Massive MIMO Handbook

Extended version

1st edition

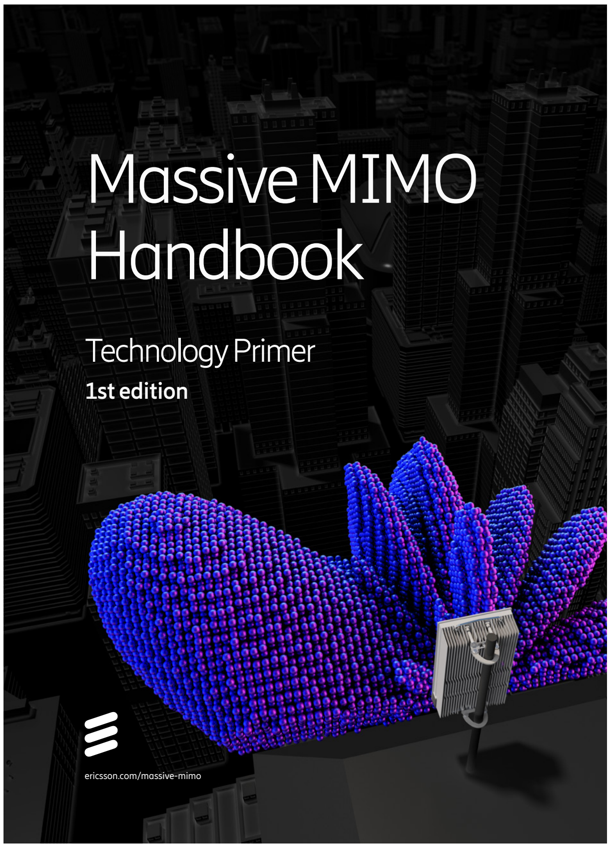
ericsson.com/massive-mimo

Massive MIMO Handbook

Extended version

This book contains two documents:

- Massive MIMO Handbook
- Massive MIMO Handbook – Technology Primer



EN/LZT 4/28701-FGB1010987 Uen Rev A

Massive MIMO Handbook

1st edition

ericsson.com/massive-mimo

Introduction

The handbook provides a guide for how to get most of Massive MIMO in a mobile network

Purpose

The main purpose of the handbook is to provide a guide for how to use Massive MIMO to meet the performance requirements in a 5G mobile networks. It should also provide a guide for how to choose suitable products in typical network deployment scenarios. The handbook shall also briefly explain key aspects of how Massive MIMO works and how the different technology components affect network performance in field.

Target readers

This handbook primarily targets the Massive MIMO stakeholders in the communications service providers´ organizations. It can also be used by internal Ericsson organizations.

Scope

The document focuses on Massive MIMO solutions, including as a means for meeting the performance requirements in the network. Focus is on products operating with time division duplex (TDD) on mid-band spectrum, typically 3.5-3.7 GHz. Conventional radio solutions are also included as an alternative where Massive MIMO is not needed or not cost efficient. Furthermore, emphasis is on the radio solution, i.e. the radio parts and the antenna parts. To keep the document focused and limited in volume, the baseband solution, site solution other than radio parts and the antenna (e.g. power, enclosure, cooling, etc.), transport solutions (backhaul and fronthaul) are not included. High-band (mm Wave) and FDD are not included in this version. The service in focus is mobile broadband (MBB) as this is the dominating service in all mobile networks.

Authors team

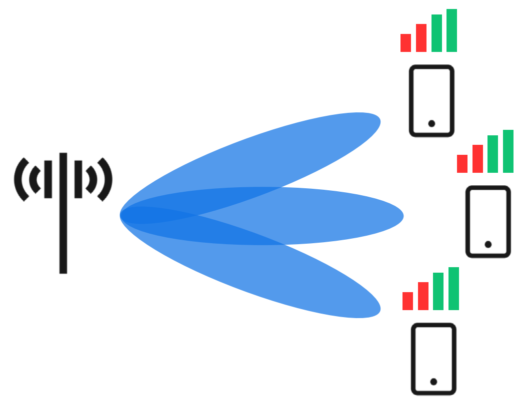
Project manager: Peter von Butovitsch

Authors: Henrik Asplund, David Astely, Peter von Butovitsch, Thomas Chapman, Sebastian Faxer, Mattias Frenne, Christer Friberg, Farshid Ghasemzadeh, Bo Göransson, Billy Hogan, George Jöngren, Jonas Karlsson, Erik Larsson

Review team: Bo Hagerman, Kjell Gustafsson, Raquel Herranz, Thomas Köhler, Peter Moberg, Tomas Sandin, Sibel Tombaz

Executive summary – Massive MIMO to meet network requirements

Massive MIMO explores the spatial domain by using beamforming, spatial multiplexing and null forming



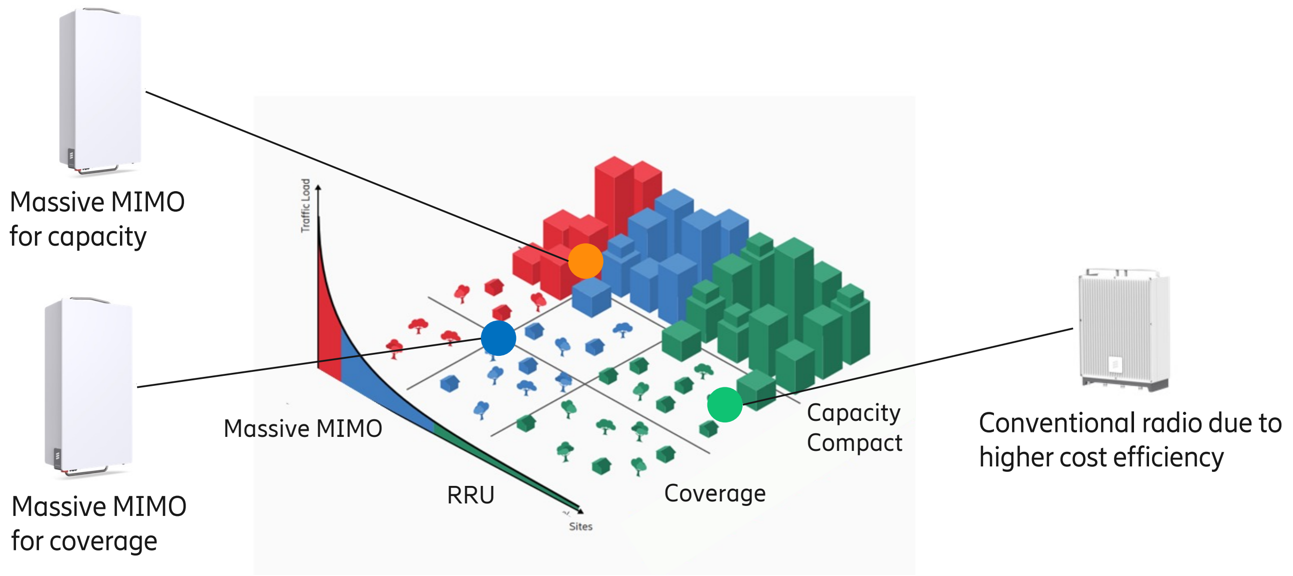
This handbook demonstrates how a service provider can use Massive MIMO to meet the 5G network requirements of today and the future. New use-cases and continuous traffic growth require considerably better user experiences (in terms of speeds and latency) and substantially higher network capacity. Service providers can evolve their mobile networks to meet these requirements with ease of deployment and in a cost-efficient and energy-efficient way with Massive MIMO.

Massive MIMO provides the means for improved coverage, capacity, and user throughput of mobile networks by exploiting the spatial domain. This is achieved by using multi-antenna technologies such as beamforming, null forming, and spatial multiplexing (MIMO), that take advantage of specific channel and antenna array properties.

The most important capability of Massive MIMO is to improve coverage on new and higher 5G frequency bands to enable the same coverage on the 5G bands as for 4G using the existing site grid.

Massive MIMO offers higher capacity and a better user experience than a conventional solution using remote radio units (RRU) and passive antennas. Massive MIMO can carry increased traffic growth over a longer time period.

Executive summary – Finding suitable Massive MIMO solutions



Service providers need to evolve their networks to meet the future needs in a cost- and energy-efficient way. In this process it is important to consider the complete hardware and software solution toolbox and understand how the efficiency of the tools depend on deployment characteristics. Massive MIMO is a vital solution, but other solutions should also be considered including conventional RRU products.

Different areas and sites of the network have different characteristics in terms of traffic load, inter site distance, spread of users in the vertical domain (e.g. low-rise vs high-rise buildings), etc. These characteristics have a profound impact on the effectiveness of different radio solutions. For example, sites with high traffic load need Massive MIMO radios capable of large bandwidths and capacity features like MU-MIMO, sites with large spread of users in the vertical domain (e.g. high-rise buildings) benefit from vertical domain beamforming, and areas with large inter-site distance require radios with high effective isotropic radiated power (EIRP) to ensure coverage. At the same time some sites have constraints on ease of deployment, e.g. size and weight of radios, that need to be considered.

To guide the service provider in the process of finding a suitable radio solution, the network sites are classified according to deployment environment, e.g. rural, suburban, dense urban, traffic load variations and site rebuild complexity issues.

Executive summary – Massive MIMO radios and features

The Massive MIMO portfolio comprises Massive MIMO radios (hardware) and Massive MIMO features (software), and it is designed to meet network performance requirements (coverage, capacity, and user throughput) as well as constraints on ease of deployment, cost-efficiency, and energy efficiency for all site types. The design is guided by the site classification summarized in the previous slide, as well as specific requests from the service providers.

The Ericsson Massive MIMO radio solutions are divided into three segments:

Capacity covers the most capacity-demanding sites and provides superior performance in all deployments ranging from dense urban high rises to rural areas. It supports many radio chains facilitating superior horizontal and vertical domain beamforming.

Coverage targets deployments with large inter-site distances, such as suburban or rural areas with a smaller spread of users in the vertical domain. It has typically fewer radio chains to lower total cost of ownership (TCO) compared to the Capacity products.

Compact is for sites where there are constraints on the deployment. It prioritizes TCO and ensures that mechanical properties are in line with site constraints, such as size and weight. These products still provide substantial performance gains compared to conventional RRUs in many deployment scenarios.

A high-level mapping of these segments into different site types can be found in the previous slide. It needs to be highlighted that capacity, coverage and compact refer here to product segments, and should not be mixed up with network properties such as coverage, capacity and user throughput. That is, there is not a one-to-one mapping between the segment names and 'corresponding' network properties. For example, products from the capacity segment will provide superior overall performance including coverage, capacity and user throughput.

Unlike Massive MIMO radio solutions, Massive MIMO features are developed continuously. They can be deployed at any time providing the Massive MIMO radio solution supports it. The Massive MIMO features are developed to meet the relevant requirements over time in alignment with the entire ecosystem. For example, coverage features are necessary in early deployments, whereas high-capacity features become more important over time.

The purpose of the handbook is to describe how Massive MIMO can be used in a mobile network to meet the network needs. It does not require deep technical understanding. For those interested in Massive MIMO technology, please refer to the book "Advanced Antenna Systems for 5G Network Deployments" [1], which describes many key concepts in detail.

In Chapter 1 the concept of Massive MIMO is introduced. Short explanations of what it is, how it is defined, why it is useful and how it works are given. The main abilities of Massive MIMO to enhance capacity, coverage and use throughput are established. The possibilities for Massive MIMO to improve total cost of ownership (TCO) in some scenarios are also introduced.

Chapter 2 describes the needs of an evolving 5G network. It is established that there are general performance needs such as capacity, end-user experience and coverage (on higher frequency bands). These needs vary across the network and they also vary over time. There are also some constraints to consider, e.g. cost efficiency and energy efficiency. It is shown that there is a good match between the 5G network needs, specifically, the needs w.r.t. capacity, coverage and user throughput, and the Massive MIMO capabilities and that this is the main reason for the high interest in Massive MIMO in 5G networks. The network needs provide input to be considered when designing a product portfolio that shall meet all network needs.

Chapter 3 explains how to design the Massive MIMO solutions, i.e. Massive MIMO radio and Massive MIMO features, to meet the 5G network needs. Firstly, there is a discussion on how to segment a radio portfolio that covers all network needs to a sufficient degree with a limited number of products. Secondly, it is shown how the Ericsson portfolio is segmented to deliver capacity, coverage, user throughput, cost-efficiency and energy efficiency for different relevant deployment scenarios.

Chapter 4 introduces a guide for how a service provider can choose radio and feature solutions to meet their needs, specifically when (typically) to use Massive MIMO and when a conventional solution meet the needs in a better way.

Table of Contents

1. Introduction to Massive MIMO

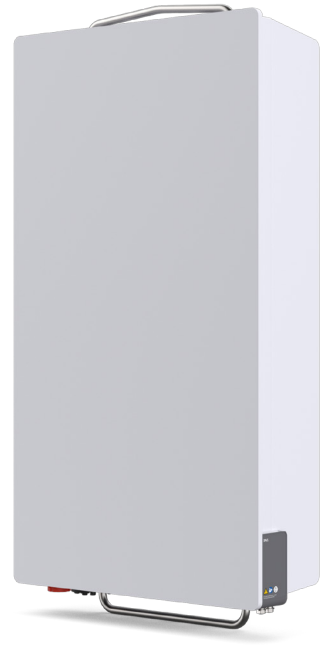
Outline

— History and future of Massive MIMO

— What is Massive MIMO

— Why is Massive MIMO used

— How does Massive MIMO work

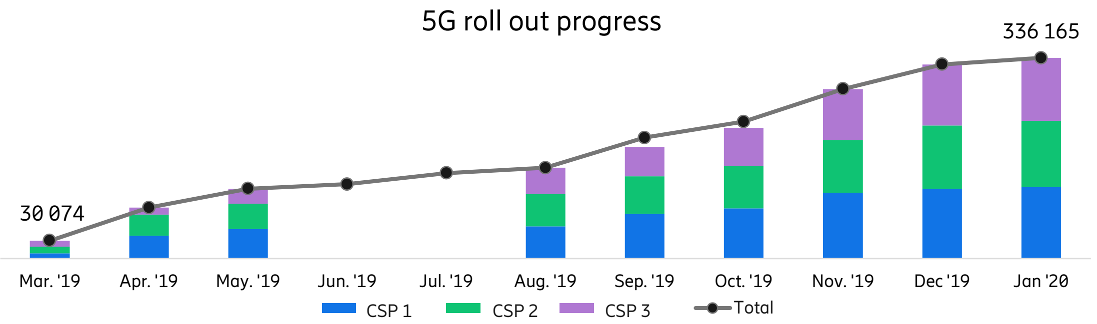


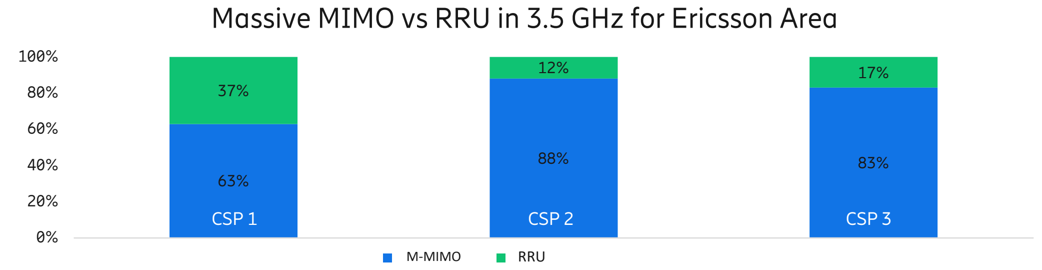
The main purpose of this introduction is to briefly explain what Massive MIMO is, why it is used and how it works.

This chapter starts with a brief history of Massive MIMO, followed by a definition of Massive MIMO and a discussion of the relation to conventional solutions. The main reason why Massive MIMO is so relevant in today's 5G mobile networks is the ability to enhance coverage, capacity and user throughput, in a cost-efficient manner, which is then shortly explained. To deliver performance enhancements, Massive MIMO makes use of multi-antenna technologies and antenna array properties. The multi-antenna technology components, viz. beamforming, MIMO and null forming, are briefly introduced and their relevance in the network is also discussed. The properties of antenna arrays and how they can influence network performance are also introduced.

This brief introduction gives a short glimpse of the very rich and fascinating topic of how Massive MIMO works. Interested readers can however find a more thorough description in [1].

Massive MIMO is here and now
Speeding ahead with 5G



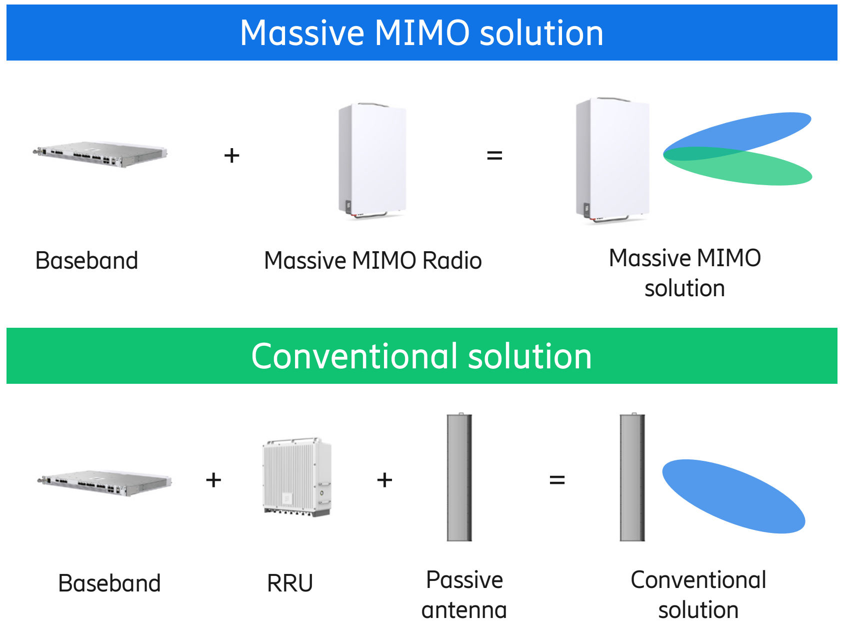


Multi-antenna technologies, such as beamforming and spatial multiplexing, have been known for many decades. The mobile industry began to explore these technologies almost 30 years ago. For 3G/WCDMA, these solutions never reached significant commercial success, and the number of deployed solutions were few. For 4G/LTE, beamforming and MIMO solutions were more advanced and more efficient. They have been used commercially in many deployments. Massive MIMO solutions, the latest evolution of these technologies have been deployed for 4G since 2017/18. With the deployment of 5G/NR, Massive MIMO is an inherent technology component from the start. There is now a significant uptake of 5G Massive MIMO solutions globally, and it is expected that the Massive MIMO share of the network will continue to grow.

