Millimeter Wave Backhaul for Ultra-Dense Wireless Networks — Analysis of Plug and Play Implementations

Arttu Hellsten

Thesis supervisor:

> Prof. Jyri Hämäläinen

Thesis advisors:

> D.Sc. (Tech.) Kari Leppänen

> Ph.D. (Tech.) Tao Cai

Aalto University
School of Electrical
Engineering

Author: Arttu Hellsten

Title: Millimeter Wave Backhaul for Ultra-Dense Wireless Networks — Analysis of Plug and Play Implementations

Date: 23.11.2015   Language: English   Number of pages: 9+99

Department of Communications and Networking

Professorship: Communications Engineering (S-72)

Supervisor: Prof. Jyri Hämäläinen

Advisors: D.Sc. (Tech.) Kari Leppänen, Ph.D. (Tech.) Tao Cai

The amount of wireless traffic and number of connected devices are expected to explode in the coming future. By the year 2020 the amount of data traffic is forecasted to grow 1000 times from 2010 levels and the amount of connected devices is expected to reach 50 billion. One reason to these numbers is massive increase in machine type communications. 5G networks have been envisioned to address these challenges.

In the 5G network concept the networks are getting denser than ever before. Millimeter wave communications play an important role in backhauling of the mobile traffic as deploying optical fiber to every small node is most likely going to be too cost intensive for operators.

Efficient deployment of an ultra-dense wireless network requires that the devices support so called "plug and play" installation. In practice it means that a mechanic installing a new radio node should only perform physical mounting of the device. Antenna alignment and link setup processes should be fully automated.

The purpose of this thesis is to study practical issues and possible solutions of realizing the plug and play installation in a cost efficient way. This study will define scenarios and functional requirements of adding access nodes to backhaul networks. Technical evaluation of link discovery process and cost analysis on plug and play installation of access nodes are conducted.

Tekijä: Arttu Hellsten

Työn nimi: Millimetriaaltopohjainen runkoyhteys ultratiheille langattomille verkoille – Itseasentuvien verkkoelementtien analyysi

Päivämäärä: 23.11.2015   Kieli: Englanti   Sivumäärä: 9+99

Tietoliikenne- ja tietoverkkotekniikan laitos

Professuuri: Tietoliikennetekniikka (S-72)

Työn valvoja: Prof. Jyri Hämäläinen

Työn ohjaajat: TkT Kari Leppänen, FT (tekn.) Tao Cai

Tulevaisuudessa langattomien laitteiden määrän ja niiden generoiman liikenteen odotetaan kasvavan räjähdysmäisesti. Vuoteen 2020 mennessä verkoissa siirrettävien datamäärien on ennustettu kasvavan tuhatkertaisiksi vuoden 2010 tasosta, ja liitettyjen laitteiden määrän odotetaan nousevan 50 miljardiin. Yksi syy näihin on koneiden välisen viestinnän massiivinen kasvu. 5G-verkkoja on kaavailtu vastaamaan edellä mainittuihin haasteisiin.

Osana 5G-konseptia verkkojen odotetaan rakentuvan tiheämmiksi kuin koskaan aiemmin. Millimetriaaltoihin pohjautuvat linkit tulevat olemaan merkittävässä roolissa mobiilidatan siirtämisessä radionoodeista runkoverkkoon, koska optisen kuidun rakentaminen jokaiselle pienelle noodille tulisi mitä luultavimmin operaattoreille liian kalliiksi.

Ultratiheiden verkkojen tehokas rakentaminen vaatii, että asennettavat laitteet tukevan niin sanottua itseasennusta. Käytännössä se tarkoittaa sitä, että asentajan täytyy ainoastaan suorittaa radionoodin fyysinen asennus. Antenniensuuntaus- ja linkinmuodostusprosessien tulisi olla täysin automatisoituja.

Tämän diplomityön tarkoituksena on tutkia kustannustehokkaaseen itseasennukseen liittyviä käytännön ongelmia sekä mahdollisia ratkaisuita. Tutkimus määrittelee skenaariot ja funktionaaliset vaatimukset radionoodien lisäämiseksi osaksi operaattorin verkkoa. Työ sisältää linkkienmuodostusprosessin teknisen evaluoinnin, sekä kustannusanalyysin tiheiden verkkojen rakentamisesta hyödyntäen itseasennustekniikkaa.

Avainsanat: 5G, millimetriaallot, mobiiliverkon runkoyhteys, itseasentuvat verkkoelementit, ultratiheät verkot

Preface

I was given a unique opportunity to write my thesis at Huawei Finland R&D center. During the past almost six months I gained several new friends, and also finally finished my thesis.

I would like to thank Professor Jyri Hämäläinen for supervising my thesis. His valuable comments and flexibility in the review process has helped me to finish the thesis before end of the year.

I would also like to thank my advisors D.Sc. Kari Leppänen and Ph.D. Tao Cai from Huawei 5G R&D group for offering me such an interesting topic for the thesis. Their support in my daily writing process has been invaluable. Also many thanks to my new friends Andrey, Jussi, Kari H., Mário, Petteri and Philip from the Huawei Finland R&D 5G team for supporting and advising me in this project.

Last but definitely not least, I would like to thank all my friends and family for supporting me – especially my dear wife Salla. I couldn't have made it this far without her.

Helsinki, 23.11.2015

Arttu Hellsten

Contents

List of abbreviations

3GPP	3rd Generation Partnership Project
ADC	Analog-to-Digital Converter
ADEV	Allan Deviation
AoA	Angle of Arrival
AoD	Angle of Departure
AVAR	Allan Variance
BBU	Baseband Unit
BS	Base Station
CAPEX	Capital Expenditure
CDM	Code Division Multiplexing
CMOS	Complementary Metal-Oxide Semiconductor
CoMP	Coordinated Multi-Point
CPICH	Common Pilot Channel
CPRI	Common Public Radio Interface
C-RAN	Cloud Radio Access Network
CRC	Cyclic Redundancy Check
CSFB	Circuit Switched Fallback
CSI	Channel State Information
D2D	Device-to-Device
DAC	Digital-to-Analog Converter
DAS	Distributed Antenna System
EDGE	Enhanced Data rates for GSM Evolution
EHF	Extremely High Frequency (30–300 GHz range)
EIRP	Effective Isotropic Radiated Power
FDD	Frequency Division Duplexing
FDM	Frequency Division Multiplexing
FDMA	Frequency Division Multiple Access
FFT	Fast Fourier Transform
GNSS	Global Navigation Satellite System
GPP	General Purpose Processor
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
HetNet	Heterogeneous Network
IC	Integrated Circuit
ICIC	Inter-Cell Interference Cancellation
IEEE	Institute of Electrical and Electronics Engineers
IMT	International Mobile Telecommunications
IoT	Internet of Things
ISD	Inter-Site Distance

ISI	Inter-Symbol Interference
ITU	International Telecommunication Union
ITU-R	ITU Radiocommunications sector
LDPC	Low-Density Parity Check
LOS	Line of Sight
LTE	Long Term Evolution
LTE-A	Long Term Evolution - Advanced
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MEC	Moving Extended Cell
MIMO	Multiple Input Multiple Output
M-MIMO	Massive MIMO
MTC	Machine-Type Communications
mMTC	Massive Machine-Type Communications
NFC	Near Field Communication
NGMN	Next Generation Mobile Network
NMT	Nordic Mobile Telephone
NOC	Network Operations Center
OBSAI	Open Base Station Architecture Initiative
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operational Expenditure
OSC	On-Site-Coding
O&M	Operations and Maintenance
PA	Power Amplifier
PAPR	Peak-to-Average Power Ratio
PHY	Physical Layer
QAM	Quadrature Amplitude Modulation
QoE	Quality of Experience
QPSK	Quadrature Phase Shift Keying
RAN	Radio Access Network
RAT	Radio Access Technology
RoF	Radio over Fiber
RRM	Radio Resource Manager
RRU	Remote Radio Unit
RS	Reed-Solomon
SCBT	Single Carrier Block Transmission
SC-FDE	Single Carrier Frequency Domain Equalization
SIM	Subscriber Identity Module
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio

TCO	Total Cost of Ownership
TDD	Time Division Duplexing
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
UDN	Ultra-dense network
UE	User Equipment
uMTC	Ultra-reliable Machine Type Communications
UMTS	Universal Mobile Telecommunications System
UWB	Ultra Wide Band
V2X	Vehicle-to-X
VCO	Voltage-Controlled Oscillator
VoIP	Voice over IP
VoLTE	Voice over LTE
WCDMA	Wideband Code Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network
xMBB	Extreme Mobile Broadband
ZF	Zero Forcing

1 Introduction

1.1 Motivation

The amount of wireless traffic and number of connected devices are expected to explode in the coming future. As shown in figure 1, by the year 2020 the amount of data traffic is forecasted to grow 1000 times from 2010 levels and the amount of connected devices is expected to reach 50 billion. One reason to these numbers is massive increase in machine type communications. Also the peak data rates are expected to reach gigabit levels. Diversified use cases will arise from the increasing coexistence of human type and machine type communications. The study of these future networks in telecommunications industry is referring to "beyond 2020" technologies, which is also referred as 5G wireless network technology. [1]
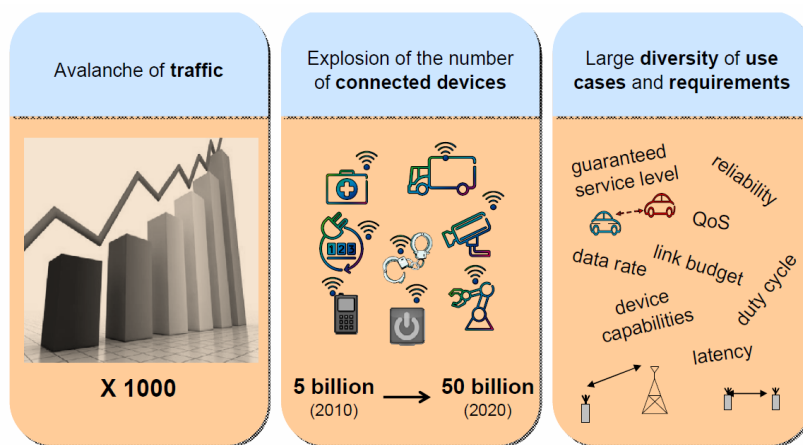
Figure 1: Forecast of development of communications requirements in the upcoming decade [1].

Today's wireless networks must be renewed completely in order to cope with the increasing amount of data traffic and the number of connected nodes. The performance of the most recently deployed technologies such as LTE-A (Long Term Evolution - Advanced) are far from those that are going to be needed to meet the future requirements.

In the 5G network concept the networks are getting denser than ever before. Millimeter wave communications play an important role in backhauling of the mobile traffic: Deploying optical fiber to every small node is most likely going to be too cost intensive for operators. Millimeter wave systems operating for example at 70-80GHz (E-Band) area can provide high throughput with very high frequency reuse by utilizing very narrow beams. The downside of narrow beams is the requirement of very careful alignment of the link while installing.

Efficient deployment of an ultra-dense wireless network (UDN) requires that the devices support so called "plug and play" installation. In practice it means that a mechanic installing a new radio node should only perform physical mounting

of the device and attaching power cabling. Antenna alignment and link setup processes should be fully automated. Also secure link establishment, positioning of the equipment and attaching as integral part of the radio access network (RAN) must be automated.

1.2 Purpose and Scope

As stated by Next generation mobile network (NGMN) Alliance in their white paper, the backhaul link setup of small cells should be plug and play. That means the small cells should be automatically connected to the a backhaul node which is within their reach [2]. The purpose of this thesis is to study practical issues and possible solutions of realizing the plug and play installation in a cost efficient way. This study will define scenarios and functional requirements of adding access nodes to backhaul networks. Possible functional requirements related to plug and play installation are listed below:

1. How an access node is registered to be owned by a certain operator before starting radio link discovery process?

2. How the backhaul node and an access node find each other, what kind of pilot signals are required, and are there some special hardware requirements?

3. How an access node securely pairs with a found backhaul node and establishes secure channel? What kind of security modules/functions should be embedded into the access and backhaul nodes?

4. How mutual time synchronization is performed between access nodes in a certain UDN area?

5. How to position an access node with very high accuracy? Are additional components required (e.g. a satellite navigation receiver)?

6. How the access node registers as part of the whole access network and how the necessary radio resource and other parameters are assigned from the radio network resource controller?

Scenarios of access node deployment are dependent on the existing status of the network:

- First node installation scenario (no existing load in the network segment)

- Network expansion scenario (existing user traffic in the network segment)

- (Optionally) Multi-hop backhaul scenario (no direct path between access node and aggregation node – access node will connect its backhaul via neighboring access node)

A techno-economic study will be conducted on plug and play installation of access nodes. The study will first focus on technical requirements and their possible solutions, followed by a cost analysis of deploying an ultra-dense network. Effects of the proposed plug and play functionality on the network deployment costs will be presented.

The questions this thesis is about to answer are:

- What are the technical options for implementing a plug and play deployment which requires the least amount of manual work?

- How different technical options relate to the total cost of the deployment of a 5G UDN? Both CAPEX (capital expenditure, e.g. hardware installation costs) and OPEX (operational expenditure, e.g. maintenance cost over 5 years) are considered.

1.3 Structure of the thesis

Chapter 2 will introduce the necessary technological background in order to understand the study of this thesis. Chapter 3 describes the scenarios and functional requirements in detail and lists the challenges for which this thesis proposes solution candidates. Chapter 4 contains the technical study of solution candidates and analyzes the feasibility of a chosen solution candidate. Chapter 5 proposes a cost model for economical evaluation of deploying an ultra-dense network. Final conclusions are given in Chapter 6 with proposals for further study.

2 Background

This chapter will give the necessary background information for understanding the 5G concept and related technologies. Also comparison to currently used technologies will be provided in order to put the new concepts into perspective and to see how major changes the "beyond 2020" thinking is about to bring.

2.1 5G concept

There has been four generations of mobile networks before reaching the so called fifth generation. The term 5G is not yet fixed as the standardization is still in pre-standardization phase. Large players in the industry are investing more and more on the research of the next generation mobile networks which are expected to launch commercially around the year 2020. This chapter will introduce the evolution path towards fifth generation networks and the new concepts that will distinguish the fifth generation from the previous generations.

2.1.1 Mobile networks evolution

Mobile radio network technologies have been categorized into different generation systems. This section describes the evolution path from first generation mobile radio networks all the way to 5th generation networks. The so called zeroth generation systems have not been included here since the first real cellular networks came with the first generation systems. The generation of a mobile radio network is not always very clear as there are several incremental steps added to existing standards. The guidelines that define each generation has been stated in IMT (International Mobile Telecommunications) recommendations by ITU-R (International Telecommunications Union Radiocommunications Sector) working party 5D (WP-5D) starting from the third generation systems with "IMT-2000" [3, pp. 37].

First generation

First generation mobile radio networks are the first wireless networks utilizing cellular concept. They were based on analog technology operating on 150MHz and above. There were no international standards available – each continent (or even country) had their own 1st generation radio network system (where deployed). Some most known 1G standards were NMT (Nordic Mobile Telephone) in Nordics and eastern Europe, AMPS (Advanced Mobile Phone System) in USA, TACS (Total Access Communications System) in UK, and C-Netz in West Germany.

As an example, NMT was launched in 1981 and operated first on 450MHz spectrum for easy wide network deployment. Later it was expanded to 900MHz spectrum to increase the capacity by adding more channels. Multiple access method used in most 1st generation mobile radio networks was based on FDMA (Frequency Division Multiple Access), where mobile stations were allocated a specific frequency channel for a call. Initially there was no encryption on the air interface so it was possible to listen to the calls by using a sophisticated radio scanner. Later the air interface

security was added at least to the NMT standard in form of scrambling, but it was not as secure as the encryptions introduced in 2nd generation digital networks. [4] [5] [6]

Second generation

Second generation mobile networks were based on digital technology which improved the quality, security, and capacity of the network tremendously. Also the power consumption of mobile terminals was greatly reduced which improved the mobility. There were several competing 2nd generation standards, but GSM (Global System for Mobile communications) became the most widely spread of all.

The multiple access method in GSM was based on combination of FDMA and TDMA (Time Division Multiple Access), where the allocated spectrum was divided to narrow carriers, which then again were divided to time slots. These resources were allocated according to need which improved the spectral efficiency from 1st generation networks.The first commercial GSM network was launched by Finnish operator Radiolinja (nowadays Elisa) in 1991. Widespread adoption of GSM standard allowed global roaming agreements by which the mobiles could be used outside of the home operators' networks. Even today it has the biggest geographical coverage of all mobile radio networks.

GSM was first designed for traditional circuit switched speech traffic only, but later packet switching was added as extension. GPRS (General Packet Radio Service), which was considered as "2.5G" technology, brought packet switching to GSM with data rates reaching from 56kbps up to 172kbps. Later the data rates were further expanded by EDGE (Enhanced Data rates for GSM Evolution) standard, so called "2.75G" technology, which used higher modulation and coding schemes (MCS) combined with timeslot aggregation. The maximum theoretical net throughput in EDGE network could reach 384kbps. [4] [5]

Third generation

Third generation networks were defined by International Telecommunications Union (ITU) in recommendation called IMT-2000. The 3rd generation standards focused more on enhancing packet switching data services while maintaining the circuit switched speech services. Additional services were for example video calls. The most widely used 3G standard was published by a consortium called 3GPP (3rd Generation Partnership Project), whose WCDMA (Wideband Code Division Multiple Access) based system was deployed internationally by all major network vendors. The 3GPP variant was called Universal Mobile Telecommunications System (UMTS) as it was designed to be used worldwide.

The initial 3GPP Release 99 (release number refers to 1999) based 3G UMTS network changed from narrow FDMA based carriers to a wide 5 MHz spectrum block from which capacity was allocated to users simultaneously using orthogonal codes (instead of frequency channels). This technology allowed the data rates to initially reach 384kbps, but later so called 3.5G standards in later 3GPP releases increased the data rates to over 20Mbps and beyond using carrier aggregation of two 5 MHz carriers and MIMO (Multiple Input Multiple Output) technology. The core network

structure was using same elements as the 2G GPRS, most differences were only in the base station and radio network controllers. [4] [5]

Fourth generation

The fourth generation mobile radio networks were defined in IMT-Advanced recommendation by ITU. The main design principle of 4G systems was to be fully packet based – circuit switched voice was not existing anymore so voice has to be carried using Voice over IP (VoIP) technology. Other major evolution step was the bit rates and latency which were to be improved a lot from previous generations.

There were initially two competing 4G standards in the industry; WiMAX (Worldwide Interoperability for Microwave Access) by IEEE (Institute of Electrical and Electronics Engineers) and LTE (Long Term Evolution) by 3GPP. The initial versions of both standards were in fact categorized as 3.9G technologies, as they weren't fully compliant with the IMT-Advanced recommendation. Later on LTE prevailed as the winner becoming the mainstream 4G network technology deployed worldwide. Later releases of LTE called LTE-Advanced (LTE-A) integrated the functions defined in IMT-Advanced to officially be categorized as a 4G network, such as coordinated multipoint transmission (CoMP) and intercell interference cancellation (ICIC). The spectrum usage was further expanded from 3G networks 5 MHz up to 20 MHz in single channel, which can offer up to 150 Mbps downlink throughput for single user. It is possible to aggregate multiple 20 MHz blocks to increase the throughput even further (e.g. 2 x 20 MHz channels can offer 300Mbps). The multiple access method was selected to be OFDMA (Orthogonal Frequency Division Multiple Access), which utilizes a range of orthogonal narrow carriers inside the (e.g. 20 MHz) channel. There are different LTE variants available designed to meet different spectrum availabilities around the world. The mainstream variant is FDD (Frequency Division Duplexing) based system which is used almost everywhere. The other variant is TDD (Time Division Duplexing) based, which is mainly designed for China and other Asian countries. FDD based system uses separate frequency bands for uplink and downlink, where TDD uses single frequency band and divides uplink and downlink based on ratio of uplink and downlink timeslots. The benefit of TDD over FDD is flexible configuration of uplink and downlink ratio, but the downside is high synchronization requirement between base stations.

The global rollout of 4G networks has been rather slow, but the releasing of lower frequency bands for 4G networks usage, such as 800 MHz in Finland, has been able to speed up the geographical coverage increase. Voice service over 4G network has not been widely used, but operators tend to use a technology called Circuit Switched Fallback (CSFB), where the network instructs the handset to change from 4G network to 3G or 2G network. Fallback to 3G network will still enable simultaneous data while calling, but in case fallback to 2G, mobile data will be put on hold until the call has finished. In any case the data performance will be greatly affected during a call when CSFB is used. After the call the handset will eventually switch back to 4G network for better data performance. Voice over LTE (VoLTE) has been already deployed by some operators, but globally many operators are still planning or testing. [4] [7]

Fifth generation

The fifth generation of mobile networks is at the moment one of the hottest research topics in the telecommunications industry. The general idea of 5th generation networks is to provide advanced enough networks that are capable of meeting demands of year 2020 and beyond. Nobody really knows yet what the 5G will bring in reality, but the generally acknowledged fact is that even the latest IMT-Advanced based networks are not capable of scaling up enough in terms of bandwidth and amount of connected terminals to meet the requirements of beyond 2020. The main difference between 5G and earlier generations is that 5G goes beyond human communications and expands to basically anything that requires connection. It has been estimated that by year 2020 there would be about 50 billion connected devices [8]. The new key requirements are to be able to provide super-high bitrates, massive machine-to-machine communications (M2M), and ultra-reliable M2M communications to support the so called wireless Internet-of-Things (IoT).

The new IMT vision document is named as IMT-2020 into which all requirements of the next generation system will be collected. Besides ITU's IMT vision project, there are also several other research projects ongoing in parallel. An example is the EU-funded "Mobile and Wireless Communications Enablers for the 2020 Information Society" (METIS) project. First METIS project was running from beginning of November 2012 until end of April 2015. Following METIS II project will be running for 24 months starting from 1st of July 2015. The following sections will further introduce the 5G concept in detail. [9] [10]

2.1.2   Target metrics of 5G

The vision for future 5G networks would be to support ubiquitously connected society of beyond 2020. Anything that requires connectivity could be connected wirelessly. The requirements on the next generation network vary depending on the use cases and location.

General targets and service types of 5G

The overall technical objectives set by METIS project are to support following performance increments without raising energy and cost levels of current systems:

- 1000x higher mobile data volumes

- 10-100x higher typical end-user data rates

- 10-100x higher number of connected devices

- 10x longer battery life of low-power devices

- 5x lower End-to-End (E2E) latency

According to the METIS project, there are three generic service types of envisioned 5G networks (see figure 2: Extreme Mobile Broadband (xMBB), Massive Machine-Type Communications (mMTC) and Ultra-reliable Machine Type Communications (uMTC).
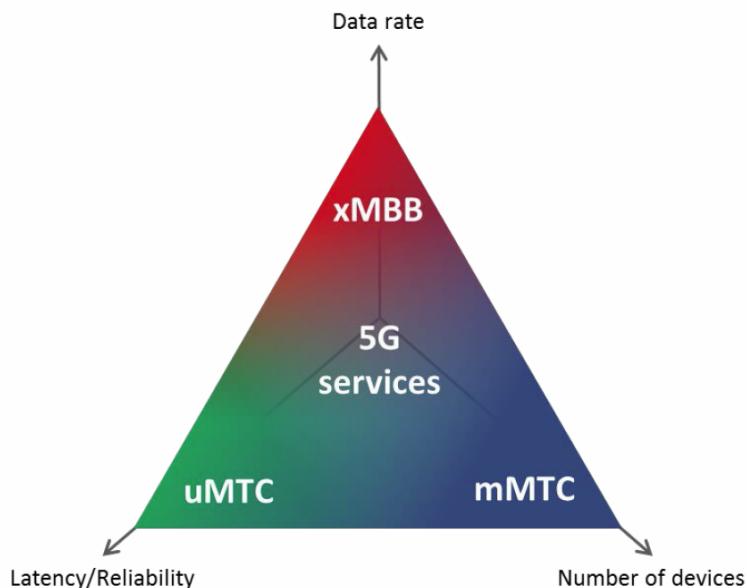
Figure 2: Three generic service types of 5G [11].

The target of xMBB is to enhance today's networks in both coverage and capacity by bringing end-user Quality of Experience (QoE) to a totally new level. Expanded capacity should address the peak data rates of end-users by raising them to 1-10 Gbps range and beyond. Meanwhile radio access network (RAN) latencies should drop to around millisecond level. An important aspect is to support more constant QoE across the whole coverage area by ensuring sufficiently high data rates (around 50-100 Mbps) at all times. Also in heavily loaded situations the network should be able to degrade the performance of users gracefully and uniformly instead of blocking some users totally from accessing the network.

The mMTC aims to scale the mobile networks to be able to support tens of billions of connected devices in the future. Wide outdoor coverage and enhanced indoor coverage are prioritized over high throughput target as opposed to xMBB. Battery life extension of small MTC nodes with lifetime reaching over 10 years requires optimization on the access network air interface and communication distances. Besides device-to-network connectivity, also device-to-device connectivity could be used to shorten transmission distances and improve coverage. Some devices could act as relay or aggregation nodes. This could effectively help saving power of the nodes requiring ultra long battery life.

uMTC is envisioned to provide ultra-reliable communications mostly for vehicle and industrial control applications. Vehicular communications are referred as Vehicle-to-X (V2X) communications, where X refers usually to vehicle as in V2V, pedestrians as in V2P, or infrastructure as in V2I. Interaction of (fast) moving vehicles with their surroundings requires ultra low latency and reliability. Device-to-device (D2D) communications play an important role complementing the network infrastructure in reducing latencies and increasing network resources by offloading traffic. Industrial

applications could comprise of control of stationary or moving equipment such as manufacturing or warehousing robots. Their requirements on delays may not be as critical as V2X communications due to slower velocities and less public working areas, but their reliability requirements are very high as well. [11]

5G network density metrics

NGMN Alliance has envisioned performance targets for the next generation networks which should be able to handle connection densities from thousands to hundreds of thousands per square kilometer. Also traffic densities are envisioned to grow even up to over ten terabits per second per square kilometer in ultra high indoor broadband access use case and also over terabit level in other dense broadband access cases. [2]

Network densification is one key enabler in supporting such vast amount of connections. Methods of densification can be for example cell splitting in terms of deploying more base stations in a given area. Cell splitting can be for instance based on use of distributed antenna systems (DAS), remote radio units (RRU) and small cells. Heterogeneous networks including macro cells complemented with small cells in traffic hot spots are estimated to play more important role in future. Ultra-dense networks (UDN) are an example of seamless coverage built using small cells or small RRUs in dense user areas. Also D2D communications can potentially increase spectral reuse and enhance QoE of proximity-based applications requiring high bandwidth and low latency. Besides adding more and expanding existing cells, spectral aggregation of new higher frequency bands and existing license free bands can provide much higher bandwidths compared with existing mobile radio networks. [12]

5G efficiency metrics

5G networks are expected to introduce enhancements in spectral efficiency compared with today's networks. The efficiency boost can be achieved for example by more efficient modulation and coding schemes and better filtering. Greater spatial reuse can enhance network area efficiency by means of Massive MIMO and UDN deployments. 5G also adds possibility of network infrastructure offloading by using D2D communications.

Enhanced signaling and radio resources control in especially xMTC cases are crucial as the networks will need to bear the signalling load from ever growing numbers of small MTC nodes. Also great battery life savings can be expected from reduced signalling which is crucial for the small MTC nodes.

