]   H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO HetNets: Modelling, ordering results and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5208–5222, Oct. 2013.

[DMP^{+}14]   E. Dahlman *et al.*, "5G wireless access: requirements and realization," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 42–47, December 2014.

[DRW+09]   K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Commun. Mag.*, vol. 47, no. 12, pp. 42–49, Dec. 2009.

[DYRJ10]   K. Doppler, C. Yu, C. Ribeiro, and P. Jänis, "Mode selection for device-to-device communication underlaying an LTE-advanced network," in *Proc. IEEE Wireless Commun. Network. Conf. (WCNC)*, Sydney, Australia, Apr. 2010.

[Eri12]   "Heterogeneous networks," White paper, Ericsson, Feb. 2012.

[Eri13]   "5G radio access - research and vision," White paper, Ericsson, Jun. 2013.

[FDM+12]   G. Fodor *et al.*, "Design aspects of network assisted device-to-device communications," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 170–177, Mar. 2012.

[FR11]   G. Fodor and N. Reider, "A distributed power control scheme for cellular network assisted D2D communications," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Houston, TX, Dec. 2011.

[GJMS09]   A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: an information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[GKH+07]   D. Gesbert, M. Kountouris, R. W. Heath, C. B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[GLP]   "GNU Linear Programming Kit (GLPK)," http://www.gnu.org/software/glpk/.

[Gol05]   A. Goldsmith, *Wireless Communications.* Cambridge University Press, 2005.

[Hae13]   M. Haenggi, *Stochastic Geometry for Wireless Networks.* Cambridge University Press, 2013.

[HRTA14]   E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," vol. 21, no. 3, pp. 118–127, June 2014.

[HS04]   H. Y. Hsieh and R. Sivakumar, "On using peer-to-peer communication in cellular wireless data networks," *IEEE Trans. Moblie Comp.*, vol. 3, no. 1, pp. 57–72, Jan 2004.

[HSS13]   I. Hwang, B. Song, and S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, 2013.

[IR09]   ITU-R, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," http://www.itu.int/pub/R-REP-M.2135/, International Telecommunication Union, Tech. Rep., 2009.

[JKR⁺09] P. Jänis, V. Koivunen, C. Ribeiro, J. Korhonen, K. Doppler, and K. Hugl, "Interference-aware resource allocation for device-to-device radio underlaying cellular networks," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, Barcelona, Spain, Apr. 2009.

[JYD⁺09] P. Jänis *et al.*, "Device-to-device communication underlaying cellular communications systems," *IEEE Trans. Inf. Theory*, vol. 2, no. 3, pp. 169–178, Mar. 2009.

[KB02]   S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 46–55, Jan. 2002.

[KLA13]   B. Kaufman, J. Lilleberg, and B. Aazhang, "Spectrum sharing scheme between cellular users and ad-hoc device-to-device users," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1038–1049, Mar. 2013.

[KLH09]   E. Karipidis, E. Larsson, and K. Holmberg, "Optimal scheduling and QoS power control for cognitive underlay networks," in *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Process. (CAMSAP)*, Aruba, Dutch Antilles, Dec. 2009, pp. 408–411.

[LAG]   X. Lin, J. Andrews, and A. Ghosh, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Trans. Wireless Commun.*, to appear.

[LKLH12] D. Lee, S. Kim, J. Lee, and J. Heo, "Performance of multihop decode-and-forward relaying assisted device-to-device communication underlaying cellular networks," in *Proc. IEEE Int. Symp. on Inf. Theory and its Applications (ISITA)*, Honolulu, HI, Oct. 2012, pp. 455–459.

[LLAH14] N. Lee, X. Lin, J. G. Andrews, and R. W. Heath Jr., "Power control for D2D underlaid cellular networks: Modeling, algorithms and analysis," *IEEE J. Sel. Areas Commun.*, 2014.

[LMU⁺14] Z. Li *et al.*, "Overview on initial METIS D2D concept," in *1st International Conference on 5G for Ubiquitous Connectivity (5GU)*, Levi, Finland, Nov. 2014.

[LPXW12]   Z. Liu, T. Peng, S. Xiang, and W. Wang, "Mode selection for device-to-device (D2D) communication under LTE-advanced networks," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Ottawa, Canada, Jun. 2012, pp. 5563–5567.

[LTW04]   J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[LZLS12]   L. Lei, Z. Zhong, C. Lin, and X. Shen, "Operator controlled device-to-device communications in LTE-advanced networks," *IEEE Wireless Commun. Mag.*, vol. 19, no. 3, pp. 96–104, Jun. 2012.