Use of Massive MIMO is growing fast with the deployment of 5G networks. Massive MIMO is used as an intrinsic component in all 5G networks from the start and the fraction of Massive MIMO solutions is high. 5G networks are built in a similar way across the world. The build out is made from the dense urban areas where the requirements are highest and gradually outwards to areas with lower traffic.

The graph shows the rollout of radio sites in a country in East Asia with the accumulated number of radio sites for the three service providers (also called communications service providers (CSP), referred to in the figure as CSP 1-3). The lower graph shows the fraction of Massive MIMO versus RRU solutions for the three, respectively. The fraction of Massive MIMO is high in all three networks, ranging from 63% to 88%. The ratio of Massive MIMO products versus RRU is usually different for different deployment environments, higher for dense urban areas and lower for suburban areas. For a given deployment environment, it is expected that the ratio of Massive MIMO compared to RRU will increase over time. Over two years, the number of 5G base stations increased by a factor of 10 from a base of 30,000 in March 2019 to more than 300,000 in March 21. Massive MIMO radios were the radio of choice for the mid-band spectrum at 3.5 GHz with about 80% Massive MIMO versus 20% conventional RRU solutions in Ericsson areas.

Massive MIMO solution



There are many definitions of Massive MIMO. In this handbook, the following definitions are used:

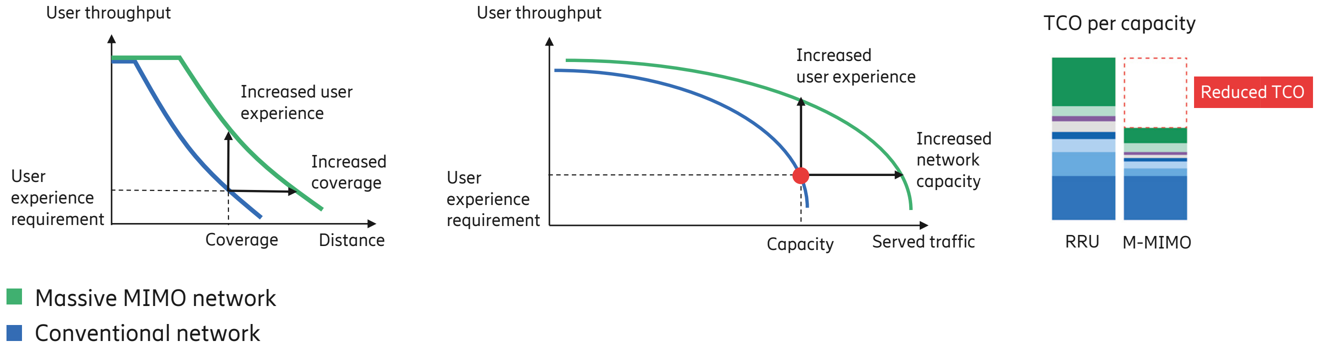
- Massive MIMO is a concept where multi-antenna techniques exploit massively many antennas with dynamically adaptable input and/or output signals.
- A Massive MIMO solution is an implementation of Massive MIMO, consisting of both hardware (Massive MIMO Radio) and software parts (Massive MIMO features).
- A Massive MIMO radio is a hardware unit that comprises an antenna array, a large number of radio transceiver chains (in this document, large is assumed to be 'significantly more than 8'), and parts of the baseband functionality, all tightly integrated, see [Ch. 3.1, p. 34-35].
- A Massive MIMO feature is a multi-antenna (software) feature, such as beamforming, null forming or MIMO or a combination of these. Massive MIMO features can be executed in the Massive MIMO radio or in the baseband* or in both locations [Ch. 3.1, p. 34-35].

There are many concepts that are similar or identical to Massive MIMO. Examples include advanced antenna systems or active antenna systems, both abbreviated as AAS. It should also be noted that the definition of Massive MIMO within academia is somewhat different to that used by industry. There are no hard boundaries between Massive MIMO solutions and conventional solutions. A conventional solution, or a non-Massive MIMO solution, typically comprises a passive antenna and a remote radio unit using a low number of radio chains** (typically 2, 4 or 8). A Massive MIMO solution is typically much more capable than a conventional solution in terms of executing the multi-antenna features.

*) In the Ericsson product portfolio, the baseband solution is either a conventional separate Baseband unit or a cloud-based solution, which is called RAN compute.

**) A radio chain is defined as radio resources required to transform a digital low power signal to an analog high-power signal that can be transmitted over the air and conversely to transform an analog signal to a digital signal on the receiver side. The number of transmitter chains (T) does not need to be the same as the number of receiver chains (R). In this document, it will be assumed that these numbers are equal and when referring to 'radio chains', both the transmitter and the receiver parts are included. The number of radio chains impacts the beamforming and MIMO capabilities. Also, it affects the requirements on signal processing and the total complexity and cost of the product.

Massive MIMO is addressing key network characteristics



Increased Capacity, user throughput and coverage in both uplink and downlink for existing and new frequency bands
Reduced Total-Cost-of-Ownership (TCO) per capacity

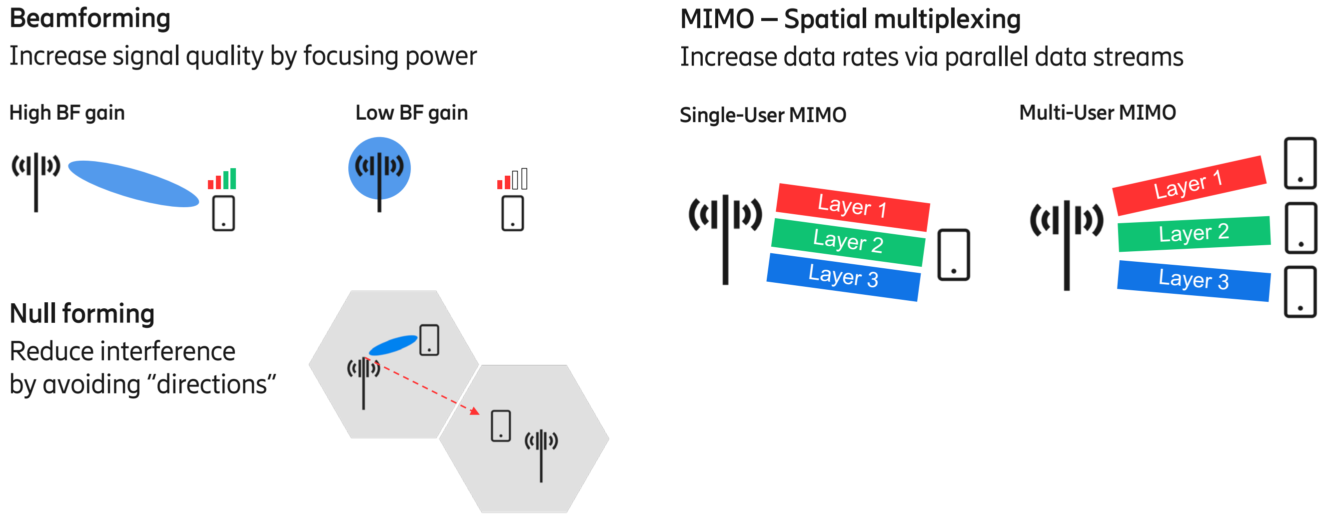
Massive MIMO can increase network coverage, capacity, and user throughput using multi-antenna technologies such as beamforming, null forming and spatial multiplexing (MIMO). These exploit spatial properties of the antenna array and the propagation channel. These techniques are briefly described in the following three pages, [Ch. 1, p. 12-14], and in more depth in [1] and [2].

The figure on the left illustrates single user throughput as a function of distance to the serving base station in an unloaded network for a conventional network and a Massive MIMO network. It is assumed that both networks have the same system settings, e.g. bandwidth and output power. Massive MIMO increases the coverage of all throughput levels. Specifically, it increases the user throughput for most distances (the peak rate is assumed to be the same for both networks). These aspects are thoroughly elaborated in [1].

The middle figure shows user throughput as a function of served traffic. When the served traffic stays the same, the benefit will be seen in terms of increased user throughputs also in a loaded network. At the same time, the figures also show that Massive MIMO can serve more traffic for a given data rate requirement. This means that the network is prepared to handle future traffic growth.

Massive MIMO can also be more cost efficient in some scenarios. In the figure on the right, different colors describe different cost contributions. Here, cost efficiency is measured as TCO per delivered capacity and Massive MIMO can be more cost efficient in this metric when the load is high or will be high during the investment period. This will be further explained in [Ch. 2, p. 27].

Massive MIMO offers a diverse set of techniques addressing the different network needs



There are basically three multi-antenna technology components that contribute to the increased performance of Massive MIMO, viz. beamforming, null forming and spatial multiplexing (also referred to as MIMO). These technologies are applicable to both downlink and uplink and are briefly introduced below.

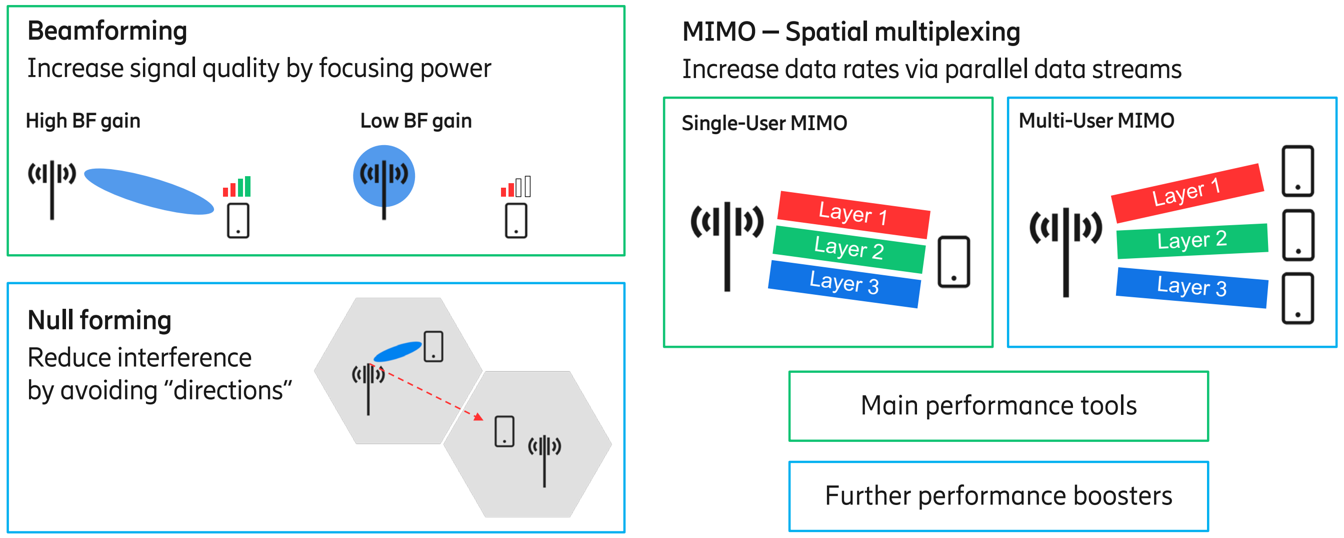
Beamforming (BF): The purpose of beamforming is to amplify transmitted/received signals more in some directions than others. A common goal is to achieve a high BF gain in the direction of the device of interest to improve link quality in terms of signal-to-interference-and-noise-ratio (SINR). Improved link quality translates into better network coverage, capacity, and user throughput. As the energy conservation principle applies, having a higher gain in some directions necessarily implies lower gain in others. Thus, a high gain beam is often narrow while a low gain beam has a more isotropic radiation pattern. Note: The beam is often depicted as one lobe directed to the receiver. In reality, the beam distributes power in many directions. This is further described in [1, Ch. 4-6].

Null forming: Null forming is a form of beamforming that strives to lower the beam gain in certain directions or reduce it to zero. By intentionally designing the beam shape to have nulls or lower gain in the directions where the victim transceivers are, interfering signals can be reduced. The picture illustrates null forming that mitigates inter-cell interference while communicating with a UE of interest in the own cell. Note: Null forming is usually a concept used for the downlink. For the uplink, the concept corresponding to null forming is interference rejection combining (IRC).

Spatial multiplexing: In massive MIMO, it is possible to multiplex several data streams on the same time-frequency symbol. This is often called spectrum re-use. The data streams are commonly referred to as layers. The multiplexed data streams can all go to the same device or to different devices. The case when all the layers belong to the same transmitting/receiving device is referred to as single-user MIMO (SU-MIMO) while the case involving spatial multiplexing of multiple transmitting and/or receiving devices is called multi-user MIMO (MU-MIMO). Spatial multiplexing can increase user throughput and network capacity.

These fundamental technology components can be combined in various ways to form a complete massive MIMO feature, as is shown in chapter 3.2. More details are found in chapter 6 of [1].

The fundamental massive MIMO techniques have different levels of importance



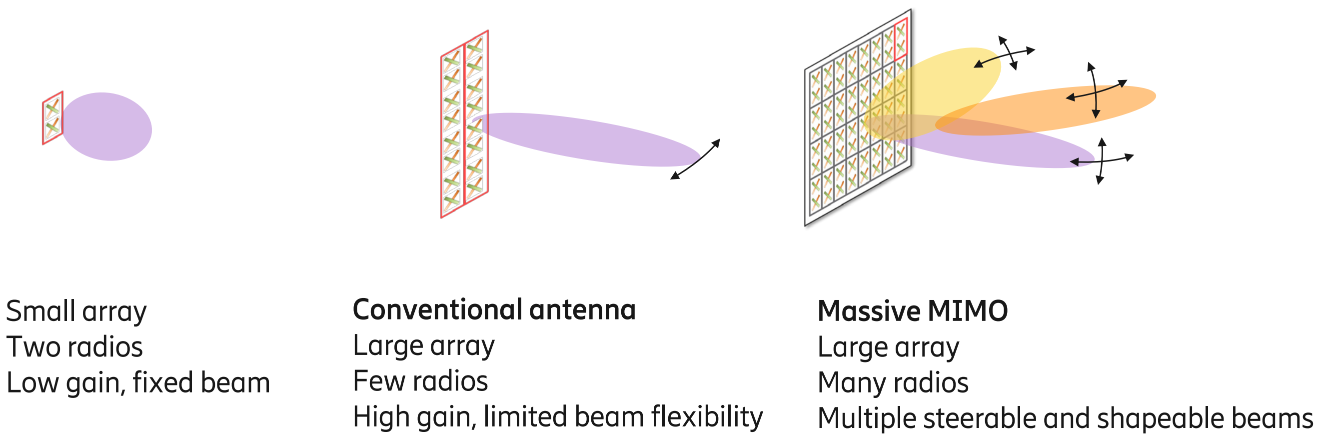
Main performance tools
Beamforming and single-user MIMO are very central to Massive MIMO. The ability of beamforming to increase the received signal level is crucial to obtain high performance. The impact on interference levels may however be small or non-existent, unless intentional null forming is used. Substantial beamforming gains can be achieved in a wide range of situations, regardless of downlink/uplink, traffic load, or if the user is in a good or bad spot. Capacity and user throughput generally improve. Improved coverage is an obvious and especially important benefit, facilitating reuse of existing site grids for higher frequencies.

Spatial multiplexing via single-user MIMO (SU-MIMO) benefits from high signal levels. Beamforming helps to improve signal levels which then can be exploited for single-user spatial multiplexing. Particularly in downlink, more than one layer to a specific user can often be supported in rather large parts of a cell. This contributes to its relatively general applicability.

Further performance boosters
MU-MIMO improves performance at high traffic loads and good channel conditions. These are conflicting requirements as high traffic loads often lead to higher inter-cell interference levels, which means worse channel conditions. Not all cells benefit significantly from MU-MIMO. Gains from beamforming are of greater importance. Compared to SU-MIMO, there are considerably more requirements on MU-MIMO to reach meaningful performance improvements. MU-MIMO is however nevertheless a great capacity enhancement tool for highly loaded cells as is also further discussed [1, Ch. 6].

Intentional null forming to selected users serves to reduce interference to those users. It is a key sub-component of multi-user MIMO to mitigate intra-cell interference and it is also commonly used on the receiver side in both uplink and downlink to suppress inter-cell interference. The inter-cell interference may be substantial in some scenarios and then gains from null forming on top of the performance given by beamforming and single-user MIMO can be significant.