Total Cost of Ownership (TCO) of 5G networks needs to be optimized in a way that the networks are able to be constructed and maintained in economically feasible way. Efficiency in network roll-out by means of automatic link establishment and optimization as well as automatic healing of run-time failures are essential aspects in minimizing the TCO. Also energy efficiency of 5G networks should be improved vastly in terms of bit/Joule so that the capacity and coverage targets could be met with about half of the energy costs than that of today's networks. This means roughly 2000 times of energy efficiency improvement. [2]

5G spectrum visions

In order to support the throughput and other targets set for the 5G networks, new frequency bands need to be allocated for the use of mobile networks. Pure spectral efficiency improvement in existing licensed mobile network bands will not be enough. Also differing requirements on the use cases of 5G require clever spectrum allocation. Thus "re-farming" of existing frequencies and allocation of new frequencies are essential in future. [2]

Existing frequencies allocated to IMT-based mobile networks operate on fairly low region of possible spectrum, mostly operating between 700 MHz and 2.6 GHz. Recent studies show that it would be possible to expand the spectrum usage up to millimeter wave area where the available spectrum is more than what is used together by all the different generations of mobile networks today. Most promising new frequency bands for access links in millimeter wave area are 28 GHz and 38 GHz bands – both offering over 1 GHz of bandwidth. The downside of using high frequencies is decreased range, but in indoor and urban environments dense radio network architecture and efficient use of smart antennas can compensate for the deficiencies. [13]

Flexible spectrum sharing between network operators has been proposed as one of the solutions for increased spectrum usage efficiency. Fixed allocation of frequency bands may lead to severe under-utilization of the sparse resources in lower frequency bands. Dynamic in-band resource sharing can potentially help raising the utilization ratio, but work on solving for example ways of handling fairness between operators and mitigation of inter-operator interference are still needed. Addition of unlicensed bands among the supported frequencies could also increase the network resources in shared manner. Especially 60 GHz millimeter wave band would provide roughly 7 GHz of shared spectrum. Short communication range of high frequency millimeter wave bands is not only a downside, but also a benefit, as it allows high possibility of coexistense of networks without severe interference leakage between operators. [14] [15]

Network densification in terms of adding more radio nodes by deployment of ultra-dense networks requires also backhaul densification. Constructing backhaul for the small radio nodes by traditional means of wired networks cannot support the set TCO optimization targets of 5G networks. Millimeter wave Line-of-Sight (LOS) links could prove to be essential in new types of wireless backhaul architectures. High spatial multiplexing and processing together with effective beamforming are among the beneficial properties of millimeter wave links. The availability of multiple gigahertz of bandwidth at millimeter wave area would be able to meet the ultra-high backhaul capacity requirements of the future networks. [12] [16]

Technology roadmap towards 5G

Roadmap towards 5G networks in the industry is presented in figure 3. ITU-R is working towards IMT-2020 requirements until beginning of 2017. 3GPP as the current main standardization body of 3G and 4G technologies is expected to continue in similar position in standardization of 5G. Specifications of 5G radio access are scheduled around end of 2018 and beginning of 2019. Roadmap presented by NGMN

Alliance targets trials of 5G networks around end of 2019 and commercially deployable solutions to be ready by end of 2020. [2] [11]
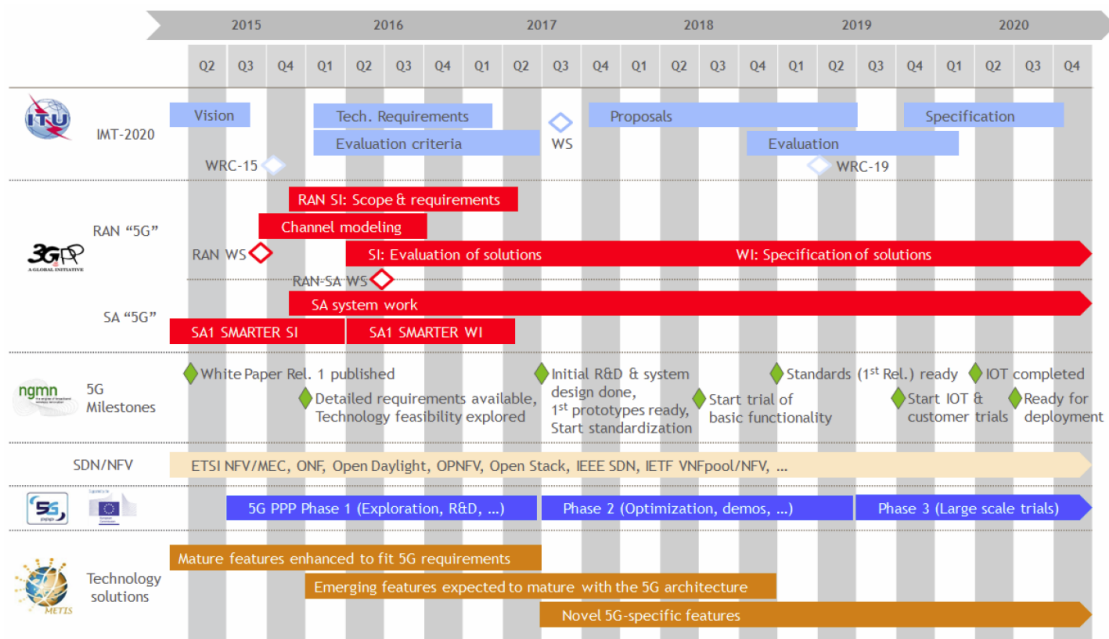

Figure 3: 5G roadmap including regulatory and standardization processes as well as industry milestone targets [11].

2.1.3 Heterogeneous Network concept

As stated in chapter 2.1.2, in order to handle the ever increasing end-user generated traffic loads, a possible way of providing more capacity to radio networks is network densification. Instead of building more and more high-power macro base stations, the introduction of small nodes such as micro, pico and femto-cells are considered as viable alternatives (see figure 4). The transmit powers of such nodes are much smaller than that of macro base stations, and they are deployed only locally based on needs of extra capacity or fixing a coverage hole in certain small areas. This leads to a network architecture called heterogeneous network (HetNet), where several types of serving nodes can work together to form seamless coverage for network users. A HetNet could consist of two or more tiers of different class of nodes. [17]

HetNet challenges
Traditional network planning methods utilize hexagonal macro-cellular grid and nearly uniform user distribution. In HetNets this kind of methodology is outdated, as one of the main purposes of constructing HetNets is the fact that users are not uniformly distributed. Instead, traffic hot spots where user densities are normally much higher are covered using small cells. Another legacy planning method is to assume that mobile terminals connect to base stations with highest signal to noise

ratio regardless of the distance. In HetNet deployment this often leads to suboptimal scenario in which the users in most cases connect to macro-cells due to their high transmit powers, even though nearby small cells with low transmit powers could provide much higher data rates or Quality of Experience. [17]
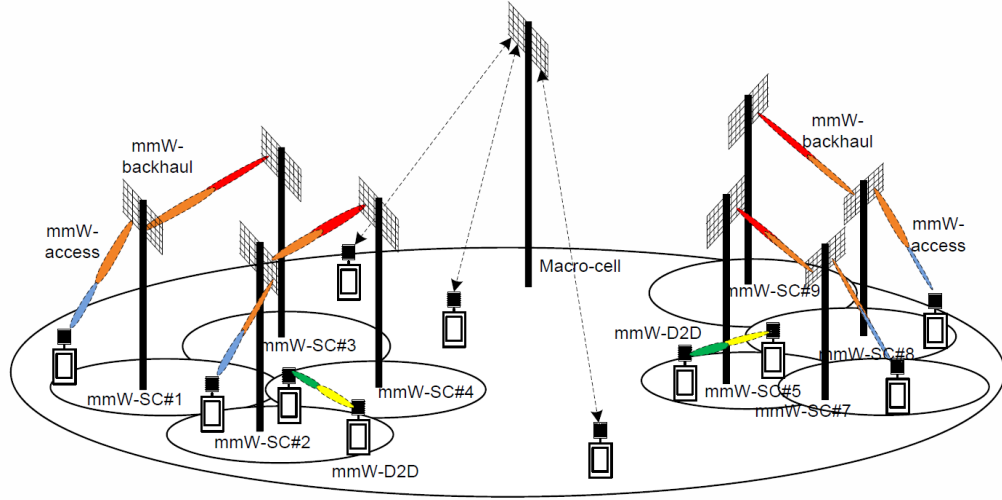

Figure 4: Example of a Heterogeneous network structure including millimeter wave systems [16].

Vehicular mobility introduces challenges for HetNet architecture. Vehicular users might experience highly variable data rates when rapidly passing through several cells. Especially handover planning and performing in HetNets is a challenging task for highly mobile users. The coverage areas of small cells are significantly smaller than in macro cells leading to increased handover requests in a network. When vehicular users are passing through multiple cells very rapidly, it has been noticed that handover failures have been increasing compared with traditional pure macro networks. User speed has been found to have significant impact on the handover performance. Especially HetNets without seamless small cell coverage suffer from macro-to-pico and pico-to-macro handovers which have been found to perform worse than pico-to-pico and macro-to-macro handovers. New handover algorithms have been studied and proposed to solve aforementioned problems. As an example, Park et al. proposed a "Two-Step Handover" algorithm which consists of early handover preparation and ping-pong avoidance functions. [18]

The envisioned Ultra-dense network (UDN) concept in 5G network architecture is likely be the next evolution step from HetNets in dense deployment areas. A possible solution to the issue of handovers due to mobility faced in HetNets could be solved by evolving into "Continuous UDNs". Continuous UDN would provide borderless experience to mobile terminals as traditional handovers would not be required within the area of the UDN. This concept is covered in further detail in chapter 2.1.4.

HetNet opportunities

Modern technologies such as self-organization and self-optimization of radio networks can help operators to easily increase the capacity of their networks. A large area can be covered by macro layer base stations, and they can be complemented with short range, low power, but high capacity small cells in areas where traffic requirements are considerably higher. The joint operation of the network nodes can be based on for example splitting control plane to macro layer and user plane to small cell. This ensures critical signaling (e.g. handovers) can be transmitted more reliably than in the small cell network which might have coverage holes. Consequently this will effectively reduce signaling towards core network. Signaling interference could be also reduced in HetNet as not all control information would have to be transmitted from the small cells, it would be sufficient to let the macro layer handle it. Other benefit of HetNet is possibility for separation of uplink and downlink transmissions. The downstream transmitting node can be selected based on the best signal conditions to maximize throughput, whereas uplink receiving node can be based on lowest pathloss reducing the required transmit power at the mobile node. [10]

2.1.4   Continuous ultra-dense networks

As mentioned in earlier sections, a network densification process is inevitable to be able to offer ultra high bandwidth in hotspot areas. Building an ultra-dense network (UDN) means adding ever smaller cells into a certain area which before could have been covered by for example a single macro base station. Constructing such networks with existing standardized cellular mobile network technologies would mean splitting the area into very small cells. This in turn means exponentially increasing work for radio network planning engineers and at the same time exponentially increasing amount of handover signalling in the mobile network. If high-speed mobility is added to the mix, the probabilities of call dropping and packet loss due to handover failures are expected to increase drastically. Continuous UDN is a proposal to solve the aforementioned issues.

An UDN could in reality mean a network layout of ultra small cells with inter-site distances (ISD) of 20 meters or less in an outdoor environment. This translates to for example installing a small cell to every lamp post. The word "continuous" in this context refers to total or partial removal of the cellular structure at least from traditional handovers point of view. The mobile terminals would be served by only few small cells in the vicinity but from their point of view the network would seem to have no cellular borders. Also the expected quality of experience in a continuous UDN would be similar as if the mobile terminal would constantly sit close to a cell center. Instead of mobile terminal tracking the network, the network would track the mobile terminals. [19]

There is no extensive research published on the topic of continuous UDN yet, but the topic is expected to gain more popularity as the expectations laid on the next generation networks (presented in section 2.1.2) need to be solved in the upcoming future.

Pleros et al. have initially presented a moving extended cell (MEC) concept in

2008 in [20], and further expanded it in 2009 to ensure seamless communication in high-mobility scenario [21]. The MEC would build on top of a cellular network utilizing a radio over fiber (RoF) based fronthauling architecture. An extended cell architecture is constructed so that user specific data is transmitted from several surrounding small cells. The MEC system recalculates the new extended cell area based on user mobility pattern. It was found that the MEC architecture performs extremely well in lowering call drop and packet loss rates and allowing high speed mobility of up to 40m/s velocity with zero loss. The MEC is illustrated in figure 5.
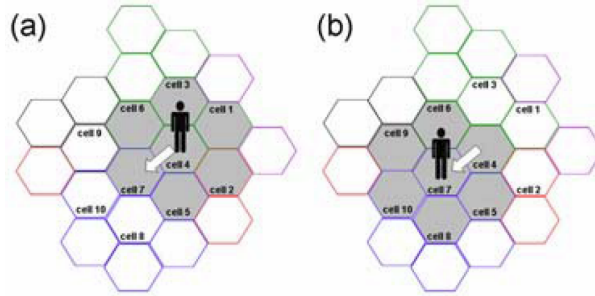

Figure 5: Moving Extended cell concept presented in [20].

Kela et al. have formulated and simulated a borderless mobility concept for 5G UDN [22]. Their idea is similar with the MEC concept to use uplink beacons so that the network can locate and track the mobile terminals, and also find specific channel responses between a mobile terminal and each serving radio node. They also proposed a scheduling algorithm for providing more uniform user throughput than in normally used maximum-throughput schedulers. They found that their proposed borderless scheduling concept achieves about 77% higher median user throughput while only sacrificing about 17% of total area throughput for high density of users with velocities of about 50km/h.

Besides solving the issues related to the access link between UDN and mobile terminals, an important aspect is to figure out how to feed power and provide high capacity backhaul links to the small radio nodes. While electricity is fairly easy to obtain from lamp posts and other street furniture, the connection towards core network is expected to be much more difficult to provide. A fiber based wired backhaul would be most ideal to ensure good enough connection to core network, but from cost point of view it most likely will not be possible. As an alternative solution millimeter wave links have been envisioned to be used in an UDN. This topic will be covered in more detail in chapters 2.2 and 2.3.

Later in this thesis a (continuous) UDN is assumed as the base scenario when analyzing plug and play construction of the network.

2.2   Millimeter wave systems

As the name indicates, millimeter wave systems operate on wavelengths in millimeter scale (around 10mm down to 1mm). Wavelength converted to frequency with the relation to speed of light ($f = c/\lambda$) leads to a range of $30 - 300$ GHz. This frequency range is also referred as Extremely High Frequency (EHF) by ITU-R [23]. The area of millimeter waves is relatively new in terms of actual products, but it has been an area of research for over a hundred years. Since the data rates in wireless communication systems have been growing very fast recently, the conventional systems operating on sub-6GHz spectrum are reaching their capacity limits. The big potential of millimeter wave area is that there is more available spectrum than what has been occupied by all wireless communication systems together until today. The main reason of low utilization of the millimeter wave area in the past is the cost of manufacturing the equipment. There are still many problems to be solved to unleash the full potential, but the latest research in chipsets and other components have been showing promising results [24, Ch. 1.1]. This will be covered in detail in section 2.2.3.

At present the frequency regulation authorities around the world have mostly allocated the available spectrum up to about 100GHz. ITU-R global frequency allocations reach up to 275 GHz in "Radio Regulations, edition 2012" document [23]. The existing research and available products have been focusing mostly on so called V and E-bands. Products in V-band (around 60 GHz) are currently available for private wireless personal area networks (WPAN), wireless local area networks (WLAN) and short range small cell backhaul links. The V-band has been assigned as license free by most frequency regulatory authorities. The reason for that is oxygen absorption spike which limits the usable range due to high attenuation, but also enables high reuse due to low interference leakage. More details will be explained in section 2.2.2 "Millimeter Wave Propagation Channel". E-band (around $70 - 90$ GHz) on the other hand requires a low cost license in most countries. It has been used as a last mile gigabit link for base stations to save in fiber deployment cost.

2.2.1   Physical layer background

This chapter introduces the background information on physical layer of millimeter wave systems. Current status of the technology and the directions of ongoing research in the industry are presented.

Modulation

In general, modulation can be either analog or digital. The modulations used in millimeter wave systems are chosen to be digital as it is the dominant mode in modern communication systems. Digital modulations are more efficient and secure when compared with analog modulations. Basic principle of digital modulation is to take bit sequences as input and map them onto set of symbols and then further map the symbols onto digital waveforms.

The amount of bits to be mapped on each symbol depends on the order of the modulation. The most commonly used base modulation in modern high bit rate communication systems is Quadrature Amplitude Modulation (QAM) and its

variants. As example of QAM usage in modern wireless communication systems, in 4G LTE the highest used modulation currently is 64-QAM, and in current microwave wireless communication systems the QAM order can reach up to 1024-QAM. The higher the modulation order, the more sensitive the system is to phase noise. In millimeter wave systems the phase noise is the most limiting factor of increasing the throughput.

Millimeter wave radio system designs have been made for both single and multi carrier architectures. Single carrier block transmission (SCBT), also known as single carrier frequency domain equalization (SC-FDE), is today's most viable option for a single carrier architecture. Multi carrier systems are commonly based on mature orthogonal frequency division multiplexing (OFDM) architecture. SC-FDE has been found to provide slightly better performance in several aspects, such as lower ADC/DAC complexity, higher SNR (signal-to-noise ratio) gain and lower peak-to-average power ratio (PAPR) requirement (meaning reduced power amplifier (PA) back-off), but OFDM is more robust against severe frequency selective fading. For example in the most recent WLAN standard IEEE 802.11ad both SC-FDE and OFDM are supported in the physical layer (PHY) specifications. Large millimeter wave antenna arrays have been proposed in recent 5G research as part of backhaul technology (mostly LOS links), where each antenna element would have its own small PA and ADC module. In this kind of scenario the single carrier architecture would be most promising choice due to lower complexity requirement of ADCs and smaller dynamic range requirement of PAs. [24, Ch. 2.3, 2.4 & 7.3]

Equalization

Wide band wireless transmission cannot avoid frequency selective fading in the propagation channel caused by e.g. multipath reflections. This generates self interference for the signal called inter-symbol interference (ISI). The process of reducing frequency selective fading is called equalization. In general equalization can increase the gain of the received signal and improve the correct detection of symbols, but as a trade-off it increases delay. Equalization can be performed in different ways; it can be analog or digital and in time or frequency domain. The scope of millimeter wave study in the industry mostly focuses on digital equalization in frequency domain. The implementation of digital frequency domain equalization is most feasible on modern modulations such as single carrier SC-FDE and multi carrier OFDM. Coding and interleaving are commonly combined with OFDM to increase performance. The details of different equalizing methods and equalizer architectures are not presented in this thesis as they are not relevant for understanding the scope of the study. [24, Ch. 2.4 & 7.3]

Error control coding

Error control coding is used to detect and correct bit errors during data transfer. It can be implemented on physical layer (PHY), medium access control (MAC) layer, or on higher layers of a protocol stack. Error control coding can be divided to two categories: error detecting codes and error correcting codes. Another way of defining error control coding is to divide them into block codes and Trellis codes. Block codes

operate on a certain length of data block and add extra bits to the transmitted block for error control purposes. They can be either binary (operating on bit level) or non-binary (operating on symbol level). Trellis codes as such are not very widely used in wireless communication systems, however convolutional codes, which are special case of Trellis codes, are widely used.

The simplest forms of error detecting binary block codes are parity bits or cyclic redundancy check (CRC) that are added to a data block of certain length. The CRC is a block of parity bits calculated from the original block of data. The detection accuracy of CRC is very good with common block length of 32 parity bits reaching detection error levels below 10^{-9}.

Error correcting codes can not only detect errors, but also correct them to some extent. An example of widely used error correcting non-binary block codes are Reed-Solomon (RS) codes. They are well suited for correcting burst errors which can happen due to e.g. symbol decoding errors or multiple symbol fading in OFDM. Low density parity check (LDPC) codes are another example of error correcting block codes. They are well suited for correcting noise generated decoding errors. Besides block codes, convolutional codes operate on a stream of data instead of blocks. Only parity bits are transmitted and then decoded back to actual data stream.

Various kinds of error control coding have been implemented in the current millimeter wave systems. It is also common to combine block codes and convolutional codes in concatenated coding techniques where convolutional codes act as the inner code and for example RS is used as the outer code. It is however worth noting that in some LOS scenarios which allow high link SNR the error control coding gives no extra gain. In fact it will limit the maximum capacity of the link with coding overhead. [24, Ch. 2.6] [25, Ch. 7]

Synchronization

In order to successfully demodulate the received data stream and read the frames of data, the receiver needs to synchronize itself with the incoming stream. There are multiple aspects in performing the synchronization [24, Ch. 2.7]:

1. It is necessary to know when a frame of data begins to be able to read the contained data. The propagation delay of a millimeter wave system might be in the magnitude of even up to tens of frames. The process of finding the beginnings of the frames is called frame synchronization.

2. The receiver also needs to synchronize its local oscillator with transmitted carrier frequency in order to downconvert the received signal successfully. The difference between transmitted signal carrier frequency and the downconverting frequency is called frequency offset. Even slight frequency offset has high effect on the error rate in the receiver. The process of removing frequency offset is called frequency synchronization.

3. Channel estimation is performed after frame and frequency synchronizations have been completed. Channel estimation is used to guide equalization.

Multiple Input Multiple Output (MIMO)

In modern wireless communication systems it is very common to utilize "Multiple Input Multiple Output" (MIMO) technology, which refers to using several antennas in transmitter and receiver sides. There are multiple possibilities of gaining from using MIMO:

- Spatial multiplexing refers to demultiplexing a single stream of data and distributing the resulting streams to multiple transmit antenna elements. At receiver the signals from multiple antenna elements are decoded jointly. The purpose of spatial multiplexing is to increase spectral efficiency by utilizing the propagation environment. Spatial multiplexing can be used to serve multiple users simultaneously with spatially isolated streams utilizing overlapping frequency-time resources.

- Spatial diversity utilizes different propagation paths to reduce the effect of small scale fading. The same stream of data can be transmitted from multiple antenna elements with for example different delays. Multiple copies of the same stream will be received, and it is very likely that some of the copies have not experienced similar fading as some others might have. So far spatial diversity has not been widely studied in the field of millimeter wave communications.

- A very commonly used MIMO technique for millimeter wave systems is beamforming. Using beamforming it is possible to generate high gain and very directive beams by electronically phasing the antenna elements in an antenna array. Multiple elements transmitting with different phases and/or amplitudes can be used to cause constructive interference towards desired directions and destructive interference towards undesired directions. The concept of beamforming is explained in more detail in the following section as it is a key technology for understaning the scope of this thesis.

- Hybrid precoding combines beamforming and spatial multiplexing technologies in such way that spatially multiplexed streams have stream based beams generated towards served users. Optionally also nulls can be directed towards other users to limit interference. Hybrid precoding for millimeter wave systems is still in early phases of study but it could be an effective technology for mobile backhaul systems especially in massive MIMO deployments.

Beamforming and Beam steering

Radiation pattern of an antenna element is dependent on its shape and size. A combination of multiple antenna elements is called an antenna array. Antenna radiation pattern consists of "beams" which represent the radiation intensity to specific directions. The radiation power is highest at the boresight of a beam. Width of a beam is normally measured as an angle between two directions where half of the maximum power can be radiated. This angle is also called half-power or 3 dB beamwidth. Figure 6 presents an example of a three-dimensional antenna radiation pattern in which maximum radiation is on positive z-axis at direction $\theta = 0^o$. Most

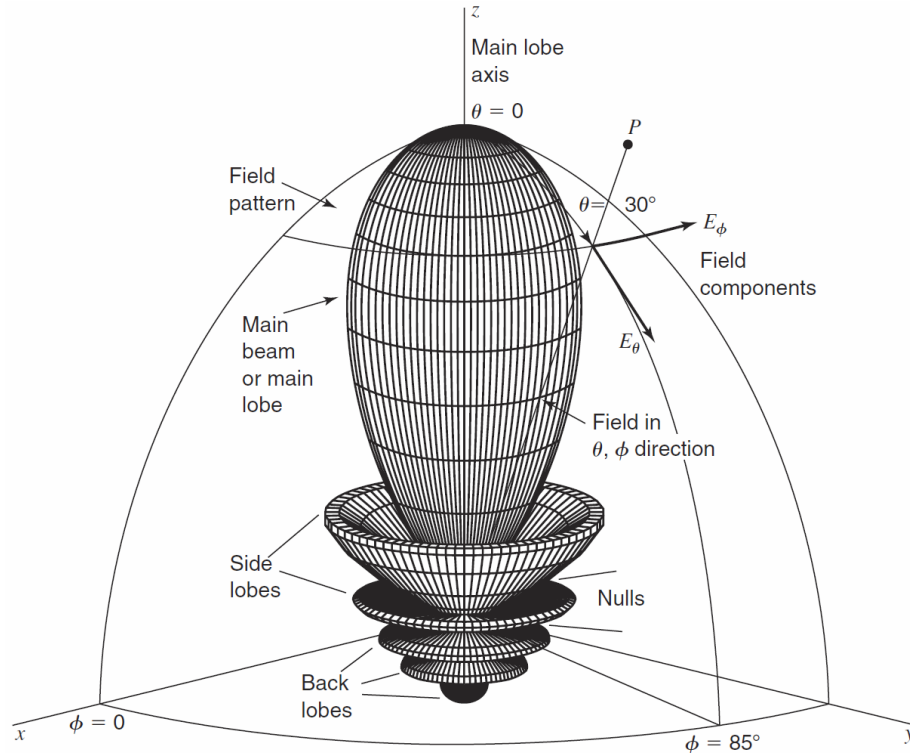of radiation is concentrated in the main lobe, the rest is radiated through side and back lobes. [26]

Figure 6: Three-dimensional example of antenna radiation pattern. Main lobe is pointed at $\theta = 0^o$ direction on positive z-axis. Direction of point P is at angles $\theta = 30^o$ and $\phi = 85^o$. [26]

The ability to efficiently radiate signals to specific directions and to control the radiation pattern dynamically can be implemented using a technology called beamforming or by using an antenna system which has a set of selectable fixed beams. Beamforming is based on coherent superposition of wavefronts, meaning that copies of signals sent with specific phases and amplitudes from individual antenna elements are added up constructively in desired directions and destructively at least in the most unwanted directions. By limiting the radiation to unnecessary directions it is possible to increase the overall energy efficiency of the radio system and reduce interference to other terminals/systems. The method of specifically generating "nulls" towards undesired directions is called Zero Forcing (ZF). Beamforming can be considered as a way of realizing spatial filtering. In beamforming system the beamwidths can be changed freely within the antenna array capabilities. One possibility is to generate a beamforming codebook where different width and direction of beams are allocated from a code tree as presented in figure 7 from [27].

Beamforming can be performed in fully analog or digital domain, but also as a combination of analog and digital called hybrid beamforming (or hybrid precoding) [24, Ch. 2.8.3 & 2.8.4]:

- Analog beamforming operates normally at the RF frequency by adjusting phasing of antenna elements. Each antenna element could have its individual phase shifter. This enables a cost efficient millimeter wave designs. Analog beamforming however is not able to efficiently control amplitudes of the antenna elements, which results in sub-optimal beamforming solution.

- Digital beamforming relies on precoding of the transmitted signals at baseband. This way both phase and amplitude of each stream can be efficiently controlled. Limitation of full digital beamforming in millimeter wave designs is the cost of implementation in both hardware and power consumption, as every antenna element should have individual transmitter/receiver chain from baseband to RF circuitry. Digital beamforming also requires channel state information for the generation of optimal precoding vectors.

- Hybrid beamforming combines benefits of both analog and digital beamforming. A hybrid beamforming system normally consists of antenna element specific analog phase shifters and fewer digital transmitter/receiver chains. It supports analog beamforming and works without good channel state information which is normally required by digital beamforming. A sub-optimal beamforming solution can be formed using analog phase shifters and then compensated using by digital precoders. Due to fewer number of required digital transmitter/receiver chains the hybrid beamforming architecture can achieve more efficient and cost-optimized design.
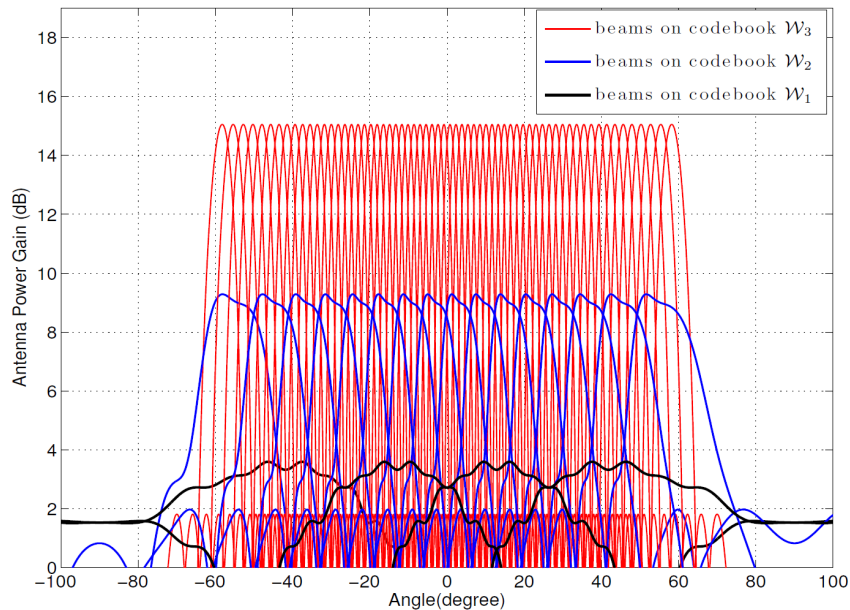

Figure 7: Beam patterns of three-level beamforming codebook [27].

Massive MIMO

A commonly used antenna element spacing in MIMO antenna arrays is $\lambda/2$. Since the operating wavelengths are in millimeter level, it is possible to pack a very large number of antenna elements in a small area. This opens a new area of research in MIMO called Massive MIMO (M-MIMO). The actual shape of the antenna array can vary depending on required directions for the signal radiation. Examples of different antenna array types for massive MIMO are presented in figure 8. The physical dimensions of the presented arrays are relatively large since the frequency range under the study in [28] was in the sub-6GHz area. In millimeter wave area the array structures could be of similar type but the physical sizes would be only a fraction of the ones in figure 8.
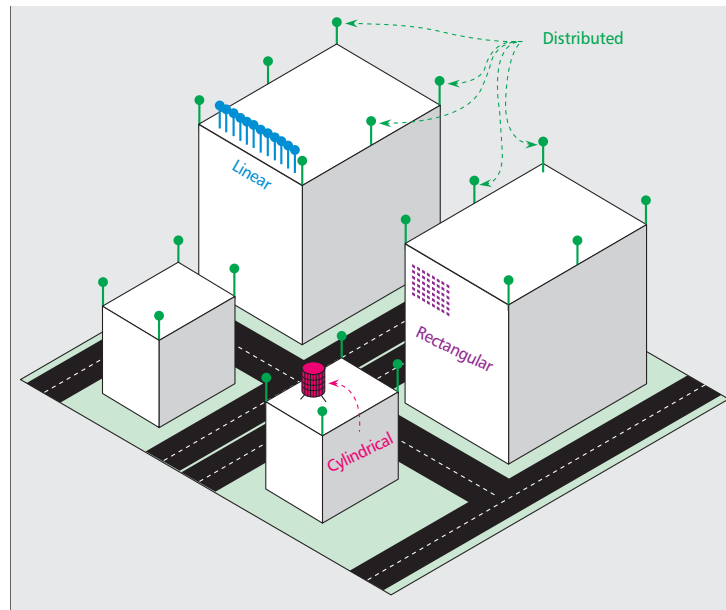
Figure 8: Examples of Massive MIMO antenna arrays [28]

The principle of massive MIMO is to further exploit the beamforming function presented above. With large enough amount of antenna elements in an array it is possible to generate such accurate beams that each served terminal can be allocated a unique beam. This would effectively limit multipath scattering based inter-cell and intra-cell interferences, which in turn means that single spectrum and time resources could be reused for every served terminal. It might be possible to increase serving radio node overall capacity by more than the times and power efficiency by up to hundreds of times compared to current state of the art systems. Main reason for much higher power efficiency is that the power can be accurately radiated towards the desired terminals and the undesired power leakage to other directions can be minimized.

In massive MIMO system each antenna element would require to be fed individually with individual amplifiers. In existing systems the amplifiers are feeding only few antenna elements so the power levels need to be up to tens of watts with

very high linearity requirements. With a very large number of antenna elements the amplifiers could operate in milliwatt levels. Also the linearity and accuracy requirements of small individual amplifiers would be much lower. Only requirement is that the simple amplifiers and antenna elements would work in a combined manner for efficient beamforming. [28]

There are multiple challenges related to implementing massive MIMO in millimeter wave systems, such as proper downlink channel estimation and link establishment without prior knowledge of where to direct a beam. The challenges are further addressed in chapter 2.2.3.

