Decoupled Cell Association Towards Device-Centric 5G Cellular Networks

Elshaer, Hisham

Awarding institution:
King's College London

Decoupled Cell Association Towards Device-Centric 5G Cellular Networks

by

Hisham Elshaer

This dissertation is submitted for the degree of
Doctor of Philosophy at
King's College London

King's College London November 2016

To Mom, Dad, Tala, and Yomna...

Acknowledgements

First and foremost, I thank Allah for giving me the strength, patience and confidence to complete my doctoral studies.

My journey with the PhD for the past three years has been, so far, the most exciting and challenging time of my life. It completely altered the way I think and look at things and made me realize that knowledge and humility are, or should be, directly proportional, and for all that I am utterly grateful.

I have to start by thanking my supervisor Prof. Mischa Dohler for his relentless support and continuous guidance. Mischa, I consider myself lucky to have worked with you because of your passion for technology and innovation, your entrepreneurial mind, unconventional way of thinking and, of course, your trademark red boots.

Special thanks to my friend and ex-Vodafone colleague Federico Boccardi who had an instrumental role in shaping my research topic with his innovative thinking and excitement for new ideas.

I am also grateful for David Lister, my manager at Vodafone, who was continuously supportive and gave me the freedom to do my own research. I would also like to thank Ralf Irmer, my previous manager, for believing in me and for helping me in shaping my research. I am also grateful to all my Vodafone colleagues including Eric, Neil, John, Ilaria, Tony, Bob, Sami, James and all the R&D team for their invaluable support.

I also want to thank my colleagues and friends at CTR, including Christoforos, Giorgos, Yacoub, Ehsan, Omar, Maria, and Massimo who helped me with my research and made my days at King's College enjoyable. Special thanks to Prof. Jeffrey Andrews who hosted me in his group at the University of Texas at Austin and gave me the chance to collaborate with exceptional people at WNCG.

I would also like to acknowledge the MITN Crossfire project for funding my PhD.

Last, but not least, I wish to thank my family, whom I always take for granted but have always stood by my side and supported me through good and bad. It is to them that I owe my success in life.

Abstract

5G will be different from previous cellular generations in the fact that it will enable the cellular industry, besides offering superior broadband services, to conquer vertical industries such as vehicular communication, factory automation, healthcare and many more. Many of these use cases have challenging and quite often contradicting requirements in terms of data rate, latency, power consumption and so on. This suggests that 5G needs to adopt a flexible architecture that can adapt to different devices and traffic requirements. Consequently, a fresh look onto how cellular networks are currently designed and deployed is needed. Historically, cellular networks have relied on the axiomatic role of cells as the cornerstone of the radio access network. Cellular systems have witnessed several recent trends such as the increased heterogeneity in infrastructure and spectrum as well as the rise of different traffic types with different requirements. These trends have called for a shift from the cell-centric architecture approach to a more device-centric architecture where a user or a device should be able to communicate with the network by exchanging information over multiple traffic flows through several sets of heterogeneous nodes. This design concept suggests to drop the rigid cell-centric concept and move to a more flexible design where information is exchanged in the most efficient way possible disregarding in which cell the device is located.

This thesis features a comprehensive study of some of the technological enablers of the device-centric 5G architecture vision where we start by motivating the need for this architectural change by presenting the envisioned use cases and requirements of 5G and how the current cellular designs are lacking the flexibility and agility to satisfy the 5G ambition.

The main contribution of the thesis is on Downlink and Uplink Decoupling (DUDe) where we pioneered the research on this disruptive 5G architectural design. The traditional way for users to associate to the cellular network is through coupled association where a device associates in both uplink (UL) and downlink (DL) to the same cell. However, the ever increasing density and heterogeneity of cellular networks have rendered the traditional design concepts such as coupled cell association obsolete

and highly suboptimal. In simple words, a device connecting in the DL to a high power macro cell from which it receives the highest signal power might want to connect in the UL to a small cell to which the pathloss is lower. Therefore, DUDe solves the UL and DL coverage imbalance problem caused by the different transmit powers from the different tiers. The UL and DL imbalance could also be caused by imbalance in the UL and DL loads, interference and traffic requirements. The concept of DUDe is ground breaking in the sense that it introduces the notion of treating the UL and DL as two separate network entities emphasizing the fact that these two entities have different transmission and traffic requirements.

The thesis features a comprehensive simulation study on DUDe using Vodafone's small cell live network deployment in conjunction with a high resolution 3D ray tracing propagation model as well as user deployments based on traffic measurements to guarantee the most realistic simulation setup. Using this setup, the superiority of DUDe compared to baseline LTE is shown in terms of data rate, outage, channel quality and many more parameters. The evaluation starts by examining the basic form of DUDe where the UL and DL associations are based on pathloss and DL received power respectively which is followed by a more complicated form of DUDe where the UL association takes into account the cell load and the backhaul capacity. An extensive theoretical evaluation of DUDe using tools from stochastic geometry is then presented where cell association and SINR/rate distributions are evaluated in great detail for a sub-6GHz deployment as well as a mixed millimeter wave and sub-6GHz deployment. In addition, the architectural aspect is extensively discussed highlighting the support of DUDe in current 4G networks as well as the changes needed for a native support in future 5G networks.

The second aspect covered in this thesis is Device-to-Device (D2D) communications. D2D allows to establish a direct link between devices in the same vicinity to exchange data instead of going through the traditional way through the network infrastructure. D2D is considered to be another important aspect of the device-centric framework as it allows devices to exchange information in the most efficient way possible through a direct communications without the need to abide by the normal cellular way of conveying data. The cell association in a D2D enabled network is studied through an optimization framework considering a decoupled access regime. In addition, novel resource management techniques for D2D communications are presented considering bio-inspired genetic algorithms.

Finally, the thesis is concluded by a summary of the findings and takeaways from the conducted research along with some directions for future work.

Table of contents

List of figures

List of tables

List of Acronyms

3GPP	Third Generation Partnership Project
ACK	Acknowledgement
ARQ	Automatic Repeat reQuest
AWGN	Additive White Gaussian Noise
API	Application Programming Interface
BS	Base Station
BLER	BLock Error Rate
CA	Carrier Aggregation
CDMA	Code Division Multiple Access
CoMP	Cooperative MultiPoint
CQI	Channel Quality Indicator
D2D	Device-to-Device
EMO	European Mobile Observatory
eNB	Enhanced NodeB
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
E-UTRAN	Evolved-UMTS Terrestrial Radio Access Network
FDD	Frequency Division Duplex
GPRS	General Packet Radio System
GSM	Global System for Mobile communications
GTP	GPRS Tunneling Protocol
HARQ	Hybrid Automatic Repeat reQuest
HLR	Home Location Register
HSS	Home Subscriber Server
IP	Internet Protocol
ISI	Inter Symbol Interference
ICT	Information and Communication Technologies
ICI	Inter Cell Interference

ICIC	Inter Cell Interference Coordination
IEEE	Institute of Electrical and Electronic Engineers
IMT-A	International Mobile Telecommunications-Advanced
IoT	Internet of Things
ITS	Intelligent Transportation System
ITU-R	International Telecommunications Union Radiocommunication Sector
KPI	Key Performance Indicator
LTE	Long Term Evolution
LTE-A	Long Term Evolution-Advanced
M2M	Machine-to-Machine
MAC	Medium Access Control
MOS	Mean Opinion Score
MCS	Modulation and Coding Scheme
MIMO	Multiple-Input Multiple-Output
MME	Mobility Management Entity
MTC	Machine-Type Communications
NAS	Non Access Stratum
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Multiple Access
ONF	Open Networking Foundation
PCRF	Policy and Charging Rules Function
PDCP	Packet Data Convergence Protocol
PUSCH	Physical Uplink Shared Channel
QCI	QoS Class Identifier
QoS	Quality of Service
QoE	Quality of Experience
QPSK	Quadrature Phase Shift Keying
RA	Random Access
RACH	Random Access Channel
RAN	Radio Access Network
RB	Resource Block
RF	Radio Frequency
RLC	Radio Link Control
RRC	Radio Resource Control
RTT	Round Trip Time
SAE	System Architecture Evolution

SC-FDMA	Single Carrier-Frequency Division Multiple Access
SIM	Subscriber Identity Module
SINR	Signal-to-Interference-and-Noise Ratio
SNR	Signal-to-Noise Ratio
SMS	Short Message Service
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
VoIP	Voice over IP

Chapter 1

Overview

Nearly every decade we witness a new generation of mobile networks and these generations have always been driven by several factors such as technological advancements, operators seeking new business opportunities through innovative services, vendors looking to increase their sales or the increased reliance on wireless services and the need for faster, more reliable networks. The same mix of factors has driven the wireless community to develop the next generation of mobile networks referred to as 5G or the 5th generation.

Mobile operators have so far functioned on the basis of being bit-pipes for wireless data transmission where the business model is based on selling contracts for voice, text and data bundles. This model will continue to evolve bringing new challenges as to how to cope with the surging data rates and volumes while offering an excellent quality of experience to customers. From a customer behaviour perspective, high data rate multimedia and mobile-cloud services are becoming increasingly popular, and thus calling for higher-capacity future networks.

However, 5G is expected to extend the bit-pipe role of operators by creating new business opportunities allowing the mobile industry to have a major role in vertical industries such as the automotive industry, healthcare, factory automation and many more. This gives rise to new requirements that 5G needs to fulfil to cater for the different and diverse needs of different industries.

5G will operate in a highly heterogeneous environment characterized by the coexistence of multiple types of access technologies, multi-tier networks, different device types, different traffic requirements and behaviours, etc. In such an environment 5G is expected to achieve a consistent user experience everywhere and at any time and with the staggering technological advancements that we see today, 5G needs to be flexible enough to cater to future applications that we do not yet know anything about. In

addition, capital and operational costs need to be reduced in 5G in order to allow for ultra-dense deployments of small cells and to be economically viable for some internet of things (IoT) applications that require extensive coverage and low cost modules.

In the next sections we will discuss the main 5G requirements and use cases and we will review the current design concepts of cellular networks and how these concepts need to evolve for 5G to be a real device-centric network.

1.1 5G requirements and use cases

1.1.1 5G envisioned use cases

In addition to supporting the evolution of the existing broadband use cases, 5G will support countless new and emerging use cases. In fact, it is very challenging to predict the new services and use cases that will emerge in the next 10-15 years with the rate of technological advancement and innovation that we are witnessing today. The NGMN have summarized the use cases into eight families in [1] including broadband access in dense areas, broadband access everywhere, higher user mobility, massive internet of things (IoT), extreme real-time communications, lifeline communication, ultra-reliable communications and broadcast-like services.

In this section we focus on some of the new services and use cases that we think will probably play an important role in the next years.

Tactile Internet

Real-time services involving interactions between humans, or between humans and machines, can be classified according to four types of physiological interactions [2]: muscular, audio, video, and tactile. Tactile interaction refers to a system that has a tactile input and audio and/or visual feedback. One example is a software or application running in the cloud in a way that the user cannot distinguish between local and remote content. A good example is moving or dragging an object on the touch screen of a device, which is connected to a cloud-server. In order to create a seamless experience, the response from the cloud-server must be fast enough to allow the object to follow the movements of the finger on the screen without allowing the user to experience a spatial discrepancy between the finger and the object. This is not feasible with current LTE networks, as the end-to-end (E2E) latency is about 10 times the tactile requirement which estimated to be around 1 ms in [2].

Controlling robots remotely

Robotics is currently undergoing a revolution where robots that are used in production lines are becoming smaller, cheaper and versatile. They are enabling manufacturing companies to move their factories from cheap labour countries to the country of origin. Automation will replace or complement human workers not only in jobs requiring basic skills or repetitive tasks (e.g. production, transportation, logistics and office/administrative support) but also in the service industry and in jobs requiring intelligence and decision making due to artificial intelligence.

In order to enable this new generation of robots to function in different environments and in a cooperative way to complete complex and delicate tasks, an underlying control network with low latency and high reliability is needed where in [2] it is stated that control signals for robots must follow reaction time intervals that are short enough to avoid mechanical resonance between, for instance, a robot and the object that it is controlling. For many scenarios in manufacturing, this has led to maximum target latencies for a communication link in the order of 100 μs, and round-trip reaction times of 1ms.

Vehicular communications

Current systems such as the IEEE 802.11p [3] or LTE have proven to be inadequate for providing the full spectrum of vehicle to vehicle and vehicle to infrastructure services. 5G should fill the gaps in these technologies in terms of providing higher reliability, higher availability and lower latencies. Autonomous driving cars are expected to cause a revolution in the car industry with expected market share of trillions of dollars in the USA only [4]. In addition, autonomous cars are expected to reduce road accident rates by over 90% which would mean huge savings for the economy. These numbers call for the urgency of assessing the role of communications in the autonomous-cars domain. For instance, can an autonomous car function solely relying on on-board sensors or does it need assistance from the surrounding environment and network? Most likely assistance will be needed but to what extent and what are the functions that the network needs to provide? In addition, new problems need to be addressed such as who is to blame in the case of an accident, is it the network provider? The car manufacturer? Or both?

Remote control of Drones

Until recently, drones were exclusive to military operations. However, these unmanned flying devices can be put to use in a wide range of applications. Some obvious examples are parcel delivery, search and rescue operations and traffic monitoring.

It is expected that drones will soon invade our airspace. In this context, 5G cellular network can play a very important role in providing the control infrastructure for these drones. Nowadays, these drones need to be controlled from the ground by professional pilots, hence, a higher degree of automation is desirable. A very good coverage, very strict latency requirements, and extremely robust security are also required.

Immersion and augmented reality

In a world where businesses are more and more global and companies are more geographically distributed, virtual reality will be essential and instrumental in reducing the distances between colleagues, customers and suppliers in a sustainable way. In the future, there will be no need for business travel and teams will have the chance to be sitting in a virtual meeting room manipulating virtual objects and having the experience of 'actually' being there.

Virtual reality will not only change how we do business. Smart glasses and virtual reality goggles will enable us to be immersed in a world where the virtual and real worlds are glued together and will be involved in every aspect of our lives where designers can view their designed buildings before being built, shoppers can read reviews on products in supermarkets just by looking at them and many more applications.

Remote surgery and remote object manipulation

Successful remote surgeries are already performed today using fixed networks such as dedicated ATM lines or fibre connections. The haptic technology necessary for providing the correct feedback to the surgeon requires very strict requirements in terms of latency and security. A related scenario is represented by robots cooperating with humans for manipulating remote objects, e.g., in disaster-recovery, exploration of dangerous and hazardous areas, displacement of radioactive material and many more use cases.

Massive Internet of Things (IoT)

The vision of the connected world of 2020 and beyond includes an exponential increase in the number of connections which will mainly be driven by machine-to-machine

communications rather than human communications. IoT includes a plethora of applications including: (a) Smart wearables which are expected to go mainstream for health attributes such as pressure, temperature, heart rate, blood pressure, body temperature and breathing rate. (b) Sensor networks where smart services will become pervasive in urban environments including metering, building lights management, environment monitoring, traffic control and other smart cities features. (c) Mobile video surveillance will evolve to be available on aircrafts, drones and cars for monitoring houses/buildings and areas of interest. All these applications require an infrastructure that can manage thousands and millions of devices generating traffic of different volume and characteristics, and operating this vast network of devices in an efficient and sustainable way is very challenging.

1.1.2 5G requirements

As discussed before, 5G should, at the same time, enable new services, reduce operational costs and improve performance for traditional services. To achieve these strategic goals, there is a need for introducing a set of technical requirements that reflect a multitude of use cases. For example, the Tactile Internet will require providing very low End-to-End (E2E) latency (in the order of 1ms) with medium to high bit-rates and no other major requirements on reliability. Autonomous cars or controlled robots or drones will require a low E2E latency, high reliability, and high-device density with moderate bit-rates. Immersive reality at distance will require very high data-rates with no major requirements on reliability. It is clear that it will be difficult for one 5G system to be able to satisfy all these requirements at the same time. However, it will be possible for different subsets of requirements to be satisfied on different carriers or sub-systems. For instance, reliability demanding applications could be deployed on low frequency bands whereas data rate demanding applications could be deployed on higher frequency bands where there is more spectrum. This can be extended to the core network where the concept of network slicing [5] could be used where different network slices would be created to satisfy the requirements of different applications or traffic streams.

In this section we discuss a selection of 5G requirements which we think are the most disruptive. A detailed list of the requirements can be found in the NGMN white paper on 5G [1].

A focus on quality of user experience

For short range communications, speeds in the order of tens of Gigabits/s will be technically possible thanks to the continuous network densification and the large bandwidths available at higher frequencies. However, there is more or less a consensus now that more focus should be put on improving the cell edge capacity and less on the peak capacity. To this end, delivering high speeds in a wider area is much more challenging. Therefore, the focus should be on technology enablers that will guarantee the minimum required rate for a satisfactory result.

It was actually found that the fluctuations in throughput highly affect the perceived quality of users experience in [6]. Lab experiments have shown for users browsing pictures stored in the cloud, the peak and average throughput were not as critical to user experience as keeping a consistent throughput level with low fluctuations. Therefore, there is a need to shift our focus towards providing a more ubiquitous and consistent service rather than optimizing the peak or even average throughput. This also implies a change in marketing strategies with a focus on the user experience instead of peak rate values.

Ultra low latency

As discussed in the use cases, some applications such as the tactile internet would require an end-to-end latency in the order of 1 ms with some use cases such as the ones involving robotics requiring 100 μs. Achieving such latencies would enable new services and new business opportunities in different industries for 5G to be involved in. A latency as low as 1 ms would allow provisioning services in automation of robotic workers in factories where the low latency would allow the robots to tightly interact with each other and even share the same working space with humans. However, imposing a latency in the order of 1 ms as a fixed requirement would lead to an over-dimensioning of the system since low latencies are only required for some specific applications.Therefore, low latency applications could be delivered on a separate carrier from the other applications. They can also use a dedicated network slice and can also make use of new architectural features such as edge computing, caching and short transmission time interval.

Reliability and availability

Today's cellular networks mainly provide best-effort services: having a data connection or coverage everywhere and at any time is not guaranteed at all. However, many

applications that are requiring much more than the best-effort performance are getting more and more popular such as enterprises that are hosting critical functionality in the cloud, public safety and traffic control services. 5G has the opportunity to create new business opportunities by enabling operators to offer service level agreements (SLA) that guarantee a certain level of reliability and availability.

Reliability can be defined as the probability of successfully receiving the transmitted message within a given time threshold [7]. Availability can be defined as the probability of successfully receiving the transmitted message in a given area under a given threshold. In reality, availability can be measured in terms of network outage. That is, while reliability focuses on delivering a service with a given latency, availability focuses on providing that service in a given area.

If 5G could offer a reliable service with ultra-low latency and a certain level of availability in a certain area, this would allow 5G to become a major player in vehicular safety applications whereas offering reliability and low latency in a restricted area could enable other applications such as medical applications.

Devices density and energy consumption

Current cellular systems have been designed to operate with a few hundred devices per base station. In addition, architecture, protocols and components have been designed to allow devices' batteries to last from a few hours to a few days. With the rise of the internet of things (IoT), these two design concepts would need a major rethinking in order to support the massive numbers of connected devices where the requirement for some applications that the battery life would be extended to several years.

Designing a system to support massive numbers of devices that sporadically transmit little amounts of data would require major changes in the communication paradigm. First, existing coding methods that rely on long code-words are not applicable for short data blocks [8]. Second, short data blocks amplify the inefficiencies associated with control and channel estimation overheads. Finally, an efficient way is needed to handle this large number of devices in an efficient way which would require changes in layer 2 and 3.

The need for low battery consumption is related to applications where there is no fixed power source and the device is not easily accessible for maintenance. Therefore, the need for long battery life is an essential requirement for the economical viability of these applications. Hence, the potential to manufacture or configure a device that can operate with an extremely long battery lifetime is a very useful breakthrough that will allow 5G to have an edge in several industries.

1.2 Cell centric architecture

Cellular networks have historically relied on the axiomatic role of cells as the cornerstone of the whole radio access network. The word 'cellular' and the conventional way of representing a cellular network with hexagonal cells are obvious indications of the dominance of the cell-centric approach in designing radio access networks. In such a network, a device is served by the network by establishing a downlink (DL) and an uplink (UL) connection, carrying both control and data traffic, with the base station controlling the cell where the device is located.

This approach was sufficient in earlier generation of mobile telephony where the network was fairly homogeneous in the sense that it was composed of macro cells with similar coverage and similar transmit powers. The traffic was mostly based on voice and messaging and the density of the network was relatively low. However, in recent years we have seen that the base station density has been rapidly increasing driven mainly by the proliferation of heterogeneous networks with base stations of different sizes and transmit powers. While heterogeneous networks were standardized in 4G, the architecture was not natively designed to support them in an efficient way. Network densification could require a major rethinking of the design concepts that were taken for granted until now. An example of the new concepts that need to be adopted is the decoupling of uplink and downlink which is the main topic of the thesis.

In addition, the spectrum scarcity suffered by cellular networks will lead to the exploration of new spectrum opportunities, particularly in the millimeter waves band which will lead to another degree of heterogeneity in spectrum. Millimeter wave bands have different characteristics than sub 6 GHz bands where millimeter waves bands have vast amounts of free spectrum but poor propagation characteristics. This leads to the concept of control and data separation where the data and control planes are separated: the control information is sent via macro cells at sub 6 GHz frequencies whereas the payload data is conveyed by small cells at millimeter wave frequencies.

Based on these trends and many more, the cell-centric design approach should evolve into a more device-centric approach as will be detailed in the sequel.

1.3 Device centric architecture

A device centric architecture is defined in [9] as an architecture where a given device (human or machine) should be able to communicate by exchanging multiple information flows through several possible sets of heterogeneous nodes. In other words, the group

of nodes that are providing connectivity to a certain device and their functions should be tailored to fit the specific needs of that device. This basically calls for blurring the rigid cell borders we have today and treat the network as a single entity where the devices are served by the node or set of nodes that fit its requirements disregarding where the device is located. The following concepts constitute the main building blocks of the device-centric vision and some of them will be covered in details in the following chapters.

1.3.1 Decoupled uplink and downlink association

The trend of cell hyperdensification with multiple tiers requires different approaches in network planning and design to meet the fundamental objectives of maximizing user rate and empowering fair spectral efficiency, among others.

Typically, small cells (Scells) have Very low DL transmit power which results in low Scell coverage, providing very little user migration from the Mcell. The DL transmit power disparities among the different nodes and cell selection based on the DL received power result in imbalance problems between both UL and DL connections. The DL/UL imbalance problem has been recognised by 3GPP in [10, 11]. A UE is said to be in this situation if the best serving cell from an UL received power perspective and the best DL serving cell are different. The UE optimal connection in the UL is to the Scell as it is closer, but in the DL it would be connected to the Mcell. Classical cell association rules have a huge impact in the load imbalance and the UL performance. Fig. 1.1 shows a graphical example of the heterogeneous networs (Hetnet) imbalance problem.

The DL coverage of the Mcell is much larger than the Scell due to the large difference in the transmit powers of both tiers. However, in the UL all the transmitters, which are battery powered mobile devices, have about the same transmit power and thus the same range. Therefore, a user equipment (UE) that is connected to a Mcell in the DL from which it receives the highest signal level might want to connect to a Scell in the UL where the pathloss is lower to that cell.

This leads us to the concept of Downlink and Uplink Decoupling (DUDe) which is the main topic of this thesis.

1.3.2 Control and data separation

The increased heterogeneity in infrastructure where macro and small cells coexist and in spectrum where higher frequency bands complement the lower bands have led to a different thinking in relation to control and data plane collocation. In this

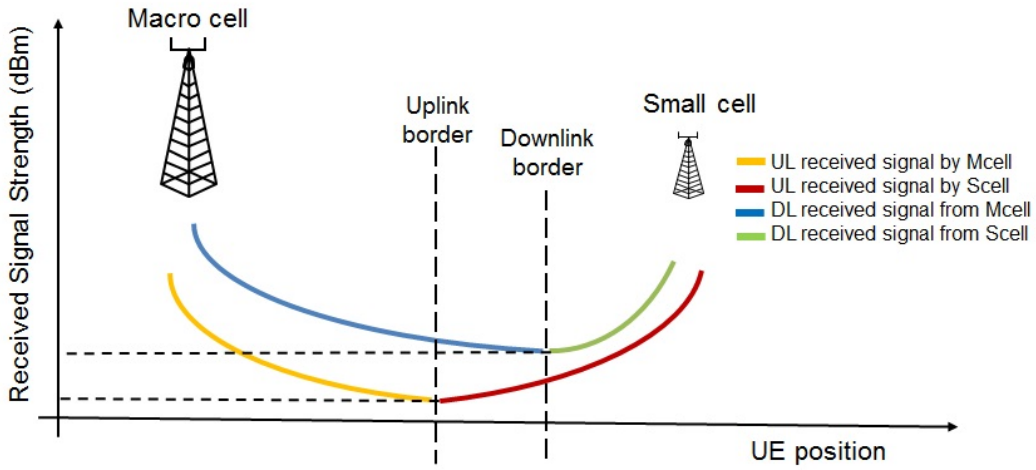


Fig. 1.1 Example of HetNet imbalance problem

context, [12] proposed the concept of a 'phantom cell' where the control and data planes are decoupled. The high power Mcells are used to send the control plane using lower frequencies whereas the data plane is delivered using low power Scells at higher frequencies where there is more bandwidth available but the coverage is less consistent than lower frequencies.

The main idea of control and data separation originates from the fact that only a limited amount of control signalling is needed to enable a ubiquitous coverage [13] where the control signalling needs to continuously be delivered reliably to guarantee coverage. Alternatively, data transmission is an on demand service where UEs send or receive data when needed only. This feature can be enabled in LTE using the dual connectivity feature [10] which allows a device to simultaneously connect to a Mcell and Scell on different frequency bands.

The control/data and uplink/downlink separation constitute a fully decoupled network where the data/control planes of a UE are connected in the UL and DL to the optimum node.

1.3.3 Device-to-Device communication

Device-to-Device (D2D) is a mode of communication that was introduced in LTE Release 12 [14] which entails the possibility of establishing a direct link between two devices without the need for the data to go through the cellular infrastructure. This mode of communication is quite important in current and future networks for the following reasons. In voice centric networks, it was conventional for parties calling

each other to not be in close proximity. However, in today's data centric networks this assumption might not hold in all cases, as there might be cases where co-located users would want to exchange content (e.g. pictures and videos) or interact with each other (gaming and social media) or even for public safety reasons. Handling these types of communications through the conventional way by connecting to the cellular infrastructure is highly inefficient for several reasons:

- Using the conventional way of transmission entails several hops to achieves what requires a single hop between the UEs. In addition to the extra hops on the radio side, on the core side the data might traverse large distances before being transmitted to the receiver depending on the network architecture.
- Using the cellular network for this type of communication is inefficient in terms of energy consumption as it entails an UL transmission of a fraction of a Watt and a DL transmission of several Watts for a connection that requires a few milli-watts to be established. This results in an inefficient use of the limited device battery power as well as higher level of interference.
- Given that the UEs are in close proximity, this means that the pathloss between the UEs might very well be less than that between the UEs and the serving BS which means that the direct link might achieve a higher spectral efficiency.

D2D is quite important for applications requiring low latency as well as having the possibility of establishing a communication link out of coverage. This makes D2D very relevant for vehicular communications and public safety applications.

D2D constitutes an important part in the device centric architecture where the data is forwarded between two co-located users through a direct link, which is the most efficient way, without the need to follow to conventional cell centric approach.

Some challenges associated with D2D include finding an efficient cell association technique for D2D as well as resource management techniques that improve the D2D achieved rate with minimal effect on the cellular transmission. These factors will be addressed in chapter 5 where we study the cell association and resource management in D2D in great details. Another important challenge is that, so far, D2D communication is restricted to the same operator which would cause problems for vehicular technologies where safety applications require seamless communication across operators. This aspect will be addressed in our future work.

1.3.4 Software defined networking and Virtualization

Software defined networks

The need for scalability, resource efficiency and flexibility in future networks has led to the adoption of software defined networks (SDN) in cellular networks [15]. SDN works by decoupling the control plane and the data forwarding plane and abstracts the network functions offering the freedom to dynamically configure the routing and forwarding logic. The control function is logically centralized in an SDN controller which acts as the brain for the network. The SDN controller has full knowledge of the network nodes and dynamically updates network policies according to flow activities. This guarantees the scalability of the network as the SDN controller has the ability to dynamically discover and reconfigure the forwarding mechanisms. In addition, application programming interfaces (APIs) are supported to link the application and control layer in order to facilitate the management of the network. The development of the SDN concepts are mainly led by collective efforts such as the Open Networking Foundation (ONF) [16].

SDN could be enabled using several protocols such as Openflow [17] which can be used as a standard communications interface between the control layer and the forwarding layers. In particular, it provides access to the forwarding plane of physical and virtual network devices (e.g. switches, routers).

Network virtualization

The idea behind network function virtualization (NFV) [18] is to run the cellular network functions as software instances on data centres, network nodes and user devices. This essentially aims at reducing the hardware dependence, the maintenance cost and accelerates the adoption of new features which will just be in the form of software updates installed on generic servers.

Virtualization leads to the decoupling between the node function and the hardware allocated to handle the processing associated with this node. This leads to flexibility in assigning resources from the 'pool of resources' to the node where they are most needed.

NFV has the potential of flexibly locating network functionalities and resources in a vendor agnostic fashion and also there is the potential of combining network functions in on-demand adaptive manner, hence creating a link with SDN.

SDN and NFV are crucial feature for device-centric architectures as they allow the creation of different slices or instances of the network in order to satisfy the different requirements for different traffic types.

1.3.5 Cooperative communications

Over the years the importance of intercell interference (ICI) has been recognized and various techniques have been used from the days of GSM to mitigate its effects. Interference averaging techniques such as frequency hopping were used. However, as technology has advanced, much tighter and more effective methods of combating and utilising the interference have gained support. LTE coordinated mutipoint transmission (CoMP) [19] is essentially a range of different techniques that enable the dynamic coordination of transmission and reception over a variety of different base stations. The aim is to improve overall quality for the user as well as improving the utilisation of the network.

The main idea of CoMP is as follows. When a UE is in the cell-edge region, it may be able to receive signals from multiple cell sites and the UE's transmission may be received at multiple cell sites regardless of the system load. Given that, if the signaling transmitted from the multiple cell sites is coordinated, the DL performance can be increased significantly. This coordination can be simple as in the techniques that focus on interference avoidance or more complex as in the case where the same data is transmitted from multiple cell sites. For the UL, since the signal can be received by multiple cell sites, if the scheduling is coordinated from the different cell sites, the system can take advantage of this multiple reception to significantly improve the link performance.

Essentially CoMP, falls into two major categories [20]:

- Joint processing: It occurs where there is coordination between multiple entities (base stations) that are simultaneously transmitting or receiving to or from UEs.
- Coordinated scheduling or beamforming: It is often referred to as CS/CB and it is a form of coordination where a UE is exchanging data and control flows with a single transmission or reception point. However the communication is made with an exchange of control among several coordinated entities.

To achieve either of these modes, highly detailed feedback is required on the channel properties in a fast manner so that the changes can be made. The other requirement is for very close coordination between the base stations to facilitate the combination of data or fast switching of the cells.

Coordinated transmission is an important aspect of device-centric architectures as it's core idea relies on the fact that the network nodes cooperate to serve the UEs in the most efficient way possible regardless of where the UE is.

1.4 Quality of experience in 5G

According[1] to the Next Generation Mobile Networks (NGMN) alliance, "5G is an end-to-end ecosystem to enable a fully mobile and connected society. It empowers value creation towards customers and partners, through existing and emerging use cases, delivered with consistent experience, and enabled by sustainable business models" [1]. 5G is a system both for "Things", i.e. cars, smart meters, machines, etc. and for human-centric devices and services. Whilst defining the exact requirements for things in terms of latency, data rate, coverage, reliability, security, etc. is important, this section has a more focused scope to human-centric services and devices, i.e. where a "Quality of User Experience" can be defined.

Overall, we see a consensus regarding "what 5G is", which in fact is envisioned as a system to serve the user and the society. Moving beyond the "edge rates" that were already discussed in the context of 4G, what should really make 5G different from previous generations of communication networks is that it will be designed not to offer rigid peak data rates anywhere, anytime and to anyone, but to perform a much more meaningful, flexible and personalized network management based on the understanding of the end-user's and the service's needs. Thus, terms such as "user experienced data rate" and "QoE", in general, are now in the spotlight.

Satisfying the customers' requirements implies the necessity for moving from system-centric to more user- or even human-centric designs. However, in order to design the 5G technology in a user-centric way, we first and foremost have to understand what it is that users really want. For instance, based on a recent study [21], the dimensions that really concern users are the following: reliability (47%), coverage (36%), and data speed (9%), (and other - 5%). These results should not be neglected; they give a clear message that the user mostly cares about reliability, i.e. consistency in the perceived experience, as well as coverage, i.e. availability and uniformity in the delivered service.

To complement the previous insights, it seems rational that in order to guarantee the best and most reliable experience to the end-users, and to ensure that this experience "always follows you", some new requirements need to be identified. These are, amongst others, customer personalization and service differentiation; indeed, different services

[1]The work in this section is a part of [6]

or use cases have different QoE requirements. Moreover, the end-users themselves are not really uniform with respect to their QoE requirements; for instance, their prior knowledge and expectations, along with the current usage context have a strong influence of their QoE needs [22]. Similarly, different services or use cases have different QoE requirements. QoS, as currently defined and implemented in LTE, is not sufficient to achieve such fine differentiations; on the contrary, these can only be efficiently addressed through QoE-aware network and application management approaches. To reach the maximum potential of this opportunity, QoE needs to be inherently supported in 5G systems, similarly to how QoS is today. Adding a new "QoE-layer" on top of a rigid 5G architecture will not do the job anymore; hence, QoE needs to be handled as a native ingredient of the upcoming 5G network designs (in a holistic sense). Overall, QoE in 5G will likely be about better insights (real-time data collection), better decisions based on these insights (big data, machine learning and deep learning analytics) and the ability to actuate on the insights (intelligent networks, through e.g. software defined radio (SDR), SDN and NFV).

It is inevitable that the 5G system has to be designed 5-10 years before it is actually used, and therefore, we can only make assumptions regarding the type of devices and services that will be available then. Hence, we cannot really nail down QoE completely today, but it is worth making an attempt towards that direction, based on current information and future visions. Therefore, the contribution of this section lies, firstly, in identifying and outlining the QoE requirements in 5G ecosystems. Our second objective is to reflect on how 5G would guarantee the required level of QoE. Thus, specific technological enablers, either existing or emerging, are identified, and their impact on shaping the perceived user experience is analyzed. Consequently, the importance of these must-have components in 5G systems is stressed. Ultimately, with this section, we expect to influence the research community in taking the described QoE requirements and technological enablers strongly into consideration while designing the 5G networks of the future.

Having discussed the device-centric architecture approach in the previous sections from a system level, we now move to a different part of the device-centric concept which is the personalized QoE.

1.4.1 QoE requirements in the 5G ecosystem

In this section, we sketch what the customer experience should look like in the 5G ecosystem by describing its desired requirements in terms of QoE. The main objective is to draw and emphasize the necessity of these requirements as the only way to provide

an excellent and solid user experience, as expected by the next generation of cellular networks. It is crucial, that these attributes are identified at this early stage, so as to push towards the design of more user-centric 5G networks, which will enable these requirements using current or emerging technologies. With this objective, we identify that the end-user experience in 5G ecosystems should have the following characteristics:

Consistency

The requirement for "consistency" has been clearly identified in the vision of NGMN [1]. It refers to the uninterrupted, seamless and invariable (or with as low variance as possible), but still excellent quality of the offered application or service. Consistency spans across many dimensions such as time, space, infrastructure, operator/vendor, end-device and application. Therefore, a 5G user should expect to receive a continuous service, without disruptions, and with as few fluctuations as possible. Some of the main obstacles in achieving this requirement in a mobile environment are the uncontrollable and unstable channel conditions, the competition over the scarce spectrum resources and the high heterogeneity and density of these networks, causing constant handovers and unpredictable interference levels. To overcome these challenges, traditional network management decisions will have to be revisited and transformed to smarter, QoE-aware mechanisms. Such mechanisms will then be able to account for the impact of various QoS-based parameters on the user experienced quality, and drive the network operations accordingly.

Transparency

Transparency refers to the requirement of the network to "hide" its complexity and efforts on delivering excellent and seamless quality to its customers. Best experience should always follow the end-user, while he/she is spectrum and system agnostic. This means, that although the end-user is considered to be the epicenter of a 5G network, his/her implicit input or intervention in any network or service management decisions should be avoided. For instance, even though providing a solid experience is a clearly subjective issue, the user is not expected to be actively involved in QoE measuring and monitoring procedures; this should be done automatically by the network either by exploiting passive feedback from the user's application or device or by using network probes and deep packet inspection techniques.

User Personalization and Service Differentiation

Services provided over a 5G network should be tailored to specific users or user profiles. The key to achieving this is through enabling QoE personalization inside the network. Netflix already distinguishes among gold/silver/bronze users, based on their subscription profiles, and configures the offered quality accordingly. However, explicitly paying for a subscription profile and, thus, receiving correlate quality is just one possibility of enabling QoE personalization.

Except for differentiation on a per-user basis, a differentiation per service and application type is expected. QoE is tightly dependent on the type of application, and different QoE requirements are needed for different applications [23]. Therefore, it is required that the 5G systems are flexible when serving diverse applications, tailoring their quality monitoring and provisioning techniques a) according to the different influence factors per application, b) according to the different impact and tolerability that the same QoS parameters have on different applications, and c) according to the applications' adaptability to varying network conditions.

To achieve this degree of personalization, the provisioned QoE in 5G networks should account for the context of each communication session, a very challenging task. The term "context" stands for anything that may have an impact on the QoE, such as the user demographics, the end-device, the current spatiotemporal environment of the user, the urgency and nature of the session, the social context, the content transmitted, etc. For instance, high demanding users (e.g. business users) may be prioritized over users who would not perceive or would not care about some extra delays during their communication sessions. Another example is to automatically setup a new connection as Device-to-Device (D2D) based on proximity-awareness of two communications entities. Reaching to such a context-awareness may enable a more meaningful network and service management and, thus, become a powerful tool of 5G networks.

Resource and Energy Efficient QoE-awareness

Adding QoE awareness and, in turn, QoE-aware service and network management mechanisms will inevitably introduce more complexity in the network. For instance, periodic QoE probing and monitoring will have to be implemented in both edge and core network nodes, increasing their battery consumption. Similarly, extra control signaling overhead will be imposed in the access network, which may cause a resource-insufficiency problem. Since enabling QoE in the network is translated to such resource

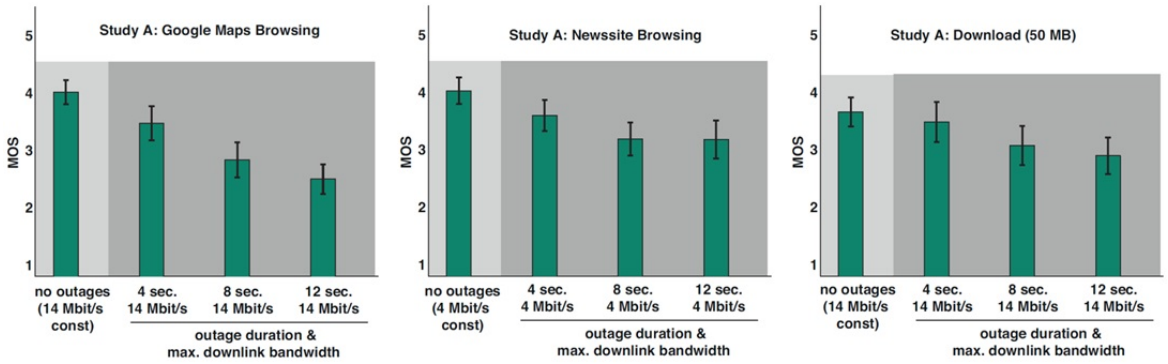


Fig. 1.2 Impact of downlink outages duration on QoE (MOS ratings) for different Web applications.

and energy costs, we need to make sure that the energy and resource costs per "QoE unit" are maintained to a reasonable, minimum level. Opportunities to control these ratios may stem from the science of human perception (psychophysics), amongst others. In the following section we describe how the aforementioned requirements may be enabled by exploiting the architecture and technological advancements that 5G brings/should bring.

1.4.2 Technological enablers

The QoE requirements described in the previous section are not just a "wish list" for the envisioned future mobile networks. 5G is already on the horizon and provides fertile ground for realizing these requirements and shaping an excellent user experience. The discussions on 5G usually revolve around Massive MIMO, Millimeter Wave, etc. Below, however, we identify and describe those main technological enablers, existing or emerging within 5G, which can contribute towards shaping the Quality of User Experience, as this was described in the previous section.

Quality Fluctuations Management

As has been stated in the previous section, enabling consistency of service delivery is a central requirement for the emerging 5G ecosystem. Consequently, understanding and managing the impact of time-varying quality (including disruptions and performance fluctuations) has become an essential building block for realizing the 5G vision. In this context, the QoE impact of time-varying network quality bears particular relevance, not only because for wireless connections' performance fluctuations actually are the norm (due to changing cell load, interference, handovers and outages), but also since

increased peak rates (as enabled by each new access technology generation) lead to increased likelihood and magnitude of performance variations [24].

However, so far attempts to quantify the relationship between network QoS and end-user QoE mainly have focused on developing metrics and KPIs based on average values: existing models and mappings for services and applications (voice, video, Web) are typically based on averaged QoS parameters (such as mean download throughput of the traffic flows or the mean round-trip time (RTT) of the access network) and QoE. This approach has the great advantage of making derived models easier to understand, compute and integrate into live network monitoring and reporting systems. However, relying on average values has also a fundamental limitation: it assumes that within a given time period end-user QoE is determined by the averaged stimulus and not by transient behaviors - an assumption that has already been disproven by several recent studies [25–27].

For example, interactive Web applications like Google Maps are heavily impaired by network outages, while less interactive applications (like browsing a news website) or file downloads tend to be less sensitive w.r.t. outages (see Fig. 1.2) [26]. Both effects (user sensitivity to fluctuations, application-dependency of QoE impact) become even more visible in results from lab studies where the average throughput in all conditions (constant, varying) were kept on identical levels (see Fig. 1.3) [27].

The above examples demonstrate that the QoE of networked services is not only dependent on high speed and low latency connectivity, but also on network stability and reliability. Given the relevance of fluctuations as a QoE influence factor, extended models, metrics and KPIs are required to enable the systematic quantification and measurement of their impact on QoE. This topic is currently being addressed by ongoing research [27, 28]. So far, approaches utilizing KPIs like the effective average downlink throughput (EADT) that captures fluctuation impact by means of a correction factor yield promising results [27]. In this context, the evaluation of various modelling approaches based on such KPIs like EADT has shown that metrics that take the duration and intensity of throughput drops into account perform significantly better in terms of QoE prediction accuracy than models based on average throughput only [28].

Consequently, a fluctuations-aware quality management has strong potential to improve the end-user QoE by, depending on the application being used, utilizing mechanisms that a) either prevent quality fluctuations (e.g. on network level) or b) mitigate their visibility and impact. Examples for a) are: novel resource management schemes, traffic shaping and network dimensioning strategies. Prime examples for b) are: buffers and caches, as well as adaptation mechanisms that adapt application-level

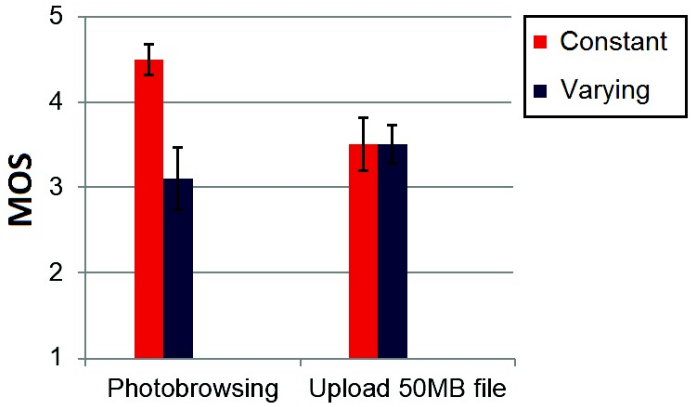


Fig. 1.3 Impact of varying throughput (alternating between 14 and 1 Mbps) on QoE for web photo gallery browsing and file upload.

performance (e.g. video picture quality) to current network QoS (as DASH does in the case of streaming video). On top of these measures, in the ideal case, networks and applications would be able to exchange QoS and QoE-related information with each other. Such approaches that – while not having been implemented in today's stacks yet – would be able to very efficiently and effectively mitigate the impact of QoE impairments like fluctuations by better matching supply and demand (cf. [29]).

QoE-aware Resource Management

The resource management procedures in 5G networks should be mapped from the QoS- to the QoE-domain. This is the only way to account for the impact that various technical parameters, such as throughput, have on the user's perception, which is the ultimate criterion for quality assessment. Numerous subjective studies have been conducted in order to understand the non-linear relationships that rule these two quantities, with perhaps the IQX and WQL hypotheses being the most characteristic ones. The former [30], describes the way in which the degradation of one QoS parameters affects the humanly perceived QoE, while the latter [31] describes the effect of a physical stimulus on the human perception. By observing such interesting relationships between QoS parameters and the user-experienced quality, opportunities emerge to design more powerful and more meaningful resource management procedures, in terms of resource and energy efficiency.

Furthermore, in order to unveil the true potential of QoE-aware scheduling, novel cross-layer approaches need to be devised. These scheduling methods should optimize both the networking and application layer status, which can be only feasible through an interaction between those two layers. The network status includes information

such as the availability in network resources, the channel conditions, fairness among users, required data rates, etc., while the latter incorporates information such as the application type and current state, level of interactivity, adaptation capabilities, or even fed-back Mean Opinion Scores (MOS), etc. For instance, in [32] a scheduling strategy is proposed, where the adaptive capabilities of Skype are exploited in order to save some bandwidth for resource-hungry YouTube flows.

Downlink and Uplink Decoupling

Cellular networks have often been designed based on the DL due to the fact that it is the most demanding link in terms of capacity. Similarly, QoE has so far been more focused on the DL since most of the traffic in the network is DL (e.g. video streaming). However, the UL is becoming more and more important with the increasing popularity of symmetric applications such as social networking, video calls and real-time video gaming. In these applications, the UL is as important as the DL in terms of QoE, as a satisfactory experience cannot be guaranteed with a poor UL connection. An enabler of improving the UL is the decoupling of the UL and DL cell association [33], which is considered as an important aspect of user-centric network architectures. The basic concept is to treat the UL and DL as two separate networks to which the users attach based on different criteria. This is motivated by the fact that in a heterogeneous network by having a large disparity in the transmit power of the different layers, there is an imbalance between the UL and DL coverage areas, as shown in Fig. 1.4.

The work in [33] considers the case where a user would perform the DL and UL association based on DL received power and pathloss, respectively. This would allow the users in the region between the DL and UL cell borders of a small cell to connect to the small cell and macro cell in the UL and DL, respectively. This approach has shown very high throughput gains in dense heterogeneous networks in the order of 200-300%. This is mainly due to the load balancing effect in the UL, which results in a more efficient usage of the resources. This study also shows that in a high traffic scenario, outage can be reduced by about 80%, which would guarantee a consistent and reliable service even in a very loaded system. Also the fact that the users associate to the nearest node in the UL would result in a more efficient usage of the users' battery power. The improvement of the UL in terms of capacity, energy and resource efficiency would translate directly in an improved QoE, especially in symmetric or UL critical applications.

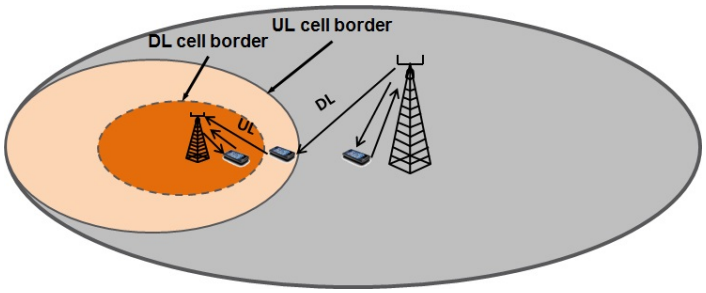


Fig. 1.4 The concept of UL/DL decoupling.