A Massive MIMO radio uses antenna arrays to enable Massive MIMO techniques



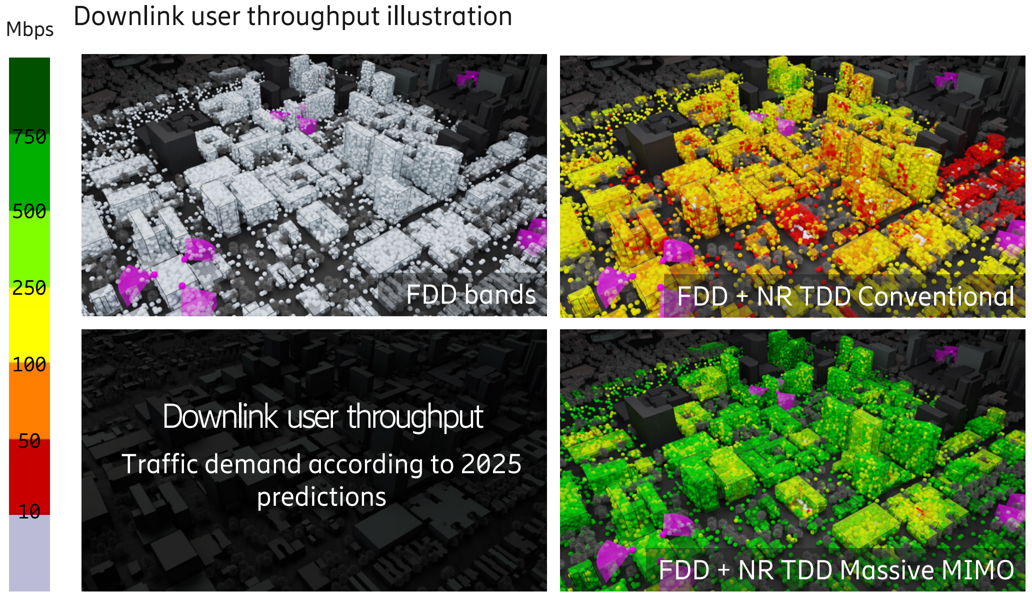
The capabilities of doing beamforming and MIMO depend on the ability to control the amplitude and phase of parts, or even single elements, of the antenna array. The dimensions and the properties of the antenna array affect the network performance and are key inputs to the product portfolio. The antenna array, its structure and its capabilities are here shortly introduced. For an extensive description of antenna array properties, see [1].

The maximum antenna gain is proportional to the total antenna array area. A small antenna provides low gain, while a large antenna provides high gain. See the figures on the left and in the middle. It follows also that the width of the main lobe, in horizontal/vertical dimension, is inversely proportional to the size of the antenna array in horizontal/vertical dimension. A wide antenna will therefore typically produce narrow lobes and a narrow antenna will produce wide lobes.

The properties of the antenna array can be shaped by partitioning the array into an array of subarrays (AOSA) where each subarray is controlled individually. The array in the middle figure has two subarrays stacked horizontally in a 1x2 AOSA and the array in the right figure consists of 32 subarrays stacked horizontally and vertically in a 4x8 AOSA structure. The beampattern of the antenna array is controlled by steering the amplitude and phase of the signals fed to the subarray. The purpose of using subarrays is to get the appropriate steerability of the array while keeping the complexity, and hence the cost, at an adequate level. The finer partition (smaller subarrays), the better steerability, but higher cost and complexity, since more radio chains are needed for the same size of the array.

The size and shape of the antenna, and how it is partitioned into subarrays, affect Massive MIMO performance. Design choices for these parameters are important to provide products with the desired properties. Also of particular importance is the EIRP, which is a combination of output power and antenna gain.

Massive MIMO – future-proof solution providing superior experience



By applying the methods outlined in previous slides, Massive MIMO can support performance that is superior to that of conventional radios. The figure illustrates downlink user throughput for an example deployment scenario given the expected traffic demand of 2025 and compares different solutions.

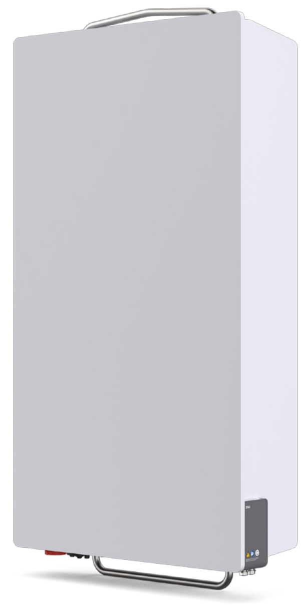
The deployment scenario is Atlanta midtown (US dense urban scenario), and results are based on multi-band simulations for three different radio deployment options:

- FDD bands only: FDD 700 MHz with 20 MHz bandwidth and 2T2R radio, and FDD 2 GHz with 40 MHz bandwidth and 4T4R radio
- FDD bands plus 3.5 GHz TDD with 100 MHz and 8T8R radios
- FDD bands plus 3.5 GHz TDD with 100 MHz and Massive MIMO (64T64R)

In this time period, typical frequency division duplex (FDD) bands cannot fully support the traffic demand. Adding time division duplex (TDD) mid-band spectrum, such as the C-band in the US, with conventional radios (RRUs) improves the performance, while the combination of TDD mid-band spectrum and Massive MIMO provides superior user experiences.

Summary – Massive MIMO improves relevant mobile network requirements

- Massive MIMO is an integral part of 5G. It has already reached general use and is growing rapidly
- Massive MIMO solutions consist of a hardware and a software part
- Massive MIMO provides higher coverage, capacity and user throughput and can increase cost efficiency
- Massive MIMO features explore the spatial domain and build on techniques such as beamforming, null forming and spatial multiplexing



Massive MIMO makes use of many antennas with adaptive input and output signals. A Massive MIMO solution consists of a radio hardware part (Massive MIMO Radio) and multi-antenna features (Massive MIMO features).

Massive MIMO techniques are applicable in both downlink and uplink and can provide better coverage, capacity and user throughput, which are useful characteristics in a mobile network with continuously growing traffic. Massive MIMO solutions can also be more cost efficient than conventional RRU solutions in many scenarios.

Massive MIMO makes use of multi-antenna features and the antenna array characteristics to reach the desired performance benefits. These characteristics are exploited by the Massive MIMO hardware and software solutions. The most important multi-antenna features are beamforming and single-user MIMO. These features provide benefits in almost all environments. Examples of other multi-antenna features are multi-user MIMO, that can provide significant benefits in some environments under certain conditions, and null forming that can suppress interference from certain identified users or suppress inter-cell interference. The antenna array characteristics are also important to shape the properties of a Massive MIMO solution, e.g. size, number of elements in horizontal and vertical dimensions and subarray partitioning.

2. Requirements in an evolving 5G network

Outline

— Requirement drivers

— Network requirements and solution options

— Deployment constraints

— Network build and site classification

The purpose of this chapter is to discuss the most important network performance requirements and constraints impacting 5G network evolutions and highlighting where Massive MIMO can make a difference.

5G networks need to be evolved to handle increased traffic growth and enhanced user experience, both for old and new services, which leads to requirements on increased coverage, capacity and user throughput.

The requirements for specific sites are also discussed as they depend on the environment in which the sites are deployed, the actual load and the load variations over time.

In addition to the network requirements, there are also constraints to consider, viz. ease of deployment, cost efficiency and energy efficiency. These will also be discussed.

A new 5G experience

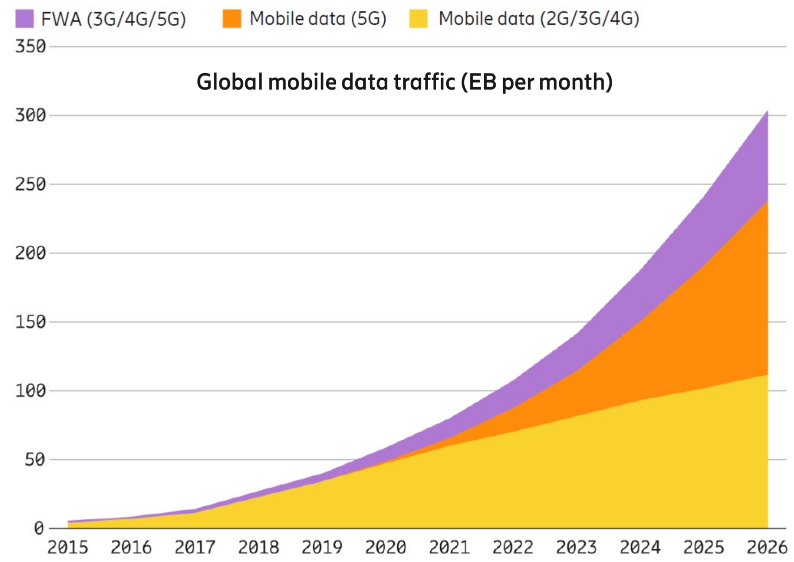
It is expected that 5G will deliver a significantly better user experience; any service, any time, anywhere, both for existing MBB services and new services*/use cases. Many new services require higher user throughput and/or have more stringent reliability requirements (time-to-content) than today's best effort MBB services. Higher reliability translates into higher capacity requirements as stricter reliability and/or latency requirements will require more radio resources than best effort MBB services to keep required margins.

To meet these expectations, the entire telecom ecosystem needs to be evolved. Several new use cases, and the functionality required to support them, are developed in standardization forums. The standardization bodies provide an extensive toolbox that network and UE providers need to choose between; what features to implement and when to do it. Implementing new network features often requires addition of software and hardware, and although the standards specify much of the functionality, there is still a possibility to differentiate beyond the basic standard support using different products. For example, as described in [Ch. 1, p. 11], Massive MIMO is one solution that can offer capabilities and performance that exceeds conventional RRUs.

* Such services can include for example, advanced gaming services or industrial control applications with low tolerance for packet loss and can include XR (extended reality), AR (augmented reality) and VR(virtual reality) and MR (mixed reality) services. Fixed Wireless Access (FWA), [3], is another service that is already available in some markets and is expected to grow.

Higher capacity is required in networks delivering a 5G experience

MBB data traffic alone is growing by 50-70% per year



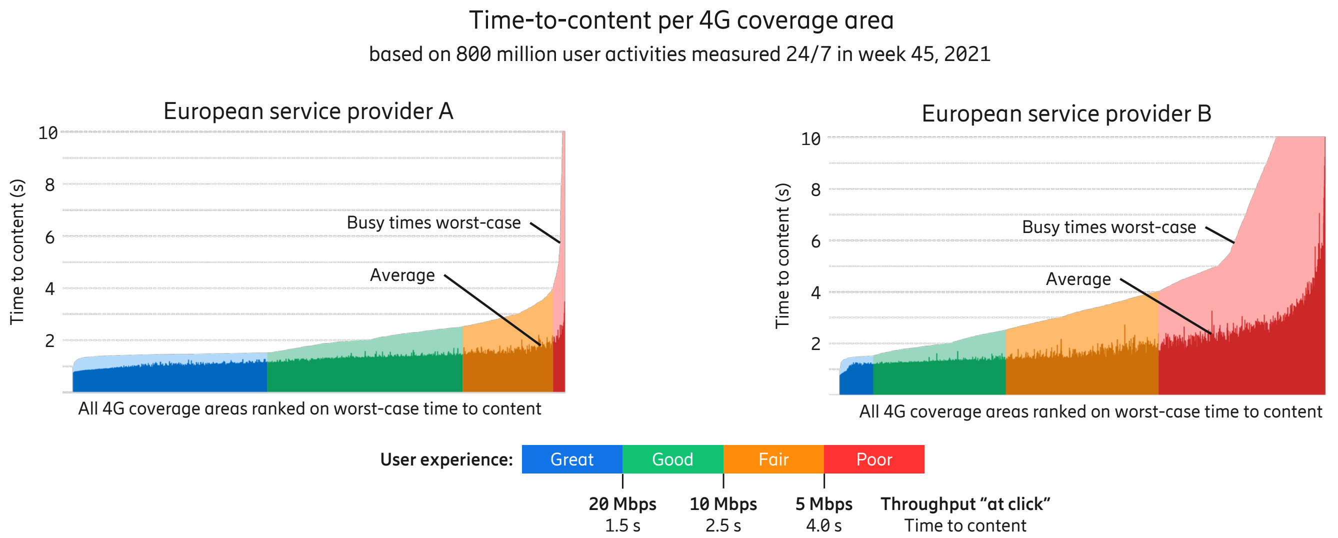
Globally, traffic in MBB* networks has increased 50-70% each year for many years, see figure**. This growth is expected to continue at a similar or higher pace in the future. To cater for this traffic growth, mobile networks must support corresponding capacity growth. As discussed on the previous page, [Ch. 2, p. 19], the introduction of new services like augmented reality (AR) and virtual reality (VR) could further accelerate the rate of load increase.

Evolving a network requires increased capacity over time to maintain high service quality.

* MBB, often defined as 'best effort' data traffic, is the traffic type that dominates in mobile networks today. MBB networks are primarily built and expanded to meet the continually growing data consumption need. Mobile broadband services include web browsing, social networking, video streaming (Netflix, YouTube), email and messaging, online gaming, audio services, and file sharing.

** The source of the graph is Ericsson Mobility Report June 2021, [4].

4G user experience challenged by congestion in busy areas during busy times



There is an important relation between user experience and network capacity. A network that is dimensioned to handle the peak load can consistently deliver a high better user experience, whereas networks that are insufficiently dimensioned will experience impacts on user performance during high load.

Time-to-content is an important metric that describes the time it takes to access a service after initialization, for example, the time it takes to load a web page after clicking the link. This metric is what really matters for the end users. Studies done by Ericsson show a clear correlation between time-to-content and end-user experience. This is illustrated by the coloring explained in the lower legend, where user experience is rated from poor to great. Furthermore, a consistently low time-to-content is crucial for online businesses; be it e-commerce, e-travel, e-banking, or other.

The figures show time-to-content as a function of all cells sorted from worst to best busy-hour time-to-content for two different networks. Each cell in the network is measured every 15 minutes over a period of one week, and the darker colors represent the median (50th percentile) and the lighter colors represent the busy-hour (90th percentile) time-to-content measurements. In total there are 800 million measurements.

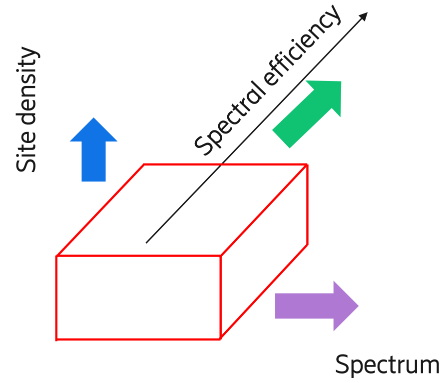
Service provider A (left figure) has a high-capacity network that is rated best-in-class. The network can therefore offer high performance in all cells at any time. Busy-hour time-to-content is rated good or great (<2.5 seconds) for most of the cells and poor performance for only about 4% of the cells. Service provider B (right figure) has a lower capacity network and time-to-content is rated Poor (>4 seconds) for about 30% of the cells.

To summarize, it is very important to have sufficiently high capacity in all cells to avoid congestion during peak hours. Congestion leads to degraded user experience and poor time-to-content.

Expanding the volume of the capacity box

Network capacity can be represented conceptually by the volume of the red box

To increase capacity, increase the size of the box in any dimension



Higher capacity can be achieved by increasing spectrum, spectral efficiency, and site density
Massive MIMO systems increase spectral efficiency (green arrow) compared to conventional systems

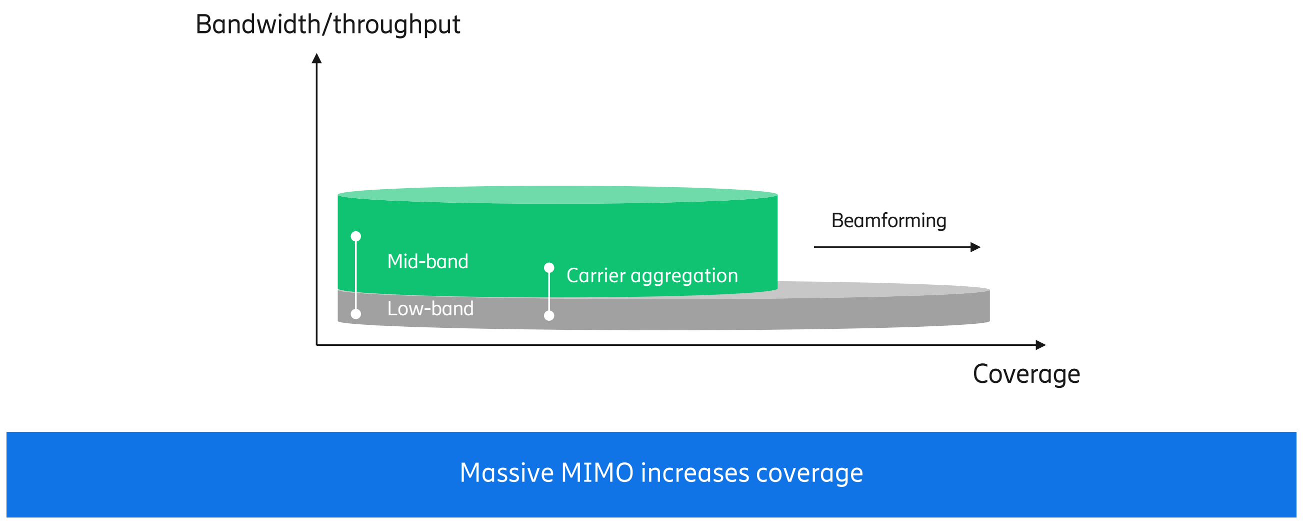
There are several ways to deliver more capacity in a mobile network. The preferred way of doing this depends on several factors, e.g. availability of spectrum, cost efficiency, CAPEX and OPEX budgets, accessibility and cost for new sites, etc. As outlined in the figure, there are essentially three ways of increasing capacity, adding spectrum, increasing spectral efficiency in the network and increasing the site density.