2.2.2 Millimeter wave propagation channel

Each frequency band have their unique flavors in the propagation channel. The properties that mostly affect the millimeter wave propagation channel are high frequency caused increase in path loss, increased rain attenuation and absorption caused by some atmospheric gases (such as O_2 at 60GHz). This chapter will introduce the characteristics of millimeter wave propagation channel and different components affecting the total link budget.

Path loss

Free space path loss gives the baseline for the losses in the propagation channel. The factors affecting the free space path loss are squares of wavelength and distance between transmitter and receiver. Friis free space path loss law (1) gives the received power as a function of transmitted power, transmit and receive gains, and distance plus wavelength dependent loss factor.

$$P_{rx} = \frac{P_{tx} G_{tx} G_{rx}}{L} \left(\frac{\lambda}{4\pi d}\right)^2 \tag{1}$$

where P_{tx} and P_{rx} are transmitted and received powers respectively, G_{tx} and G_{rx} are transmitter and receiver gains respectively. L represents the system losses (e.g. feeder loss and impedance mismatch), and $\left(\frac{\lambda}{4\pi d}\right)^2$ defines the wavelength (λ) and distance (d) dependent free space path loss component. It is common to present path loss calculations in decibel format as all absolute value multiplications reduce to simple additions in decibel format. Excluding antenna gains and other losses, the path loss in decibel format is

$$L_{PL} = 10 \log \frac{P_{tx}}{P_{rx}} = -10 \log \left(\frac{\lambda}{4\pi d}\right)^2. \tag{2}$$

Wavelength can be translated into frequency assuming the waves propagate at the speed of light using relation $\lambda = c/f$. The path loss in decibel format with wavelength converted to frequency is

$$L_{PL}(d, f) = 20 \log(f) + 20 \log(d) + 20 \log \left(\frac{4\pi}{c}\right). \tag{3}$$

The constant term $20 \log \left(\frac{4\pi}{c} \right)$ can be replaced with its rounded value *-147.55 dB*. Further modifying the equation to take distance in meters and frequency in megahertz, the constant will increase by $20 \log(10^6) = 120$ giving us a bit simpler form

$$L_{PL}(d, f) = 20 \log(f[MHz]) + 20 \log(d[m]) - 27.55dB \tag{4}$$

Since the above equation is only modeling the free space path loss, it is a bit too general for modeling actual path losses in real environments. So according to [25] an option for a more suitable way of modeling path loss in millimeter wave area could be to introduce path loss exponent component α. A log-normal shadowing model is

$$L_{PL}(d) = L_{PL}(d_0) + 10\alpha \log \left(\frac{d}{d_0} \right) + \chi_\sigma \tag{5}$$

where d_0 represents a reference distance meeting $d \geq d_0$ condition. $L_{PL}(d_0)$ is the pathloss at reference distance d_0 at given carrier frequency f. α represents the pathloss exponent, which based on measurements could be close to $\alpha = 2$ (free space loss) for a LOS link, and up to $\alpha = 4$ for severe multipath scenario. χ_σ represents large scale shadowing component, following log-normal distribution with 0dB mean and standard deviation $\sigma[dB]$.

The frequency dependent loss component will not necessarily become a limiting factor with state of the art systems. If assuming that the antenna physical size is kept the same over different frequencies, it is possible to pack a much larger array of $\lambda/2$ sized dipole elements on the same physical area as with lower frequency dipole antennas. As introduced in chapter 2.2.1 "Beamforming" section, it is possible to generate narrow high gain beams to compensate (and even overcome) the high frequency losses with a large amount of antenna elements in an array. An example given by Akdeniz et al. in [29] indicates that a 4×4 array at 28 GHz and an 8×8 array at 73 GHz would take approximately same physical area with $\lambda/2$ sized elements resulting in similar SNR and rate distributions.

Atmospheric effects

Distance and frequency dependent path loss is not the only attenuating factor on the total path loss in millimeter wave systems. Extra attenuation caused by absorption of atmospheric molecules, especially oxygen and water, affect millimeter waves more severely than for example centimeter waves (microwave frequencies). As seen in figure 9, the absorption by for example O_2 molecule is very severe – about 7–15.5 dB/km around 60 GHz. Other severe attenuation spikes can be found around 180 GHz and 380 GHz by H_2O molecules absorption. The absorption is highest at sea level as the air pressure forces more air molecules in same space. H_2O caused absorption is relative to the concentration of water vapor in the air, i.e. dry air absorbs less than humid air. [30]

Frequencies experiencing highest absorption rates presented above are most likely going to be utilized in non-licensed systems as the co-channel interference can be effectively reduced by even short distances – compared to for example 2.4 GHz WLAN frequency band being too crowded due to omni-directional transmissions from almost every apartment, office, or other public building nowadays.
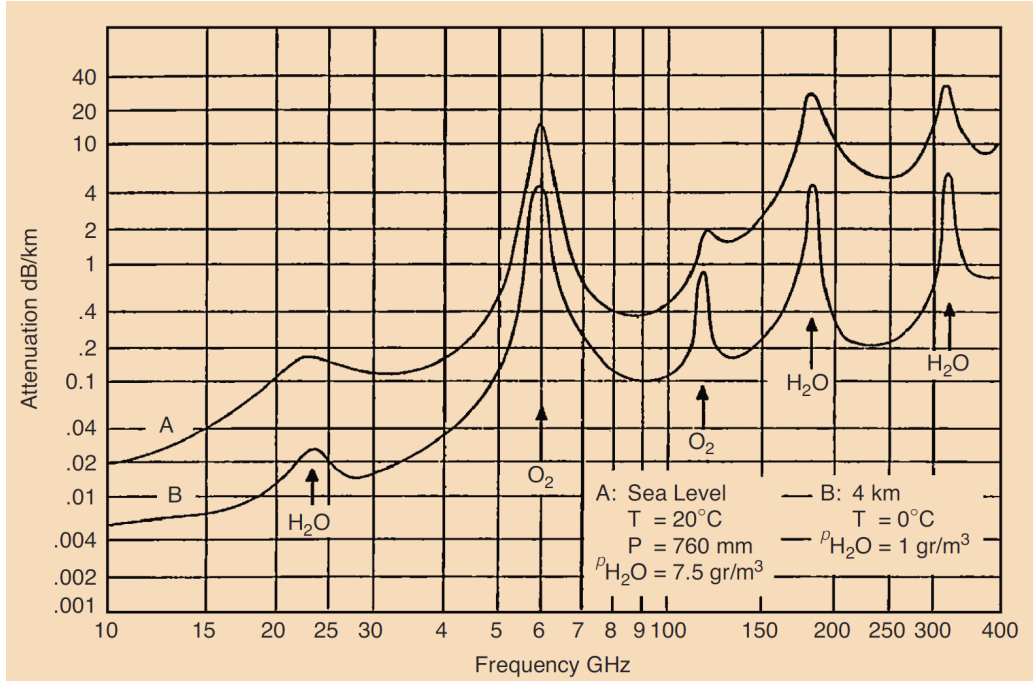
Figure 9: Specific attenuation due to atmospheric abrorption. [30]

Millimeter wave bands normally under regulation are the ones that are least affected by additional attenuation due to absorption of atmospheric gases. E-band around 70-90 GHz is one of the frequency bands that have been recently introduced for licensing in most countries. At the moment it is being used by mobile operators as point-to-point fiber replacement for base station or small cell "last mile" link.

Rain attenuation at millimeter wave range is an important factor affecting the atmospheric losses as the wavelengths are roughly the same size as the size of raindrops. This similarity in size causes extensive scattering of signals from the raindrops. Rain rate is used to indicate the rainfall intensity in a given time. The rain rate is normally measured in millimeters per hour. One active research topic has been to model the rain attenuation differences raindrop size distributions. The rain attenuation is dependent on the used distribution for the prediction, for example rainfall rate of 50 mm/h (heavy rain) could introduce attenuation between 8dB/km and 18dB/km at 60 GHz frequency. ITU-R model for frequency and rain rate specific attenuation is given in [31]:

$$L_{rain/km}(f[GHz], R[mm/h]) = k(f)R^{a(f)} \left[\frac{dB}{km} \right] \tag{6}$$

$$k(f) = 10^{1.203 \log(f) - 2.290}$$

$$a(f) = 1.703 - 0.493 \log(f).$$

An example of specific rain attenuation curves for different rain rates is presented in figure 10 for frequencies from 1 GHz up to 1 THz [30]. Besides rain, hail stones
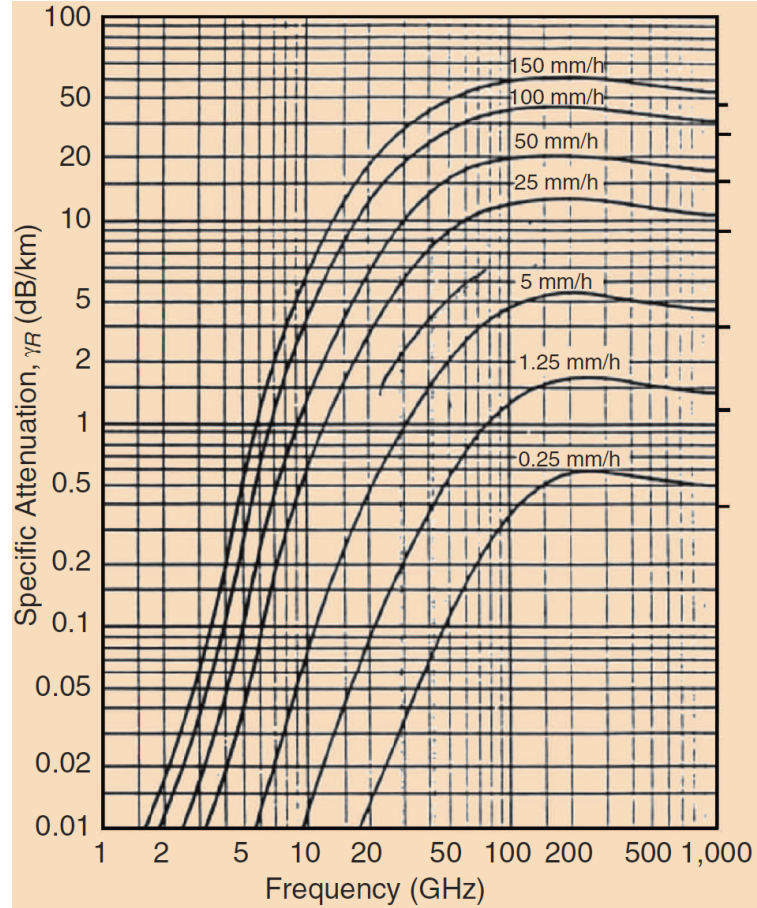
Figure 10: Specific attenuation due to rain. [30]

and snowflakes also cause severe attenuation, but the detail study on the effects is not presented in this thesis. [32]

Fading

When defining total link budget, it is necessary to also include margin for fading. Fading can be flat or frequency selective, also fast or slow. Flat fading refers to scenario where all frequency components of the signal will experience similar fading. Frequency selective fading will result in only some frequency components being faded due to constructive or destructive interference of different multipath components of the signal. Fast fading refers to fast varying movements on the signal paths causing fast changes on the received signal level. Slow fading, also referred as shadowing, is normally caused by larger and more stationary objects such as buildings.

Since LOS paths are likely to exist in the millimeter wave communications, the fading can be modeled as Ricean instead of Rayleigh which is used to model fading in Non-LOS paths. [32]. According to several studies (summarized by Rappaport et al. [24, Ch. 3.2.7]), it has been found that foliage and other vegetation can cause heavy attenuation at millimeter wave range. At 57.6 GHz the signal can attenuate as much

as 50 to 80 dB when propagating through 20 to 80 m of foliage. The attenuation is hardly ever constant due to wind induced swaying and moving of the trees and leaves. The attenuation is better modeled as random process with variance depending on the used carrier frequency and wind speed.

Another element causing possible fading in millimeter wave systems is beam misalignment. In millimeter wave communications the beams are normally very narrow, possibly only few degrees of half-power beamwidth. Narrow beam combined with short link distance requires the beam aligning to be very accurate. When considering an outdoor deployment on for example lamp poles, there might be movement of the radio nodes itselves due to wind instead of elements between the nodes. This could lead to problems with today's millimeter wave equipment having fixed mechanical antennas without beam tracking functionality. [33]

There has been studies how the antenna array could track the moving antenna. In case of Integrated Lens Antenna (ILA) the strongest beams need to be detected and switched accordingly, and in case of electrically adjusted beamforming a more flexible beam steering is available. Beam direction and width can be adjusted based on detected angle of arrival (AoA) and required angle of departure (AoD). Veijalainen proposes in his thesis [34] three methods of tracking the moving remote node; orientation sensor in remote node with feedback, RSSI measurement method of detecting adjacent beams signal strength, and the combination of the two methods. His findings indicate that both orientation sensing and RSSI measurement method have their flaws, but the combination of both methods yields the best results as they can address the weak points of each others.

Link budget example for millimeter wave system

The purpose of calculating a link budget of a radio system is to find the received power margin at the receiver. It can be compared with the receiver sensitivity to find if the power is high enough for the receiver to detect the signal sufficiently. Also the used modulation and coding scheme (MCS) can be decided based on the calculated signal to noise ratio (SNR) according to the MCS requirements.

Total link budget of a radio system consists of the following

+ Power fed to antenna at transmitter

+ Transmitter and receiver antenna gains

− Path loss (where path loss exponent α set according to scenario)

− Atmospheric losses (including gaseous absorption and rain/snow)

− Transmitter and receiver system losses (e.g. impedance mismatch and feeder losses)

− Fading margin (e.g. fast fading and shadowing)

The equation for calculating received power according to above listed items is as follows (in decibel format):

$$P_{RX} = P_{TX} + G_{TX} + G_{RX} - L_{PL} - L_{atm} - L_{TX-sys} - L_{RX-sys} - M_{fading} \qquad (7)$$

Distance and frequency dependent component L_{PL} is calculated using equation 4 or 5 (depending on requirement).

Sensitivity of a receiver can be calculated from

$$P_{RX-sens} = \left[\frac{S}{N}\right]_{dB} + N_0 \tag{8}$$

where S/N states the SNR requirement for specific MCS and N_0 is the thermal noise level which can be derived as follows

$$N_0 = 10 + log(k_B T_{syst}) + 10\log(B) + F \tag{9}$$

In the noise level calculation k_B is the Boltzmann constant, T_{syst} is the receiver system temperature in Kelvin, B is the used bandwidth, and F indicates noise figure of the receiver. A normally used value for the system temperature dependent part is -174 dBm/Hz which results using a system temperature value of 17^oC (= 290.15 K).

When received signal power P_{RX} and N_0 are known, it is possible to determine the SNR of a noise constraint system in decibel domain by modifying the sensitivity calculation in (8) as follows:

$$SNR = P_{RX} - N_0. \tag{10}$$

When a system is interference constrained, the calculation can be based on P_{RX} and I (I could be also written as $P_{RX,interference}$. The resulting Signal-to-Interference ratio (SIR) is defined in decibel domain:

$$SIR = P_{RX} - I. \tag{11}$$

When both thermal noise and interference are included in calculation, Signal-to-Interference+Noise ratio (SINR) is then written as:

$$SINR = P_{RX} - (I + N_0). \tag{12}$$

The SNR (or SINR) requirement of a system is dependent on the bit error ratio requirement, which for a "quasi-error-free" system after Viterbi/Reed-Solomon decoding could be $\sim 10^{-11}$. For commonly utilized modulations of QPSK (Quadrature Phase Shift Keying), 16-QAM and 64-QAM the resulting SNR levels would be 10.7 dB, 16.7 dB and 21.7 dB, respectively. [35]

2.2.3 Challenges of using millimeter wave links

This section will discuss several challenges related to implementing and using millimeter wave links. The challenges relate to the nature of millimeter waves from propagation to atmospheric effects. Status of millimeter wave chip development challenges in history and today's industry are also covered.

Challenge of directivity
High directivity of millimeter wave links can be seen as a benefit and a challenge.

The use of narrow beams result in high gain and low spatial power leakage, which are considered as positive effects. The negative effect is that the links require very accurate alignment. Considering an initial link establishment process – if no prior knowledge of where to point a beam is available, it might be hard to establish the link with reasonable effort. A tree-based beamforming codebook search process could be used as presented in "Beamforming and beam steering" section of chapter 2.2.1. The search is started with wide (low gain) beams and continued by searching with more narrow (higher gain) beams in the tree until the best leaf of the tree has been found.

In fixed (non-steerable) links the link establishment is also a very delicate process. If for example half-power beamwidth of 3 degrees is assumed, only 1.5 degrees misalignment to either side of the boresight results in loss of half of the transmitted/received power. In practical terms – if an antenna is fixed on a pole with diameter of 8 cm, 1.5 degrees misalignment will be caused by only 1 mm twist on the pole surface. At 100m distance an error of 1.5 degrees in alignment means roughly 2.6 meters "off the target".

In "Fading" section of 2.2.2 wind-induced movement of nodes causing misalignment was introduced. A possible solution is to have beam tracking capabilities in the nodes. An example of a beam tracking method was presented by Veijalainen in his thesis [34]. A paper by Nitsche et al. proposes a method of "blind beam steering" for reducing the required overhead on narrow beam training. Their study focuses on latest wireless LAN (WLAN) standard IEEE 802.11ad which operates on millimeter wave frequencies. It is assumed that the WLAN device supports also legacy technologies on sub 6GHz band operating with omni-directional transmissions. They propose to use an antenna array to measure the angle of arrival (AoA) of received legacy WLAN signals and use the same estimate for steering the narrow beam to be used with millimeter wave system.[36]

Challenge of attenuation

High attenuation on millimeter wave links is result of several affecting components. The basic component is distance and frequency specific free space path loss, presented in (4). Besides free space path loss, other components such as rain and atmospheric molecules absorption are causing high attenuation as presented in "Atmospheric effects" section of 2.2.2. They are commonly measured as specific attenuation per certain distance (e.g. dB/km). In short range links ($\sim$10–100m) the attenuation is not such a big problem, but in longer spans the attenuations might become unbearable.

Challenge of phase noise

Crystal oscillators are used to generate local oscillator frequencies that are essential for frequency conversions. Normal way is to use voltage-controlled oscillators (VCO) which can adjust the frequency based on given control voltage. All oscillators are unstable, which is the base for the whole oscillation phenomenon. A control loop is used to adjust and maintain the oscillation at desired frequency as stable as possible. Basic crystal oscillators base frequency is normally much lower than the millimeter wave carrier frequencies. In order to generate a millimeter wave range

carrier, the base frequency of an oscillator is multiplied by several times. The higher the carrier frequency, the more frequency multiplications are required. Phase noise of an oscillator close to its base frequency might not be an issue, but when the noise is multiplied by several times with the carrier, it will most likely become an issue. [24, Ch. 5.11.4] [37]

Phase noise effects on a communication system are seen as random rotations of the signal constellation. This in turn affects the observed carrier-to-noise (or other words signal-to-noise) ratio. Especially in systems utilizing high-order QAM modulations the random rotations of constellation can cause very severe symbol detection issues. Lower order constellations can withstand more severe phase noise levels, but in case of for example 256-QAM even slight rotations coupled with slight amplitude changes are likely to cause high amount of symbol detection errors. The approach in today's millimeter wave design has been to compensate the reduction in modulation orders by using much wider bandwidths in order to reach high bitrates. [38]

Challenge of limited transceiver dynamic range
Peak-to-average power ratio (PAPR) defines the dynamic range requirements of for example amplifiers and analog-to-digital converters (ADC). The higher the PAPR, the higher dynamic range is required. As noted in "Modulation" section of 2.2.1, single carrier systems tend to have lower PAPR requirements than multi-carrier systems. Reduced dynamic range requirement allows more efficient power amplifier (PA) operations and reduces the required ADC complexity. [39]

Challenge of downlink channel estimation
Massive MIMO system (introduced in 2.2.1) requires on knowledge about the channels on both uplink and downlink. Uplink channel estimation is relatively easy as terminals can send uplink beacons based on which a massive MIMO base station estimates the channel state information (CSI) [22]. Downlink CSI is much more difficult to obtain in a massive MIMO base station as the amount of beacons is proportional to the used antenna elements. After the beacons has been received by a terminal, it must report the measured CSI back to the base station. Compared with for example LTE, there might be couple of decades more information to be sent. A possible solution to this problem is to use TDD mode instead of FDD, as channel reciprocity between uplink and downlink channels could be utilized. [28]

Challenge of millimeter wave circuit design and manufacturing
Early millimeter wave tests to carry telecommunication traffic were conducted already in 1930s by using klystron vacuum tubes as frequency sources and circular waveguides as the link medium. The frequency source was far too unstable for long term operation. Also the waveguides for millimeter waves were later replaced by optical waveguides (fibers).

Silicon based microwave integrated circuits (IC) arose in the 1960s, followed later by millimeter wave ICs. In early stages difficulties in fabricating silicon substrates with high resistivity led to the path of using GaAs microwave ICs. In 1970s the

GaAs based microwave ICs were started to be used on a large scale.

In the beginning of 21st century silicon based complementary metal-oxide semiconductor (CMOS) based millimeter wave oscillators were demonstrated to work which paved the way for the era of millimeter wave CMOS use on a large scale. On-chip antennas at sub-millimeter dimensions were presented to achieve far greater level of integration. With large phased arrays in small form factor enable efficient beam steering capabilities for millimeter wave applications. As the CMOS development enables smaller and smaller production processes, it in turn enables efficient ICs at ever higher frequencies. The continuous investments in the industry to the development of affordable CMOS ICs will enable affordable millimeter wave applications in the near future. [40]

The issue of producing millimeter wave ICs in general is that as frequency increases, the available effective gain of a transistor decreases. This fact makes it difficult to design for example amplifiers for millimeter wave area. The current practice of industry is to design and manufacture fully integrated millimeter wave ICs including all elements of a tranceiver architecture from antennas to amplifers and analog-to-digital converters (ADC). Different CMOS processes can be used for different elements on a layered desing. Figure 11 shows a layered CMOS architecture using higher cost substrate used for active RF chip and lower cost substrate for passive antenna chip. The layers utilize different CMOS production processes. [41]
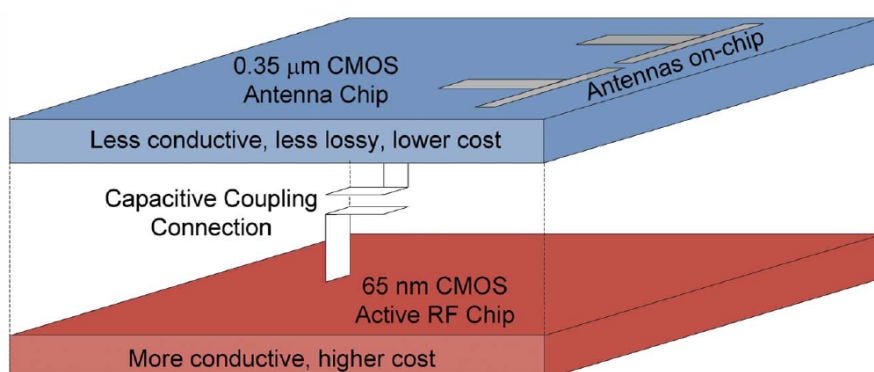
Figure 11: Layered CMOS architecture with chips using different CMOS production processes coupled together with capacitive coupling connection. [41]

2.2.4   Example of today's millimeter wave system: 802.11ad WLAN

Previous IEEE 802.11 WLAN standards have been focusing on sub-6GHz bands. Most commonly used unlicensed bands have been 2.45 GHz and 5 GHz. The bandwidth available especially in 2.45 GHz area is far from enough for today's requirements, and the band has been heavily congested already for some time. 5 GHz area has been a little less congested as result of fewer deployed 5 GHz access points and limited penetration through obstacles (e.g. walls). The most recent commercial devices utilize 802.11ac standard which enhances the operation on 2.45 and 5 GHz bands by

utilizing up to 8 antennas MIMO processing, higher order QAM modulations and expanding the used bandwidth.

Most feasible way of expanding the troughput beyond 802.11ac capabilities is to utilize much wider channels. This is where 802.11ad standard comes to the picture with capability to operate at 60 GHz range where several gigaherz of bandwidth is available. Operation at 60 GHz requires the use of state-of-the-art CMOS technology. In order to make the hardware architecture flexible depeding on required use cases, there are multiple options for the physical layer (PHY) operation of 802.11ad:

1. Control PHY

2. OFDM Multi-carrier PHY

3. Single-carrier PHY

4. Low-power single-carrier PHY

The control PHY is designed for maximal interoperability between different vendors' implementations. Similarly as introduced in chapter 2.2.1, multi-carrier PHY enables most spectral efficient operation but with most complex hardware implementation. Single-carrier PHY enables the best tradeoff between efficiency and hardware simplicity. Low-power PHY enables most power-efficient operation and minimal complexity in hardware design – naturally with reduced performance.

To be able to overcome the high path loss of millimeter wave (and especially 60 GHz) area, highly directional beams are required to be used. The 802.11ad implementation of beam adaptation includes sector sweeping with wider beams and refinement using narrow beams. Once the initial beam direction has been found, beam tracking will be performed during operation to maintain proper alignment.

Coverage of 60 GHz 802.11ad WLAN will be very limited as the low-power signals will have hard time penetrating any obstacles. Mostly LOS or near-LOS links are required for proper operation. Coverage extension by means of relaying has been included in the standard. Relaying has been designed to work in two modes: either by switching to relay path if direct path has been obstructed, or by using cooperative relay operation of transmitting to both the relay unit and the receiving unit. [13] [24, Ch. 8.3.2, 9.6]

2.3   Fronthaul and backhaul technologies

Classification of a link to be fronthaul or backhaul is defined based on the position of the link in a base station (BS) architecture.

The roles of equipment for example in a conventional LTE network are such that a base station (BS) includes RF and baseband modules, meaning that it downconverts, demodulates, and decodes the received symbols from a user equipment (UE) to a stream of actual payload bits. Then the BS sends the decoded data bits over a link called backhaul towards the mobile core network.

Fronthaul link on the other hand is a relatively new concept, which is located between RF unit and a baseband processing unit. It conveys raw in-phase and

quadrature components of received and transmitted radio signals. Initially this concept was utilized within in a single site with dark fiber, but an emerging concept of Cloud Radio Access Network (C-RAN) can unleash the full potential of fronthauling.

Following sections will introduce C-RAN and distributed RAN architectures in more detail, and finally a summary of fronthaul and backhaul technologies will be presented.

2.3.1 Cloud RAN Architecture

A widely used recent innovation in the mobile network architecture is to separate baseband unit (BBU) from remote radio unit (RRU). RF jumper cables can be shortened tremendously as a RRH unit can be placed next to an antenna effectively limiting power loss in the RF cable. A BBU would be located within reasonably short range from the RRH. The link between BBU and RRU is called fronthaul, and in currently available systems it uses one of the competing fronthaul protocols Open Base Station Architecture Initiative (OBSAI) or Common Public Radio Interface (CPRI). Raw in-phase and quadrature components (I/Q) of received and transmitted radio signals are transmitted over the fronthaul link, which means the amount of transmitted data is huge compared to the actual UE payload data. Physically the link is most often fiber due to very high bandwidth and low delay requirements, but most recent millimeter wave radio links have the necessary performance to carry this so called radio over fiber (RoF) signal also over the air for 3G and 4G networks. [42]

The split of RRU and BBU enables a new type of RAN architecture where all baseband processing could be moved to a cloud platform using general purpose processors (GPP) and a hypervisor platform. In the latest research at least the following benefits of a C-RAN architecture have been identified [43] [44]:

1. Flexible allocation of baseband processing resources from a centralized BBU pool. The RRUs can be kept relatively simple as no baseband processor is required which allows more CAPEX optimized wide deployment.

2. Radio control data combining and superfast information exchange between virtual BBUs allow much better inter cell interference coordination (ICIC) and Coordinated MultiPoint (CoMP) transmission. These two technologies enable a more uniform user experience in the coverage area and enhance the network capacity by efficiently controlling leaked interference.

3. Dynamic resource allocation for multi radio access technology (RAT) deployment.

4. Efficient power consumption control of a heterogeneous network (HetNet) by switching on/off certain RRUs and allocating users between macro cells and small cells.

5. Control plane and user plane decoupling for more efficient management of resources. It could be possible to run control plane through macro cells while user plane is running over small cells. Also uplink and downlink separation

between different RRUs is possible by utilizing more complete radio environment view in the centralized control point.

Actual implementation of fronthaul links in a C-RAN architecture can be realized in several ways. A common implementation way today is to digitize the I/Q samples using an ADC and transport them as high speed digital stream. This method can be also referred as digital fronthaul. Another way is to transfer the I/Q radio data purely over analog link without performing AD and DA conversions for each MIMO channel at the endpoints of the fronthaul link. Both methods will be briefly presented below.

Digital fronthaul

Digital fronthaul link is based on digitizing I/Q samples of individual received channels and transporting them to BBU for processing. In a MIMO system the number of received MIMO channels equals the amount of received I/Q streams. Currently there are two competing open standards for digital fronthaul, CPRI and OBSAI. The bitrate of the fronthaul stream is dependent on the sampling rate and size of each sample. As an example, CPRI uses sample size of 30 bits where 15 bits are allocated for I and Q branches separately. For example in OFDMA based 4G LTE system the sampling rate is calculated as follows:

$$F_s = FFT\_size * OFDM\_subcarrier\_spacing. \qquad (13)$$

The FFT_size is 2048 for 20 MHz LTE channel and $OFDM\_subcarrier\_spacing$ is 15 kHz. The resulting F_s is then 30.72 MHz. The total digitized I/Q data rate of LTE system with MIMO and without control overhead is as follows:

$$R_{CPRI(LTE\&MIMO)} = MIMO\_channels * F_s * Sample\_size. \qquad (14)$$

As an example the bitrate of a CPRI link for non-MIMO 20MHz LTE results in $R = 921.6Mbit/s$. 8-channel MIMO would increase the rate to $R = 8*921.6Mbit/s = 7372.8Mbit/s$. Adding the control overhead results in bitrate of just below 10 Gbps. This kind of data rate is still possible to be carried with common 10GE optics or most recent commercial millimeter wave links. [45]

Challenges on the feasibility of CPRI-based digital fronthaul for 5G systems are presented in chapter 3.1.1.