Big Data Analytics

A large amount of user -and network- related data is flowing through the telecom's network at any time. It is thus to no surprise that Big Data approaches are gaining in importance, where the shift is from structured, non-real-time data towards unstructured real-time data. The aim of using big data analytics is largely to improve internal operations of the network, as well as improve interaction with the customer. Both have profound impact on the QoE of the end-user. Specific benefits from running big data analytics in real-time are i) using insights into customer behavior and usage to develop new services and products; ii) be better in identifying fraud, and optimize routing and technical QoS by analyzing network traffic in real time.

Furthermore, from a network point of view, big data allows configuring the system more intelligently which in turn translates into a better QoE. Take the example of a user catching up with last night's football using BBC's iPlayer whilst driving down the Strand in Central London. Big data allows matching connectivity coverage holes with the direction and speed of the user; this, in turn, allows pre-emptively downloading parts for the user by intelligently invoking all required layers and networking entities in a future 5G system. Overall, big data allows the network to be less spontaneous and more intelligent, i.e. able to take decisions beneficial in the near future based on prior historic insights.

Another relevant point that is worth mentioning is the vision that the 5G operator will be able to enhance the user and service experience via securely collecting information and performing such data analytics [1]. This triggers a potential to expand data analytics towards realizing build-in QoE prediction, via the collection of meaningful, end-to-end quality KPIs. Such KPIs could ensure that network operations are more focused on QoE, rather than KPIs that don't really matter to end customers (e.g. average downlink data rate values versus number of stallings when watching a YouTube video).

Software Defined Networking

Flexibility, scalability and service-oriented management have been identified both by [34] and [1] as the main drivers of the 5G architecture development. For instance, [34] envisions that 5G systems will be able to steer network capabilities on demand. A technological paradigm that is highly promising to enforce these main drivers is Software Defined Networking (SDN) [35]. SDN enables the creation of a centralized, abstract view of the underlying network's physical topology and network state, and sequentially facilitates its agile, adaptive and flexible control. This is made possible via the decoupling of the control plane (decision-making) and the data plane (traffic forwarding). This makes SDN an important enabler towards providing intelligence and innovation in the network, in a broad sense.

As a logical extension, SDN may prove to be the key technology also for the QoE management and provisioning functions. This could be realized, in general terms, by creating a virtualized control plane of the network infrastructure (network slicing) and then, dynamically imposing management decisions to the affected network nodes, by providing respective programmable instructions (via e.g. a "QoE orchestrator"). The main benefits of controlling QoE via SDN would be to allow for a unified, less complex, less expensive, and adjustable configuration of the network behavior, in a simple, software-based way. As a consequence, the faster introduction of innovation in QoE-centric policy enforcement and network management will be enforced. Another possibility offered by SDN, is the flexible acquisition of QoE-related information from the network infrastructure, towards the enhancement of QoE for Over-The-Top (OTT) on-demand services or for OTT premium users. Such an approach will be able to unlock additional business value, not only for the OTT providers, but also for the network operators, who by using such a paradigm could enter the revenue loop between the former and their customers.

1.4.3 Challenges

Apart from the technical challenges towards implementing or exploiting the enablers discussed in the previous section, there exist business and legal issues that need to be addressed. Especially, if we consider the plethora of different stakeholders potentially involved throughout the end-to-end service provisioning chain, such as OTT providers, different mobile network operators or Internet Service Providers (ISPs), content providers, vendors, etc., we see that achieving end-to-end quality becomes a very challenging task in business terms. Often, different or even contradicting

interests need to be addressed. First and foremost, we see that contradicting interests emerge between operators and content providers (such as Facebook, Google, etc.). The latter design their new technologies and services with security (i.e. encryption/content labeling) in mind. Encryption, however, is an "enemy" for QoE-awareness and in turn, for flexible QoE delivery. Unencrypted data, on the contrary, can be a powerful tool for the operators as the source of information to ensure, enhance or adjust QoE, or to make service differentiation. To achieve such a differentiation, the operator actually needs to know the application type, its current state, etc. As an example, by having access to the buffer state information of a video playout, the operator can prioritize the limited resources available at a specific time and location in order to maximize a certain utility function, e.g. maximize the number of satisfied users/services.

Network Neutrality (NN) is another major issue when it comes to QoE provisioning. From the operator's perspective, NN may not leave enough space for innovation and investment in the networks in terms of QoE. Furthermore, even though the recently voted NN regulatory framework [36] has been welcomed by most OTT/content providers as a way to allow flawless access to their services, it is not a black or white issue. For instance, the dynamic allocation of "fast lanes" may no longer be allowed by the network providers to pass, say, Netflix content to premium users or to do any other type of service differentiation. Moreover, there are business implications: with NN it may not even be allowed or meaningful to do customer differentiation and/or market segmentation at all and thus, inevitably block any possibilities for service personalization.

Finally, Service Level Agreements (SLAs) i) between the operator and the end-users, as well as ii) between the operator and the OTT providers need to be revisited. Presently, the requirements in such SLAs are described using QoS terms, which however do not directly imply a proportionally satisfactory QoE level for the operator's or the OTT provider's customers. Consequently, new types of SLAs, or so-called ELAs (Experience Level Agreements), may be considered that define the required quality using QoE terminology. The great challenge in this new approach is finding a way to clearly define the various QoE classes using a "vocabulary" that ensures common understanding between all stakeholders involved [37]. Then, it is further required to devise indisputable methods of measuring the QoE at the various interconnection points, of checking it against the SLAs' QoE requirements and of finding which side of an interconnection is responsible in case of QoE deficiencies.

1.5 Goals and Organization of Thesis

5G is expected to create new business opportunities allowing the Telecom industry to evolve from just offering voice and mobile broadband services to having a pivotal role in vertical industries such as the automotive industry, healthcare, factory automation massive machine-to-machine and many more verticals where the limitations of 4G and previous generations prevented the cellular industry to play a leading role in these industries. In order for 5G to fulfil this ambition, new ways of designing the network need to be thought of. One crucial postulate that 5G is expected to break is the cell-centric design concept where the focus is more on cells and less on individual users. A move towards a device-centric architecture is essential where the network nodes and their services are tailored to serve the specific needs of users. In other words, the user has to be at the epicentre of the network's focus and the network must be flexible and agile enough to be able to assign to the users resources that satisfy their needs whether these resources are network nodes, bandwidth or cooperative transmission to name a few.

In this thesis, we study in great details two of the main building blocks of device-centric architectures, namely Downlink and Uplink Decoupling (DUDe) and Device-to-Device communications where the two techniques are considered to revolutionize the traditional way of communication in cellular networks.

In traditional cellular networks, it is practically an axiom that the uplink connection is always associated with the same base station that has been selected for downlink reception In this thesis we revisit this axiom and introduce the features of downlink/uplink decoupling (DUDe), a new architectural paradigm where downlink and uplink are not constrained to be associated with the same BS. This is becoming especially relevant in the wake of the densification expected in future 5G cellular networks, where each terminal has multiple access nodes in proximity. In a heterogeneous network, an imbalance between the uplink and downlink coverages arises from the difference in transmit powers of different tiers which results in users tending to connect to different cells in the uplink and downlink which makes DUDe very relevant. In chapter 2 we present a detailed simulation study of DUDe based on Vodafone's LTE testbed in London and using a 3D ray tracing algorithm. We show huge uplink gains for baseline DUDe with pathloss uplink association as well as the optimized DUDe where the load of the cells and the backhaul capacity are taken into account in the uplink association. In chapter 3 we present a detailed analysis of DUDe using tools from stochastic geometry. We start by analysing DUDe in a sub 6 GHz HetNet where we study the cell association probabilities and prove that the association trends are

more dependent on the density of the deployment rather than the process used to generate it. Subsequently, we study DUDe in a mixed millimeter wave and sub 6 GHz network where we provide detailed analysis on the association trends and the SINR and rate coverage trends considering cell biasing. Chapter 4 contains a detailed study of the enabling architectures of DUDe in 4G and 5G and interoperability of DUDe with other emerging technologies.

In chapter 5 we turn our attention to D2D where the direct communication is enabled between devices in proximity without the need for the traffic to go through the infrastructure. This technique is very crucial for latency sensitive applications such as vehicular communications. It also has major advantages in terms of spectral efficiency and energy saving. We start by studying cell association in a D2D enabled network where we show, using an optimization framework, that decoupled association results in reduced interference and improved energy efficiency. Subsequently, we study the radio resource management in relay aided D2D networks where we use genetic algorithms to reach a near optimal allocation of resources in terms of achieved rate.

Chapter 2

Simulation study of Downlink and Uplink Decoupling

This chapter[1] aims at introducing the main benefits and gains of Downlink and Uplink Decoupling (DUDe) supported by an extensive simulation based study demonstrating the performance improvements in the UL resulting from DUDe. The chapter starts by presenting the main motivations and benefits that arise from DUDe in Section 2.1. In Section 2.2 the previously discussed benefits are proven using a realistic simulation setup considering the basic implementation of DUDe where UL and DL cell associations are based on pathloss (PL) [2] and DL received power respectively. Finally, in Section 2.3 an enhanced version of DUDe that takes into account the UL cell load and backhaul capacity is introduced and simulation results that reveal its superiority to the basic DUDe implementation are provided.

2.1 Why to decouple the uplink and downlink?

2.1.1 Introduction

From the first to the fourth generation (4G) of mobile networks, the DL and UL of a given communication session have been coupled: the UE must associate with the same BS in both the DL and UL. Historically, this was a nearly optimal approach, since the strongest BS-UE connection was the same in both directions. However, this conventional approach has recently [33] come under scrutiny given the possible

[1] The work in this chapter is partly included in [33, 38, 39].

[2] Pathloss is the reduction in power density of an electromagnetic wave as it propagates through space. In a simplified way, pathloss can be calculated by subtracting the received power at the receiving end from the transmitted power radiated from the transmitter.

gains that can be achieved by decoupling the association in the context of a dense heterogeneous cellular network, wherein different BSs can have highly variable transmit powers and deployment topologies.

The arguments in favour of the coupled status quo are several. From a pure network design perspective, the logical, transport and physical channels are easier to design and operate; this pertains particularly to the synchronization of the acknowledgements, the call admission and handover procedures, DL/UL radio resource management, and power control, among others. Decoupling both links also requires strong synchronization and data connectivity (e.g. via fibre) between the BSs. From a deployment and topology perspective, until just a few years ago cellular systems have been designed and deployed under the assumption of a homogeneous network with macro cells all transmitting with about the same power. From a traffic point of view, the load in both directions has been approximately the same in voice-centric 2G and early 3G systems. Moreover, 3.5G (e.g. HSPA) and 4G systems are dominated by downlink traffic, justifying the use of DL-centric association procedures rather than UL or decoupled ones.

The emergence of Heterogeneous Networks (HetNets) [2], where small cells at higher carrier frequencies and/or smaller transmit powers are deployed within the coverage area of macro cells, calls for revisiting the coupled association approach. Range extension has been included in 4G to add a bias in the cell association to offload more traffic from macro to small cells. Data and control plane separation has been introduced in [3]: the control information is sent by high-power nodes at lower frequencies, whereas the payload data is conveyed by low-power nodes at possibly higher frequencies. However, both range extension and data/control plane separation are based on a coupled DL/UL association, where DL and UL are associated to the same BS.

The motivation for DUDe emerges from a holistic view on the two-way (DL/UL) traffic and the association procedure of a UE, rather than adopting a coupled association a priori and then optimizing separately DL and UL transmissions. Since a coupled association is a particular sub-case of a decoupled one, a well-designed association policy based on Downlink/Uplink Decoupling can in principle outperform a coupled association. But by how much? And at what cost?

More specifically, the main questions tackled in this section are:

- What recent trends in cellular network deployment and applications make the gains from DUDe more relevant now than in the past?

- What are the key benefits of a decoupled association in terms of throughput gain, reliability, and power conservation? What are the challenges? How can these gains be realized in current (e.g. LTE-A) and future 5G cellular networks?
- How disruptive will these changes be to the network architecture? Are the gains large enough to be worth the trouble?

We begin the discussion in this section with the five key arguments in favour of DUDe, these arguments are then proven with extensive simulation results in Section 2.2 for pathloss based DUDe and subsequently in Section 2.3 for backhaul and load aware DUDe.

2.1.2 Increased uplink SNR and reduced transmit power

In a typical HetNet scenario the downlink coverage area of a macro cell is much larger than that of a low power BS. The coverage area disparity is primarily attributable to the differences in downlink transmit powers, but is also due to the BS heights and antenna gains. In contrast, in the uplink all transmitters have roughly the same maximum transmit power. Therefore, a device that is associated to a macro cell in the downlink might instead wish to be associated to a small cell in the uplink, to take advantage of the reduced path loss. The positive effects are twofold. For UEs that are transmitting at the maximum power, a connection to a closer BS provides a higher SNR. Moreover, for a fixed target SNR, the reduced path loss alternatively allows transmit power reduction via power control.

2.1.3 Improved Interference Conditions

DUDe also decreases the uplink interference, due to multiple complementary effects. First, and as an obvious consequence of the transmit power reduction discussed in the previous section, the UL interference generated to other base stations is correspondingly largely reduced. This is quite significant especially for the low SINR UEs in the uplink, since at low SINR in a dense network, decreasing the interference by 3dB implies an approximate doubling of data rate.

Second, DUDe provides the ability to independently select the association that minimizes interference at both the UE and the BS. Uplink interference in a given spectral band is an aggregation of many different UEs' transmissions in different cells, as received by a given BS, say BS0. The interference generated by each of these UEs depends on its location relative to its own desired BS, the amount of power control, its

distance to BS0 and the uplink precoding weights. In contrast, the downlink interference at a given UE depends on the BSs' transmit power, the downlink beamforming weights and the distance to the different BSs. On top of this, the nearly independent scheduling and loading in each of the DL and UL causes further randomness in the interference. For all these reasons, average interference levels can be quite different in the downlink and uplink resources. Therefore, a decoupled association that allows the UE/network to seek out the best interference environment in the two links independently can be expected to substantially outperform a coupled association, which must "split the difference".

Third, DUDe will also prove a boon for Device-2-Device (D2D) communication which, as of 3GPP Rel. 12, will take place in the uplink bands. By lowering the UL transmit power and generating less interference, DUDe will create a more benign environment for D2D receivers and thus allow more D2D transmissions to take place.

Finally, in addition to reducing the amount of average interference, DUDe also allows a reduction of the uplink SINR variance, as will be shown in Section 2.2 which translates into more efficient and effective UL schedulers and performance gains [40].

2.1.4 Improved Uplink Data Rate

Unsurprisingly, increasing the desired received power and decreasing the interference leads to higher SINR, and hence a higher spectral efficiency and data rate. However, there are additional factors which can complicate the effect of DUDe on the uplink rate.

For example, consider an LTE HetNet with small cell range expansion and biasing. On average the optimal downlink bias is in the neighbourhood of 5-10 dB, although with blanking or interference avoidance, up to 18-20 dB may be used in certain scenarios [41, 42]. DL biasing leads to a better association in both directions even with coupled association, since by expanding the DL small cell coverage region, more UEs associate with the nearby small cells in the UL as well, which is also the main point of DUDe.

Nevertheless, we still observe very substantial rate gains for DUDe even when compared with biased coupled associations which will be detailed in Section 2.2.

2.1.5 Different load balancing in the uplink and the downlink

The load of a given BS in the UL may be different from the load of the same BS in the DL. This implies that it is not optimal to have the same set of UEs connected to the

same BS in both uplink and the downlink, such that at least some of the UEs should use decoupled access.

Additionally, DUDe allows pushing more UEs to under-utilized small cells in the UL only since it is not limited by interference as is the case in the DL.

2.1.6 Low deployment costs with RAN centralization

Implementing a decoupled cell association in a real network requires excellent connectivity and modest cooperation between different BSs. As we will discuss in the subsequent section, the main requirement DUDe imposes is a low latency connection between the downlink and the uplink base stations, to allow fast exchange of control messages, like HARQ messages. We emphasize that unlike the most sophisticated forms of Cooperative Multi-Point (CoMP), like joint processing, where a high throughput backhaul connection between BSs is required to allow rapid data exchange, DUDe does not impose a tight requirement on the backhaul capacity. Put another way, DUDe allows gains similar to uplink joint processing (about 100% edge and average throughput gain), but with lower deployment costs. Compared to using MIMO or new spectrum to increase the throughput, the cost comparison is even more favourable to DUDe [39].

The ongoing trend towards using partial or full Radio Access Network (RAN) centralization in deployments where a high-speed backhaul is available, will be an enabler for downlink and uplink decoupling, as signalling will be routed to a central processing unit with low-latency connections. In particular, partial centralization refers to those local deployments (e.g. indoor) where the transmission points serving the same local area are all connected to the same baseband processing central unit. Full centralization, often referred as Cloud-RAN, extends this approach to larger areas, where a large number of RF units are connected to the same baseband processing central unit. Given this already ongoing trend towards more centralized RAN architectures, which are underpinned by low-latency connectivity between BSs, the incremental cost of DUDe appears negligible in such scenarios.

2.2 Pathloss based decoupled access (DUDe 1.0)

In this section we provide a detailed simulation study to corroborate the decoupling benefits listed in the previous section. We study the gains that can be achieved by the DUDe technique in terms of UL capacity and throughput as well as the effects that this

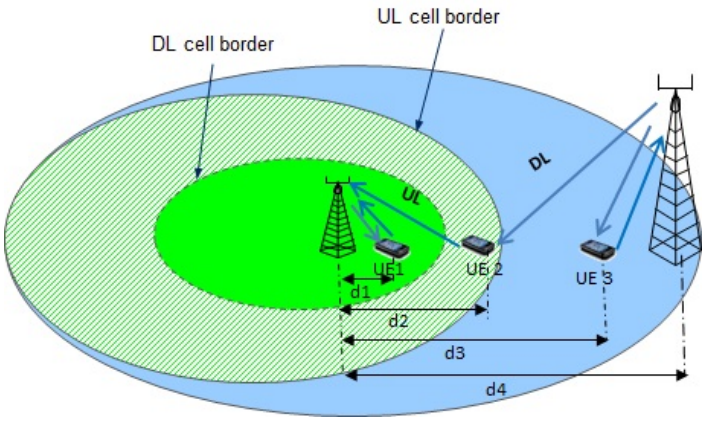


Fig. 2.1 Illustration of the decoupling concept. The large and small BS represent the Mcell and Scell respectively.

approach has on interference. We use a realistic scenario of a cellular network based on real-world planning/optimisation tools which adds a lot of value and credibility to this work.

2.2.1 Simplified example

In this study we drop the traditional UL/DL cell association based on DL received power (RP). We assume that while the downlink association is still based on DL RP, the uplink association is in fact based on pathloss. This apparently simple assumption in reality leads to radical changes in system design and architecture.

DUDe results in different cell boundaries in the UL and DL in a HetNet scenario where a UE in the region between the UL and DL cell boundaries will be connected to the Scell and Mcell in the UL and DL respectively as shown in Fig. 2.1. We will focus on the gains in the UL as this is the main motive for applying this technique. Note that DL capacities are not affected since the association remains unchanged.

We consider a two cell network model composed of a Mcell and a Scell to present the advantages of DUDe in a simplified way. The model is used to study two cases; the first case is a noise limited scenario with only one UE, to show the benefits in terms of uplink UE capacity. The second case is an interference limited scenario where there are three UEs in the network to show the benefits in reducing the interference. The two cases are explained in details below.

Case 1 (noise limited): In this case we have one UE moving from the Scell vicinity towards the Mcell and the UE UL rate is calculated for two cases; the first is the conventional case where cell selection is based on the DL received power so the UE

performs a Handover (UL & DL) from the Scell to the Mcell when passing the DL cell border (shown in Fig. 2.1) and the second case is where the UL cell selection is based on the pathloss where the UE is still connected to the Scell until passing the UL cell border which represents the DUDe technique. Neglecting, for simplicity, fading and shadowing and normalizing various quantities, the UL rate calculation is based on the below equations:

$$R = BW \log_2(1 + SNR) \tag{2.1}$$

$$SNR = \frac{P_{ue}}{\sigma^2 d^{\alpha}}, \tag{2.2}$$

where R is the rate; SNR is the signal to noise ratio, P_{ue} is the UE transmit power, σ^2 is the noise power and BW is the bandwidth which is considered to be unity for simplicity. The distance based PL is dependent on the distance d and the pathloss exponent α. We now calculate the UL rate for a UE moving from the Scell towards the Mcell for the two cell association methods, assuming, P_{ue} to be 20 dBm and the Scell and Mcell to have a pathloss exponent of 3.6 and 4 respectively. Finally, the Mcell and Scell have a transmit power of 46 and 23 dBm respectively. Fig. 2.2 illustrates the UL normalized rate for the pathloss and RP cell association cases. It shows that the pathloss case has a higher performance in the area between the DL cell border and the UL cell border since in that area the UE has a lower pathloss to the Scell, thus obtaining a higher rate when connected to the Scell.

Case 2 (interference limited): In this case, we have the same setup as the previous one but with three UEs instead of only one UE as shown in Fig. 2.1. We calculate the overall UL rate of the network using the pathloss based cell association where UE2 is connected to the Scell in the UL and then using the RP based cell association where UE2 is connected to the Mcell in the UL. UE1 is always connected to the Scell in the UL and UE3 is always connected to the Mcell in the UL.

$$R = BW \log_2(1 + SIR) \tag{2.3}$$

$$SIR = \frac{P_{ue}}{I d^{\alpha}}. \tag{2.4}$$

The UL rate is calculated based on the above equation, where SIR is the Signal to Interference Ratio (we neglect the noise for simplicity). The total normalized UL rate (R_T) is the sum of the normalized UL rate at the Mcell (R_M) and the Scell (R_S) which means the UL rate of the whole system ($R_T = R_M + R_S$). We use the same parameters as case 1 and setting d1, d2, d3, and d4 in Fig. 2.1 to 10, 25, 80, and 100 respectively.

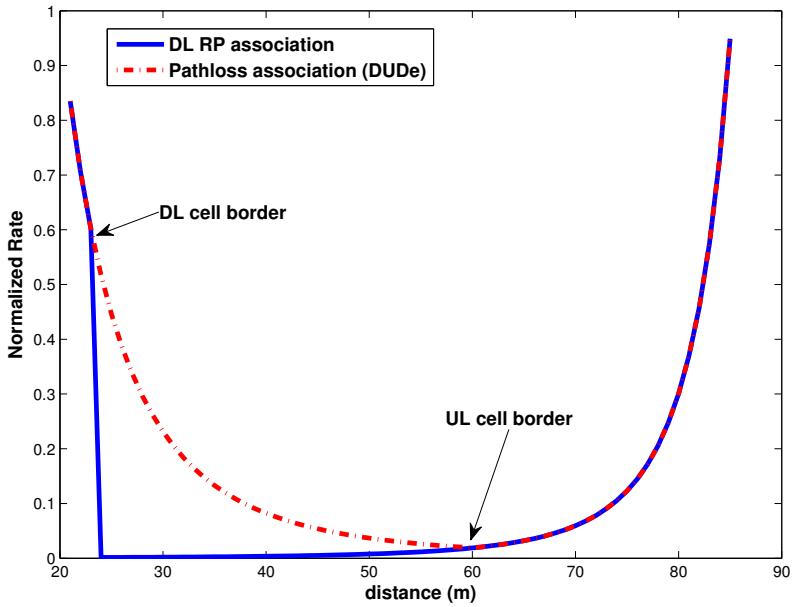


Fig. 2.2 UE rate comparison between the DL Received Power (RP) case and the Pathloss (PL) case.

So calculating RT in the pathloss case yields $R_T = 0.46 + 0.54 = 1$ and in the RP case $R_T = 0.34 + 0.33 = 0.67$. We can see that R_T is almost 50% higher in the pathloss case for the following reasons.

- UE2 in the pathloss case has a lower pathloss to the Scell which means that UE2 has a better channel to the Scell and in turn gets a better rate when connected to it.
- UE2 causes less interference to the Mcell in the pathloss case than the interference it causes to the Scell in the RP case for the same reason as above, so the interference level in the network is lower and in turn the rate is higher.

In the next section we present our realistic simulation setup which is based on an existing cellular network and we use this setup to validate our findings and illustrate the gains from the studied concept.

2.2.2 System model and simulation setup

In our simulation we use the Multi-technology radio planning tool Atoll [43] in conjunction with a high resolution 3D ray tracing pathloss prediction model [44]. The

model takes into account clutter, terrain and building data. This guarantees a realistic and accurate propagation model. Atoll has the capability of performing system level simulations where a simulation is a snapshot of the LTE network. For each simulation, it generates a user distribution using a Monte Carlo algorithm. The user distribution is based on traffic data extracted from the real network. Resource allocation in each simulation is carried out over a duration of 1 second (100 frames).

As deployment setup, we use a Vodafone LTE small cell test bed network that is up and running in a specific area in London. The test network covers an area of approximately one square kilometer. We use this existing test bed to simulate a relatively dense HetNet scenario. The considered network is shown in Fig. 2.3 where the black shapes are macro sites and the small circles are small cells which are considered to be pico cells. We consider a realistic user distribution based on traffic data from the field trial network in peak times. The distribution is up-scaled to simulate a high user density. We use an uplink power control algorithm where each cell has a predefined interference upper limit. If the UL received interference at a cell is higher than this limit the cell signals the neighboring cells to lower the UL transmit power of their UEs in order to lower the interference level at that cell.

DUDe is compared with conventional LTE operation where both UL and DL associations are coupled and based on the DL received power (RSRP) which is the conventional LTE procedure [45]. Two LTE baseline cases are considered where Scells are treated as Pico cells and Femto cells respectively. The difference between Pico and Femto cells is mainly the transmit power where Pico cells have a higher transmit power than Femto cells. These two cases are termed as Pico-Baseline and Femto-Baseline respectively. We do this to understand the gains of the pathloss approach compared to the DL RSRP approach with different Scell sizes. The transmit powers of Macro, Pico, Femto cells are 46, 30, 20 dBm, respectively. As pointed out before, all the results in the next section will focus on the UL performance. The simulation parameters are listed in Table 3.1 where we consider an LTE deployment.

In our simulations we define a UE minimum and maximum throughput demand where basically a UE has to reach the minimum throughput requirement to be able to transmit its data otherwise it is considered in outage. On the other hand, the maximum throughput demand puts a limit to the amount of throughput that each UE can get, therefore setting a high value for it helps in simulating a highly loaded network. The used scheduler tries first to satisfy the minimum throughput requirements for all the UEs and then distributes the remaining resources among the UEs to satisfy the maximum throughput demand of each UE according to the proportional fair criterion.

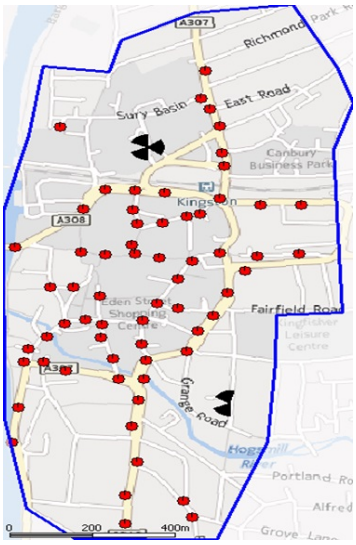


Fig. 2.3 Vodafone small cell LTE test network for an area in London.

One deployment issue is that a UE connected to different nodes in the UL and DL needs a way to send Acknowledgment, pilot and relevant control signalling to its DL node with which it has no UL established. A possible way is to route the data to the UL node and through the backhaul to the DL node and vice versa with receiving control signals from the UL node this issue will be discussed in details in the architecture part in Chapter 4. We assume an ideal backhaul where control signals are delivered with no notable delay. Non-ideal backhaul operation will be considered in Section 2.3.

2.2.3 Simulation results

In this section we show extensive simulation results covering all the aspects in Section 2.1.

Coverage

In a HetNet, the UL and DL coverages [3] are quite different as discussed previously, therefore basing the UL and DL associations on the same criterion which is the DL received power is highly suboptimal. In this part, the great difference between the UL coverage in the LTE baseline cases and DUDe is illustrated.

[3]The coverage of a base station is the geographic area where the BS can communicate where a UE that is able to receive and decode data for a BS is considered in the coverage area of this BS. In a multi BS scenario, the DL coverage area of a BS is the area around the BS (ignoring shadowing) where the received power from that BS is higher than the received power from the adjacent BSs. The UL coverage of a BS is defined similarly but considering the UL received power at the BS.

Table 2.1 Notation and simulation parameters

Parameter	**Value**
Operating frequency	2.6 GHz (co-channel deployment)
Bandwidth	20 MHz (100 frequency blocks)
Network deployment	5 Mcells and 64 Scells distributed in the test area as shown in Fig. 2.3
User distribution	560 UEs distributed according to traffic maps read from a live network
Scheduler	Proportional fair
Simulation time	50 simulation runs with 1 second each
Traffic model	Full buffer
Propagation model	3D ray-tracing model
Max. Tx. power	Macro=46 dBm High power Pico = 30 dBm Low power Pico = 20 dBm UE= 20 dBm
Antenna system	Macro: 2Tx, 2Rx, 17.8 dBi gain Pico: 2Tx, 2Rx, 4 dBi gain UE: 1Tx, 1Rx, 0 dBi gain
UEs mobility	Pedestrian (3km/h)
Supported UL modulation schemes	QPSK, 16 QAM, 64 QAM

Fig. 2.4 illustrates the UL coverage of the Mcell and Scell layers for the three cases in comparison. The UL coverage of Scells is shown to be very small in the Femto-Baseline and Pico-Baseline cases. However, in the DUDe case Scells have a much larger UL coverage which is shown to be much less dominated by Mcells than in the LTE baseline cases. This effect results in a more homogeneous distribution of UEs between the nodes which, in turn, leads to a much more efficient use of resources as will be demonstrated in the following results.

Fig. 2.4 Uplink coverage of the Femto-Baseline (left), Pico-Baseline (middle) and DUDe (right) cases where green and red represent the Mcells and Scells coverage respectively.

Association probabilities

Fig. 2.5 and 2.6 illustrate the UEs association probability as a function of the Scells density. Case 1 and 3 represent the probability of the coupled association (UL and DL based on RSRP) to Mcells and Scells respectively. Case 2 represents the probability of the decoupled association where the UL and DL are associated to the Scell and Mcell respectively which represents the DUDe case. Finally Case 4 represents the decoupled association where UL and DL are associated to Mcell and Scell which is always zero.

Considering Fig. 2.5, as the Scells density increases, Case 2 representing DUDe is dominating where the probability of association saturates at around 70% starting from 20 Scells. However, Case 4 is increasing at a much slower rate which is a result of the very limited DL coverage of Femto cells. On the contrary, Case 1 is decreasing as the number of Scells increases which makes sense as more and more UEs are more likely to connect to Scells in the UL and DL.

Now, looking at Fig. 2.6, a similar trend as Fig. 2.5 can be noticed. However, Case 2 is dominating with a probability of 50% up to a certain number of Scells (around 35) where Case 4 surpasses Case 2. It can be noticed that for Case 4 the curve is increasing with a higher slope than in Fig. 2.5 which can be explained by the fact that Pico cells have a larger coverage area, such that after a certain density of most of the devices are connected to the Picos in the UL and DL. Another point to note in these two figures is the comparison between the DL coverage of Picos and Femtos, which is shown by comparing Case 4 in Fig. 2.5 and Fig. 2.6, where we see that 10 Pcells deliver the same level of coverage as 50 Fcells. This comparison also shows that the probability of

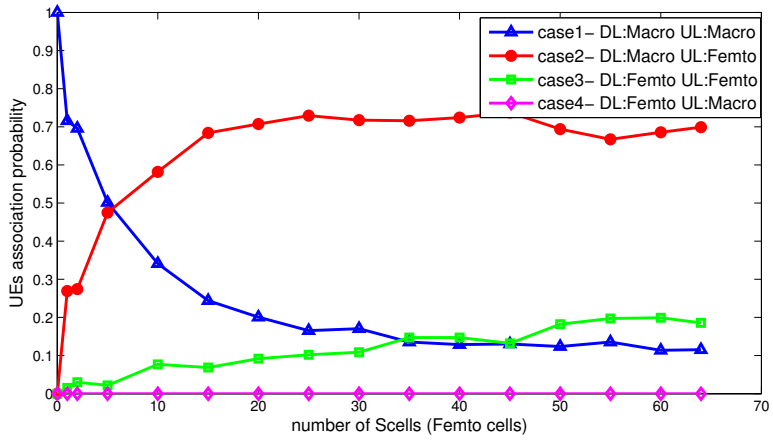


Fig. 2.5 UEs association probability.

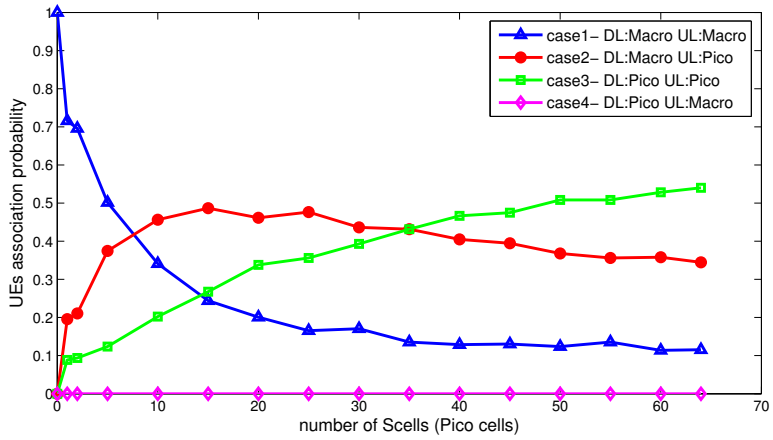


Fig. 2.6 UEs association probability.

association in UL and DL to a Pcell is more than twice that to a Fcell with 64 Scells deployed. The trends in these figures will be analytically proven in the next chapter.

Reliability and load balancing

Reliability is becoming one of the most important requirements of future cellular systems due to the proliferation of the internet of things (IoT) which is an umbrella for several applications such as sensor networks, factory automation and many more. The reliability of wireless systems can be affected by several factors including congestion and coverage loss where a device could have its connection dropped due to lack of available resources or due to very low link quality resulting from the device being out

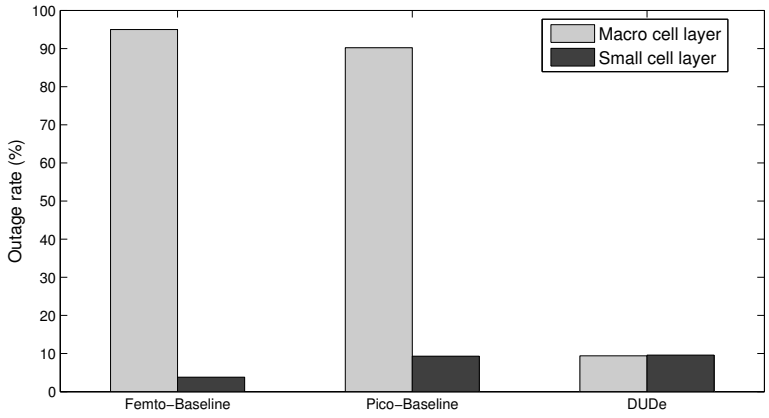


Fig. 2.7 Throughput outage.

of coverage or being subject to high level of interference. DUDe plays a crucial role in improving the UL reliability by reducing the outage rate as will be shown next.

Fig. 2.7 illustrates the average outage rate for the Mcell and Scell layers for the three cases where a high traffic scenario is simulated by setting the minimum requested throughput for each UE to 1 Mb/s. The outage is defined as the percentage of UEs that cannot achieve the 1 Mb/s minimum throughput. Since this scenario is considered to be a high traffic congestion scenario it requires a very efficient use of resources in order to satisfy the high throughput requirements of the UEs. As can be noticed from the figure the Macro layer has a very high outage rate in the LTE baseline cases which is explained by the fact that the Macro layer is very congested in the UL and Mcells cannot serve all the UEs with the required throughput which results in a high number of dropped UEs. However, in the DUDe case, UEs are distributed more evenly among the nodes so the outage rate can be drastically reduced to less than 10% on the Mcell and Scell layers.

Another trend that affects reliability is outage caused by the loss of coverage or poor channel conditions. Figure 2.8 shows the outage rate against the maximum transmit power of the UE. Outage here is defined by the fraction of UEs whose UL signal quality is lower than the one needed to access the lowest LTE modulation and coding scheme (MCS) and therefore are dropped. This result shows that for the same maximum UE transmit power DUDe can reduce the UL outage rate by upto 25% compared to baseline LTE. Alternatively, for the same outage rate the maximum transmit power can be reduced by upto 10 dB using DUDe. These results are very crucial for IoT applications where either link reliability or battery life or both are of paramount importance and it can be seen from the figure that both can be drastically improved using DUDe.

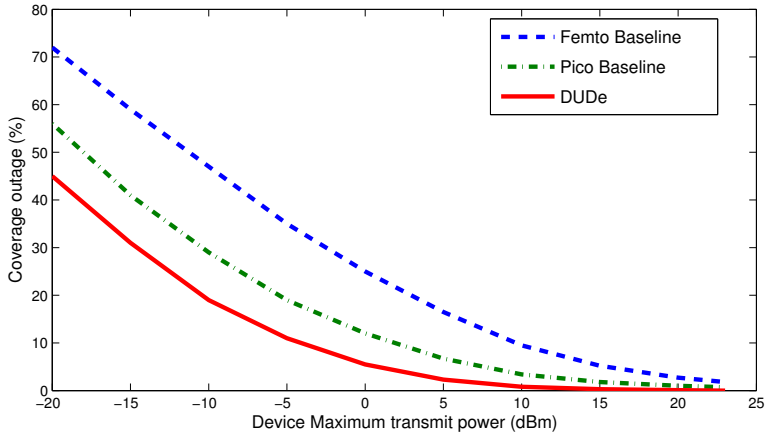


Fig. 2.8 Power outage.

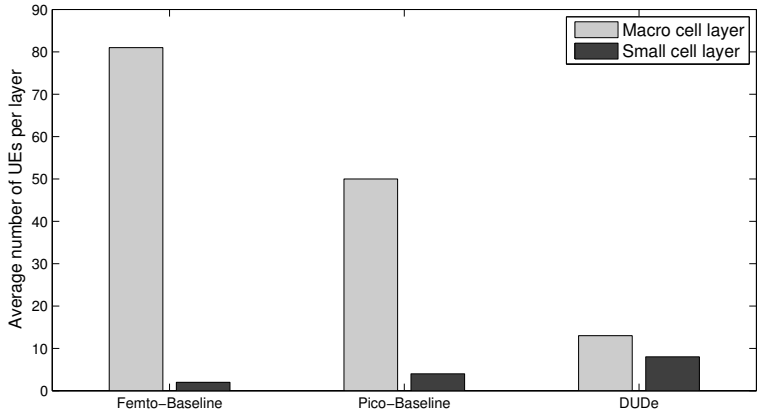


Fig. 2.9 Load balancing efect where the average number of UEs per layer (Macro and Small cell layers).

The load at a given BS in the UL maybe different than the load in the DL at the same BS. This implies that it is not optimal, in terms of load balancing, to have the same set of UEs connected to the same BS in the UL and DL.

In addition, DUDe has been shown in the previous results to improve the UL coverage of Scells which results in a much better distribution of the UEs among the different tiers of the cellular network. This load balancing effect is illustrated in Figure 2.9 where the average number of UEs per cell in each tier is shown for the three cases in comparison. DUDe results in a much more homogeneous distribution of UEs between the different node types which is directly translated into a drastically improved spectral efficiency and outage reduction as shown in the previous results.

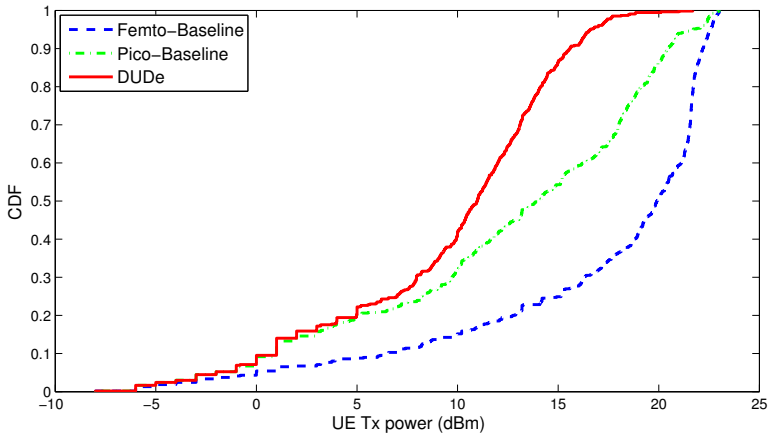


Fig. 2.10 Power CDF.

improved UL SNR

As discussed in Section 2.1.2, Dude allows a flexible association in the UL and DL where a device can connect to a closer Scell in the UL and in the DL to the Mcell from which the UE receives the highest power. The gain from this technique is twofold, for a UE transmitting with maximum power, a connection to a closer cell provides higher SNR as shown in Fig. 2.8. On the other hand for a fixed target SNR, the reduced pathloss allows to reduce the transmit power of the UE. In Fig. 2.10, we observe the decrease in transmit power via DUDe by comparing it to the other LTE baseline cases where DUDe can achieve a transmit power reduction of more than 9 dB compared to the Femto-Baseline case. This power saving feature is very important especially for IoT applications where the battery life of devices can be prolonged by more than 6 times.

Interference variation

As a result of the UL association that tends to connect the UEs in the UL to their closest node or the node to which they have the best UL received power, this can be translated into a reduction in UEs transmit power as shown in Fig. 2.8 and emphasized in [39]. This has the effect of reducing the UL interference to other base stations which is quite significant especially for UEs with low UL SINR.

In addition to reducing the average amount of interference, DUDe allows a reduction in the standard deviation of the UEs UL SINR over time as shown in Fig. 2.11. Reducing the variance of the SINR means that the channel is more stable and predictable which has a substantial impact on reducing the complexity of radio resource management (RRM) and self-organizing network (SON) functions.

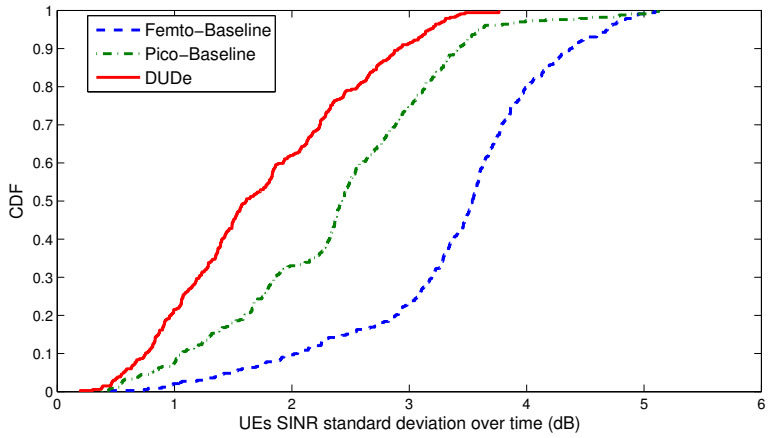


Fig. 2.11 SINR STD.

UL rate improvement

The homogeneous coverage has a large effect on the throughput and specially the cell edge UEs throughput which is represented by the 5th percentile throughput. Fig. 2.12 shows the effect of increasing the number of Scells in the simulated scenario on the 5th percentile throughput. In the DUDe case, it can be noticed that the 5th percentile throughput is improving quickly with the number of Scells. This is partly due to the increased UL coverage of Scells in DUDe as shown in Fig. 2.4. As the number of Scells increases, the 5th percentile UEs throughput starts to saturate as these UEs become limited by the channel quality and transmit power and the gains start to be more pronounced in the 50th and 90th percentiles. Looking at the Femto-Baseline and Pico-Baseline cases, it can be noticed that adding Scells has little effect on the 5th percentile throughput due to the very limited coverage of Scells in both cases which makes them more effective for the 50th and 90th percentile UEs. In addition, in these two cases the throughput is fluctuating as the number of Scells is increased. This is due to the high interference that the Scells UEs create to the Mcell cell edge UEs as these UEs are closer to the Scells so they suffer from a high level of interference. This effect is emphasised more in the Femto-Baseline case as the 5th percentile throughput starts to decrease after a certain point whereas in the DUDe case the throughput increases in a more stable way since UEs always connect to the node to which they have the best UL channel which guarantees a lower interference level.

Fig. 2.13 shows the 5th, 50th and 90th percentile throughput results for the three cases in comparison. The 5th percentile UL throughput in DUDe is increased by more than 200% and 100% compared to Femto-Baseline and Pico-Baseline respectively.

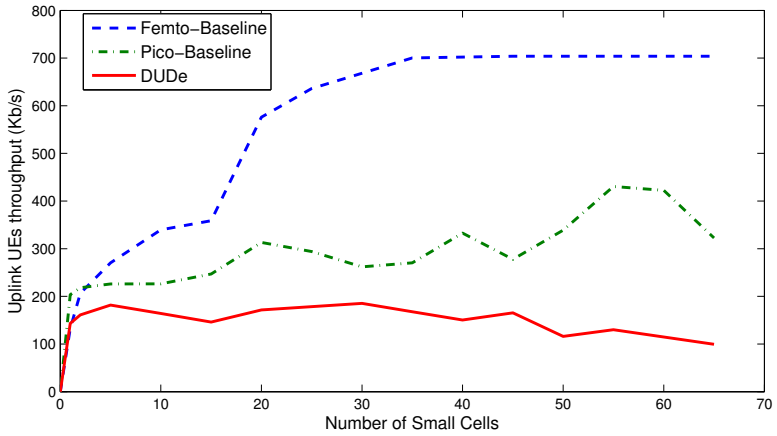


Fig. 2.12 Uplink 5th percentile throughput improvements.

Similarly, the 50th percentile throughput is improved in the DUDe case by 600% and 100% compared to Femto-Baseline and Pico-Baseline. The gains result from the improved coverage of Scells in the DUDe case which results in a better distribution of UEs between the nodes giving way to a more efficient resources utilization. In addition, the fact that UEs connect to the node to which they have the best UL channel results in an improved Signal to Interference and Noise Ratio (SINR) which allows UEs to use a higher modulation and coding scheme and in turn achieve a better utilization of resources and a higher throughput. As for the 90th percentile throughput, DUDe does not achieve the highest throughput which makes sense since DUDe aims at improving the load balancing effect which will naturally result in a reduction of the peak data rate.

Additionally, it can be seen that Pico-Baseline achieves the highest 90th percentile throughput which seems counter intuitive since it would be expected that the Femto-Baseline would be the one achieving the highest peak data rate. However, looking at the 98th percentile throughput, the throughput reaches 15 Mb/s and 10 Mb/s in the Femto-Baseline and Pico-Baseline cases, respectively. This shows that the effect of Scells in the Femto-Baseline case is limited to a small number of UEs.

In this last result, we introduce small cell biasing onto the Pico-Baseline and Femto-Baseline cases. Biasing, which is called Range Extension (RE) in LTE, is a feature that was introduced in LTE release 10 as a way to offload traffic from Mcells to Scells [46]. The idea is to add a cell selection offset to the reference signals of Small cells which allows to expand the coverage area of Scells. However, using offsets greater than 3-8 dB may lead to high interference levels in the DL which is why techniques – like enhanced Inter-Cell Interference Coordination (eICIC) – have been developed to try

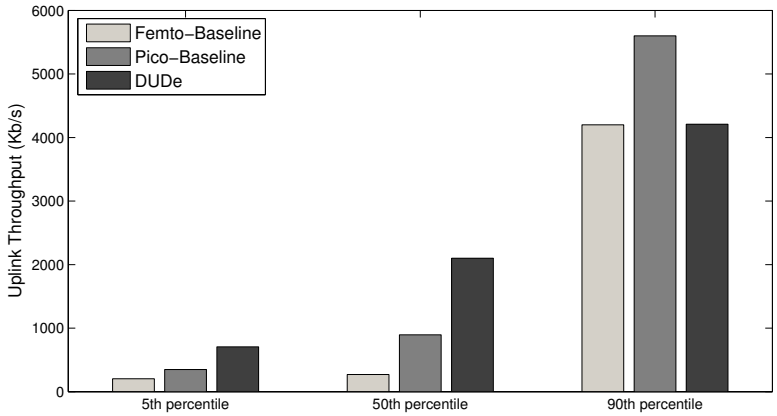


Fig. 2.13 Uplink throughput improvements.

to combat this type of interference [47]. Nevertheless, the RE technique is limited to moderate offset values due to the harsh interference in the DL. Hence, biasing allows a similar gain to DUDe but is limited by the interference in the DL. Therefore, we still observe very substantial rate gains for DUDe even when compared with biased coupled associations. Detailed breakdowns of these rate gains in various configurations are given in [33] for the simulation results and [48] for the analytical results, with our findings summarized in Table 2.2. Here, we use a slightly different setup than the previous results where 21 instead of 64 Scells are used and the UEs density is $330/km^2$ instead of $560/km^2$. Furthermore, we add a bias of 6 and 8 dBs to Pico and Femto cells respectively which have been shown in [42] to be reasonable. The gains result mainly from the improved channel quality and also from the biasing as discussed before, which gives cell-edge (5th percentile) and median (50th percentile) UEs access to more resources which results in a higher UL rate. It is quite encouraging that two very different models and approaches to evaluating the rate gains both result in the conclusion that gains in the range of 100-200% are within reach, although the gains do erode somewhat with biasing since the baseline improves. Finally, we note that a recent paper based on optimization theory with a yet different model also finds significant gains from DUDe [49].