Adding spectrum on new or existing frequency bands is a very efficient way to add capacity. Many loaded sites use all available spectrum already. With the introduction of 5G, however, most countries release new spectrum, in most cases on mid-band around 3.5 GHz. The amount of new 5G spectrum is considerable, typically 100 MHz, and can hence provide much capacity. To increase capacity in a 5G network, it is therefore imperative to make use of the available mid-band spectrum.

Spectral efficiency can be improved in many ways. A very efficient way of doing this is however by exploiting the spatial domain with multi-antenna features such as beamforming and MIMO. Thus, Massive MIMO provides important tools to further increase network capacity. The effectiveness of these features vary as indicated in [Ch. 1, p. 13] and extensively elaborated in [1; Ch. 13].

Increasing site density in network is a third way of increasing capacity, either by densifying the network with sites on roof-top level, which is often preferred, or with street-level sites. Radio sites are however often expensive to acquire and to maintain. Hence, site densification is usually exploited when the other options are exhausted.

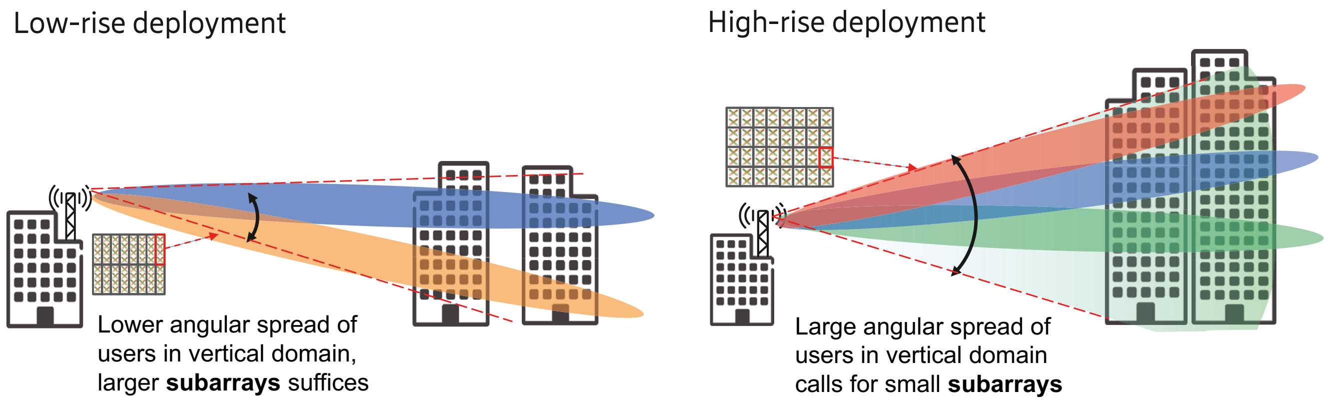
Coverage on new 5G bands provides network capacity



Existing networks are typically built to provide coverage 'essentially everywhere' using either 1.8 GHz (GSM) or 2.1 GHz (3G). Many networks are also densified to provide coverage 'essentially everywhere' in certain areas also for 2.6 GHz. For 5G networks and higher frequency bands, coverage using existing site grid will however be more challenging, [1; Ch. 3]. There are two main reasons for why coverage will be more challenging. Firstly, radio propagation becomes more challenging the higher the frequency (the physical reality). Secondly, 5G services often have higher throughput requirements, that is, 5G networks need to be dimensioned for higher data rates compared to existing 2G-4G networks. The latter is often referred to as data coverage as opposed to access coverage that relates to the ability to connect to the network and have sufficient coverage for low-rate signaling and control channels.

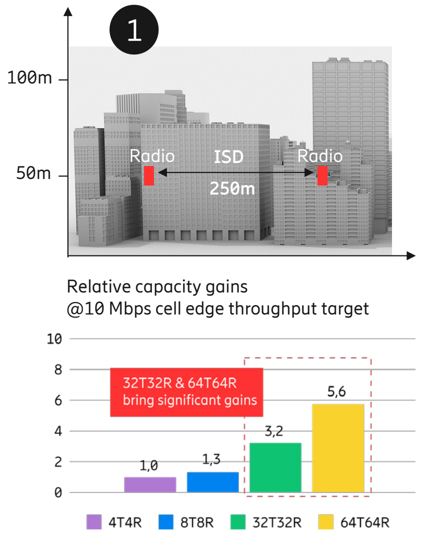
New mid-band frequencies allocated for 5G are located around 3.5 GHz and do hence not have the same coverage as networks built to provide coverage at 2.6 GHz or lower frequencies. In order to make full use of the new 5G bands in the entire existing coverage area, there is a need for improved coverage on the 5G frequency band. Massive MIMO can improve both uplink and downlink coverage by improving the antenna gain through beamforming.

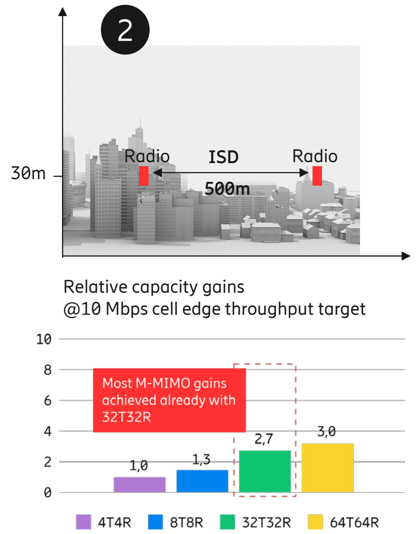
Massive MIMO provides coverage where needed (spatial flexibility)

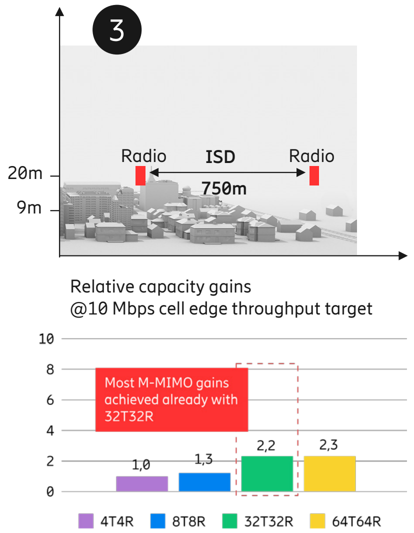


All mobile networks must provide coverage in the entire service area (the area where users are located). For most Macro networks, the service area typically include the entire horizontal domain, while which part of the vertical domain that must be covered is highly scenario dependent, ranging from small range of vertical angles in suburban/rural environments to potentially a large range of vertical angles in dense high-rise environments. This observation has an impact on choice of Massive MIMO product, where smaller subarrays are needed to have a wide vertical coverage area, leading to many radio chains in order to maximize antenna area, hence total antenna gain, while larger subarrays suffice for a narrower vertical coverage area.

Deployment scenarios and load affect the performance of the different products







Beamforming in the vertical dimension also affects capacity, which is illustrated by the following example.

The figures show downlink capacity gains of four different products in three different deployment scenarios, dense urban high-rise (figure 1), urban mid-rise (figure 2) and suburban (figure 3). The products are 4T (1x2 AOSA) RRU, 8T (1x4 AOSA) RRU, 32T (2x8 AOSA) Massive MIMO and 64T (4x8 AOSA) Massive MIMO, and gains are relative to the 4T RRU product.

In a dense urban high-rise deployment (figure 1), there are substantial gains from Massive MIMO compared to RRUs, and 64T64R provides significant gains over 32T32R. Dense urban high-rise is a deployment where a capable Massive MIMO radio product with advanced beamforming features excels. Due to significant spread of users in the vertical domain (cells are almost as tall as wide), both horizontal beamforming and vertical domain beamforming are useful. Note: In some high-rise deployments, indoor systems are often installed in the high-rise buildings. This reduces much of the need to cover these buildings from the outdoor radio sites and the benefit of vertical beamforming is in practice lower in those cases.

In the urban mid-rise deployment (figure 2), there are still substantial gains from Massive MIMO compared to RRUs but gains from 64T64R compared to 32T32R are significantly lower. The main reason that 32T32R performs almost as well as 64T64R is that the spread of users in the vertical domain is much lower, so larger subarrays can be used and gains from vertical domain beamforming are smaller. In this deployment, the lower complexity 32T32R product has similar performance as the 64T64R product but is more cost-efficient, which can be a deciding factor when choosing product. It should also be noted that the performance of both Massive MIMO configurations is still significantly better than the conventional system regarding offered capacity.

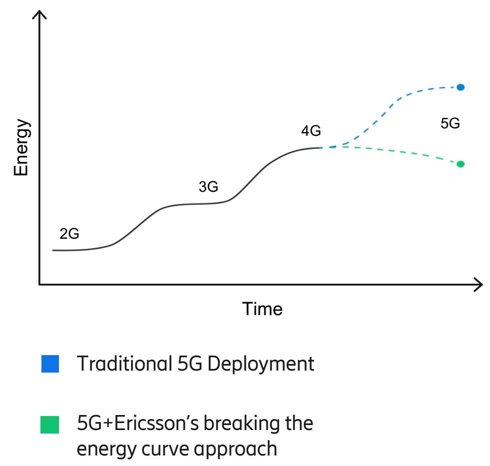
In the suburban deployment (figure 3), the gains from Massive MIMO compared to RRU are still significant, but smaller than for urban mid-rise. Gains from 64T64R compared to 32T32R are also smaller. Hence, 32T32R is more cost efficient than 64T64R also in this case. RRU can also be an option for sites with low requirements on capacity.

Three constraints for network evolution

Ease of deployment

TCO per capacity

Energy efficiency



In addition to fulfilling the performance requirements outlined in previous slides, the solutions must also meet some deployment related constraints to be attractive to service providers. The most significant and generally applicable constraints are ease of deployment, cost efficiency and energy efficiency:

Ease of deployment: Having products that are easy to install and fit well into the environments where they are deployed is of key importance in many cases. In many deployments, requirements include physical size, ease of installation, and visual impact. These may be driven by site building constraints, esthetical requirements, wind load, and site accessibility. In some cases, they represent determining factors in the product selection process. In many cases, it can be a difficult process to seek permission to rebuild a site regardless if the site is owned by a landlord or an infrastructure company where the site lease agreement is regulated in site rental fees. For this case, small and discrete product variants will be easier to deploy and can reduce the site rental fee.

Cost efficiency: All service providers have limited CAPEX and OPEX budgets. They need to build mobile networks that are cost efficient over time. It is suggested to measure cost efficiency as TCO per delivered capacity, as cost efficiency over the whole investment period is of importance. To reach cost efficiency requires an appropriate choice of product for efficient use over a longer time period. The typical service provider investment cycle is typically around 5-7 years. Therefore, it is reasonable to seek optimum product solutions for the corresponding time period. In the figure on the right, different colors describe different cost contributions.

Energy efficiency: Adding more equipment that operates on new frequency bands requires additional use of energy. The cost for the network energy consumption is significant already today and even further increased energy consumption when adding 5G to the network is unwanted. Service providers look for minimum use of energy for 5G equipment and reduced use of energy on other equipment wherever possible. There is also an ever-increasing requirement on reducing the CO_2 footprint which also relates to minimizing energy consumption. New standalone 5G solutions must consume low levels of energy. New site solutions that also include replacement of parts of the existing solution must minimize the energy consumption for the whole site solution.

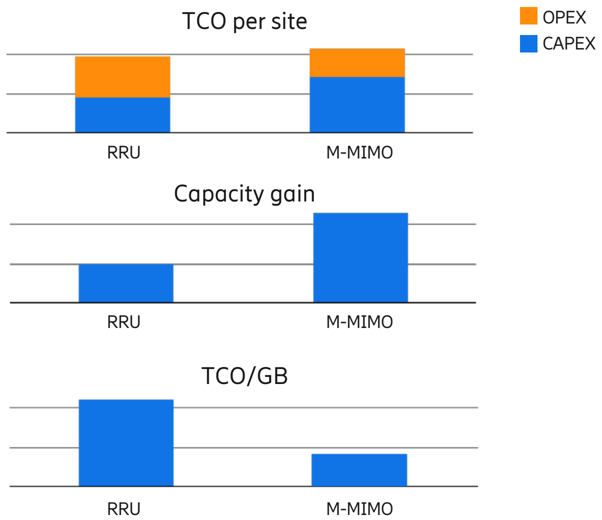
These constraints need to be considered in all deployments and are therefore expected both for Massive MIMO and conventional RRU. They may also be deciding factors concerning the product to choose. This is further elaborated in [Ch. 4, p. 59].

TCO per delivered capacity matters for investment decision

Cost per site
Small differences in TCO for remote radio units (RRU) and Massive MIMO (M-MIMO)

Capacity needs vs. product potential
Massive MIMO efficiency is scenario dependent

Cost per delivered capacity
Massive MIMO has superior total cost per capacity



A per-site total cost of ownership (TCO) analysis is an important step in the process of deciding a recommended radio solution. TCO calculations depend on many local and factors specific to the service provider. This slide illustrates three steps of the TCO analysis. It highlights that what matters most for investment decisions is TCO per delivered capacity.

The figure illustrates the three main steps in the TCO analysis where an 8T RRU product is compared with a 64T Massive MIMO product in an urban deployment scenario using mid-band 3.5 GHz.

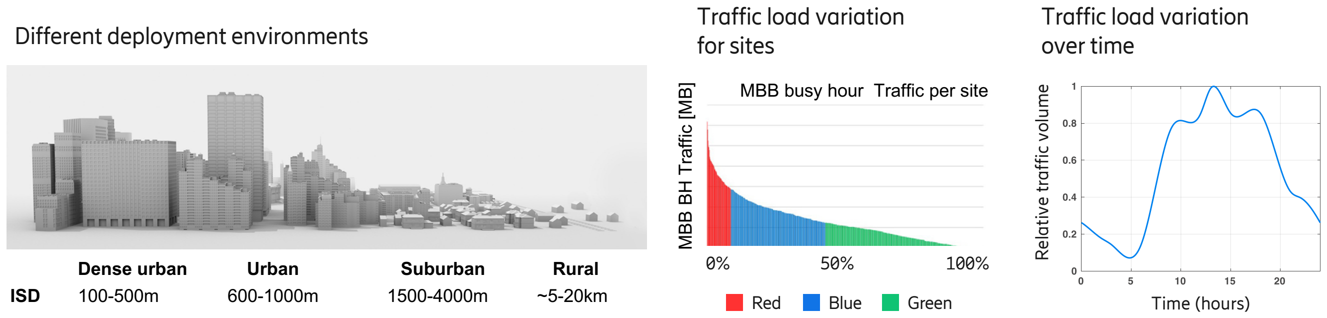
First step is to run a traditional TCO analysis of the cost parts, where both CAPEX and OPEX are evaluate over a typical depreciation time. Although the hardware cost for Massive MIMO is higher than for conventional solutions, this step often demonstrates that the TCO difference between deploying mid-band with a conventional solution and a more advanced Massive MIMO product is rather small. This is mainly because the conventional solution has higher cost related to baseband capacity, network roll out and site rental.

Second step is to determine the offered capacity of the products. As illustrated on the previous page, the capacity benefits of Massive MIMO over a conventional solution depend on many aspects. In many deployment scenarios it is reasonable to assume ~2-3 times more capacity from Massive MIMO than a conventional system. The figure illustrates that 64T Massive MIMO provides twice the capacity compared to 8T RRU in an urban mid-rise deployment scenario.

Third step is then to normalize the TCO results from the first step with the predicted capacity gain from the second step. As the TCO for Massive MIMO is only slightly higher than for RRU, but Massive MIMO provides roughly twice the capacity compared to the RRU, the TCO per capacity is reduced with more than 40% if Massive MIMO is used.

The analysis illustrated here indicates that Massive MIMO is a cost-efficient solution for mid-band deployments in many cases.

Requirement variations in the network



Coverage 'essentially everywhere' and capacity where needed
Different requirements in different parts of the network

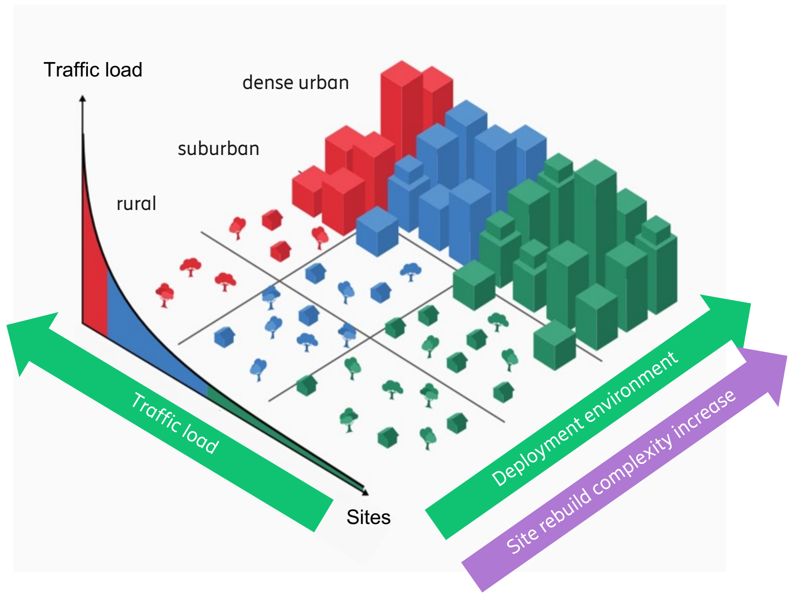
Both network performance requirements and constraints can vary significantly between different sites and different parts of the network. This will affect the choice of radio solutions.

Different deployment areas have different characteristics in terms of typical building height distribution, inter-site distances (ISD), subscriber density, use of services, building wall thickness, window glass characteristics, and occurrence of vegetation. The characteristics of the deployment scenarios have an impact on which products serve future network requirements in the most cost-efficient manner. For example, the distribution of users in the vertical domain impacts both capacity and the need for coverage in the vertical domain, as shown in [Ch. 2, p. 24-25].