Analog fronthaul

There are multiple ways of realizing analog fronthauling. The most intuitive way would be to mix the received signals from each antenna element as separate subcarriers on an ultra wide band (UWB) fronthaul system. The system could be optical such that the channels would be separate wavelengths in the fiber. Alternatively the combined UWB signal could be transmitted over a millimeter wave link. This way the received signals would be transmitted as-is. As an example, if the access link bandwidth is 200 MHz and there are 5 antenna elements - ideally the mixed fronthaul link bandwidth would be $5 * 200MHz = 1GHz$. In practice some guard bands are

most likely necessary – unless the guard bands in the access link 200 MHz channels would be enough to protect against adjacent channel interference on the millimeter wave or optical area.

Another way of realizing analog fronthaul would be to use a method called on-site-coding (OSC) proposed by Alwan et al. in [46]. As opposed to the digital fronthaul case where each antenna element would have its own ADC and the data stream of each ADC would be carried separately, the method proposed by Alwan et al. uses code division multiplexing (CDM) to combine the received signals per antenna element. This way the antenna elements do not require individual ADCs, but instead a single ADC can be shared by all the streams. On the baseband processor it would be then possible to identify each stream uniquely based on the spreading codes. This ensures combining gain can be achieved. Due to some non-orthogonalities between the codes, the combining gain suffers minor losses compared to the case where each stream would be carried separately. To utilize this method for UWB applications such as 5G, Alwan et al. propose to split the wide band signals into multiple sub bands and dedicate single ADC per sub band. Figure 12 presents the on-site-coding receiver architecture for UWB signal reception. The analog fronthaul link would sit between the on-site-encoder and the combiner in the figure.

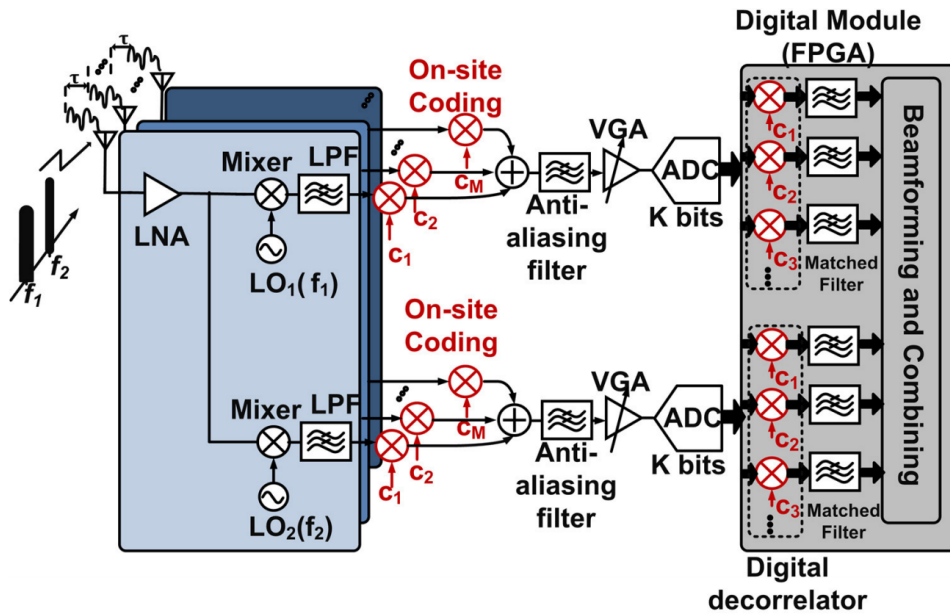Figure 12: Proposed on-site-coding receiver (OSCR) architecture for UWB applications by Alwan et al. in [46]

2.3.2 Distributed architecture

Distributed architecture refers to the architecture where baseband processing is fully distributed to the base stations. All received symbols will be decoded and transmitted symbols encoded in the base station baseband module. The link from base station to

upper network (after digital baseband processing) is called backhaul. Pure payload data is transmitted in a backhaul link, meaning a non-encoded stream of bits. The backhaul link between the BS and mobile core network can be basically any that can meets necessary end-to-end bandwidth and delay constraints. The rate of backhaul link

Distributed architecture has at least following options:

- Fully distributed architecture: baseband and control in radio nodes. The bandwidth requirement on the backhaul link is minimized, but the lack of joint processing of control data can reduce the performance of for example coordinated multipoint transmissions (CoMP).

- Hybrid architecture: partial baseband processing in the access nodes, control data processing is centralized.

  - Since user plane is not so delay sensitive, it is beneficial to perform baseband processing for it in the access node. This will heavily decrease the bandwidth and delay requirements of the upstream link. It is especially helpful when the link is based on wireless technology (e.g. a millimeter wave link).

  - The control plane is very sensitive to additional delays due to pre-coding computation in a central processing node and distribution of pre-coding vectors to the access nodes. Control plane information would thus be transmitted "as is" to the central radio resource processor.

The capacity requirement of backhaul links are close to the actual payload traffic transmitted through the access nodes. This makes the backhaul attractive option which could be possibly carried even over today's millimeter wave links. The payload traffic flowing through a single access node is most likely going to stay below 10 Gbps at all times according to the 5G network metrics listed in chapter 2.1.2. Average throughput most likely much lower.

2.3.3 Fronthaul and backhaul summary

It is not possible to clearly say which one should be chosen, Cloud RAN or distributed RAN model. The choice depends much on the available link technologies and possibly required central processing capabilities.

Background of Cloud RAN architecture was presented in chapter 2.3.1. The usage of centralized Cloud RAN architecture enables combining gains and makes radio resource management more efficient by having a "full picture" of the radio environment at a central resource management entity. The drawback of using Cloud RAN technology is the requirement on the fronthaul link capacity and latency. Analog fronthaul could enable more efficient bandwidth usage compared with digital fronthaul (e.g. today's CPRI).

Distributed architecture (presented in 2.3.2) reduces the capacity requirements heavily on backhaul link compared with analog and digital fronthauls by distributing

baseband processing to each access node. The distribution of baseband processing units will not only increase the cost of hardware in access nodes but also prevent having instantaneous radio resource view at a central radio resource manager (RRM). This in turn will result in sub-optimal radio resources allocation within the network.

Hybrid backhaul architecture could resolve the problem of lacking an instantaneous radio view at a central RRM. Control plane could be separated from baseband processing and directly fed to the central RRM. This increases the latency requirement to similar level as in fronthauling, but the most capacity consuming user plane would still use fully distributed baseband processing. This approach would enable efficient CoMP on downlink as in fronthauling.

Table 1: Comparison of fronthaul and backhaul technologies carried over millimeter wave link

	Analog fronthaul	Digital fronthaul	Backhaul	Hybrid backhaul
Capacity requirement	Moderate	High	Low	Low/Moder.
Latency requirement	High	High	Low	High
Spectrum requirement	Moderate	High	Low	Low/Moder.
Access node cost	Low	Moderate	High	High
CoMP gain in access link	High	High	Low	High in DL Low in UL

Few key aspects of fronthaul and backhaul technologies are compared in table 1. Further comparison of fronthaul and backhaul technologies will be presented in chapter 3.1.1 by elaborating on the challenges related to both technologies.

2.4 Definitions and practical deployment of access and aggregation nodes

This chapter provides definitions of access and aggregation nodes which are essential elements of this study of 5G networks. Additionally, the work related to installing a large number of small access nodes and relatively smaller number of somewhat larger aggregation nodes are introduced. Figure 13 shows network elements of an UDN deployment. Besides access and aggregation nodes, network controller is also an essential element for managing the nodes, and macro base station can be used to provide overall coverage.

2.4.1 Definition of an access node

An access node can be thought of as an evolution step of today's pico or femto cell. Coverage area of an access node is assumed to be very limited, meaning that a continuous coverage (as introduced in chapter 2.1.4) requires large amount of access nodes. Hardware of the access node is dependent on the network architecture –
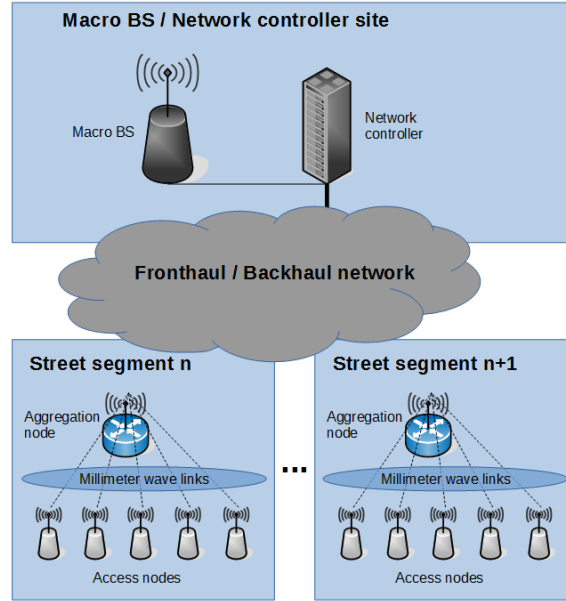
Figure 13: Network architecture of an UDN deployment

centralized or distributed. It could be acting only as a remote radio head (centralized architecture), or it can include all baseband processing functions (fully distributed architecture).

An example of what an access node could look like is shown in figure 14. The picture is from a concept node developed jointly by Ericsson and Philips: They call it as "Zero site". The thick part of the pole contains all base station radio hardware and antennas.

Considering similar structure for the acces node in this thesis, the upper part of the cylindrical unit could be reserved for uplink purposes and lower part for access link purposes. Antenna arrays could be split respectively. Millimeter wave fronthaul/backhaul antenna array could reside in the upper part of the cylinder as most commonly the opposite end of the link (an aggregation node) would reside at same of higher level from ground. Lower part would be natural to reserver for access link as pedestrians and vehicular users would reside mostly closer to ground level.

2.4.2 Definition of an aggregation node

An aggregation node could be thought of as an evolved version of the existing V or E-band millimeter wave link devices. Today's commercial devices are mostly equipped with fixed direction antennas which are capable of running only single point-to-point link per device. Beam steering is possible by manual adjustments on-site.

The aggregation nodes of the envisioned 5G networks could employ a massive MIMO antenna array capable of serving several (around 10) point-to-point connections and steering beams (quasi) freely towards served access nodes. The antenna

Figure 14: Zero site concept released together by Ericsson and Philips. The thick part of the pole contains base station hardware and antennas. [47]

array could consist of hundreds of small antenna elements. This is enabled by short wavelengths and advancements in millimeter wave integrated component manufacturing.

Backhaul connection of an aggregation node towards mobile core networks would be in most cases optical fiber due to high capacity requirements. Aggregated traffic of around 10 access nodes could reach 100+ Gbps, which makes fiber the most viable transport media.

2.4.3 Deployment considerations of an access node

There are no such large scale deployment examples of small radio nodes which could be used as an example, so the analysis of practical deployment will rely on other civil engineering and telecommunication equipment installation projects. The deployment process steps and skills required to perform the deployment will vary depending on the installed hardware plug and play capabilities. Two scenarios of deployment will be analyzed here: access node hardware without plug and play link discovery capability and access node with such capability.

In chapter 2.1.4, a concept of continuous ultra-dense network (UDN) was presented. In this concept a very likely mounting location of an access node could be on the same poles as the street lamps. Since there are no existing references of such large scale small radio equipment installation projects, a project of modernizing lamp posts from traditional lamps into LEDs will be used as an example. City of Milan

in Italy decided to modernize all over 140 000 street lamps into LEDs during years 2014 and 2015. The project total cost was around 38 million euro, which translates to slightly below 300 euro per street lamp. The cost includes all civil engineering work to replace the lamp and the cost of the new hardware. It also includes around 500 control panels and remote control system of the lights. [48] [49]

Access node with plug and play capability

When plug and play link discovery capability is included in the hardware to be installed, the manual installation work can be done by any basic mechanic who can perform the mechanical installation and connecting electricity. The links would be automatically set up and no link alignment skills would be required from the mechanics. Installation work amount of an access node with plug and play capability could be rather close to the installation work of aforementioned LED lamps. Installation process of single lamp pole could be as follows (assuming all required tools and materials are available at the installation site):

1. Cut power from the lamp pole (unless whole street has been cut off from main switch).

2. Lift mechanic(s) to the correct height with needed tools and hardware using e.g. a manlift.

3. Route power cable for the new node preferably inside (but could be outside) of the pole.

4. Mount access node hardware on the pole and connect newly router power cable.

5. Restore power to the lamp pole (if possible locally) and make sure the node turns on and starts discovery process.

Other steps such as node integrity/security checks and possible pre-pairing to correct aggregation node over legacy radio (e.g. WiFi or LTE) has been omitted in the above list of steps. They could be added before and after the listed steps, respectively.

Access node without plug and play capability

In case an access node would not have plug and play capability for link discovery process, it would be required to manually align the link. The negative effect of manual aligning of millimeter wave links is that the beams are likely to be very narrow and any misalignment might result in poor operation of the link. Due to precision work requirement the technician installing the equipment should have special knowledge about how to align the link and how to configure some link parameters. The labor cost of such technician is supposed to be somewhat higher than that of a basic mechanic, and the work amount due to precision alignment could be even doubled. Similar installation process of single lamp pole will be required as presented above, but with the added steps of aligning the link before powering on the node and verifying the link operation after powering on the node.

Other steps such as node integrity/security checks and possible pre-configuration of the node have been omitted here. Also the last step of link verification is not possible to be performed in case the power per lamp pole cannot be controlled individually. In worst case it might be required to return for post-installation checks after all hardware along the street have been installed and power is turned back on.

2.4.4 Deployment considerations of an aggregation node

An aggregation node could be installed on a street side as explained in chapter 2.1.4 – for example on a wall of a street canyon side. The physical installation of an aggregation node could be thought to be similar as installing today's microwave/millimeter wave link equipment. Difference between installing today's nodes and a 5G aggregation node could be that link alignment is expected to be performed electronically by means of beamforming using a phased antenna array. That means the technician do not have to perform it manually. As the aggregation node will most certainly connect to a wired uplink connection, also the configuration of the equipment could be pushed from a configuration server instead of local configuration. The installation steps of an aggregation node to be installed on a wall could be as follows:

1. Route power cable and uplink fiber from inside of the building to the outer wall.

2. Use manlift or crane to lift technician(s) to the correct installation spot.

3. Mount the aggregation node hardware on the wall

4. Connect pluggable optics and fiber to the node

5. Connect power cable to the node

6. Turn on power and make sure the node starts up properly

7. (Optional) Configure initial uplink and other parameters on the aggregation node

As mentioned earlier, the last step could be omitted if the aggregation node can obtain all necessary configuration directly from a configuration server. That would be useful in the sense that if no configuration is required at the site, the technicians do not have to possess any special skills to install the node (i.e. a basic mechanic could do the job as well).

2.4.5 Other network deployment considerations

The installation of both node types in city-wide scale will require careful planning of the project. Logistics of such project are in an important role to have correct materials available in right place at right time. Required traffic signs and possible detours have to be planned in advance if the deployment operations block a lane or in worst case a whole street from traffic. The installation itself requires at least a manlift to raise one or two technicians (depending on case) to the installation height. Details of deployment of each node type were covered in previous sections.

3 Challenges review and research introduction

3.1 Challenges review

This section will list the challenges of using today's backhaul and fronthaul technologies if applied to the small cells of an envisioned 5G UDN. The challenges are considered from technological, operational and cost perspectives.

3.1.1 Challenges of fronthaul and backhaul technologies

Background of fronthaul and backhaul technologies were covered in chapter 2.3. From joint processing point of view the fronthaul based cloud RAN architecture is the most feasible, but it is also the most capacity demanding. The most capacity optimized solution is to use backhaul with fully distributed RAN architecture where baseband processing is included in every radio node. Full distribution of baseband processing will prevent the use of joint processing and leads to sub-optimal radio resource allocations. Also extra hardware features in radio nodes naturally increase their cost. This section will introduce the above mentioned challenges in detail.

Fronthaul

Chapter 2.3.1 introduced the possible fronthaul methods; digital and analog. The challenges related to both are explained in this section.

Digital fronthaul link capacity requirement was presented using commonly used CPRI standard as an example. Comparing the calculated values for LTE with those of an exemplary 5G system model by Kela et al. in [22] shows the differences between today's LTE system and the envisioned 5G UDN. The FFT size is actually lower, 1024 compared with 2048 of LTE, but the subcarrier spacing is much larger, 312.5 kHz. The resulting F_s of this 5G model is then 320 MHz. Multiplying the sampling rate and sample size of 30bits gives a total data rate of $R = 320MHz * 30bit = 9600Mbps$ for single MIMO channel. Conservatively multiplying this with 8 MIMO channels gives us already a data rate of 76.8 Gbps. In reality the amount of MIMO channels are expected to grow even larger thus increasing the required fronthaul data rate even more. Even if the sampling rate and sample sizes could be optimized, the resulting data rates are still very high. This kind of data rates are not easily (and especially cost effectively) carried over today's bearer networks. To make the numbers even more hideous, the expected number of nodes per square kilometer is going to increase by at least tenfold, so the amount of collected fronthaul data from a square kilometer area would increase by multiple decades. There are no solutions presented yet how to solve this challenge, but clearly much optimization would be needed.

Two options were presented for analog fronthaul. First option is to move low carrier frequency signals to millimeter wave area while stacking multiple carriers side by side. The benefit of simply moving multiple carriers is the bandwidth efficiency, e.g. 5 pcs of 200 MHz carriers would result in 1 GHz total bandwidth occupancy. The available spectrum in millimeter wave area can be up to 5 GHz so in that sense there should be no issues. The issue might arise from the millimeter wave practical system architecture. In millimeter wave area the amplifiers used need to operate on

very small dynamic range in order to maintain power-efficient operation. Stacking multiple carriers side by side would result in multi-carrier millimeter wave system causing possible issues in peak-to-average power ratio (PAPR) (as introduced in chapter 2.2.3). Second option presented was to use on-site-coding (OSC) based solution. To avoid SNR degradation, it was found in [50] that the length of used spreading code (i.e. spreading factor) needs to be at least double of the number of codes (i.e. channels) used in the system. Using the same example of 5 x 200 MHz streams as above, the resulting code length would be minimum of 10. Spreading the 200 MHz signal with spreading factor of 10 would result at least 2 GHz of required spectrum. Moreover due to the proposal of splitting the wideband signal into several narrow band signals before coding introduces the same limitation as in the first option of analog fronthaul: multiple carriers result in higher PAPR which might be an issue with millimeter wave systems.

Backhaul

Chapter 2.3.2 proposed two types of backhaul links; "fully distributed architecture" and "hybrid architecture" which separates control and user plane baseband processing. Backhaul link requirement on capacity perspective will be roughly the same as the payload traffic on the access nodes with some overhead. This fact would enable the use of even the millimeter wave links that are available today. The challenge in introducing distributed baseband processing in the access nodes would be the resulting baseband processing delay. Baseband processing also prevents the full advantage of combining gain in coordinated transmissions as the process would benefit most by combining the radio samples before the baseband processing. Coordinated multipoint (CoMP) transmissions using backhaul links requires very high synchronization accuracy of the network and very low delay backhaul links. Hybrid backhaul combines the good aspects of both fronthaul and backhaul:

- User plane as most capacity intensive element would be processed in the access nodes and only baseband data would be sent over the hybrid backhaul link.

- Control plane would consist of sending channel measurements to a centralized radio resource processor. The centralized processor would see the "big picture" and it would allocate optimal precoding vectors for all access nodes in a given area to optimize the usage of radio resources.

Also the current simulations by Kela et al. in [22] refer to scenario where coordinated processing of radio samples is not employed, but instead interference is centrally controlled by utilizing narrower beams and predicting how the beams should be aligned. The probability of causing high interference is lowered when "pencil beams" are utilized.

Summary of backhaul and fronthaul challenges

The above sections introduced challenges related to different fronthaul and backhaul architectures. Each architecture have their advantages and disadvantages, but from tradeoff point of view the most prospecting architecture would be hybrid backhaul.

As concluded in chapter 2.3.3, hybrid backhaul would enable similar performance as fronthauling in downlink direction allowing efficient CoMP. In uplink direction joint reception is not possible, but already in today's systems downlink is the dominating direction of data traffic [51].

3.1.2 Challenges of different link media

The backhaul link can be based on wireless or wired link. This section will list challenges related on most feasible link types.

Microwave and millimeter wave links

Traditional microwave links have been widely used as last mile backhaul links for 2G, 3G and 4G mobile technologies. The capcacity requirement in these mobile technologies grows with the generation, but even a 4G macro base station with multiple sectors can survive with a 500 Mbps to 1 Gbps link. Shorter range millimeter wave links have been introduced with higher spatial reuse capability to cover short hops mostly in urban environments. These links can also provide higher data rates as due to much wider spectrum availability in millimeter wave area they can utilize tenfold of bandwidth per link compare with traditional microwave links.

Millimeter wave links possess the same fundamental properties as microwave links or any other wireless link. Any blockade on the signal path may and will cause attenuation on the link. The higher the frequency, the more severe the effect is. Heavy rain is also one big source of attenuation on microwave links, but due to short range of millimeter wave links it is rather easily compensated by few decibels of margin in link budget.

Deployment of today's microwave and millimeter wave links require skilled technicians. The alignment of links is very delicate process and requires high precision. If links are poorly aligned, it will result in high attenuations and low attainable bit rates. Also the reliability of poorly aligned links suffer, as rain and other items interfering with the link path will cause more severe fading. [33]

Direct fiber

Fiber links have been used widely for macro cell backhauling. It has been popular due to its high reliability and basically limitless capacity. Laying fiber is rather expensive because firstly it requires opening the ground, and secondly it requires skilled technicians terminating the fibers by splicing "pigtails" to both ends of every fiber cord. The explosion in number of nodes happens when small cells are deployed instead of macro cells, which at the same time rises the amount of required backhaul connections. Deploying fiber to every small cell is most likely economically totally prohibitive. [33]

Relaying

There has been also discussion of using the access nodes as relays for other access nodes. This could be useful when some access nodes do not have any reasonable means of having direct connectivity with fiber or even an aggregation node site. It

can also provide redundancy in the network instead of relying on single feeder only. [33]

Challenge of using relaying is that the access node acting as a relay needs to have uplink capacity worth all the relayed nodes plus itself. With fronthauling this capacity requirement might not be feasible at all. With backhauling the capacity requirement could be (at least closer to) acceptable.

Summary of link type challenges

The selection of the kind of backhaul link should be selected based on feasibility of the different technologies and should be evaluated case by case. If fiber is readily available, it should be prioritized due to its high capacity and reliability. It is also possible to use the node at fiber termination location as wireless aggregation node for other small cells which are out of the fibers reach. As the number of nodes is expected to grow exponentially in an UDN layout, it might become very cost intensive to use skilled technicians carefully aligning every small cell backhaul links. Also having only a single static link will not provide redundancy in case of any failure in the feeding aggregation node. An automatic link discovery method in access nodes would help in the installation of the nodes by providing plug and play connectivity, and also it can scan for redundant aggregation nodes in case of serving node failure.

3.1.3   Physical installation challenges of access and aggregation nodes

Chapter 2.4 introduced the physical installation of access and aggregation node hardware. The challenges of the installation are related to the complexity of required work – especially due to the high number of installed nodes and delicate alignment of narrow beams (in case without plug and play function). Basic principle of the 5G network design from deployment point of view is to find trade-off between hardware cost and labor cost in order to optimize the total CAPEX. Labor cost can be reduced to some extent by adding more capabilities to the hardware/software (e.g. link auto-discovery and auto-configuration). On one hand the amount of labor can be reduced, and on the other hand the requirement of the labor skill level can be also reduced. In today's network deployment projects one of the cost intensive parts of total CAPEX is related to human labor. Plug and play deployment capability has been set as the target for next generation networks. The nodes should be able to establish secure link towards core network and register as part of the RAN automatically with minimum human interaction. The details of proposed plug and play methods will be introduced in following chapters.

3.2   System assumptions

This chapter will introduce the assumptions related to the study in chapters 4 and 5.

3.2.1   Network layout assumptions

The network under study is assumed to follow the concept of Continuous UDN (presented in chapter 2.1.4). The aggregation nodes are assumed to be positioned on

walls of a street canyon, at same level or higher than the access nodes. This should ensure mostly line-of-sight paths towards access nodes.

Access nodes are assumed to be installed on lamp posts along the street. From lamp posts there are good chances of having line-of-sight or near-line-of-sight paths towards pedestrian and vehicular users, and good chance of line-of-sight towards at least single aggregation node.

3.2.2 Link technology assumptions

This study focuses on millimeter wave links operating at E-band (around 70 GHz). The band was selected based on three reasons:

1. The band is in global scale either lightly licensed or non-licensed. This will keep the licensing costs down even though the amount of links is expected to grow very high as result of network densification.

2. There are multiple gigahertz of bandwidth available at the E-band.

3. There are no significant attenuation spikes caused by atmospheric gases at the E-band.

Other possible frequency band could be the V-band (around 60 GHz). It is license free in most countries, but attenuation caused by oxygen absorption at 60 GHz might cause issues on longer link spans. The positive side effect of higher attenuation is lower level of spatially leaked interference.

The nodes are assumed to have antenna arrays with beamforming/beamsteering capabilities (background covered in chapter 2.2.1). It is assumed that the antenna arrays have large enough number of elements to be able to create very narrow beams.

Links between access and aggregation nodes are assumed to be line-of-sight (LOS). Near-LOS and Non-LOS links could work as well if the signals can for example bounce off a properly reflecting surface. Basically any propagation path with losses not too much greater than in free-space propagation environment could do.

3.2.3 Other related assumptions

The assumption of network state is not only restricted on initial network deployment scenario, but also expansion of the network. It is possible that in the first stage of network deployment only minimum amount of nodes is deployed to meet coverage targets. Later more nodes might need to be added for extended capacity purposes. The analyzed plug and play features of node deployment need to be able to function in both scenarios. The details of possible scenarios will be explained in the next chapter.

3.3 Scenario descriptions

The scenarios related to installing an access node can be considered from two dimensions: deployment scenarios and backhaul scenarios.

3.3.1 Deployment scenarios

Deployment scenarios cover the first dimension by specifying the network status when installation takes place:

1. **First node installation** (no load in aggregation node)

   - Initial deployment of the network
   - The network has no active users, so there is no possibility of interfering existing traffic

2. **Network expansion** (existing user traffic in aggregation node)

   - Existing network with active users
   - New nodes added for increased capacity or coverage

3. **Multi-hop backhaul** (both with load and without load in aggregation node)

   - Case when an access node does not have direct path to and aggregation node
   - The access node will rely for example on other access nodes acting as relays

First node installation can be thought of a special case of capacity expansion. The process is most likely going to be the same, but the main difference is that the aggregation node is serving other already existing access nodes with payload traffic streams. The process of adding new access nodes should cause minimum disturbance to the existing nodes. More detailed description of these functional requirements will be given in 3.4 and performance requirements in 4.1.

3.3.2 Backhaul scenarios

Backhaul scenarios cover the second dimension by defining the used fronthaul/backhaul technology (presented in chapter 2.3). The link types have different properties that relate to the operational and cost aspects. Differences and challenges of each link type have been covered in chapters 2.3 and 3.1.1. Possible scenarios are:

1. Analog fronthaul

2. Digital fronthaul

3. Backhaul (fully distributed architecture)

4. Hybrid backhaul (user plane distributed, control plane centralized)

The choise of fronthaul/backhaul technology will dictate the performance requirements of the used link medium. The suitability of different fronthaul/backhaul technologies with millimeter wave links was compared in chapters 2.3.3 and 3.1.1.

3.4   Detailed description of functional requirements

Six functional requirements were identified to be involved in the plug and play process of adding an access node to an ultra-dense network:

1. Pre-configuration of access node

2. Automatic link discovery and setup

3. Secure authentication and validation of access node

4. Time synchronization of access node

5. Accurate positioning of access node

6. Attaching access node as part of radio access network

Functional requirement 1: Pre-configuration of access node

The first functional requirement describes requirements on methods of pre-configuration of access nodes before or during installation. In principle the more intelligence is embedded in the node, the less manual work is required. In case a generic access node hardware is used, some manual work is required at some stage of network delivery process.

Assuming that generic access node hardware is shipped to site without any specific pre-configuration, there needs to be some way of defining which nodes need to attach to which network in locations where multiple operators are running similar equipment. The registration can be based on the access nodes itself or on the management system where the nodes will be attached. In general it may be more feasible to move the registration upper in the network as much fewer nodes needs to be touched.

A solution providing proper trade-off between efficiency (low cost) and high data integrity (few mistakes) need to be found. Security aspects are not considered here, but some degree of security is required on one hand to prevent unauthorized rogue access nodes from attaching to a network, and on the other hand to prevent theft of existing installed access nodes to be used elsewhere.

Functional requirement 2: Automatic link discovery and setup

The second functional requirement describes the requirements for the access/aggregation node search and link establishment processes. As mentioned earlier in section 2.2.3, there are challenges of using millimeter wave links in a dynamic environment. Even the discovery of other radio nodes requires certain intelligence in the aggregation and access nodes. Current mobile networks use low enough frequencies to support omni-directional antennas for communication, but in millimeter wave area it is mandatory to utilize beamforming and intelligent beam steering techniques. Possible technical solutions for implementing the radio node discovery process using electrical beam steering will be analyzed. The required elements and process will be described in detail and the discovery process will be simulated using MATLAB in chapter 4. The

simulation results and technical requirements will be further used in chapter 5 to estimate the cost effects of plug and play process to the overall network deployment costs.

The requirement for radio discovery is that the aggregation and access nodes need to find each other using a beam steering/alignment technique in order to establish a communication channel. The requirements for link discovery process are presented below:

- In order to achieve close to total plug and play installation, an access node should not require any specific configuration during the hardware installation and initial power on. This will reduce the required labor time during installation.