Table 2.2 Summary of predicted uplink rate gains averaged over all UEs in the network, as a result of DUDe. Picocells have transmit power of 30 dBm while femtocells 20 dBm. We note that DUDe outperforms the baseline also when downlink biasing is used.

	Pico (RE=0)		Pico (RE=6dB)		Femto (RE=0)		Femto (RE=8dB)	
	Anal.	Sim.	Anal.	Sim.	Anal.	Sim.	Anal.	Sim.
5^{th} percent.	115%	90%	50%	30%	270%	260%	140%	95%
50^{th} percent.	95%	150%	30%	60%	260%	230%	120%	180%

2.3 Load and backhaul aware decoupled access (DUDe 2.0)

We have so far assumed that the decoupled cell association strategy is based on the link quality in each direction. That is, the decision to handover the DL is (and has been) based solely on the DL received power; whereas the decision to handover the UL is based solely on the UL pathloss. The system assumptions were to some extent ideal in that neither the cell load nor the backhauling capabilities have been taken into account – both of which have an impact onto the actual performance gains under more realistic operating conditions. This shortcoming is addressed in this section, where we proceed to outline prior related art as well as a summary of our technical contributions.

In [50] and [51], backhaul aware cell association was considered but only from a DL perspective. In [52] and [53], load aware cell association was studied in the DL as well. In [54], the authors study UL cell association in a game theoretic approach to optimize the packet success rate of the UEs.

In this section, we argue that UL pathloss alone is not sufficient to efficiently apply DUDe. Notably, the association algorithm ought also to consider the overall load of the cell(s). Furthermore, since DUDe requires significant backhauling support, we also condition association with backhauling capacity. Therefore, instead of taking the decision based only on link quality, the system now considers the link quality, the cell load and the cell backhaul capacity. We use a similar simulation scenario as the previous section. We give special attention to UL power control where we show that the performance depends greatly on the power control settings. We use a flow level traffic model that is more realistic than the full buffer model considered in prior art.

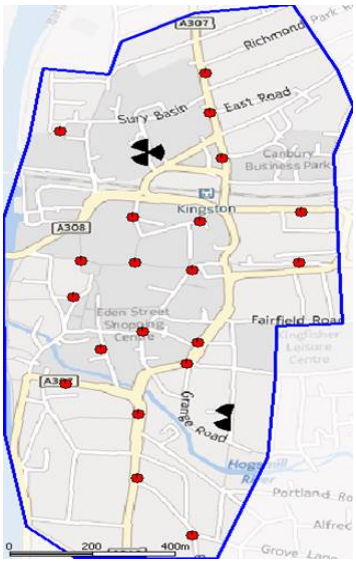


Fig. 2.14 Vodafone small cell LTE test network in London (with fewer small cells).

2.3.1 System model

We consider the UL of a HetNet where, as deployment setup, we use the same deployment as the one used in Section 2.2 with fewer small cells as shown in Fig. 2.14 where in total we have B cells.

We consider a realistic user distribution based on traffic data from the field trial network in peak times where the total number of users is N_u.

Network traffic is modelled on a flow level where flows represent individual file or data transfers e.g. video, audio or generic file uploads. This model reflects a much more realistic traffic model than the full buffer model considered in [33]. We assume that a flow of size ρ (Mbits) arrives to a UE's queue after a certain period "wait time" T_W. T_W and ρ are exponentially distributed with certain mean values. UEs experience a different T_W each time a flow transmission is finished. The radio link quality is determined by many factors including pathloss, fading, interference, and transmit power of the UEs. The UL SINR of UE i connected to BS j is given by

$$SINR_{ij} = \frac{h_{ij}P_i}{\sigma^2 + I}, \tag{2.5}$$

where P_i is the i^{th} UE transmit power, h_{ij} incorporates pathloss, shadowing and fast fading between UE i and BS j, σ^2 is the noise power and I is the UL intercell interference power. We characterize the achievable data rate using the Shannon formula as follows

$$C_{ij}^{Access} = BW \log_2(1 + SINR_{ij}), \tag{2.6}$$

where BW is the system bandwidth.

Uplink power control for the UEs follows the 3GPP specifications [55], where we consider open loop power control which is given by

$$P_{UE} = \min\{P_{MAX}, 10\log_{10}(M) + P_0 + \alpha L\}, \tag{2.7}$$

where P_{MAX} is the maximum permitted transmit power of the UE, M is the number of physical resource blocks (PRBs) assigned to the UE, P_0 is a normalized power, α is the pathloss compensation factor and L is the pathloss towards the serving cell.

However, the power control algorithm does not account for inter-cell interference which, as we will show in the results, affects greatly the UL performance. The effect is more pronounced when load balancing takes place since UEs connect to a suboptimal cell so they are more vulnerable to interference. Therefore we will use an interference aware power control algorithm which sets a limit to the transmit power of the UEs depending on the interference level that the UE causes to the closest neighboring cell. Similar algorithms have been proposed in the literature such as [56]. The algorithm is as follows

$$P_{UE} = \min\{P_{MAX}, 10\log_1 0(M) + P_0 + \alpha L, I_0 + L_s + 10\log_{10}(M)\}, \tag{2.8}$$

where I_0 represents the UL interference power spectral density (PSD) target for the UE and L_s is the pathloss towards the most interfered cell by the UE. This allows us to control the interference level in the system by changing I_0.

In a real world deployment, the Scell backhaul is always an issue since outdoor Scells are usually mounted on street furniture where there is no guaranteed wired connection or line-of-sight to the Mcell. Furthermore with the increasing bit rates provided by access technologies the bottleneck is moving slowly from the access network to the backhaul. We consider that all cells in the test network have a limited backhaul capacity C_j^{bk} where, naturally, Scells would have tighter backhaul constraints than Mcells.

2.3.2 Cell association algorithm

In our previous study [33] we have considered the UL cell association to be based on pathloss (PL) which showed very high performance improvements that were mainly due to the load balancing effect and the improved link quality of the UEs. We extend this approach to include the cell load and backhaul capacity in the decision criterion;

consequently instead of taking the decision based only on link quality the UE considers the link quality, the cell load and the cell backhaul capacity. This approach makes sense since in real networks users are distributed in a non-uniform way where a UE that is close to a congested cell might be better off connecting to a cell that is further but less congested.

We consider a cell association criterion that was considered in [52] in the DL. We extend this by applying it to the UL and including the backhaul capacity so that the optimal BS chosen by UE i is given by

$$s(i) = arg\ max_{j \in B}(1 - \eta_j)C_{ij}^{Max}, \tag{2.9}$$

where $C_{ij}^{Max} = \min(C_{ij}^{Access}, C_j^{bk})$, η_j is the j^{th} BS load which is reflected in [52] as being the average resource utilization per cell. We found that this approach for η_j works fine in the DL whereas in the UL the situation is different since the UEs are power limited which means that a UE with bad channel conditions would not be able to transmit on a large number of resource blocks. This would result in a low utilization of the resources of the cell even though this cell could be serving many UEs. Therefore the cell utilization is a poor metric to characterize the cell load in the UL and we resort to a different way of estimating the load. Notably, since the flow arrival is exponentially distributed and assuming the system to be stationary. The average number of flows is then given by

$$E[N_j] = \frac{\eta_j}{1 - \eta_j}. \tag{2.10}$$

This yields

$$\eta_j = \frac{E[N_j]}{1 + E[N_j]}. \tag{2.11}$$

Inserting η_j in (2.9) yields

$$s(i) = arg\ max_{j \in B} \frac{C_{ij}^{Max}}{E[N_j] + 1}. \tag{2.12}$$

The cell association criterion in (2.12) will be used for the rest of the paper. We consider a fully distributed algorithm where the main idea is that a UE does not need to stay connected to one BS in the UL all the time. Instead, a UE can keep its anchor DL cell and every time the UE has data (flow) to transmit in the UL, the UE connects to the cell with the highest criterion according to (2.12).

The algorithm thus functions as follows: The BSs broadcast their load $E[N_j]$ and backhaul capacity C_j^{bk}. All UEs in the system start with an exponentially distributed

wait time (T_W) after which a UE has a flow of size ρ to transmit. The UE uses the criterion in (2.12) to find the best cell to connect to and after finishing its transmission the UE disconnects from the cell and goes idle for another random period T_W; thereupon the operation is repeated. The steps are detailed in Algorithm 1.

Algorithm 1 Load/backhaul aware UL cell association

BSs broadcast $E[N_j]$ and C_{bk} periodically.
UEs (1,...,N_u) are idle for a random $T_W(1, ..., N_u)$.
for Number of subframes **do**
    for each idle UE **do**
        if $T_W = 0$
        $UE_{queue} = \rho$
        UE connects to BS (i) according to (2.12)
        UE is scheduled in BS (i) until $UE_{queue} = 0$.
        UE goes idle for a random T_W.
        else
        $T_W = T_W - 1$
        end if
    end for
end for

2.3.3 Simulation setup

In our simulations we use the deployment setup in Fig. 2.14 which consists of 5 Mcells and 21 outdoor Scells. The propagation model is based on a high resolution 3D ray tracing pathloss prediction model. The model takes into account clutter, terrain and building data. The user distribution is based on traffic data extracted from the real network.

We consider three power control settings:

- Loose power control with full pathloss compensation. We use (2.7) where (α, P_0) are set to (1, -80 dBm). This is referred to as **Setting 1**.
- Conservative power control with partial pathloss compensation. We use (2.7) where (α, P_0) are set to (0.6, -70 dBm). This is referred to as **Setting 2**.
- Interference aware power control where we use (2.8) where (α, P_0) are set to (1, -80 dBm) and I_0 to -100 dBm.

We compare three UL cell association cases:

- Cell association based on the DL Reference Signal Received Power (RSRP)[4] which is the conventional LTE procedure [55]. This case is termed **DL-RSRP**.
- Cell association based on the pathloss which represents the DUDe algorithm as considered in [33] and is termed as **DUDe**.
- Cell association based on Algorithm 1 which considers the cell load and backhaul capacity on top of the conventional DUDe. This case is termed **DUDe-Load**.

As pointed out before, all the results in the next section will focus on the UL performance. The simulation parameters are listed in Table 2.3 where we consider an outdoor LTE deployment.

2.3.4 Simulation results

Initially we assume having an ideal backhaul (i.e. no limit on the backhaul capacity) on all the cells in order to study the load balancing effect. We start by comparing the throughput results with different power control settings according to Setting 1 and Setting 2.

The throughput results for the three cases in comparison are shown in Fig. 2.15. Comparing DUDe to DL-RSRP, we see similar gains as in [33] where the 5th and 50th percentiles are increased by more than 100% and 150% respectively for both power settings. The gains are due to the load balancing effect of DUDe and the better link quality as UEs connect to the cells to which they have the lowest pathloss. The 90th percentile throughput is less in DUDe than DL-RSRP as in the latter case only a few UEs are served by the Scells; therefore these UEs achieve a high throughput.

We notice also that DUDe-Load is more affected, in terms of 5th and 50th percentiles, by the change in the power settings than DUDe. This is due mainly to the fact that UEs connect to a suboptimal cell in terms of pathloss due to the load balancing effect which makes these UEs more vulnerable to interference and more affected by the other UEs transmit power.

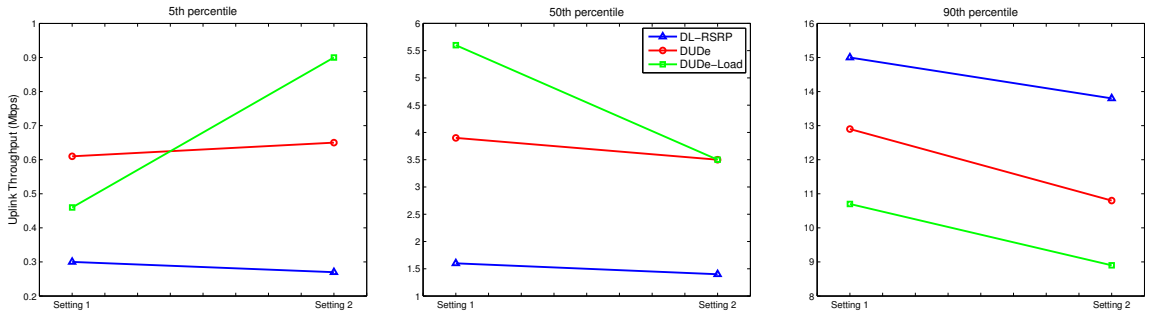
We then compare DUDe and DUDe-Load starting with Setting 1 where we see that the 5th percentile throughput is reduced by about 20% in the DUDe-Load case while the 50th percentile is increased by 40% compared to DUDe. The loose power control causes the interference level to increase which has a negative effect on the cell edge UEs as explained below.

[4]RSRP is the measured received power at the UE of the reference signals transmitted by the BS.

Table 2.3 Notation and simulation parameters

Parameter	**Value**
Operating frequency	2.6 GHz (co-channel deployment)
Bandwidth	20 MHz (100 frequency blocks)
Network deployment	5 Mcells and 21 Scells distributed in the test area as shown in Fig. 2.14
User distribution	330 UEs distributed according to traffic measurements read from a live network
Scheduler	Proportional fair
Simulation time	10 seconds (10,000 subframes)
Traffic model	Flow level traffic Mean flow size = 1 Mbit Mean wait time = 100 ms
Propagation model	3D ray-tracing model
Max. Tx. power	Mcell = 46 dBm Scell = 30 dBm UE = 20 dBm
Antenna system	Mcell: 2Tx, 2Rx, 17.8 dBi gain Scell: 2Tx, 2Rx, 4 dBi gain UE: 1Tx, 1Rx, 0 dBi gain
Supported UL modulation schemes	QPSK, 16 QAM, 64 QAM

This result shows that cell edge UEs (5th percentile) are better connected to a loaded cell to which they have the better link quality than connecting to an unloaded cell with a worse channel. On the other hand the 50th percentile UEs can afford a reduced channel quality and with the higher power headroom they actually achieve a high gain by using the extra resources provided by the load balancing effect of DUDe-Load. Finally, the figure also shows a loss of about 20% in the 90th percentile throughput which is logical since load balancing is always a trade-off between peak and (cell-edge/average) throughput.



(a) (b) (c)

Fig. 2.15 Throughput percentiles for the three cases with power control Setting 1 and 2.

Then we compare DUDe and DUDe-Load for Setting 2 where the 5th percentile throughput in DUDe-Load is improved by about 40% over DUDe whereas the 50th percentile throughput is almost the same. This result shows how power control affects the network performance greatly. The used power control scheme sets a lower limit on the transmit power of the UEs than the one used in Setting 1; this causes the UL interference level in the network to be lower than the previous case which, in turn, allows the cell edge UEs to achieve a higher throughput when connected to a suboptimal cell in terms of pathloss.

On the other hand, the 50th percentile UEs do not achieve a higher throughput with the load balancing effect due to the lower bound on the UEs transmit power. These UEs hence might not be able to use all the resources available to them; therefore, these UEs achieve a relatively low gain from the higher resource availability whereas the lower link quality to the suboptimal cell reduces the throughput. Consequently, both effects almost even out and there is no gain in terms of 50th percentile throughput.

The main message in Fig. 2.15 is that cell edge UEs are mostly interference limited whereas 50th percentile UEs are power limited so having power control Setting 1 would benefit the 50th percentile UEs but would be harmful for cell edge UEs while power control Setting 2 has the opposite effect.

Fig. 2.16 shows a CDF of the standard deviation of the UEs UL SINR over time for Setting 1 where interference is quite high. DUDe and DUDe-load show a reduction of interference variance of about 1 dB at 50% CDF compared to DL-RSRP. The lower variance reflects a more stable interference scenario in DUDe where the lower variance of DUDe-Load results from the improved load balancing effect which improves the resource utilization and, in turn, helps in stabilizing the interference. DUDe-load shows

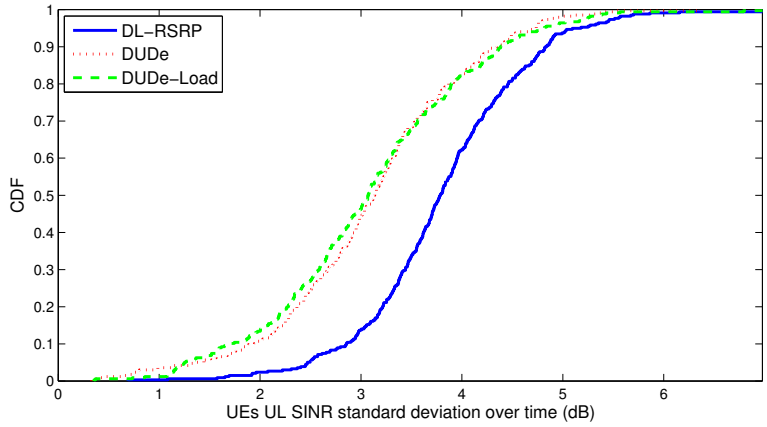


Fig. 2.16 Throughput outage.

a slightly lower variance than DUDe until 70% which is a result of the more emphasized load balancing in DUDe-load. This is a very important feature since UL interference is known to be very volatile and dynamic and this result shows that radio resource management (RRM) and self-organizing network (SON) operation in general can be facilitated using DUDe.

In Fig. 2.17, we show throughput results for the interference aware power control in (2.8). The aim here is to try to find a trade-off between 5th and 50th percentile performance. We see, indeed, that using this power control setup we achieve a similar or even higher 5th percentile throughput as in Setting 2 in Fig. 2.17 where DUDe-Load outperforms DUDe by 15%. Also, in the 50th percentile the performance is similar to Setting 1 in Fig. 2.15 where DUDe-Load outperforms DUDe by 20%. The better performance of DUDe-Load in the 5th and 50th percentile throughputs results from the fact that the interference aware power control affects more the UEs that cause higher interference, mostly cell edge UEs, to neighboring cells while allowing the other UEs, 50th and 90th percentile UEs, to transmit with a higher power. This results in a lower interference scenario which benefits the cell edge UEs that are interference limited and also allows the higher achieving UEs to transmit with a higher power and, in turn, exploit the extra resources resulting from load balancing.

In the results in Fig. 2.18 we study the throughput behaviour in the 3 cases while changing the backhaul capacity of Scells from 1 to 100 Mbps. The Mcells backhaul capacity is assumed to be 100 Mbps in all cases. We present the results for the interference aware power control setup used in Fig. 2.17.

In the 5th percentile result the DUDe-load case shows the highest throughput since the UEs know of the backhaul and load capabilities of the cells. The DL-RSRP case

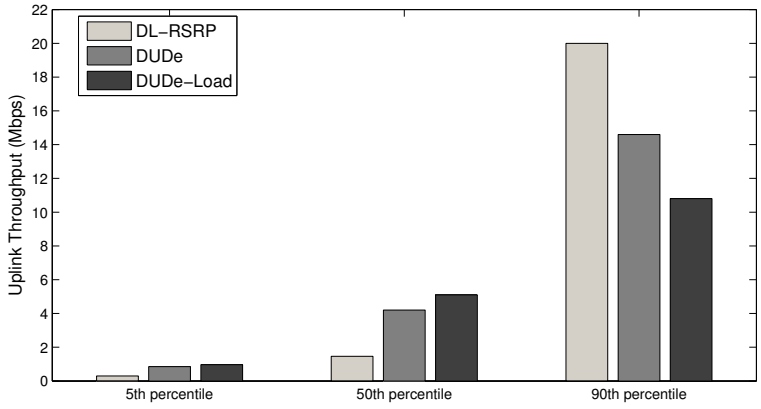


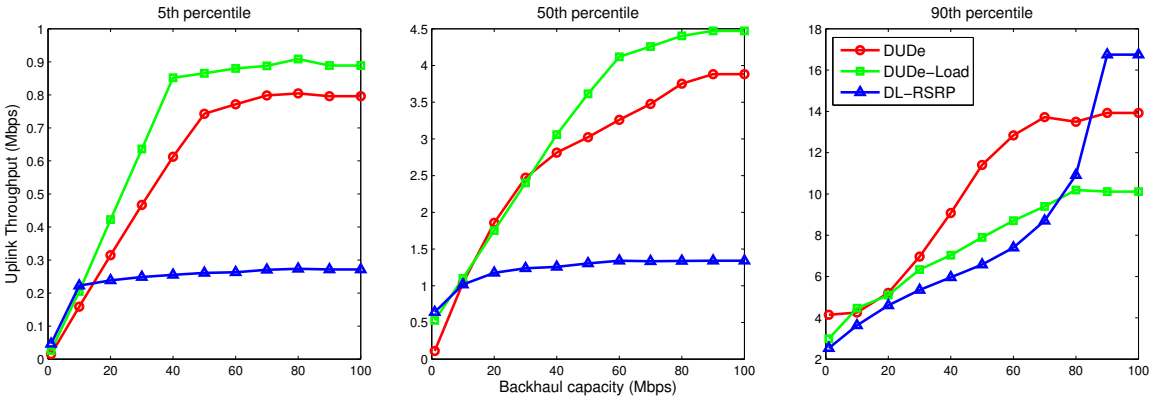
Fig. 2.17 Throughput outage.

performs better than the DUDe case up to a backhaul capacity of 10 Mbps after which DL-RSRP saturates and DUDe keeps on increasing. Similarly, in the 50th percentile the DL-RSRP case is performing almost the same as DUDe-Load for very low Scell backhaul capacities since in the former case the UEs are mostly connected to the Mcells but as the Scell backhaul capacity increases DL-RSRP starts saturating and DUDe-Load surpasses it. Also the DUDe-load case is outperforming DUDe for the different capacities where the gain increases as the backhaul capacity of Scells increases as with the increase of Scell capacity DUDe-Load can have more options to assign UEs to Scells in a more efficient way.

Finally for the 90th percentile throughput, DUDe outperforms both DUDe-load and DL-RSRP since it has the lowest number of UEs connected to the Mcells. These UEs can get very high throughputs, up to a certain point where DL-RSRP surpasses DUDe. The reason is that Scells in DL-RSRP serve fewer UEs than the other 2 cases. Therefore after a certain backhaul capacity Scells can provide very high data rates to these UEs.

Looking at the DUDe-load case, with lower Scell backhaul capacities the UEs are pushed more towards the Mcells but still DUDe-load has less UEs connected to Mcells than DL-RSRP which explains why DUDe-load outperforms DL-RSRP at the beginning but as the Scells backhaul capacity increases the load balancing role is stronger which stops the 90th percentile throughput of DUDe-load from increasing as explained before.

Finally, in order to have some insight on the load balancing effect of DUDe and DUDe-load we compare the variance of the number of UEs per cell in the three cases. This measure gives an indication of how UEs are distributed among the cells. A high



(a) (b) (c)

Fig. 2.18 Throughput percentiles against backhaul capacity.

variance indicates low load balancing effect and vice-versa. The variance of the number of UEs is 470, 83 and 21 for DL-RSRP, DUDe and DUDe-load, respectively. The DUDe case shows a clear improvement of load balancing over DL-RSRP which is shown by a dramatically reduced variance which, in turn, shows that the variation in the number of UEs/cell is small. The DUDe-load case shows an even lower variance (i.e. better load balancing) than DUDe as it is not only restricted on balancing the UEs between Mcells and Scells but it also improves the load balancing among Scells which is a very important feature in future ultra-dense Scell networks.

2.4 Summary

The decoupling of the downlink and uplink, referred to as DUDe, is an emerging paradigm shown to improve capacity significantly for cell edge users. The underlying principles of DUDe relate to a proper and independent association of the uplink and downlink. In this chapter we have presented an extensive simulation study of DUDe using highly realistic network deployment, propagation model and user distribution. The chapter started with a discussion on the main reasons for decoupling the UL and DL and the benefits of using this 5G architecture design. The benefits included the improved UL SNR, interference conditions, UL data rate and better load balancing. In Section 2.2 we compared the simplest form of DUDe, where UL and DL associations are based on pathloss and RSRP respectively, with LTE baseline cases. The gains are very high in a dense HetNet deployment where this technique can achieve between 100% and 200% improvement in the 5th percentile UL throughput and even more than that in the 50th percentile throughput. Also, we have shown that the outage rate is

decreased dramatically in networks with high minimum throughput requirements where the outage rate is decreased from 90% to below 10% on the Macro layer. Subsequently, in Section 2.3 we introduced a more complicated version of DUDe where the cell load and backhaul capacity is considered in addition for an interference aware power control algorithm. Our findings confirm that the enhanced DUDe achieves a reduced UL SINR variance over baseline LTE which facilitates RRM and SON operations. Results for our load-aware DUDe show that the system throughput improves even further compared to the prior introduced baseline DUDe approach. The performance improvement depends very much on the power control mechanism used. We have shown performance results for different power control settings where throughput gains of the load aware DUDe over baseline DUDe are 15% and 20% in the 5th and 50th percentile throughput respectively.

After establishing the superior performance of DUDe from a simulation perspective, in the next chapter we delve into proving these gains analytically using tools from stochastic geometry to derive tractable expressions for the SINR and rate in a DUDe enabled cellular network.

Chapter 3

Theoretical analysis of DUDe

In the previous chapter, we presented an extensive simulation study of DUDe showing the main benefits and gains behind this technique. In this chapter[1] we provide a rigorous analytical study of DUDe to confirm the previously discussed gains and advantages in a tractable way which we believe adds a lot of value to our work in which we are trying to narrow down the gap between simulation and analysis showing that similar results and trends can be attained using such different models and assumptions as was shown for the data rate in chapter 2 and will be later shown for the association probabilities later in this chapter. We start by introducing the basics of modelling cellular networks using stochastic geometry. Thereafter, we derive the association probabilities for a sub-6GHz HetNet in Section 3.2 followed by an extensive study of biased cell association in a mixed millimeter wave and sub-6GHz HetNet and the resulting SINR and rate coverages in Section 3.3.

3.1 Overview of Stochastic geometry

The study and design of cellular networks often tended towards two extremes. For analysis and academic research, very simplistic models are usually implemented in order to maintain the tractability of the system. The conventional way to model cellular networks was by placing the base stations on a 2-D regular hexagonal lattice or simply a square lattice. Tractable analysis can be done for a fixed UE with a small number of interfering base stations with the UE placed, for example, at the cell corner to simulate a worst case scenario and finding the SINR as in [59] [60]. The resulting SINR is a random variable in the case of shadowing and fading. Such an approach

[1] The work in this chapter is partly included in [57, 58].

results in a very pessimistic view of the network that doesn't provide much guidance on the actual performance of most UEs in the cellular network.

On the other hand, for industrial research more complex system level simulations with a very large number of parameters and that are based on Monte Carlo simulations are used where random drops of UEs are averaged in order to get some meaningful result that reflects the performance of UEs as done in the landmark capacity paper [61]. This approach does not result in tractable expressions for the SINR so more general results that provide guidance into typical SINR or the probability of outage over the entire cell or network must rely on complex and time consuming simulations. in addition to being cumbersome to build and run, simulation tools suffer from the lack of repeatability and transparency.

In addition, although widely acceptable, grid-based models are highly idealized and are increasingly inaccurate and less representative for heterogeneous networks where we are moving away from carefully planned homogeneous networks, where we can get away with the grid model, to a more random and unplanned deployments of small cells which cannot be represented by a grid but rather a random deployment of base stations. In heterogeneous networks cell radii vary considerably due to differences in transmit power, antenna patterns, tower hight and user density.

Stochastic geometry is a tool that allows to solve this lack of a tractable way to model cellular networks by introducing an extra source of randomness which is the position of the base stations. Stochastic geometry not only accounts for and captures the randomness in the topology of cellular networks but also leads to tractable analytical expressions. It has been applied to ad hoc networks for more than three decades [62] to model systems with random channel access.

In stochastic geometry analysis, the network is abstracted to a certain point process (PP) which captures the network characteristics. That is, according to the network type and MAC layer properties a suitable PP is selected to model this particular network. The most popular PPs are: Poisson point process (PPP), Binomial point process, Hard core point process, Poisson cluster process. A formal definition of each of these PPs can be found in [63–65].

Our focus will be on the poisson point process which due to its independence properties, is the most tractable and most important PP from the above. In this chapter we model the locations of base stations of each tier of a heterogeneous network using independent homogeneous PPP with a certain density for each tier. This approach for modelling BSs locations has been used in early works such as [66–68]. However, the key metrics for coverage such as SINR distribution have been identified in the seminal

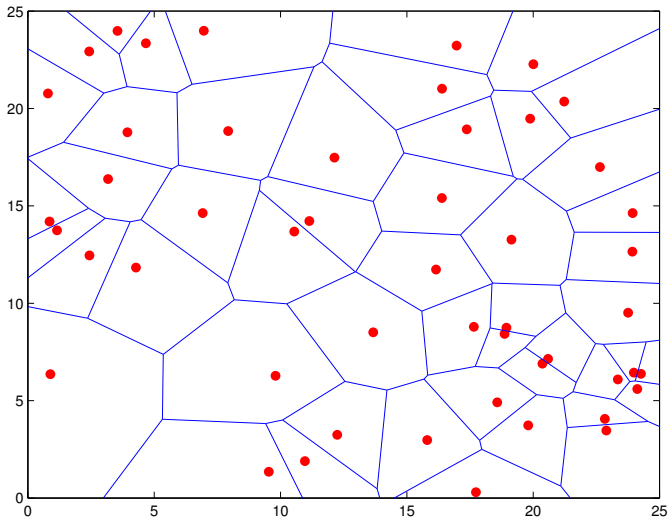


Fig. 3.1 Poisson distributed base stations where cell boundaries form a Voronoi tessellation.

work in [69] which will be discussed shortly. The main advantage of this approach is that the base stations locations are independent which allows the use of many tools from stochastic geometry. Assuming UEs are randomly distributed in a 2D plane and that UEs connect to their closest cell we would end up with the deployment in Fig. 3.1 where the dots represent the BSs and the lines represent the cell boundaries which simply form a Voronoi tessellation. That is, a line bisecting the distance between two adjacent BSs will act as the coverage border between the two.

The authors in [69] modelled both UEs and BSs as independent homogeneous PPPs where the analysis was done for the user at the origin. Note that, according to Slivnyak's theorem, the statistics seen from a PPP are independent of the test location [63–65]. Hence, no generality is lost in studying the statistics seen by a UE at the origin. The main findings in [69] are: (a) the PPP provides a tight bound for the performance of actual networks, (b) simple expressions for the rate and coverage probability are possible to derive, (c) in interference networks (i.e. networks where noise is negligible as compared to interference and therefore neglected) the SIR statistics are independent of the density of BSs. This latter insight is quite interesting as it means that coverage probabilities in a cellular network are independent of the density of BSs, so increasing the density of BSs neither improves nor degrades the coverage probability in a cell. This can be explained as follows: as the intensity of BSs increases, the distance to the serving cell decreases which results in increasing the received signal power at the UE.

However, the increase of density of BSs means also that the interfering BSs density is increased and that they are closer to the UE. In essence, these two effects cancel out leaving the coverage probability independent from the density of the deployment. Although these results are valid only for PPP network models, they give useful insights into the worst case network performance. More specifically, deploying more BSs, in the worst case, does not lead to deteriorating the SINR statistics. Another insight in [69] was that they have shown that the random deployment of a PPP is very similar to a realistic operator deployed network.

In the sequel, we extend the work in [69] to study the effect of DUDe on the network performance in sub-6GHz networks as well as mmWave networks.

3.2 Analysis of DUDe in sub 6 GHz heterogeneous networks

3.2.1 Introduction

HetNets have been recognized as a, if not the, most promising approach to yield required communications rates in 4G and emerging 5G systems [70, 71]. Heterogeneity here typically spans across Macro cells (Mcells) and different types of Small cells (Scells), such as Micro cells (Micell), Pico cells (Pcells) and Femto cells (Fcells); in the future, the inclusion of 3GPP-alien systems, such as WiFi, shall also be considered in this setting. However, the current 3GPP communications architecture and protocols were designed with Mcells in mind and heterogeneity was just an afterthought. Having heterogeneity in place, it is important to present a fresh look on how the networks should be deployed and operated, as well as which fundamental improvements should be made in order to have efficient operation of the heterogeneous deployment model as a part of 5G wireless.

Notably, the increasing heterogeneity is dramatically changing our traditional notion of a communication cell [9]: as the number of Base Stations (BSs) becomes comparable to the number of devices and the deployment pattern of the BSs is rather irregular, there are multiple BSs from which a device can select one to associate with. Given that the uplink traffic is gaining in importance, transmission powers and interference levels differ significantly between available systems, and the link quality experience is very different between downlink (DL) and uplink (UL). The current policy for association to the BS is *DL Reference signal received Power (DRP)* and works such that the device is associated to a BS both for DL and UL transmissions, provided that the selected

BS offers the highest downlink received power. One of the main contribution of this chapter is to challenge the current policy used by a device to associate to a BS using tools from stochastic geometry.

The main contribution of this section is to derive, based on stochastic geometry and prior derived association probabilities in [72], the achievable rates of a decoupled system. The derived expressions are easily evaluated numerically. Furthermore we compare the analytically derived association probabilities with simulation results.

It was very interesting to note that the analysis with stochastic geometry and real-world experimental data show the same trend in the association probabilities. This led us to verify the trend by performing additional simulations over a third deployment model, in which the BSs are placed in a regular grid. The trend was confirmed with the third model as well, thereby leading to a conclusion that the association probability depends chiefly on the deployment density and less on the actual deployment process.

3.2.2 System model

The system model represents heterogeneous cellular network, consisting of two tiers, Mcell tier and Scell tier. A co-channel deployment is assumed, i.e. the BSs of the two tiers re-use the same frequency band. The locations of BSs and the locations of devices are modelled by independent homogeneous Poisson Point Processes (PPPs). Each PPP is denoted as Φ_v and has intensity measure λ_v, where $v = M$ for Mcells, $v = S$ for Scells and $v = d$ for devices. A point in $\mathbb{R}^2$ that results of a realization of Φ_v is denoted as $x_v = (x_{v_1}, x_{v_2})$. The transmit power of Mcells, Scells and devices is P_M, P_S and P_d, respectively. Without loss of generality, the analysis is performed on a typical device located at the origin, i.e. $x_d = (0, 0)$. By Slivnyak's theorem, a PPP conditioned on a presence of a typical point in the origin has the same distribution as the original PPP [73].

We analyze the association probability, where we consider both directions, DL and UL. Subsequently, we analyse the spectral efficiency, where we focus on the UL only. The signal power received from BS located at $x_v \in \Phi_v$ is denoted as S_v^{DL} and the signal power received at the same BS in UL is denoted as S_v^{UL}. The signals are given by

$$S_v^{DL} = P_v h_{x_v} \chi_v \|x_v\|^{-\alpha} \tag{3.1}$$

$$S_v^{UL} = P_d h_{x_v} \chi_v \|x_v\|^{-\alpha}, \tag{3.2}$$

where h_{x_v} describes Rayleigh fading and is an exponentially distributed random variable with unit mean. $\|x_v\|$ is the distance from x_v to the origin and α is the path loss exponent

($\alpha > 2$). χ_v is lognormal shadowing defined as $\chi_v = 10^{\frac{X_v}{10}}$, where $X_v \sim N(\mu_v, \sigma_v^2)$. We are using the approach elaborated in Lemma 1 in [74], where the authors include the shadowing in a transparent way by using the displacement theorem [75]. The received signals given by (3.1) and (3.2) can be transformed in terms of the displaced points y_v as follows,

$$S_v^{DL} = P_v h_{x_v} \left\| \chi_v^{-1/\alpha} x_v \right\|^{-\alpha} = P_v h_{x_v} \left\| y_v \right\|^{-\alpha} \tag{3.3}$$

$$S_v^{UL} = P_d h_{x_v} \left\| \chi_v^{-1/\alpha} x_v \right\|^{-\alpha} = P_d h_{x_v} \left\| y_v \right\|^{-\alpha}. \tag{3.4}$$

By displacement theorem, the points y_v are obtained by independent realization of equivalent PPP $\widetilde{\Phi}_v$ with intensity $\widetilde{\lambda}_v$, which is related to the intensity of the original PPP by the fractional moment of χ_v

$$\begin{aligned} \widetilde{\lambda}_v &= \mathbb{E}\left[\chi_v^{2/\alpha}\right] \lambda_v \\ &= \exp\left(\frac{ln10}{5}\frac{\mu_v}{\alpha} + \frac{1}{2}\left(\frac{ln10}{5}\frac{\sigma_v}{\alpha}\right)^2 \right) \lambda_v. \end{aligned} \tag{3.5}$$

In the remaining of the section, we will use the equivalent processes $\widetilde{\Phi}_v$, where $v \in \{M, S, d\}$.

Let D_v denote the distance from the closest point from $\widetilde{\Phi}_v$ to the origin. The nearest point distance distribution of PPP is completely defined by the null probabilities of the process [73]. The probability density function (pdf) and cumulative distribution function (cdf) are given by

$$f_{D_v}(x) = 2\pi\widetilde{\lambda}_v x e^{-\pi\widetilde{\lambda}_v x^2}, x \geq 0 \tag{3.6}$$

$$F_{D_v}(x) = 1 - e^{-\pi\widetilde{\lambda}_v x^2}, x \geq 0. \tag{3.7}$$

We will use distributions in (3.6) and (3.7) to derive distance distributions to the serving BS in UL later on.

Recall that $\widetilde{\Phi}_d$ is PPP that describes the locations of devices. We assume that each BS avoids the interference among the devices associated to it through orthogonal resource allocation. Therefore, in UL the interference arises from devices associated to different BSs and transmit to the same resource unit. Having one interfering device from each BS, the number of interfering devices is equal to the number of BSs. Since the number of devices is larger than the number of BSs, then only a fraction of all devices $\widetilde{\Phi}_d$ causes UL interference. We are modeling the interfering devices by thinning the set $\widetilde{\Phi}_d$, sampling it randomly with probability $p = \frac{\widetilde{N}_{MS}}{\widetilde{N}_d}$, where $\widetilde{N}_{MS} = \widetilde{\lambda}_M A + \widetilde{\lambda}_S A$

is the average number of BSs in the area A and $\widetilde{N}_d = \widetilde{\lambda}_d A$ is the average number of devices in the area. This representation is an approximation due to the dependence among the actual set of interfering devices created by the fact that each device needs to be associated to a different BS. However, as shown in [76], this dependence is weak and the random thinning is justified. The thinned process is denoted by $\widetilde{\Phi}_{I_d}$ and has an intensity of $\widetilde{\lambda}_{I_d} = p\widetilde{\lambda}_d$. Modelling interfering devices by thinning a point process is already used in [72], where the authors show that it is as accurate as modeling all devices, associating them with BSs and randomly selecting one interferer per BS.

Using the notion of a typical device located at the origin, one should calculate the Signal-to-Interference-plus-Noise Ratio (SINR) in the UL at a BS located at $y_v \in \widetilde{\Phi}_v$, which is not at the origin. This problem is simplified by translating the points from all point processes such that the associated BS in UL becomes located at the origin [72]. A homogeneous PPP is stationary, which means that the original and the translated versions have the same distribution for all points in $\mathbb{R}^2$ [73]. Using the definition for signal power in (3.4), the UL SINR can be written as:

$$SINR^{UL} = \frac{P_d h_{y_v} \left\| y_v \right\|^{-\alpha}}{\sum\limits_{y_j \in \widetilde{\Phi}_{I_d}} P_d h_{y_j} \left\| y_j \right\|^{-\alpha} + \sigma^2}, \tag{3.8}$$

where σ^2 is constant noise power at the receiver.

3.2.3 Cell association probabilities

In this section we provide a brief derivation of the joint association probabilities derived in [72] as these association probabilities will be used in the rate derivation in the next section. The association decision is based on the average received signal in DL/UL, averaged over fading. By averaging (3.3) and (3.4), we obtain the signal powers that are used to decide the association:

$$\mathbb{E}_h \left[S_v^{DL} \right] = \mathbb{E}_h \left[P_v h_{x_v} \left\| y_v \right\|^{-\alpha} \right] = P_v \left\| y_v \right\|^{-\alpha} \tag{3.9}$$

$$\mathbb{E}_h \left[S_v^{UL} \right] = \mathbb{E}_h \left[P_d h_{x_v} \left\| y_v \right\|^{-\alpha} \right] = P_d \left\| y_v \right\|^{-\alpha}. \tag{3.10}$$

Using the contact distribution for PPP and the displacement theorem, the remaining of the section describes the derivation of the four possible joint association probabilities cases.

- Case 1: $DL \rightarrow Mcell; UL \rightarrow Mcell$

The joint probability that a device will associate to Mcell in both directions, DL and UL, is defined by the following joint events,

$$\begin{aligned} P_M \left\|y_M\right\|^{-\alpha} > P_S \left\|y_S\right\|^{-\alpha} \cap P_d \left\|y_M\right\|^{-\alpha} > P_d \left\|x_S\right\|^{-\alpha} \\ \left\|y_M\right\|^{-\alpha} > \frac{P_S}{P_M} \left\|y_S\right\|^{-\alpha} \cap \left\|y_M\right\|^{-\alpha} > \left\|y_S\right\|^{-\alpha}. \end{aligned} \tag{3.11}$$

The power of the Scells is significantly smaller than the transmit power of Mcells and hence, the ratio P_S/P_M is less than one. The joint probability reduces to satisfying the second event because if $\left\|y_M\right\|^{-\alpha} > \left\|y_S\right\|^{-\alpha}$ than also $\left\|y_M\right\|^{-\alpha} > \frac{P_S}{P_M} \left\|y_S\right\|^{-\alpha}$.

The joint association probability for case 1 is calculated as

$$\begin{aligned} P(case1) &= P(\left\|y_M\right\|^{-\alpha} > \left\|y_S\right\|^{-\alpha}) \\ &= P(\left\|y_S\right\| > \left\|y_M\right\|) \\ &= \int_0^\infty (1 - F_{Y_S}(y_M)) f_{Y_M}(y_M) \mathrm{d}y_M \\ &= \int_0^\infty e^{-\pi \widetilde{\lambda}_F y_M^2} 2\pi \widetilde{\lambda}_M e^{-\pi \widetilde{\lambda}_M y_M^2} \mathrm{d}y_M \\ &= \frac{\widetilde{\lambda}_M}{\widetilde{\lambda}_M + \widetilde{\lambda}_F}. \end{aligned} \tag{3.12}$$

- Case 2: $DL \rightarrow Mcell; UL \rightarrow Scell$

  The joint association probability for Mcell in DL and Scell in UL is given by the following joint events,

$$\begin{aligned} P_M \left\|y_M\right\|^{-\alpha} > P_S \left\|y_S\right\|^{-\alpha} \cap P_d \left\|y_S\right\|^{-\alpha} \geq P_d \left\|y_M\right\|^{-\alpha} \\ \left\|y_M\right\|^{-\alpha} > \frac{P_S}{P_M} \left\|y_S\right\|^{-\alpha} \cap \left\|y_M\right\|^{-\alpha} \leq \left\|y_S\right\|^{-\alpha}. \end{aligned} \tag{3.13}$$

  The domain that satisfies both conditions is $\frac{P_S}{P_M} \left\|y_S\right\|^{-\alpha} < \left\|y_M\right\|^{-\alpha} \leq \left\|y_S\right\|^{-\alpha}$.

  The joint association probability for case 2 is calculated as

$$
\begin{aligned}
P(case2) &= P(\frac{P_S}{P_M}\left\|y_S\right\|^{-\alpha} < \left\|y_M\right\|^{-\alpha} \leq \left\|y_S\right\|^{-\alpha}) \\
&= P(\left\|y_S\right\| \leq \left\|y_M\right\| < \frac{P_M}{P_S}^{1/\alpha}\left\|y_S\right\|) \\
&= \int_0^\infty (F_{Y_M}(\frac{P_M}{P_S}^{1/\alpha} y_S) - F_{Y_M}(y_S)) f_{Y_S}(y_S)\mathrm{d}y_F \\
&= \int_0^\infty (e^{-\pi\widetilde{\lambda}_M y_S^2} - e^{-\pi\widetilde{\lambda}_M \frac{P_M}{P_S}^{2/\alpha} y_S^2}) 2\pi\widetilde{\lambda}_S e^{-\pi\widetilde{\lambda}_S y_F^2}\mathrm{d}y_M \\
&= \frac{\widetilde{\lambda}_S}{\widetilde{\lambda}_S + \widetilde{\lambda}_M} - \frac{\widetilde{\lambda}_S}{\widetilde{\lambda}_S + \frac{P_M}{P_S}^{2/\alpha}\widetilde{\lambda}_M}.
\end{aligned}
\tag{3.14}
$$

- Case 3: $DL \rightarrow Scell; UL \rightarrow Mcell$

  The joint association probability for associating to Scell in DL and Mcell in UL is given by the following joint events,

$$
\begin{aligned}
P_S\left\|y_S\right\|^{-\alpha} > P_M\left\|y_M\right\|^{-\alpha} \cap P_d\left\|y_M\right\|^{-\alpha} \geq P_d\left\|y_S\right\|^{-\alpha} \\
\left\|y_M\right\|^{-\alpha} < \frac{P_S}{P_M}\left\|y_S\right\|^{-\alpha} \cap \left\|y_M\right\|^{-\alpha} \geq \left\|y_S\right\|^{-\alpha}.
\end{aligned}
\tag{3.15}
$$

  A domain that satisfies both conditions does not exist. The probability that the device will choose Scell in DL and Mcell in UL is equal to zero.

- Case 4: $DL \rightarrow Scell; UL \rightarrow Scell$

  The joint probability that a device will associate to Scell in both directions, DL and UL, is defined by the following joint events,

$$
\begin{aligned}
P_S\left\|y_S\right\|^{-\alpha} > P_M\left\|y_M\right\|^{-\alpha} \cap P_d\left\|y_S\right\|^{-\alpha} > P_d\left\|y_M\right\|^{-\alpha} \\
\left\|y_S\right\|^{-\alpha} > \frac{P_M}{P_S}\left\|y_M\right\|^{-\alpha} \cap \left\|y_S\right\|^{-\alpha} > \left\|y_M\right\|^{-\alpha}.
\end{aligned}
\tag{3.16}
$$

  Since the ratio P_M/P_S is greater than one, the joint probability reduces to satisfying the first event, i.e. $\left\|y_S\right\|^{-\alpha} > \frac{P_M}{P_S}\left\|y_M\right\|^{-\alpha}$.

  The joint probability for case 4 is calculated as

$$
\begin{aligned}
P(case4) &= P(\|y_S\|^{-\alpha} > \frac{P_M}{P_S}\|y_M\|^{-\alpha}) \\
&= P(\|y_M\| > \frac{P_M}{P_S}^{1/\alpha}\|y_S\|) \\
&= \int_0^\infty (1 - F_{Y_M}(\frac{P_M}{P_S}^{1/\alpha} y_S)) f_{Y_S}(y_S)\mathrm{d}y_S \\
&= \int_0^\infty e^{-\pi\widetilde{\lambda}_M \frac{P_M}{P_S}^{2/\alpha} y_S^2} 2\pi\widetilde{\lambda}_S e^{-\pi\widetilde{\lambda}_S y_S^2}\mathrm{d}y_S \\
&= \frac{\widetilde{\lambda}_S}{\widetilde{\lambda}_S + \frac{P_M}{P_S}^{2/\alpha}\widetilde{\lambda}_M}.
\end{aligned}
\tag{3.17}
$$

We note that the conventional association policy, here referred to as DRP (DL Reference signal received Power), is implemented in the following way: If the DL received power from a Mcell is higher than that from a Scell, then the device is associated to a Mcell in *both* DL and UL; otherwise the device is associated to a Scell, again in both DL and UL.