Traffic load varies between sites in real networks. A few sites carry a significant amount of the total network traffic, whereas most sites are modestly loaded. The data in the graph in the middle figure show a typical load distribution in a 4G network. In this graph, the sites are ordered from the most loaded site to the left (RED) to the modestly loaded sites in the middle (BLUE) to the least loaded site to the right (GREEN). The value on the vertical axis is the level of load in the site. This load distribution is a function of the distributions of sites, users and the use of services in different parts of the network. It is therefore very similar for 3G and 4G and is expected to remain roughly similar also for 5G. The distribution looks similar in most networks. The high-load and low-load sites can be found in any part of the network. The performance requirements will be different as they depend on the actual site load. This load distribution is an important input to portfolio segmentation. Also, it impacts the product selection of the sites. Later in this chapter, different load segments will be defined which will be useful when selecting products for the sites, respectively.

There are also significant load variations in the cells, as illustrated in the figure to the right. During peak hours, the load can be considerably higher than during off-peak. If the capacity in the cell is not sufficient, the cell will be congested and there will be an impact to the end-user experience, as shown in [Ch. 2, p. 21].

Site classification – Traffic load and deployment environments



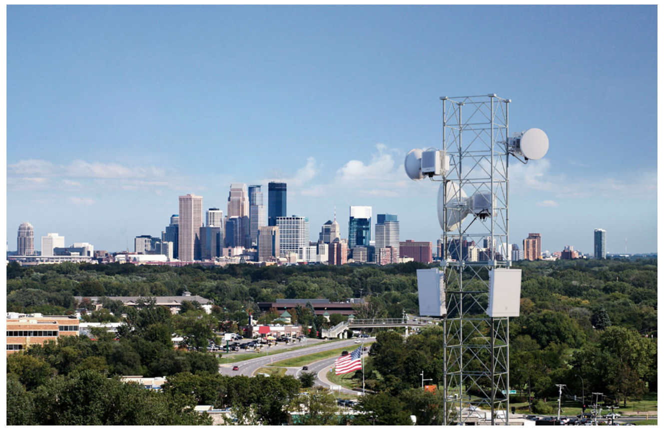
To guide the design of a radio product portfolio and to understand where to use each product, the requirements on a site level are now mapped to different deployment scenarios and traffic load levels.

As discussed, a key factor determining the effectiveness of different Massive MIMO configurations is the deployment scenario and in particular the distribution of users in the vertical domain. The site density (inter-site distance) is also an important deployment parameter that effects coverage.

Another important parameter is the traffic load or capacity need, that is known to vary between different sites, as shown on the previous page, [Ch. 2, p. 28]. Three different load categories are considered, high (RED), medium (BLUE) and low (GREEN) load. These load categories are then mapped onto the different deployment scenarios, ranging from rural to dense urban high-rise. A third important dimension is site rebuild complexity that relates to ease of deployment.

Summary

- New services, improved user experience and increased network traffic load drive network requirements
- 5G networks will require significant enhancements of coverage, capacity and user throughput
- 5G networks must be easy to deploy and be cost- and energy efficient
- Each site is unique with respect to performance requirements and deployment constraints



A mobile network that delivers 5G experiences must not only enhance user experiences but also handle exponentially increasing traffic loads. This is achieved by increasing capacity, enhancing user throughput, and improving coverage on the new and higher 5G frequency bands (to make use of the full capacity on the 5G band in the whole coverage area using the existing site grid).

In addition to these network performance requirements, there are also some constraints to consider when choosing the appropriate product solution for a specific site. Some environments put requirements on ease of deployment, e.g. requirements on small size, low weight and/or low visual impact. Cost efficiency and energy efficiency are general constraints that all service providers need to consider. All service providers have constraints on both CAPEX and OPEX and for most service providers there are both regulatory and legal requirements, political goals or incentives and corporate strategies to limit energy consumption and the production of CO_2. These constraints are thus highly relevant evaluation criteria in the process of product selection.

The requirements and constraints outlined above serve as input to the design of radio and feature solutions portfolios. As can be seen, the capabilities of Massive MIMO outlined in Chapter 1 meet many of the requirements outlined here in Chapter 2. Hence, Massive MIMO is an attractive solution for many 5G deployment scenarios. The principles for how a portfolio can be structured and examples of how it is done by Ericsson are outlined in Chapter 3.

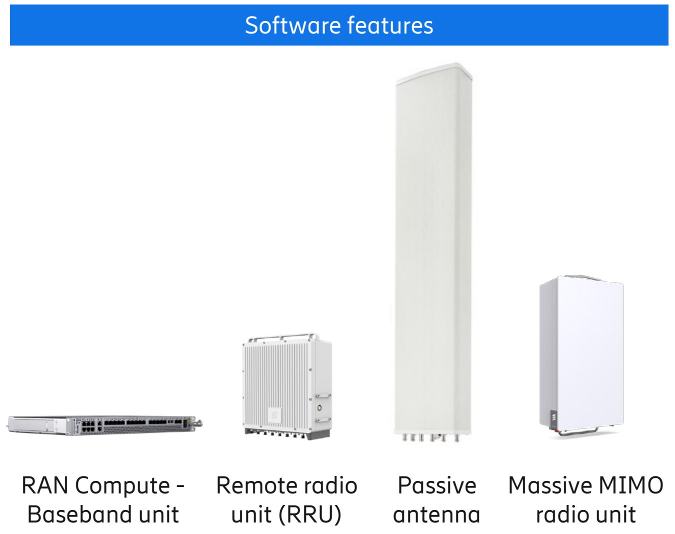
3. Massive MIMO solutions

Outline

— **Radio architectures** differences between Massive MIMO radio and conventional radio

— Ericsson **Massive MIMO radio portfolio**

— Ericsson **Massive MIMO feature portfolio**



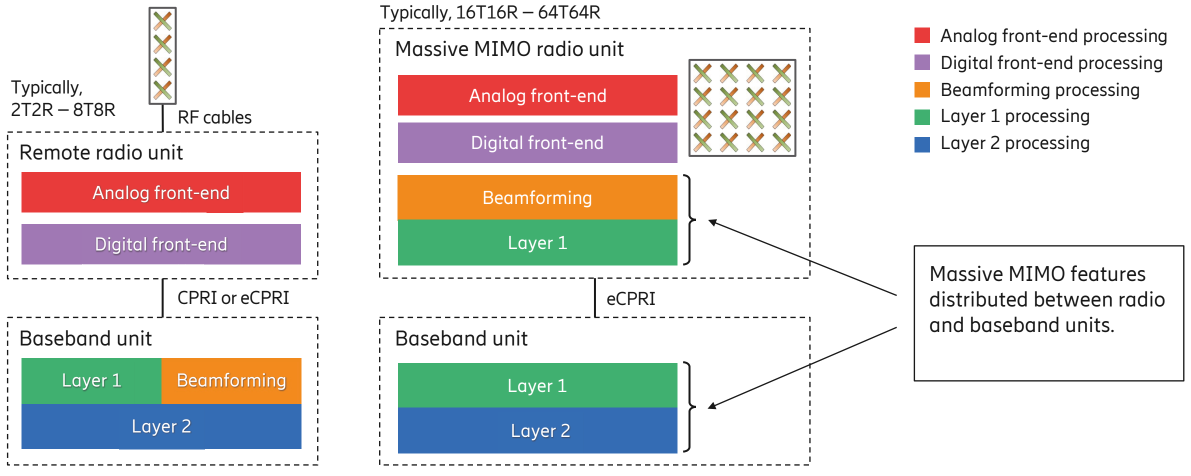
Ericsson Radio System (ERS) is designed to meet the requirements of past, current and future mobile networks. It provides a versatile toolbox that includes Radio, RAN Compute, Radio System Software/Solutions, Transport, Antenna Systems and Radio Site System. Here, three key components of a radio solution, namely baseband*, radio units, and antennas, are in focus.

This chapter is outlined as follows. The first section describes differences in architecture between Massive MIMO and conventional RRUs. Thereafter follows two sections describing the Massive MIMO radio solution and Massive MIMO feature solution, respectively. Discussions on how the radio and feature portfolios can be segmented to address the network requirements and constraints are also provided

* In Ericsson implementation, the baseband functionality are implemented either as a Baseband unit, shown in the figure, or as a cloud-based RAN compute solution.

3.1. Massive MIMO solutions – Solution architectures

Different functional distribution Remote radio units vs. Massive MIMO radio units

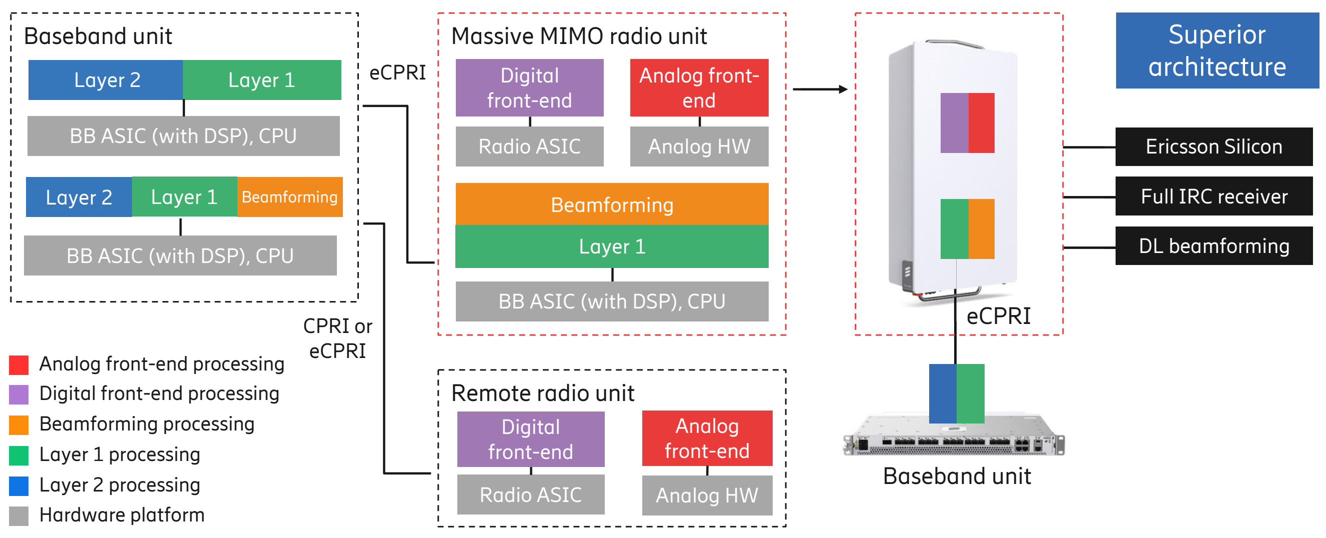


Massive MIMO includes hardware and software, as described on page [Ch. 1, p. 10]. Functional distribution in Massive MIMO solutions is somewhat different than in conventional solutions. The reason for this is that a Massive MIMO radio unit has more radio chains and therefore a need for transferring substantially more information to the baseband unit. For example, a 64T64R Massive MIMO radio unit would require 16x the capacity of the Common Public Radio Interface (CPRI) link than a 4T4R radio unit. However, even though the functional distribution is different, the same Ericsson baseband solution can handle RRUs and Massive MIMO radio units.

There are two ways of solving the interface issue: either provide more CPRI capacity to the baseband or move functionality to the radio to aggregate information already in the radio. The solution chosen by Ericsson is to move some parts of the Massive MIMO functionality to the radio unit, which is made possible by using the eCPRI interface (a Common Public Radio Interface with support for 5G fronthaul). eCPRI is an evolution of CPRI that uses the packet-based Ethernet protocol and enables more flexibility and data compression, such as moving the beamforming function to the radio unit. The use of Ethernet enables switched fronthaul which is a reason for also the RRUs to use eCPRI. The data compression ratio that eCPRI provides is on a high level related to the fact that data rate over eCPRI typically scales with the user traffic bitrates, in contrast to CPRI, where the data rate is fixed and scales with the number of antennas. The reduction of front haul bitrates can be to the order of ten.

In addition, to avoid extremely high bit-rate links, there are also benefits of reduced total processing time, such as processing in radio, transmission time from radio to baseband, processing in baseband, transmission time back to the radio, processing in radio, by locating functionality in the radio. Certain time-critical and transmission-volume-consuming data handling is in the radio's uplink receiver (also called the Uplink Booster, [5]), which greatly benefits from having access to instantaneous channel information.

Design efficiency enabled by Ericsson Silicon



The digital functions of the baseband units and the radio units are implemented by software running on application-specific integrated circuits (ASIC). Compared to conventional radio units, mid-band Massive MIMO radio units have the same type of ASICs as the baseband units, including the software that implements a subset of the baseband functions that are implemented in the baseband unit for RRUs.

In addition to the baseband functions, the Massive MIMO radio also includes ASICs to perform digital radio processing. Since the Massive MIMO radio has many more radio chains than an RRU there is considerably more processing needed. Thus, there is more hardware and more functionality in the Massive MIMO radio compared with an RRU. The Massive MIMO radio is a more complex but also a more capable product.

The trend with more radio chains and more total configured bandwidth (TCBW) is driving the processing requirements. To cope with this increase in complexity and to keep the power consumption on a manageable level, Ericsson has developed a set of custom ASICs, Ericsson Silicon. Inhouse development makes it possible to target the design of the silicon for best-in-class algorithms.

From a portfolio perspective, the radio portfolio comprises the hardware of the Massive MIMO radios (AIR products), and the software features, Massive MIMO features and other common features, comprise the software, which is located both in the Baseband and the Radio. The remainder of the chapter will elaborate on how the radio portfolio and the Massive MIMO features can be designed to cover all requirements in the 5G network.

3.2. Massive MIMO solutions – Massive MIMO radio

Ericsson radio product segmentation

Capacity

Targeting urban (incl high rise) to suburban deployments. Mast and rooftop deployment.

Capacity (bandwidth, Multi-band) and EIRP optimized

Compact

Targeting urban and suburban deployments. Rooftop and wall mounted deployment.

Footprint optimized capacity, adressing cost, ease of deployment and energy consumption

Coverage

Targeting urban to rural deployments with large inter-site distance and non-high rise. Mast and Rooftop deployment.

EIRP and bandwidth (Multi-Operator RAN) optimized coverage, adressing TCO and sustainability

The radio portfolio of Ericsson has been categorized into three product segments, capacity, compact and coverage, where the products in the respective categories target different purposes and deployment scenarios, and thereby have different characteristics. The most important design parameters used for defining these segments are:

Radio parameters such as number of radio chains, output power, number of frequency bands and total configured bandwidth (TCBW)

ntenna parameters such as antenna size, number of antenna elements and subarray structure cost and building practice parameters such as size and weight.

The first category, Capacity, targets the most capacity demanding sites and provides superior performance in all deployments ranging from dense urban high-rise to rural. It supports multiple frequency bands, large bandwidth, high EIRP, and many radio chains facilitating superior horizontal and vertical domain beamforming.

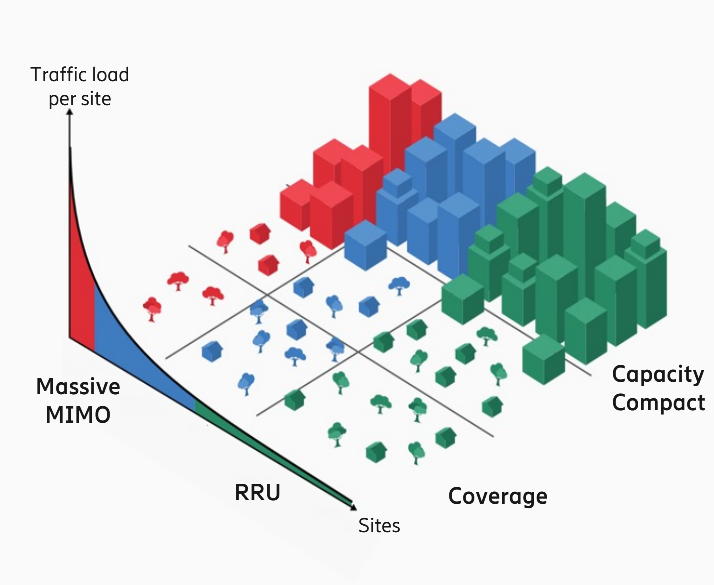
The second category, Coverage, targets deployments with larger inter-site distance like suburban or rural. It has high EIRP (high output power and many antenna elements) to ensure coverage and typically fewer radio chains to lower total cost of ownership (TCO) compared to the Capacity segment. Fewer radio chains is well motivated as targeted deployments have less spread of users in the vertical domain, hence less use of vertical domain beamforming. Note also that large antennas are desired to maximize both downlink and uplink coverage.

The third category, Compact, targets deployments where ease of deployment is important. It prioritizes TCO and ensures that mechanical properties are in line with common site constraints, such as size and weight. Still, the products can provide substantially performance gains compared to conventional RRUs in many deployment scenarios.

It needs to be highlighted that capacity, coverage and compact refer here to product segments, and should not be mixed up with network properties such as coverage, capacity and user throughput. That is, there is not a one-to-one mapping between the segment names and 'corresponding' network properties. For example, products from the capacity segment will provide superior overall performance including coverage, capacity and user throughput.

See also Start simple with the Ericsson Radio System, [6] for a general introduction of the Ericsson radio portfolio.

Portfolio mapping in network deployment



Capacity
Optimized for capacity

Compact
Optimized for site deployment

Coverage
Optimized for longer ISD

The aim of the portfolio is that the product segments shall meet the performance requirements on site level in different parts of the network given relevant constraints. The radio product segments introduced in the previous slide can be roughly mapped to site classes introduced on [Ch. 2, p. 29] and shown again here in the figure, where one axis shows the load distribution over sites, and the other axis represents different deployment environments.