- Person installing the access node could have a tool (e.g. a mobile terminal with specific software) which has capability of communicating over e.g. LTE with a RAN controller responsible of managing the access and aggregation nodes. An identifier tag (based on e.g. NFC) of an access node could be read and reported to the RAN controller using the tool.

Functional requirement 3: Secure authentication and validation of access node

The third functional requirement defines requirements for secure pairing of access and aggregation nodes once the radio discovery process of previous step has been completed. Before transmitting any protected data over the established link, the remote node should be identified as an allowed node and the communications link should be possibly encrypted.

Secure registration could have at least the following two steps:

1. Whitelist based filtering could be used in aggregation node. If the identifier (similar to IMEI in today's mobile terminals) of the access node is not in the whitelist, the security pairing attempt would be rejected. This function would effectively increase the security by rejecting rogue access nodes from being attached to the network (assuming the identifier is not known by the attacker).

2. Secure key exchange between aggregation and access nodes is required in order to establish the encrypted communications channel. Possibilities are at least:

   - SIM-card based authentication (this could use physical SIM-card / embedded-SIM / soft-SIM with remote provisioning functions) [52]

   - Certificate based authentication, where pre-delivered certificates would be used to ensure secrecy

Functional requirement 4: Time synchronization of access node

Use of TDD and many coordinated transmission functions (such as enhanced inter cell interference coordination) require highly accurate phase and time synchronization of the network to be able to function properly [53]. TDD is a possible choice for

access layer duplexing method, but could be also used for the uplink connectivity. Sophisticated coordinated functions are also expected to be added to 5G systems. Thus, it is required that the access nodes are able to perform accurate phase and time synchronization between each other to support these functions.

Functional requirement 5: Position determination of an access node

Terminal positioning in 5G networks has been set as one of the design targets to enable and enhance existing location and context aware services [2]. Satellite based positioning in urban street canyons and indoor locations will result in poor accuracy and availability. Dammann et al. proposed a cooperative method for mobile terminals to position each other by not only signaling with base stations, but also directly with each other [54]. The same methodology could be applied to plug and play deployment of access nodes so that the access nodes would cooperatively position each other. In order to obtain absolute positions, some nodes need to have their absolute positions known acting as "anchors" [55]. It could be assumed that aggregation nodes know their locations accurately by measurements during installation. Also every access node should be able to communicate at least with two aggregation nodes for redundancy purposes. In case there are too few anchor nodes in the area, some access nodes could be given accurate positions during installation so that they can also act as anchors.

Functional requirement 6: Attaching radio node as part of an ultra-dense network

Once an access node has a secure communication link to the UDN backbone and the location of the node has been determined, the access node has to be registered as a working part of the whole UDN access layer. This means that the node need to be included in radio resource management and mobility management processes. Terminals attached to the UDN cannot be served before the access node has been properly attached to the UDN. This functional requirement is relatively wide and can be split into several sub-requirements (e.g. signaling requirements between UDN and mobile core network).

3.5 Scope of analysis work

3.5.1 Selection of scenarios and functional requirements for analysis

To limit the scope of the thesis, not all listed scenarios and functional requirements will be covered in this study. The scope of the study will be to analyze a set of selected scenarios and functional requirements as follows:

- **Scenarios**

  - The choise of fronthaul/backhaul technology has no big impact on the study of this thesis, but one technology will be selected for completeness. Based on the comparison of fronthaul and backhaul technologies, hybrid

backhaul will be selected as it has been found to be the most promising compromise for the future.

– The installation scenario is selected to be network expansion, but the same methodology can be applied to first node installation scenario. As the system dimensioning will be done based on a loaded situation, it should perform similarly – most likely even better – in a non-loaded situation (first node installation).

- **Functional requirements**

  – The analysis will be conducted on the second functional requirement – automatic link discovery and setup. The study will focus on finding suitable technical methods for radio link discovery and establishment process based on the selected scenario and other relevant assumptions.

3.5.2 Outline of the analysis

The study will cover both technical and cost related aspects. The outline of this study is as follows:

- **Technical analysis** (covered in chapter 4)

  1. Performance requirements are defined for the selected functional requirement
  2. Possibilities of technical options will be listed
  3. Technical analysis model will be constructed on the most feasible solutions identified in the previous step
  4. Results of the technical analysis will be presented

- **Cost analysis** (covered in chapter 5)

  1. A cost model will be formulated to analyze deployment costs of an ultra-dense network
  2. The cost effects of plug and play deployment of access nodes will be included in the cost analysis

Finally conclusions will be given based on results of above analyses, and required further study will be proposed.

4 Technical solutions and evaluation

This chapter proposes technical solutions and evaluation of selected scenarios and functional requirements involved in the plug and play radio node addition process for ultra-dense networks. Detail selections of scenarios and functional requirements were presented in chapter 3.5.

This thesis focuses on the study of the automatic link discovery and setup process which was introduced in chapter 3.4. This chapter will propose detail technical requirements and their analysis.

4.1 Performance requirements

The requirements for plug and play installation can be considered from complexity, reliability and time requirement perspective. As stated in chapter 3.4, the complexity of physical installation should be kept minimum in order to reduce the costs on manual labor. The absolute minimum amount of work is to physically mount the access node for example onto a pole or wall, connect power cable, and switch the device on (in case a physical power switch is present).

The reliability aspect in plug and play performance requirements refers to the success rate of the pairing process between access and aggregation nodes. Should the process fail completely for example due to missing LOS path or strong enough reflected path, the only solution is to re-send a mechanic/technician to manually fix the problems. In a UDN case (explained in chapter 2.1.4) the amount of nodes is likely to be very large, so the increasing failure rate of the pairing process would have severe increase in labor costs. It might also delay the rollout of the whole network.

Time requirement on the plug and play process completion is not very strict. It is acceptable to spend more time on the discovery process if the reliability of the pairing can be kept sufficiently high. The process will be started immediately after powering on the device. It would be useful to be able to finish the process while the mechanic/technician is still at the site to be able to verify the process completion before moving to next site. 5 minutes will be set as a target maximum time for the discovery process, but preferably it should be less.

Also interference caused to possibly ongoing payload traffic must be minimized. If the plug and play pilot signal requires relatively high transmission power compared with an ongoing payload signal, it will affect the overall performance of the link. A target occupancy of 0.1 – 0.5% of total radiated power (i.e. EIRP) or time (in case of time division multiplexing) will be set for the pilot signal. This way minimum of 99.5% of total node capacity will be reserved for the possible payload traffic. This kind of target setting would in fact allow the pilot to be always on without consuming system resources and power too much.

4.2 Solution derivation

The main target of the automatic link discovery and setup process is to reduce manual configuration work as much as possible. The installed radio nodes need to

be able to autonomously figure out with which nodes they need to inter-operate with. Fully autonomous configuration is probably impossible to achieve, but with few boundary conditions a high degree of automatic operation is possible. Newly installed access nodes should be able to determine where a possible aggregation node resides in relation to its own position. This is required to point a narrow beam towards the access node for backhaul link establishment purposes.

Detailed analysis of the radio discovery process requires certain architectural choices to be made before modeling. These architectural aspects will be presented in below sections with justification of the made choices.

4.2.1   Architectural selections

This section will cover different architectural choices which the automatic link discovery function analysis will be based on.

Choice of pilot transmission sequence

The link discovery process involves two types of nodes: access node and aggregation node. A choice should be made which one should transmit pilot first and which one should listen. In general the process could consist of two steps. Assuming nodes are named as "Node 1" and "Node 2" the steps would be as below:

1. Node 1 will transmit pilot signal omnidirectionally as no prior knowledge about locations is assumed. Node 2 should scan the area with narrow beams trying to find the pilot signal source direction.

2. Once the beam training is completed for Node 2, it should start transmitting pilot with narrow beam towards Node 1. Node 1 should scan the area with narrow beam trying to find direction of Node 2.

If an access node would transmit the pilot first (i.e. act as Node 1), having no prior knowledge about the direction/distance of aggregation nodes it would have to transmit omnidirectionally with high power. This would potentially lead to causing interference to existing payload streams in the area as the amount of "shouting" access nodes can be large. Another problem is that how does the access node learn when an aggregation node has found it, how many aggregation nodes have found it, and when it should start listening for aggregation nodes.

In case aggregation node would act as the Node 1, it could transmit the pilot omnidirectionally (in fact hemispherically) with low proportion of total resources in order to avoid interfering with existing payload streams possibly running through itself. Access nodes acting as Node 2 would only listen until aggregation node directions are found. Then the pilot transmission from access nodes is very controlled, as they are only transmitting with narrow beams towards desired directions.

Hence, for this analysis the latter option will be selected (aggregation node transmitting and acces node listening in first step).

Choice between plain power detection and coherent pilot signal detection

Two options for determining direction of beams are compared: pure power detection and coherent pilot signal detection.

Benefit of plain power detection is that it is very simple and it does not necessarily require a dedicated pilot signal to be transmitted. It could rely on sidelobes leaking power from running payload transmissions. Only if there are no access nodes attached to an aggregation node, it would have to transmit some waveform. The signal power from each direction could be integrated for longer time in order to ensure proper detection. Problem of using plain power detection is that there is no way of knowing for sure if the signal is received from an aggregation node – and even if it is, the node could belong to a neighboring operator.

With coherent pilot signals the disadvantage is that a specific pilot waveform/code needs to be transmitted. On the other hand the detector can determine that the received signal indeed is the one we want. Coherent detection also allows longer integration times, so the transmitted pilot signal can be of relatively low level.

For this analysis the coherent pilot signal detection will be assumed. The link discovery process will be much more reliable as the node transmitting a pilot signal can be identified.

Choice between single carrier and multi carrier architectures

A single carrier system is assumed in this analysis due to lower PAPR than in multi carrier systems. Millimeter wave systems with large antenna arrays do not perform very efficiently with large dynamic range requirements, as every antenna element is assumed to have an individual small PA and ADC/DAC. Large dynamic range requirement decreases efficiency and increases power consumption of the components. This is further explained in chapters 2.2.1 and 2.2.3.

Choice between FDD and TDD methods

The choice of duplexing method could in principle follow the duplexing method chosen for access link, as it would be natural to allocate similar capacity symmetry for uplink and downlink for both links. With TDD the symmetry between DL and UL can be modified easily based on requirement. With FDD the adjustment (at least on the fly) is not usually possible due to fixed filters for DL and UL bands.

Besides capacity allocation, latency is another aspect when considering the selection of duplexing method. FDD has benefit over TDD in latency, as FDD supports full-duplexing whereas TDD only supports half-duplexing. Instantaneous data rates can however reach higher levels with TDD as whole bandwidth can be occupied per direction. Latency should also not be a big issue in TDD as the radio frames are supposed to be very short.

So the radio discovery process should in principle work similarly with both duplexing methods, but TDD mode might cause some delay to the process. A link discovery pilot signal could be constantly active when using FDD mode (except

when using TDM as multiplexing architecture), so it will be chosen for this study for simplicity.

Choice between different multiplexing methods

Three alternatives of multiplexing methods will be compared: frequency divison multiplexing (FDM), time division multiplexing (TDM), and code division multiplexing (CDM). The choice will be made by evaluating each method against general applicability for millimeter waves (discussed in chapters 2.2.1 and 2.2.3), and the performance requirements set in chapter 4.1. The suitability of each method for the payload signals transmission is not compared here. For more comprehensive analysis both transmission types should be considered jointly.

- **Summary of performance requirements:**

  - Omnidirectional (in reality closer to hemispherical) pilot transmission
  - Always-on pilot transmission
  - Pilot resource allocation around 0.1 - 0.5% of total available resources
  - Discovery process time constraint of 5 minutes or less

- **Frequency division multiplexing:**

  Frequency division multiplexing (FDM) method could in principle facilitate always-on omnidirectional pilot signal transmission by allocating certain amount of the total used bandwidth for the pilot signal. If the total signal bandwidth is assumed to be 500 MHz, it would mean that with 0.1% resource allocation a 500 kHz wide block would be allocated for the pilot signal.

  An issue with narrowband transmission (without frequency hopping) is its sensitivity against frequency selective fading. This however should not be an issue if the link is assumed to be mostly line-of-sight.

  Bigger issue in this case is the high dynamic range requirement of the transmitter – particularly per antenna element transmitter chain. While the differences on the levels between subcarriers in the far-field of the antenna array would not seem very high, on per antenna element they are. The reason is that only the pilot signal requires to be transmitted omnidirectionally, and the other subcarriers (payload signals) only need to be sent using very narrow beams. That would mean the total transmitted energy of the pilot subcarrier is much higher than the other subcarriers. Assuming every antenna element would have individual transmitter chain from baseband to RF, it would mean the pilot signal should be generated using same baseband modules with payload streams. If single pilot subcarrier needs to be very much higher than rest of the subcarriers, it would create problems by requiring very high dynamic range from the baseband fast Fourier transform (FFT) modules.

  To summarize, FDM could meet the performance requirements defined in chapter 4.1, but it fails in keeping dynamic range requirement at suitable levels for millimeter wave systems.

- **Time division multiplexing:**

  Time division multiplexing (TDM) performs better in terms of dynamic range requirements than FDM, so from that perspective it could be a valid candidate technology. The issue with TDM is that the pilot resources need to be allocated as fraction of total time. It means that the pilot would be sent with high energy, but only about 0.1% of the time. For example during a 0.5 second time period the pilot would remain active for 500 microseconds.

  This kind of allocation could possibly result in high discovery process time, as the receiving access node would have to keep its beam stationary at each step for the total duration of the time period within which the pilot should appear. Using the mentioned 0.5 seconds and 500 microseconds as the assumed values, it would mean that the beam should be kept at least for 0.5 seconds at each step. Depending on the beam steer step sizes and the horizontal and vertical search angles, the amount of steps could be several thousands up to tens or hundreds of thousands.

  To summarize, TDM could meet the general millimeter wave system design recommendations, but it might fail to meet the time constraint on the performance requirements defined for the radio discovery process.

- **Code division multiplexing:**

  Code division multiplexing (CDM) divides resources by allocating different lengths of codes depending on the resource requirement. Used spreading factor defines the power allocation of the code. Using this property, CDM could facilitate always on omnidirectional transmission with low resource allocation.

  For the purposes of automatic link discovery, a code should be transmitted omnidirectionally with such level that the EIRP of the pilot signal at a boresight direction of a payload signal would be around 0.1% of the payload signal EIRP.

  To summarize, CDM based architecture seems to fit both the general requirements of millimeter wave systems and the performance requirements set for the automatic link discovery process. Hence, CDM is assumed in this study. The feasibility of using a single wide beam with CDM in the aggregation node will be analyzed in chapter 4.3 by performing a link budget analysis.

Method for triggering pilot transmission

Possible methods of transmitting pilot signal downstream from an aggregation node are:

1. Always-on pilot transmission: The pilot signal is transmitted always in the background without interfering other ongoing payload traffic.

   - Single low-energy pilot signal broadcasted using a wide beam
   - Multiple low-energy pilot signals broadcasted using multiple narrow beams to increase gain of each beam and limit the directions to reduce interference.

This method naturally uses more system resources as multiple beamformers are required.

2. Periodical pilot transmission: The plug and play pilot signal could be transmitted on certain times of day/week to avoid generating interference during busy hours.

   - This could potentially cause long delay of newly installed access nodes being able to register to the network.
   - Exact time frame of the search process might be hard to set efficiently. For examoke if a mechanic/technician should verify on-site that the plug and play process is successful after installation, he would need to perform installations at certain time (or make second visit after the periodic process has run).

3. Manually triggered pilot transmission: The pilot could be triggered manually by network maintenance staff when network expansion is scheduled. This option is most effective on network resources usage, but it requires more labor on network planning and operation.

The option of "always-on pilot transmission" with single low-energy pilot is selected to be studied further in next sections. Reasons are the benefit of low resource utilization during pilot transmission and it being as close to real plug and play as possible. If the solution is found feasible, it could be also adapted to the other two solutions. This way in any case the pilot would not interfere with existing payload signals.

4.2.2 Detailed link discovery process

This section will describe the process steps related to the functional requirements of automatic link discovery process defined in chapter 3.4. The process steps are to some extent dependent on the result of the first step of the plug and play process "pre-configuration of access node" (also introduced in chapter 3.4). Since it has not been analyzed further in this thesis, following assumptions need to be made:

1. Identifiers have been installed to the access nodes during manufacturing:

   - ID of the access node is stored in a tag attached to the access node (e.g. QR-code or NFC-tag).
   - Code of a certain operator (who ordered the device) has been stored in the access node. This allows the access node to find only aggregation nodes belonging to this specific operator.

2. Mechanic/technicial has scanned the access node ID-tag during installation with an installation tool.

   - The tool has reported the scanned IDs over e.g. LTE connection to a network controller in charge of the aggregation units.

- Additionally the location of the node could have been reported during the scanning (based on e.g. some Global Navigation Satellite System (GNSS) technology).

3. Aggregation nodes in the area are transmitting pilot signal and scanning for the reported access node IDs.

   - Network controller will instruct the aggregation nodes to listen for specific pilot sequences after a device ID has been reported to be transmitted by an access node.

   - If locations of access nodes were reported while scanning the tags, it is possible to trigger the search process in a constrained area.

Process steps following above assumptions (flowchart in figure 15):

1. Once an access node has been powered on, it will search for aggregation node pilot signals from all directions using a narrow beam (by utilizing beamforming)

   - The search is only based on an operator specific known waveform/code
   - All found aggregation node directions will be recorded

2. Once beam directions for possible aggregation nodes have been found, the access node will start transmitting its unique pilot signal towards the aggregation node with strongest received pilot signal.

3. An aggregation node will find the pilot transmitted by the access node whose ID was reported.

4. Once the aggregation unit has found an access node they will establish a communication channel

   - Access node will report to network controller how many aggregation nodes were found during the scanning

   - Network controller will instruct the access node to start pairing with the next strongest aggregation node. This will be repeated until link has been established with all found aggregation nodes.

5. When an access node has finished the link establishment with possible aggregation nodes, the network controller will instruct the access node to switch to a serving aggregation node

   - The serving aggregation node is decided by network controller. It could be the aggregation node with strongest signal or another aggregation node with less load.

   - Access node will store the link information about rest of the aggregation nodes in database for redundancy purposes

6. Search process is finished for this specific access node. The network controller informs remaining aggregation nodes in the area to stop searching for this specific access node.
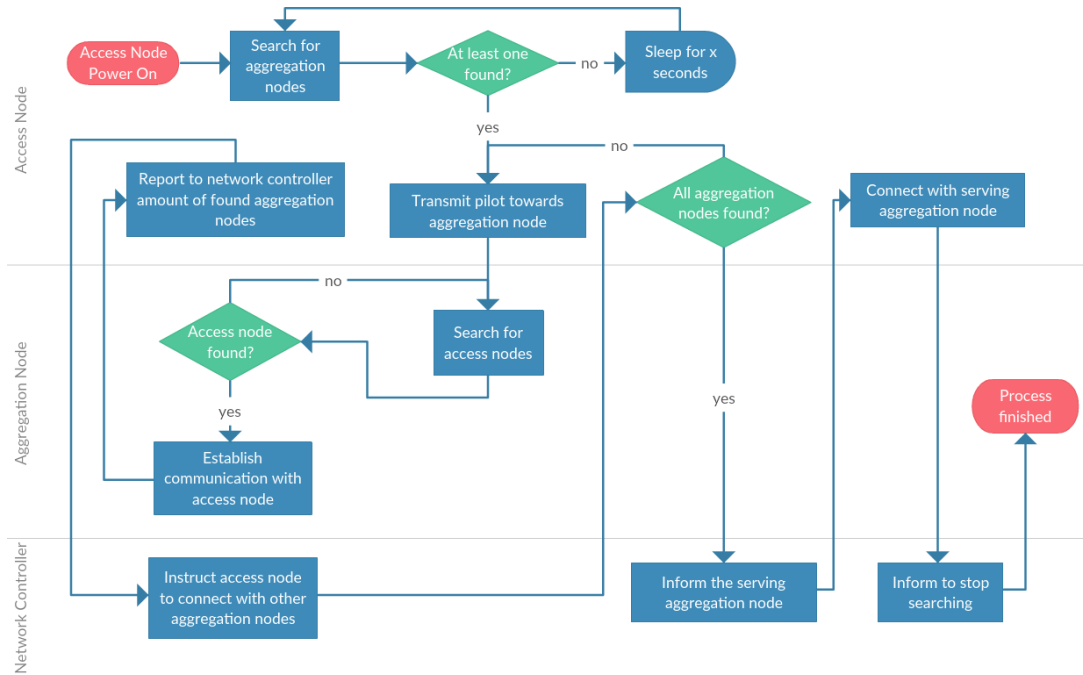

Figure 15: Automatic link discovery process flowchart

4.3 Technical analysis model construction

This section describes the environment and methods of analysis of the chosen technical model. The technical analysis is conducted using Matlab simulation and spreadsheet calculations.

4.3.1 Geometry model

The simulator is constructed to operate in 3D space with Cartesian coordinates. The simulated scenario consists of single aggregation node and single access node which are dropped in the x-y-z grid according to following assumptions:

- Environment assumptions:
  - The assumed environment of the UDN under study is positioned along a straight road where an aggregation node is located at a street side and is feeding multiple access nodes which are installed on lamp poles
  - The street is assumed to run straight along y-axis, lamp poles are evenly distributed along the street (middle of the road or on one side of the road).

- Aggregation node assumptions:

  – The shape of the aggregation node antenna group is assumed to be a panel, which in theory enables beamforming in nearly hemispherical area. An example of such antenna can be seen in figure 8.

  – The aggregation node is assumed to be positioned along a street side so that the boresight of the panel antenna points on positive (or negative) x-axis – depending on which side of the street the node is located.

  – The location of the node is fixed, for example at origin of x-y-plane with height set somewhere in positive z-axis (0 meaning ground level). The height is assumed to be same or higher than the heights of the access nodes (minimum around 2m but at least 10m is more reasonable for better visibility to multiple access nodes). The height assumption can be used to limit the search area on both nodes, and also split access node antenna group in such way that the upper part is used for backhaul link and lower part for access link. Figure 16 depicts the geometry of the deployment, where aggregation node is on the left side.

- Access node assumptions:

  – The shape of an access node is assumed to be cylindrical so that it is "wrapped around" z-axis. This enables beamforming of 360 degrees on horizontal plane and nearly half circle on vertical plane. See figure 14 for illustration.

  – The access node is assumed to be positioned on a lamp pole. This effectively sets the minimum on the inter node distances and the minimum distance to an aggregation node. Also the maximum distances can be specified to make sure the link budget remains valid. Line-of-Sight is assumed towards an aggregation node.

  – The actual coordinates on x-y-plane can be fixed or randomized with the inter node distance and minimum distance to aggregation node as the boundary conditions. The node's height on positive z-axis is assumed to be at least 2m from ground. Figure 16 depicts the geometry of the deployment where access node is on the right side.

  – The initial horizontal boresight direction of the access node can be fixed or randomized. The initial vertial direction is assumed to be fixed at highest possible position from where the scanning will progress downwards.

  – The antenna array size is assumed to be 32×32 elements which can facilitate very narrow beams, e.g. 3 degrees in horizontal plane and 5 degrees in vertical plane [56]. A narrow beam will ensure high enough gain and enables the use of a wide beam on the aggregation node.
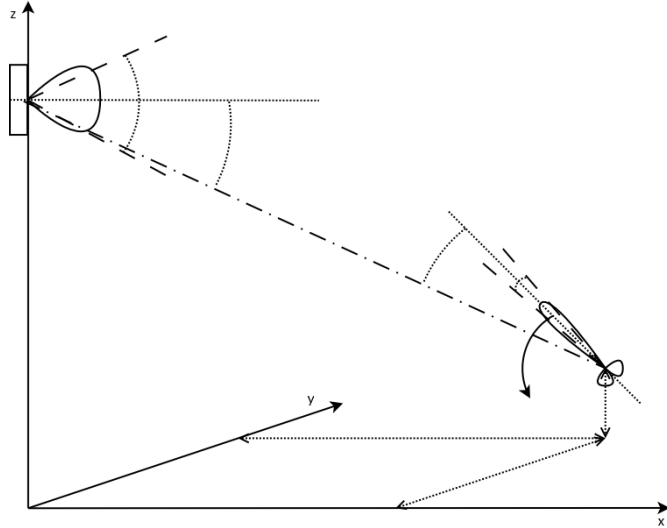
Figure 16: Illustration of the geometry model of aggregation and access nodes in Cartesian coordinates. Aggregation node is transmitting with wide beam on the left side and access node is searching using narrow beam on the right side.

4.3.2 Pilot signal model

The radiated power of a pilot signal is set to be only a fraction of a payload signal stream in order to cause as little interference as possible to the payload streams. In this analysis $0.1 - 0.5\%$ of radiated power of a single payload stream is studied. This means that the pilot signal level stays under the noise floor level from payload signal point of view. Also the payload signal can be considered as a raise in thermal noise floor from the pilot signal point of view (assuming no correlation between the signals). This kind of dimensioning meets the performance requirement of minimizing effect on the link performance even if the pilot signal would always remain active in the background.

Based on the multiplexing method choice of code division multiplexing in chapter 4.2.1, the transmitted pilot signal is assumed to be a known waveform which can be coherently integrated – similar method as used in 3GPP WCDMA system's Common Pilot Channel (CPICH) for initial cell search procedure [57].

In order to find required SNR of the received pilot signal, a probability of false alarm P_{FA} and a probability of detection P_D need to be defined. A detector analyzes the received signal and tries to determine if pilot signal is present within the background thermal noise and possible interference. Common method for this kind of signal detection relies on a set threshold level – if the signal exceeds the threshold, it is declared as a detection. A false alarm is caused by noise and/or interference exceeding the set threshold. The probability of detection P_D can be defined as function of P_{FA} and SNR [58]:

$$P_D = Q\left[\sqrt{2\,SNR}, \sqrt{2\,ln(\frac{1}{P_{FA}})}\right], \tag{15}$$

where Q denotes Marcums Q-function, defined in (16). In this study it has been approximated numerically using a method presented by Parl in [59].

$$Q[\alpha, \beta] = \int_{\beta}^{\infty} \zeta I_0(\alpha\zeta)e^{-(\zeta^2+\alpha^2)/2} \, d\zeta \tag{16}$$

The value for P_{FA} has been selected to be relatively low, (10^{-7}), and for P_D relatively high (0.9) for good quality detection. Choosing the values in this manner naturally results in high required SNR threshold, but it has been compensated by increasing processing gain by coherent integration of the known pilot waveform.

In theory the longer the coherent integration is, the better the resulting SNR for detection is (i.e. the more processing gain is introduced). Below equation shows the SNR improvement of coherent integration of n times longer integration time [60, pp. 41]:

$$\left(\frac{S}{N}\right)_{CI} = n\left(\frac{S}{N}\right) \tag{17}$$

In practice, however, it is not possible to continue integrating infinitely long time. In a practical receiver architecture a local oscillator introduces phase noise in the receiver system. Phase noise is reflected as random rotations of received symbol constellation, which leads to SNR degradation. Practical limit of the integration time is found using the following equation [61, pp. 178]:

$$T_{CI} = \frac{\sqrt{-6ln\beta}}{2\pi\nu_0\sigma_y}, \tag{18}$$

where β denotes fractional reduction in SNR $(0 < \beta < 1)$, ν_0 denotes carrier center frequency and σ_y denotes so called "Allan deviation" of the local oscillator under study.

Allan deviation (ADEV) is square root of Allan variance (AVAR), which is also known as a two-sample variance. IEEE and ITU recommend this to be used as measure of frequency instability of clocks and oscillators. Frequency instability is mostly caused by different types of phase noise. Allan variance is convergent for all the noise types, whereas classical variance is not. Mathematical definition of Allan variance is shown in (19), and Allan deviation in (20).