3.2.4 Spectral efficiency

By decoupling UL and DL, we achieve improvement in the UL by adapting UL association to the actual conditions in the UL. Therefore, the analysis of spectral efficiency will be focused on UL only. Moreover, it is focused only on the devices with suboptimal association (devices that have closer Scell but receive higher downlink signal power from Mcell) because the other devices are not affected by path-loss based association in UL. The objective of this section is to analyze the spectral efficiency of the fraction of devices that have suboptimal association with DRP. Those are the devices that, using DRP, are associated to Mcell in both UL/DL, regardless of the fact that they have a closer Scell. Using DUDe, those devices are associated to a Scell in the UL and to a Mcell in the DL.

We derive the distribution of the distance to the serving BS in UL for both DRP and DUDe-based association. From [72], the association region that corresponds to decoupled access is $\frac{P_S}{P_M} D_S^{-\alpha} < D_M^{-\alpha} \leq D_S^{-\alpha}$, where $P_S/P_M < 1$. The distance to the serving BS is denoted as $D_{v,2}$, where $v = M$ with DRP and $v = S$ with DUDe. The second subscript, 2, describes the conditioning on Case 2. When using DUDe, the complementary cumulative distribution function (ccdf) of the distance $D_{S,2}$ to the

serving BS is derived as:

$$\begin{aligned} F^{c}_{D_{S,2}}(x) &= \Pr\left(D_S > x \mid \frac{P_S}{P_M} D_S^{-\alpha} < D_M^{-\alpha} \leq D_S^{-\alpha}\right) \\ &= \frac{\Pr\left(D_S > x; D_S \leq D_M < \left(\frac{P_M}{P_S}\right)^{1/\alpha} D_S\right)}{\Pr(\text{Case 2})} \\ &= \frac{\int_{x}^{\infty} \left(e^{-\pi\widetilde{\lambda}_M x_s^2} - e^{-\pi\widetilde{\lambda}_M \left(\frac{P_M}{P_S}\right)^{2/\alpha} x_s^2}\right) f_{D_S}(x_s)\mathrm{d}x_s}{\Pr(\text{Case 2})}. \end{aligned} \tag{3.18}$$

where Pr(Case 2) is given by (3.14). The cdf of the distance is $F_{D_{S,2}}(x) = 1 - F^{c}_{D_{S,2}}(x)$. By differentiating the cdf, we derive the pdf of the distance to the serving BS when DUDe is used, conditioned on Case 2,

$$\begin{aligned} f_{D_{S,2}}(x) &= \frac{dF_{D_S}(x)}{dx} \\ &= \frac{\left(e^{-\pi\widetilde{\lambda}_M x^2} - e^{-\pi\widetilde{\lambda}_M \left(\frac{P_M}{P_S}\right)^{\frac{2}{\alpha}} x^2}\right) f_{D_S}.(x)}{\Pr(\text{Case 2})} \end{aligned} \tag{3.19}$$

Following the same procedure, the distribution of the distance to the serving base station using DRP, conditioned on Case 2 is

$$f_{D_{M,2}}(x) = \frac{\left(e^{-\pi\widetilde{\lambda}_S \left(\frac{P_S}{P_M}\right)^{\frac{2}{\alpha}} x^2} - e^{-\pi\widetilde{\lambda}_S x^2}\right) f_{D_M}(x)}{\Pr(\text{Case 2})}. \tag{3.20}$$

When DUDe is used, the distance to the serving base station in (3.8) has a pdf defined by (3.19). In the remaining part of this section we will derive the spectral efficiency for the decoupled access. The results for DRP association follow by substituting the distribution of the distance to the associated base station by (3.20).

The spectral efficiency, or equivalently, the normalized throughput with DUDe is defined as:

$$C_{DUDe} = \mathbb{E}\left[\log_2\left(1 + SINR^{UL}\right)\right] \tag{3.21}$$

For $T > 0$, $\mathbb{E}[T] = \int_0^\infty \Pr(T > t)\mathrm{d}t$. Applying this property in equation (3.21), the spectral efficiency reduces to:

$$\begin{aligned} C_{DUDe} &= \frac{1}{\ln(2)} \int_0^\infty \Pr\left(\ln\left(1 + SINR^{UL}\right) > t\right) \mathrm{d}t \\ &= \frac{1}{\ln(2)} \int_0^\infty \Pr\left(SINR^{UL} > e^t - 1\right) \mathrm{d}t \end{aligned} \tag{3.22}$$

The integrand in (3.22) is basically a definition for coverage probability, with SINR threshold set to $e^t - 1$. It is evaluated as:

$$\begin{aligned} &\Pr\left(SINR^{UL} > e^t - 1\right) = \\ &= \Pr\left(\frac{P_d h_{y_S} D_{S,2}^{-\alpha}}{I_y + \sigma^2} > e^t - 1\right) \\ &= \mathbb{E}_y\left[P\left(h_{y_S} > (e^t - 1)y^\alpha \left(I_y + \sigma^2\right) | D_{S,2} = y\right)\right] = \\ &= \int_0^\infty \mathbb{E}_{I_y}\left[e^{-(e^t-1)y^\alpha I_y}\right] e^{-(e^t-1)y^\alpha \sigma^2} f_{D_S,2}(y)\mathrm{d}y = \\ &= \int_0^\infty L_{I_y}\left((e^t - 1)y^\alpha\right) e^{-(e^t-1)y^\alpha \sigma^2} f_{D_S,2}(y)\mathrm{d}y \end{aligned} \tag{3.23}$$

where $L_{I_y}\left((e^t - 1)y^\alpha\right)$ is the Laplace Functional (LF) of the interference, derived as

$$\begin{aligned} &L_{I_y}(s) = \mathbb{E}_{I_y}\left[e^{-sI_y}\right] = \\ &= E_{\widetilde{\Phi}_{I_d}}\left[\prod_{y_j \in \widetilde{\Phi}_{I_d}} \mathbb{E}_h\left[e^{-sh_{y_j}\|y_j\|^{-\alpha}}\right]\right] \\ &= \exp\left(-2\pi\widetilde{\lambda}_{I_d} \int_0^\infty \left(1 - \frac{1}{1 + sv^{-\alpha}}\right) v\mathrm{d}v\right) \end{aligned} \tag{3.24}$$

Combining (3.19), (3.23) and (3.24), we derive the final expression for spectral efficiency with decoupled access conditioned on devices with suboptimal association (Case 2), given in (3.25). The expression for spectral efficiency with DRP for devices with suboptimal association has a similar form to (3.25) and is given by (3.26).

$$C_{DUDe} = \frac{\log_2(e)}{\Pr(\text{Case 2})} \int_0^\infty \int_0^\infty e^{-\pi\widetilde{\lambda}_{I_d}(e^t-1)^{\frac{2}{\alpha}}y^2 \int_0^\infty \left(\frac{1}{1+v^{\alpha/2}}\right)\mathrm{d}v} e^{-\frac{(e^t-1)y^\alpha\sigma^2}{P_d}} \times \left(e^{-\pi\widetilde{\lambda}_M y^2} - e^{-\pi\widetilde{\lambda}_M\left(\frac{P_M}{P_S}\right)^{\frac{2}{\alpha}}y^2}\right) 2\pi\widetilde{\lambda}_S y e^{-\widetilde{\lambda}_S\pi y^2}\mathrm{d}t\mathrm{d}y \tag{3.25}$$

$$C_{DRP} = \frac{\log_2(e)}{\Pr(\text{Case 2})} \int_0^\infty \int_0^\infty e^{-\pi\widetilde{\lambda}_{I_d}(e^t-1)^{\frac{2}{\alpha}}y^2 \int_0^\infty \left(\frac{1}{1+v^{\alpha/2}}\right)\mathrm{d}v} e^{-\frac{(e^t-1)y^\alpha\sigma^2}{P_d}} \times \left(e^{-\pi\widetilde{\lambda}_S\left(\frac{P_S}{P_M}\right)^{\frac{2}{\alpha}}y^2} - e^{-\pi\widetilde{\lambda}_S y^2}\right) 2\pi\widetilde{\lambda}_M y e^{-\widetilde{\lambda}_M\pi y^2}\mathrm{d}t\mathrm{d}y \tag{3.26}$$

3.2.5 Numerical results

The analysis presented in previous sections is validated by numerical results. The association probabilities for joint association in DL and UL are analyzed for two types of Scells, high power Scells (Pcells) and low power Scells (Fcells); here the high/low power refers to the power used by the base station of the particular Scell. Fig. 3.2 shows the association probabilities for a Mcell-Fcell and Mcell-Pcell two-tier network. In both cases, it is visible that increasing the number of Scells rapidly increases the probability for decoupled access, which also corresponds to the percentage of devices with decoupled access. It is important to notice that for the Mcell-Fcell network, this percentage goes nearly to 75%. This corroborates that a high percentage of devices in today's networks have suboptimal association and thus achieve suboptimal performance. For Mcell-Pcell networks, the percentage of devices with decoupled access is slightly over 50% because high power Scells force the devices to connect to Scells in DL, thus increasing the probability for Case 4 (DL/UL access with Scells).

By further increasing the Scells density, the probability for decoupled access decreases at the expense of increased probability for Case 4. Basically, the difference between the power levels of the two tiers reflects in the association process, making trade-off only between the devices with decoupled access and devices with DL/UL association to Scells. The higher the disparity between P_M and P_S, the higher the probability of decoupled access. The percentage of devices that associate to Mcells in both directions constantly decreases.

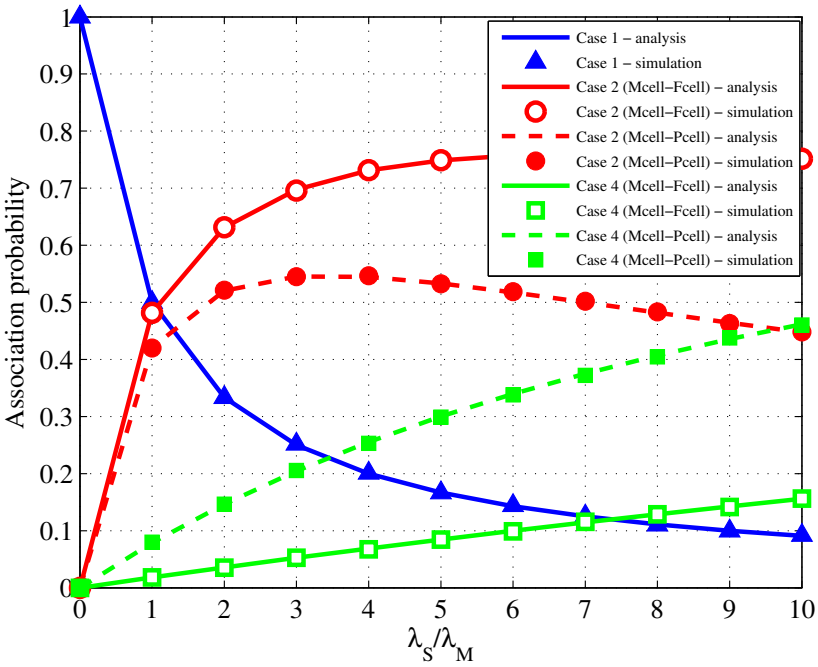


Fig. 3.2 Association probability for Mcell-Fcell and Mcell-Pcell heterogeneous network ($P_M = 46$ dBm, $P_S = 20$ dBm for Mcell-Fcell; $P_S = 30$ dBm for Mcell-Pcell, $P_d = 20$ dBm, $\alpha = 3$).

The distance distributions to the serving BS with DUDe and with DRP, derived in Section 3.2.4, are shown in Fig. 3.3. Both network settings are included, Mcell-Fcell and Mcell-Pcell, respectively. By decoupling DL/UL, the distance distribution becomes narrower and shifts on the left towards smaller distances, i.e. there is higher probability that the serving BS will be closer to the device. This conclusion holds for both Fcells and Pcells. For the same BS density the pdf of the distance to the serving BS with DUDe for Mcell-Fcell network is narrower and is shifted to the left compared to the same distribution for Mcell-Pcell network. This is due to the fact that Pcells, with their higher transmit power, are able to associate more devices than Fcells and hence only the farthest devices remain in Case 2. On the other hand, Fcells have smaller coverage and therefore, there are many devices that are close to the Fcell but remain in Case 2. This percentage of devices increases the probability that the serving BS is closer to the devices that belong to Case 2 for Mcell-Fcell network.

By using DUDe, one can achieve a significant improvement in UL signal power at the associated BS. This results in increased SINR and thus improved spectral efficiency in the system. The results in Fig. 3.4 compare the UL throughput with DUDe and DRP for both network settings. The UL throughput with DUDe is calculated by multiplying

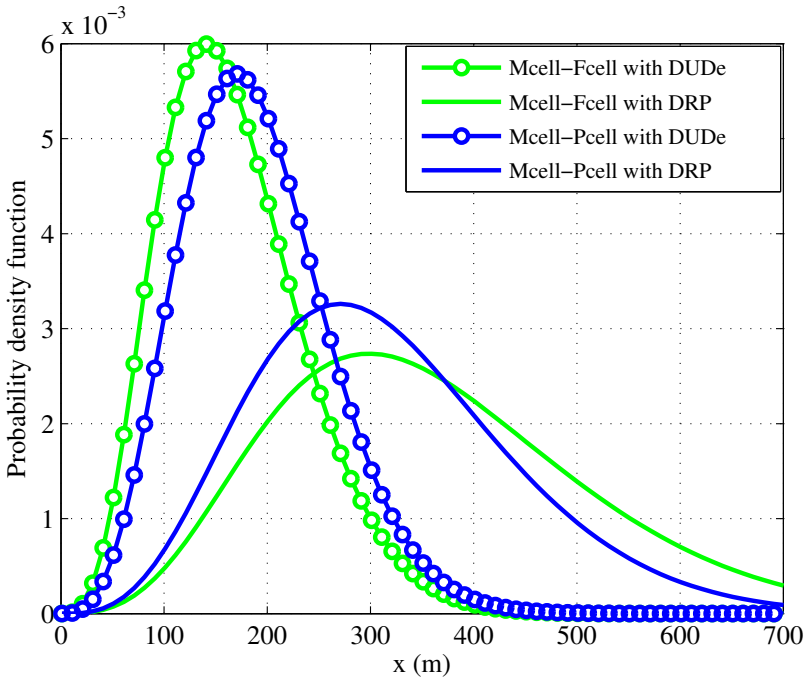


Fig. 3.3 Probability density function of the distance to the serving BS ($P_M = 46dBm, P_S = 20dBm$ for Mcell-Fcell; $P_S = 30dBm$ for Mcell-Pcell, $\lambda_S = 5\lambda_M, \alpha = 4$).

the spectral efficiency by B/N_a, where B is the system frequency bandwidth and N_a is the average number of associated devices per BSs. Given the average number of devices in the area N_d and the average number of BSs $N_{MS} = N_M + N_S$, $N_a = N_d/N_{MS}$. When DRP association is applied, N_a is calculated using the association probabilities, i.e. $N_a = N_d(\Pr(\text{Case 1}) + \Pr(\text{Case 2}))/N_M$.

The abscissa shows the UL throughput with DRP and the ordinate shows the value with decoupled access for same BS density. It can be noted that the gains are significant. For instance, for Mcell-Fcell network, the UL throughput of 10 kbps with DRP maps into $>$ 150 kbps with DUDe. For the Mcell-Pcell network, the UL throughput of 30 kbps with DRP maps into 500 kbps with DUDe.

Another interesting observation is that Mcell-Pcell network achieves higher throughput with DRP than Mcell-Fcell network, for the same BS density. On the other hand, Mcell-Fcell is superior in UL throughput with DUDe. If we observe the point for $\lambda_S = 15\lambda_M$, we can see that with DRP, Mcell-Pcell achieves 25 kbps, while Mcell-Fcell network achieves 7 kbps. Using DUDe leads to the opposite situation, Mcell-Fcell achieves 500 kbps and Mcell-Pcell achieves 400 kbps. This phenomenon is a consequence of the distance distributions presented in Fig. 3.3. Basically, using high power Scells (Pcells), we force the devices to associate to Scells in both directions, leaving

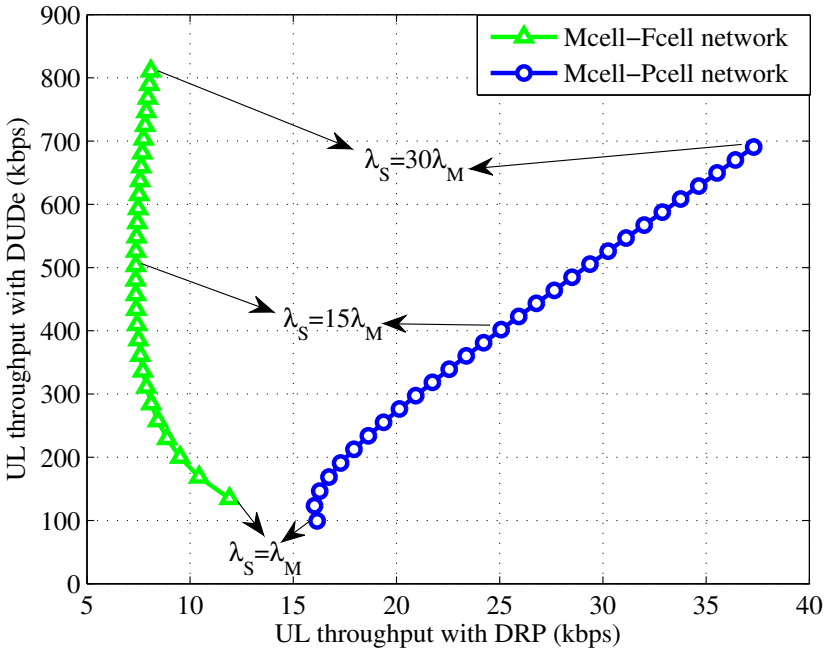


Fig. 3.4 Spectral efficiency for Case 2 devices with and without decoupled access ($P_d = 20dBm, \alpha = 4, \lambda_d = 10^4/km^2, B = 20MHz$).

the worst case devices with DUDe. When using low power Scells (Fcells), we have additional percentage of devices in Case 2, which are closer to Scells and contribute to higher UL throughput with DUDe.

In this chapter and the previous one, we have elaborated the concept of decoupled access with two particularly different settings, one based on theoretical model in the chapter at hand and one based on experimental data in Chapter 2. Although both models have different assumptions and parameters, the trends in association probabilities are observable in both settings. This leads us to the conjecture that *the association probability depends chiefly on the density of the deployment, but not the process used to generate the deployment geometry.*

In order to test this conjecture, we have also evaluated the association probabilities in a *third* scenario, in which the BSs are deployed in a regular grid. The grid model is generated in the following way: N_{MS} BSs are positioned in a grid such that they cover the same area A as the BSs in the stochastic geometry model; each of them is assigned with transmit power P_M with probability Q and with transmit power P_S with probability $B = 1 - Q$. Using the values of Q and B we can manipulate with the densities of the Mcells and Scells. The results for the association probability are shown in Fig. 3.5 and it is clearly visible that they are favorable to our conjecture. Using grid

model to prove particular trend derived with stochastic geometry has been already used in [77]; in our case this verification has a significant additional value to the match with the trend in the experimental setting.

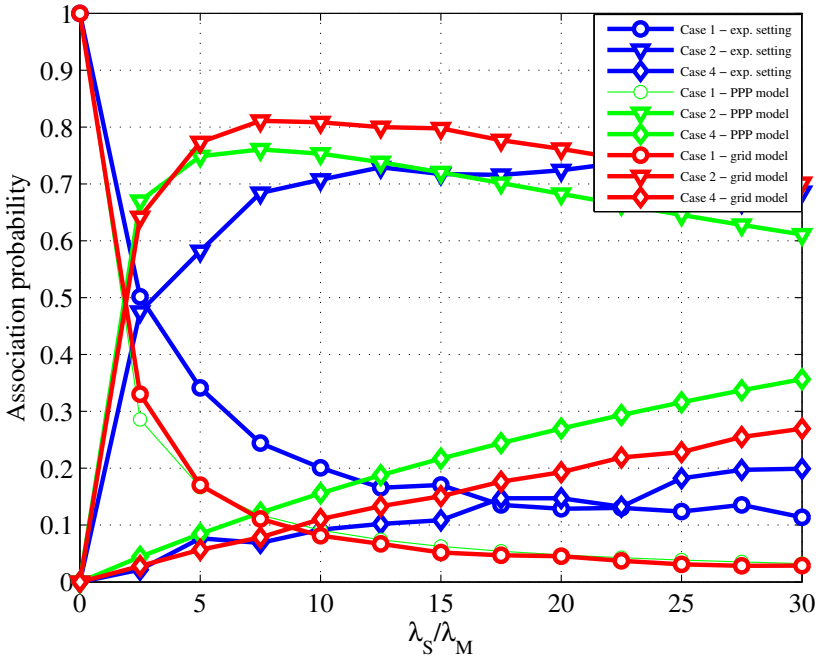


Fig. 3.5 Association probability for three different network settings: experimental model, PPP model and grid model.

3.3 Analysis of DUDe in millimeter wave heterogeneous networks

In this section we introduce a different type of heterogeneity in our work. Whilst so far our work was based on heterogeneous networks in terms of the cell type/size, we introduce spectrum heterogeneity in this section where we study cell association in a mixed millimeter wave (mmWave) and sub 6 GHz (sub-6GHz) scenario. MmWaves have different characteristics than conventional sub-6GHz frequencies as will be discussed in detail later on and the interplay between these two different frequency ranges is very interesting to study.

3.3.1 Overview of millimeter waves

The ever increasing cellular network traffic demands is expected to grow by at least 40% to 70% year over year [78]. This growth implies that within the next decades cellular networks are expected to deliver more than 1000 times the capacities of today. At the same time, we see the rise of the internet of things (IoT) where cellular networks are not exclusive to the conventional broadband and voice services but extend to all sorts of applications and industries which is expected to add millions and millions of new connected devices in the next years. These trends have lead to several initiatives in cellular networks including the densification of the network and the emergence of heterogeneous networks where small cells are playing a key role in providing capacity in hotspots and dense urban areas. Another approach was to explore new unused frequency bands that would help in satisfying the projected traffic needs of 2020.

Wireless communications in millimeter waves (mmWaves) bands, between 30 and 300 GHz, is not new. In fact, the first mmWaves communication setup was demonstrated by Bose more than 100 years ago [79]. Nowadays, mmWaves are being mostly used for satellite communications [80] and cellular backhaul [81, 82]. More recently, mmWaves have been used for high throughput wireless local area networks (LAN) and personal area networks (PAN) systems in the unlicensed 60 GHz band [83–86] where rates of more than 1 Gb/s can be achieved. However, it is believed that mmWave communications is typically for short range or point-to-point line of sight (LOS) conditions.

Mmwaves have the advantage of very wide bandwidths which can easily be 200 times greater than all cellular allocations today (under 3 GHz) [83, 87]. In addition, the small wavelengths of mmWaves combined with advances in low-power complementary metal-oxide-semiconductor (CMOS) RF circuits have enabled large number of miniaturized antennas to be integrated into a relatively small area. These multiple antenna systems can be used to form very high gain electrically steerable beams at the base station side, at the UE side or even on a chip [88, 89]. These advances have lead to tremendous interest in mmWaves in the recent years as an attractive solution to the spectrum churn in beyond 4G and 5G cellular networks.

However, the development of cellular networks in the mmWave band still faces major technical challenges and the applicability of mmWaves in real life requires careful assessment. We will now further detail the main challenges in mmWaves deployments.

Challenges of millimeter waves

Despite the great prospects and large bandwidths available at mmWaves, there are still quite a few challenges to realize the vision of cellular networks in such high bands.

- **Range and directional transmission**: Friis' transmission law [59] states that that the free space omnidirectional path loss grows with the square of the frequency. However, the smaller wavelength of mmWaves enables having a larger antenna gain for the same antenna size. In fact, the higher frequency of mmWave does not mean an increase in the propagation loss as long as the antenna size is kept constant and directional transmission is being used. The measurements in [90] confirm these insights. However, relying on directional transmission would need certain design principles to be considered that will be discussed in the next subsection.

- **Shadowing**: A more serious concern for mmWaves than range is that mmWaves are extremely sensitive to shadowing or blockage. For instance, materials such as brick can attenuate the signals by as much as 40 - 80 dB [91, 87, 92, 93] and the human body could result in attenuations in the order of 20-35 dB [94]. Rain and foliage are also common problems for mmWaves.

- **Rapid channel variations**: Channel coherence time is linear in the carrier frequency for a certain velocity [59]. This means that the coherence time in mmWaves is very small, hence the channel will change in the order of hundreds of microsconds which is much faster than current cellular networks. Additionally, the high sensitivity to blockage means that the sudden appearance of an obstacle could result in major fluctuations in the channel quality although beam steering may overcome that [95]. Mmwaves will mostly be associated to Scells which means more handovers. All these factors would result in a highly intermittent communications that would need fast adaptation techniques.

- **Mutiuser coordination**: Current applications of mmWaves are mostly limited to point to point transmissions such as backhaul, LAN or PAN applications with limited number of UEs. However, with the expected high spatial reuse of spectrum, there will be simultaneous transmissions on the same band which would require new MAC mechanisms to handle the interfering links. MAC layer aspects of mmWave networks are discussed in [96].

- **Processing power consumption**: A significant challenge in leveraging the multi antenna and wide bandwidth features of mmWaves is the power consumption

of the analog-to-digital A/D conversion. Power consumption generally scales linearly with the sampling rate and exponentially with the number of bits per sample [83, 97], this makes high resolution quantization at wide bandwidths and large number of antennas hard to accomplish for low power, low cost devices.

- **Uplink transmit power**: Recent studies on electromagnetic field exposure [98] show that to be compliant with applicable exposure limits at frequencies above 6 GHz, the maximum transmit power in the uplink might have to be several dB below the power levels used for current cellular technologies. This makes it challenging for the uplink in mmWave as the transmit power has an important impact on uplink coverage, in particular for sounding over a non-precoded channel.

Design issues for millimeter waves in 5G

The above mentioned challenges show that while mmWaves promise much higher capacities than conventional sub 6 GHz, cellular systems need to be significantly redesigned to cope with the properties and challenges of mmWaves. In this section we identify the main design issues that need to be addressed from a system perspective in order to achieve to full gain from mmWaves.

- **Directional transmission**: The main implication of using mmWaves is that performance gains depend on directional transmissions. Directionality with suitable beamforming can fully compensate the frequency dependent propagation losses. In this case, the SINR distribution of mmWaves could be better than that of sub 6 GHz frequencies since interference plays a smaller role in mmWaves as will be shown later. However, high directionality could result in challenges in the design of synchronization and broadcast signalling used in the initial cell search. Both BSs and UEs may need to scan different angles before finding the right beam alignment which could incur a lot of delay in the initial access as well as in handovers. This problem was highlighted in [99] and solutions were discussed in [96].

- **Multiple access and duplexing consideration**: Time domain duplexing (TDD) is an attractive duplexing mechanism for mmWaves. One practical reason for that is that it is hard and maybe inefficient to find large contingent UL and DL bandwidths to do FDD. However, an issue that is related to directional transmission arises which is how to support FDMA within the TDD time slots. The current cellular systems use digital processing for MIMO and beamforming.

However, with the high bandwidth and the large number of antennas at mmWaves, it is not practical from a complexity, power or cost to use high resolution wideband A/D converters at each antenna element in mmWaves. Most commercial solutions have considered a phased array approach where signals are combined in free space or RF with phase shifters [100] prior to the A/D conversion. A limitation of this approach is that FDMA cannot be used in the TDD time slots which means that only one UE can be served in each time slot. This limitation results in a few problems: (a) in terms of UL power, using only TDMA means that the power of only one UE can be received at each time slot, with UEs at the cell edge having limited power, this would result in a cap on the capacity due to power limitation rather than bandwidth. This can be largely improved with the use of FDMA as shown in [90]. (b) Support for small packets: the support for multiuser transmission is essential for an efficient transmission of short messages as well as low latency applications where it is not possible to wait for several time slots to transmit. This lack of FDMA in mmWaves will make these two applications quite hard to support efficiently. (c) Power consumption: from a baseband power consumption point of view it might be easier for UEs to process a small chunk of the bandwidth (say 50 MHz) during one time slot instead of processing the whole bandwidth which can be in the order of 1 GHz. Therefore supporting multiple access in mmWaves is essential for several 5G applications.

- **Directional relaying**: A key design aspect of mmWaves is the support for relaying. Relays are a feature that emerged in LTE as a solution to either extend coverage or, to a lesser extent, to improve capacity [101]. However, the gains from relays have been quite modest in current cellular networks therefore they are not widely adopted. On the other hand, due to the intermittent nature of a mmWave channel, the sensitivity to blockage and the poor propagation characteristics, relays could be an attractive solution to solve the coverage problems using directional relays to serve certain UEs that have poor coverage or even indoors.

- **Noise limited**: Cellular networks have always been limited by interference where noise can sometimes be ignored, this lead the wireless communities to come up with several techniques to mitigate the interference in cellular networks such as intercell interference coordination (ICIC), interference alignment and coordinated multipoint transmission. However, due to the high directionality of mmWaves transmissions the beams became more or less isolated and we have moved towards a more noise limited scenario where the previously mentioned

schemes have limited or no gains at all. This implies that point-to-point, rather than network, technologies will be more relevant in the future.

- **Heterogeneous networks issues**: mmWaves cannot be deployed in a standalone mode due to their limited coverage and poor propagation characteristics. Therefore, they will have to coexist with the conventional macro cell network that will provide the umbrella coverage which is essential in cellular networks. This will require the exploration of new techniques to make the best of this new heterogeneous network, the aspect of cell association is studied in the next section. Carrier aggregation is an attractive option to keep a seamless connection to the network using the sub 6 GHz network and provide high data rates in hotspots corresponding to mmWaves coverage.

- **Licensing of Spectrum**: The current spectrum licensing regime where operators are granted exclusive rights for certain chunks of spectrum might not be the best or most efficient route to take with mmWaves. The noise limited nature of mmWaves could allow several operators to transmit on the same spectrum in the same vicinity without causing noticeable interference to each other, this aspect was recently studied in [102–104]. Furthermore, the large bandwidths available at mmWaves are not expected to be fully used all the time mainly because of the limited coverage which results in fewer UEs connected to each mmWave Scell. Therefore, the mmWave spectrum could be idle most of the time which is another argument for resource sharing between operators.

3.3.2 Related work and Contribution

Related Work

Downlink and uplink associations are typically coupled, i.e. a UE connects to the same BS in the DL and UL. In the context of a heterogeneous network, downlink-uplink decoupling (DUDe), as discussed in details before, has been recently shown to significantly improve the network capacity (especially in the UL) by considering different association criteria for the UL and DL [33]. DUDe has been discussed in [9, 71, 105] as an interesting component for future cellular networks. Significant improvement in throughput and signal-to-interference-and-noise-ratio (`SINR`) have been shown in [33] with realistic simulations, while [72, 49, 48] reached similar conclusions from a theoretical perspective.

Meanwhile, starting with [69], modeling and analyzing cellular networks using stochastic geometry has become a popular and accepted approach to understanding their performance trends. Most relevant to this study, mmWave networks were analyzed assuming a Poisson point process (PPP) for the base station (BS) distribution in [106–108]. In [106] a line-of-sight (LOS) ball model was considered for blockage modeling where BSs inside the LOS ball were considered to be in LOS whereas any BS outside of the LOS ball was treated as NLOS. In [107], this blocking model was modified by adding a LOS probability within the LOS ball, and this approach was shown to reflect several realistic blockage scenarios. Therefore we consider the same approach in this section. Decoupled association in a mixed sub-6GHz and mmWave deployment was very recently considered in [108] from a resource allocation perspective. However, there is no complete or analytical study to our knowledge on downlink-uplink decoupling for mmWave networks or the mmWave-sub-6GHz hybrid network considered in this section which motivates our following study.

Contributions

We model a cellular network with sub-6GHz macrocells (Mcells) and mmWave small cells (Scells) each distributed according to an independent Poisson point process. A UE can in general independently connect to either type of BS on the UL and DL. The key technical contributions of this section are the following.

Cell association probabilities. We derive the cell association probabilities based on the UL and DL maximum biased received power where the different parameters that affect the association trends are highlighted and discussed in detail. Subsequently, a similar analysis based on the UL and DL maximum achievable rate is given. The role of decoupled access is discussed in detail.

Coverage and rate trends. The UL and DL `SINR` and rate coverage probabilities are derived, where a special emphasis is put on Scell biasing. We show that high biasing values can be used for mmWave Scells due to the abundant bandwidth in the mmWave bands. The altered UL and DL `SINR` and rate coverage with the biasing value are also studied.

System design insights. The analytical results, which employ a number of simplifying approximations, are validated later in the section. Design insights are highlighted which include:

- Decoupled access plays a key role in mmWave deployments and the gains of decoupling are more pronounced in less dense urban environments.

- Scell beamforming gain improves the association probability to Scells dramatically and therefore needs to be considered in the association phase.
- Aggressive values of small cell biasing are possible thanks to the wide bandwidth offered by mmWaves. Supporting these large biasing values requires having robust low modulation and coding techniques to allow UEs to operate in very low SINR.

3.3.3 System model

A two-tier heterogeneous network is considered where Mcells and Scells are distributed uniformly in $\mathbb{R}^2$ according to independent homogeneous Poisson point processes (PPP) Φ_m and Φ_s with densities λ_m and λ_s respectively. Specifically, a deployment of sub-6GHz Mcells overlaid by mmWave Scells is considered. The UEs are also assumed to be uniformly distributed according to a homogeneous PPP Φ_u with density λ_u. The analysis is done for a typical UE located at the origin where the BS serving the typical UE is referred to as the tagged BS [2]. The notation is summarized in Table 3.1. The inclusion of sub-6GHz Scells is left for future work.

Propagation assumptions

The received power in the DL at a UE at location $u \in \Phi_u$ from a sub-6GHz Mcell (m) at $x \in \Phi_m$ or a mmWave Scell (s) at $y \in \Phi_s$ is given by $\mathrm{P}_m h_{x,u} \beta_m G_m L_m (x-u)^{-1}$ or $\mathrm{P}_s h_{y,u} \beta_s G_s(\theta) L_s (y-u)^{-1}$, respectively. Here, L is the pathloss where for the typical UE at the origin $L_m(x) = \|x\|^{\alpha_m}$ and $L_s(y) = \|y\|^{\alpha_s(y)}$, α is the pathloss exponent (PLE) where $\alpha_s(y)$ equals α_l if the link is LOS and α_n otherwise, h is the small scale fading power gain where in this study we consider Rayleigh fading, β is the the near-field pathloss at 1 m and G is the antenna gain. UEs are assumed to have omni-directional antennas so the antenna gains are only accounted for at the BS side. All mmWave Scells are equipped with directional antennas with a sectorized gain pattern assuming a simplified rectangular antenna pattern that was used in [107] where a UE receives a signal with $G_{s_{max}}$ if the UE's angle (θ) with respect to the best beam alignment is within the main beamwidth (θ_s) of the serving cell and $G_{s_{min}}$ otherwise. This is formulated by

$$G_s(\theta) = \begin{cases} G_{s_{max}} & if\ |\theta| \leq \frac{\theta_s}{2} \\ G_{s_{min}} & otherwise \end{cases} .$$

[2] The analysis of the typical UE is enabled by Slivnyak's theorem.

The UL received signal powers are derived by replacing P_m or P_s by P_{um} or P_{us}, and interchanging x or y with u, respectively. Shadowing is ignored in this study since for mmWaves the blockage model introduces a similar effect to shadowing. As for the sub-6GHz network, as shown in [69], the randomness of the PPP BS locations emulates the shadowing effect, therefore shadowing is ignored in the sub-6GHz model as well.

All UEs served by the Scells are assumed to be in perfect alignment with their serving cells whereas the beams of all interfering links are assumed to be randomly oriented with respect to each other and hence the gain on the interfering links is considered to be random. Results for the association probabilities considering different antenna gains are subsequently shown in order to study how important it is to have antenna alignment in the cell association phase.

Table 3.1 Notation and simulation parameters

Notation	**Parameter**	**Value (if applicable)**
Φ_m, λ_m	Mcells PPP and density	$\lambda_m = 5$ per km^2
Φ_s, λ_s	Scells PPP and density	$\lambda_s = 50$ per km^2
Φ_u, λ_u	UEs PPP and density	$\lambda_u = 200$ per km^2
f_m, f_s	sub-6GHz and mmWave carrier frequencies	2 GHz, 70 GHz
W_m, W_s	sub-6GHz, mmWave bandwidth	20 MHz, 1 GHz
P_m, P_s	Mcell and Scell transmit power	46 dBm, 30 dBm
P_{um}, P_{us}	UE transmit power to Mcell and Scell	23 dBm
K_{UL}, K_{DL}	UL and DL association tiers	
T_s, T'_s	DL and UL association bias of mmWave Scells	
T_m, T'_m	DL and UL association bias of sub-6GHz Mcells	
α_m	Pathloss exponent for Mcells	3
α_l, α_n	LOS and NLOS pathloss exponent for Scells	2, 4
$G_{s_{max}}$, $G_{s_{min}}$, θ_s	Main lobe gain, side lobe gain and 3 dB beamwidth for mmWave	18 dBi, -2 dBi, 10°
G_m	Mcell antenna gain (omni-directional)	0 dBi
ω, μ	Fractional LOS area ω in a ball of radius μ	0.11, 200 m
N_m, N_s	Load of serving Macro or Small cell	
$\mathcal{A}$, $\mathcal{B}$	Association probabilities based on max. biased received power and max. rate	
h	Small scale fading	$h \sim \exp(1)$
β	$\beta = \left(\frac{\text{carrier wavelength}}{4\pi}\right)^2$ is the pathloss at 1m	
σ_m^2, σ_s^2	Noise powers for sub-6GHz and mmWave	-174 dBm/Hz + $10\log_{10}$(W) + 10 dB

Blockage model

A simple yet accurate blockage model that was proposed in [107] is used where a UE within a distance μ from a Scell is assumed LOS with probability ω and 0 otherwise. The parameters ω and μ are environment dependent; the Manhattan scenario from [107] is considered for this study. Results for other values of ω and μ are shown in Section 3.3.7 to study their effect on cell association.

Biased uplink and downlink cell association

It is assumed that the UL and DL cell associations are based on different criteria, namely the UL and DL biased received powers, respectively. The typical user associates with BS at $x^* \in \Phi_l$, where $l \in \{s, m\}$, in UL if and only if

$$\mathrm{P}_{ul}\mathrm{T}'_l\psi_l L_l(x^*)^{-1} \geq \mathrm{P}_{uk}\mathrm{T}'_k\psi_k L^{-1}_{\min,k},\ \forall k \in \{s, m\}, \tag{3.27}$$

where $\psi_k = G_k\beta_k$ is the combination of antenna gain and near-field pathloss and G_k is equal to $G_{s_{max}}$ or G_m in the mmWave or sub-6GHz cases, respectively. $L_{\min,k} = \min_{x\in\Phi_k} L_k(x)$ is the minimum pathloss of the typical UE from the k^{th} tier and T' and T are the UL and DL cell bias values respectively. Similarly, the typical user associates with BS at $x^* \in \Phi_l$ in DL if and only if

$$\mathrm{P}_l\mathrm{T}_l\psi_l L_l(x^*)^{-1} \geq \mathrm{P}_k\mathrm{T}_k\psi_k L^{-1}_{\min,k},\ \forall k \in \{s, m\}. \tag{3.28}$$

The assumption that large bandwidth mmWave networks are noise-limited has been considered and motivated in [107]. We show that this assumption holds even for high densities of mmWave Scells. Henceforth, this assumption will be considered for this study and is validated later on with simulation results. Consequently and in order to simplify the analysis, the signal-to-noise-ratio (`SNR`) is considered instead of the `SINR` for the mmWave links. With no interference between the two tiers due to the orthogonality of both frequency bands, the UL/DL sub-6GHz `SINR` and mmWave `SNR` of a typical UE at the origin are given by

$$\mathtt{SINR}_{\mathrm{UL},m} = \frac{\mathrm{P}_{um}\psi_m h_{0,x^*} L_m(x^*)^{-1}}{I_{\mathrm{UL},m} + \sigma_m^2},\quad \mathtt{SINR}_{\mathrm{DL},m} = \frac{\mathrm{P}_m\psi_m h_{x^*,0} L_m(x^*)^{-1}}{I_{\mathrm{DL},m} + \sigma_m^2},$$

$$\mathtt{SNR}_{\mathrm{UL},s} = \frac{\mathrm{P}_{us}\psi_s h_{0,x^*} L_s(x^*)^{-1}}{\sigma_s^2},\quad \mathtt{SNR}_{\mathrm{DL},s} = \frac{\mathrm{P}_s\psi_s h_{x^*,0} L_s(x^*)^{-1}}{\sigma_s^2}, \tag{3.29}$$

where $I_{\mathrm{UL},m} = \sum_{y \in \Phi_{Iu}} \mathrm{P}_{um} \psi_m h_{y,x^*} L_m(y - x^*)^{-1}$, $I_{\mathrm{DL},m} = \sum_{x \in \Phi_m \setminus x^*} \mathrm{P}_m \psi_m h_{x,0} L_m(x)^{-1}$ and Φ_{Iu} is the point process denoting the locations of UEs transmitting in the UL on the same resource as the typical UE. It is assumed that each BS has at least one UE in its association region. With this assumption, the realizations of Φ_{Iu} have one point randomly chosen from the association cell of each BS other than the serving BS, which represents the interfering UE (y) from that cell in the UL. Furthermore, the queues in the UL and DL are assumed to be always full and resources are on average equally distributed among the UEs (e.g. by proportional fair or round robin scheduling). The DL rate of the typical UE connected to a Mcell or Scell is given by

$$R_{\mathrm{DL},m} = \frac{\mathrm{W}_m}{N_m} \log(1 + \mathtt{SINR}_{\mathrm{DL},m}), \quad R_{\mathrm{DL},s} = \frac{\mathrm{W}_s}{N_s} \log(1 + \mathtt{SNR}_{\mathrm{DL},s}), \tag{3.30}$$

where N_m and N_s are the loads on the serving Mcell and Scell respectively. $R_{\mathrm{UL},m}, R_{\mathrm{UL},s}$ are defined similarly.

3.3.4 Received power based cell association

In this section, the UL and DL cell association probabilities are derived for four different cases where K_{DL} and K_{UL} denote the DL and UL association tiers of the typical UE. Hence, the below cases denote the probability of the UE associating to the Mcell and Scell in the UL and DL assuming a decoupled UL and DL association approach.

- Case 1: $\mathbb{P}(K_{\mathrm{DL}} = Mcell)$
- Case 2: $\mathbb{P}(K_{\mathrm{UL}} = Mcell)$
- Case 3: $\mathbb{P}(K_{\mathrm{DL}} = Scell)$
- Case 4: $\mathbb{P}(K_{\mathrm{UL}} = Scell)$

Note that the sum of probabilities of Case 1 and 3 equals 1 and similarly for Case 2 and 4. The association probabilities are derived in this subsection maximizing the biased DL/UL received power whereas in the next subsection the association probabilities are derived maximizing the DL/UL rate. Subsequently, the outcomes from the two association strategies are compared in the results subsection.

In order to derive the association probabilities, we first characterize the point process formed by the pathloss between each BS and the typical UE at the origin. Assuming a BS at $x \in \mathbb{R}^2$, the pathloss point process is defined as $\mathcal{N}_l := \{L_l(x) = \|x\|^{\alpha_l}\}_{x \in \Phi_l}$, where $l \in \{m, s\}$. Making use of the displacement theorem, $\mathcal{N}_l$ is a Poisson point

process with intensity measure denoted by $\Lambda_l(.)$ similar to [107, 109]. Since the pathloss in the sub-6GHz and mmWave cases has different characteristics, we will have two independent pathloss processes for mmWave and sub-6GHz given by $\mathcal{N}_s$ and $\mathcal{N}_m$ respectively. Therefore, the intensities, probability distribution function (PDF) and complementary cumulative distribution function (CCDF) will be derived separately for mmWave and sub-6GHz.

Lemma 1. *The distribution of the pathloss from the typical UE to the tagged BS is such that* $\mathbb{P}(L_l(x) > t) = \exp(-\Lambda_l((0,t)]$, *where* $l \in \{m, s\}$, *the intensity measures for pathloss in mmWave and sub-6GHz are given by*

$$\Lambda_s((0,t)] = \pi\lambda_s \Bigg(\left(\omega t^{\frac{2}{\alpha_l}} + (1-\omega)t^{\frac{2}{\alpha_n}}\right) \mathbb{1}(t < \mu^{\alpha_l}) + \left(\omega\mu^2 + (1-\omega)t^{\frac{2}{\alpha_n}}\right) \times \mathbb{1}(\mu^{\alpha_l} \leq t \leq \mu^{\alpha_n}) + t^{\frac{2}{\alpha_n}} \mathbb{1}(t > \mu^{\alpha_n}) \Bigg) \tag{3.31}$$

$$\Lambda_m((0,t)] = \pi\lambda_m t^{\frac{2}{\alpha_m}}. \tag{3.32}$$

Proof: See Appendix A.1. ■

Since $\mathcal{N}_l$ is a PPP, the CCDF of pathloss to the tagged BS is $\bar{F}_l(t) = \mathbb{P}(L_l(x) > t) = \exp(-\Lambda_l((0,t)]$ and the PDF is given by $f_l(t) = \frac{-d\bar{F}_l(t)}{dt} = \Lambda_l'((0,t)] \exp(-\Lambda_l((0,t)]$ for $l \in (m, s)$. The expressions for the pathloss process CCDFs for mmWave and sub-6GHz are given by

$$\bar{F}_s(t) = \exp\Bigg(-\pi\lambda_s \Bigg(\left(\omega t^{\frac{2}{\alpha_l}} + (1-\omega)t^{\frac{2}{\alpha_n}}\right) \mathbb{1}(t < \mu^{\alpha_l}) + \left(\omega\mu^2 + (1-\omega)t^{\frac{2}{\alpha_n}}\right) \times \mathbb{1}(\mu^{\alpha_l} \leq t \leq \mu^{\alpha_n}) + t^{\frac{2}{\alpha_n}} \mathbb{1}(t > \mu^{\alpha_n}) \Bigg) \Bigg) \tag{3.33}$$

$$\bar{F}_m(t) = \exp\left(-\pi\lambda_m t^{\frac{2}{\alpha_m}}\right) \tag{3.34}$$

and the corresponding PDFs by

$$f_s(t) = 2\pi\lambda_s \frac{t^{\frac{2}{\alpha_n}-1}}{\alpha_n} \Bigg(\left(\frac{\alpha_n \omega t^{\frac{2}{\alpha_l}-\frac{2}{\alpha_n}}}{\alpha_l} + (1-\omega) \right) \exp\left(-\pi\lambda_s \left(\omega t^{\frac{2}{\alpha_l}} + (1-\omega) t^{\frac{2}{\alpha_n}} \right) \right) \times \mathbb{1}(t < \mu^{\alpha_l}) + (1-\omega) \exp\left(-\pi\lambda_s \left(\omega\mu^2 + (1-\omega) t^{\frac{2}{\alpha_n}} \right) \right) \mathbb{1}(\mu^{\alpha_l} \leq t \leq \mu^{\alpha_n}) + \exp\left(-\pi\lambda_s t^{\frac{2}{\alpha_n}} \right) \mathbb{1}(t > \mu^{\alpha_n}) \Bigg) \tag{3.35}$$

$$f_m(t) = \frac{2\pi\lambda_m t^{\frac{2}{\alpha_m}-1}}{\alpha_m} \exp\left(\pi\lambda_m t^{\frac{2}{\alpha_m}} \right). \tag{3.36}$$

All the needed components to derive the association probabilities specified above are now available.

We then move to deriving the UL and DL association probabilities maximizing the biased UL and DL received power respectively. This method is referred to as maximum biased received power (Max-BRP). It is assumed that the DL and UL serving cells are chosen based on the biased DL and UL received powers respectively. The association probabilities are defined in the following definition and the final expressions are given in Lemma 2.