In the deployment environments dimension, the Coverage segment is primarily intended to address sites in rural and sparse suburban environments and the Capacity segment addresses sites in the suburban, urban and dense urban environments. The Compact segment is a complement to the portfolio for sites where there are restrictions on deployment, e.g. size, weight, wind load, visual impact, and/or cost efficiency.

In the traffic load dimension, Massive MIMO addresses primarily sites where the traffic is medium to high and RRU addresses sites where the traffic is low and is expected to remain low throughout the time the RRU is intended to be deployed.

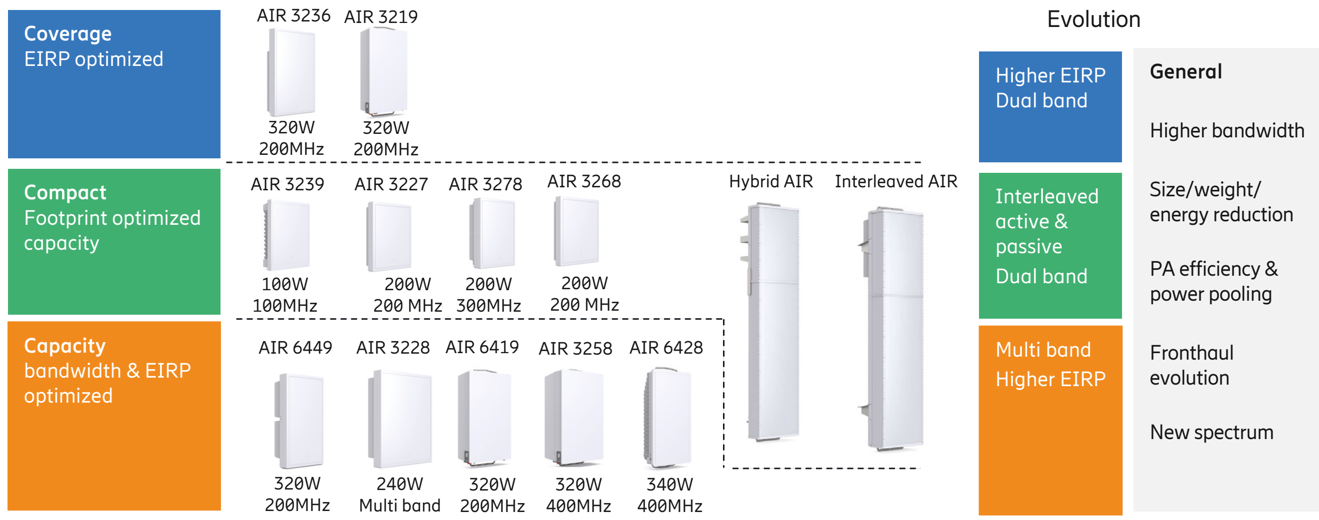
There is no clear-cut classification, so the mapping acts as a high-level guide. Products targeting highly loaded sites need to provide superior capacity meaning many frequency bands, large bandwidths and often excellent multi-antenna support. In these deployments, Massive MIMO typically outperforms RRU. Products targeting site with low load can use slightly less capable products but instead excel in terms of TCO and ease of deployment. In this case, RRU might be a cost-efficient alternative to Massive MIMO.

Similarly, products in dense urban high rise deployments benefit from highly capable beamforming in both horizontal and vertical domain, many Ts and Rs, whereas products in rural deployments need less vertical domain beamforming, fewer Ts and Rs. Products in rural deployments also need to have large EIRP (large antennas and higher power spectral density) to achieve good coverage.

One thing to note is that sites with all different load levels can occur in all deployment scenarios, e.g. highly loaded sites in rural deployments. A highly loaded site in rural requires a slightly different Massive MIMO product compared to dense urban high-rise, e.g. lower degree of vertical domain beamforming in rural compared to dense urban high-rise.

Depending on market requirements, the product availability is different for different bands and hence all products are not available for all frequency bands. The portfolio will however be more complete over time as market requirements develop.

M-MIMO portfolio TDD mid-band



A summary of the Ericsson M-MIMO portfolio, as of 2021H2, is presented below.

One product in the Capacity segment, is the M-MIMO model AIR 6419 which provides high EIRP (320 W, 192 Antenna Element) and high beam steering capability in vertical domain (64 TRX). The weight is below 19 kg (for B42 band variant). This is a newer and more improved version of the previous variant in this capacity segment, AIR 6449, with its 37 kg.

In the Compact segment the model AIR 3268 is has been optimized with respect to its mechanical properties and TCO for the operators with constrained site deployment in focus. The output power is 200 W, the footprint is very small and the weight is only 12 kg. The product is a good compromise between high capacity and small size which is useful in many network deployments around the world within urban and sub-urban environments. The cost effectiveness, TCO per capacity, is most attractive for many service providers. As with many products, the latest product variants deliver similar or better performance compared to earlier variants. In this case, the newer AIR 3268 has the same radio performance as the earlier version, AIR 3227, but the weight is reduced from 22 kg down to 12kg. In this product segment, there are also the combination of active and passive antennas. There are antennas available that include both M-MIMO for mid-band deployment and many FDD bands for capacity expansions (new bands, support for FDD 4T4R)

In the Coverage segment the main focus is to provide high coverage in wider areas with limited capacity needs. These products have been optimized to provide a lower TCO in those deployment scenarios. The high requirements on coverage require high EIRP (320 W total output power, 192 Antenna Element). Due to the lower demand for capacity and lower demand for beamforming in the vertical dimension and hence 32T32R rather than 64T64R is sufficient. To manage the network optimization, Remote Electrical Tilt, RET, is supported. In total, the Coverage solution offers high coverage, but to lower price and reduced energy consumption compared with similar product in Capacity segment (total power, antenna size).

To the right in the picture, examples of how the product segments will evolve over time are provided. Different product segments have some specific segment properties for that segment (i.e Compact segment: focus on evolving the interleaved active and passive products and add dual-band products. Also indicated are some properties expected to be common improvements over all product segments (i.e. mechanical reduction, energy consumption improvements). The purpose with this information is to show how the portfolio will evolve over the next years to come.

Comprehensive Radio Portfolio

RRU Massive MIMO Street and indoor

The picture above illustrates the entire radio portfolio (as of 2021H2) and products that are available for TDD mid-band spectrum range are marked in yellow. For further descriptions on how to build radios and specifically Massive MIMO, see [7] and [8].

When service providers are considering to deploy mid-band spectrum, new radios are needed. Massive MIMO products provide higher capacity than RRUs. They also provide better coverage which is required to get mid-band coverage using the existing site grid, [Ch. 2, p. 23]*. For site reuse in macro network deployment with limited capacity demand and modest capacity growth, conventional RRU, with 2TRX up to 8 TRX that are connected to passive antennas, is a good complement to Massive MIMO. Similar as for Massive MIMO, RRUs are categorized into the product segments capacity, coverage and compact. The RRU with 8T8R is found in the Capacity segment and RRU 2T2R and 4T4R in the Compact segments. These radios need to be connected to passive antenna in order to get a complete radio site solution.

When considering other deployments than traditional macro networks, there are also the micro and street segment containing radios and AIR products (built in antenna) to be used 'below rooftop deployment' and in special hotspot deployments. For dedicated indoor deployments, there are variants of Ericsson Radio Dot System (RDS), where single or multi-band RDS can be used for efficient mid-band indoor deployment.

When finding a suitable mid-band solution, the entire radio product portfolio should be considered. The scenarios where Massive MIMO products will be the best long-term solution will be discussed in Chapter 4.

* The complete Ericsson Radio System portfolio has a lot of Massive MIMO product variants. The variants come from market requests that define band variants and properties per deployment scenarios and are grouped into our product segments. The result of market-driven product requirements is band combinations that do not have all different product segments realized. Therefore, it is important that service providers engage with Ericsson and share their plans and requests to secure that Ericsson will plan for new product variant in time for the demand.

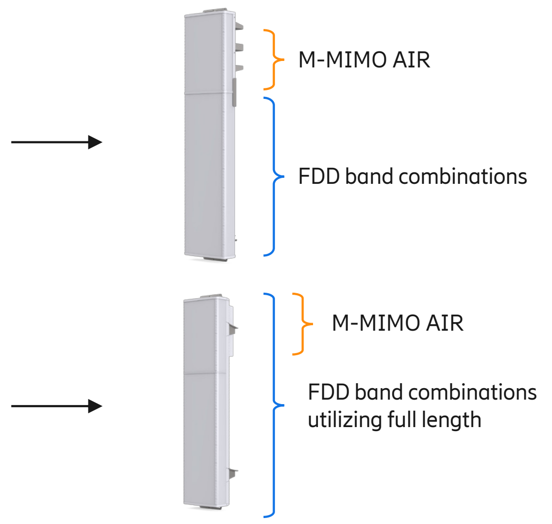
Add mid-band with M-MIMO and remain with only one antenna after expansion

Hybrid AIR

- Combines low band antennas with a Massive MIMO AIR on top, utilizing the wide range of passive antenna modules from Ericsson Antenna Systems (EAS) portfolio
- Option to use variants of 32TRx and 64TRx radio modules from the Ericsson portfolio

Interleaved AIR

- A single antenna form factor that uses interleaving technology for maintained low band antenna properties for same height of antenna
- Enable future expansion to M-MIMO mid-band within the same site lease contract



The products presented on the previous page assume that a new product is introduced to cover the requirements on mid-band without affecting the equipment already existing on site.

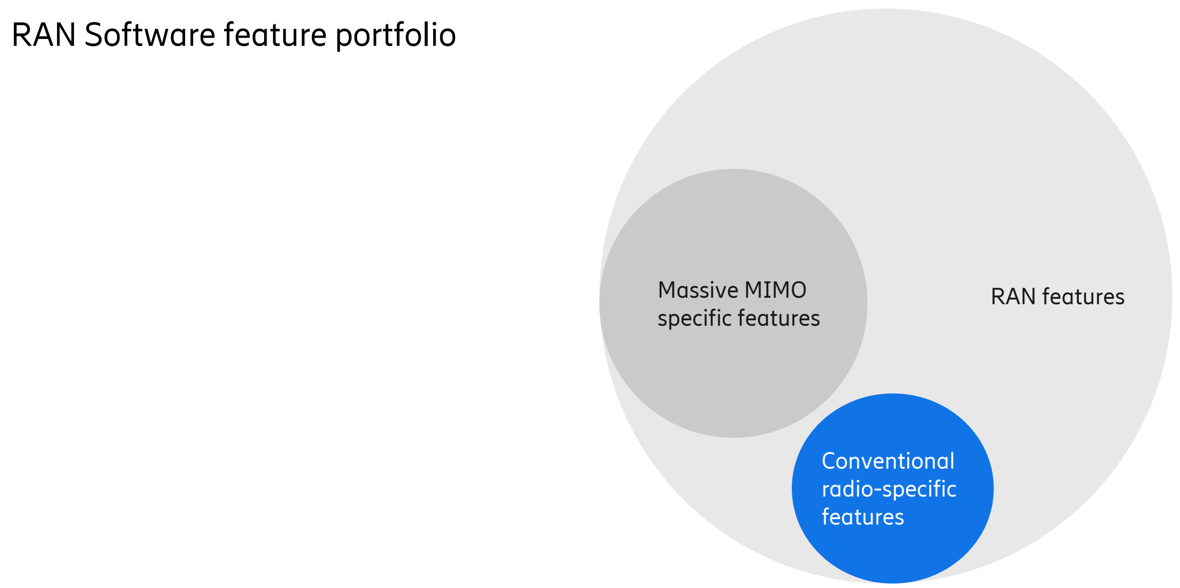
In many deployments it would be highly attractive to be able to have a product that adds Massive MIMO on mid-band, and at the same time modernizes existing FDD bands. Hence, one product that replaces existing products and adds Massive MIMO mid-band support. Ericsson has two product variants to support the vast demand of different product characteristics on FDD antennas on global scale. In the product called Hybrid AIR, there is a modular building practice that maximize the selection of Massive MIMO AIR products in combination with large number of passive antenna models for FDD. A more integrated solution is also designed where the full length of the total antenna product can be used both for mid-band TDD and the low band FDD parts. This product is called Interleaved AIR where the top part of the FDD antenna is transparent towards the TDD Massive MIMO radio embedded behind the low band part of FDD antennas.

The main advantages of the combination of an active/passive radio solution as described here is of course that this will allow service providers to simplify the renegotiations on mid-band expansion on the site where the final solution will only contain one antenna radome supporting both FDD modernization, additions of new bands and addition of mid-band TDD with a Massive MIMO solution.

The product of choice for service providers depends on cost and performance, but both product variants are key solutions for service providers that need expansion on site where only one antenna per sector is allowed, or for other reasons, where only one antenna is preferred.

3.3. Massive MIMO solutions – Massive MIMO features

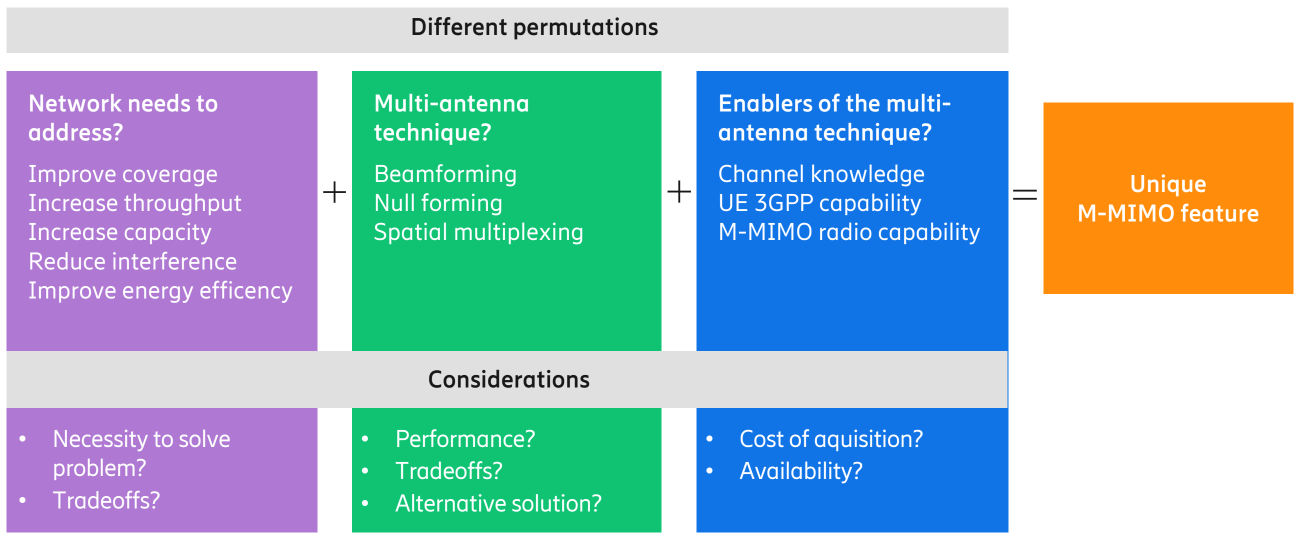
What is a Massive MIMO software feature?



The RAN software portfolio consists of many different kind of features. Generally, these features are agnostic to the radio being used. This means that they can be used on conventional as well as Massive MIMO radios. However, Massive MIMO adds the capability to utilize a new range of software features that are Massive-MIMO specific (or features which substantially increase in benefit when using a Massive MIMO), denoted Massive MIMO (specific) features, which will be further elaborated on in this chapter.

For Ericsson products, most software features can be used with most radios presented in [Ch. 3.2, p. 40]. Some features, like some Massive-MIMO-specific features, may require hardware support and can only be used with Massive MIMO products.

What is a Massive MIMO software feature?



A Massive MIMO feature can be described by a combination of three different categories

- Which network requirement does the Massive MIMO feature intend to address? For a list of requirements, see [Ch. 2].
- What channel knowledge is available?
- Which multi-antenna technique, [Ch. 1, p. 12] (or combination of) can be applied using the channel knowledge (2) to solve the problem in (1).

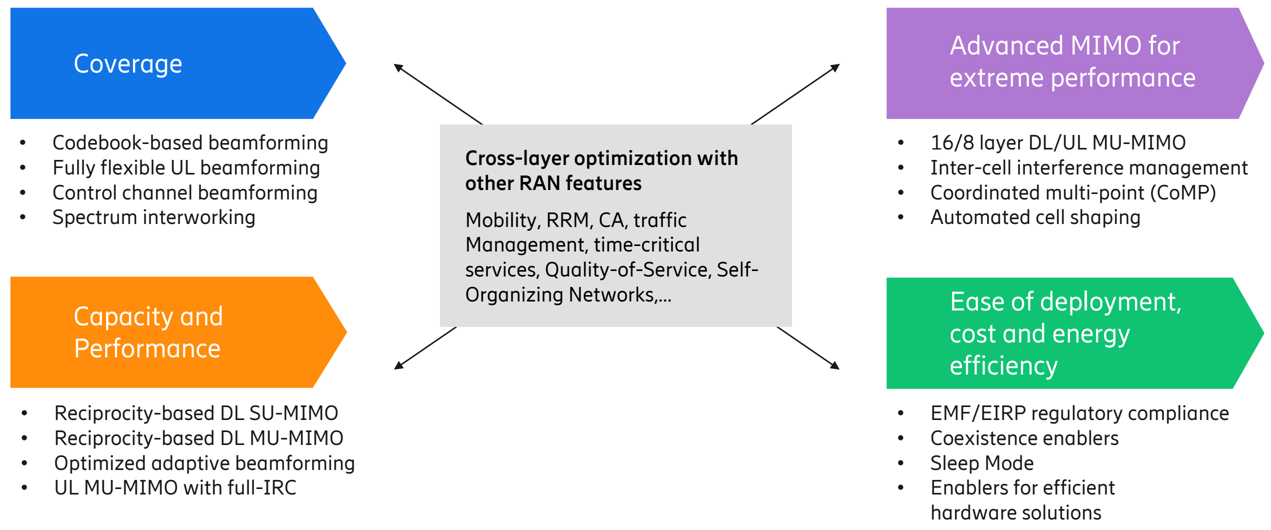
Different permutations of these three categories will yield unique Massive MIMO features – potentially with different tradeoffs and applicability to different conditions.