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2}\left\langle\left(\Delta^2 x\right)^2\right\rangle = \frac{1}{2\tau^2}\left\langle(x_{n+2} - 2x_{n+1} + x_n)^2\right\rangle = \frac{1}{2}\left\langle(\Delta y)^2\right\rangle \tag{19}$$

$$\sigma_y(\tau) = \sqrt{\sigma_y^2(\tau)} \tag{20}$$

In (19) the brackets $\langle \, \rangle$ denote infinite time average. x_n denotes n^{th} error measurement sample of a frequency source and τ denotes time interval between the two samples. $\Delta y = y_{n+1} - y_n$, where $y_n = (x_{n+1} - x_n)/\tau = \Delta x_n/\tau$. [62]

Quartz crystal oscillators are the most common frequency sources in nowadays devices due to cheap price and small form factor. Figure 17 shows Allan deviation of three example crystal oscillators (more details are shown in Appendix A). The Allan deviation values are found to be around 10^{-6} to 10^{-8} for integration time of 10^{-6}
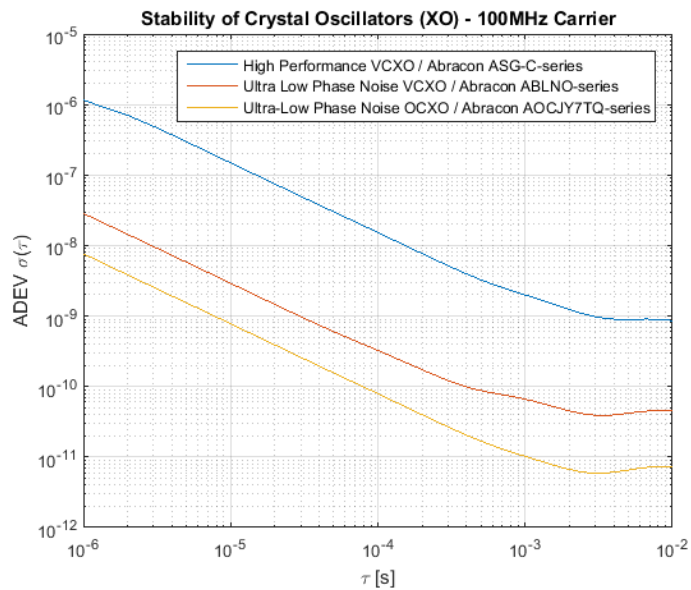
Figure 17: Allan deviation values of three different crystal oscillators. Details of these oscillators are presented in Appendix A.

seconds. Figure 18 shows attainable processing gain as function of integration time with Allan deviation values of 10^{-6}, 10^{-7}, and 10^{-8}. Further analysis on the effect of Allan deviation caused integration time constraints are presented in the results section.
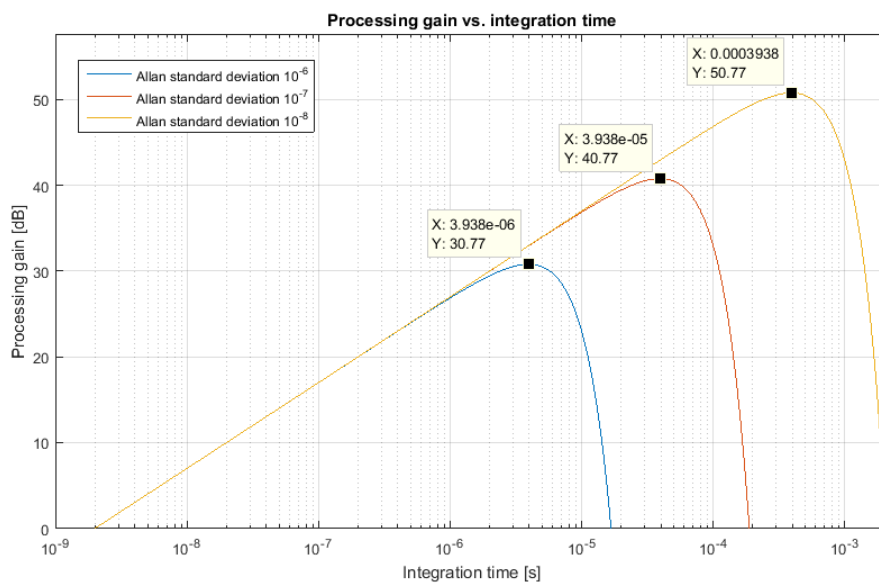

Figure 18: Processing gain as function of integration time for Allan deviation values of 10^{-6}, 10^{-7}, and 10^{-8}. The carrier is at 70 GHz and used bandwidth 500 MHz.

4.3.3 Antenna model

Antenna beam patterns are modeled in 3D using combination of 2-dimensional horizontal and vertical beam patterns. Large antenna arrays are assumed for efficient beamforming and beam steering capabilities as introduced in chapter 2.2.1. For simplicity, the beams are in this analysis generated so that a side lobe floor covers all other directions than the main lobe. A more realistic approach would be to use modified sinc function which would generate also the side lobes and back lobes. Examples of used horizontal and vertical beam patterns are illustrated in figures 19 and 20. The scale in aforementioned figures is between 0 and 1, where 1 means maximum gain at the beam boresight direction and 0.5 indicates half of the maximum antenna level. 3dB (half power) beamwidths can be seen from the figures as the angle between the points where the beam curve intersects with the 0.5-marker. These figures represent a non-tilted scenario, where the beams are pointing along x-axis. In the simulator setup the beam directions will naturally vary.
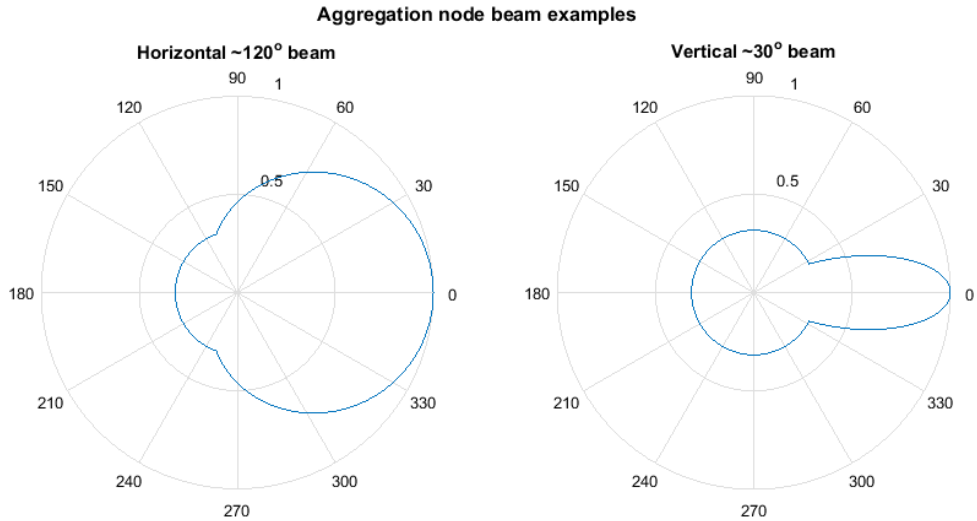

Figure 19: Examples of aggregation node horizontal and vertical beams. Fixed wide beams are used for transmitting pilot signal all around the visible area without beam scanning.

Gains of antenna beam patterns are estimated using following method:

$$G = kD \tag{21}$$

where G denotes the antenna gain, k is efficiency factor with value between 0 and 1. D denotes directivity of the antenna (i.e. gain of an ideally efficient antenna when $k = 1$). Since half-power beamwidths of antenna patterns are set based on simulator assumptions, the directivity D of an antenna can be approximated using following
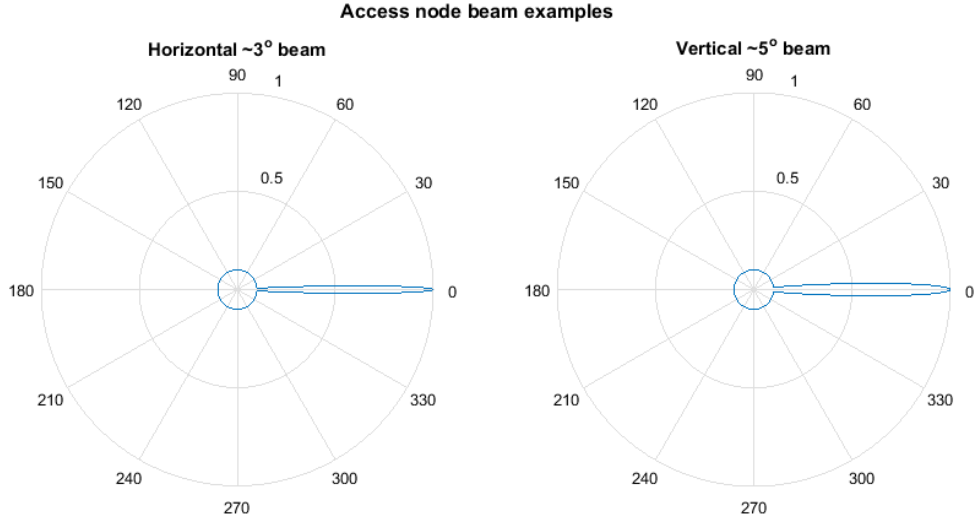
Figure 20: Examples of access node horizontal and vertical beams. Narrow beams are used and the area is scanned to find best direction towards an aggregation node.

equation:

$$D = \frac{40000^{o^2}}{\theta^o_{HP}\phi^o_{HP}} \tag{22}$$

where θ^o_{HP} and ϕ^o_{HP} denote horizontal and vertical half-power beamwidths in degrees, respectively. In well designed antennas the efficiency factor k can be close to 1. An approximation of it was be found from [63] where a millimeter wave "system on a package" design was used for 60 GHz range. The efficiency was found to be greater than 91% (translating to $k > 0.91$). [26, pp. 24]

Downstream pilot signal need to be sent out from an aggregation node with a wide beam (see figure 19) and with constant power to all possible directions, as no prior detail knowledge of access node locations is assumed. The access nodes can reside basically anywhere in the visible area of the aggregation node. A possible method for synthesizing such wide beam in a practical transmitter is to use "Phase Only Excitation Law" presented by Petrolati et al. in [64] utilizing "Code-Fed Array Theory". A more simple way to generate a wide beam is to use random phase weights across antenna elements in an array as proposed by Sayidmarie and Sultan in [65]. The access node will scan the visible area as mentioned earlier in "Geometry model" section.

4.3.4 Link budget model

A link budget analysis is conducted in order to establish a working radio link between aggregation and access nodes. The beam steering for link establishment works in two steps, so single direction link budget is analyzed at a time. The link budget used

in this analysis follows the one presented in chapter 2.2.2. General assumptions in the link budget model are presented in table 2.

Table 2: Example of general assumptions of link budget model for downstream and upstream links

Parameter	Value
Link Frequency [MHz]	70000
Link Bandwidth [MHz]	500
Path loss exponent	2.5
Noise Figure [dB]	6
Rain rate [mm/h]	20
Payload target SNR [dB]	30
Antenna efficiency	0.91
DL pilot power allocation of EIRP	0.001
UL pilot power allocation of payload RX power	0.005
Number of active payload beams	10

Knowing the requirement of SNR of a payload stream, it is possible to calculate the transmitted payload signal power backwards from the SNR and relevant gains and losses using link budget equations (7) and (10). For dimensioning a maximum required (worst case) payload signal level, the path loss L_{PL} in the link budget should be calculated using assumed maximum distance. Table 3 shows a "reversed" link budget calculation according to

$$P_{TX} = SNR - G_{TX} - G_{RX} + L_{TX,sys} + L_{RX,sys} + L_{PL} + L_{ATM} + L_{rain} + L_{other} + N_0 \quad (23)$$

for finding necessary total transmit power of a payload stream. Once the necessary TX power is found, the radiated power (called effective isotropic radiated power or EIRP) is defined by the TX power, TX antenna gain and TX system losses:

$$EIRP = P_{TX} + G_{TX} - L_{TX}. \quad (24)$$

In downstream direction the EIRP pf pilot signal is dimensioned to be roughly 0.1 – 0.5% of the EIRP of a payload signal as stated earlier in "Pilot signal model" section. Equation (24) can be used first for finding $EIRP_{payload}$ by using payload related gain and loss, then $EIRP_{pilot}$ is found by taking 0.1 – 0.5% fraction of $EIRP_{payload}$, and finally $P_{TX,pilot}$ by reversing calculation of (24) with pilot related gain and loss.

To find resulting pilot signal SINR at the receiver, it is required to use equation (7) to first calculate RX power of the pilot signal $P_{RX,pilot}$. What is left to calculate is interference power leaked from active payload streams. In table 2 called "Number of active payload beams" states how many payload streams is assumed to interfere with the pilot signal. Now, assuming a "clever" placement of the nodes in the UDN, we can assume that only side lobes of payload beams might be directed towards an access node under the downstream pilot analysis. It is assumed that maximum

Table 3: Example of reversed link budget calculation to find single payload stream transmit power from SNR constraint and relevant gains and losses

Parameter	Value [dB]
Payload target SNR	30
Gains	
TX-antenna gain	33.98
RX-antenna gain	33.98
Losses	
TX-system losses	2
RX-system losses	2
Path loss	129.45
Rain loss	2.32
Atmospheric loss	0.13
Other losses	3
Thermal noise power [dBm]	-81.01
Required TX-power [dBm]	19.92
Single payload stream EIRP [dBm]	51.90

sidelobe gain is -20dB below the boresight gain of a payload beam. The resulting interference power in downstream can be calculated in dB domain:

$$I_{total,DS} = P_{RX,payload,DS} - L_{sidelobe} + 10\log(No.\ of\ active\ payload\ beams), \quad (25)$$

where $P_{RX,payload}$ is calculated using (7) at the access node location. It is worth noting that this calculation will give the worst case interference scenario where all the payload streams are served with maximum power and a strongest side lobe is pointed towards the interfered node. Downstream SINR at the receiver can be now defined using (12) as all needed parameters $P_{RX,pilot,DS}$, $I_{total,DS}$ and N_0 are known.

In upstream analysis all the active payload beams have their boresights pointed towards the receiver, so the sidelobe loss is not included in the calculation of upstream pilot SINR. Transmit power control is assumed to be used, so the RX power level of a payload stream is in simplified way calculated from target SNR and thermal noise floor

$$P_{RX,payload,US} = SNR + N_0. \quad (26)$$

Again to minimize the interference to the payload streams, the received pilot signal power should be only a fraction of received power of a single payload stream:

$$P_{RX,pilot,US} = P_{RX,payload,US} + 10\log(fraction\ of\ payload\ US\ RX\ power). \quad (27)$$

Total interference to the upstream pilot is calculated using $P_{RX,payload,US}$ multiplied by the number of active payload beams (e.g. 10) and assumed side lobe loss (same as for downstream):

$$I_{total,US} = P_{RX,payload,US} + 10\log(No.\ of\ active\ payload\ beams) - L_{sidelobe} \quad (28)$$

Finally upstream SINR is calculated using (12) with $P_{RX,pilot,US}$, $I_{total,US}$ and N_0.

4.4 Technical analysis results

This section presents the results and findings of the technical analysis. Some calculations have not been shown in utmost detail but the equations and methods are described for proper understanding of the analysis.

The link budget of the discovery process has been dimensioned so that it will have sufficient margin to reach distance of about 250m while withstanding rainfall of 20 mm/h (25 mm/h is considered as heavy rainfall). The required SNR for detection was dimensioned using detection probability of 0.9, which is assumed to be sufficiently high. It is very likely that over several horizontal sweeps with slightly different vertical tilt there will be detection occurring. Assuming this is the case, then the horizontal direction can be determined fairly accurately and further beam fine-refinement can be made even during normal link operation. The refinement can use for example the RSSI measurement method described in Veijalainen's thesis [34].

4.4.1 Downstream link budget analysis results

When access node performs beam sweeping across its visible area, it will search for the strongest signal direction and also perform coherent integration of the received signal in every direction. Figure 21 shows an example of the received power levels during the access node beam sweeping process at assumed maximum distance of $\sim 250m$.
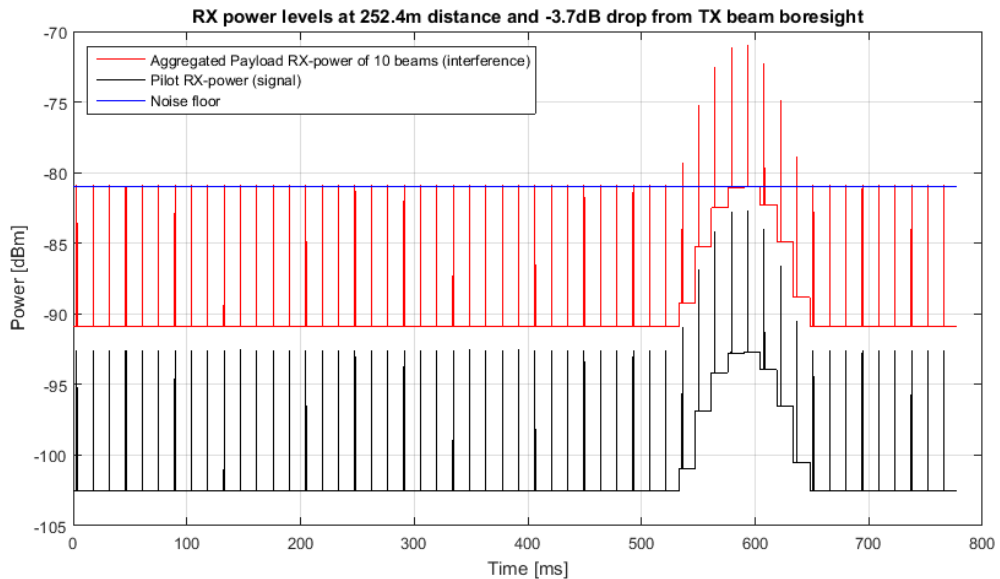

Figure 21: RX power levels of pilot and interference with assumed maximum distance ($\sim 250m$) with 0.1% of payload stream EIRP allocated to pilot TX power. Loss of 3.7dB is caused by the access node residing side from the main lobe of the aggregation node pilot TX beam.

The resulting target SNR after coherent integration and the integration time per beam direction are dependent on the performance of the local oscillator. A range of oscillators with different performance were analyzed. When the oscillator performance is very good, coherent integration can bring more gain to the detection process, and as a result less power needs to be allocated to the pilot signal. When a worse oscillator, e.g. one with Allan deviation of 10^{-6}, is used, processing gain from coherent integration is about 20 dB lower than with one having Allan deviation of 10^{-8}.

To complete the technical feasibility analysis, a link budget analysis of the downstream link is conducted. The analysis is split into following steps:

1. Define required transmit power of single payload stream $P_{TX,payload}$ using (23) and corresponding $EIRP_{payload}$ using (24). The "reversed" link budget has been illustrated in table 3.

2. Find maximum allowed $EIRP_{pilot,DS}$ from $EIRP_{payload}$ and downstream pilot power allocation coefficient. $EIRP_{pilot,DS}$ is also included in table 3. In this analysis maximum allowed EIRP of pilot signal $EIRP_{pilot,DS} = EIRP_{payload} + 10\log(0.001) = 21.90dBm$ which gives maximum transmit power of $P_{TX,pilot,DS} = EIRP_{pilot,DS} - GTX, pilot + L_{sys,TX} = 13.72dBm$.

3. Calculate received power of pilot signal $P_{RX,pilot,DS}$ from $EIRP_{pilot,DS}$ and aggregated received power of interfering payload signals $I_{total,DS}$ at access node location using (25). Calculation of above are presented in table 4.

4. Define SINR of the pilot signal using (12) and above calculated $P_{RX,pilot,DS}$ and $I_{total,DS}$. Due to requirement of summing $I_{total,DS}$ and N_0, the SINR calculation is not possible in decibel domain: $SINR_{pilot,DS} = P_{RX,pilot,DS} - 10\log(10^{I_{total,DS}/10} + 10^{N_0/10}) = -23.72dB$. SINR values at multiple distances are presented in table 5, where slight changes in pilot TX beam alignment are ignored. It can be seen that in the working distance (below 250m) the system is interference limited. The reason is very low pilot signal power compared to multiple interfering payload signals.

Table 6 shows comparison of target SNRs before and after coherent integration with certain oscillator performance values. These values are compared with the results from downstream link budget calculations with the worst case resulting SINRs.

It can be seen that with local oscillators with Allan deviation of 10^{-7} or better the processing gain of coherent integration can compensate the large gap between the target and actual values when only 0.1% of payload EIRP is allocated to the pilot. When oscillator performance is worse than 10^{-7} then more power is required to be allocated. For example with Allan deviation of 10^{-6} about 0.5% of power is required to be allocated. This is still acceptable, but if the power requirement grows much more, the normal system performance will be degraded.

Figure 22 shows an example of link discovery process in downlink direction (from aggregation node to access node) in terms of required amount of beam directions (beam steer steps) and received SNR. Target SNRs are drawn as horizontal lines for

Table 4: Downstream link budget calculation with maximum distance ($\sim 250m$) between access and aggregation nodes

Parameter	Payload signal	Pilot signal
Transmit power [dBm]	19.92	13.72
Gains		
TX-antenna gain [dB]	33.98	10.18
RX-antenna gain [dB]	33.98	33.98
Losses		
TX-beam misalignment [dB]	20.00 (sidelobe)	3.68
RX-beam misalignment [dB]	0.09	0.09
TX-system losses [dB]	2	2
RX-system losses [dB]	2	2
Path loss [dB]	129.41	129.41
Rain loss [dB]	2.31	2.31
Atmospheric loss [dB]	0.13	0.31
Other losses	3	3
Thermal noise power [dBm]	-81.01	-81.01
RX-power [dBm]	-61.04 (10 beams)	-83.72

Table 5: Distance vs. SINR study shows that within the assumed link distances 10–250m the system is interference limited.

Distance [m]	Path loss [dB]	SINR [dB]
10	94.35	-23.68
50	111,83	-23,68
100	119,35	-23,68
150	123,75	-23,69
200	126,88	-23,70
250	129,30	-23,72
1000	144,35	-24,86
2000	151,88	-28,12

different oscillators (with maximum allowed integration times). Red line indicates the threshold without processing gain. Green lines indicate detection threshold SNRs with processing gain from coherent integration over the defined time per beamstep. When SNR raises over the detection threshold, a detection is expected to happen according to the designed probability of detection. It can be seen that detection without processing gain is not possible.

Table 6: Target SNR versus estimated downstream minimum SINR values with 0.1% and 0.5% pilot power allocation levels and a range of local oscillator Allan deviation values

Oscillator performance	Proc. gain	Target SNR	0.1% allocation		0.5% allocation	
			Actual SNR	ΔSNR	Actual SNR	ΔSNR
N/A	0	13.91	-23.72	-37.63	-15.09	-29.00
ADEV 10^{-6}	30.77	-16.86	-23.72	-6.86	-15.09	1.77
ADEV 10^{-7}	40.77	-26.86	-23.72	3.14	-15.09	11.77
ADEV 10^{-8}	50.77	-36.86	-23.72	13.14	-15.09	21.77
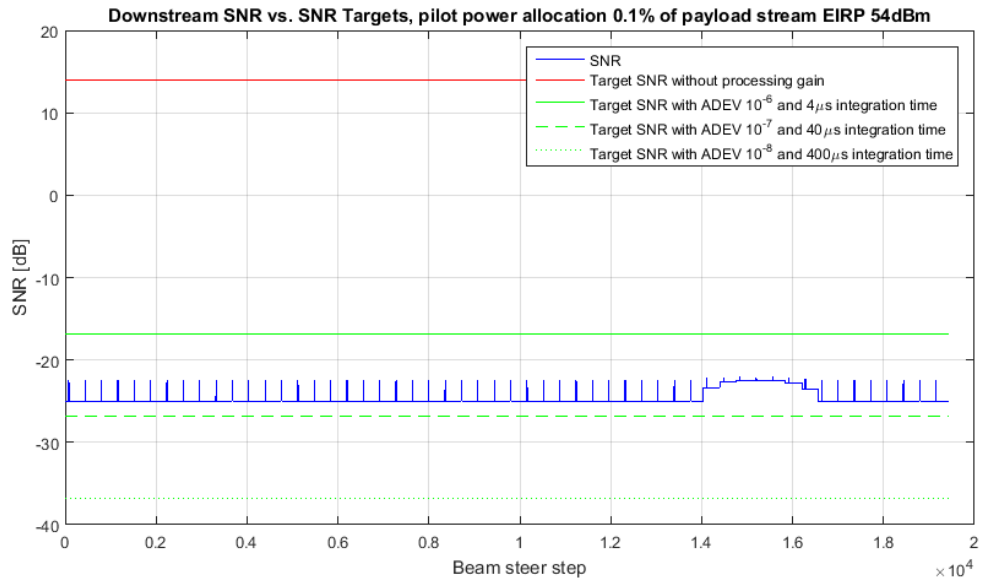
Figure 22: Access node beam search process: received SNR vs. target SNR with and without processing gain of coherent integration. Processing gain is calculated using receiver local oscillator phase noise Allan deviation of 10^{-8} and maximum effective integration time of $\sim 400 \mu s$. 0.1% of total aggregation node TX power is allocated for the pilot signal. Link distace is $\sim$250m.

4.4.2 Upstream link budget analysis results

Uplink pilot signal will be activated by the access node once the beam training process has been completed for downstream pilot. The aggregation node beam training process can be triggered by a management system when access nodes are known to be being added to a UDN. Alternatively if the same methodology is followed as for downlink pilot to maintain the pilot power level as low as possible, the beam search process could possibly run always in the background. Difference with downlink pilot is that now the required transmit power level can be assumed to be known based on

downlink signal measurements. Due to this assumption, the transmit power level can be adjusted to match the individual requirements instead of dimensioning according to "worst case".
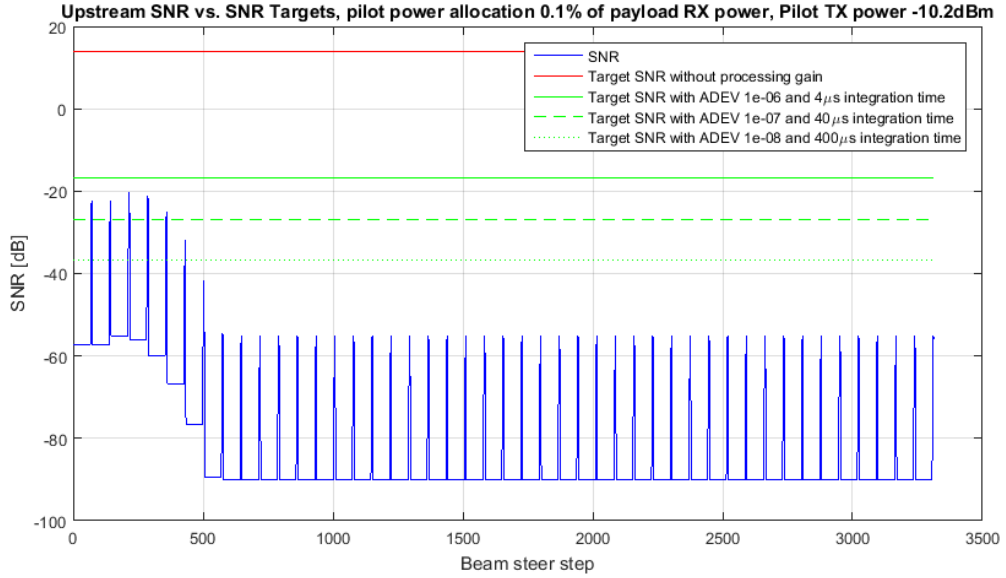

Figure 23: Aggregation node beam search process: received SNR vs. target SNR with and without processing gain of coherent integration. Processing gain is calculated for multiple local oscillator phase noise Allan deviations and respective maximum integration times. Transmit power of the pilot is controlled so that 0.1% of received payload stream power is allocated for the pilot signal. Link distance is ∼250m, but due to power control it has only effect on the transmit power but not on the received SINR.

Link budget analysis of the upstream links are split into following steps:

1. Define minimum required receive power of single payload stream $P_{RX,payload,US}$ using (26): $P_{RX,payload,US} = 30 + (-81.01) = -51.01 dBm$.

2. Find maximum allowed receive power of an upstream pilot signal $P_{RX,pilot,US}$ using (27) by inserting $P_{RX,payload,US}$ and upstream pilot power allocation coefficient: $P_{RX,pilot,US} = -51.01 + 10 \log(0.001) = -81.01 dBm$.

3. Calculate received power of interfering aggregated payload signals $I_{total,US}$ using (28) by inserting $P_{RX,payload,US}$, number of active payload streams and assumed side lobe loss: $I_{total,US} = -51.01 + 10 \log(10) - 20 = -61.01 dBm$.

4. Define received SINR of the pilot signal using (12) and above calculated values: $SINR_{pilot,US} = -81.01 - 10 \log(10^{-61.01/10} + 10^{-81.01/10}) \approx -20.04 dBm$. Note that due to the assumption of uplink pilot power control the received SINR is constant regardless of the distance. Changing parameters of e.g. pilot

power allocation, number of active payload streams or required SNR of payload streams would have effect on the pilot SINR. Figure 23 shows the received SINR of aggregation node beam search process.

From table 7 it can be seen that when upstream pilot power control is utilized and 0.5% power is allocated to the pilot, it is possible to reach sufficient SINR with regular oscillators. 0.1% allocation however requires an oscillator with Allan deviation performance of 10^{-7} or better. Since this oscillator resides in the aggregation node, it can be justified to use slightly more expensive components than in access nodes as the amount of aggregation nodes in a UDN are expected to be around a decade less than the amount of access nodes.

Table 7: Target SNR versus estimated upstream minimum SINR values with 0.1% and 0.5% pilot power allocation levels and a range of local oscillator Allan deviation values

Oscillator performance	Proc. gain	Target SNR	0.1% allocation		0.5% allocation	
			Actual SNR	ΔSNR	Actual SNR	ΔSNR
N/A	0	13.91	-20.04	-33.95	-13.05	-26.96
ADEV 10^{-6}	30.77	-16.86	-20.04	-3.18	-13.05	3.81
ADEV 10^{-7}	40.77	-26.86	-20.04	6.82	-13.05	13.81
ADEV 10^{-8}	50.77	-36.86	-20.04	16.82	-13.05	23.81

4.4.3   Technical analysis conclusions

The quality of an oscillator is reflected in its stability – more stable oscillators are naturally more expensive. A balance between quality and cost need to be found, which will be studied in chapter 5. From the above analysis it can be concluded that in order to maintain reasonably low pilot power occupancy, Allan deviation of $\sim 10^{-6}$ or better should be set as the baseline. This would result in pilot power allocation of $\sim 0.5\%$ or less, which is still acceptable for "always on" pilot not interfering with actual payload signals. Mutual interference of the payload signals was not included in the link budget calculations, but only interference between payload and pilot signals. Total interference on any signals need to be carefully analyzed for more complete analysis.

The time taken by the search process is determined from integration time per beam direction multiplied by number of beam directions tested (beam steer steps). The amount of beam steer steps is dependent on the total scanned angle and the turn angle of each beam steer step. The smaller the turn per step, the more accurate beam direction can be found. Example analysis of the search processes are shown in table 8.

Time required redirecting the beam has not been included in calculation, as it is assumed to be significantly smaller than the integration time per beam step. To optimize the search process, it would be possible to stop horizontal search once

Table 8: Illustration of beam search process total time analysis

		Access node (accurate)	Access node (coarse)	Aggregation node
Horizontal scan area [^{o}]		360	360	180
Horizontal step size [^{o}]		0.5	1.5	1.5
Total horizontal steps		**720**	**240**	**120**
Vertical scan area [^{o}]		90	90	70
Vertical step size [^{o}]		0.5	2.5	2.5
Total vertical steps		**180**	**36**	**28**
Total steps		**129600**	**8640**	**3360**
Total time [s]	$4\mu s$ per step	0.518	0.035	0.013
	$40\mu s$ per step	5.18	0.35	0.13
	$400\mu s$ per step	51.8	3.5	1.3

the likely direction has been found and continue directly with vertical direction refinement. This optimization has not been implemented in the analysis, as according to the performance requirements set in section 4.1, the time taken by the discovery process is not critical. Instead, it is critical to make sure the process will not fail. Moreover, the time required by the process is very short even when longest feasible integration time of $400\mu s$ and small beam steer step size are used (as shown in table 8). Justification for sweeping whole area is to make sure a global maximum of received power will be found and direction of a sidelobe or reflection is not selected. Also in access node beam search process it is possible that there are multiple aggregation nodes in the area and the access node should find as many as possible for redundancy purposes.