Definition 1. Max-BRP Association probabilities. The probabilities of the typical UE associating to a sub-6GHz Mcell or mmWave Scell based on the maximum biased received power in the downlink or uplink is defined as

$$\mathcal{A}_{\mathrm{DL},m} \triangleq \mathbb{P}\left(\mathrm{P}_m \mathrm{T}_m \psi_m L_{\min,m}^{-1} > \mathrm{P}_s \mathrm{T}_s \psi_s L_{\min,s}^{-1} \right) \tag{3.37}$$

$$\mathcal{A}_{\mathrm{UL},m} \triangleq \mathbb{P}\left(\mathrm{P}_{um} \mathrm{T}'_m \psi_m L_{\min,m}^{-1} > \mathrm{P}_{us} \mathrm{T}'_s \psi_s L_{\min,s}^{-1} \right) \tag{3.38}$$

$$\mathcal{A}_{\mathrm{DL},s} \triangleq \mathbb{P}\left(\mathrm{P}_s \mathrm{T}_s \psi_s L_{\min,s}^{-1} > \mathrm{P}_m \mathrm{T}_m \psi_m L_{\min,m}^{-1} \right) \tag{3.39}$$

$$\mathcal{A}_{\mathrm{UL},s} \triangleq \mathbb{P}\left(\mathrm{P}_{us} \mathrm{T}'_s \psi_s L_{\min,s}^{-1} > \mathrm{P}_{um} \mathrm{T}'_m \psi_m L_{\min,m}^{-1} \right). \tag{3.40}$$

Lemma 2. *The uplink and downlink association probability to a sub-6GHz Mcell or mmWave Scell are given below.*

$$\begin{aligned}\mathcal{A}_{c,m} = &\frac{2\pi\lambda_m}{\alpha_m a_c^{\frac{2}{\alpha_m}}} \int_0^\infty l^{\frac{2}{\alpha_m}-1} \exp\left(-\pi\lambda_m \left(\frac{l}{a_c}\right)^{\frac{2}{\alpha_m}}\right) \\ &\times \Bigg(\exp\left(-\pi\lambda_s(\omega l^{\frac{2}{\alpha_l}} + (1-\omega) l^{\frac{2}{\alpha_n}})\right) \mathbb{1}(l < \mu^{\alpha_l}) \\ &+ \exp\left(-\pi\lambda_s\left((1-\omega) l^{\frac{2}{\alpha_n}} + \omega\mu^2\right)\right) \mathbb{1}(\mu^{\alpha_l} \le l \le \mu^{\alpha_n}) \\ &+ \exp\left(-\pi\lambda_s l^{\frac{2}{\alpha_n}}\right) \mathbb{1}(l > \mu^{\alpha_n})\Bigg) \mathrm{d}l \\ \mathcal{A}_{c,s} = &\, 1 - \mathcal{A}_{c,m},\end{aligned} \tag{3.41}$$

where $c \in \{\mathrm{UL}, \mathrm{DL}\}$, $a_{\mathrm{DL}} = \frac{\mathrm{P}_s \mathrm{T}_s \psi_s}{\mathrm{P}_m \mathrm{T}_m \psi_m}$ and $a_{\mathrm{UL}} = \frac{\mathrm{P}_{us} \mathrm{T}'_s \psi_s}{\mathrm{P}_{um} \mathrm{T}'_m \psi_m}$.

Proof: The proof for $\mathcal{A}_{\mathrm{DL},m}$ is given below.

$$\begin{aligned}\mathcal{A}_{\mathrm{DL},m} &= \mathbb{P}\left(\mathrm{P}_{um}\mathrm{T}_m\psi_m L_{min,m}^{-1} > \mathrm{P}_{us}\mathrm{T}_s\psi_s L_{min,s}^{-1}\right) = \mathbb{P}\left(L_{min,s} > a_{\mathrm{DL}} L_{min,m}\right) \\ &= \int_0^\infty \bar{F}_s(a_{\mathrm{DL}} l_m) f_m(l_m) \mathrm{d}l_m,\end{aligned}$$

where the last step follows from the fact that $\mathbb{P}(X > Y) = \int_0^\infty \mathbb{P}(X > y) f_Y(y) \mathrm{d}y$. Changing variables as $l = a_{\mathrm{DL}} l_m$ yields

$$\mathcal{A}_{\mathrm{DL},m} = \frac{1}{a_{\mathrm{DL}}} \int_0^\infty \bar{F}_s(l) f_m\left(\frac{l}{a_{\mathrm{DL}}}\right) \mathrm{d}l.$$

This directly results in $\mathcal{A}_{\mathrm{DL},m}$, and $\mathcal{A}_{\mathrm{UL},m}$ follows similarly. ■

Corollary 1. *The association probabilities can be acquired in closed form for the special case where $\alpha_l = 2$ and $\alpha_n = \alpha_m = 4$ and with simple mathematical manipulation, $\mathcal{A}_{\mathrm{DL},m}$ can be expressed by*

$$\begin{aligned}\mathcal{A}_{\mathrm{DL},m} = &\frac{\pi\lambda_m}{\sqrt{a_{\mathrm{DL}}}} \Bigg(\frac{\sqrt{\pi} e^{\frac{c_2^2}{4c_1}}}{2\sqrt{c_1}} \left(Q\left(\frac{c_2}{\sqrt{2c_1}}\right) - Q\left(\frac{2\mu c_1 + c_2}{\sqrt{2c_1}}\right)\right) \\ &+ e^{-\mu^2 c_1}\left(\frac{e^{-\mu c_2}}{c_2} - \frac{c_1 e^{-\mu^2 c_2}}{c_2(c_1+c_2)}\right)\Bigg),\end{aligned} \tag{3.42}$$

where $c_1 = \pi\lambda_s\omega$, $c_2 = \pi\lambda_s(1-\omega) + \frac{\pi\lambda_m}{a_{\mathrm{DL}}^{2/\alpha_n}}$ and $Q(.)$ is the Q-function. Similarly, the other three cases can be obtained in closed form.

3.3.5 Rate based cell association

In this part the UL and DL association probabilities are derived where the association criteria are the UL and DL rates respectively. The sub-6GHz Macro DL association probability is given by

$$\mathcal{B}_{\mathrm{DL},m} = \mathbb{P}\left(\frac{\mathrm{W}_m}{N_{\mathrm{DL},m}}\log_2(1+\mathtt{SINR}_{\mathrm{DL},m}) > \frac{\mathrm{W}_s}{N_{\mathrm{DL},s}}\log_2(1+\mathtt{SINR}_{\mathrm{DL},s})\right).$$

It is assumed that $\mathtt{SINR}_{\mathrm{DL},m} \approx \mathtt{SIR}_{\mathrm{DL},m}$ and $\mathtt{SINR}_{\mathrm{DL},s} \approx \mathtt{SNR}_{\mathrm{DL},s}$ for simplicity since sub-6GHz frequencies are interference limited whereas mmWaves are rather noise limited. In order to simplify the expressions, an approximation[3] that was proposed in [42] is used where the cell load is characterized by the average number of UEs per cell on the corresponding tier. The average load on the serving BS of tier l for the UL and DL is given by

$$\bar{N}_{c,l} = 1 + \frac{1.28\lambda_u\mathcal{B}_{c,l}}{\lambda_c} \quad \text{for } l \in \{m,s\} \text{ and } c \in \{\mathrm{UL},\mathrm{DL}\}. \tag{3.43}$$

This approximation results in

$$\mathcal{B}_{\mathrm{DL},m} = \mathbb{P}\left(\mathtt{SIR}_{\mathrm{DL},m} > (1+\mathtt{SNR}_{\mathrm{DL},s})^{\left(\frac{\mathrm{W}_s(\lambda_m+1.28\lambda_u\mathcal{B}_{\mathrm{DL},m})\lambda_s}{\mathrm{W}_m(\lambda_s+1.28\lambda_u\mathcal{B}_{\mathrm{DL},s})\lambda_m}\right)} - 1\right). \tag{3.44}$$

Having $\mathcal{B}_{\mathrm{DL},m}$ on both sides of the equation makes it very hard to solve. Therefore we resort to a simple approximation by neglecting the load term in the rate expression (setting N_m and N_s to 1). In other words deriving the association probability based on the maximum achievable rate in the UL and DL. This approach is suboptimal but it results in a tractable expression for the association probability and also suffices our purpose of showing different decoupling trends as compared to Max-BRP. The association trends resulting from this approximation are also validated in Fig. 3.13(b). Henceforth this method is referred to as *Max-Rate* and the corresponding association probabilities are now defined.

Definition 2. Max-Rate Association Probability. The association probabilities in the UL and DL to a sub-6GHz Mcells and mmWave Scells for the Max-Rate

[3]This approximation was proposed for sub-6GHz in [42] and was later verified for mmWaves in [107].

case are given by

$$\mathcal{B}_{c,m} \triangleq \mathbb{P}\left(\mathtt{SIR}_{c,m} > (1+\mathtt{SNR}_{c,s})^{\left(\frac{\mathrm{W}_s}{\mathrm{W}_m}\right)} - 1\right) \tag{3.45}$$

$$\mathcal{B}_{c,s} \triangleq \mathbb{P}\left(\mathtt{SNR}_{c,s} > (1+\mathtt{SIR}_{c,m})^{\left(\frac{\mathrm{W}_m}{\mathrm{W}_s}\right)} - 1\right), \tag{3.46}$$

where $c \in \{\mathrm{UL}, \mathrm{DL}\}$.

Using power control in the UL would complicate the UL coverage expression where the derived expressions in [76] and [48] include two to three integrals. Furthermore, it is assumed that UEs transmit with their maximum power on mmWaves since mmWaves are coverage limited, therefore to make the analysis consistent and fair we assume that UEs transmit with their maximum power on sub-6GHz as well. With the assumption of no power control the UL and DL coverage expressions (neglecting noise and considering exponential fading) are the same. In order for this assumption to be valid we also need to assume that the interferers in the UL are PPP distributed and that the exclusion region around the typical UE/BS in the DL/UL are the same. Although the latter might seem to be a strong assumption it will be shown in Fig. 3.9(a) that the derived rate based association probability matches very well the simulation results, verifying that the above assumptions are valid.

The final expressions for the Max-Rate based association probabilities based on Definition 2 are given in the following Lemma.

Lemma 3. *The DL and UL association probability based on the maximum achievable rate for mmWave Scells and sub-6GHz Mcells are given below.*

$$\mathcal{B}_{c,m} = \int_0^\infty \frac{f_{\mathtt{SNR}_{c,s}}(z)}{1+\rho\left((1+z)^{\frac{\mathrm{W}_s}{\mathrm{W}_m}} - 1\right)} \mathrm{d}z \tag{3.47}$$

$$\mathcal{B}_{c,s} = 1 - \mathcal{B}_{c,m}, \tag{3.48}$$

where $c \in \{\mathrm{UL}, \mathrm{DL}\}$, $f_{\mathtt{SNR}_{\mathrm{DL},s}(z)} = \frac{\sigma_s^2}{\mathrm{P}_s \psi_s} \int_0^\infty l \exp\left(\frac{-z\sigma_s^2 l}{\mathrm{P}_s \psi_s}\right) f_s(l) dl$, $\rho(t, \alpha_m) = t^{\frac{2}{\alpha_m}} \int_{t^{\frac{-2}{\alpha_m}}}^\infty \frac{\mathrm{d}u}{1+u^{\frac{\alpha_m}{2}}}$ and $f_{\mathtt{SNR}_{\mathrm{UL},s}}$ is the same as $f_{\mathtt{SNR}_{\mathrm{DL},s}}$ exchanging P_s by P_{us} .

Proof: See Appendix A.2. ■

After deriving the association probabilities, the UL and DL SINR and rate coverage probabilities are derived in the next section.

3.3.6 SINR and rate distributions: Downlink and Uplink

In this section the SINR and rate coverage distributions are derived for the DL and UL in the mmWave and sub-6GHz cases. These distributions would help in studying the effect of the different association strategies on the SINR and rate of the whole system. The SINR and rate CCDFs will be derived for the Max-BRP association case only as the derivation for the Max-Rate association is quite complicated and will be left for future work. However, we use biasing in the results for SINR and rate coverage probabilities to validate some of the trends that result from the Max-Rate association strategy.

SINR coverage

The SINR coverage can be defined as the average fraction of UEs that at any given time achieve SINR τ. The SINR coverage is the CCDF of the SINR over the entire network which, due to the assumption of stationary PPP for the UEs and BSs, can be characterized considering the typical link between the typical UE at the origin and its serving BS. Since mmWave networks are usually noise limited (i.e. $\texttt{SINR} \approx \texttt{SNR}$), we consider the SNR coverage for mmWaves while still considering SINR for sub-6GHz. The SINR/SNR coverage in the sub-6GHz and mmWave cases is expressed as: $\mathcal{P}_m \triangleq \mathbb{P}(\texttt{SINR} > \tau)$ and $\mathcal{P}_s \triangleq \mathbb{P}(\texttt{SNR} > \tau)$ respectively. Since there is no interference between the mmWave and sub-6GHz BSs, the SINR/SNR coverage can be derived separately for sub-6GHz and mmWave.

Similar to the previous section, since UL transmissions on mmWaves are assumed to be at maximum power, the sub-6GHz UL SINR coverage is derived assuming maximum UL transmit power (no power control) for simplicity and fairness. The final expression for the UL coverage probability with fractional pathloss compensation power control is given below.

Theorem 1. *The* SINR *coverage probability for the typical UL and DL links based on the Max-BRP association criterion is given by*

$$\begin{aligned}\mathcal{P}_{\mathrm{DL}}(\tau) &= \mathcal{P}_{\mathrm{DL},m}(\tau) + \mathcal{P}_{\mathrm{DL},s}(\tau) \\ &= \int_0^\infty \exp\left(\frac{-\tau\sigma_m^2 l}{\mathrm{P}_m\psi_m}\right) \exp\left(\frac{-2\pi\lambda_m}{\alpha_m}\int_l^\infty \frac{t^{\frac{2}{\alpha_m}-1}}{1+\frac{t}{\tau l}}\mathrm{d}t\right) \bar{F}_s(a_{\mathrm{DL}} l) f_m(l)\mathrm{d}l \\ &+ \int_0^\infty \exp\left(\frac{-\tau\sigma_s^2 l}{\mathrm{P}_s\psi_s}\right) \bar{F}_m\left(\frac{l}{a_{DL}}\right) f_s(l)\mathrm{d}l\end{aligned} \tag{3.49}$$

$$\begin{aligned}\mathcal{P}_{\mathrm{UL}}(\tau) &= \mathcal{P}_{\mathrm{UL},m}(\tau) + \mathcal{P}_{\mathrm{UL},s}(\tau) \\ &= \int_0^\infty \exp\left(\frac{-\tau\sigma_m^2 l}{\mathrm{P}_{um}\psi_m}\right) \exp\left(\frac{-2\pi\lambda_m}{\alpha_m}\int_l^\infty \frac{t^{\frac{2}{\alpha_m}-1}}{1+\frac{t}{\tau l}}\mathrm{d}t\right) \bar{F}_s(a_{\mathrm{UL}} l) f_m(l)\mathrm{d}l \\ &+ \int_0^\infty \exp\left(\frac{-\tau\sigma_s^2 l}{\mathrm{P}_{us}\psi_s}\right) \bar{F}_m\left(\frac{l}{a_{\mathrm{UL}}}\right) f_s(l)\mathrm{d}l,\end{aligned} \tag{3.50}$$

where $\bar{F}_s, \bar{F}_m, f_s,\ f_m,\ a_{\mathrm{DL}}$ and a_{UL} have been derived/defined in Section 3.3.4.

Proof: See Appendix A.3. ■

The final expression of $\mathcal{P}_{\mathrm{UL}}$ with fractional pathloss compensation power control is given by

$$\begin{aligned}\mathcal{P}_{\mathrm{UL}}(\tau) &= \int_0^\infty \exp\left(\frac{-\tau\sigma_m^2 l^{1-\epsilon}}{\mathrm{P}_{um}\psi_m}\right) \\ &\times \exp\left(\frac{-2\pi\lambda_m}{\alpha_m}\int_l^\infty \left(1 - \int_0^\infty \frac{2\pi\lambda_m u^{\frac{2}{\alpha_m}-1} e^{-\pi\lambda_m u^{\frac{2}{\alpha_m}}}}{\alpha_m(1+\tau l^{1-\epsilon}u^\epsilon t^{-1})}\mathrm{d}u\right) t^{\frac{2}{\alpha_m}-1}\mathrm{d}t\right) \\ &\times \bar{F}_s(a_{\mathrm{UL}} l) f_m(l)\mathrm{d}l + \int_0^\infty \exp\left(\frac{-\tau\sigma_s^2 l^{1-\epsilon}}{\mathrm{P}_{us}\psi_s}\right) \bar{F}_m\left(\frac{l}{a_{\mathrm{UL}}}\right) f_s(l)\mathrm{d}l,\end{aligned} \tag{3.51}$$

where ϵ is the pathloss compensation factor. The proof for $\mathcal{P}_{\mathrm{UL},m}$ follows along the same lines as in [76] therefore the proof is omitted. The inclusion of the power control adds an extra integral to the Mcell coverage expression which makes it quite complex. Therefore we stick to the assumption of no power control and use the expression in Theorem 1.

Rate coverage

In order to derive the rate coverage, the load on both Mcell and Scell tiers needs to be characterized. We resort to the same approximation used in the previous section where the load is given by (3.43). This approximation is validated with simulation results in Fig. 3.10(b).

Definition 3. The rate coverage probability is defined as

$$\mathcal{R}(\rho) = \mathbb{P}(R > \rho) = \mathbb{P}\left(\frac{\mathrm{W}}{N}\log_2(1+\mathtt{SINR}) > \rho\right) = \mathbb{P}\left(\mathtt{SINR} > 2^{\frac{\rho N}{\mathrm{W}}} - 1\right).$$

Using the above definition, the UL and DL rate coverage probabilities are given by

$$\begin{aligned}\mathcal{R}_{\mathrm{DL}}(\rho) &= \mathcal{R}_{\mathrm{DL},m}(\rho) + \mathcal{R}_{\mathrm{DL},s}(\rho) \\ &= \mathcal{P}_{\mathrm{DL},m}\left(2^{\frac{\rho\bar{N}_{\mathrm{DL},m}}{\mathrm{W}_m}} - 1\right) + \mathcal{P}_{\mathrm{DL},s}\left(2^{\frac{\rho\bar{N}_{\mathrm{DL},s}}{\mathrm{W}_s}} - 1\right)\end{aligned} \tag{3.52}$$

$$\begin{aligned}\mathcal{R}_{\mathrm{UL}}(\rho) &= \mathcal{R}_{\mathrm{UL},m}(\rho) + \mathcal{R}_{\mathrm{UL},s}(\rho) \\ &= \mathcal{P}_{\mathrm{UL},m}\left(2^{\frac{\rho\bar{N}_{\mathrm{UL},m}}{\mathrm{W}_m}} - 1\right) + \mathcal{P}_{\mathrm{UL},s}\left(2^{\frac{\rho\bar{N}_{\mathrm{UL},s}}{\mathrm{W}_s}} - 1\right).\end{aligned} \tag{3.53}$$

3.3.7 Performance evaluation

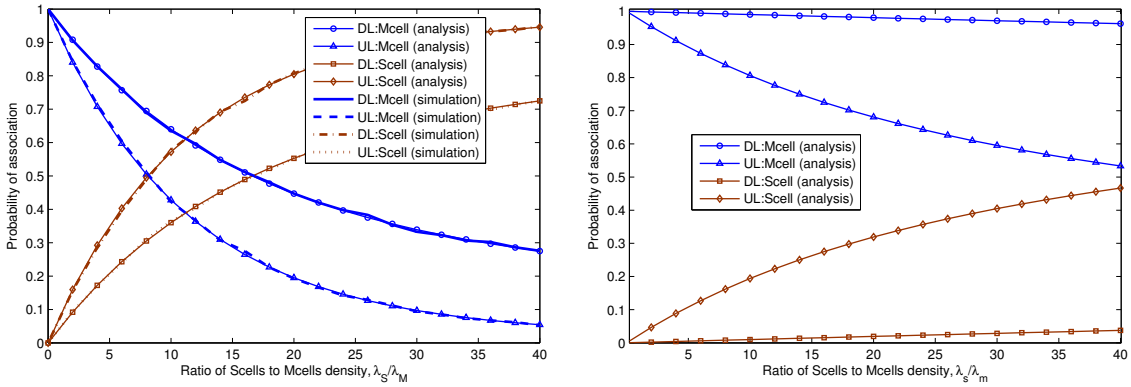
In this section we validate our analysis with Monte Carlo simulations where in each simulation run, UEs and BSs are dropped randomly according to the corresponding densities. All UEs are assumed outdoor. The association criteria, propagation and blockage model are as described in Section 3.2.2 and the simulation parameters follow Table 3.1. Using the analytical results, the different factors that affect the association probability are studied. Special emphasis is placed on the Downlink and Uplink Decoupling (DUDe) [33] to understand if decoupling is still useful in the case of mmWave networks. Furthermore, the `SINR` and rate coverage trends are illustrated considering the special case of biased DL received power association where the effect of small cell biasing on both `SINR` and rate trends is studied with a reflection on the implications in real deployments.

The parameter values in Table 3.1 are used as a baseline. Some of the parameters are altered in some figures in order to understand their effect on the association probability.

Association probability

We start by looking into the Max-BRP association probabilities derived in Section 3.3.4 and the different factors that affect these probabilities. It is assumed that $G_s = G_{s_{max}}$ in the association phase. The UL and DL association bias values are unity (0 dB) unless otherwise stated.

Association analysis validation. Fig. 3.6(a) illustrates the association probabilities derived in Lemma 2 against the ratio of Scells to Mcells density and compared with simulation results. It can be seen that the simulation and analysis results have a very close match which validates our analysis and gives confidence in using the analysis for the following results. Furthermore, there is a difference between the DL and UL association probabilities for Scells and Mcells, this difference represents the decoupled



(a) (b)

Fig. 3.6 (a) Association probability and validation of the analysis with simulation results for $G_s = 23$ dBi. (b) Association probability with $G_s = 0$ dBi.

access where UEs prefer to connect to different cells in the UL and DL. We refer to this difference as the *decoupling gain* for the rest of the paper. In the Scell case, it can be noticed that the UL association probability is always higher than the DL one, this is because the UL coverage of Scells is larger than its DL coverage and vice versa with Mcells. The figure shows that more than 20% of the UEs have decoupled access at a ratio of Scells to Mcells of 40, in other words the decoupling gain is 20%.

Antenna gain's effect on the association probability. Fig. 3.6(b) shows the association probability where $G_s = 0$ dBi, i.e. there is no antenna gain. Predictably, there is very low Scell association probability in the UL and DL. On the other hand, Fig. 3.6(a) shows a high Scell association probability with $G_s = 23$ dBi.

Another observation is that higher G_s leads to lower decoupling gain. This stems from the blocking model which is represented by a LOS ball. Most of the DL coverage is inside the LOS ball (with a certain probability of low pathloss exponent (PLE) ($\alpha_l = 2$)) while the UL coverage extends to the NLOS area (with higher PLE ($\alpha_n = 4$)). Therefore increasing the antenna gain expands the DL coverage at a faster rate than the UL coverage due to the difference in the PLE between the LOS and NLOS areas. This in effect reduces the difference between the UL and DL coverage of Scells which, in turn, reduces the decoupling gain. It is worth noting that this trend could be seen with other blockage models as it only depends on the fact that UEs that are closer to the mmWave Scells have higher LOS probability than UEs that are further away which is a general characteristic that would be included in most blockage models.

Fig. 3.7 illustrates the ratio of Scell to Mcell density λ_s/λ_m at which the crossing point between the Scell and Mcell UL and DL association curves occurs versus the

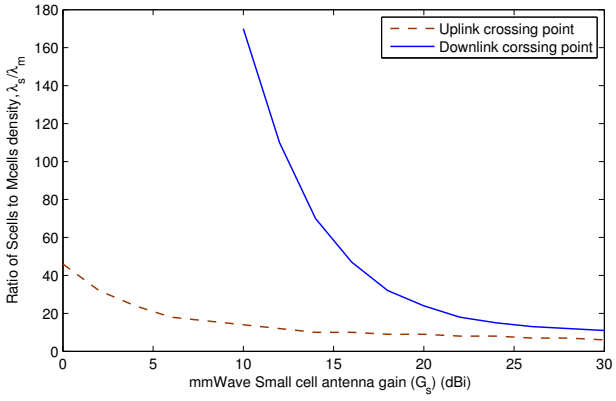


Fig. 3.7 The density λ_s/λ_m at which the crossing between the Mcell and Scell UL and DL association probability curves occurs versus the Scell antenna gain G_s.

antenna gain G_s. The difference between the two curves is an indication of the decoupling gain which is shown to decrease with the increase of G_s which confirms the trend in Fig. 3.6. In addition, the decreasing tendency of the curves indicate that the crossing point happens at a lower density of Scells the more the gain is increased which results in more and more UEs associating to the Scells.

Pathloss exponent and LOS ball parameters effect on the association probability. Fig. 3.8 shows the effect of the PLE and the LOS ball parameters on the decoupling gain. Having a higher α_n and α_m than α_l results, as in the previous figure, in reducing the difference between the DL and UL coverages of the Scell since the DL coverage is assumed mostly in the LOS ball which makes the DL coverage expand as α_l gets smaller resulting in reducing the gap between the DL and UL coverage borders and, in turn, decreasing the decoupling gain. Therefore the higher the difference between α_n or α_m with α_l, the lower the decoupling gain.

On the other hand, having a higher LOS ball radius (μ) results into a higher decoupling gain since as μ gets larger more of the UL coverage of Scells area is included in the LOS region which helps in expanding the UL coverage of Scells and, in turn, increases the decoupling gain. The lower PLE and larger μ are characteristics of a low density urban environment and indicate that decoupling is more relevant in such a scenario.

Max-Rate association probability validation and trends. The results in Fig. 3.9 are based on the Max-Rate association probability derived earlier. Fig. 3.9(a) illustrates the comparison between the analysis and simulation where the very close match between them validates our analysis and the assumption of having the same exclusion region for UL and DL. The rate based association results into more offloading

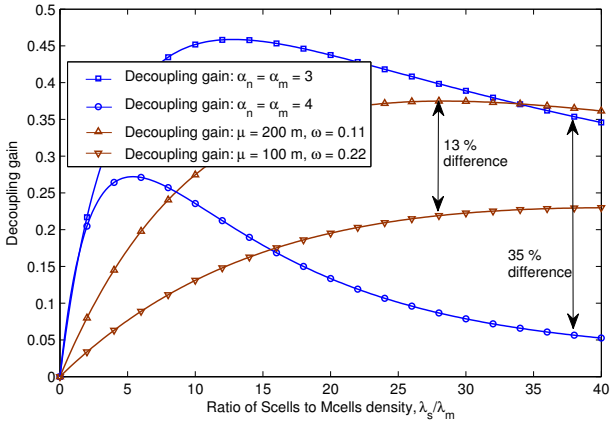


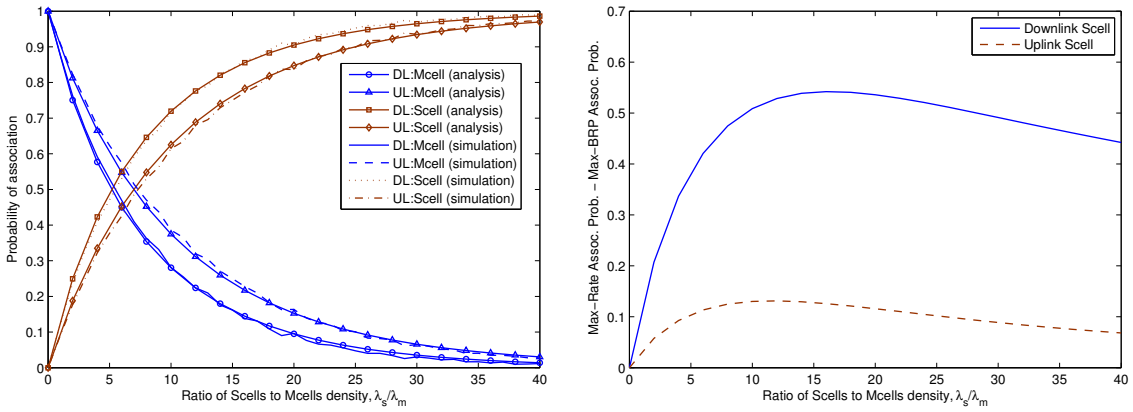
Fig. 3.8 Variation of the decoupling gain with the pathloss exponents (α) and the LOS ball parameters.

of UEs towards mmWave Scells as compared to Max-BRP in Fig. 3.6. This is a direct result from the much wider bandwidth at mmWave Scells. It can also be noticed that there is a decoupling gain in the Max-Rate association as well. However, in this case the decoupled association results from UEs tending to connect to a Scell in the DL and to a Mcell in the UL which is shown by the superior Scell DL association probability compared to the UL and vice versa with the Mcell case. This behaviour is opposite to the Max-BRP association in Fig. 3.6. This is a result of the higher bandwidth at mmWave Scells which pushes more UEs to connect to the Scells in the UL and DL and since –in general– the UL range is more limited than that of the DL then the mmWave Scells can afford to serve more UEs in the DL than in the UL. This effect is amplified the further the UEs are from the Scell. At a certain point the UL connection towards the Scell is too weak whereas the DL one is relatively stronger and this is the point where the decoupling happens.

This effect is further clarified in Fig. 3.9(b) where the increase in the DL association probability in the Max-Rate case over the Max-BRP case is more than 40% higher than the UL increase. This important result will be further confirmed in the subsequent results.

SINR and rate coverage results

In this part we present several results for the SINR and rate coverage to illustrate the effect that the mixed sub-6GHz/mmWave deployment has on the SINR and rate distributions and how the bias can affect these distributions. From this point onwards, we consider that $\mathrm{T}'_s = \frac{\mathrm{P}_s \mathrm{T}_s}{\mathrm{P}_{us}}$ and $\mathrm{T}'_m = \frac{\mathrm{P}_m \mathrm{T}_m}{\mathrm{P}_{um}}$ where $\mathrm{T}'_m = \mathrm{T}_m = 0$ dB. In other words,



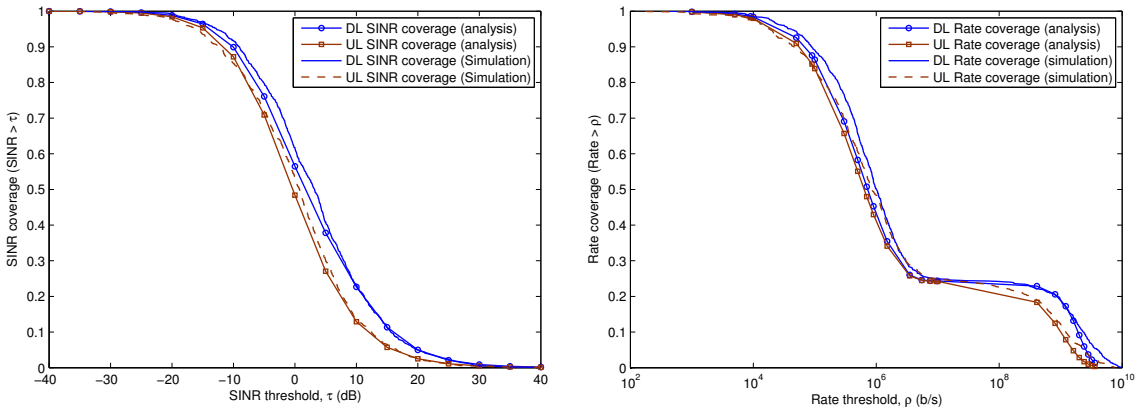
(a) (b)

Fig. 3.9 (a) Max-Rate Association probability analysis plot compared with simulation. (b) The difference between the UL/DL Scell association probability based on Max-Rate and Max-BRP.

we assume that the UL and DL cell associations are based on the DL biased received power where biasing is only assumed for Scells. The reason behind this is to clearly show the effect of small cell biasing on both UL and DL SINR and rate distributions based on the same association mechanism used in LTE systems where biasing is done jointly for UL and DL and is based on biasing the DL received power. This will help in drawing conclusions related to how currently deployed systems need to be changed and this setup will also be used to confirm our insights regarding the Max-Rate association as will be shown later on.

SINR and rate coverage analysis validation. Fig. 3.10 shows the SINR and rate distributions with no bias ($T_s = 0$ dB) where the derived SINR and rate analysis expressions are compared with simulation results. The figure shows that the analysis gives quite accurate results that match very well the simulation results, this allows us to use the analysis for further insights in the coming results. Furthermore, Fig. 3.10(b) has a flat area between 10^7 and 10^9(b/s) rate threshold, this area separates the sub-6GHz UEs below 10^7(b/s) from the mmWave UEs with very good channel above 10^9(b/s). This shows the substantial difference in rate that the larger bandwidth in mmWave could offer.

in addition to the previous figure, we compare the considered carrier frequency and bandwidth (70 GHz, 1 GHz) with a different setting (28 GHz, 200 MHz) in Fig. 3.11. The figure shows the accuracy of the derived rate coverage expressions for both carrier frequencies. It can be seen that even though the 70 GHz case has more bandwidth



(a) SINR coverage (b) Rate coverage

Fig. 3.10 SINR (a) and rate (b) distribution comparison from simulation and analysis.

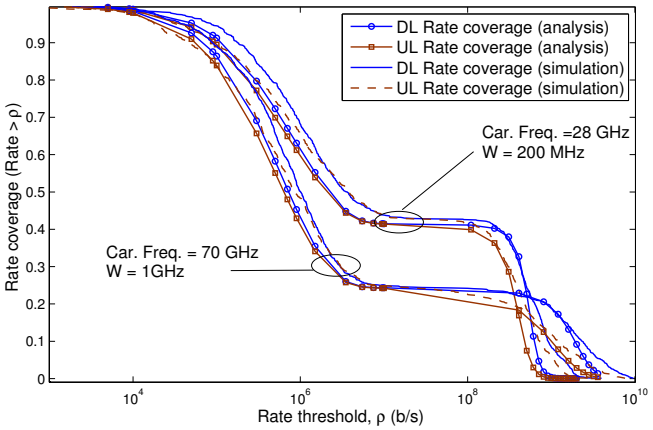


Fig. 3.11 Validation of rate coverage analysis with simulations for 70 GHz and 28 GHz carrier frequencies.

than the 28 GHz case, the 28 GHz case offers superior rate for almost 80% of the UEs, thanks to the larger coverage of Scells in the 28 GHz case which allows them to serve more UEs with higher SINR. This highlights the importance of biasing in extending the coverage of Scells.

(SINR ≈ SNR) validation. Fig. 3.12 shows simulation results for the CCDF of the mmWave UEs UL and DL SINR and SNR for two different mmWave Scells densities. It can be seen from the figure that for $\lambda_s = 30/\text{km}^2$ the SINR and SNR are almost overlapping and even at $\lambda_s = 200/\text{km}^2$ the difference between SINR and SNR is very small. This result confirms our assumption that interference in mmWave has a minimal impact on coverage for the mmWave Scells densities considered in our scenario. This,

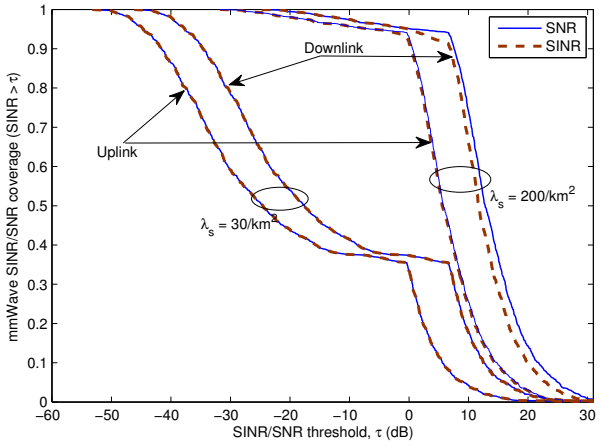
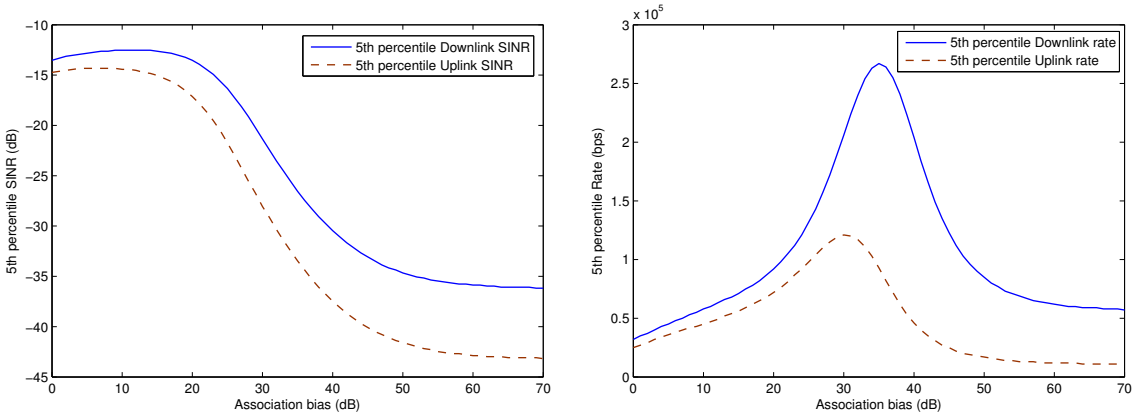


Fig. 3.12 Simulation results for the distribution of the mmWave SINR and SNR which validates the assumption (SINR $\approx$ SNR).

in turn, confirms that the SINR can be approximated by the SNR for mmWaves which is quite different than the trend in sub-6GHz networks where SINR $\approx$ SIR. Furthermore, the break point in the curves at 30% and 90% of the CCDF for λ_s of 30 and 200 shows how the SNR starts degrading quickly after a certain point which is a result of the LOS ball blockage model which assumes that beyond a certain distance between the UE and the BSs all the UEs are considered non line of sight. In addition, the degradation affects fewer UEs at $\lambda_s = 200/\text{km}^2$ since at a higher density fewer UEs are expected to be outside the LOS ball of the mmWave Scells.

Scell biasing effect on SINR and rate trends. Several previous studies have shown the importance of cell biasing in Hetnets [42, 53, 47]. However, the different propagation characteristics of sub-6GHz and mmWaves and the high imbalance in the available resources in both bands could result in different conclusions when it comes to biasing. Hence, the following results focus on the effect of biasing on the system's SINR and rate coverage and the optimal value of biasing for UL and DL.

Fig. 3.13 illustrates the UL and DL 5th percentile SINR τ_{95} and rate ρ_{95} against the Scell association bias (T_s) where the relation between the 5th percentile SINR and the SINR coverage is ($\mathcal{P}(\tau_{95}) = 0.95$) and the same for rate. In Fig. 3.13(a) the DL SINR increases slightly and then starts decreasing beyond 5 dB bias, the UL SINR behaves similarly. The slight increase in SINR at the beginning is due to the negligible interference in mmWave networks, therefore although the SNR is reduced the overall SINR is slightly increased. On the other hand, the 5th percentile rate in Fig. 3.13(b) is peaking at a bias of 30 and 35 dB for the UL and DL respectively. However, the



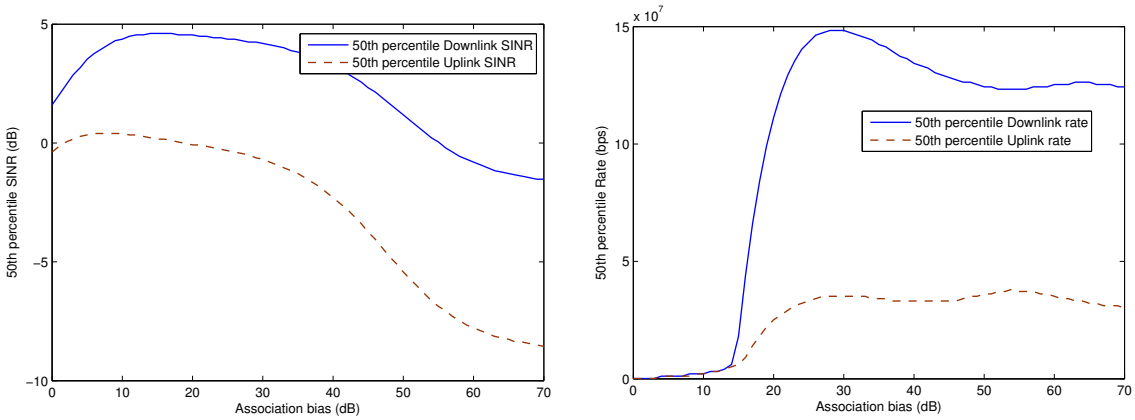
(a) 5th percentile SINR (b) 5th percentile rate

Fig. 3.13 5th percentile SINR (a) and rate (b) against the small cell bias value in dB.

corresponding SINR values with such large bias are around -30 dB, which is extremely low. The reason for the high rate despite the low SINR is obviously the much higher bandwidth at mmWaves. These bias values are over 100x the typical values seen in sub-6GHz scenarios in [42, 53, 47].

The design insight behind this result is that very robust modulation and coding schemes need to be considered for mmWave networks so that they can operate at very low SINR. Another insight is that the UL 5th percentile rate peaks at a lower bias value than the DL rate which means that a fraction of the UEs would tend to connect to the Scell and Mcell in the DL and UL respectively. This confirms the trend resulting from the Max-Rate association in Fig. 3.9(a) about the *reversed* decoupling behaviour since at the optimal bias value UEs are assumed to be connected to their rate optimal cell. This also confirms that the association probability in Lemma 2 (with $N_m = N_s = 1$) results in the same trend as the optimal rate results (considering the cell loads) in Fig. 3.13(b).

Fig. 3.14 illustrates the 50th percentile SINR and rate where a similar behaviour to Fig. 3.13 can be noticed. However the increase in the 50th percentile SINR in Fig. 3.14(a) is much higher than in the 5th percentile SINR, also the decline starts at a higher bias than the 5th percentile SINR, this is because the 50th percentile UEs typically are closer or have a better channel to their serving cells. Therefore, a degradation in their SINR would require a higher bias value. Looking at the rate in Fig. 3.14(b), it can be noticed that it peaks at around 30 dB bias for the UL and DL which corresponds to an SINR of 3 dB for the DL and -2 dB for the UL which is still considered low for the median UEs, therefore the need for robust modulation and coding still applies in



(a) 50th percentile SINR (b) 50th percentile rate

Fig. 3.14 50th percentile SINR (a) and rate (b) against the small cell bias value in dB.

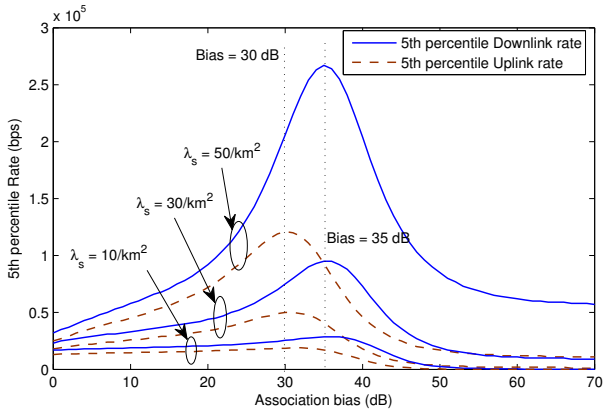


Fig. 3.15 UL and DL 5th percentile rate with variable mmWave Scell densities.

the 50th percentile UEs case. The fluctuation in the 50th percentile rate beyond 50 dB bias results from UEs moving from their serving mmWave Scells to less loaded mmWave Scells which results in a slight increase in the rate.

Impact on infrastructure density. In Fig. 3.15 the impact of the density of mmWave Scells on the 5th percentile UL and DL rate is illustrated. It can be observed that the optimal bias in terms of achieved rate is 30 and 35 dB for the UL and DL respectively and these values are the same for all densities. It has been shown in [42] that the optimal bias considering resource partitioning decreases with the increase in the Scells density because of the increased interference on the range expanded UEs. However, in our scenario it was already shown that mmWave operation is noise limited, therefore interference has a marginal effect which explains the invariance of the optimal bias with the Scell density.

3.3.8 System design implications

In this part we summarize the system design and deployment implications based on the results shown previously in this section:

- The association probability of mmWave Scells is dramatically improved as the Scell beamforming gain is taken into account during the association phase. This indicates the importance of having highly directional beams in the association phase.
- Decoupled access is still relevant in mmWave/sub-6GHz HetNets from a maximum received power as well as rate based associations. It was shown that DUDe is key for improved performance in both UL and DL whether the optimization criterion is received power or rate where decoupling occurs in different directions for the two criteria.
- Decoupled access is more relevant in less dense urban environments. This is reflected in Fig. 3.8 by the higher decoupling gain with a smaller α_m and α_n and a higher LOS ball radius (μ) and both features characterise low density urban scenarios.
- Aggressive values of mmWave Scell biasing can be beneficial in terms of rate as shown in Fig. 3.13(b). This would result in UEs having to operate in very low `SINR` which gives rise to a need for robust modulation and coding techniques that would allow the UEs to operate in these low `SINR` regimes to harvest the benefits of mmWaves.
- It was shown that from a rate perspective, UEs are more probable to connect to sub-6GHz Mcell in the UL and mmWave Scell in the DL. In addition, recent studies on electromagnetic field exposure [98] have shown that the maximum UL transmit power on frequencies above 6GHz will need to be several dB lower than sub-6GHz to be compliant with exposure limits. These trends could lead to allocating the UL on sub-6GHz Mcells and the DL on mmWave Scells as discussed in [39].

3.4 Summary

In this chapter, we have presented a rigorous and extensive analytical study of the decoupled uplink and downlink concept using tools from stochastic geometry. In the

first part, DUDe was studied in a sub-6GHz heterogeneous network where a co-channel deployment (Mcells and Scells are deployed on the same frequency) is assumed. Based on previously derived cell association probabilities, we characterized the capacity gains from the DUDe concept where these gains were validated using simulation results. We have also compared cell association trends for three completely different simulation scenarios. Namely, a deployment based on Vodafone's small cell test network, a Poisson point process based random deployment and a regular grid deployment. Surprisingly, the trends of these three deployments were matching very closely which lead to the following conjecture: the association probability depends chiefly on the density of the deployment, but not the process used to generate the deployment geometry.

In the second part, we proposed a detailed analytical framework for cell association in sub-6GHz-mmWave heterogeneous networks considering a decoupled uplink and downlink association. The analysis considered a maximum biased received power as well as a maximum achievable rate approach highlighting the main differences between them. The results show that there is a different trend in decoupling between the two approaches where in the rate based approach devices tend to connect in the UL to the sub-6GHz Mcells which is opposite to the decoupling trend in previous studies. The `SINR` and rate coverages are also derived where we put special emphasis on Scell biasing in the results showing that quite high Scell bias values are possible which has implications on the modulation and coding schemes in future networks. This work could be extended in numerous ways including the consideration of UL power control, indoor users and mobility in a mmWave scenario. Considering the cell load in the rate based association is an interesting extension as well. In addition, the inclusion of sub-6GHz small cells and allowing users to have multiple decoupled connections in the uplink and downlink to different base stations is quite interesting and will be left for future work.

Chapter 4

Architecture and system design of DUDe

In the previous chapters we have provided an extensive simulation and theoretical analysis of DUDe illustrating the substantial gains resulting from this technique. An important question is: Can DUDe be supported in today's 4G networks? If not, what are the changes that need to be done to introduce the support of DUDe in 4G and future 5G networks? This is a very crucial point to discuss as if it turns out that substantial changes in the architecture are needed, this could be prohibitive from a complexity or cost perspective and would limit the applicability of DUDe. In this chapter[1], we start by exploring the interoperability of DUDe with some of the most important emerging technologies. Then we tackle the architectural aspect by exploring the support of DUDe in 4G networks and what components are needed in 5G networks in order to efficiently support DUDe.

4.1 Interoperability of DUDe with emerging technologies and trends

The deployment of a reliable high speed spectral efficient network needs the inclusion of a variety of innovative features, provided that link level solutions have evolved to near Shannon limit capacity with advanced modulation and coding schemes (MCS). Given this, 4G and the forthcoming 5G, must offer pioneering solutions or improved versions of earlier releases in spectrum management and cooperative communications.

[1]The work in this chapter is partly included in [39].