[Mar10]   T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[MHR+14a]   S. Mumtaz, K. M. S. Huq, A. Radwan, J. Rodriguez, and R. L. Aguiar, "Energy efficient interference-aware resource allocation in LTE-D2D communication," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Sydney, Australia, June 2014.

[MHR14b]   S. Mumtaz, K. M. S. Huq, and J. Rodriguez, "Direct mobile-to-mobile communication: Paradigm for 5G," *IEEE Wireless Commun. Mag.*, vol. 21, no. 5, pp. 14–23, Oct. 2014.

[MLPH11]   H. Min, J. Lee, S. Park, and D. Hong, "Capacity enhancement using an interference limited area for device-to-device uplink underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 3995–4000, Dec. 2011.

[MSL+11]   H. Min, W. Seo, J. Lee, S. Park, and D. Hong, "Reliability improvement using receive mode selection in the device-to-device uplink period underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 413–418, Feb. 2011.

[NDA13]   T. D. Novlan, H. S. Dhillon, and J. G. Andrews, "Analytical modeling of uplink cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2669–2679, Jun. 2013.

[NLM13]   H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[OBB+14]   A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Trans. Commun.*, vol. 52, no. 5, pp. 26–35, May 2014.

[PBM+13]   P. Popovski, V. Braun, H.-P. Mayer, P. Fertl *et al.*, "ICT-317669-METIS/D1.1 scenarios, requirements and KPIs for 5G mobile and wireless system," Tech. Rep., 2013. [Online]. Available: http://www.metis2020.com/

[PdC08]   P. Popovski and E. de Carvalho, "Improving the rates in wireless relay systems through superposition coding," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4831–4836, Dec. 2008.

[PL13]   Y. Pei and Y. Liang, "Resource allocation for device-to-device communication overlaying two-way cellular networks," in *Proc. IEEE Wireless Commun. Network. Conf. (WCNC)*, Shanghai, China, Apr. 2013, pp. 3375–3380.

[PP02]   A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes.*   McGraw-Hill, 2002.

[Rea10]   "Strategies for mobile network capacity expansion," White paper, Real Wireless Ltd, Nov. 2010.

[RPB+13]   F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[SBK+15a] S. Shalmashi, E. Björnson, M. Kountouris, K. W. Sung, and M. Debbah, "Energy efficiency and sum rate analysis of massive MIMO systems with underlaid device-to-device communications," *IEEE Trans. Commun.*, May 2015, submitted.

[SBK+15b] S. Shalmashi, E. Björnson, M. Kountouris, K. W. Sung, and M. Debbah, "Energy efficiency and sum rate when massive MIMO meets device-to-device communication," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, London, UK, June 2015.

[SBSD14]   S. Shalmashi, E. Björnson, S. B. Slimane, and M. Debbah, "Closed-form optimality characterization of network-assisted device-to-device communications," in *Proc. IEEE Wireless Commun. Network. Conf. (WCNC)*, Istanbul, Turkey, Apr. 2014.

[Sha14]   S. Shalmashi, "Cooperative spectrum sharing and device-to-device communications," Licentiate thesis, KTH Royal Institute of Technology, Stockholm, Sweden, June 2014.

[SMHS14]   S. Shalmashi, G. Miao, Z. Han, and S. B. Slimane, "Interference constrained device-to-device communications," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Sydney, Australia, Jun. 2014.

[SMS13]   S. Shalmashi, G. Miao, and S. B. Slimane, "Interference management for multiple device-to-device communications underlaying cellular networks," in *Proc. IEEE Int. Symp. on Personal, Indoor, Mobile Radio Commun. (PIMRC)*, London, UK, Sep. 2013.

[SNHHar]   L. Song, D. Niyato, Z. Han, and E. Hossain, *Wireless Device-to-Device Communications and Networks*. Cambridge University Press, to appear.

[Sol78]   H. Solomon, *Geometric Probability*. Society for Industrial and Applied Mathematics, 1978.

[SS12]   S. Shalmashi and S. B. Slimane, "Performance analysis of relay-assisted cognitive radio systems with superposition coding," in *Proc. IEEE Int. Symp. on Personal, Indoor, Mobile Radio Commun. (PIMRC)*, Sydney, Australia, Sep. 2012, pp. 1226–1231.

[SS13]   S. Shalmashi and S. B. Slimane, "On secondary user transmission schemes in relay-assisted cognitive radio networks," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, Dresden, Germany, Jun. 2013.