It can be initially concluded that the proposed method of link discovery could work using state of the art millimeter wave designs. Further analysis of the feasibility is required when millimeter wave systems become more mature and their practical limitations are well known. In this analysis it was found that one of the most limiting components in the proposed design is the oscillator of the receiver architecture. If phase stability of the oscillator can be kept in sufficiently high level in practical designs, the coherent integration can provide enough processing gain even when the pilot signals are radiated with low power levels. Also inaccuracy of uplink power control was not considered here which will have some impact on the aggregation node search process. Only FDD system was considered for fully "always-on" function. In reality the backhaul link might as well adopt TDD. Then the synchronization must be taken into account more carefully to avoid causing extra interference. Also the link discovery process total duration might be affected.

5   Cost comparisons of technical solutions

In this chapter ultra-dense network (UDN) construction costs are evaluated. Section 5.1 proposes a cost analysis model and section 5.2 applies the model on two scenarios, where the second breaks into two possible sub-scenarios:

1. Heterogeneous network deployment using today's network technologies. This will serve as a base scenario for the cost analysis.

2. UDN deployment scenario

   - without plug and play capability
   - with plug and play capability

The plug and play capability (automatic link discovery function) was introduced in chapter 3.4 and analyzed in detail in chapter 4. The purpose of this analysis is to find a reasonable cost for access node hardware in both UDN scenarios. Also the overall cost of deploying an UDN is analyzed.

5.1   Cost analysis model and assumptions

This section proposes a cost analysis model for an UDN deployment. The analysis methods are explained, and then deployment area assumptions for each scenario are listed.

5.1.1   Cost analysis methods

The method for analyzing the cost of deploying and UDN is considered through a base scenario. The base scenario is assumed to be a heterogeneous network deployment using today's radio technology (e.g. GSM/UMTS/LTE). All scenarios are assumed to be greenfield deployments for easier cost comparison. Also all the fixed investments of the scenarios are assumed to be done at the starting point of the study period, and the following years will only have fixed yearly operational costs. The study period in this thesis is five years.

Once the cost of the base scenario is known, it can be used as the maximum allowed cost of the UDN. This assumption is based on the requirement set in chapter 2.1.2 "General targets and service types of 5G", that the cost of 5G network should not be higher compared with today's networks. Some relaxation can be assumed on the cost with the idea that in future the UDNs are most likely going to be shared by multiple operators. The capacity of the networks is assumed to be much higher than today's netwoks, allowing the sharing between operators. Also it might not make commercially sense that each operator would build such dense networks at same areas leading to low utilization of resources (and low cost efficiency).

Defining types of costs

To evaluate the cost of deploying and running a network, different cost items need to be defined. These are commonly divided to capital expenditure (CAPEX) and

operational expenditure (OPEX). CAPEX commonly means mostly investment related costs. OPEX then contains the operational and other running costs. Examples of both are listed below from network operator's perspective [66]:

1. Examples of what could be counted as CAPEX:

   - Network planning related costs
   - Network equipment related hardware costs
   - Network deployment/site construction related costs

2. Examples of what could be counted as OPEX [67]:

   - Site lease
   - Leased lines (backhaul)
   - Operations and maintenance (O&M)
   - Electricity

In this study only network equipment and network deployment costs are considered from above CAPEX list. Site lease and O&M costs are considered from OPEX. The leased line cost is shifted from OPEX to CAPEX – idea is that the line is owned by the same network operator. The fixed cost is assumed be roughly worth of 3 times of the yearly cost of a leased line (assuming the leasing operator would have 3 years payback time on the investment). Electricity cost evaluation is also excluded here, as based on the target metrics of 5G defined in chapter 2.1.2 it should not be higher than in today's networks.

To compare the investments and operational costs over a certain period of time, it is common to conduct a discounted cash flow analysis. The costs per year are discounted to so called net present value (NPV) using a certain discount rate [66]. For simplicity this analysis is excluded in this study due to relatively short time span of five years and the assumption of fixed yearly OPEX and all CAPEX allocated to the starting point of the study period.

5.1.2 Deployment scenario assumptions

This section lists general assumptions common to all scenarios. Also scenario-specific assumptions are presented.

General assumptions

The cost analysis will be conducted on a $1km^2$ area (depicted in figure 24). In order to find the amount of nodes in the area, following assumptions are made:

- The area is assumed to be a Manhattan grid, a $1km \times 1km$ square.
- Single building footprint is assumed to be a $150m \times 150m$ square.
- Streets between the buildings are assumed to be $30m$ wide

Base scenario assumptions

As a baseline for the cost evaluation, a network deployment using today's radio technology (e.g. GSM/UMTS/LTE) is shown in figure 24. It consists of one macro base station (BS) providing coverage overlay and several micro BSs providing capacity at the streets. One micro BS per street segment is assumed, leading to inter-site distance (ISD) of 180m (along the street). The total amount of street segments in the proposed grid is calculated by first finding out the number of buildings using

$$N_{buildings/km^2} = \left(\frac{L_{area-side}}{W_{building} + W_{street}} \right)^2, \tag{29}$$

which gives $N_{buildings/km^2} = (1000m/(150m + 30m))^2 \approx 30.864$. It is then multiplied by two taking into account horizontal and vertical sides of the building using

$$N_{street-segments/km^2} = 2 * N_{buildings/km^2}, \tag{30}$$

which gives $N_{street-segments/km^2} = 2 * N_{buildings/km^2} \approx 61.728$. The resulting amount of micro base stations for this base scenario is thus $N_{micro-BS/km^2} = N_{street-segments/km^2} \approx 61.728$. The node quantities for the baseline deployment scenario are summarized in table 9.

Table 9: Summary of node quantities in Base deployment scenario

Base scenario	
Node type	**Nodes/km^2**
Macro BS	1
Micro BS	61.728

UDN deployment scenario assumptions

To model the cost of the ultra-dense network (UDN) scenarios, a reference deployment similar to above needs to be decided for the cost evaluation. Following assumptions will be made:

- One macro BS will be serving a $1km^2$ area. The macro BS site will also serve as a network controller site in charge of managing the access and aggregation nodes in the UDN (illustrated in figure 13).

- The number of aggregation nodes will be set to correspond the number of micro BSs in the baseline design ($N_{agg-node} = N_{micro-BS}$).

- Inter-site distance (ISD) of access nodes in an UDN is assumed to be between $30m - 90m$.

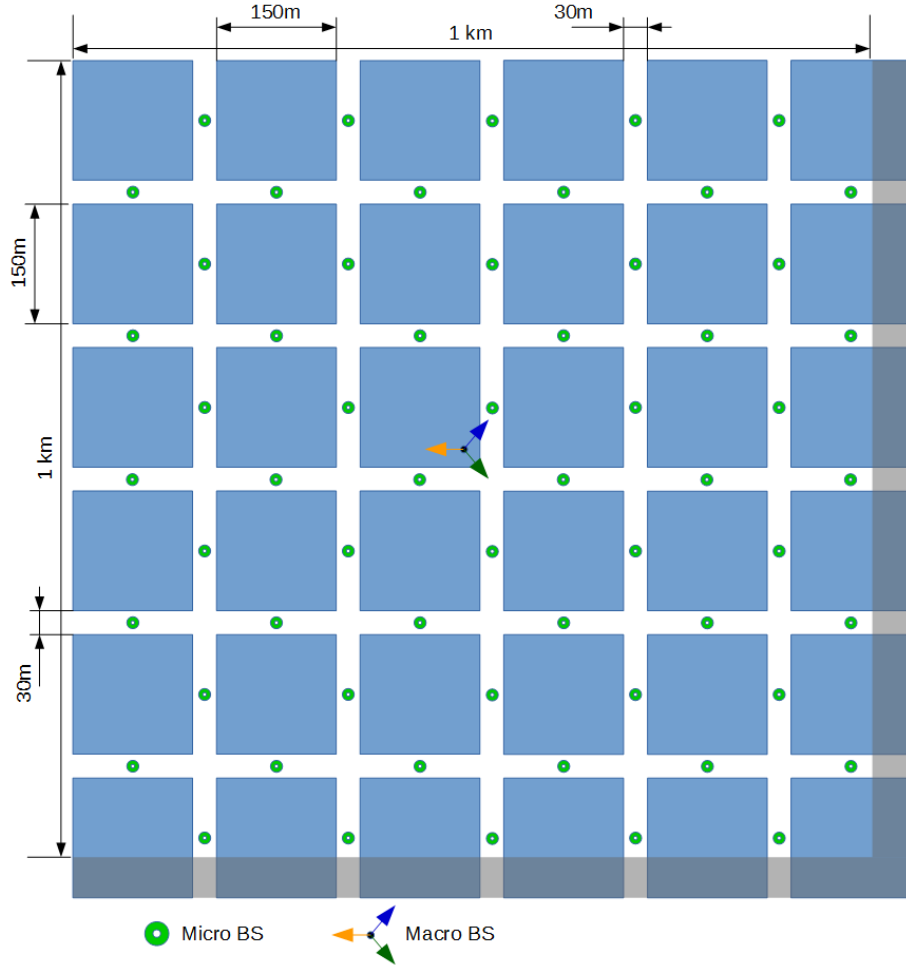- The UDN infrastructure can be shared between multiple service operators.

Figure 24: Network layout consisting of single macro base station and several micro base stations using today's network technologies.

- The costs of macro BS, network controller and aggregation node are assumed to be same for both sub-scenarios – with and without plug and play capability. The sub-scenario related costs are assumed to be reflected only on the CAPEX and OPEX of access nodes.

To model a "worst case" (i.e. most expensive) scenario, the shortest ISD of $30m$ will be selected for the access nodes. This case is depicted in figure 25. An aggregation node is in charge of connecting to 5-6 access nodes (meaning an average of 5.5 access nodes). The total number of access nodes is calculated using (30) with

$$N_{acc-node/km^2} = N_{street-segments/km^2} * N_{acc-node/segment,avg.} \tag{31}$$

which gives $N_{acc-node/km^2} = 61.728 * 5.5 \approx 339.5$. The node quantities for the UDN deployment scenario are summarized in table 10.

Table 10: Summary of node quantities in UDN deployment scenario

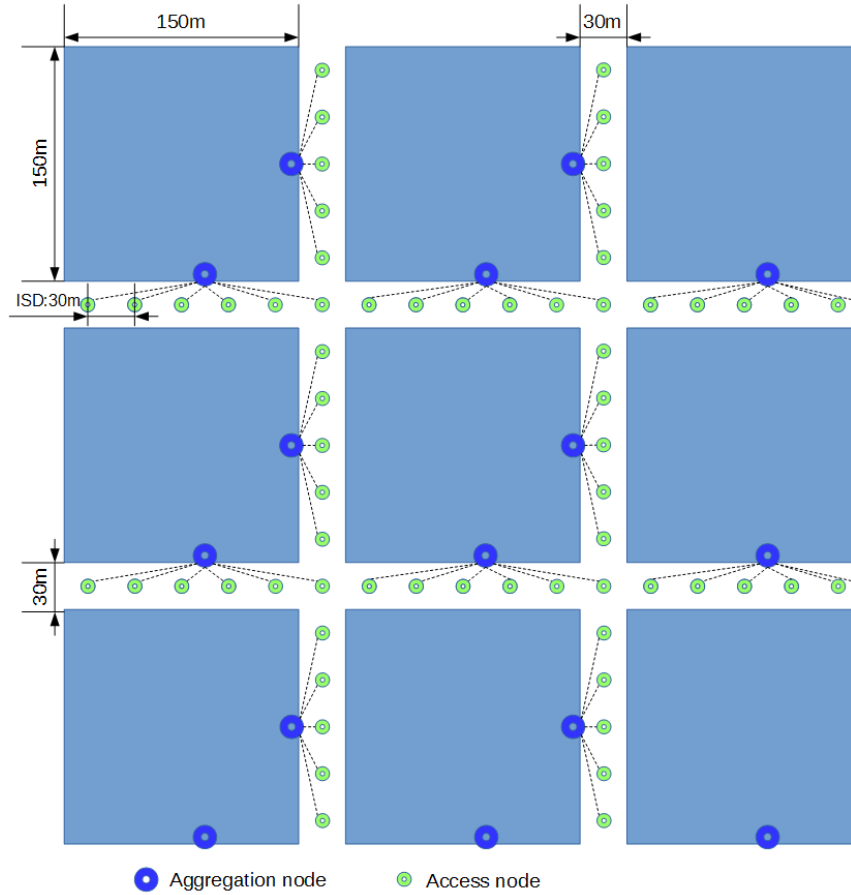| UDN scenario	
Node	**Nodes/km^2**
Macro BS	1
Network controller	1
Aggregation node	61.728
Access node	339.504

Figure 25: A possible network layout with two aggregation nodes per building and 30m inter-site distance between access nodes (along the street).

5.1.3 Unit cost assumptions

In this section unit cost assumptions are proposed for each cost item. The cost assumptions are listed per node type for each scenario.

Table 11 lists the unit costs assumptions of nodes in the base scenario, namely macro and micro BSs. The costs are taken from assumptions in doctoral dissertation

by Yunas [68]. Only exception is that the lease line cost has been moved from OPEX to CAPEX assuming a 3-year payback period (and without conducting the discounted cash flow analysis). So the 2 k€ yearly cost has been translated as 6 k€ one-time investment.

Table 11: CAPEX and OPEX of Base scenario elements from [68]. Only difference is that the backhaul lease has been shifted from OPEX to CAPEX assuming 3-years payback period of the investment.

CAPEX	Macro BS	Micro BS
Equipment (hardware)	10 k€	2.5 k€
Site deployment (installation)	5 k€	0.5 k€
Backhaul link deployment	6 k€	6 k€
Total CAPEX	21 k€	9 k€
OPEX		
Site rent	5 k€/year	1 k€/year
O&M	7 k€/year	1.5 k€/year
Total OPEX	12 k€/year	2.5 k€/year

Table 12 shows assumptions of CAPEX and OPEX of nodes related to the UDN scenario. The macro BS is assumed to have equal costs with the base scenario. Network controller as a new element is assumed to be roughly half of the macro BS CAPEX-wise and slightly smaller yearly O&M costs. As noted earlier, it is co-located with the macro BS, so the backhaul link and site rent are assumed to be shared with the macro BS. Aggregation nodes are assumed to be similar deployment-wise as the micro BS in the base scenario. The yearly OPEX is assumed to be equal, but the millimeter wave based aggregation node capable of running several ultra-high-speed point-to-point links is assumed to have double the equipment cost.

5.2 Comparison of selected scenarios

In this chapter the comparison of scenarios are conducted. Base and UDN scenarios are first compared, and then comparison of UDN sub-scenarios (with and without plug and play capability) is conducted.

5.2.1 Initial comparison of base scenario and UDN

The required node quantities of the base scenario for a square kilometer area were presented in table 9. These are multiplied by the CAPEX and OPEX assumptions for the base scenario (presented in table 11). The OPEX is assumed to run for five years. Total costs of the base scenario are presented in table 13.

Same methodology is used to calculate the total costs of the UDN scenario based on tables 10 and 12. The results are presented in table 14.

Table 12: CAPEX and OPEX of UDN scenario elements (without access nodes). Since the Network controller is assumed to be co-located with macro BS, it is assumed that the backhaul link and site rental costs can be shared.

CAPEX	Macro BS	Network controller	Aggregation node
Equipment (hardware)	10 k€	5 k€	5 k€
Site deployment (installation)	5 k€	2.5 k€	1 k€
Backhaul link deployment	6 k€	0 k€	6 k€
Total CAPEX	21 k€	7.5 k€	12 k€
OPEX			
Site rent	5 k€/year	0 k€/year	1 k€/year
O&M	7 k€/year	5 k€/year	1.5 k€/year
Total OPEX	12 k€/year	5 k€/year	2.5 k€/year

Table 13: Total costs of base scenario.

CAPEX	Macro BS		Micro BS		Total
	Qty	Cost	Qty	Cost	
Equipment	1	10 k€	61.73	2.5 k€	164.33 k€
Site deployment	1	5 k€	61.73	0.5 k€	35.87 k€
Backhaul link	1	6 k€	61.73	6 k€	376.38 k€
Total CAPEX	21 k€		555.57 k€		**576.57 k€**
OPEX					
Site rent	1*5y	5 k€/y	61.73*5y	1 k€/y	333.65 k€
O&M	1*5y	7 k€/y	61.73*5y	1.5 k€/y	497.98 k€
Total OPEX	60 k€		771.63 k€		**831.63 k€**
Total cost	81 k€		1327.20 k€		**1408.2 k€**

From the resulting total costs of base scenario and UDN scenario without access nodes (in tables 13 and 14), it can be seen that the cost of UDN will be higher than that of the base scenario. Since the access nodes are not included in this comparison, it can be expected that the total costs of the UDN will be significantly higher.

In the UDN scenario assumptions of section 5.1.2 it was noted that the network could be shared between multiple operators for more cost optimized solution. This method will be applied in the UDN scenario cost analysis to see if sharing between two or three operators would be sufficient to keep the resulting allowed cost of access nodes at reasonable level. Table 15 shows the resulting cost differences between the scenarios.

The remaining cost difference is divided by the assumed number of access nodes

Table 14: Total costs of UDN scenario based network (without access nodes).

CAPEX	Macro BS		Network controller		Aggregation node		Total
	Qty	Cost	Qty	Cost	Qty	Cost	
Equipment	1	10 k€	1	5 k€	61.73	5 k€	323.65 k€
Site deployment	1	5 k€	1	2.5 k€	61.73	1 k€	69.23 k€
Backhaul link	1	6 k€	1	0 k€	61.73	6 k€	376.38 k€
Total CAPEX	21 k€		7.5 k€		740.76 k€		**769.26 k€**
OPEX							
Site rent	1*5y	5 k€/y	1*5y	0 k€/y	61.73*5y	1 k€/y	333.5 k€
O&M	1*5y	7 k€/y	1*5y	5 k€/y	61.73*5y	1.5 k€/y	522.8 k€
Total OPEX	60 k€		25 k€		771.63 k€		**856.63 k€**
Total cost	81 k€		32.5 k€		1512.39 k€		**1625.9 k€**

Table 15: Cost comparison of base scenario and UDN scenario shared by multiple operators. The assumption here is that no capacity increase is required in the UDN to share the network.

Scenario	No network sharing	Network shared with 2 operators	Network shared with 3 operators
Cost of base scenario	1408.2 k€	2816.4 k€	4224.6 k€
Cost of UDN scenario	1625.9 k€	1625.9 k€	1625.9 k€
Cost difference	**-217.7 k€**	**1190.5 k€**	**2598.7 k€**

in the square kilometer area (in table 10). The results of the division are shown in table 16.

5.2.2 Comparison between UDN scenarios

Two UDN scenarios were presented at the beginning of chapter 5: UDN deployment with and without plug and play capability. In this section the total cost of single access node is divided into the same CAPEX and OPEX components as the other nodes in section 5.1.3. The target is to find out the feasibility of the two UDN scenarios in relation with the network sharing between two or three operators. Same five years study period is assumed here, so first the total OPEX is estimated. Once the split between CAPEX and OPEX is done, they need to be divided into individual components. The split between the CAPEX and OPEX elements will be affected by the presence or absence of the plug and play capability.

Table 16: The remaining cost difference in the network sharing cases are allocated for the access nodes.

	Network shared with 2 operators	Network shared with 3 operators
Total cost of access nodes	1190.5 k€	2598.7 k€
Access nodes (qty/km^2)	339.50	339.50
Total cost of single access node	**3.5 k€**	**7.7 k€**

The plug and play capability will not only allow easy establishment of the link, but it also allows constant tuning of the link to prevent misalignment of the beams. This will help reducing the operation and maintenance costs of running the network. If the function is not included, it can decrease the hardware cost, but in turn will increase the site deployment cost. The study of effects on CAPEX and OPEX are presented next.

OPEX assumptions for access nodes

The operation and maintenance cost of the access nodes consists of remote operations and field maintenance. Remote operations include the network monitoring and adjustment from a network operations center (NOC). Field operations include manual re-alignment of the links and replacement of broken hardware.

As the plug and play capability will help also adjusting the alignment of the millimeter wave beams during runtime, it is likely that no (or very few) field maintenance visits are required for that purpose. Also the NOC burden can be reduced by automatic adjustments of the network.

Evaluation of the operation and maintenance costs of an access node could be conducted using a similar size/type of equipment as a reference. Markendahl and Mäkitalo evaluated in 2010 that the O&M costs of a femtocell would be roughly 500 €/year [67]. For an access node without plug and play capability 25% of the O&M cost of an aggregation node ($0.25 * 1.5ke/year* = 0.375ke/year$) is assumed here. The assumption is considering that the technology has evolved since 2010 to allow lower O&M costs per small radio node. Further O&M cost savings are assumed to be available if the access node would have the plug and play capability. Electronic equipment will always have some failure rate, so the field maintenance cost cannot be totally neglected. The O&M cost of an access node with the plug and play capability is assumed to be only 25% of the O&M cost of the node without it ($0.25 * 0.375ke/year \approx 0.094ke/year$).

The other contributor for the total OPEX in this analysis is site rent. Paying site rent for hundreds of sites per square kilometer area could lead to high running costs of the network. Assuming that almost every lamp post in a certain city area would be covered with access nodes, it would make sense for the operator(s) to find a better agreement with the city that could benefit both. The operator(s) could for example modernize the lamps to LEDs while rolling out the access nodes. It

might be also possible that the access node and LED street lamp hardware would be integrated as single package. Then a possible win-win agreement between operator(s) and the city could be that the operator(s) take the cost of hardware and the city provides electricity. From field maintenance and site rent point of view it could also make sense that the operator(s) and the city would organize joint field maintenance operations which could take care of the maintenance of the network and the lamps at the same time. The effects on the possible cost savings are left for further study. Summary of the access node OPEX assumptions are listed in table 17.

Table 17: OPEX breakdown of an access node. Here "plug and play" has been abbreviated with PnP.

	Without PnP capability	**With PnP capability**
Site rent	0 k€/y	0 k€/y
O&M	0.375 k€/y	0.094k€/y
Total yearly OPEX per access node	0.375 k€/y	0.094 k€/y
Total OPEX over 5 years period	**1.875 k€**	**0.47 k€**

CAPEX assumptions for access nodes

To finalize the access node unit cost evaluation, CAPEX breakdown between equipment cost and site deployment (installation work) cost needs to be decided. There are no references of such dense radio network deployments, so a similar project reference needs to be used to evaluate the cost of a mass-rollout of small nodes. In chapter 2.4.3 practical installation of access nodes were considered with and without plug and play capability. Also a reference from city of Milan was presented where 140 000 street lamps were modernized as LEDs. The total cost of the project was around 38 million euro, which translates to slightly below 300 euro per lamp. [49]

Installation process of a LED street lamp could be rather close to the installation of an access node with plug and play capability. Unfortunately the references didn't reveal detail breakdown of the cost per lamp between hardware and installation work, so a roughly 50/50 division is assumed ($0.5 * 300e/unit = 150e/unit$). The installation work of an access node without plug and play capability is assumed to be more expensive as the person installing needs to have special skills to align the link and set up the node configuration. Also the installation would be much more time consuming. If the cost increase from basic mechanic to a technician would be around 50%, and the time consumption of installation would be doubled, the total cost of installation would be roughly three times higher ($3 * 150e/unit = 450e/unit$).

The remaining part of the total cost is allocated for the hardware. The allowed maximum hardware cost is dependent on the assumption how many operators are sharing the network, and if the node is capable of plug and play or not. Table 18 provides a "reversed" cost calculation to find the allowed hardware cost in each of

the four cases. The calculation is started from the total access node cost after which all assumed costs are deducted.

Table 18: Reversed cost calculation of an access node to find the maximum allowed hardware cost in each scenario.

5 years period	2 operators sharing		3 operators sharing	
	Without PnP	**With PnP**	**Without PnP**	**With PnP**
Total cost	3.5 k€		7.7 k€	
Total OPEX	1.875 k€	0.47 k€	1.875 k€	0.47 k€
Total CAPEX	1.68 k€	3.03 k€	5.83 k€	7.23 k€
Site deployment	0.45 k€	0.15 k€	0.45 k€	0.15 k€
Total HW cost	**1.2 k€**	**2.9 k€**	**5.4 k€**	**7.1 k€**

The resulting maximum allowed cost of access node hardware seems to settle at a reasonable level even when only two operators are sharing the network infrastructure. The allowed cost of the hardware that supports plug and play is over double of the one that doesn't support it. Higher cost of the hardware can be explained with the requirement of millimeter wave antenna array instead of a fixed non-steerable antenna element. Also higher quality oscillator requirement that was found in the technical study in chapter 4.

In reality the cost difference might not be so huge, so using the plug and play capable hardware would potentially provide cost savings compared with the baseline scenario. Especially if the network is run beyond the assumed 5 years period, lower yearly OPEX would bring cost savings over the scenario without plug and play. The allowed access node hardware cost in the "three operator sharing" scenario exceeds the assumed cost of aggregation node hardware, which makes it seem unrealistic.

5.3   Summary and findings of cost analysis

Cost of deploying and running an ultra-dense network was analyzed using today's heterogeneous network deployment and 5G target metrics as the base. Cost of heterogeneous network was calculated as a base scenario, which was used to cap the allowed cost of an UDN deployment. Relaxation on the cost was provided by assuming network sharing between two or three operators. The unitary costs of the HetNet deployment were taken from recent cost analysis of by Yunas in his doctoral dissertation [68], but since there are not many studies conducted on the cost of an UDN deployment, several assumptions had to be made on the unitary costs of the UDN network elements.

As a conclusion, the UDN deployment shared by two operators could already provide feasible cost structure, and sharing between three operators would potentially offer cost savings over the baseline HetNet scenario. Power consumption of the

elements were not considered in any of the scenarios, but according to the target metrics of 5G (presented in chapter 2.1.2), future networks should not consume more energy than today's networks.

Additionally, it was found during the study that further cost-optimizations could be possible by gradual network expansion instead of full density deployment at day one. This requires different method of placing the aggregation nodes in the Manhattan grid. Instead of placing the nodes at the sides of the buildings (as shown in figure 25), it would be possible to cover two street segments using single aggregation node if the node is positioned at a corner of a building (shown in figure 26 (a)). Further capacity expansion would be possible by adding the second aggregation node later to the opposite corner of the building (shown in figure 26 (b)). A challenge in this was found to be the separation between two adjacent beams. If the beams are too close to each other or the used beamwidths are too wide, it might potentially cause excessive interference between the millimeter wave beams. The separation of the beams is dependent on dimensions of the Manhattan grid and the placement of the access nodes on the streets. By increasing the installation height difference between the access and aggregation nodes it is possible to increase the separation between the beams.
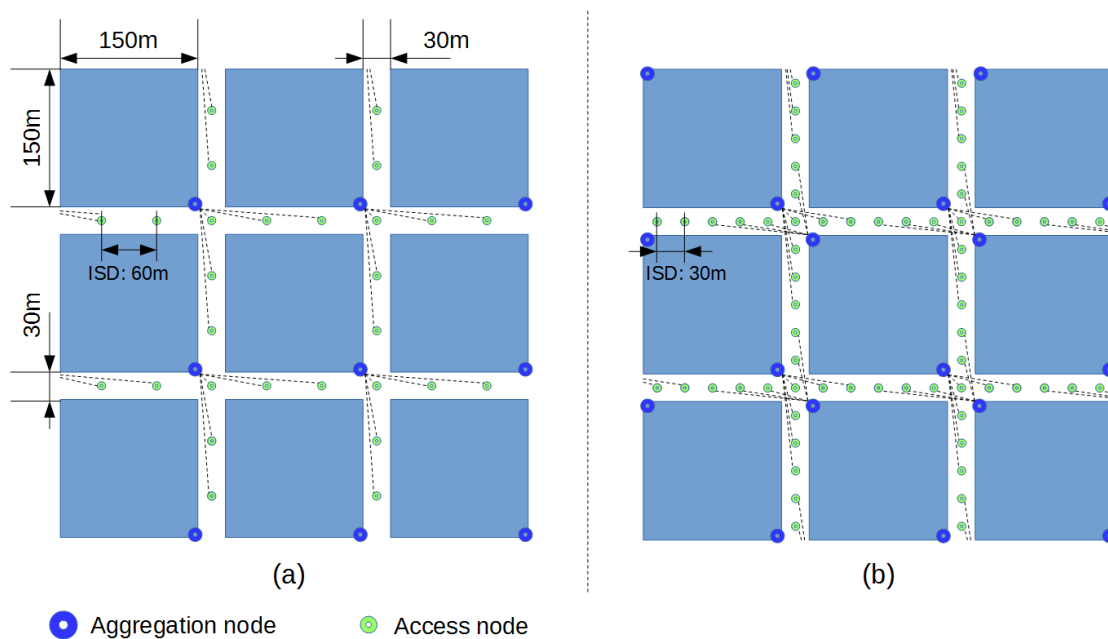
Figure 26: Optimization of node placement: aggregation nodes at corners are able to cover access nodes placed on two street segments (a). Further capacity addition can be achieved by adding new aggregation nodes to opposite corners (b)

6 Conclusions

The amount of wireless traffic and number of connected devices are expected to grow faster than ever before in the near future. By the year 2020 the amount of data traffic is forecasted to grow 1000 times from 2010 levels, and the amount of connected devices is expected to reach 50 billion. Diversified use cases will arise from the increasing coexistence of human type and machine type communications. A shift from fixed to wireless networks in last mile access is constantly ongoing in all areas of communications. Biggest drivers are the free mobility of wireless technology and the constant evolution of better performing wireless networks. The emergence of 5G wireless network technology is envisioned to address the challenges of the future wireless communications.