In particular, the interoperability of DUDe with other radio access technologies can lead to an overall improvement of these features. The reduced interference variability, the enhanced network flexibility or the reduced transmit power are some of the advantages that help to make the most out of the radio management techniques.

4.1.1 Inter-band Carrier Aggregation

Carrier Aggregation (CA) provides great advantages when carried out in HetNets. In particular, the inclusion of CA in such context has been recognised as a feasible way of providing the multi-site radio resource allocation feature that was first introduced in Release 11 [110]. Also, CA in general (multi or same site aggregation) allows to improve capacity by extending the available bandwidth and support for mobility and interference management techniques.

Several studies have focused specifically on the implementation of CA in the UL context, where power capabilities of the UEs constitute the most limiting constraint [111, 112]. It is observed that the potential gains of CA transmissions are strongly related to the power demanded, which essentially depends on the bandwidth allocations and UE pathloss. Cell association determines which eNB is serving the UE, less power is needed to transmit data when the UE is closer to its serving eNB. To assure a correct performance of the transmission on the aggregated bands in the UL, it is crucial to account for the UE's maximum transmit power. In Mcell-only deployments, cell edge UEs are less likely to transmit in CA, however in heterogeneous deployments, the distance to the eNB is shorter given the higher cell density. In such a context, if UEs are associated based on the DL RSRP, the UL CA transmission is going to be highly restricted. Decoupling strategies are more lax in adopting aggregated transmissions; mainly, this is owed to the improvement in UL power availability brought by decoupled associations. This is important, since CA is intended to be applied in both UL and DL, and with traditional DL received power association rules, the applicability of CA in the UL is seriously conditioned by the lack of power availability.

4.1.2 Cooperative Multi-Point

Base station cooperation, in the form of Coordinated Multi-Point (CoMP) transmission or reception, has gained popularity in the context of HetNets as a means to increase the UE achievable throughput. eNBs within the same cluster communicate via backhaul links (i.e., via the X2 interface) with the objective of minimizing the inter-cell interference and capitalizing on the benefits of distributed antenna systems. In fact,

interference within a cooperation cluster can be effectively cancelled [113, 114]. This level of coordination and cooperation can be carried out in both UL and DL, and the realization of such coordination relies strongly in the availability of sufficient backhaul capacity, first to serve the UE in the cell cluster, and second to communicate with other cells in the cooperation cluster. This backhaul dependency can be very limiting in situations of high load, and in capacity limited links.

The increased flexibility provided by decoupled UL and DL associations provides advantages when selecting UL and DL coordinated transmissions or receptions. In particular, there is no need to have both UL and DL simultaneous connection to the entire cooperating set of base stations and the UE could have unequal UL and DL active links (as in the case of CA). This flexible association inside the cluster, and the interoperability of DUDe with CoMP goes one step further towards a device-centric network, since the UE can select independently the number and position of DL and UL serving cells, according to several input parameters such as backhaul capacity, power limitation, throughput maximization, among others.

4.1.3 Millimeter Waves

The ever increasing network traffic demands have lead to several trends in cellular networks including the densification of the network and the emergence of heterogeneous networks where small cells are playing a key role in providing capacity in hotspots and dense urban areas. Another approach was to explore new unused frequency bands that would help in satisfying the projected future traffic needs. Popular sub 6 GHz frequencies are becoming scarce and would no longer be able to cope with the increasing network traffic.

A solution to this problem is to resort to higher frequencies in the millimeter-wave bands where a significant amount of spectrum is underutilized or completely unused. The several GHz of available spectrum promise a spectacular increase in capacity which qualifies millimeter-wave technologies as one of the main enablers of future 5G networks.

In reality, millimeter-wave bands will not replace sub 6 GHz bands, at least initially, where sub 6 GHz will still be needed to provide coverage and ubiquitous service since millimeter-wave frequencies have poor propagation properties and are quite sensitive to blockages. The existence of millimeter-wave networks as an overlay to the conventional sub 6 GHz networks would require a change to the conventional cell association techniques. DUDe is expected to play a key role in millimeter-wave networks for several reasons.

Recent studies on electromagnetic field exposure [115] have shown that to be compliant with the applicable exposure limits at frequencies above 6 GHz, the maximum transmit power of devices would have to be reduced by several dBs below the conventional power levels in current cellular networks. This has significant implications on the UL since the UL coverage depends on the transmit power of devices and since mmWaves have unfavourable propagation properties a reduction in the transmit power of devices could result in serious degradations in the link quality in the UL. A possible solution to this problem would be to decouple the UL and DL where for some UEs the UL will be served by sub-6GHz macro cells with a better link budget whereas the DL would be served from the mmWave small cells. That is, while previously DUDe was discussed in the context of associating UEs to macro cells in the DL and Scells in the UL, for mmWaves the opposite strategy might prove useful. This aspect was extensively explored in the theoretical analysis of Chapter 3.

In addition, mmWave small cells are expected to have a very limited coverage area and consequently the conventional association techniques based on the DL received power will leave the mmWave small cells very much under-utilized considering the vast amounts of spectrum available in the millimeter-wave bands which could go beyond the available spectrum below 6 GHz. Small cell biasing or range extension is a technique that was introduced in LTE release 10 where effectively a cell selection bias is applied to small cells in order to expand the coverage of small cells and improve their offloading effect. A similar approach could be employed for millimeter-wave small cells to solve the aforementioned problem by using aggressive range extension values to attract as many UEs as possible in order to make use of the large spectrum chunks available at millimeter-wave frequencies.

DUDe would play an important role in this setup as it would allow to set different values of range extension for UL and DL in order to meet the requirements of both links. For instance, if the UL is required to be highly reliable whereas the DL can be less reliable but is more demanding in terms of throughput, a high value of DL Scell range extension can be used while not applying any range extension in the UL. Therefore DUDe would offer the flexibility to cater for the different needs for the UL and DL in millimeter-wave scenarios.

4.1.4 Different Duplexing Techniques

DUDe can function with both FDD and TDD, with different implications from a system level and spectrum point of view.

TDD allows much more flexibility in trading DL and UL resources as compared to FDD. With DUDe, it was demonstrated that fewer UL resources are needed to achieve the same UL rate as compared to coupled operation. This could lead the way to DUDe having a positive effect on the DL rate by allowing the DL to use more resources via dynamic TDD.

Another benefit of TDD is the possibility of estimating the DL channel via UL reference signals. This is quite important especially for channels with large dimensionality like in the case of massive MIMO. However, when DUDe is used this reciprocity is broken as DL and UL transmissions originate and terminate at different BSs. Much of the existing spectrum is paired FDD spectrum, therefore for these two reasons massive MIMO may need to be supported without channel reciprocity.

In the medium to long term, DUDe along with different emerging technology trends could lead to a rethinking of the traditional FDD/TDD paradigms. DUDe, hyper-densification, the use of millimeter wave frequencies and highly directional antennas, could allow for duplexing approaches over the spatial domain. For instance, the same band could be used for two different devices located in different locations one is receiving in the DL from a certain BS and the other is transmitting in the UL to another BS. Effectively, assuming a spatial UL/DL coordinated scheduling mechanism which would allow full-duplex like gains without the complicated interference cancellation mechanisms of full-duplex. In addition, once analog/digital interference-cancellation mechanisms are realistically realisable to support full temporal duplex, DUDe can be beneficial as it allows a generalized decoupled access that would allow the support of a DL and not necessarily the same UL user in the same frequency band.

4.2 The support of DUDe in 4G

4.2.1 Main existing techniques

DUDe can, depending on the deployment scenario and backhaul properties, already be supported by the existing LTE/LTE-A specifications. Illustrated in Fig. 4.1, three specific embodiments are discussed below.

Centralized Processing: in a deployment scenario with multiple radio units with a different cell-ID connected to a centralized node (like in the case of a Centralized Radio Access Network C-RAN), DUDe is possible in LTE-A without additional standardization support (see Fig.4.1a). The BS used for downlink transmission to a specific UE is selected using conventional means, typically based on downlink signal strength

measurements. Uplink transmissions are received by one, or if macro diversity is desirable, multiple radio units as the specifications do not mandate the reception node. Uplink decoding could either be performed at the radio unit (or at the set of radio units) or sampled analog data could be forwarded to the centralized unit via a Common Public Radio Interface (CPRI) interface for further processing. Uplink-related control signalling (including e.g. hybrid Automatic Repeat Request (ARQ) and power control commands) needs to be transmitted from the downlink node. In the same way, downlink-related control signalling from the terminal needs to be received by the uplink node and forwarded to the downlink node over the infrastructure.

Shared Cell-ID: An interesting extension of the approach described above is the so-called shared cell-ID approach [116] (see Fig. 4.1b), where radio units all belong to the same cell (i.e. have the same cell-ID). Here, Channel State Information (CSI) enhancements and quasi-co-location mechanisms introduced in Release 11 as part of the CoMP work are used to rapidly, independently and, from a terminal perspective, transparently switch transmission and reception points for a given terminal. This is a step away from the traditional cell-oriented paradigm towards viewing the antenna points as resources to be used in the best possible way to maximize performance. Furthermore, node association and mobility are handled via proprietary (non-standardized) solutions, transparent to the mobile terminal, and providing better mobility robustness in dense networks compared to methods relying on UE-centric measurements. Although conceptually straightforward, both centralized processing and shared-ID approaches require a fairly low-latency backhaul to meet the timing requirements (e.g. to send hybrid-HARQ messages). In a practical LTE-A rollout, the deployment is thus limited to remote radio units connected to a centralized baseband processing node.

Dual connectivity: While the two solutions described above require a very low-latency backhaul, usually achieved via connecting radio units to the same central unit, DUDe can also be implemented with a less ideal backhaul. Dual Connectivity, an extension first introduced in Release 12, allows for a terminal to be simultaneously connected to two cells and can be used for DUDe (Fig. 4.1c). We note that in Release 12, DUDe using dual connectivity is limited to inter-frequency deployments, i.e. to deployments where the two cells transmit over different frequency bands; nevertheless, later releases may add support for intra-frequency band deployments. The two cells operate separately, handling their own scheduling and control signaling (e.g. H-ARQ message) and thereby significantly relaxing the backhaul requirements compared to the centralized baseband approach and enabling the standardized X2 interface to be used

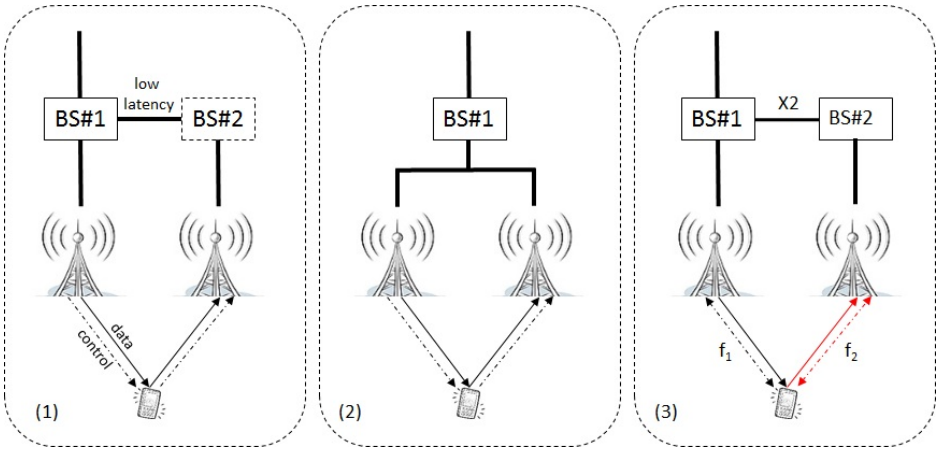


Fig. 4.1 The three discussed embodiments of DUDe are: (a) centralized processing unit; (b) shared cell-ID; and (c) the dual connectivity option.

for inter-BS communication. This solution has advantages and disadvantages. On one hand, a low latency backhaul connection for the signaling is not needed. On the other hand, mobility must be handled using standardized mechanisms and the possibilities for proprietary optimization are limited.

In the following subsection we provide a detailed architectural description of the above solutions with an emphasis on messaging flows and latencies in each option.

4.2.2 Architectural details of the available solutions

The access network configuration presents several challenges to enable the use of decoupled connections, while maintaining interference and energy efficiency at reasonable levels. Holding more than one UL connection is less power efficient for users that are placed near the cell edge, which are the ones more likely for decoupling [117, 118]; also, the use of one carrier to exclusively transmit control signals will lead to poor spectral efficiency. In particular, the most challenging part of devising architectures that support DUDe is that an increased amount of control information needs to be signalled back to the corresponding serving cell.

3GPP has proposed several architectural alternatives for DL dual connectivity and the architecture needed to support the U-Plane aggregation from different eNBs is expected to be very similar to those proposed for dual connectivity, based on the bearer split concept [10]. Alternatively, those architectures that enable a full UL and DL decoupling should support a feasible cooperation among both serving cells while not jeopardising the improvements in the UL in terms of reliability and capacity. To

achieve this it is necessary to assure the delivery of layer 1 and layer 2 control signals while maintaining the RAN latency requirements.

Network Procedures

When a UE accesses the system and associates with an eNB, a radio resource control connection is set up. The RRC is the protocol that handles all the control plane signalling of Layer 3 between the UE and the EUTRAN. Among the different tasks, the RRC is in charge of the connection configuration, radio bearer establishment and release, mobility procedures and configuration of power control mechanisms. The amount of resources the RRC consumes dynamically changes, depending on the cell traffic type, number of UEs being served and the connection timers, short timer values generate high RRC signalling overhead whereas high timer values result in poor use of idle mode [119].

Layer 1 and layer 2 control channels are crucial to support the user plane data transmission. PDCCH and PUCCH physical control channels are in charge of delivering information related to the active transmission, such as the scheduling decisions or the acknowledgements of the transmitted information, and the channel state information to perform accurate link adaptation. In particular, RAN control signals that are of paramount importance to handle scheduling and other MAC layer procedures are:

- Control information carried in the PUCCH: DL CQI, buffer status reports, scheduling requests and power headroom reports.
- Downlink control information (DCI) carried in the PDCCH in charge of indicating, among others, both UL and DL physical resource blocks for transmission (UL-SCH and DL-SCH), as well as link adaptation forms and transmit power for the UL.
- Downlink HARQ acknowledgment messages carried in the Physical Hybrid-ARQ Indicator Channel (PHICH).
- RRC messages that configure the UE connection and release, as well as the PUCCH position and resources and SRSs for UL channel state information configuration.

Among all these control information, the one most stringent in terms of latency requirement is the HARQ RTT, which is approximately 8 ms, considering information processing at both sides and frame transmission. The HARQ in the UL follows a

synchronous process: with a periodicity of 8 sub-frames the same HARQ id process is used, and no explicit information is exchanged about the process id. This means that there is a strict relation among the sub-frame number and the HARQ process id, which results in strong delays if one of the acknowledgements cannot be sent in the corresponding sub-frame. The architecture that enables the use of DUDe shall be the one that minimizes delay in the acknowledgment synchronization process and assures that the RAN RTT is maintained.

DUDe with Assisting Connections

The use of supporting UL and DL connections in order to transmit the control signals is an immediate solution to feasibly implement DUDe. The terminal is connected to both cells and aggregates the data flows; in such a way, all signalling layer 1/layer 2 and RRC can be handled locally. Architecture alternatives in this line where presented in [120].

A simple way to support this is with the use of CA, where each carrier component is configured separately to carry a shared and a control channel, this configuration is shown in Figure 4.2. This solution allows to keep the RAN latency at desired levels since no further delay is sensibly introduced. However, potential disadvantages of this configuration are: first, having two simultaneous UL connections can seriously jeopardise the UL performance in terms of UE energy efficiency. Second, this configuration does not maximize capacity over the available spectrum, since one component carrier is exclusively used to handle control information.

User Plane Bearer Split for DUDe

Latest releases of LTE-A (Release 12 and 13) consider new architecture alternatives for dual connectivity, with the introduction of the novel bearer split concept, an alternative that eases having two simultaneous transmissions in different eNBs, a Mcell and a Scell for instance [10]. On the other hand, the UL feasibility of adopting the bearer split can be argued in terms of power consumption, and UL data should be either transmitted directly to the Mcell, or forwarded to the Mcell by the Scell [117].

Based on this, in a DUDe context, assisting PUSCH or PUCCH connections may not be carried out in the Mcell; and PDSCH and PDCCH connections may not be carried out in the Scell. Having no control information being signaled back to the corresponding serving cell through the user interface (i.e., using supporting UL and DL connections), the delivery of layer 1 and layer 2 signaling and RRC relies on the the non ideal backhaul connection, the X2 interface, between both serving cells.

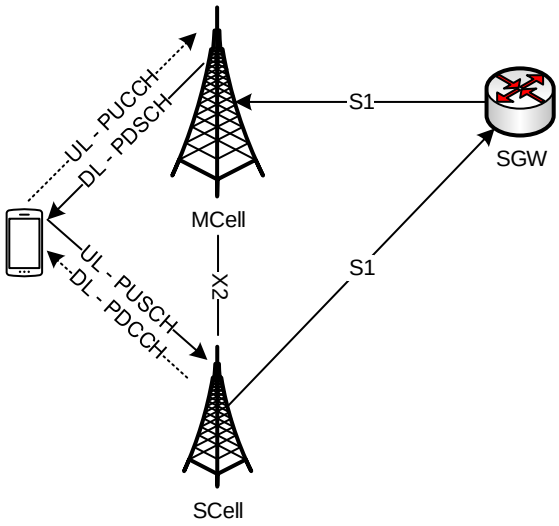


Fig. 4.2 Protocol stack information flow for radio bearer managed at SGW

Following the LTE architecture of distributed cells and using the bearer split concept two possible solutions arise for the decoupling:

1. The radio bearer is managed at the serving gateway (SGW), and the DL flows through the S1 from the SGW to the Mcell and the UL flows through the S1 to the SGW from Scell. This option has reduced flow control among both serving eNBs, since only part of the control signals, for example HARQ acknowledgements, needs to be forwarded through the X2. Figure 4.2 shows a diagram of the information flow for DL and UL.

2. The Radio bearer is managed at the Mcell, so the configuration is the master/slave MAC for UL and DL control feedback. This means that real-time MAC PDUs need to be forwarded to the corresponding eNB via the X2 backhaul interface, while respecting the 8 ms HARQ round trip time requirement. Master/slave configuration is for UL and DL, and the processing of each MAC PDU is done on the corresponding cell. Figure 4.3 shows a diagram of the information flow for DL and UL.

Both configurations allow to maximize the spectrum usage since all carriers can be configured for U-Plane information transmission. Current heterogeneous networks pose a big challenge to meet the RAN RTT fixed by the HARQ process, since delays that range from 5 ms to 30 ms are expected in the X2 interface [10].

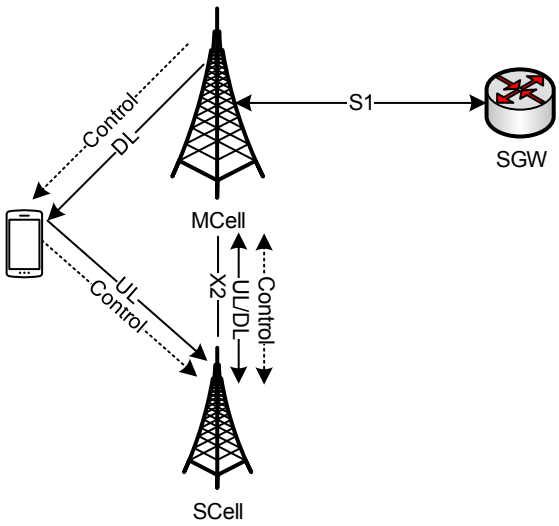


Fig. 4.3 Protocol stack information flow for radio bearer managed at Mcell

Similar to what it had been done in the user plane, the control plane that enables the full DUDe is analysed based on the RRC protocol architecture presented in [10] for dual connectivity. Two alternatives have been presented:

- based on prior negotiation of parameters between Scell and Mcell, the Mcell generates the final RRC message, this alternative is good as only one RRC connection is being active, but requires an assisting UL connection with the Mcell;
- in this case, the RRC message is built by the Scell based on input parameters provided by the Mcell, this solution adds complexity in the UE side as it needs to handle parallel RRC procedures [121].

In light of this, to completely decouple UL and DL poses a huge effort in the RRC connection.

Centralized solution

Previous architectural solutions present limitations in terms of spectrum usage, power efficiency and meeting latency requirements. If both eNBs, Mcell and Scell, are able to share the same BBU, then complete DUDe can be handled, and mitigates most of the drawbacks presented hitherto. The concept under the BBU sharing is the centralized

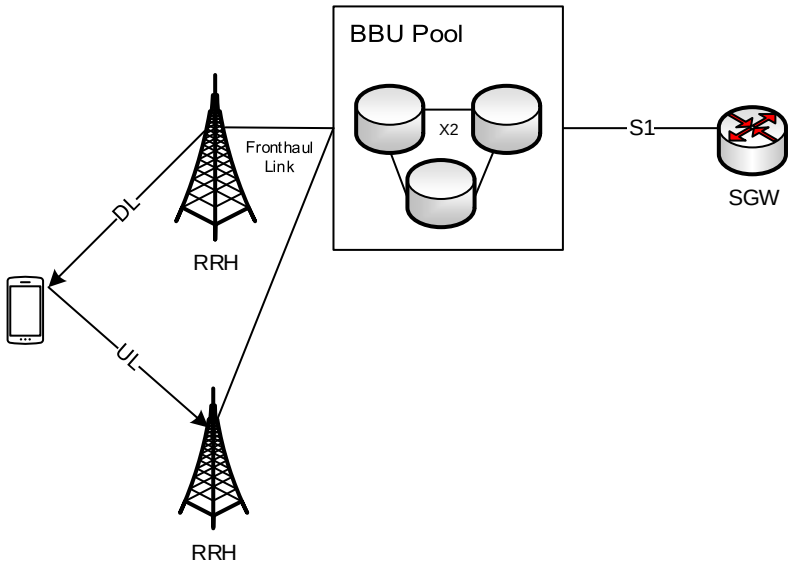


Fig. 4.4 C-RAN Architecture with DUDe

based architecture, also known as C-RAN, which breaks the static relationship between BBU and RRH, and each RRH does not necessarily needs to be mapped with one specific BBU. In particular, the radio signals from/to a RRH can be processed by a virtual eNB, which facilitates enhanced real-time cooperation among them. Given this, virtualization technology (network function virtualization, NFV) will maximize the flexibility in the RAN network allowing different levels of coordinated transmissions, or separated associations. Figure 4.4 shows the C-RAN architecture for DUDe.

The C-RAN architecture needs to use a new fronthaul interface that allows the communication among the RRH and the BBU, which can potentially increase the RAN latency. Different protocol functional split options can be recognised which variate the delays and capacity requirements over the fronthaul interface [122–124]. There is a benefit in locating part of the signal processing functions near the RRH, the bandwidth and latency requirements are brought to a level that can be fulfilled by cost-effective transport networks (i.e, dark fiber, wireless or other technologies) and at the same time enable the possibility of having enhanced cooperative radio resource management procedures. Realistic fronthaul delays can range from 250 μs to 4 ms [122].

When the DL serving RRH and the UL serving RRH share the same BBU, one RRC connection is held. To assume this situation the BBU has one single physical cell identifier (PCI).

4.3 The support of DUDe in 5G

The next few years will witness a surge in research and development on the different areas of 5G. The ITU and 3GPP are ramping up their activities on requirements and standardization of 5G where 5G is expected to be rolled out commercially by the early 2020's. In addition, there is an increasing consensus about the main requirements of 5G regarding capacity and latency as well as the key technical features of 5G, including ultra base station densification, massive MIMO, millimeter wave frequencies and possibly a "cell-less" architecture [125, 9].

With this view of 5G, does the new 5G standard need to include extra features to natively support DUDe?

An important question to answer is whether a simple evolution of current 3GPP architecture design would be able to efficiently support DUDe in future heterogeneous 5G deployments. In the previous subsections, a discussion was undertaken on how the LTE-A architecture already supports an implementation of DUDe when different BSs are linked via fiber to the same baseband unit, it was also discussed how the support of DUDe in 4G could be limited to different frequencies in the case of carrier aggregation or dual connectivity. Intuitively, any future 5G releases in 3GPP should allow for same-frequency dual connectivity, which despite having implications on resource and interference management is not considered to be a major upgrade.

Another issue to address is to ensure proper encryption of all data and control channels, specially when communication via the X2 interface is used between BSs. In LTE, each BS can support tens of IP Security (IPSec) tunnels. However, the management of security via IPsec is so problematic that operators tend to deploy only a few IPsec gateways per country where most of the gateways are deployed near the SGW. This means that traffic that logically goes via the X2 is actually routed via the SGW which adds a lot of delay that renders the DUDe operation rather inefficient. Whilst, LTE-A has more IPSec gateways deployed closer to the mobile edge, 5G architecture designs need to consider efficient encryption of the X2 traffic to reduce the extra latency in 4G.

In addition, some integration work is needed for some emerging techniques that have proved efficient for coupled systems. The integration of DUDe with decoupled Control/Data planes and License Assisted Access (LAA) will require some architecture changes. Self-organizing networking (SON) will also be instrumental in managing and coordinating the increased degrees of freedom introduced by DUDe.

Given, the above discussion, it can be concluded that a native support of DUDe does not require major design changes in 5G from an architectural perspective.

4.4 DUDe in higher layers

In this section, the main focus is on the transport layer and in particular on the TCP, given that is the prominently used transport layer protocol to achieve reliable end-to-end data transfer in IP based networks. Despite the fact that TCP has been initially designed for elastic applications it is currently commonly used in various popular streaming applications. It is worthwhile noting that Real Media and Windows Media, the two dominant streaming media applications, both are based on TCP streaming. In this section we briefly try to address the support available at the TCP level to enable DL/UL decoupling.

One of the challenges facing DUDe architecture is the possibility of having multiple TCP flows, that can handle the separate DL and UL connections. One of the major existing efforts in the Internet is the MultiPath TCP (MPTCP) [126]. MPTCP enables TCP to present a regular TCP interface to applications, while several IP interfaces are used simultaneously. In other words, data communication of a flow is spread over multiple *subflows*. The MPTCP connections begin regularly, and if there are extra paths available, additional TCP sessions, termed *subflows*, are created on these paths, and are combined with the existing session, which continues to appear as a single connection to the applications at both ends. Therefore MPTCP will allow for possibility of switching between coupled and decoupled DL/UL for different traffic flows. While MPTCP could be a great enabler for DUDe architecture, the challenge here is the availability of two different IP addresses. In other words, MPTCP can only be used to handle the DL and UL, if the UE is multihomed. Example of other efforts in running multiple and parallel TCP flows include MulTCP [127], which behaves as if it was a collection of multiple virtual TCP connections. Such initiative in the Internet domain are often for increasing data rate and better utilization of the end-to-end bandwidth.

4.5 Summary

In this chapter we tackled the architecture considerations of DUDe in 4G and 5G. We started by studying the interoperability of DUDe with several emerging techniques such as carrier aggregation, cooperative transmission, millimeter waves and different duplexing techniques. We then looked into how DUDe can be supported in existing LTE networks using centralized processing, the shared cell-ID concept and dual connectivity as main candidates and we delved into architectural details of each of these solutions. Subsequently, the support of DUDe in 5G was discussed highlighting the main limita-

tions of current 4G networks and reaching the conclusion that no fundamental changes to existing networks are needed for DUDe to be supported. Finally, we discussed the challenges facing DUDe in higher layers and mainly on the transport layer where we discussed different TCP techniques.

Chapter 5

Optimizing cell association and resource allocation in Device-to-Device communication

Until this point in the thesis, we have studied Downlink and Uplink Decoupling as one of the main constituents of the device centric architecture in 5G. We now shift our attention to another component in this 5G architecture vision which is Device-to-Device (D2D) communications [1]. D2D allows to exchange data through a direct link between two collocated UEs without the need to go through the network infrastructure in the conventional way. D2D fits very well in the device centric picture as the concept focuses more on relaying data between UEs in the most efficient way possible which is the core idea of our thesis. In the following sections we will study the cell association and resource management in a D2D enabled cellular network.

5.1 Overview of Device-to-Device communications

The exponentially increased demand for higher data rates and the emergence of data intensive applications along with the spectrum crunch have lead the research community to look for innovative paradigms to revolutionize the traditional cellular communications. D2D communications is one of these components and is the focus of this chapter.

D2D communications in cellular networks is defined as the direct communication between two mobile devices without the need for the data to be routed through the

[1] The work in this chapter was partly published in [128, 129].

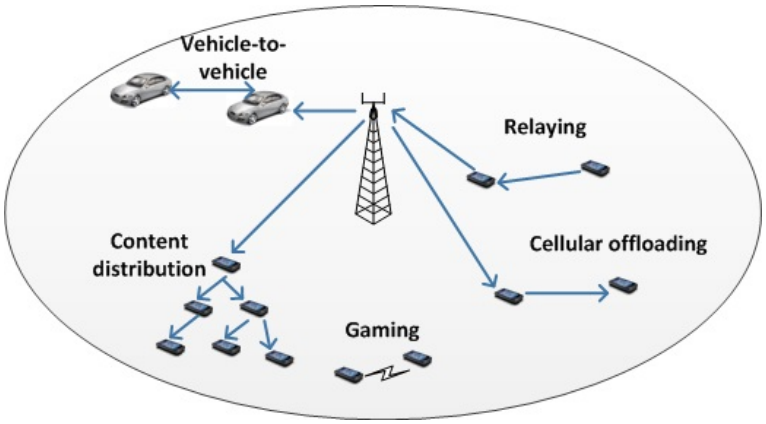


Fig. 5.1 Different D2D applications.

cellular infrastructure as it would normally need to in conventional cellular communications. D2D is not necessarily transparent to the cellular network and it can happen on cellular spectrum (inband) or in unlicensed spectrum (outband). In traditional cellular networks all communications must go through the cellular infrastructure even for UEs that are in proximity or collocated in the same geographical area. This traditional approach could be well suited for low data rate services such as voice and text messages where UEs are not close enough to have a direct communication. However, in today's networks UEs use high data rate services such as video sharing, gaming, proximity based social networking and many more where, if the UEs are close enough, the conventional way of communication is not efficient. Instead, these UEs could directly communicate in a D2D fashion which would result in a higher spectral efficiency and potentially improve throughput, energy efficiency and delay [130].

In academic research D2D was first introduced in [131] as a way to enable multi-hop relay communication in cellular networks. D2D was then extensively studied in the literature in terms of improving spectral efficiency [132], multi-casting [133], peer-to-peer [134], video streaming [135] and many more applications. The different D2D applications are summarized in Fig. 5.1.

In addition to academia, 3GPP has also investigated D2D communications for Proximity Services (ProSe). In particular, this use case in LTE has been studied in [136] and the required architecture requirements to support this use case was explored in [14]. D2D is limited in Release 12 to broadcast based sublic safety services but will be extended to cover more use cases in Release 13 and 14.

Fig. 5.2 illustrates the different modes in which D2D can operate in terms of spectrum. In Inband mode, D2D communication takes part in the licensed cellular spectrum. The motivation behind choosing inband communications is mostly related

to the fact that cellular spectrum is highly controlled in terms of interference and scheduling. Interference in unlicensed bands is not controlled which could impose some QoS challenges. Inband communication can be further divided into underlay and overlay communication. In underlay D2D communication, cellular and D2D communications share the same radio resources, i.e. that D2D UEs can reuse the already occupied cellular spectrum. This approach has the advantage of higher spectral efficiency since resources are reused instead of being blocked for a certain link. on the other hand, this approach results in higher interference between D2D and cellular UEs since UEs reusing the same resources are in the same cell, therefore tight interference coordination is required to reduce the interference effect. The second type of inband communications is overlay where D2D and cellular communications take part in dedicated non-overlapping resources. The advantage of this approach is the reduced interference effect between D2D and cellular UEs whereas the disadvantage lies in the low spectral efficiency as some resources are blocked for exclusive use by the D2D UEs.

The second type of spectrum usage shown in Fig. 5.2 is the outband D2D communication. The motivation behind using outband is to eliminate the interference issue between cellular and D2D communication. Using unlicensed spectrum usually requires different interfaces or wireless technologies such as WiFi Direct [137], ZigBee [138] or bluetooth [139]. In addition, the use of unlicensed spectrum in outband mode results in the D2D communication taking part in an uncontrollable environment which could be a problem in terms of QoS and reliability.

In this chapter we focus on inband D2D communication. In Sec. 5.2 we study D2D cell association in overlay mode where we use the decoupled uplink and downlink feature as an extra degree of freedom. In Sec. 5.3 we solve the resource allocation problem for D2D and cellular communication links in underlay mode using Bio-inspired Genetic Algorithms (GAs).

5.2 Decoupled cell association in D2D

5.2.1 Introduction

The ever increasing cellular network traffic has led to a shift from single-tier homogeneous networks to multi-tier heterogeneous networks (HetNets) in an attempt to increase the network capacity in hotspots in an efficient and scalable way. The HetNet solution helps in improving the capacity of cellular networks and bringing the network closer to the user equipments (UEs). Device-to-Device (D2D) communication

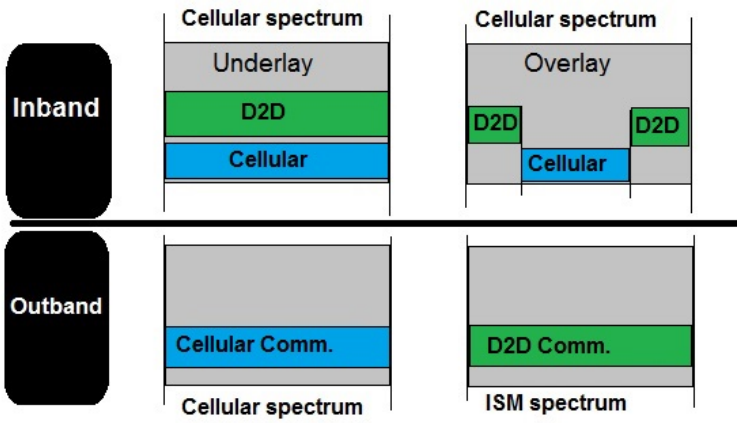


Fig. 5.2 Different D2D spectrum allocations.

introduces similar benefits arising from the proximity of UEs to each others that is exploited by enabling direct communication between UEs without the need for the data to be routed via the fixed infrastructure network [140].

Until the fourth generation of cellular networks, cell association has been based on the downlink (DL) received signal power only. It was shown in [141] that associating both uplink (UL) and DL based on the DL received power in a HetNet is highly suboptimal and that the decoupling of both UL and DL results in substantial gains in the UL. D2D UEs are expected to have cellular and direct D2D communications in subsequent time instants or subframes. Therefore, D2D cell association needs to take the nature of cellular transmission into account. As per 3GPP [142], D2D communication will take place in the UL licensed band which makes the decoupled association strategy very relevant to the D2D cell association problem. To the best of our knowledge, cell association has been extensively studied in macro-cellular systems, but only recently in heterogeneous networks [143]. However, D2D-aware cell association is still an open issue for research and needs to be well investigated [144].

The aim of this proposal is to study the different cell association (CAS) strategies for inband D2D communications in a heterogeneous network. D2D technology is expected to yield numerous overall benefits that mainly arise from the proximity gain they offer. Therefore, meticulous enhancements need to be included that will make full use of its merits. We focus on the inband overlay communication where D2D and cellular communications take part in the licensed frequency band and there is no overlap in resource block (RB) usage between D2D and cellular communication. The contribution of this work is the optimization of D2D-based cell association using the decoupled UL and DL association concept developed in prior art [33]. To this end,

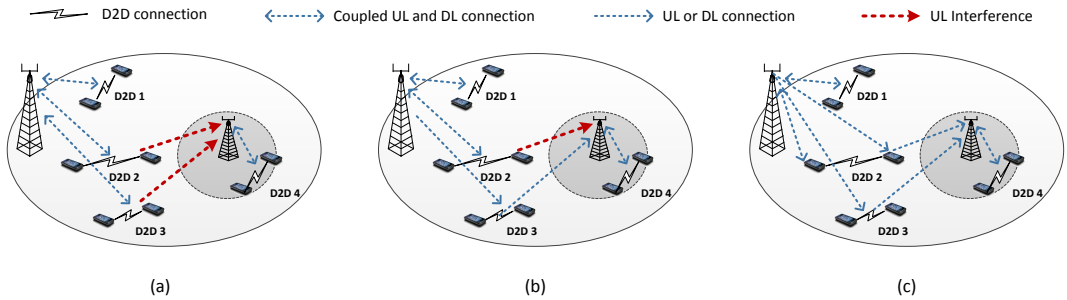


Fig. 5.3 Considered cell association scenarios: (a) Joint-Coupled, (b) Joint-Decoupled, (c) Disjoint-Decoupled.

integer linear programming (ILP) optimization tools are applied to introduce efficient D2D cell association that aims at minimizing the interference caused by D2D devices onto cellular communications as well as improve the efficiency of D2D resources usage.

Considering the D2D communication paradigm, the interference management among D2D and cellular transmissions in inband is very challenging. Furthermore, commonly applied power control and interference management solutions within the literature usually resort to high complexity resource allocation methods, as stated in [130]. D2D and decoupled uplink and downlink have been identified as main building blocks of future 5G networks in [9]. To this end, effective interference limitation with respect to resource utilization needs to be devised in order to improve the overall network welfare.

5.2.2 Problem Description

The principle aim of this section is to firstly give a glimpse of the investigated cell association techniques and then dive into detailed analysis of their realization. Abiding by the milestones of LTE Release 12 and the prospective integration of D2D notion as a technological component to current and emerging networks, we present a number of general design assumptions according to up-to-date standardization working documents [142]:

- D2D connections will utilize the uplink (UL) resources.
- The inband scenario of D2D communications is taken into consideration, where the interference from D2D devices onto cellular links in a neighbouring cell (either Macro-eNB (MeNB) or Small-eNB (SeNB)) could be substantial.
- The transmit power of D2D devices will be controlled by the serving cell (MeNB or SeNB) based on fractional path-loss compensation power control [145] as done

with cellular users. Formally speaking, the transmit power of a D2D device u associated with BS b is given by

$$P_t^{lu} = min\{P_{Max}, 10\log_{10}(M) + P_0 + \alpha L_{blu}\}, \tag{5.1}$$

where P_{Max} is the maximum transmit power of the device, M is the number of physical resource blocks (PRB) assigned to the device, P_0 is a normalized power value (in dB), α is the pathloss compensation factor and L_{blu} is the pathloss between the device u of link l and its serving cell b.

As mentioned in the previous section, D2D-based cell association algorithms need to consider the fact that D2D devices can have subsequent cellular and direct D2D transmissions in adjacent subframes to satisfy their communication needs. According to the current trend, a UE is primarily connected to a BS that provides the highest DL received power. However, the decoupled UL and DL association proposed in [33] has shown substantial gains by allowing users to choose different cells in the UL and DL considering a cellular heterogeneous network.

This idea constitutes the basis of this paper where it will be shown that the same concept is applicable to D2D-centric association as well. For the different cell association techniques that will be analysed, interference is the main validation criterion as it results from the ongoing cells' densification [146].

In the upcoming subsections, we provide an ILP optimization framework based on the different association policies by taking into account the notion of decoupling and the ability of the devices of a D2D pair to connect to different serving cells. Without loss of generality, unicast D2D connections are assumed. We will further compare these strategies in terms of transmit power efficiency, resource utilization and interference metrics. The considered cases are listed below.

- **Joint-Coupled (JC)**: The baseline case where devices of the same D2D pair are only allowed to connect to one cell (Joint). Furthermore, the D2D devices have the same UL and DL serving cell based on the DL received power (Coupled).
- **Joint-Decoupled (JD)**: The devices of a D2D pair connect to the same serving cell but the UL and DL associations are decoupled where the UL serving cell is the one that minimizes the UL interference to cellular communication.
- **Disjoint-Decoupled (DD)**: The devices of a D2D pair are allowed to connect to different serving cells with the same association technique as the Joint-Decoupled case.

- **Hybrid-Decoupled (HD)**: In this case we combine both the Joint-Decoupled and the Disjoint-Decoupled cases to strike a balance between minimizing the interference and the resource usage.

Considering the three last cases, we allow D2D UEs to be associated with different serving cells in the UL (decoupled access) based on the minimum UL interference metric. For the rest of this subsection we will focus on the UL association optimization for the involved D2D UEs. To this end, before we detail the applied optimization framework, we need to define the set of deployed BSs as B (including both MeNBs and SeNBs), the set of randomly distributed D2D links $\mathcal{L}$, and lastly, U is the set of UEs that constitute these links.

Joint-Coupled CAS

In this scheme we assume that D2D UEs that constitute a link are associated with the same BS according to DL received power estimations. This approach is the baseline method as it is the technique used in LTE. However, the interference exerted by the D2D UEs that follow this association method can cause harmful effects on the cellular links, as clearly illustrated in Fig. 5.3a. In this figure, *D2D 2* and *D2D 3*, both associated (coupled) with the related MeNB can severely interfere with the proximate SeNB UEs active transmissions.

Joint-Decoupled CAS

This scheme is realized by associating the D2D links in the UL with the BS minimizing the link's UL interference. Fig. 5.3b represents this case. In this scenario *D2D 3* is served in the UL by the SeNB which results in the reduction of the transmit power of *D2D 3* as the couple is closer to the SeNB. However due to the joint association constraint, *D2D 2* is still associated to the MeNB.

Applied in the authors' prior work [144], we extend the cell association optimization logic for D2D links, where the paired devices are both connected to the same serving BS [147]. For this reason, we define the following binary decision variable

$$y_{bl} = \begin{cases} 1, & \text{if D2D link } l \text{ is associated with BS } b \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

Further, in order to view the problem of interference minimization caused by D2D UEs' potential transmissions, we need to define as $I_{bl} = mean\{I_{blu_1}, I_{blu_2}\}$ the average

of the maximum interference generated by the two paired devices (u_1 and u_2 of link l) which are both associated with BS b. The corresponding interference term for a D2D device u of link l is given by $I_{blu} = max(P_t^{lu}\mathbf{G}_{B'lu})$ where $\mathbf{G}_{B'lu}$ is the matrix of link gains between user u and all BSs that belong to the set $B' = B - b$. Herein, P_t^{lu} accounts for the transmission power of the UE u of link l according to (5.1) and depends on its associated BS.

The interference-based optimization problem can be then formulated as follows

$$\min \sum_{b\in B}\sum_{l\in\mathcal{L}} I_{bl}y_{bl} \tag{5.3}$$

$$\text{s.t.} \sum_{b\in B} y_{bl} = 1,\ \forall l \in \mathcal{L} \tag{5.3a}$$

$$\sum_{l\in\mathcal{L}} y_{bl} \leq K_b,\ \forall b \in B \tag{5.3b}$$

$$y_{bl} \in \{0,1\},\ \forall b \in B, l \in \mathcal{L} \tag{5.3c}$$

where constraint (5.3a) requires the sole association of a D2D link l to BS b, and (5.3b) provides an upper bound of the number of user links that can be associated with every BS b. The difference of this scheme compared to the Joint-Coupled baseline strategy is the decoupling of DL and UL for the D2D links located in the topology. Intuitively, but as also proven in the sequel, this method is very efficient in terms of resource utilization by blocking (utilizing) one RB only from its associated BS that controls the D2D transmission. On the other hand, this method lacks intelligence in terms of interference controllability as it associates both devices of a D2D link to one BS without giving the flexibility for separate association of the nodes that could be less harmful.

Disjoint-Decoupled CAS

In this decoupled D2D scenario the paired devices can be also connected to different serving cells as shown in Fig. 5.3c. We anticipate that, in terms of interference, this is a very efficient strategy as every device connects to its closest serving BS. However, this scheme is not efficient in terms of resource usage, simply because if the devices of a D2D pair are connected to two different BSs, the resources used by these devices have to be allocated (blocked) for the D2D connection in both cells as opposed to the case when both devices are served by the same BS where the resources will be allocated (blocked) only in one cell. Therefore this scheme is interference optimal

but it uses twice as much resources as the Joint schemes. To this end, we provide an optimization setting that aims to minimize the introduced interference caused by the D2D transmissions.

First, we consider the following binary decision variable that indicates each UE's association with a BS

$$y_{blu} = \begin{cases} 1, & \text{if user } u \text{ of link } l \text{ associates with BS } b \\ 0, & \text{otherwise.} \end{cases} \tag{5.4}$$

where $b \in B$, $l \in \mathcal{L}$, and $u \in U$. Additionally, each link l constitutes a direct link between two proximate devices (i.e. D2D devices u_1 and u_2) that, as already mentioned, can be either both associated with a serving BS [144] or disjointly (loosely) connected with two separate BSs.

Therefore, the interference minimization problem for the disjoint decoupled D2D cell association can be set as follows

$$\min \sum_{b\in B}\sum_{l\in\mathcal{L}}\sum_{u\in U} I_{blu}y_{blu} \tag{5.5}$$

$$\text{s.t.} \sum_{b\in B} y_{blu} = 1,\ \forall l \in \mathcal{L},\ u \in U \tag{5.5a}$$

$$\sum_{l\in\mathcal{L}}\sum_{u\in U} y_{blu} \leq K_b,\ \forall b \in B \tag{5.5b}$$

$$y_{blu} \in \{0,1\},\ \forall b \in B, l \in \mathcal{L}, u \in U \tag{5.5c}$$

where I_{blu} is the maximum interference generated by user device u of D2D link l if associated with BS b; its power part is again estimated according to (5.1).

Hybrid-Decoupled CAS

In this case, we propose an interference-aware optimization problem with an objective to achieve resource usage efficiency. An effective and controllable resource utilization on top of an interference-aware method may well entail in balanced interference mitigation and resource efficiency impact. The Disjoint-Decoupled approach might be optimal in terms of interference but it is not efficient in terms of resource usage. On the other hand, the Joint-Decoupled approach is optimal in the sense of resource usage but lacks of satisfactory interference performance compared to the two methods mentioned above.

Hence, the Hybrid-Decoupled problem tries to strike the balance between interference and resource utilization.

In order to realize this hybrid problem, an additional decision variable needs to be defined that will act as an indication of joint association for two devices that construct a D2D pair. This can be written as follows

$$z_{bl} = \begin{cases} 1, & \text{if link } l \text{ associates with BS } b \\ 0, & \text{otherwise.} \end{cases} \tag{5.6}$$

Therefore, we propose a resource usage optimization problem that considers interference and formulate it as follows

$$\max \sum_{b \in B} \sum_{l \in \mathcal{L}} z_{bl} \tag{5.7}$$

$$\text{s.t.} \sum_{b \in B} y_{blu} = 1, \ \forall l \in \mathcal{L}, \ u \in U \tag{5.7a}$$

$$\sum_{l \in \mathcal{L}} \sum_{u \in U} y_{blu} \leq K_b, \ \forall b \in B \tag{5.7b}$$

$$\sum_{u \in U} I_{blu} z_{bl} \leq I_{th}, \ \forall b \in B, l \in \mathcal{L} \tag{5.7c}$$

$$2 z_{bl} \leq \sum_{u} y_{blu}, \ \forall b \in B, l \in \mathcal{L} \tag{5.7d}$$

$$\sum_{b \in B} z_{bl} \leq 1, \ \forall l \in \mathcal{L} \tag{5.7e}$$

$$y_{blu}, z_{bl} \in \{0, 1\}, \ \forall b \in B, l \in \mathcal{L}, u \in U \tag{5.7f}$$

As shown, the main objective is the maximization of the number of joint connections for the distributed D2D paired devices with respect to interference. Constraints (5.7a) and (5.7b) are defined as in problem (5.5). In (5.7c), a threshold that constrains the levels of interference if the devices of a link are jointly connected to a BS is added. This threshold can act as a weighting factor to decide if the focus of the algorithm should be interference (low I_{th}) or resource efficiency (high I_{th}). For this constraint, we limit the search to the n closest BSs to reduce the search space and consequently the complexity and size of the inequality matrix. Furthermore, constraint (5.7d) indicates that only if both devices of a link l will be associated with the same BS b, the value of z_{bl} variable equals to one (joint case). Lastly, (5.7e) stands for the restriction that each link's users can be associated with only one BS in the case of joint connection ($z_{bl} = 1$). Differently, they are disjointly connected to two separate BSs ($z_{bl} = 0$).