Firstly, the problems the feature should solve must be determined, e.g. improve coverage, increase capacity or increase throughput. Potentially, one feature can solve multiple problems at once, while in other cases, there may be trade-offs to be made. For instance, in order to improve energy efficiency, the capacity may be negatively affected. Thus, it is important to assess what network requirements are important for a certain cell at a certain time. For instance, during off-peak hours, the capacity demand in the cell is low and so it may be acceptable or even desirable to apply a feature that sacrifices capacity to improve energy efficiency.

All Massive MIMO features can essentially be described by applying a combination of the three basic multi-antenna techniques, beamforming, null forming and spatial multiplexing, to some physical channel or signal, using some available channel knowledge in order to solve a certain problem. This may sound simple, but there are multiple things to consider which results in a wide variety of potential features. First, the network requirements to acquire channel knowledge in order to know how to perform beamforming, null forming or spatial multiplexing. This can be achieved in several ways, but it is important to understand that there is always a cost associated with acquiring channel state information (CSI), for example, increased overhead.

There is also a problem of availability of CSI. Different sounding and feedback methods are available in the 3GPP standard and different UEs may have different capabilities and support different modes. Hence, the network must support several Massive MIMO features in parallel. Even if a UE supports a certain CSI mode, that CSI may not be available at a certain instance in time. For instance, when a UE first connects to a cell, no channel information is generally available and measurement or sounding configurations will need to be set up via RRC configurations implying that there is a lead time before such CSI is available to the network. Different sets of MIMO features are needed when no or limited CSI is available compared to when CSI is available.

Categories of Massive MIMO features



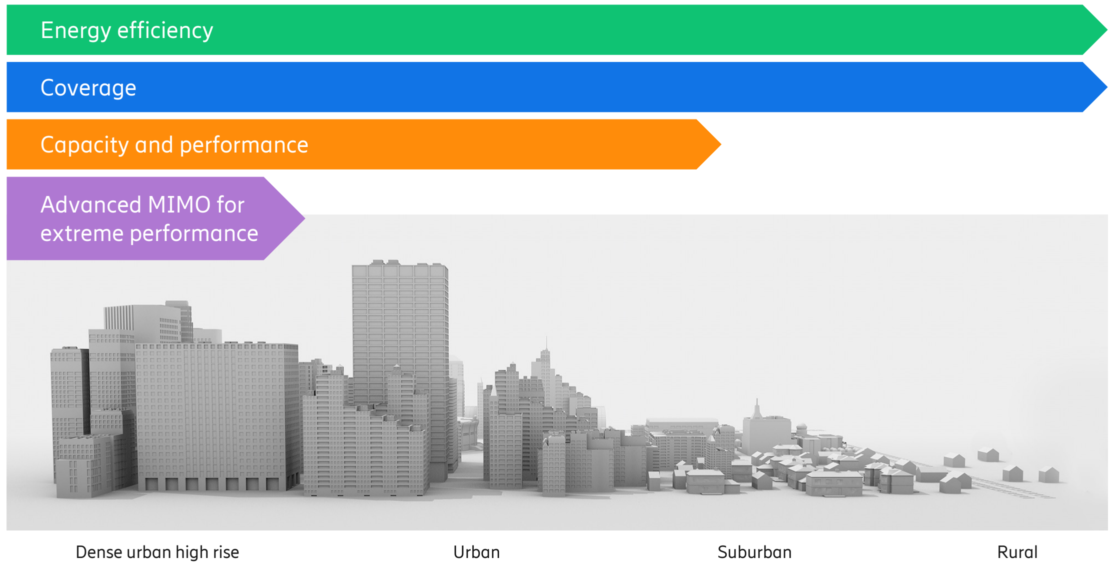
The Massive MIMO features can be put in four different categories, illustrated with examples in the figure above. These feature categories will be further elaborated in the coming slides.

- Coverage
- Capacity and performance
- Advanced MIMO feature for extreme performance
- Ease of deployment, cost efficiency, and energy efficiency

Coverage features improve either the access coverage or the data channel coverage typically by applying beamforming to boost the signal strength. This can be done both in uplink and downlink. Capacity and performance features on the other hand typically aim to improve peak user throughput and/or capacity of the data channels and generally use a combination of beamforming, null forming and spatial multiplexing. Advanced MIMO features target more specific deployments or scenarios which may have an extreme capacity need, very high load or some other special characteristics. For these cases, more advanced features such as inter-cell coordination features or higher-order spatial multiplexing can be applied (which would not be required or provide additional benefit for e.g. less capacity-demanding cells).

Features in the category Ease of deployment, cost efficiency, and energy efficiency address non-performance related aspects [Ch. 2, p. 26] such as fulfilling regulatory requirements on electromagnetic field (EMF) emissions, which may require dynamically adapting the power level depending on the beamforming characteristics of the cell or reducing the amount of beamforming that is performed in order to conserve energy.

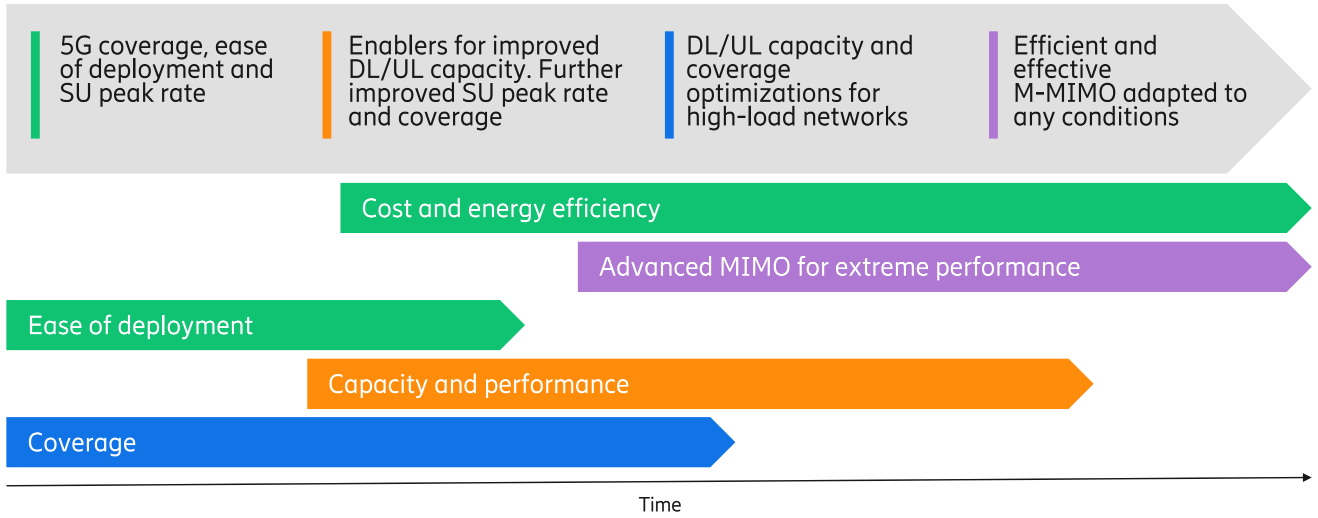
Where is a solution most useful?



Different categories of Massive MIMO features may be more applicable to some parts of the network than others, depending mainly on capacity demand and deployment density. This is like how different hardware solutions apply to different parts of the network as discussed in [Ch. 3.2, p. 38]. Capacity and performance features are generally useful for a large part of the network, except for less capacity-demanding rural sites. Coverage enhancement features are most useful in rural and suburban deployments with larger inter-site distances but are still needed for dense urban deployments due to e.g. deep indoor users. The Advanced MIMO features category targets specific deployments with very high capacity requirements such as dense urban deployments with short ISD, or capacity hotspots such as stadiums, shopping centers, or trains stations.

Energy efficiency features are applicable to all deployment scenarios as even very high-capacity demanding sites may have a traffic pattern that varies over 24 hours (e.g. sites around office buildings which may be empty during evenings but full during daytime).

Feature evolution of Ericsson Massive MIMO over time



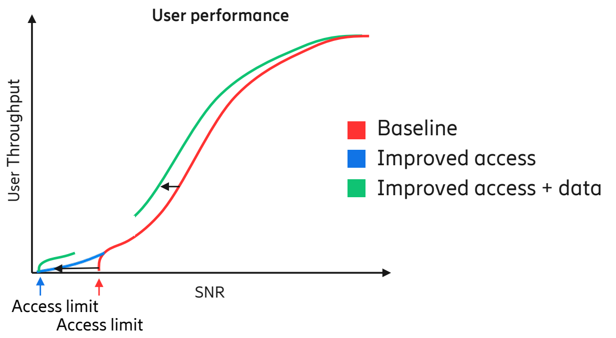
As the different categories of software features address different network requirements, it makes sense to deploy features step-wise depending on which network need is most urgent at different times of the network evolution. For instance, when first rolling out the mid-band spectrum, focus should be on ease of deployment features and coverage to make sure the Massive MIMO radios can be deployed in various conditions and that the mid-band coverage can be extended so that as many users as possible can access the mid-band spectrum.

Focus should shift to features which improve single-user performance and throughput, and as the traffic network load gradually increases on mid-band, capacity features need to be introduced in order to cater that demand, eventually shifting focus to more advanced MIMO features which can provide extreme capacity and performance for high-load scenarios.

Coverage solutions

Access coverage
- Initial access
- Control channel

Data channel (App) coverage



Ways to improve coverage
- Improving SNR
- Reducing interference (improving SINR)
- Improving detection reliability given an SNR
- Spectrum interworking

Access might have limited CSI → restricted Bf required

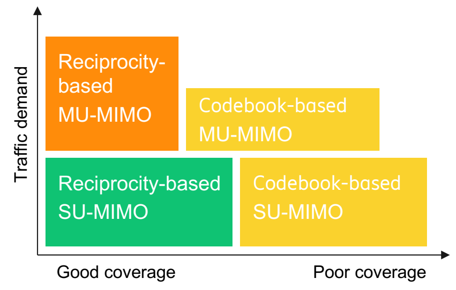
When addressing coverage enhancements, it is important to differentiate between access coverage between access and data channel coverage. Access coverage is the coverage of the cell-defining signals and messages in the initial access procedure as well as the control channel coverage, which is the coverage of the supporting control channels needed for scheduling and providing feedback for the data channels, while data channel coverage is the coverage of the data channel itself. Access coverage should be at least as good as data channel coverage. However, having access coverage that surpasses data channel coverage does not necessarily result in actual improved downlink or uplink coverage, but rather it means that some radio network resources are wasted.

Massive MIMO has the potential to further increase the coverage by utilizing beamforming and interference reduction capabilities of the Massive MIMO to improve the signal-to-interference-and-noise-ratio (SINR) on the receiver side. A Massive MIMO radio has the possibility to use beamforming to create narrow beams that maximize the beamforming gain in a certain direction, but it also has the possibility to create wider beams (and thus lower antenna gain). The benefit with using wider beams is that fewer beams are required to cover the entire intended cell service area, which in some cases may be desirable, on the other hand the downside is that the experienced beamforming gain is lower.

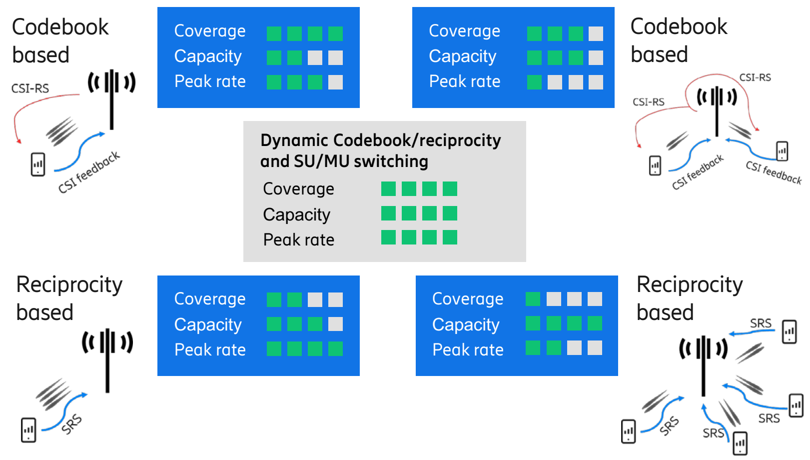
Using fully flexible beamforming, with many narrow candidate beams covering the entire cell service area whenever possible is beneficial in both downlink and uplink in order to maximize the beamforming gain and thus improving the coverage. However, there are some exceptions where restricted beamforming or even a single wide beam may be unavoidable or more suitable, e.g. when additional coverage improvement is not necessary or when is not CSI available.

Capacity solutions

In which scenario is each scheme most useful?



Overview of features and benefit for DL data channel

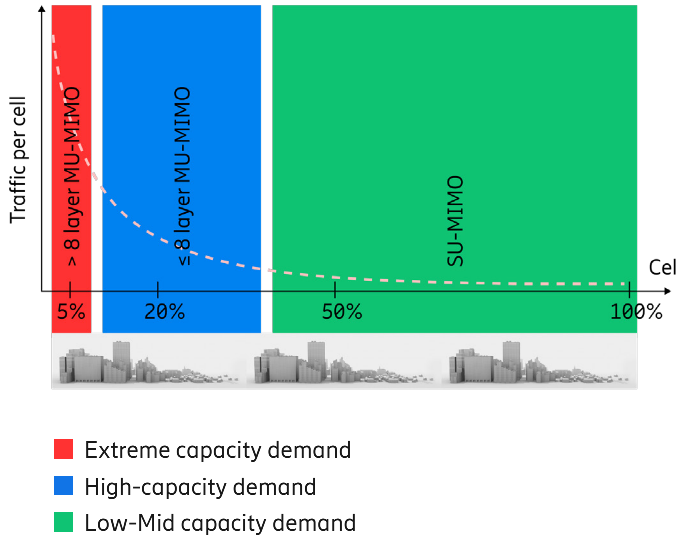


On a high level, the basic features for downlink performance and capacity in a Massive MIMO solution can be classified firstly according to the codebook-based or reciprocity-based channel information used and secondly whether single-user or multi-user MIMO is applied. This results in four distinct high level downlink Massive MIMO features. Typically, these four features would be activated in a cell, but which transmission scheme is used at a certain point in time would depend on for example traffic conditions and UE characteristics. There are different options for how reciprocity-based and codebook-based SU/MU-MIMO can be implemented, both from what is available in the 3GPP standard and from proprietary algorithm perspective, however on a high-level the various options still share the same characteristics with respect to use case and performance.

By comparing the four Massive MIMO features with respect to the network KPIs of interest, coverage, capacity and single-user peak rate, they exhibit different strengths and weaknesses. Codebook-based beamforming has an advantage in coverage over reciprocity-based beamforming. Similarly, SU-MIMO has a coverage advantage over MU-MIMO as the available transmit power needs to be split between multiple users in the latter case. When it comes to capacity, reciprocity-based MU-MIMO has the highest potential. However, since the coverage of MU-MIMO is limited, the maximum benefit only occurs in dense deployments with small cell sizes so that most of the users in the cell have coverage. To fully utilize the potential of a Massive MIMO solution, it is necessary to dynamically switch between the four above schemes so that coverage, capacity and peak rate jointly can be maximized, which is also how the typical Massive MIMO solutions are designed.

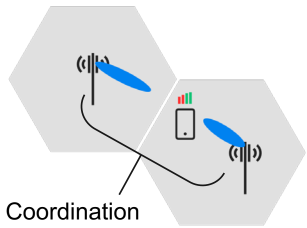
Advanced MIMO for extreme performance

High-order MU-MIMO



Cell-shaping

Coordinated multi-point



Advanced MIMO features target more specific deployments or scenarios which may have an extreme capacity need, very high load or some other special characteristics.

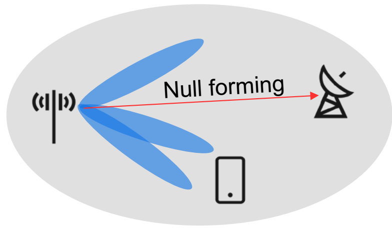
For instance, while MU-MIMO is a key feature to provide capacity in a Massive MIMO solution, the number of MU-MIMO layers required to fulfill the capacity demand is generally quite low in practice. In fact, in a large portion of the network, the capacity demand can be fulfilled with only SU-MIMO. Even for cells with high-capacity demand, it is generally sufficient with 8 downlink layers and 4 uplink layers to get most of the MU-MIMO gains with MBB traffic. However, the high-capacity Massive MIMO radios are dimensioned to support 16 downlink layers and 8 uplink layers, since this could be required in certain cell sites with extreme capacity demands, such as traffic hotspots or stadium/arena deployments with a lot of people in a small area.

Another example of advanced Massive MIMO feature are the Coordinated Multi-Point (CoMP) features wherein multiple nodes, on the same or different sites, coordinate their transmission/reception in order to mitigation interference or boost the signal energy.