[SS14]   S. Shalmashi and S. B. Slimane, "Cooperative device-to-device communications in the downlink of cellular networks," in *Proc. IEEE Wireless Commun. Network. Conf. (WCNC)*, Istanbul, Turkey, Apr. 2014.

[TSM13]   A. Thanos, S. Shalmashi, and G. Miao, "Network-assisted discovery for device-to-device communications," in *Proc. IEEE Global Telecommun. Conf. Workshops*, Atlanta, GA, Dec. 2013.

[TVZ11]   S. Tombaz, A. Vastberg, and J. Zander, "Energy- and cost-efficient ultra-high-capacity wireless access," *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 18–24, Oct. 2011.

[WA12]   S. Weber and J. G. Andrews, "Transmission capacity of wireless networks," *Foundations and Trends® in Networking*, vol. 5, no. 2â€"3, pp. 109–281, 2012. [Online]. Available: http://dx.doi.org/10.1561/1300000032

[WCDT01]   H. Wu, Q. Chunming, S. De, and O. Tonguz, "Integrated cellular and ad hoc relaying systems: iCAR," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2105–2115, Oct. 2001.

[WXS+13]   F. Wang, C. Xu, L. Song, Q. Zhao, X. Wang, and Z. Han, "Energy-aware resource allocation for device-to-device underlay communication," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Budapest, Hungary, June 2013.

[WZZY13]   S. Wen, X. Zhu, X. Zhang, and D. Yang, "QoS-aware mode selection and resource allocation scheme for device-to-device (D2D) communication in cellular networks," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, Budapest, Hungary, June 2013, pp. 101–105.

[XH10]   H. Xing and S. Hakola, "The investigation of power control schemes for a device-to-device communication integrated into OFDMA cellular system," in *Proc. IEEE Int. Symp. on Personal, Indoor, Mobile Radio Commun. (PIMRC)*, Istanbul, Turkey, Sep. 2010, pp. 1775–1780.

[XHA15]   L. Xingqin, R. W. Heath Jr., and J. G. Andrews, "The interplay between massive MIMO and underlaid D2D networking," *IEEE Trans. Wireless Commun.*, 2015, to appear. [Online]. Available: http://arxiv.org/abs/1409.2792

[XSH+12]   C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, and B. Jiao, "Interference-aware resource allocation for device-to-device communications as an underlay using sequential second price auction," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Ottawa, Canada, Jun. 2012, pp. 425–429.

[XYHY12]   Y. Xu, R. Yin, T. Han, and G. Yu, "Interference-aware channel allocation for device-to-device communication underlaying cellular networks," in *Proc. IEEE Int. Conf. on Commun. in China (ICCC)*, Beijing, China, Aug. 2012, pp. 422–427.

[YASCA14] Q. Ye, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "A tractable model for optimizing device-to-device communications in downlink cellular networks," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Sydney, Australia, June 2014.

[YDRT11]   C. Yu, K. Doppler, C. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2752–2763, Aug. 2011.

[YK12]   E. Yaacoub and O. Kubbar, "Energy-efficient device-to-device communications in LTE public safety networks," in *Proc. IEEE Global Telecommun. Conf. Workshops*, Anaheim, CA, Dec. 2012.

[YLHZ+12] W. Yu, L. Liang, S. J. H. Zhang, J. C. F. Li, and M. Lei, "Performance enhanced transmission in device-to-device communications: Beamforming or interference cancellation?" in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Anaheim, CA, Dec. 2012.

[YM13]   H. Yang and T. L. Marzetta, "Total energy efficiency of cellular large scale antenna system multiple access mobile networks," in *Proc. IEEE*

Online Conference on Green Commun. (OnlineGreenCom), Piscataway, NJ, Oct. 2013.

[YTDR09]   C. H. Yu, O. Tirkkonen, K. Doppler, and C. B. Ribeiro, "On the performance of device-to-device underlay communication with simple power control," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, Barcelona, Spain, Apr. 2009.

[ZCS10]   M. Zulhasnine, S. Changcheng, and A. Srinivasan, "Efficient resource allocation for device-to-device communication underlaying LTE network," in *Proc. IEEE Int. Conf. on Wireless and Mobile Computing, Networking, and Commun. (WiMob)*, Niagara Falls, ON, Canada, Oct. 2010.

[ZdV05]   A. Zemlianov and G. de Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 3, pp. 657–667, March 2005.

[ZK01]   J. Zander and S. L. Kim, *Radio resource management in wireless networks.*   Artech House, 2001.