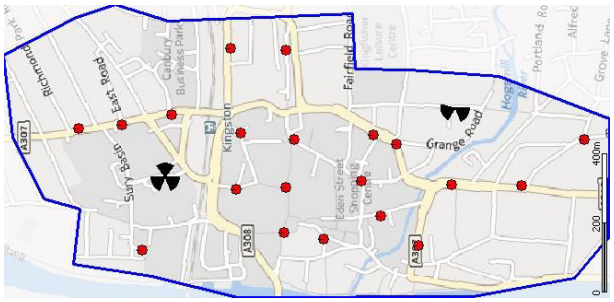


Fig. 5.4 Vodafone Small cell LTE test network.

5.2.3 Simulation setup

As deployment setup, we use the Vodafone LTE small cell test bed network deployment shown in Fig. 5.4. The test network covers an area of approximately one square kilometre and includes two Macro sites and 21 SeNBs represented by the black shapes and red dots respectively. We use this existing test bed to simulate a relatively dense HetNet scenario. The propagation model is based on a high resolution 3D ray tracing pathloss prediction model. This model takes into account clutter, terrain and building data and it guarantees a realistic and accurate propagation model. The user distribution is based on real traffic data extracted from the live network. We assume an inband operation of D2D where D2D UEs use the same UL frequency band assigned for cellular transmission. However, D2D and cellular UEs are scheduled on different resources which is termed as 'overlay' operation in the literature. The results are based on Monte Carlo simulations where the results are averaged over 100 simulation runs.

The operating frequency is 2.6 GHz. The maximum transmit powers of Mcells, Scells and UEs are 46, 30 and 23 dBm respectively. The fractional pathloss compensation power control algorithm in (5.1) is assumed with $P_0 = -90$ dBm and $\alpha = 0.8$. An average number of links of 336 is considered. I_{th} is set to -130 dB. The next section features a set of results evaluating the proposed cell association methodologies proposed in Section 5.2.2. Finally, we assume that each D2D pair is allocated one resource block (RB) per base station.

5.2.4 Simulation results

In this section, a set of numerical evaluations is presented to investigate the proposed optimization schemes. Fig. 5.5 shows the mean UL interference exerted by the D2D UEs onto cellular UL transmission against the D2D link length. The interference values are normalized relative to the DD case to show the different interference levels

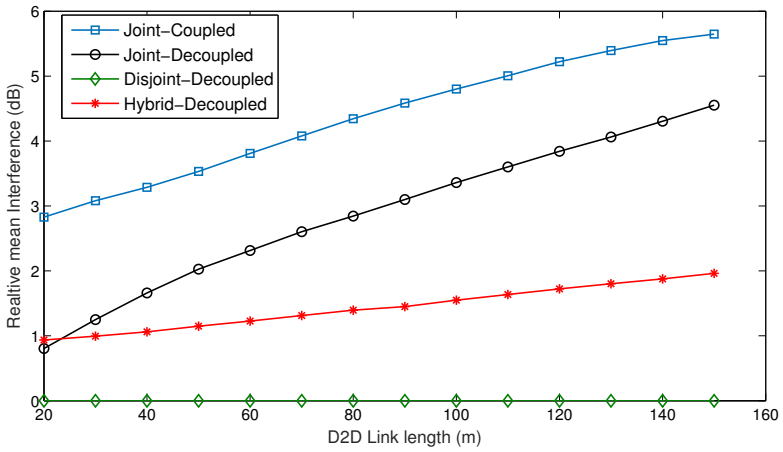


Fig. 5.5 Mean UL interference from D2D devices onto cellular transmissions.

compared to this interference optimal scheme. The JC and JD schemes show an increasing interference trend with the link length where the interference levels are around 3 dB (twice) and almost 6 dB (4 times) more than the DD scheme at 100 m and 150 m link length respectively. This is logical as the more the link length increases the more suboptimal the joint association schemes are as forcing distant devices to connect to the same BS results in a higher transmit power of these devices and a higher interference to neighbouring cells. The HD scheme introduces a trade-off between the Joint and DD schemes as it maintains an almost constant interference level that is around 1 dB higher than the DD scheme. This is due to the intelligence in the HD scheme that allows it to jointly/dis-jointly allocate D2D pairs depending on the interference level.

As explained earlier, if a D2D pair is served by one BS it is assumed to use only one RB over the whole network as this resource is reserved for this D2D pair in this BS only. However, if the devices of a pair are associated to different BSs then it is assumed that this pair is using two RBs over all the network since one RB has to be allocated for that pair in both BSs. Fig. 5.6 illustrates the average D2D resource usage per BS against the link length. The figure shows a constant resource usage for the JC and JD association schemes. This trend can be explained by the fact that the D2D pairs are jointly associated to the same BS regardless of the link length. Hence each D2D link uses 1 RB independent of the link length. However, the DD scheme shows an increasing RB usage with the link length. This can be explained by the fact that the probability of disjoint association increases with the link length and so as the D2D resource usage in the whole network since the disjoint D2D link uses twice

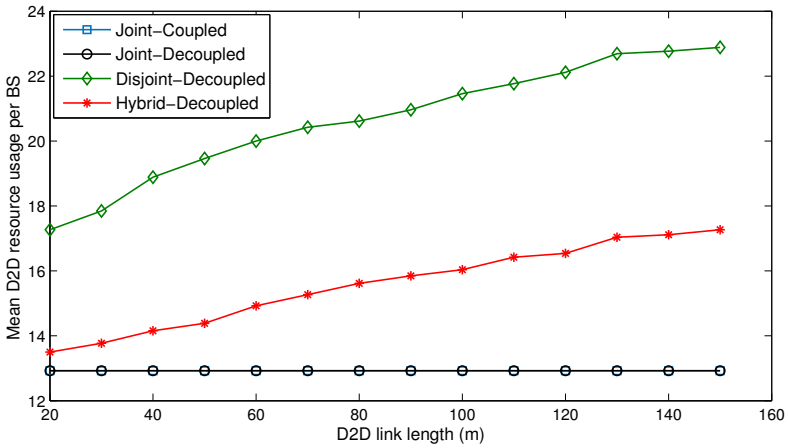


Fig. 5.6 Mean resource utilization for D2D per base station.

as much RBs as the joint one. The HD scheme -again- offers a compromise between the joint and disjoint schemes as the main scope of the optimization problem is to improve the resource usage efficiency with a constraint on the interference. The HD method achieves a reduction of resource usage of about 45% at 150 m link length compared to DD. Thus, it can be noted that the HD scheme offers a trade-off between the UL interference and resource efficiency which can be controlled by setting the I_{th} accordingly.

The cumulative distribution function (CDF) of the D2D UEs transmit power is shown in Fig. 5.7. The figure shows that the JC and DD schemes have the highest and lowest transmit power distributions with a difference of more than 5 dB at 50% of the CDF which increases the higher the transmit power is. The HD has a distribution that fits mid-way between the JC and JD distribution and that gets closer to the DD the higher the transmit power is. This shows that the HD scheme can result in a reduction of transmit power that varies between 3-5 dBs which is deemed crucial for battery powered devices.

In the next section we look at the resource management problem in D2D and try to solve it using Bio-inspired genetic algorithms.

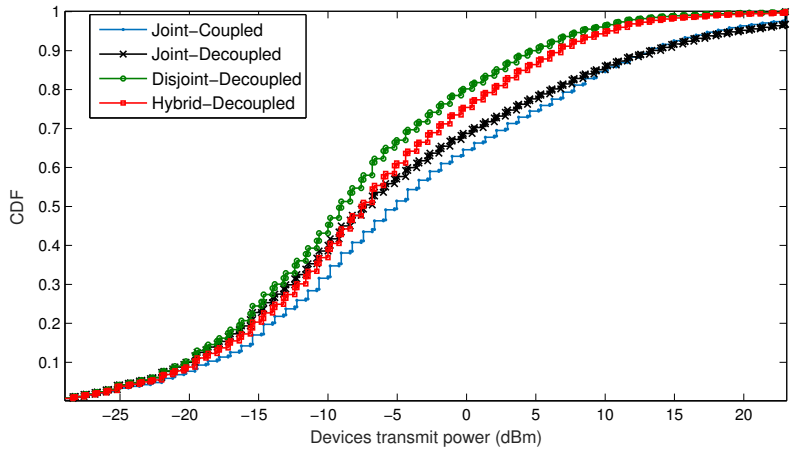


Fig. 5.7 CDF of the devices' transmit power.

5.3 Resource allocation in D2D using genetic algorithms

5.3.1 Introduction

D2D communication has proven to be an attractive solution for the surging cellular traffic increase and scarcity of spectrum in cellular networks by taking advantage of the proximity of UEs to each other. D2D enables the establishment a direct communication link between UEs in proximity without the need to route the data in the conventional way through the network infrastructure [140]. D2D communication is considered to be an enabler for a large number of proximity based applications such as public safety, peer-to-peer communication, local advertisement, multi-player gaming and many more.

D2D communication as an underlay in cellular networks allows the reuse of the spectrum assigned for cellular communications. D2D also allows the offloading of cellular traffic and enables more reliable and high throughput links between users in close proximity. For this reason, and following the prediction for network densification in future 5G networks, D2D is expected to play a crucial role in spectrum and resource management since in several cases the number of D2D connections can be very high and the resources would need to be carefully managed. However, some challenges that D2D communication faces need to be addressed in order to get the full benefit of this technology. initially, the potential D2D UEs may not be in close proximity which may render the establishment of a reliable connection between the D2D UEs challenging. In addition, the high spectral efficiency of underlay operation comes at the price of high

levels of interference to and from cellular UEs (CUEs) which could affect the quality of service (QoS) of D2D as well as cellular transmission.

Related work

Bio-inspired genetic algorithms (GAs) [148] have emerged as a popular approach in solving resource allocation problems in wireless networks [149–151] owing to their versatility, scalability and computational simplicity which make GA a very attractive method to solve the resource allocation problem as will be shown in Section 5.3.4. Resource allocation for D2D communications has been a hot topic in recent years. In [152], a proportionally fair utility maximization approach is used to allocate resources to both D2D UEs (DUEs) and cellular UEs (CUEs). In [153] the mode selection and resource allocation problems for underlay D2D communication are investigated and solved using particle swarm optimization. Further, an efficient graph-theoretical approach is proposed in [154] to perform channel allocation for DUEs. Resource allocation in relay-aided D2D scenario has been studied in [155].

Contribution

In this section[2], we study the joint resource allocation for cellular and relay-aided underlay D2D communications where DUEs share the UL resources with CUEs. We consider that a UE could act as a relay node in order to enhance the link quality between DUEs that are relatively far apart or the channel quality between them is poor. All DUEs have the choice to either communicate directly with their peer or via a relay. Usually, relays are used to enhance network coverage where needed. However, in order to offload the traffic that should traditionally be routed via the BS, a relay can also become the intermediate node that assists two UEs to communicate, without adding extra burden on the BS side. In this case, our proposal considers the use of GAs in order to find a near-optimal allocation of resources for CUEs and DUEs that can maximize the sum-rate. We compare the GA performance with a heuristic algorithm that prioritizes the D2D resource allocation as well as with a random allocation scheme. Differently from [155] which considered that all traffic flows are routed through L3 standard relays, in our study the choice of direct or relayed D2D communication based on the achievable rate is part of the optimization problem.

[2]The work in this section is included in [129].

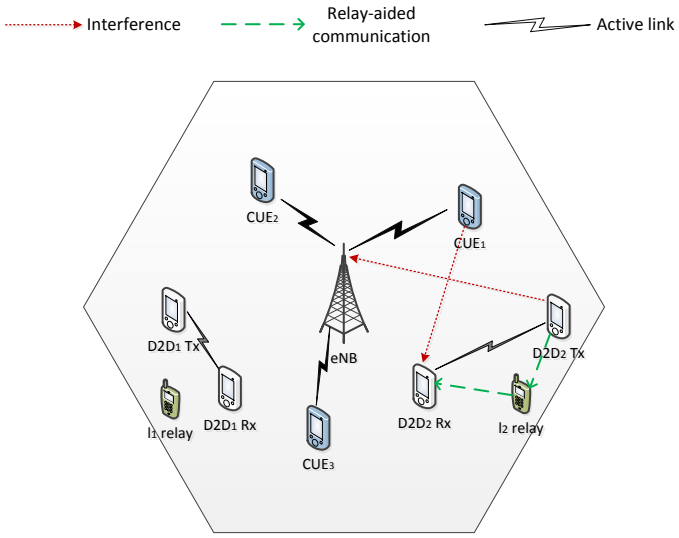


Fig. 5.8 Uplink scenario of relay-aided D2D communications as an underlay to the cellular network.

This study combines the optimization of the mode selection between direct and relayed D2D operation with the aim of achieving a joint resource allocation of cellular and D2D communication maximizing the aggregate throughput.

5.3.2 System model

The resource allocation problem in cellular networks is a widely studied area that falls within the nature of NP-hard problems which cannot be solved in real-time. Well known integer relaxation methods have been used to reduce the time complexity of resource allocation problems but do not render it a real-time solution for network operators. In this section we define important preliminary notations and parameters that will help us further formulate the relay-aided D2D/cellular resource allocation optimization setting and subsequently introduce for our proposal.

First, we consider the uplink case scenario of D2D underlaying a cellular network where interference from a CUE to a receiving DUE and from the transmitting DUEs to the CUE UL transmission occurs, as shown in Fig. 5.8. In this figure, interference exerted from the cellular user CUE_1 towards the $D2D_2$ pair and vice versa might be harmful not only for the reliability of the link, but also for the aggregate network throughput. Therefore, these two transmissions should occupy different Resource Blocks (RBs) to avoid mutual interference. In addition, we assume that CUEs are directly transmitting to the serving BS, whereas the communication mode between two

DUEs can be either direct or through a closely located relay. An important assumption is that only one proximate UE to a D2D pair can be used as relay. Now, before detailing the problem formulation, we need to define the following sets:

- $\mathcal{N} = \{1, 2, \ldots, N\}$: set of available RBs.
- $\mathcal{D} = \{1, 2, \ldots, D\}$: set of D2D links.
- $\mathcal{C} = \{1, 2, \ldots, C\}$: set of cellular links.
- $\mathcal{L} = \{1, 2, \ldots, L\}$: set of relays.

Also, in order to formulate this problem, we need to further define the decision variables of the optimization setting that will be valued according to the assignment of a RB to a specific user, either for a cellular, a direct or relayed D2D communication. The binary variable that corresponds to CUEs, relayed D2D or direct D2D RB allocation are defined by (5.8)-(5.10) respectively.

$$x_c^n = \begin{cases} 1, & \text{if CUE } c \in \mathcal{C} \text{ transmits on RB } n \in \mathcal{N} \\ 0, & \text{otherwise.} \end{cases} \tag{5.8}$$

$$y_{ij}^n = \begin{cases} 1, & \text{if DUE (relay) } i \text{ sends to relay (DUE) } j \text{ via } n \\ 0, & \text{otherwise.} \end{cases} \tag{5.9}$$

$$z_d^n = \begin{cases} 1, & \text{if D2D pair } d \text{ communicates directly with RB } n \\ 0, & \text{otherwise.} \end{cases} \tag{5.10}$$

We consider a deterministic model where the signal-to-interference-and-noise-ratio (SINR) between nodes i and j over RB n, denoted as γ_{ij}^n, can be expressed as

$$\gamma_{ij}^n = \frac{P_i G_{ij}}{I_{j,n} + \sigma^2}, \tag{5.11}$$

where $I_{j,n}$ is the interference received by user j over resource block n, P_i is the transmitted power of node i, G_{ij} is the link gain between node i and j, and lastly, σ^2 is the power of background/thermal noise. The D2D interference to the UL transmission of CUE c to BS b over RB n is denoted by $I_{c_b,n}$ and is given by

$$I_{c_b,n} = \sum_{d \in \mathcal{D}} \left(P_d G_{db} z_d^n + \sum_{l \in \mathcal{L}} (P_d G_{db} + P_l G_{lb}) y_{dl}^n \right). \tag{5.12}$$

For the rest of the section, the i, j indexes in G_{ij} (or y_{ij}) correspond to the transmitter and the receiver respectively. Also, from now on, we use y_{dl} to refer to the link between a DUE of pair d to relay l and vice versa.

The uplink channel rate of the cellular user c over resource block n, denoted by R_c^n, is given by

$$R_c^n = B \log_2 \left(1 + \frac{P_c G_{cb} x_c^n}{I_{c_b,n} + \sigma^2}\right), \tag{5.13}$$

where B is the RB bandwidth (180 kHz), P_c is the transmit power of CUE c. Finally, the overall data rate for this CUE is

$$R_c = \sum_{n \in \mathcal{N}} R_c^n. \tag{5.14}$$

Similarly, the interference affecting the D2D receiver of a pair d can be from the cellular user c or the other DUEs/relays that are transmitting over the same resource. If the resource block n is assigned to d, the received interference power for d, denoted by $I_{d,n}$, is given by

$$I_{d,n} = \sum_{c \in \mathcal{C}} P_c G_{cd} x_c^n \; + \sum_{i \in \mathcal{D} \setminus \{d\}} \left(P_i G_{id} z_i^n + \sum_{l_i \in \mathcal{L}} (P_i G_{id} + P_{l_i} G_{l_i d}) y_{il_i}^n \right). \tag{5.15}$$

The rate of the direct D2D communication of link d is then given by

$$R_{\text{direct},d} = B \sum_{n \in \mathcal{N}} \log_2 \left(1 + \frac{P_d G_d z_d^n}{I_{d,n} + \sigma^2}\right), \tag{5.16}$$

where G_d is the channel gain for the D2D pair d transmission.

If the relay-based communication is used for D2D pair d via a relay l, the link capacity of the first and second hop respectively are given by

$$R_{dl}^n = B \sum_{n \in \mathcal{N}} \log_2 \left(1 + \frac{P_d G_{dl} y_{dl}^n}{I_{l,n} + \sigma^2}\right), \tag{5.17}$$

$$R_{ld}^n = B \sum_{n \in \mathcal{N}} \log_2 \left(1 + \frac{P_l G_{ld} y_{ld}^n}{I_{d,n} + \sigma^2}\right), \tag{5.18}$$

where $I_{l,n}$ is defined as the interference power from CUE and the other D2D users exerted to relay node l, and is given by exchanging the subscript d by l in (5.15).

Finally, if we consider that relays are operating in full-duplex (FD) mode in amplify-and-forward communication [156], R_l^n, given below, denotes the total achieved rate for

a relay-aided D2D communication over RB n where l refers to the relay that assists the considered D2D pair d.

$$R_l^n = \min\{R_{dl}^n,\ R_{ld}^n\}. \tag{5.19}$$

5.3.3 Problem formulation

We define the sum-rate maximization problem in a scenario where D2D UEs underlay cellular communications:

$$\max \sum_{n\in\mathcal{N}} \left[\sum_{c\in\mathcal{C}} R_c^n x_c^n + \sum_{d\in\mathcal{D}} \left(R_{\text{direct},d}^n z_d^n + \sum_{l\in\mathcal{L}} R_l^n y_{dl}^n \right) \right] \tag{5.20}$$

$$\text{s.t.} \sum_{n\in\mathcal{N}} R_c^n x_c^n \geq R_{th},\ \forall c \in \mathcal{C} \tag{5.20a}$$

$$\sum_{n\in\mathcal{N}} \left(R_{\text{direct},d}^n z_d^n + \sum_{l\in\mathcal{L}} R_l^n y_{dl}^n \right) \geq R_{th},\ \forall d \in \mathcal{D} \tag{5.20b}$$

$$\sum_{c\in\mathcal{C}} x_c^n = 1,\ \forall n \in \mathcal{N} \tag{5.20c}$$

$$\sum_{n\in\mathcal{N}} x_c^n = 1,\ \forall c \in \mathcal{C} \tag{5.20d}$$

$$\sum_{n\in\mathcal{N}} \left(z_d^n + \sum_{l\in\mathcal{L}} y_{dl}^n \right) = 1,\ \forall d \in \mathcal{D} \tag{5.20e}$$

$$x_c^n, y_{dl}^n, z_d^n \in \{0,1\},\ \forall n \in \mathcal{N},\ d \in \mathcal{D},\ l \in \mathcal{L},\ c \in \mathcal{C}. \tag{5.20f}$$

Constraints (5.20a), (5.20b) restrict the rate to be above a predefined threshold for all communications, i.e. direct, relayed D2D and cellular connections. Following the LTE standard, (5.20c) imposes the orthogonal assignment of the cellular users. Also, constraint (5.20d) signifies the allocation of each cellular user c with a single RB, whereas (5.20e) applies the same RB limitation for the D2D communication and also implies that only one relay can be potentially assisting each D2D link. Thus, the role of the binding variables z, y in the latter constraint is to restrict each D2D to communicate only in either direct or relay mode.

5.3.4 Genetic Algorithm

GA is one of the most popular bio-inspired algorithms and is used to solve real world NP-hard optimization problems. In general, bio-inspired algorithms imitate the natural

evolution of biological organisms to provide a robust, near optimal solution for various problems. GA is inherently an evolutionary process that involves chromosome encoding, population initialization, fitness function depiction, crossover and selection mechanisms. These operations will be briefly explained in Section 5.3.4. A detailed analysis of GAs can be found in [148]. Initially, we introduce the following two important definitions.

Problem mapping: The starting point for the GA is to define a data structure that is appropriate for the solution representation. Since our problem space corresponds to CUE or DUE channel allocation, an integer based chromosome coding mechanism will be used. Based on this, each individual can directly map to a potential channel allocation for CUEs and DUEs where a channel allocation for a UE is represented by a chromosome; a set of chromosomes forms an individual. The initial population consists of a certain number of individuals, denoted by M. A common method to initialize the population is to randomly generate the chromosomes of each individual. In addition, the feasibility of each individual should be ensured to accelerate the convergence process. Thus, we first randomly generate two feasible vectors for each node, according to the representation scheme. Once all vectors are available, they will be combined to form a feasible individual with length equal to $(C + D + L)$. This is repeated until M individuals are generated. The formed population then acts as the very first generation that starts the subsequent evolving steps.

$$\begin{aligned} f = \sum_{n\in\mathcal{N}} \left[\sum_{c\in\mathcal{C}} R_c^n x_c^n + \sum_{d\in\mathcal{D}} \left(R_{\text{direct},d}^n z_d^n + \sum_{l\in\mathcal{L}} R_l^n y_{dl}^n \right) \right] \\ + \sum_{c\in\mathcal{C}} \alpha_1 \min\left(R_{th} - \sum_{n\in\mathcal{N}} R_c^n x_c^n, 0 \right) \\ + \sum_{d\in\mathcal{D}} \alpha_2 \min\left(R_{th} - \sum_{n\in\mathcal{N}} \left(R_{\text{direct},d}^n z_d^n + \sum_{l\in\mathcal{L}} R_l^n y_{dl}^n \right), 0 \right) \end{aligned} \tag{5.21}$$

Fitness function: To this end, we firstly need to interpret the objective of the optimization problem in (5.20) to a fitness function that evaluates the quality of a given individual. In this case, to formulate this we apply a penalty function to ensure that constraints (5.20a) - (5.20b) are satisfied. In addition, the D2D mode selection (i.e. direct or relayed) is also optimized during the fitness evaluation. The fitness function is defined in (5.21).

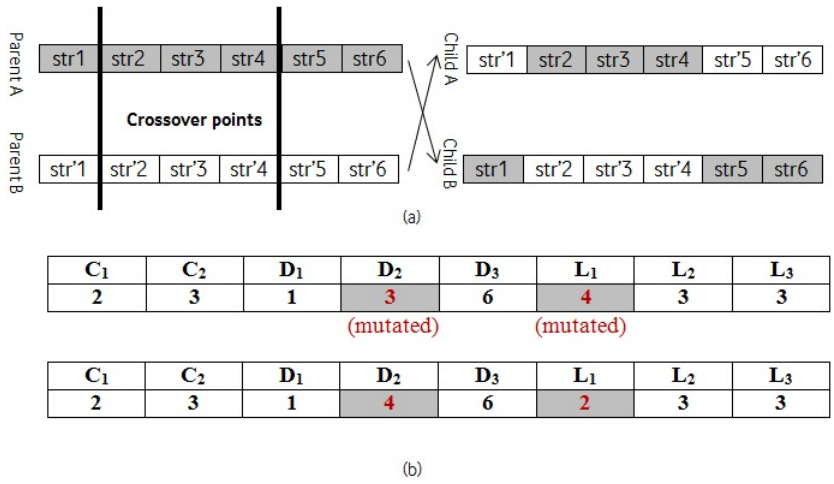


Fig. 5.9 (a) Two-point crossover example. (b) Mutation example.

GA operation

1) Selection: An operation used for choosing individuals to participate in reproduction. In this study, the roulette wheel selection model is used where the chosen probability is proportional to the individual fitness evaluation function. Its selection probability for individual i is defined as

$$p_i = \frac{f(i)}{\sum_{i \in M} f(i)}. \tag{5.22}$$

2) Crossover and mutation: Crossover mixes the current solution so as to find better ones whereas mutation helps the GA avoid local optima. We use one and two points (OP and TP) crossover cases in our results for comparison. An example of a two point crossover is illustrated in Fig. 5.9(a). The mutation operation works by randomly making minor changes in the chromosomes after the crossover operation is performed. In our algorithm, we view each chromosome as a single gene. We define a trivial probability p_m as the likelihood of a gene to mutate. If a gene is determined to mutate, one digit of the vector will be randomly selected and replaced with a different value as shown in Fig. 5.9(b).

3) Replacement: After generating a new population, an elitist based replacement model is employed to modify the old population with a certain number of new individuals. The worst individuals in the parental population are replaced by their children in the next generation.

The algorithm works as follows: an initial population is initialized. Subsequently, the reproduction process starts, including mutation and crossover. The worst individuals are replaced with fitter ones based on the fitness function and this process is repeated

until the maximum number of generations is reached. Considering the run-time performance of the GA, it is dependant on the three mentioned procedures. It is proven that GA scales well in terms of time complexity compared to ILP that are unable to run for highly dense topologies [157]. It is worth noting that GA does not necessarily result in an optimal solution. In our future studies we will focus on comparing the performance of GA to the optimal solution (using ILP for example) as well as the computational complexity of both solutions to demonstrate the trade-off between optimality and computation complexity offered by GA.

5.3.5 Heuristic Algorithm

In this section, we describe an algorithm that prioritizes the D2D users to achieve the maximum rate performance with respect to the cellular throughput. A basic assumption is that cellular users are initially allocated with orthogonal resources to satisfy their UL transmissions. Subsequently, we iterate over all D2D links and pre-calculate for each one of them their potential rate performance (according to Shannon capacity formula) on each RB, based on the interference from cellular UEs. Then, we identify the best combination of D2D UE and RB that gives the maximum among all rate as a starting point. Recall that the maximum rate of a UE on a specific RB can result from either direct or relayed communication. Then, we update the rate matrices ($\mathbf{d}_{\mathrm{m}}$ for direct and $\mathbf{r}_{\mathrm{m}}$ for relayed transmission) with the former step's allocation and iterate over all UEs by taking into account the interference deriving from this RB assignment. Last, after all UEs are served, we estimate the rate that each UE achieves through the final allocation pattern and consequently the overall throughput. The algorithmic steps are analytically shown in Algorithm 2.

5.3.6 Evaluation results

In this section, a set of numerical investigations is presented to evaluate the performance of the GA-based resource allocation method. The results are based on Monte Carlo simulations with 100 iterations. Each transmission is assumed to occupy one RB. The simulation parameters are shown in Table 5.1.

We compare the proposed GA techniques (one-point (OP) and two-points (TP) crossover) with the heuristic RA algorithm that was described in Section 5.3.5 and a random RA method. The random method works as follows: after the allocation of orthogonal RBs to cellular UEs takes place, DUEs are also randomly assigned resources

Algorithm 2 Sum-rate maximization algorithm

Input: $\mathcal{C}$, $\mathcal{D}$, $\mathcal{L}$, $\mathcal{N}$ (with their corresponding cardinalities C, D, L, N) / users' location.
Output: Aggregate throughput: R_{tot}
for c := 1 `to C` **do**
    allocate random orthogonal RB n to user c;
    $\mathcal{N}_{cellular} = \mathcal{N}_{cellular} - \{\mathrm{n}\}$;
end for
for i := 1 `to D` **do**
    for n := 1 `to N` **do**
        calculate $\mathbf{d}_{\mathrm{m}}(\mathrm{i},\mathrm{n})$;
        calculate $\mathbf{r}_{\mathrm{m}}(\mathrm{i},\mathrm{n})$;
    end for
    $\mathbf{d}_{\mathrm{m}}^{\max}(\mathrm{i}) = \max\Big(\mathbf{d}_{\mathrm{m}}(\mathrm{i},:)\Big)$;
    $\mathbf{r}_{\mathrm{m}}^{\max}(\mathrm{i}) = \max\Big(\mathbf{r}_{\mathrm{m}}(\mathrm{i},:)\Big)$;
end for
$\mathbf{S} = zeros(D, 2)$;
j = 1;
while j $\leq D$ **do**
    find $<\mathrm{d},\mathrm{n}>$ combination that gives the maximum rate among all elements in $\mathbf{d}_{\mathrm{m}}^{\max}$ and $\mathbf{r}_{\mathrm{m}}^{\max}$ matrices;
    $\mathbf{S}(\mathrm{d},:) = [\mathrm{d},\mathrm{n}]$;
    Repeat
    update the rates on the assigned RB n $\forall \mathrm{u} \in \mathcal{D} - \{\mathrm{d}\}$ for both $\mathbf{d}_{\mathrm{m}}$, $\mathbf{r}_{\mathrm{m}}$;
    update $\mathbf{d}_{\mathrm{m}}^{\max}(\mathrm{u})$ & $\mathbf{r}_{\mathrm{m}}^{\max}(\mathrm{u})$; $\mathbf{d}_{\mathrm{m}}(\mathrm{d},:) = 0$; $\mathbf{r}_{\mathrm{m}}(\mathrm{d},:) = 0$;
    Until all matrices' rows are updated
    j = j + 1;
end while
for d := 1 `to D` **do**
    calculate achieved rate for direct or relayed D2D comm. for user d (R_d);
end for
for c := 1 `to C` **do**
    calculate achieved rate for cellular link c (R_c);
end for
$R_{tot} = \sum_{c\in\mathcal{C}} R_c + \sum_{d\in\mathcal{D}} R_d$;

Table 5.1 Simulation Parameters

Parameter	**Value**
User distribution	Uniform
Macro cell radius	250 m
D2D link length	$[20, 150]$ m
Number of CUEs in cell	30
Number of relays/D2D links	50
Path-Loss model	$128.1 + 37.6 \log_{10} d$
UE/relay Tx power (*fixed*)	20 dBm
Noise power spectral density	-174 dBm/Hz
System bandwidth (BW)	10 MHz

from the available RB pool and satisfy their transmission needs by selecting either relay or direct mode, depending on which of the two modes provides better rate performance.

An important factor that needs to be taken into account is the convergence point of the applied GA methods. This point can be interpreted as the number of generations that results in the optimal achievable aggregate rate. The box plot in Fig. 5.10 shows that the TP-GA technique converges almost 1.5 times faster compared to the OP-GA (the medians of convergence points in relation to the number of generations are 290 and 412, respectively). Also, the horizontal edges of each box (25 and 75 percentiles) show a bigger gap in the second case where the TP-GA can achieve a faster convergence on average. This can be justified by the TP crossover's ability to ensure a more diverse initial population and encoding that can entail faster convergence to the optimal rate.

Fig. 5.11 shows a sample of the sum-rate performance tendency for a designated number of generations. In this case, the TP-GA not only converges faster (i.e. 210 generations less needed) but also the achievable rate is notably high compared to the heuristic (almost 10%) and clearly better than the OP-GA method. It should be noted that, in this case study, the TP-GA method provides a higher capacity performance even from the second generation and beyond, while OP-GA converges in its optimal point in the 468^{th} generation but with rather sub-optimal throughput. Last, for this simulation run, TP-GA outperforms the random method with almost 21% gain in terms of sum rate performance.

Fig. 5.12 illustrates the sum-rate performance of the proposed methods when the D2D transmitter and receiver are separated by fixed distances for each evaluation point. The TP crossover GA method achieves an average sum-rate gain of 4%, 24% and 43%

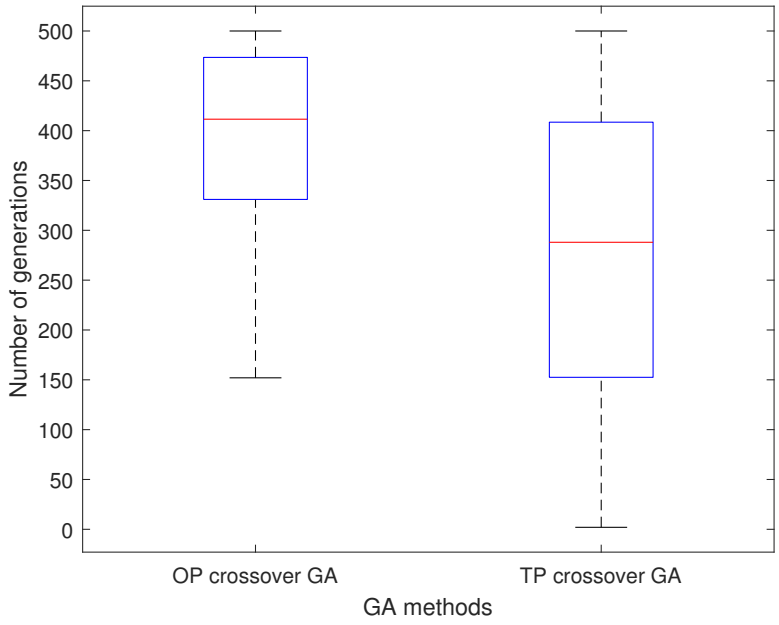


Fig. 5.10 Average convergence points for the case of (i) one-point (OP) crossover GA, and (ii) two-points (TP) crossover GA.

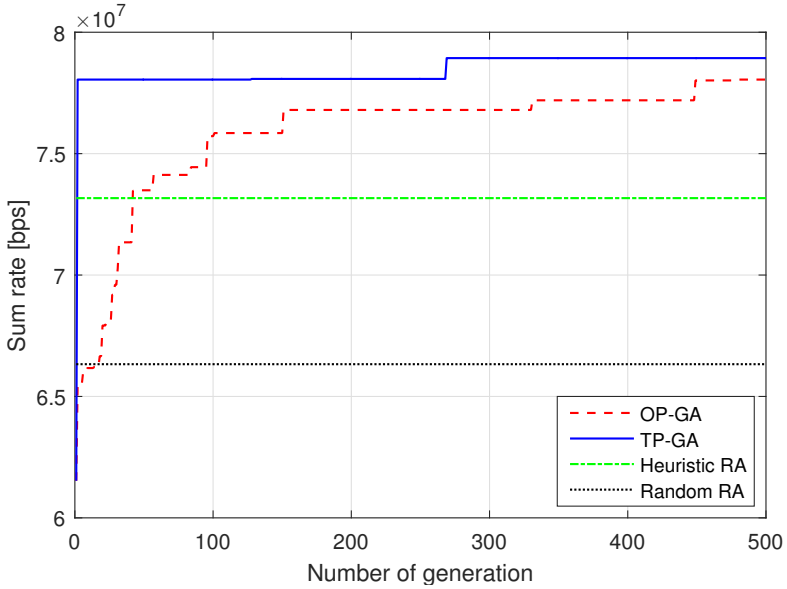


Fig. 5.11 An example of the GA's convergence to the maximum rate values.

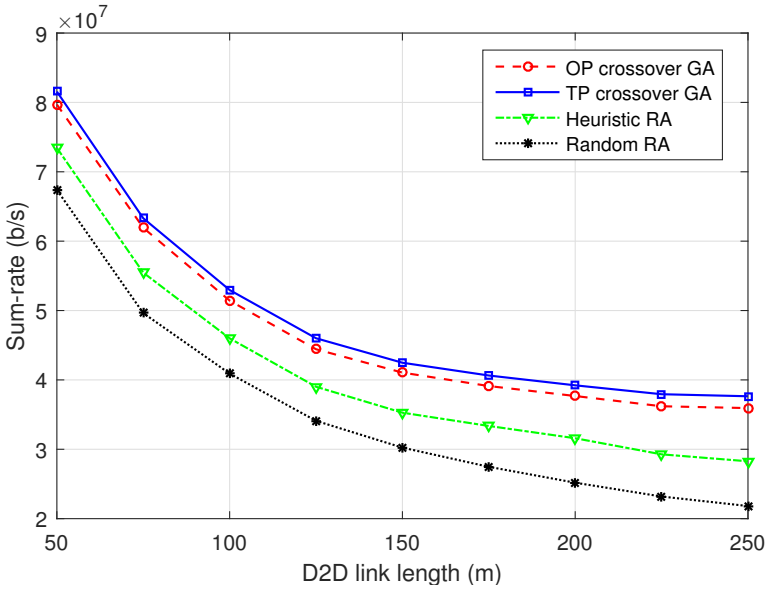


Fig. 5.12 Aggregate throughput in relation to varying D2D link lengths.

compared to the OP-GA, heuristic and random allocation techniques, respectively. The plot shows that even though the rate drops proportionally with the increase of the D2D link range, the performance gap of the GA proposed algorithms in comparison to the two RA schemes becomes larger. At 250 meters, the TP-GA method provides a rate improvement of 37% and 72% compared to the heuristic and the random methods, respectively, i.e. a more efficient resource and mode (direct, relayed) selection for D2D communications.

Finally, in Fig. 5.13 we investigate the received interference by D2D UEs for all the considered cases. Note that, this interference can result from both a cellular and other D2D/relay transmissions that reuse the same spectrum. As shown in the figure, the GA methods achieve a lower interference level where at the 50th percentile, the interference level in the GA case is 4.7 and 10 dB lower than the heuristic and random methods respectively, and at the 90th percentile the GA interference reduction is 9.4 and 15.7 dB compared to the baseline methods.

5.4 Summary

In this chapter, we presented an extensive optimization framework of the access procedures and resource management in a D2D enabled cellular network. In the first part, we presented different cell association strategies for D2D communications in cellular networks. Based on the notion of decoupled UL and DL connections, we

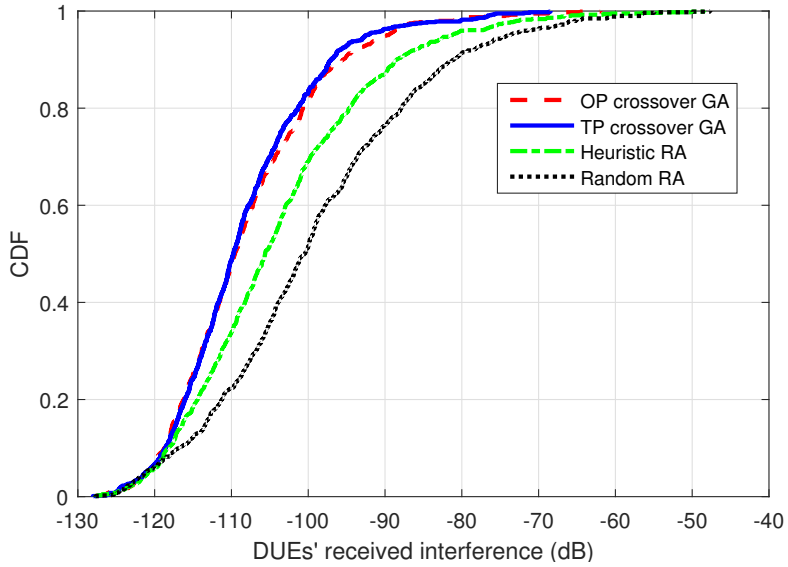


Fig. 5.13 CDF of DUEs' received interference.

proposed an integer linear programming (ILP) optimization framework that aimed at achieving efficient D2D cell association with respect to interference reduction as well as resource utilization. Extensive simulations show the significant gains of the applied methods. The Disjoint-Decoupled (DD) optimization technique achieves more than twice in UL interference reduction as well as more than 5 dBs reduction in devices transmit power compared to baseline cell association methods. However, DD results in an inefficient use of D2D resources. Therefore, we introduced the Hybrid-Decoupled (HD) technique which achieves a balance between the interference reduction and resource utilization efficiency. HD results in a slightly worse interference performance than the DD scheme but with a 45% improvement in resource usage efficiency.

In the second part, we presented a resource allocation methodology for relay-aided D2D communications underlying a cellular network. By exploiting the robustness and scalability of bio-inspired meta-heuristic techniques, we proposed a low-complexity genetic algorithmic framework that aimed at maximizing the network throughput performance with respect to interference. Numerical results highlight the gains from the investigated GA methods. The proposed one-point and two-points crossover GA techniques provide significant rate improvement compared with baseline RA methods (more than 20% on average) as well as ensure the least exerted interference towards D2D transmissions with an average gain of more than 4 dBs.

Chapter 6

Conclusions and future work

6.1 Concluding remarks

5G is expected to redefine cellular networks in the sense that the telecom industry will not solely be about providing mobile broadband -although this will still be the core business- but will be expanded to many other use cases and services. 5G will enable new business opportunities across several vertical industries such as automotive, healthcare and factory automation In order for 5G to accomplish this vision, a fresh look at the traditional way cellular networks are designed and deployed is needed. Until the fourth generation of mobile telephony, cellular networks have been relying on the axiomatic role of cells where a device is simply served in both uplink and downlink by the base station controlling the cell where the device is located. 5G is expected to witness increased levels of heterogeneity in terms of infrastructure (macro, pico and femto cells), spectrum (sub-6GHz, millimeter waves) and traffic requirements (data rate, reliability and latency). With this in mind, the current cell-centric design needs to be changed. A shift from a cell-centric to a device-centric approach is expected to be a main characteristic of future 5G networks. A device centric architecture is one where a device should be able to communicate by exchanging several information flows through several possible heterogeneous nodes where these nodes and their functions are tailored to fit the specific requirements of the device.

In this thesis, we tackle a few aspects that constitute this device-centric architecture concept. The main contribution of the thesis is on Downlink and Uplink Decoupling (DUDe). In a heterogeneous network where macro and small cells coexist there is an uplink and downlink coverage imbalance arising from the difference in the transmit power between the different tiers of the network. This gives rise to the concept of DUDe where an independent association in the uplink and downlink is proposed. An

extensive simulation study of DUDe based on realistic deployments and channel models has been presented in Chapter 2 where initially association in the UL and DL are based on the pathloss and downlink received power respectively. Several benefits in terms of uplink data rate, SNR, interference and load balancing have been illustrated. A comparison with conventional coupled association techniques has shown throughput gains that range between 100-600% as well as a large improvement in network outage due to the load balancing effect of DUDe. Taking the concept a step forward, an optimization framework for load and backhaul aware DUDe is presented and shown to offer superior performance to baseline DUDe.

In Chapter 3, we presented a comprehensive theoretical study of DUDe where in the first part we derived the UL capacity for a sub-6GHz heterogeneous network as well as show the same trends in association probability for different evaluation methods. In the second part we studied the association probabilities in a mixed sub-6GHz and millimeter waves scenario. We have derived the association probabilities based on the DL and UL received power as well as the UL and DL maximum achievable rate where different decoupling trends were witnessed in both techniques. The SINR and rate distributions were also derived where special emphasis was put on how biasing affects the SINR and rate behaviours. It was found that aggressive values of bias are favourable to harness the gains from the wide bandwidth at millimeter waves.

Chapter 4 features a study of the architectural aspect of DUDe with a discussion on the interoperability of DUDe with different emerging techniques. The main focus however was in studying how DUDe can be supported in current 4G networks and what are the changes that need to be done for a more efficient support in future 5G networks.

In Chapter 5, we move towards a different component of the device-centric concept which is Device-to-Device (D2D) which is a feature that enables to establish a direct communication link between devices in the same vicinity instead of the data being relayed through the infrastructure. This approach improves the spectral efficiency as well as reduce the latency and the energy used to transfer the data from one point to another. In the first part of the chapter, we studied the D2D cell association problem where we showed the superiority of the decoupled association using an optimization framework based on Integer linear programming (ILP). In the second part we turn our attention to the resource management problem where we use bio-inspired genetic algorithms to develop a near optimal real-time resource management algorithm.

We summarize some of the main takeaways and conclusions in the following:

- DUDe provides multiple folds of improvement in the UL rate as compared to coupled access as well as gains in terms of outage and load balancing.
- DUDe results in improvements in power savings at the terminal side which makes it an attractive solution for internet of things applications.
- the cell association probability depends chiefly on the density of the deployment, but not the process used to generate the deployment geometry.
- In a mixed millimeter waves and sub-6GHz deployment, from a maximum rate perspective, UEs tend to connect in the UL and DL to sub-6GHz Mcells and mmWaves Scells respectively which is the opposite to the decoupling trends witnessed using UL and DL received power association.
- In the previously mentioned setup, aggressive values of Scell biasing (100x normal values in sub-6GHz networks) are desirable to harness the gains from the larger bandwidth at mmWaves. In addition, more robust modulation and coding schemes are needed to compensate the much reduced SINR due to the large bias values.
- DUDe does not require many changes in current architectures to be efficiently supported which makes it a very attractive feature to implement in a native way in future 5G networks.
- Decoupled access is quite relevant in D2D enabled networks where in an overlay D2D deployment, DUDe is shown to improve the resource utilization, power efficiency and interference effects.
- Bio-inspired genetic algorithms are a very good fit for D2D radio resource management, especially for delay sensitive applications due to their computational efficiency and fast convergence.

6.2 Future work

6.2.1 Decoupled Access

Decoupled access in mmWave networks with mobility

In our analysis on cell association in millimeter waves we considered a static scenario. The main challenges of mmWaves are: the high near field pathloss which results in

a limited coverage of mmWaves and the sensitivity to blockage. The first challenge can be compensated using high gain directional antennas which could actually over-compensate the propagation losses in mmWaves. However, the second challenge which is the sensitivity to blockage is the main problem especially in dynamic scenario where UEs and obstacles are moving which creates a continuously changing coverage and fluctuating signal quality of mmWaves. It would be very interesting to study the effect of mobility on cell association as well as the SINR and rate distributions.

Control and data separation in millimeter waves

Control and data separation has been a hot topic lately where the idea is to transmit the control signals on lower frequencies that provide more consistent coverage whereas the data plane is transmitted on higher frequencies where more bandwidth is available. However, lower frequencies are getting more and more crowded so congestion or even collision in the case of random access could be a possibility in an extremely dense scenario. Exploring the possibility of transmitting some control signals on mmWaves could be interesting where the noise limited properties of mmWaves would mean a better SNR than on lower frequencies but with less consistency in the channel. Therefore a study of control and data separation in a mmWaves/sub-6GHz network could make sense where a UE could opportunistically choose the link on which the SNR or reliability is better to transmit control signals and the link with more bandwidth for rate demanding data signals. This would constitute a fully decoupled system where UL and DL as well as control and data are decoupled.

DUDe and the Internet of Things

We have shown in our results in Chapter 2 that using DUDe, the device transmit power can be reduced by more than 9 dB compared to the conventional coupled association which means a reduction of 8 times. This is extremely relevant for Internet of things (IoT) applications where some applications are UL centric and the devices are battery powered and are expected to last for years. DUDe could mean huge savings by prolonging the battery life of IoT devices. A more in-depth study on IoT specific association mechanisms based on DUDe and how can DUDe be modified to further improve the power efficiency for IoT applications.

6.2.2 D2D communications

Reliablity in mobility enabled D2D communication

One of the major applications of D2D is Vehicle-to-Vehicle (V2V) communications. In V2V there are -mainly- two types of traffic: reliability and latency demanding traffic such as beacons and traffic alerts as well as data rate demanding traffic such as media and see-through data. These different traffic types require new radio resource management algorithms that take the different traffic requirements into account. Another aspect related to V2V is the interoperability between operators which is needed to be studied carefully for the V2V use case to be applicable.

D2D in millimeter waves

Using millimeter waves for D2D communications is an interesting topic to explore. The narrow beams characterizing mmWaves would result in reducing the interference effect between D2D and cellular communications which could result in a more relaxed setup where a coordination between D2D and cellular communications is not necessary. Also the short range nature of mmWaves is very suitable for D2D communications. A study on the advantages and disadvantages of using mmWaves for D2D communications would make sense.