Ever increasing demand for higher data rates and low latencies in wireless networks are not able to be met by traditional macro base stations alone. Network densification is inevitable in traffic hot spots such as crowded cities. Network densification by means of deploying smaller and smaller cells introduces a new concept called ultra-dense network (UDN) Inter-site distances in an ultra-dense network could be as short as distance between two lamp posts. Deploying such dense networks introduces new challenges for network operators on the deployment costs.

Efficient deployment of an UDN requires that the devices support so called "plug and play" installation. This thesis introduced functional requirements of plug and play installation of small access nodes. Detail technical study focused on the link discovery process between access and aggregation nodes, followed by cost comparison of UDN deployment with and without the proposed plug and play capability. Following chapters will introduce conclusions from both technical and cost analyses. Finally some discussion and further study proposals will be presented.

6.1 Conclusions of technical analysis

A link discovery method was proposed which uses very small proportion of total power for downlink pilot, thus allowing the pilot to remain always active in the background. The proposed process utilizes similar method as used in WCDMA cell-search process: a known waveform is transmitted which the receivers are trying to detect. The waveform detection is combined with narrow beam search in order to align the beams correctly. It was found that the proposed process works in system-level study – assuming state-of-the-art millimeter wave designs are applied. One of the critical components was found to be local crystal oscillator. The stability of the oscillator needs to be sufficiently high in order to maintain SNR high enough during the waveform integration. It was also noted that if the oscillator stability cannot be ensured to be at sufficient level, it is possible to compensate this deficiency by allocating higher power for the pilot signal. This is, however, not desired, as raising the pilot power might interfere with existing payload traffic, and maintaining the pilot always active would also consume more power.

Examples of three suitable oscillators are introduced in appendix A. Using the cheapest oscillator of the three may result in having to increase the pilot power from

0.1% to 0.5% of payload EIRP. The second oscillator provides sufficient stability to maintain the EIRP proportion at 0.1%. The third was chosen as reference if utmost stability is required, but the cost will also jump radically.

The study considered only interference between payload and pilot signals, but didn't take into account possible changes in radiated powers when for example inter-payload stream interference is present. Also inter-pilot interference was not covered. The probabilities of inter-payload stream and inter-pilot interferences occurring (or at least causing harm) should be rather low, as in every case at least one side, receiver or transmitter, is utilizing narrow beams. High power omnidirectional transmissions are not present in the proposed system.

6.2   Conclusions of cost analysis

Cost evaluation method and analysis were proposed for ultra-dense networks. The main driver of the study was to find if the proposed plug and play functionality of the technical analysis would be beneficial from cost perspective. A base scenario was developed assuming a heterogeneous network deployment utilizing today's mobile network technologies. The total cost of the HetNet was used as maximum allowed cost of the UDN. It was found that maximum density UDN (access nodes 30m away from each other) is not cost effective to be built by single operator alone. By sharing the network infrastructure between at least two operators it was found to be feasible. Sharing between three or more operators could potentially bring cost savings over the HetNet scenario. The plug and play functionality proved to be useful – it offered possibility of doubling the hardware cost, which can be allocated for additional functionality (or considered as cost savings). Also the longer the network is running, the more beneficial the plug and play functionality will become due to smaller operational costs.

Power consumption of the elements were not considered in any scenario, but according to the target metrics of 5G (presented in chapter 2.1.2), future networks should not consume more energy than today's networks. Based on this assumption, including the power consumption of the UDN should not severely affect the conducted analysis.

One aspect that was also excluded was the discounted cash flow analysis. It would be fairly straightforward to perform, but it was considered to be a bit off the main topic of this thesis so it was left out. Also sensitivity analysis could be performed using different discount rates.

As a conclusion, the UDN deployment shared by two operators could provide feasible cost structure, and sharing between three operators would potentially offer cost savings over the baseline HetNet scenario. Plug and play functionality offers cost savings in labor by moving the cost partly to more intelligent hardware. OPEX savings are guaranteed with almost zero-maintenance hardware.

6.3   Discussion and further study

The proposed link discovery process was shown to work based on link budget analysis. A more thorough analysis should be performed to verify the results. Multipath caused interference should be taken into account in the analysis by for example utilizing novel ray tracing methods.

Cost analyses of 5G networks have so far not been conducted extensively. When the standardization of the technologies progresses, cost effectiveness will be one of the key factors of successful deployments. The technological advancements will provide more insight on the possible costs of the components, and the future functional requirements will enable further cost studies on the installation and operational costs. In the end, everything is about money in network business. The analysis in this thesis has just scratched the surface.

Other serious aspects for further study are the health effects of millimeter wave communications. Assuming the access nodes would be placed to several lamp posts along a street, it is impossible to avoid millimeter wave radiation from aggregation nodes to hit nearby pedestrians. Initial studies of the health effects have been conducted e.g. by Samsung [69], but more in-depth studies are highly recommended before the millimeter wave systems hit the mainstream.

References

[1] A. Osseiran, V. Braun, T. Hidekazu, P. Marsch, H. Schotten, H. Tullberg, M. Uusitalo, and M. Schellman, "The Foundation of the Mobile and Wireless Communications System for 2020 and Beyond: Challenges, Enablers and Technology Solutions," in *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pp. 1–5, June 2013.

[2] NGMN Alliance, "5g White Paper," Feb. 2015.

[3] ITU-R, "ITU-R Study Groups," May 2013.

[4] M. R. Bhalla and A. V. Bhalla, "Generations of Mobile Wireless Technology: A Survey," *International Journal of Computer Applications*, vol. 5, pp. 26–32, Aug. 2010.

[5] L. Harte, *Introduction to Mobile Telephone Systems, 2nd Edition, 1G, 2G, 2.5G, and 3G Technologies and Services*. Fuquay-Varina, NC: Althos Publishing, 2 edition ed., Oct. 2006.

[6] NMT-Group, "Nordic Mobile Telephone - System Description (NMT DOC 450-1)," Aug. 1997.

[7] Y. Sung and Y.-H. Yang, "Challenges from voice-over-LTE to video-over-LTE," in *Network Operations and Management Symposium (APNOMS), 2014 16th Asia-Pacific*, pp. 1–4, Sept. 2014.

[8] Ericsson, "More than 50 billion connected devices," white paper, Feb. 2011.

[9] H. Tullberg, H. Droste, M. Fallgren, P. Fertl, D. Gozalvez-Serrano, E. Mohyeldin, O. Queseth, and Y. Seien, "METIS research and standardization: A path towards a 5g system," in *Globecom Workshops (GC Wkshps), 2014*, pp. 577–582, Dec. 2014.

[10] ITU-R, "Future technology trends of terrestrial IMT systems," Report ITU-R M.2320-0, International Telecommunication Union, Geneva, Switzerland, 2014.

[11] METIS, "Final report on the METIS 5g system concept and technology roadmap," Deliverable D6.6, Apr. 2015.

[12] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5g," *IEEE Communications Magazine*, vol. 52, pp. 82–89, Feb. 2014.

[13] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter Wave Mobile Communications for 5g Cellular: It Will Work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[14] J. Luo, J. Eichinger, Z. Zhao, and E. Schulz, "Multi-carrier waveform based flexible inter-operator spectrum sharing for 5g systems," in *2014 IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN)*, pp. 449–457, Apr. 2014.

[15] R. Baldemair, T. Irnich, K. Balachandran, E. Dahlman, G. Mildh, Y. Selén, S. Parkvall, M. Meyer, and A. Osseiran, "Ultra-dense networks in millimeter-wave frequencies," *IEEE Communications Magazine*, vol. 53, pp. 202–208, Jan. 2015.

[16] J. Karjalainen, M. Nekovee, H. Benn, W. Kim, J. Park, and H. Sungsoo, "Challenges and opportunities of mm-wave communication in 5g networks," in *2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, pp. 372–376, June 2014.

[17] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. Thomas, J. Andrews, P. Xia, H. Jo, H. Dhillon, and T. Novlan, "Heterogeneous cellular networks: From theory to practice," *IEEE Communications Magazine*, vol. 50, pp. 54–64, June 2012.

[18] H.-S. Park, A.-S. Park, J.-Y. Lee, and B.-C. Kim, "Two-Step Handover for LTE HetNet mobility enhancements," in *2013 International Conference on ICT Convergence (ICTC)*, pp. 763–766, Oct. 2013.

[19] Huawei Technologies Co., Ltd., "5g: New Air Interface and Radio Access Virtualization," white paper, Apr. 2015.

[20] N. Pleros, K. Tsagkaris, and N. Tselikas, "A moving extended cell concept for seamless communication in 60 GHz radio-over-fiber networks," *IEEE Communications Letters*, vol. 12, pp. 852–854, Nov. 2008.

[21] N. Pleros, K. Vyrsokinos, K. Tsagkaris, and N. Tselikas, "A 60 GHz Radio-Over-Fiber Network Architecture for Seamless Communication With High Mobility," *Journal of Lightwave Technology*, vol. 27, pp. 1957–1967, June 2009.

[22] P. Kela, J. Turkka, and M. Costa, "Borderless Mobility in 5g Outdoor Ultra-Dense Networks," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2015.

[23] ITU-R, "Radio Regulations, Edition of 2012," 2012.

[24] T. S. Rappaport, R. W. H. Jr, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Prentice Hall, Sept. 2014.

[25] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, N.J: Prentice Hall, 2 edition ed., Jan. 2002.

[26] J. D. Kraus and R. J. Marhefka, *Antennas For All Applications*. New York: McGraw-Hill Science/Engineering/Math, 3 edition ed., Nov. 2001.

[27] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, and A. Ghosh, "Multilevel millimeter wave beamforming for wireless backhaul," in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, pp. 253–257, Dec. 2011.

[28] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, pp. 186–195, Feb. 2014.

[29] M. Akdeniz, Y. Liu, M. Samimi, S. Sun, S. Rangan, T. Rappaport, and E. Erkip, "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, pp. 1164–1179, June 2014.

[30] M. Marcus and B. Pattan, "Millimeter wave propagation; spectrum management implications," *IEEE Microwave Magazine*, vol. 6, pp. 54–62, June 2005.

[31] P. Smulders and L. Correia, "Characterisation of propagation in 60 GHz radio channels," *Electronics Communication Engineering Journal*, vol. 9, pp. 73–80, Apr. 1997.

[32] F. Giannetti, M. Luise, and R. Reggiannini, "Mobile and Personal Communications in the 60 GHz Band: A Survey," *Wireless Personal Communications*, vol. 10, pp. 207–243, July 1999.

[33] Small Cell Forum, "Backhaul Technologies for Small Cells: Use Cases, Requirements and Solutions," Feb. 2013.

[34] T. Veijalainen, *Beam steering in millimeter wave radio links for small cell mobile backhaul.* Master's Thesis, Aalto University, June 2014.

[35] H. Yang, P. Smulders, and M. Herben, "Channel Characteristics and Transmission Performance for Various Channel Configurations at 60 GHz," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, p. 019613, May 2007.

[36] T. Nitsche, A. B. Flores, E. W. Knightly, and J. Widmer, "Steering with Eyes Closed: mm-Wave Beam Steering without In-Band Measurement," in *The 34th IEEE International Conference on Computer Communications (IEEE INFOCOM 2015)*, Jan. 2015.

[37] R. Daniels and R. Heath, "60 GHz wireless communications: emerging requirements and design recommendations," *IEEE Vehicular Technology Magazine*, vol. 2, pp. 41–50, Sept. 2007.

[38] E. P. Serna, S. Thombre, M. Valkama, S. Lohan, V. Syrjälä, M. Detratti, H. Hurskainen, and J. Nurmi, "Local Oscillator Phase Noise Effects on GNSS Code Tracking," *Inside GNSS*, vol. 5, pp. 52–62, Dec. 2010.

[39] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, pp. 101–107, June 2011.

[40] A. Abidi, "CMOS microwave and millimeter-wave ICs: The historical background," in *2014 IEEE International Symposium on Radio-Frequency Integration Technology (RFIT)*, pp. 1–5, Aug. 2014.

[41] T. Rappaport, J. Murdock, and F. Gutierrez, "State of the Art in 60-GHz Integrated Circuits and Systems for Wireless Communications," *Proceedings of the IEEE*, vol. 99, pp. 1390–1436, Aug. 2011.

[42] J. Bartelt and G. Fettweis, "Radio-over-Radio: I/Q-stream backhauling for cloud-based networks via millimeter wave links," in *2013 IEEE Globecom Workshops (GC Wkshps)*, pp. 772–777, Dec. 2013.

[43] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): a primer," *IEEE Network*, vol. 29, pp. 35–41, Jan. 2015.

[44] R. Wang, H. Hu, and X. Yang, "Potentials and Challenges of C-RAN Supporting Multi-RATs Toward 5g Mobile Networks," *IEEE Access*, vol. 2, pp. 1187–1195, 2014.

[45] Ericsson AB, Huawei Technologies Co., Ltd., NEC Corporation, Nokia Networks, and Alcatel-Lucent, "Common Public Radio Interface (CPRI); Interface Specification," Specification V6.1, July 2014.

[46] E. Alwan, W. Khalil, and J. Volakis, "Ultra-wideband on-site coding receiver (OSCR) for digital beamforming applications," in *2013 IEEE Antennas and Propagation Society International Symposium (APSURSI)*, pp. 620–621, July 2013.

[47] Ericsson and Philips, "Zero Site," Feb. 2014.

[48] Lumileds Holding B.V., "In Milan, Sustainable Lighting is for Life," May 2015.

[49] European Commission, "Milan turns to LED," 2014.

[50] E. Alwan, S. Venkatakrishnan, A. Akhiyat, W. Khalil, and J. Volakis, "Code Optimization for a Code-Modulated RF Front End," *IEEE Access*, vol. 3, pp. 260–273, 2015.

[51] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, "A First Look at Traffic on Smartphones," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC '10, (New York, NY, USA), pp. 281–287, ACM, 2010.

[52] GSMA Intelligence, "Understanding SIM evolution," tech. rep., Mar. 2015.

[53] T. Pearson and K. Shenoi, "A case for assisted partial timing support using precision timing protocol packet synchronization for LTE-A," *IEEE Communications Magazine*, vol. 52, pp. 136–143, Aug. 2014.

[54] A. Dammann, R. Raulefs, and S. Zhang, "On prospects of positioning in 5g," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, pp. 1207–1213, June 2015.

[55] H. Wymeersch, J. Lien, and M. Win, "Cooperative Localization in Wireless Networks," *Proceedings of the IEEE*, vol. 97, pp. 427–450, Feb. 2009.

[56] T. Cella, P. Orten, and T. Ekman, "Design of a Practical and Compact mm-Wave MIMO System with Optimized Capacity and Phased Arrays," *International Journal of Antennas and Propagation*, vol. 2014, June 2014.

[57] 3GPP, "Physical channels and mapping of transport channels onto physical channels (FDD)," Specification TS 25.211, Rel-12, v12.1.0, Dec. 2014.

[58] B. R. Mahafza, *Radar Systems Analysis and Design Using MATLAB*. CRC Press, Jan. 2002.

[59] S. Parl, "A new method of calculating the generalized Q function (Corresp.)," *IEEE Transactions on Information Theory*, vol. 26, pp. 121–124, Jan. 1980.

[60] G. R. Curry, *Radar Essentials: A Concise Handbook for Radar Design and Performance Analysis*. Raleigh, NC: Scitech Pub Inc 2011-05-25, 2011.

[61] J. A. Zensus, P. J. Diamond, and P. J. Napier, *Very long baseline interferometry and the VLBA: proceedings of a summer school held in Socorro, New Mexico, June 23-30, 1993*. Astronomical Society of the Pacific, 1995.

[62] Hewlett Packard, "The Science of Timekeeping," Application Note 1289, June 1997.

[63] T. Seki, N. Honma, K. Nishikawa, and K. Tsunekawa, "Millimeter-wave high-efficiency multilayer parasitic microstrip antenna array on teflon substrate," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 2101–2106, June 2005.

[64] D. Petrolati, P. Angeletti, and G. Toso, "A novel phase only excitation law for omnidirectional arrays," in *2010 Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, Apr. 2010.

[65] K. Sayidmarie and Q. Sultan, "Synthesis of wide beam array patterns using random phase weights," in *2013 International Conference on Electrical, Communication, Computer, Power, and Control Engineering (ICECCPCE)*, pp. 52–57, Dec. 2013.

[66] T. Smura, *Techno-Economic Modelling of Wireless Network and Industry Architectures*. No. Doctoral dissertations 23/2012 in Aalto University publication series, Jan. 2012.

[67] J. Markendahl and O. Mäkitalo, "A comparative study of deployment options, capacity and cost structure for macrocellular and femtocell networks," in *2010 IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communications Workshops (PIMRC Workshops)*, pp. 145–150, Sept. 2010.

[68] S. F. Yunas, *Capacity, Energy-Efficiency and Cost-Efficiency Aspects of Future Mobile Network Deployment Solutions.* Oct. 2015.

[69] Samsung Electronics Co., Ltd., "5g Vision," white paper, June 2015.

[70] Rohde & Schwarz, "Time Domain Oscillator Stability Measurement Allan Variance," application Note, Feb. 2009.

A Reference crystal oscillators

This appendix includes data of three crystal oscillators selected for reference for the cost analysis in chapter 5. Figure A1 presents basic specifications and pricing of the three oscillators. The data has been obtained from *www.digikey.fi* electronic component webstore.

The manufacturer of these oscillators (Abracon LLC) provides only phase noise data in terms of *dBc/Hz* at certain offset frequencies. In order to evaluate the suitability of the oscillators for the study in thesis, the frequency domain phase noise data needs to be converted to Allan deviation (ADEV). The conversion method has been presented in [70]. Spectral density function $S_y(f)$ can be calculated from single side-band phase noise traces $L(f)$ (presented in figures A3, A6 and A9) using (A1)

$$S_y(f) = 2\frac{f^2}{\nu_0^2}L(f) \tag{A1}$$

where f denotes offset frequency and ν_0 denotes oscillator carrier frequency. Allan variance can be calculated from $S_y(f)$ using (A2):

$$\sigma_y^2(\tau) = 2\int_0^{f_h} S_y(f)\frac{\sin^4(\pi\tau f)}{(\pi\tau f)^2}df \tag{A2}$$

where f_h denotes a high frequency cut-off of a low-pass filter.

Allan variance is converted to Allan deviation by calculating square root of the variance, as shown in equation (20). Allan deviation data presented in figures A4, A7 and A10 have been calculated numerically using Wolfram Mathematica software.

Image			
Digi-Key Part Number	535-12963-ND	ASG-C-V-A-100.000MHZ-T-ND	ABLNO-V-100.000MHZ-T-ND
Manufacturer Part Number	AOCJY7TQ-X-100.000MHZ-1	ASG-C-V-A-100.000MHZ-T	ABLNO-V-100.000MHZ-T
Manufacturer	Abracon LLC	Abracon LLC	Abracon LLC
Description	OSC OCXO 100.000MHZ SNWV PC PIN	OSC VCXO 100.000MHZ LVCMOS SMD	OSC VCXO 100.000MHZ LVCMOS SMD
Quantity Available ?	0	0	0
Unit Price ? **EUR**	802.46000	4.69780	7.47120
Minimum Quantity ?	1	2,000 Non-Stock ?	1,000 Non-Stock ?
Series	AOCJY7TQ	ASG-C	ABLNO
Packaging	Tray ?	Tape & Reel (TR) ? Alternate Packaging	Tape & Reel (TR) ? Alternate Packaging
Type	OCXO	VCXO	VCXO
Frequency	100MHz	100MHz	100MHz
Function	-	Tri-State (Output Enable)	-
Output	Sine Wave	LVCMOS	LVCMOS
Voltage - Supply	12V	3.3V	3.3V
Frequency Stability	±100ppb	±35ppm	±18ppm
Operating Temperature	-40°C ~ 70°C	-40°C ~ 85°C	-40°C ~ 85°C
Current - Supply (Max)	-	45mA	35mA
Ratings	-	-	-
Mounting Type	Through Hole	Surface Mount	Surface Mount
Size / Dimension	1.004" L x 0.984" W (25.50mm x 25.00mm)	0.276" L x 0.197" W (7.00mm x 5.00mm)	0.563" L x 0.343" W (14.30mm x 8.70mm)
Height	0.500" (12.70mm)	0.075" (1.90mm)	0.256" (6.50mm)
Package / Case	5-SIP	6-SMD	4-SMD, No Lead (DFN, LCC)

Figure A1: Comparison of three oscillators (data from *www.digikey.fi* electronic component webstore).

Abracon ASG-C Voltage Controlled Crystal Oscillator

Programmable - High Performance SMD Crystal Oscillator

ASG-C Series

Moisture Sensitivity Level (MSL) - This product is Hermetically Sealed and not
Moisture Sensitive; therefore MSL = N/A (Not Applicable)

Pb | RoHS Compliant

ASG

7.0 x 5.0 x 2.0mm

▷ **FEATURES:**

- ASG series is a High Performance crystal based oscillator; available either as an XO or a VCXO
- Frequency range from 10MHz to 250MHz with LVCMOS output
- Available from 10MHz to 1.50GHz with LVDS or LVPECL output
- Offered with either 2.50V or 3.30V bias voltage
- Quick turn, 1~5 business days for small quantity orders

▷ **APPLICATIONS:**

- Networking, SONET/SDH
- WiMax / WLAN
- Computing
- Phase Locked Loops
- Direct Digital Synthesis (DDS)
- DSL/ADSL
- Base Terminal Stations

Figure A2: Features and applications of Abracon ASG-C crystal oscillators

Offset (Hz)	Amplitude (dBc/Hz)
10	-63.65
100	-89.98
1,000	-116.24
10,000	-122.58
100,000	-123.48
1,000,000	-141.55
5,000,000	-147.50
20,000,000	-153.59

Phase Noise Plot: 100MHz carrier, 12kHz to 20MHz BW
Calculated rms Phase Jitter = 760.080 fs
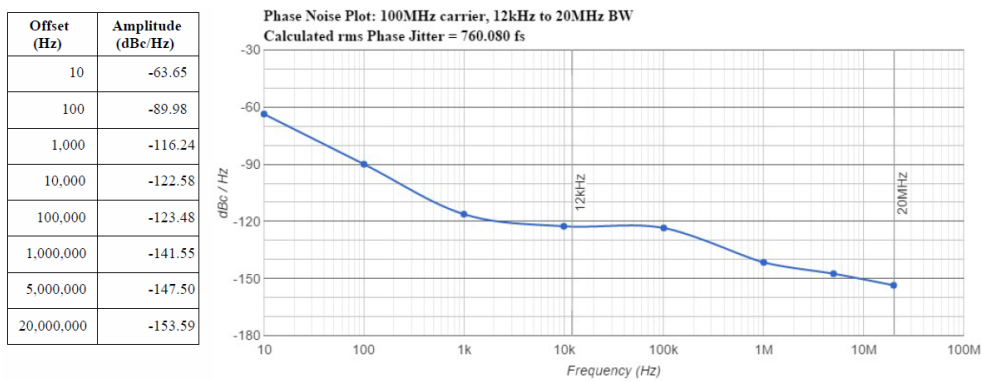
Figure A3: Phase noise data of Abracon ASG-C crystal oscillators at 100 MHz
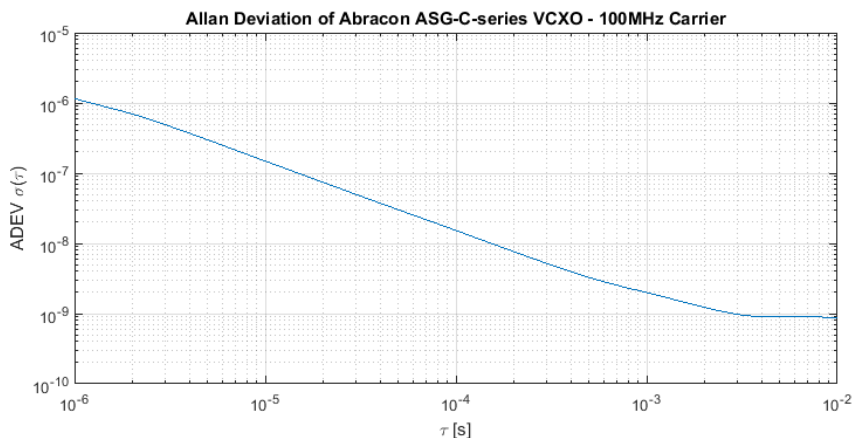
Figure A4: Allan deviation of Abracon ASG-C crystal oscillators at 100 MHz
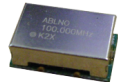
Abracon ABLNO Voltage Controlled Crystal Oscillator

Figure A5: Features and applications of Abracon ABLNO crystal oscillators
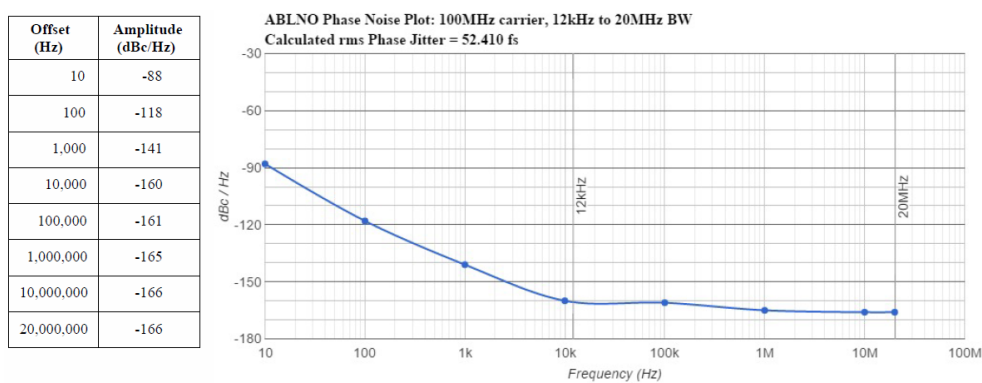

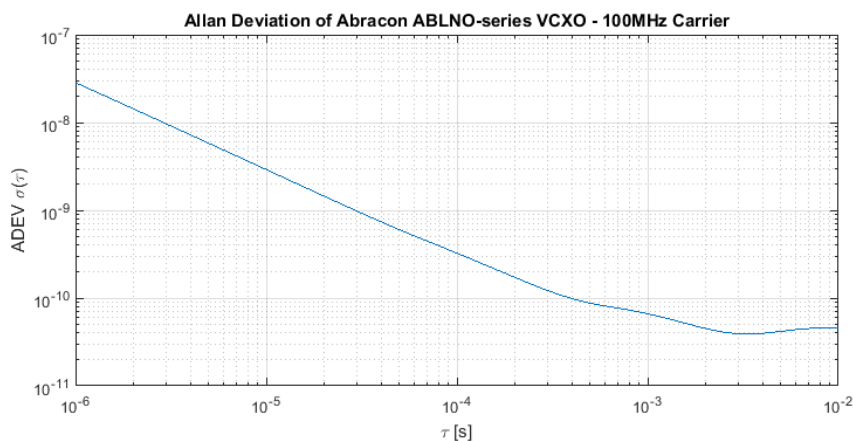Figure A6: Phase noise data of Abracon ABLNO crystal oscillators at 100 MHz

Figure A7: Allan deviation of Abracon ABLNO crystal oscillators at 100 MHz

Abracon AOCJY7TQ Oven Controlled Crystal Oscillator

Figure A8: Features and applications of Abracon AOCJY7TQ crystal oscillators

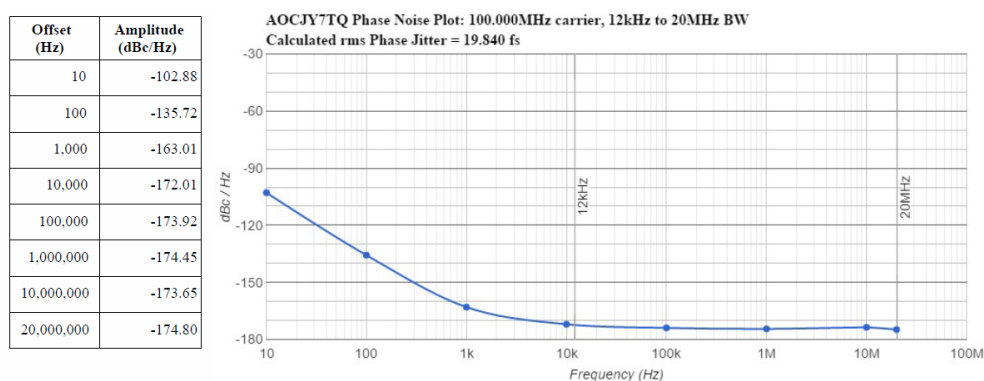| Offset (Hz)	Amplitude (dBc/Hz)
10	-102.88
100	-135.72
1,000	-163.01
10,000	-172.01
100,000	-173.92
1,000,000	-174.45
10,000,000	-173.65
20,000,000	-174.80

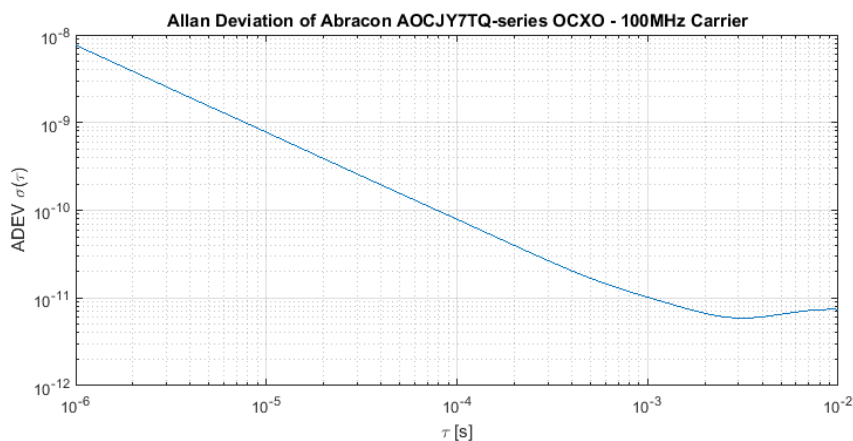Figure A9: Phase noise data of Abracon AOCJY7TQ crystal oscillators at 100 MHz

Figure A10: Allan deviation of Abracon AOCJY7TQ crystal oscillators at 100 MHz