6.2.3 Software defined networks and Virtualization

Software defined networking (SDN) and Network function virtualization (NFV) constitute major components in the device-centric architecture vision. SDN allows to decouple the control and data planes at the core network side which allows more flexibility in handling the network and makes it more scalable. NFV leads to decoupling the node function and the hardware allocated to handle the processing associated with the node. This allows the flexibility of assigning resources where they are most needed in a seamless way. NFV and SDN allow the flexibility of creating different network slices or instances to serve different requirements for different traffic types. Linking the decoupled UL and DL and control and data on the radio side with the enabling concepts on the core network side such as SDN and NFV is quite interesting and would give a holistic view of a fully dynamic and device-centric end-to-end network. Studying the dependencies and relationship between the core and radio concepts is a very interesting topic for the future.

References

[1] N. Alliance, "5G white paper," *Next Generation Mobile Networks, White paper*, 2015.

[2] G. P. Fettweis, "The tactile internet: Applications and challenges," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, March 2014.

[3] D. Jiang and L. Delgrossi, "Ieee 802.11p: Towards an international standard for wireless access in vehicular environments," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, May 2008, pp. 2036–2040.

[4] T. Litman, "Autonomous vehicle implementation predictions," *Victoria Transport Policy Institute*, vol. 28, 2014.

[5] M. Iwamura, "NGMN View on 5G Architecture," in *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, May 2015, pp. 1–5.

[6] E. Liotou, H. Elshaer, R. Schatz, R. Irmer, M. Dohler, N. Passas, and L. Merakos, "Shaping QoE in the 5G ecosystem," in *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, May 2015, pp. 1–6.

[7] M. Fallgren, B. Timus *et al.*, "Deliverable D1.1 Scenarios, requirements and KPIs for 5G mobile and wireless system," *METIS*, 2013.

[8] G. Durisi, T. Koch, and P. Popovski, "Towards massive, ultra-reliable, and low-latency wireless: The art of sending short packets," *arXiv preprint arXiv:1504.06526*, 2015.

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

[10] 3GPP, "Study on Small Cell Enhancements for E-UTRA and E-UTRAN; Higher Layer Aspects," 3rd Generation Partnership Project (3GPP), TR 36.842, Sep. 2014. [Online]. Available: http://www.3gpp.org/dynareport/36842.htm

[11] S. Ericsson, "Physical Layer Aspects of Dual Connectivity," 3GPP TSG-RAN, Tech. Rep. R1-130566, Feb. 2013.

[12] Y. Kishiyama, A. Benjebbour, T. Nakamura, and H. Ishii, "Future steps of LTE-A: evolution toward integration of local area and wide area systems," *IEEE Wireless Communications*, vol. 20, no. 1, pp. 12–18, Feb. 2013.

[13] A. Capone, A. F. dos Santos, I. Filippini, and B. Gloss, "Looking beyond green cellular networks," in *Wireless On-demand Network Systems and Services (WONS), 2012 9th Annual Conference on*, Jan 2012, pp. 127–130.

[14] 3GPP TR 23.703, "Study on architecture enhancements to support proximity services (ProSe) (Release 12)," Tech. Rep., Dec. 2013.

[15] S. Sezer, S. Scott-Hayward, P. K. Chouhan, B. Fraser, D. Lake, J. Finnegan, N. Viljoen, M. Miller, and N. Rao, "Are we ready for SDN? Implementation challenges for software-defined networks," *IEEE Communications Magazine*, vol. 51, no. 7, pp. 36–43, July 2013.

[16] O. Committee *et al.*, "Software-defined networking: The new norm for networks," *Open Networking Foundation*, 2012.

[17] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[18] B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, "Network function virtualization: Challenges and opportunities for innovations," *IEEE Communications Magazine*, vol. 53, no. 2, pp. 90–97, Feb 2015.

[19] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[20] Q. Li, R. Q. Hu, Y. Qian, and G. Wu, "Cooperative communications for wireless networks: techniques and applications in LTE-advanced systems," *IEEE Wireless Communications*, vol. 19, no. 2, Apr. 2012.

[21] "Enders Analysis/TNS RI-Survey," *Next Generation Mobile Networks, White paper*, May 2014.

[22] A. Sackl and R. Schatz, "Evaluating the impact of expectations on end-user quality perception," in *Proceedings of International Workshop Perceptual Quality of Systems (PQS)*, 2013, pp. 122–128.

[23] R. Schatz, T. Hoßfeld, L. Janowski, and S. Egger, "From packets to people: quality of experience as a new measurement challenge," in *Data traffic monitoring and analysis*. Springer, 2013, pp. 219–263.

[24] D. Raychaudhuri and N. B. Mandayam, "Frontiers of wireless and mobile communications," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 824–840, 2012.

[25] S. Egger, B. Gardlo, M. Seufert, and R. Schatz, "The impact of adaptation strategies on perceived quality of http adaptive streaming," in *Proceedings of the 2014 Workshop on Design, Quality and Deployment of Adaptive Video Streaming*. ACM, 2014, pp. 31–36.

[26] A. Sackl, S. Egger, and R. Schatz, "The influence of network quality fluctuations on Web QoE," in *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on*, Sep. 2014, pp. 123–128.

[27] P. Casas, A. Sackl, R. Schatz, L. Janowski, J. Turk, and R. Irmer, "On the quest for new kpis in mobile networks: The impact of throughput fluctuations on qoe," in *Communication Workshop (ICCW), 2015 IEEE International Conference on*, Jun. 2015, pp. 1705–1710.

[28] A. Sackl, P. Casas, R. Schatz, L. Janowski, and R. Irmer, "Quantifying the impact of network bandwidth fluctuations and outages on web qoe," in *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, May 2015, pp. 1–6.

[29] T. Zinner, T. Hoßfeld, M. Fiedler, F. Liers, T. Volkert, R. Khondoker, and R. Schatz, "Requirement driven prospects for realizing user-centric network orchestration," *Multimedia Tools and Applications*, vol. 74, no. 2, pp. 413–437, 2015.

[30] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, March 2010.

[31] P. Reichl, B. Tuffin, and R. Schatz, "Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience," *Telecommunication Systems*, vol. 52, no. 2, pp. 587–600, 2013.

[32] F. Wamser, S. Deschner, T. Zinner, and P. Tran-Gia, "Investigation of different approaches for QoE-oriented scheduling in OFDMA networks," in *Mobile Networks and Management*. Springer, 2013, pp. 172–187.

[33] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and uplink decoupling: A disruptive architectural design for 5G networks," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2014.

[34] J. F. Monserrat, G. Mange, V. Braun, H. Tullberg, G. Zimmermann, and Ö. Bulakci, "METIS research advances towards the 5G mobile and wireless system definition," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, pp. 1–16, 2015.

[35] "Open Flow Foundation, SDN Architecture, Issue 1," Jun. 2014.

[36] "FCC website:," http://www.fcc.gov/openinternet.

[37] K. De Moor, M. Fiedler, P. Reichl, and M. Varela, "Quality of experience: From assessment to application (dagstuhl seminar 15022)," *Dagstuhl Reports*, vol. 5, no. 1, 2015.

[38] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Load and backhaul aware decoupled downlink/uplink access in 5G systems," in *2015 IEEE International Conference on Communications (ICC)*, Jun. 2015, pp. 5380–5385.

[39] F. Boccardi, J. Andrews, H. Elshaer, M. Dohler, S. Parkvall, P. Popovski, and S. Singh, "Why to decouple the uplink and downlink in cellular networks and how to do it," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 110–117, Mar. 2016.

[40] N. Abu-Ali, A.-E. Taha, M. Salah, and H. Hassanein, "Uplink Scheduling in LTE and LTE-Advanced: Tutorial, Survey and Evaluation Framework," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1239–1265, Dec. 2014.

[41] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.

[42] S. Singh and J. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 888–901, Feb. 2014.

[43] "Atoll," http://www.forsk.com/web/EN/11-atoll-overview.php.

[44] J. McKown and J. Hamilton, R.L., "Ray tracing as a design tool for radio networks," *IEEE Netw.*, vol. 5, no. 6, pp. 27–30, Nov 1991.

[45] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*, 1st ed. Academic Press, 2011.

[46] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 10)," TR 36.213, 2011.

[47] Y. Wang and K. Pedersen, "Performance Analysis of Enhanced Inter-Cell Interference Coordination in LTE-Advanced Heterogeneous Networks," in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, May 2012, pp. 1–5.

[48] S. Singh, X. Zhang, and J. Andrews, "Joint rate and SINR coverage analysis for decoupled uplink-downlink biased cell associations in HetNets," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5360–5373, May 2015.

[49] H. Boostanimehr and V. Bhargava, "Joint downlink and uplink aware cell association in hetnets with QoS provisioning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5388–5401, Oct 2015.

[50] A. de Domenico, V. Savin, and D. Ktenas, "A backhaul-aware cell selection algorithm for heterogeneous cellular networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, Sep. 2013, pp. 1688–1693.

[51] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Cache-aware user association in backhaul-constrained small cell networks," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2014 12th International Symposium on*, May 2014, pp. 37–42.

[52] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, "Distributed alpha - optimal user association and cell load balancing in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 177–190, Feb 2012.

[53] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, Jun. 2013.

[54] W. Saad, Z. Han, R. Zheng, M. Debbah, and H. Poor, "A college admissions game for uplink user association in wireless small cell networks," in *INFOCOM, 2014 Proceedings IEEE*, April 2014, pp. 1096–1104.

[55] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures ," TS 36.213.

[56] H. Zhang, N. Prasad, S. Rangarajan, S. Mekhail, S. Said, and R. Arnott, "Standards-compliant LTE and LTE-A uplink power control," in *Communications (ICC), 2012 IEEE International Conference on*, Jun. 2012, pp. 5275–5279.

[57] K. Smiljkovikj, H. Elshaer, P. Popovski, F. Boccardi, M. Dohler, L. Gavrilovska, and R. Irmer, "Capacity analysis of decoupled downlink and uplink access in 5g heterogeneous systems," *arXiv preprint arXiv:1410.7270*, 2014.

[58] H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and uplink cell association with traditional macrocells and millimeter wave small cells," *Submitted to IEEE Trans. Wireless Commun.*, 2016. Available: http://arxiv.org/abs/1601.05281.

[59] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. Prentice Hall PTR New Jersey, 1996, vol. 2.

[60] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[61] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, and C. E. Wheatley, "On the capacity of a cellular CDMA system," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 2, pp. 303–312, May 1991.

[62] L. Kleinrock and J. Silvester, "Optimum transmission radii for packet radio networks or why six is a magic number," in *Proceedings of the IEEE National Telecommunications Conference*, vol. 4, 1978, pp. 1–4.

[63] M. Haenggi, *Stochastic geometry for wireless networks*. Cambridge University Press, 2012.

[64] M. Haenggi and R. K. Ganti, *Interference in large wireless networks*. Now Publishers Inc, 2009.

[65] S. Weber and J. G. Andrews, "Transmission capacity of wireless networks," *arXiv preprint arXiv:1201.0662*, 2012.

[66] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev, "Stochastic geometry and architecture of communication networks," *Telecommunication Systems*, vol. 7, no. 1-3, pp. 209–227, 1997.

[67] F. Baccelli and S. Zuyev, "Stochastic geometry models of mobile communication networks," *Frontiers in queueing*, pp. 227–243, 1997.

[68] T. X. Brown, "Cellular performance bounds via shotgun cellular systems," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 11, pp. 2443–2455, Nov 2000.

[69] J. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[70] M. Dohler, R. W. Heath, A. Lozano, C. B. Papadias, and R. A. Valenzuela, "Is the PHY layer dead?" *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 159–165, Apr. 2011.

[71] J. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 136–144, Mar. 2013.

[72] K. Smiljkovikj, P. Popovski, and L. Gavrilovska, "Analysis of the decoupled access for downlink and uplink in wireless heterogeneous networks," *IEEE Commun. Lett.*, vol. 4, no. 2, pp. 173–176, Apr. 2015.

[73] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

[74] H. S. Dhillon and J. G. Andrews, "Downlink rate distribution in heterogeneous cellular networks under generalized cell selection," vol. 3, no. 1, pp. 42–45, Feb. 2014.

[75] F. Baccelli and B. Blaszczyszyn, *Stochastic geometry and wireless networks: Theory*. Now Publishers Inc, 2009, vol. 1.

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

[77] C. Li, J. Zhang, and K. Letaief, "Throughput and energy efficiency analysis of small cell networks with multi-antenna base stations," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2505–2517, May 2014.

[78] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019," Whitepaper, available at: http://goo.gl/xxLT.

[79] J. C. Bose *et al.*, "Collected physical papers," 1927.

[80] D. Roddy, "Satellite communications," 2006.

[81] N. Alliance, "Small cell backhaul requirements," *white paper, June*, 2012.

[82] J. Hansryd, J. Edstam, B.-E. Olsson, and C. Larsson, "Non-line-of-sight microwave backhaul for small cells," *Ericsson Review*, vol. 22, 2013.

[83] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-ghz integrated circuits and systems for wireless communications," *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug 2011.

[84] E. Perahia, C. Cordeiro, M. Park, and L. L. Yang, "IEEE 802.11ad: Defining the Next Generation Multi-Gbps Wi-Fi," in *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*, Jan 2010, pp. 1–5.

[85] S. J. Vaughan-Nichols, "Gigabit Wi-Fi is on its way," *Computer*, no. 11, pp. 11–14, 2010.

[86] T. Baykas, C. S. Sum, Z. Lan, J. Wang, M. A. Rahman, H. Harada, and S. Kato, "IEEE 802.15.3c: the first IEEE wireless standard for data rates over 1 Gb/s," *IEEE Communications Magazine*, vol. 49, no. 7, pp. 114–121, July 2011.

[87] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[88] C. H. Doan, S. Emami, D. A. Sobel, A. M. Niknejad, and R. W. Brodersen, "Design considerations for 60 GHz CMOS radios," *IEEE Communications Magazine*, vol. 42, no. 12, pp. 132–140, Dec 2004.

[89] Y. P. Zhang and D. Liu, "Antenna-on-chip and antenna-in-package solutions to highly integrated millimeter-wave devices for wireless communications," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 10, pp. 2830–2841, Oct 2009.

[90] S. Rangan, T. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[91] F. Khan and Z. Pi, "mmWave mobile broadband (MMB): Unleashing the 3-300GHz spectrum," in *Sarnoff Symposium, 2011 34th IEEE*, May 2011, pp. 1–6.

[92] H. Zhao, R. Mayzus, S. Sun, M. Samimi, J. K. Schulz, Y. Azar, K. Wang, G. N. Wong, F. Gutierrez, and T. S. Rappaport, "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city," in *Communications (ICC), 2013 IEEE International Conference on*, June 2013, pp. 5163–5167.

[93] K. Allen, N. DeMinco, J. Hoffman, Y. Lo, and P. Papazian, "Building penetration loss measurements at 900 mhz, 11.4 ghz, and 28.8 mhz, ser," *NTIA report–94-306. Boulder, CO: US Dept. of Commerce, National Telecommunications and Information Administration*, 1994.

[94] J. Lu, D. Steinbach, P. Cabrol, and P. Pietraski, "Modeling the impact of human blockers in millimeter wave radio links," *ZTE Commun. Mag*, vol. 10, no. 4, pp. 23–28, 2012.

[95] T. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[96] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: A MAC layer perspective," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3437–3458, Oct 2015.

[97] T. B. Cho, D. W. Cline, C. S. G. Conroy, and P. R. Gray, "Design considerations for low-power, high-speed CMOS analog/digital converters," in *Low Power Electronics, 1994. Digest of Technical Papers., IEEE Symposium*, Oct 1994, pp. 70–73.

[98] D. Colombi, B. Thors, and C. Tornevik, "Implications of EMF exposure limits on output power levels for 5G devices above 6 GHz," *IEEE Antennas Wireless Propag. Lett.*, vol. 14, pp. 1247–1249, Feb. 2015.

[99] Q. C. Li, H. Niu, G. Wu, and R. Q. Hu, "Anchor-booster based heterogeneous networks with mmwave capable booster cells," in *Globecom Workshops (GC Wkshps), 2013 IEEE*, Dec 2013, pp. 93–98.

[100] K. J. Koh, J. W. May, and G. M. Rebeiz, "A millimeter-wave (40-45 GHz) 16-element phased-array transmitter in 0.18- μ m SiGe BiCMOS technology," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 5, pp. 1498–1509, May 2009.

[101] S. W. Peters, A. Y. Panah, K. T. Truong, and R. W. Heath Jr, "Relay architectures for 3GPP LTE-advanced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, no. 1, pp. 1–14, 2009.

[102] A. K. Gupta, J. G. Andrews, and R. W. Heath Jr, "On the feasibility of sharing spectrum licenses in mmwave cellular systems," *arXiv preprint arXiv:1512.01290*, 2015.

[103] F. Boccardi, H. Shokri-Ghadikolaei, G. Fodor, E. Erkip, C. Fischione, M. Kountouris, P. Popovski *et al.*, "Spectrum pooling in mmwave networks: Opportunities, challenges, and enablers," *arXiv preprint arXiv:1603.01080*, 2016.

[104] M. Rebato, M. Mezzavilla, S. Rangan, and M. Zorzi, "Resource sharing in 5G mmwave cellular networks," *arXiv preprint arXiv:1603.02651*, 2016.

[105] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs, "LTE release 12 and beyond," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 154–160, Jul. 2013.

[106] T. Bai and R. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, Feb. 2015.

[107] S. Singh, M. Kulkarni, A. Ghosh, and J. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2196–2211, Oct. 2015.

[108] J. Park, S.-L. Kim, and J. Zander, "Tractable resource management with uplink decoupled millimeter-wave overlay in ultra-dense cellular networks," *IEEE Trans. Wireless Commun.*, 2015. Submitted, available at: http://arxiv.org/abs/1507.08979.

[109] B. Blaszczyszyn, M. K. Karray, and H.-P. Keeler, "Using Poisson processes to model lattice cellular networks," *in Proc. IEEE INFOCOM*, pp. 773–781, 2013.

[110] 3GPP, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 11)," 3rd Generation Partnership Project (3GPP), TS 36.300, Sep. 2012. [Online]. Available: http://www.3gpp.org/dynareport/36300.htm

[111] M. A. Lema, M. Garcia-Lozano, S. Ruiz, and D. G. Gonzalez, "Improved Component Carrier Selection Considering MPR Information for LTE-A Uplink Systems," in *2013 IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 2191–2196.

[112] M. Lema, M. Garcia-Lozano, and S. Ruiz, "MPR-Aware Scheduler for Carrier Aggregation Transmissions in LTE Uplink," *Wireless Personal Communications*, vol. 84, no. 2, pp. 1417–1438, 2015. [Online]. Available: http://dx.doi.org/10.1007/s11277-015-2695-5

[113] 3GPP, "Coordinated Multi-Point Operation for LTE Physical Layer Aspects," 3rd Generation Partnership Project (3GPP), TS 36.819, Sep. 2012. [Online]. Available: http://www.3gpp.org/dynareport/36819.htm

[114] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, December 2010.

[115] D. Colombi, B. Thors, and C. Tornevik, "Implications of EMF Exposure Limits on Output Power Levels for 5G Devices Above 6 GHz," *Antennas and Wireless Propagation Letters, IEEE*, vol. 14, pp. 1247–1249, 2015.

[116] S. Parkvall, E. Dahlman, G. Jöngren, S. Landström, and L. Lindbom, "Heterogeneous Network Deployments in LTE," *Ericsson review*, Feb. 2011.

[117] Huawei, "Handling of UL Traffic of a DL Split Bearer," 3GPP TSG-RAN, Tech. Rep. R2-140054, Feb. 2014.

[118] K. Smiljkovikj, P. Popovski, and L. Gavrilovska, "Analysis of the decoupled access for downlink and uplink in wireless heterogeneous networks," *IEEE Wireless Communications Letters*, vol. 4, no. 2, pp. 173–176, April 2015.

[119] 3GPP, "LTE Radio Access Network (RAN) Enhancements for Diverse Data Applications," 3rd Generation Partnership Project (3GPP), TR 36.822, Sep. 2012. [Online]. Available: http://www.3gpp.org/dynareport/36822.htm

[120] S. Ericsson, "Further Discussions on UL/DL Split," 3GPP TSG-RAN, Tech. Rep. R2-131678, May 2013.

[121] ZTE, "Comparison between CP solution C1 and C2," 3GPP TSG-RAN, Tech. Rep. R2-132383, Aug. 2013.

[122] NGMN, "Further Study on Critical C-RAN Technologies," NGMN, Deliverable, 2015.

[123] U. Dötsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, and P. Sehier, "Quantitative analysis of split base station processing and determination of advantageous architectures for LTE," *Bell Labs Technical Journal*, vol. 18, no. 1, pp. 105–128, 2013. [Online]. Available: http://dx.doi.org/10.1002/bltj.21595

[124] A. Maeder, M. Lalam, A. De Domenico, E. Pateromichelakis, D. Wubben, J. Bartelt, R. Fritzsche, and P. Rost, "Towards a flexible functional split for cloud-RAN networks," in *Networks and Communications (EuCNC), 2014 European Conference on*, June 2014, pp. 1–5.

[125] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[126] A. Ford, M. Handley, and O. Bonaventure, "TCP Extensions for Multipath Operation with Multiple Addresses," *IETF RFC 6824*, January 2013.

[127] P. Mulroy, S. Appleby, M. Nilsson, and B. Crabtree, "The use of MulTCP for the delivery of equitable quality video," in *International Packet Video Workshop*. IEEE, May 2009.

[128] H. Elshaer, C. Vlachos, V. Friderikos, and M. Dohler, "Interference-aware decoupled cell association in Device-to-Device based 5G networks," *Accepted in IEEE Vehicular Technology Conference (VTC) Spring. Available: http://arxiv.org/abs/1601.05603*, 2016.

[129] C. Vlachos, H. Elshaer, V. Friderikos, and M. Dohler, "Bio-Inspired resource allocation for relay-aided Device-to-Device communications," *Accepted in IEEE Vehicular Technology Conference (VTC) Fall*, 2016.

[130] A. Asadi, Q. Wang, and V. Mancuso, "A survey on Device-to-Device communication in cellular networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 1801–1819, Apr. 2014.

[131] Y.-D. Lin and Y.-C. Hsu, "Multihop cellular: a new architecture for wireless communications," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, Mar. 2000, pp. 1273–1282 vol.3.

[132] B. Kaufman and B. Aazhang, "Cellular networks with an overlaid device to device network," in *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, Oct. 2008, pp. 1537–1541.

[133] J. Du, W. Zhu, J. Xu, Z. Li, and H. Wang, "A compressed HARQ feedback for Device-to-Device multicast communications," in *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, Sep. 2012, pp. 1–5.

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

[135] P. Phunchongharn, E. Hossain, and D. I. Kim, "Resource allocation for device-to-device communications underlaying lte-advanced networks," *IEEE Wireless Communications*, vol. 20, no. 4, pp. 91–100, Aug. 2013.

[136] 3GPP TR 22.803, "Feasibility study for Proximity Services (ProSe) (Release 12)," Tech. Rep., Jun. 2012.

[137] W.-F. A. T. Committee *et al.*, "P2P Task Group,"Wi-fi peer-to-peer (P2P) technical specification, ver. 1.1,"," WiFi Alliance, Tech. Rep, Tech. Rep., 2010.

[138] Z. Specification, "Zigbee document 053474r06 version 1.0," *Zigbee Alliance Std*, 2004.

[139] S. Bluetooth, "Bluetooth specification version 1.1," *Available HTTP: http://www. bluetooth. com*, 2001.

[140] X. Lin, J. Andrews, A. Ghosh, and R. Ratasuk, "An Overview of 3GPP Device-to-Device Proximity Services," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 40–48, Apr. 2014.

[141] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and Uplink Decoupling: A Disruptive Architectural Design for 5G Networks," in *IEEE GLOBECOM*, December 2014, pp. 1798–1803.

[142] 3GPP TR 36.843, "Technical Specification Group Radio Access Network; Study on LTE device to device proximity services," Tech. Rep., March 2014.

[143] C. Liu, P. Whiting, and S. Hanly, "Joint resource allocation and user association in downlink three-tier heterogeneous networks," in *IEEE GLOBECOM*, December 2014, pp. 4232–4238.

[144] C. Vlachos and V. Friderikos, "Optimal Device-to-Device cell association and load balancing," in *IEEE ICC*, June 2015, pp. 5441–5447.

[145] 3GPP TS 36.213, "Evolved Universal Terrestrial Radio Access (EUTRA) Physical Layer Procedures (Release 10), v10.0.1," Tech. Rep., Dec. 2010.

[146] T. Ihalainen, P. Janis, Z. Li, M. Moisio, V. Nurmela, M. Uusitalo, C. Wijting, and O. Yilmaz, "Flexible scalable solutions for dense small cell networks," in *WWRF*, April 2013.

[147] O. Yilmaz, Z. Li, K. Valkealahti, M. Uusitalo, M. Moisio, P. Lunden, and C. Wijting, "Smart mobility management for D2D communications in 5G networks," in *IEEE WCNC Workshops*, April 2014, pp. 219–223.

[148] J. Johnson and Y. Rahmat-Samii, "Genetic Algorithm Optimization of Wireless Communication Networks," *IEEE Antennas and Propagation Society International Symposium*, Jun. 1995.

[149] Y. Zhenhua, Y. Guangwen, L. Shanwei, and Z. Qishan, "A Modified Immune Genetic Algorithm for Channel Assignment Problems in Cellular Radio Networks," in *Int. Conference on Intelligent System Design and Engineering Application (ISDEA)*, vol. 2, Oct. 2010, pp. 823–826.

[150] M. A. C. Lima, A. F. R. Araujo, and A. C. Cesar, "Adaptive genetic algorithms for dynamic channel assignment in mobile cellular communication systems," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2685–2696, Sep. 2007.

[151] T. Fang and L.-P. Chau, "GOP-based channel rate allocation using genetic algorithm for scalable video streaming over error-prone networks," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1323–1330, Jun. 2006.

[152] S. Lembo, A. Dowhuszko, and O. Tirkkonen, "Joint resource allocation in mobile networks with macro cellular and Device-to-Device communication," in *IEEE ICC Workshop*, Jun. 2015, pp. 585–590.

[153] L. Su, Y. Ji, P. Wang, and F. Liu, "Resource Allocation Using Particle Swarm Optimization for D2D Communication Underlay of Cellular Networks," in *IEEE WCNC*, Apr. 2013, pp. 129–133.

[154] S. Maghsudi and S. Stanczak, "Hybrid centralized-distributed resource allocation for Device-to-Device communication underlaying cellular networks," *IEEE Transactions on Vehicular Technology*, no. 99, Apr. 2015.

[155] M. Hasan and E. Hossain, "Distributed resource allocation for relay-aided Device-to-Device communication: A message passing approach," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6326–6341, Nov. 2014.

[156] T. Riihonen, S. Werner, and R. Wichman, "Comparison of Full-Duplex and Half-Duplex modes with a fixed amplify-and-forward relay," in *IEEE WCNC*, Apr. 2009, pp. 13–17.

[157] L. Badia, A. Botta, and L. Lenzini, "A genetic approach to joint routing and link scheduling for wireless mesh networks," *Ad Hoc Networks*, vol. 7, no. 4, pp. 654 – 664, 2009.

[158] K. Imai, *Lectures on Expectation and Functions of Random Variables*. Department of Politics, Princeton University, Mar. 2006.

Appendix A

Theoretical derivations

A.1 Appendix 1

Derivation of Lemma 1: Starting with the mmWave case, the propagation process $\mathcal{N}_s := \{L_s(x) = \|x\|^{\alpha_s(x)}\}$ on $\mathbb{R}^+$ for $x \in \Phi_s$ has intensity

$$\Lambda_s((0,t)] = \int_{\mathbb{R}^2} \mathbb{P}(L_s(x) < t)\mathrm{d}x = 2\pi\lambda_s \int_0^\infty \mathbb{P}(r^{\alpha_s(r)} < t) r \mathrm{d}r. \tag{A.1}$$

In the previous equation α is distance dependent as it has different values for LOS and NLOS links and according to the blockage model in Section 3.3.3 the intensity can be expressed as

$$\begin{aligned}
\Lambda_s((0,t)] &= 2\pi\lambda_s \Big(\omega \int_0^\mu r \mathbb{1}(r^{\alpha_l} < t)\mathrm{d}r + (1-\omega) \int_0^\mu r \mathbb{1}(r^{\alpha_n} < t)\mathrm{d}r \\
&\qquad + \int_\mu^\infty r \mathbb{1}(r^{\alpha_n} < t)\mathrm{d}r \Big) \\
&= 2\pi\lambda_s \Big(\omega \int_0^\mu r \mathbb{1}(r < t^{\frac{1}{\alpha_l}})\mathrm{d}r + (1-\omega) \int_0^\mu r \mathbb{1}(r < t^{\frac{1}{\alpha_n}})\mathrm{d}r \\
&\qquad + \int_\mu^\infty r \mathbb{1}(r < t^{\frac{1}{\alpha_n}})\mathrm{d}r \Big) \\
&= 2\pi\lambda_s \Bigg(\omega \int_0^{\min(\mu, t^{\frac{1}{\alpha_l}})} r\mathrm{d}r + (1-\omega) \int_0^{\min(\mu, t^{\frac{1}{\alpha_n}})} r\mathrm{d}r \\
&\qquad + \int_\mu^{t^{\frac{1}{\alpha_n}}} r \mathbb{1}(t^{\frac{1}{\alpha_n}} > \mu)\mathrm{d}r \Bigg).
\end{aligned} \tag{A.2}$$

Solving the integrals yields

$$\Lambda_s((0,t)] = \pi\lambda_s\Bigg(\omega\left(\mu^2\mathbb{1}(t>\mu^{\alpha_l}) + t^{\frac{2}{\alpha_l}}\mathbb{1}(t\leq\mu^{\alpha_l})\right) + (1-\omega)\Big(\mu^2\mathbb{1}(t>\mu^{\alpha_n}) + t^{\frac{2}{\alpha_l}}\mathbb{1}(t\leq\mu^{\alpha_n})\Big) + (t^{\frac{2}{\alpha_n}}-\mu^2)\mathbb{1}(t>\mu^{\alpha_n})\Bigg). \tag{A.3}$$

Finally, rearranging the terms yields the final expression for the pathloss process intensity in (3.31). For the sub-6GHz Mcells case, deriving the pathloss process intensity is straight forward as blockage is not considered for sub-6GHz. The propagation process $\mathcal{N}_m := \{L_m(x) = \|x\|^{\alpha_m}\}$ on $\mathbb{R}^+$ for $x \in \Phi_m$ has intensity

$$\begin{aligned}\Lambda_m((0,t)] &= \int_0^\infty \mathbb{P}(L_m(x)<t)\mathrm{d}x = 2\pi\lambda_m\int_0^\infty \mathbb{P}(r^{\alpha_m}<t)r\mathrm{d}r \\ &= 2\pi\lambda_m\int_0^{t^{\frac{1}{\alpha_m}}} r\mathbb{1}(r<t^{\frac{1}{\alpha_m}})\mathrm{d}r = \pi\lambda_m t^{\frac{2}{\alpha_m}}.\end{aligned} \tag{A.4}$$

■

A.2 Appendix 2

Derivation of Lemma 3: The derivation of the Max-Rate association probabilities starts with the downlink association probability to a sub-6GHz Mcell $\mathcal{B}_{DL,m}$ which is given by

$$\begin{aligned}\mathcal{B}_{\mathrm{DL},m} &= \mathbb{P}\left(\mathtt{SIR}_{\mathrm{DL},m} > (1+\mathtt{SNR}_{\mathrm{DL},s})^{\left(\frac{\mathrm{W}_s}{\mathrm{W}_m}\right)} - 1\right) \\ &= \mathbb{E}_{(\mathtt{SNR}_{\mathrm{DL},s}=S)}\left[\bar{F}_{\mathtt{SIR}_{\mathrm{DL},m}}\left((1+S)^{\frac{\mathrm{W}_s}{\mathrm{W}_m}} - 1\right)\right] \\ &= \int_0^\infty f_{\mathtt{SNR}_{\mathrm{DL},s}}(z)\bar{F}_{\mathtt{SIR}_{\mathrm{DL},m}}\left((1+z)^{\frac{\mathrm{W}_s}{\mathrm{W}_m}} - 1\right)\mathrm{d}z,\end{aligned} \tag{A.5}$$

where $\bar{F}_{\mathtt{SIR}_{\mathrm{DL},m}}(k)$ is the DL coverage probability $\mathcal{P}(\mathtt{SIR} > k)$ and $f_{\mathtt{SNR}_{\mathrm{DL},s}}(z)$ is the PDF of $\mathtt{SNR}_{\mathrm{DL},s}$. For $\bar{F}_{\mathtt{SIR}_{\mathrm{DL},m}}(k)$ we use the expression derived in [69] for the coverage probability in the no noise and exponential fading case which is given by

$$\bar{F}_{\mathtt{SIR}_{\mathrm{DL},m}}(t) = \frac{1}{1+\rho(t,\alpha_m)}, \tag{A.6}$$

where $\rho(t,\alpha_m) = t^{\frac{2}{\alpha_m}} \int\limits_{t^{\frac{-2}{\alpha_m}}}^{\infty} \frac{\mathrm{d}u}{1+u^{\frac{\alpha_m}{2}}}$. $f_{\mathtt{SNR}_{\mathrm{DL},s}}(z)$ can be derived from $\bar{F}_{\mathtt{SNR}_{\mathrm{DL},s}}(z)$ as follows

$$\begin{aligned}\bar{F}_{\mathtt{SNR}_{\mathrm{DL},s}}(z) &= \mathbb{P}(\mathtt{SNR}_{\mathrm{DL},s} > z) = \mathbb{P}\left(\frac{\mathrm{P}_s\psi_s h_{x^*,0} L_s(x^*)^{-1}}{\sigma_s^2} > z\right) \\ &= \int\limits_0^{\infty} \exp\left(\frac{-z\sigma_s^2 l}{\mathrm{P}_s\psi_s}\right) f_s(l)\mathrm{d}l. \end{aligned} \tag{A.7}$$

$$\begin{aligned} f_{\mathtt{SNR}_{\mathrm{DL},s}}(z) &= \frac{-\mathrm{d}\ \bar{F}_{\mathtt{SNR}_{\mathrm{DL},s}}(z)}{\mathrm{d}z} = \frac{-\mathrm{d}}{\mathrm{d}z}\int\limits_0^{\infty} \exp\left(\frac{-z\sigma_s^2 l}{\mathrm{P}_s\psi_s}\right) f_s(l)\mathrm{d}l \\ &= \frac{-\mathrm{d}}{dz}\int\limits_0^{\infty} f_s(z,l)\mathrm{d}l, \end{aligned} \tag{A.8}$$

where $f_s(l)$ is given in (3.35). In order to simplify the previous expression we exchange the order of the differentiation and integral using Leibnitz Rule [158]. The following two conditions need to be satisfied in order for this rule to be applicable.

- $|\frac{\mathrm{d}f_s(z,l)}{\mathrm{d}z}| \leq g(l)$, which means that the LHS expression is differentiable.
- $\int\limits_0^{\infty} g(z,l)\mathrm{d}z < \infty$, $g(l)$ is defined below.

$$|\frac{\mathrm{d}f_s(z,l)}{\mathrm{d}z}| \leq \frac{\sigma_s^2}{\mathrm{P}_s\psi_s}\ l\ f_s(l) = g(z,l), \tag{A.9}$$

which satisfies the first condition.

$$\begin{aligned}\int\limits_0^{\infty} g(l)\mathrm{d}l &= \frac{\sigma_s^2}{\mathrm{P}_s\psi_s}\int\limits_0^{\infty} l f_s(l)\mathrm{d}l = \frac{\sigma_s^2}{\mathrm{P}_s\psi_s}\int\limits_0^{\infty}\mathbb{P}(L_s > x)\mathrm{d}x = \frac{\sigma_s^2}{\mathrm{P}_s\psi_s}\int\limits_0^{\infty} e^{-\Lambda_s(0,x)}\mathrm{d}x \\ &\overset{(a)}{=} \int\limits_0^{\mu^{\alpha_l}} e^{-\Lambda_s(0,x)}\mathrm{d}x + \int\limits_{\mu^{\alpha_l}}^{\mu^{\alpha_n}} e^{-\Lambda_s(0,x)}\mathrm{d}x + \int\limits_{\mu^{\alpha_n}}^{\infty} e^{-\Lambda_s(0,x)}\mathrm{d}x \\ &= \text{constant} + \int\limits_{\mu^{\alpha_n}}^{\infty} e^{-x^{\frac{2}{\alpha_n}}}\mathrm{d}x, \end{aligned} \tag{A.10}$$

where (a) follows from (3.31). The first two integrals are bounded so we examine the third integral. Typically, $\alpha_n \leq 10$, therefore

$$\begin{aligned}\int_{\mu^{\alpha_n}}^{\infty} e^{-x^{\frac{2}{\alpha_n}}} \, \mathrm{d}x &\leq \int_{\mu^{\alpha_n}}^{\infty} e^{-x^{0.2}} \\ \int_{\mu^{\alpha_n}}^{\infty} e^{-x^{0.2}} &= 5e^{-x^{0.2}}(\mu^{0.8\alpha_n} + 4\mu^{0.6\alpha_n} + 12\mu^{0.4\alpha_n} + 24\mu^{0.2\alpha_n} + 24) < \infty, \quad \text{(A.11)}\end{aligned}$$

which satisfies the second conditions and allows us to write $f_{\mathtt{SNR}_{DL,s}}(z)$ as follows

$$\begin{aligned} f_{\mathtt{SNR}_{\mathrm{DL},s}}(z) &= \frac{-\mathrm{d}}{\mathrm{d}z} \int_0^{\infty} \exp\left(\frac{-z\sigma_s^2 l}{\mathrm{P}_s \psi_s}\right) f_s(l)\mathrm{d}l \\ &= -\int_0^{\infty} \frac{\mathrm{d}}{\mathrm{d}z} \exp\left(\frac{-z\sigma_s^2 l}{\mathrm{P}_s \psi_s}\right) f_s(l)\mathrm{d}l \\ &= \frac{\sigma_s^2}{\mathrm{P}_s \psi_s} \int_0^{\infty} l \exp\left(\frac{-z\sigma_s^2 l}{\mathrm{P}_\mathrm{s} \psi_\mathrm{s}}\right) f_s(l)\mathrm{d}l. \end{aligned} \quad \text{(A.12)}$$

which concludes this proof and $\mathcal{B}_{UL,m}$ is derived similarly. ■

A.3 Appendix 3

Derivation of Theorem 1: The DL SINR coverage for sub-6GHz Mcells is first derived. As shown in Section 3.3.4, the condition for association to a Mcell in the DL is $L_{\min,s} > a_{DL} L_{\min,m}$.

$$\begin{aligned} \mathcal{P}_{DL,m}(\tau) &= \mathbb{P}(\mathtt{SINR}_{\mathrm{DL},m} > \tau; K_{\mathrm{DL}} = m) \\ &= \int_0^{\infty} \mathbb{P}(\mathtt{SINR}_{\mathrm{DL},m} > \tau; K_{\mathrm{DL}} = m | L_{\min,m} = l) \; f_m(l)\mathrm{d}l \\ &= \int_0^{\infty} \mathbb{P}(\mathtt{SINR}_{\mathrm{DL},m} > \tau; L_{\min,s} > a_{\mathrm{DL}} l | L_{\min,m} = l) \; f_m(l)\mathrm{d}l \\ &\overset{(a)}{=} \int_0^{\infty} \mathbb{P}(\mathtt{SINR}_{\mathrm{DL},m} > \tau | L_{\min,m} = l) \; \mathbb{P}(L_s > a_{\mathrm{DL}} l | L_{\min,m} = l) \; f_m(l)\mathrm{d}l \\ &= \int_0^{\infty} \mathbb{P}(\mathtt{SINR}_{\mathrm{DL},m} > \tau | L_{\min,m} = l) \bar{F}_s(a_{\mathrm{DL}} l) f_m(l)\mathrm{d}l, \end{aligned} \quad \text{(A.13)}$$

where (a) follows from the assumption that Φ_s and Φ_m are independent. Since $\bar{F}_s$ and f_m are known, we now derive the first part of the equation which is given by

$$\begin{aligned}\mathbb{P}(\texttt{SINR}_{\mathrm{DL},m} > \tau | L_{\min,m} = l) &= \mathbb{P}\left(\frac{\mathrm{P}_m \psi_m h_{x^*,0} l^{-1}}{I + \sigma_m^2} > \tau | L_{\min,m} = l\right) \\ &= \exp\left(\frac{-\tau\sigma_m^2 l}{\mathrm{P}_m\psi_m}\right) \mathbb{E}_I\left[\exp\left(\frac{-\tau I l}{\mathrm{P}_m\psi_m}\right)\right] \\ &= \exp\left(\frac{-\tau\sigma_m^2 l}{\mathrm{P}_m\psi_m}\right) \mathcal{L}_{I_l}\left(\frac{-\tau l}{\mathrm{P}_m\psi_m}\right),\end{aligned} \tag{A.14}$$

where $\mathcal{L}_I(s)$ is the Laplace transform of the interference and is given by

$$\begin{aligned}\mathcal{L}_{I_l}(s) &= \mathbb{E}[e^{-sI}] = \mathbb{E}_{\Phi,h_{x,0}}\left[\exp\left(-s \sum_{x\in\Phi_m\setminus x^*} \mathrm{P}_m\psi_m h_{x,0} L_m(x)^{-1}\right)\right] \\ &= \mathbb{E}_{\Phi,h_{x,0}}\left[\prod_{x\in\Phi_m\setminus x^*} \exp\left(-s\mathrm{P}_m\psi_m h_{x,0} L_m(x)^{-1}\right)\right] \\ &= \exp\left(-\int_l^\infty \left(1 - \mathbb{E}_{h_{x,0}}\left[\exp\left(-s\mathrm{P}_m\psi_m h_{x,0} t^{-1}\right)\right]\right) \Lambda_m(\mathrm{d}t)\right),\end{aligned} \tag{A.15}$$

where $\Lambda_m(\mathrm{d}t)$ is given by deriving the expression in (3.32) with respect to t.

$$\mathcal{L}_{I_l}(s) = \exp\left(\frac{-2\pi\lambda_m}{\alpha_m} \int_l^\infty \frac{t^{\frac{2}{\alpha_m}-1}}{1 + \frac{t}{s\mathrm{P}_m\psi_m}} \mathrm{d}t\right). \tag{A.16}$$

Finally,

$$\mathcal{P}_{\mathrm{DL},m}(\tau) = \int_0^\infty \exp\left(\frac{-\tau\sigma_m^2 l}{\mathrm{P}_m\psi_m}\right) \exp\left(\frac{-2\pi\lambda_m}{\alpha_m} \int_l^\infty \frac{t^{\frac{2}{\alpha_m}-1}}{1 + \frac{t}{\tau l}} dt\right) \bar{F}_s(a_{\mathrm{DL}} l) f_m(l) \mathrm{d}l. \tag{A.17}$$

$\mathcal{P}_{\mathrm{UL},m}$ is derived in the same way replacing P_m and a_{DL} by P_{um} and a_{UL} respectively.

We then derive the DL SNR coverage for mmWave Scells. Similarly, the condition for association to a Scell in the DL is $L_{\min,m} > \frac{L_{\min,s}}{a_{\mathrm{DL}}}$.

$$\begin{aligned}\mathcal{P}_{\mathrm{DL},s}(\tau) &= \mathbb{P}(\texttt{SINR}_{\mathrm{DL},s} > \tau; K_{\mathrm{DL}} = s) \\ &= \int_0^\infty \mathbb{P}(\texttt{SINR}_{\mathrm{DL},s} > \tau; K_{\mathrm{DL}} = s | L_{\min,s} = l)\ f_s(l) \mathrm{d}l \\ &= \int_0^\infty \mathbb{P}(\texttt{SINR}_{\mathrm{DL},s} > \tau | L_{\min,s} = l) \bar{F}_m\left(\frac{l}{a_{\mathrm{DL}}}\right) f_s(l) \mathrm{d}l.\end{aligned} \tag{A.18}$$

$$\begin{aligned}\mathbb{P}(\mathtt{SNR}_{\mathrm{DL},s} > \tau | L_{\min,s} = l) &= \mathbb{P}\left(\frac{\mathrm{P_s}\psi_s h_{x^*,0} l^{-1}}{\sigma_s^2} > \tau | L_{\min,s} = l\right) \\ &= \mathbb{P}\left(h_{x^*,0} > \frac{-\tau\sigma_s^2 l}{\mathrm{P_s}\psi_s} | L_{\min,s} = l\right) \\ &= \exp\left(\frac{-\tau\sigma_s^2 l}{\mathrm{P_s}\psi_s}\right).\end{aligned} \tag{A.19}$$

Finally,

$$\mathcal{P}_{\mathrm{DL},s}(\tau) = \int_0^\infty \exp\left(\frac{-\tau\sigma_s^2 l}{\mathrm{P_s}\psi_s}\right) \bar{F}_m\left(\frac{l}{a_{\mathrm{DL}}}\right) f_s(l)\mathrm{d}l. \tag{A.20}$$

$\mathcal{P}_{\mathrm{UL},s}(\tau)$ is derived similarly by exchanging P_s and a_{DL} by P_{us} and a_{UL} respectively in the previous derivation. ■

Appendix B

Related publications

B.1 Journal papers

1. H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and uplink cell association with traditional macrocells and millimeter wave small cells,", *Submitted to IEEE Trans. Wireless Commun.*, Available: http://arxiv.org/abs/1601.05281, 2016.

2. F. Boccardi, J. Andrews, H. Elshaer, M. Dohler, S. Parkvall, P. Popovski, and S. Singh, "Why to decouple the uplink and downlink in cellular networks and how to do it," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 110–117, Mar. 2016.

B.2 Conference papers

1. H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and uplink decoupling: A disruptive architectural design for 5G networks," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2014.

2. H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Load and backhaul aware decoupled downlink/uplink access in 5G systems," in *2015 IEEE International Conference on Communications (ICC)*, Jun. 2015.

3. H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and uplink cell association in sub-6GHz and millimeter wave 5G heterogeneous networks," *Submitted to IEEE Global Communications Conference (GLOBECOM)*, Dec. 2016.

4. H. Elshaer, C. Vlachos, V. Friderikos, and M. Dohler, "Interference-aware decoupled cell association in device-to-device based 5G networks," *IEEE Vehicular Technology Conference (VTC) Spring*, May 2016.

5. C. Vlachos, H. Elshaer, V. Friderikos, and M. Dohler, "Bio-inspired resource allocation for relay-aided Device-to-Device communications," *Accepted in IEEE Vehicular Technology Conference (VTC) Fall*, Sep. 2016.

6. E. Liotou, H. Elshaer, R. Schatz, R. Irmer, M. Dohler, N. Passas, and L. Merakos, "Shaping QoE in the 5G ecosystem," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, May 2015.

B.3 Book chapters

1. H. Elshaer, M. A. Lema, T. Mahmoodi, M. Dohler, "Chapter 8: Decoupled Uplink and Downlink Access in Heterogeneous Networks". 5G Wireless Technologies, *IET press, to be published.*

B.4 Patents

1. H. Elshaer, N. Scully, L. Anaya. "Carrier aggregation mode selection." GB 1412266.7.

2. H. Elshaer, N. Scully, L. Anaya. "Resource management for inter-E-NodeB carrier aggregation." GB 1412264.2

3. H. Elshaer, N. Scully, L. Anaya. "Carrier aggregation activation". GB 1412265.9

4. H. Elshaer, E. Bouton, F. Boccardi, I. Thibault. "UL Cell Selection criterion for nodes with different MIMO configuration". GB 1407726.7.

5. H. Elshaer, E. Bouton, F. Boccardi, I. Thibault. "Decoupled access with load balancing and non-ideal backhaul". GB 1420131.3.

6. H. Elshaer, E. Bouton, F. Boccardi, I. Thibault. "Arrangement for choosing transceiver nodes in a D2D enabled Mobile Telecommunications Network." GB 1417366.0.

7. H. Elshaer, S. Subramani. "Uplink-optimized cell selection/ reselection in low complexity IoT networks." GB1602592.6.

B.5 Awards

1. Best paper award at the IEEE Globecom conference 2014 for H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and uplink decoupling: A disruptive architectural design for 5G networks,".

2. Recipient of the FP7 MITN Marie-Curie Fellowship.

3. Finalist at the EPSRC UK ICT Pioneers Competition 2015.