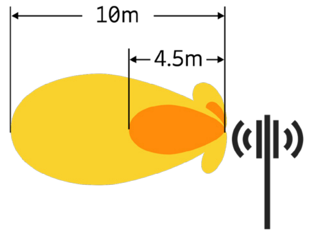
Another example is cell shaping that can be used to solve various problems. For example, the cell-border with its interference impairments can be moved to avoid being in the middle of a location where there are typically many simultaneously connected UEs, known as a hotspot, as illustrated in the upper-right figure. An example of such a hotspot is a train station. Cell shaping also offers the possibility for load balancing among cells by moving UEs from one cell to another by altering the uptake areas of cells.

Ease of deployment, cost and energy efficiency

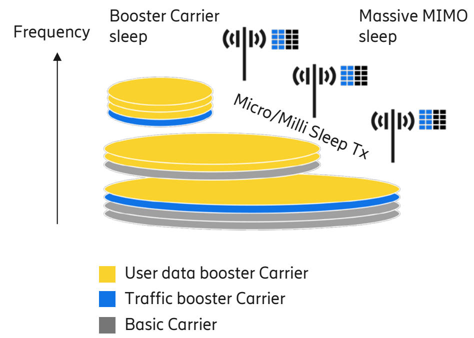
Directional interference mitigation



Electromagnetic field (EMF) exposure handling



Energy efficiency M-MIMO features



Ease of deployment features can be used to facilitate deployments in some specific cases. Some mid-band TDD bands in certain regions have spectrum coexistence requirements with regulatory limitations on generated interference towards other users of the same spectrum, such as satellite earth station receivers, which may be operating either in-band or on adjacent band.

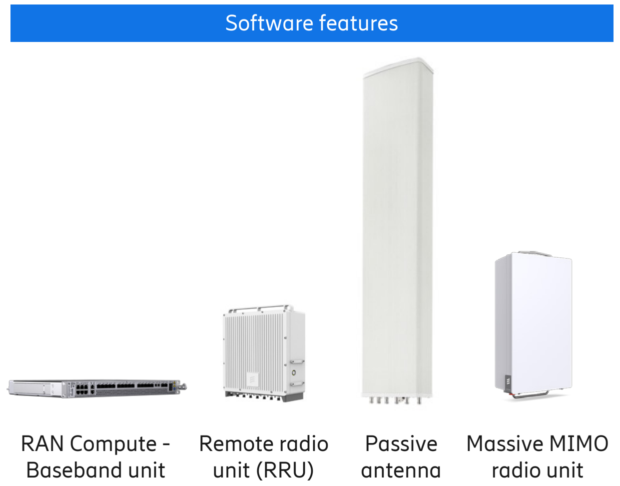
All regulatory requirements need to be fulfilled and it is imperative to have features that ensures this. One example is requirements on electromagnetic field (EMF) emissions, which may require dynamic adaptation of the power level depending on the beamforming characteristics of the cell.

Massive MIMO solution provides high capacity and performance which is useful for handling high traffic load scenarios such as busy hours and occasional peak loads. In most traffic load scenarios, even in the most loaded areas, there are still periods with low traffic load where the available capacity is not required. During these periods, power can be saved by deactivating capacity that is not needed. By switching off radio unit components power consumption is reduced.

3.4. Massive MIMO solutions – Summary

Summary

- In the Massive MIMO radio, the beamforming processing capabilities are integrated with the antenna to improve performance and reduce fronthaul capacity needs
- The Ericsson **Massive MIMO radio portfolio** is segmented to address different site requirements in terms of capacity, coverage and ease of deployment
- The Ericsson **Massive MIMO feature portfolio** is segmented to address requirements relating to capacity, coverage, extreme performance requirements and ease-of-deployment, cost- and energy efficiency



The Massive MIMO product architecture is different from conventional RRU. Some baseband functionality is moved from the baseband unit to the Massive MIMO radio to reduce the requirements to transfer very large amounts of data between the Massive MIMO radio and the baseband and to improve latency.

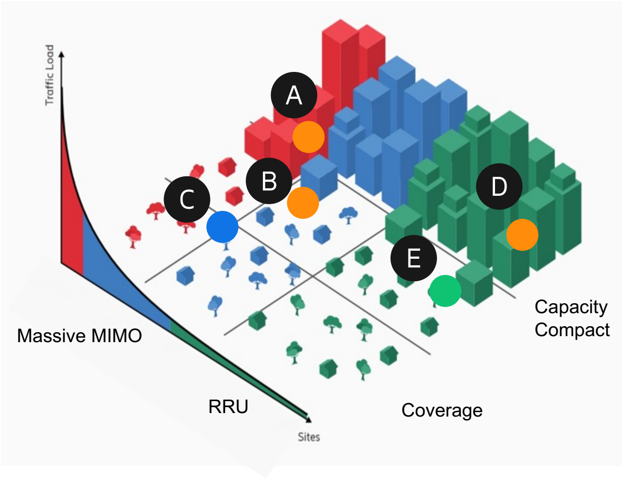
The radio portfolio is divided into three product segments, capacity, compact and coverage, to meet all performance requirements and constraints. The capacity segment targets the most capacity demanding sites and provides superior performance in all deployment ranging from dense urban high-rise to rural. The coverage targets environments with larger inter-site distance like suburban or rural and has high EIRP to ensure coverage, but typically fewer radio chains, and therefore lower TCO, compared to the Capacity segment. The compact targets sites where ease of deployment is important. It prioritizes TCO and ensures that mechanical properties, e.g. size and weight, comply with site constraints.

The feature portfolio is segmented into four classes, viz. coverage, capacity and performance, advanced MIMO for extreme performance, and ease of deployment, cost and energy efficiency, to address the same requirements and constraints as the radio products. The segmentation of the features is somewhat different than the segmentation of the radio portfolio, since features can be deployed on any site and at any time and independent of each other.

4. Guiding principles for how to select a radio solution

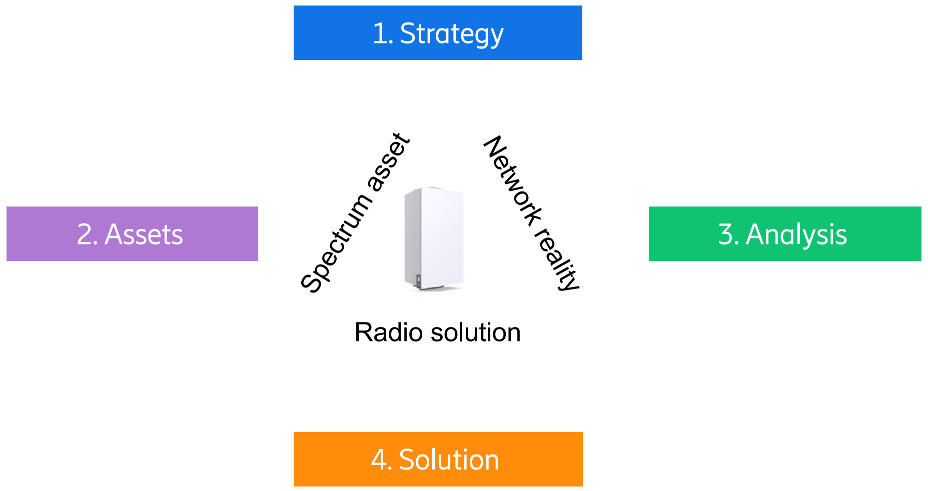
Outline – Guiding principles on how to select a radio solution

- Steps in the guiding principles and details of each step
- Guide for how to find cost-efficient solutions for different network conditions
- Applying the guide



The purpose of this chapter is to provide guiding principles for how to select cost-efficient radio solutions that meets the network requirements for typical site conditions. The guiding principles are built on four network components: strategy, assets, analysis and solutions. The steps are first outlined. The details in each step are elaborated and then the different steps are presented in a process flow. Finally, examples of how the guiding principles are applied to typical site deployments are provided.

Guiding principles on how to select a radio solution



The steps in the guiding principles are here outlined to show why they are needed and how they fit together. Since the service provider sets their own strategy objectives, some assumptions are made on what is typically stipulated with respect to network performance, ease of deployment and cost efficiency. These assumptions solely serve to illustrate how the guide can be applied.

1. It is assumed that the service provider has developed a network strategy for how to achieve their business objectives and what their mobile network shall do.
2. It is also assumed that the network strategy includes how to make use of the available network assets, including spectrum, sites and equipment.
3. Next, a network analysis is performed to find the performance requirements and constraints relevant for each site of the network over the investment cycle.
4. Finally, network solutions are made based on the available radio and feature portfolio that meet the performance requirements and deployment constraints from the analysis in step 3.

Details of each step are elaborated on the following pages.

The strategy defines service provider's long-term network objectives

1. Strategy

- Services to offer and where to offer
- Capacity and user experience to be provided
- How to use network- and financial assets

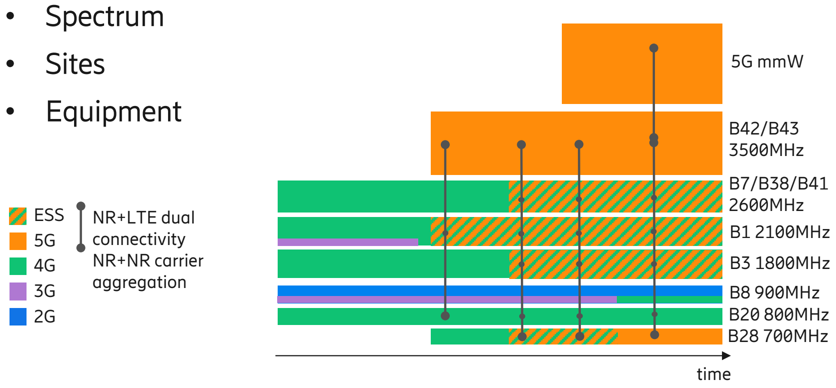
What ultimately decides how a service provider will evolve their network is the overall objectives of the service provider's business and the strategy for how to reach these objectives. Objectives and strategies can vary considerably. As part of these objectives is the position on the markets and how the service provider wants to be positioned in relation to other service providers in these markets, e.g. to be the leader or a follower. The strategy for how to reach these objectives is expected to include what services to offer, e.g. MBB, FWA, what service quality to offer, e.g. best in the market, and where in the network to offer these services, e.g. in cities over a certain size or essentially everywhere, and which price to offer. The strategy has implications on the network requirements in technical terms, i.e. coverage, capacity and user throughput, as outlined in Chapter 2. What also is expected from the strategy is how the service provider shall use its existing network, e.g. spectrum, sites and existing equipment, and financial assets, e.g. CAPEX and OPEX budgets, and hence the financial limits for the network investments.

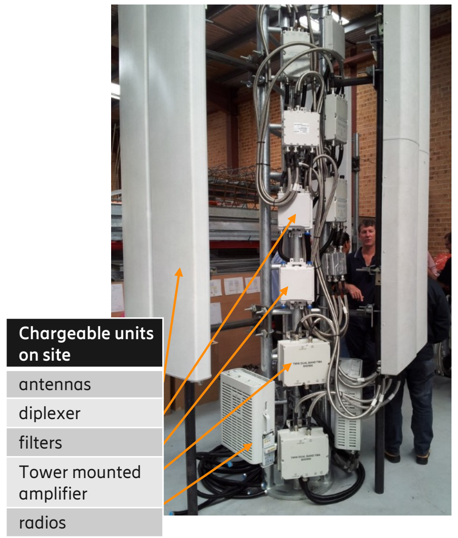
Other relevant examples of what could be contained in the business strategy is the use of different RATs (LTE, NR) and different frequency bands. In most cases the service provider is expected to provide ubiquitous coverage within a certain geographical area and sufficient capacity where and when needed within that area. The strategy may also provide guidance concerning which parts of the network are of strategic value, i.e. the sites identified as "most valuable sites", and hence where the performance should not be compromised. Another important guidance from the network strategy concerns the investment horizon, e.g. whether the network expansion is expected to last for a full investment period of 5-7 years or whether the plan is to make a radio site deployment for a shorter period, e.g. 2-3 years for capacity upgrade, and then re-evaluated. This will again guide the selection of radio solution for site expansion.

Best use of network assets

2. Assets

- Spectrum
- Sites
- Equipment





The most important radio network assets, the spectrum assets, the radio sites and site equipment, are crucial for several reasons.

Firstly, they are the tools and resources by which the service provider delivers the service to the end users. Secondly, they are also important financially, as they are tangible assets that carry an actual value.

Spectrum assets: Spectrum for all technologies (2G-5G) are typically expensive to acquire. The license to use spectrum is an investment the service provider must make, and the investment is expected to pay off over the licensing period. Hence, the usage of available spectrum is very important for all service providers. The existing bands, prior to launching 5G, are usually occupied for existing radio technologies 2G-4G. Over time, these bands are gradually migrated to the latest technology, i.e. 5G, to be used in the most efficient way. With 5G introduction, new frequency bands become available both in mid-band TDD and in the high-band spectrum (mmWave). The new 5G spectrum is often in the range of 100 MHz, which in many case will add a large capacity potential when used on site.

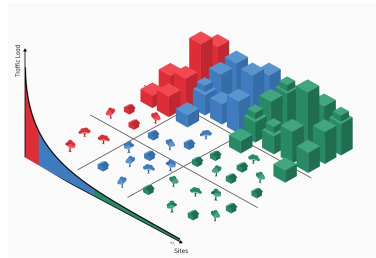
Radio sites: Radio sites are often difficult and expensive to acquire and to keep. Service providers typically want to maximize the use of the available sites before acquiring new sites. A very common situation today is that the radio sites have been sold to infrastructure companies and then leased back from the service provider. This is often financially attractive and it also separates the core business of the service provider and the expertise from companies focus on site infrastructure. In this new scenario site OPEX is very related to site lease contract agreements.

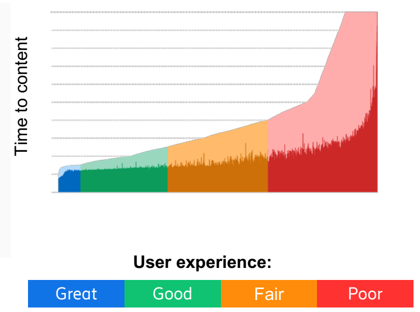
Radio site equipment: On existing radio sites, there are typically already equipment delivering services on 2G, 3G and 4G. The number of frequencies used on sites is driven from capacity requirements. For sites in high- and medium-traffic segments there are often up to 5-7 bands in operation already. Every band usually needs a dedicated radio. Therefore, for many sites where there are high-capacity demands, many radio units are needed. These radios are often connected to advanced multiband antennas. When deploying new equipment, it is often a good opportunity to also consider re-engineering the whole site and modernize existing equipment. Factors that affect this decision include the network requirements and constraints, but also the new product(s) to be deployed, the age of existing equipment and current lease conditions.

Network analysis to define site needs

3. Analysis

- Network load and traffic growth
- Load distribution in the network
- Service quality and evolution
- Deployment environment and performance requirements
- Site build constraints
- Total cost of ownership





Having considered the overall strategy with 5G deployment and the network assets, the next step is to gather and make use of information from the existing network. From such information, it is possible to assess the current load on each site and to predict what capacity is needed during the economic lifetime of a new investment, and specifically where and when the highest load levels occur. Some key considerations are discussed below.

Network load and traffic growth: The current network load and the predicted growth provide input on the general growth rate over time that the service provider must plan to handle.

Load distribution: As explained in Chapter 2 [Ch. 2, p. 28], the traffic growth and load per site is unevenly distributed, and thus some sites are highly loaded and require high-capacity solutions, while other sites are less loaded and can manage with low-capacity solutions. The high-capacity solutions are often more highly valued, since they carry a substantial amount of the total traffic.

Service quality: Time-to-content is usually a good measure of service quality, [Ch. 2, p. 21]. Long time-to-content is normally an indication of network congestion and that the site capacity is too low. This may be observed in highly loaded sites during busy hour. Quality during peak hours requires special consideration, particularly in markets where this is an important differentiator between service providers. It is also often the case that the requirements on time-to-content evolve over time.

Deployment environment and network performance requirements: Whether the considered site is deployed in dense urban, urban, suburban or rural environment is important information to understand how to best evolve it.

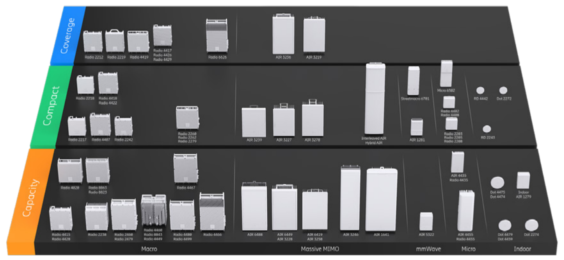
Site build constraints: The actual conditions on the sites are often setting constraints for how site expansion can be done. Some sites in urban areas may be using equipment on all frequency bands. It may therefore be difficult and/or expensive to add more equipment on-site. In other areas, there may be restrictions on visual impact. For example, there may be cultural buildings or objects which should not be impacted or there may be esthetic reasons for why antennas should not be notably visible.

TCO: Last, but certainly not least, considerations should be made concerning cost efficiency, or as suggested here TCO per capacity. Almost all service providers have requirements on return on investments. Hence, the benefits of the product deployed must motivate the costs spent. Exactly which requirements apply and how they are calculated are decided by the service provider.

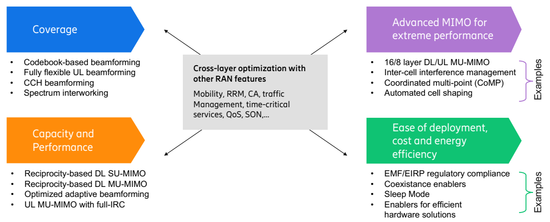
Select radio solution to meet the needs

4. Solution

- RRU or M-MIMO
- Product variants
- SW features



Categories of Massive MIMO features



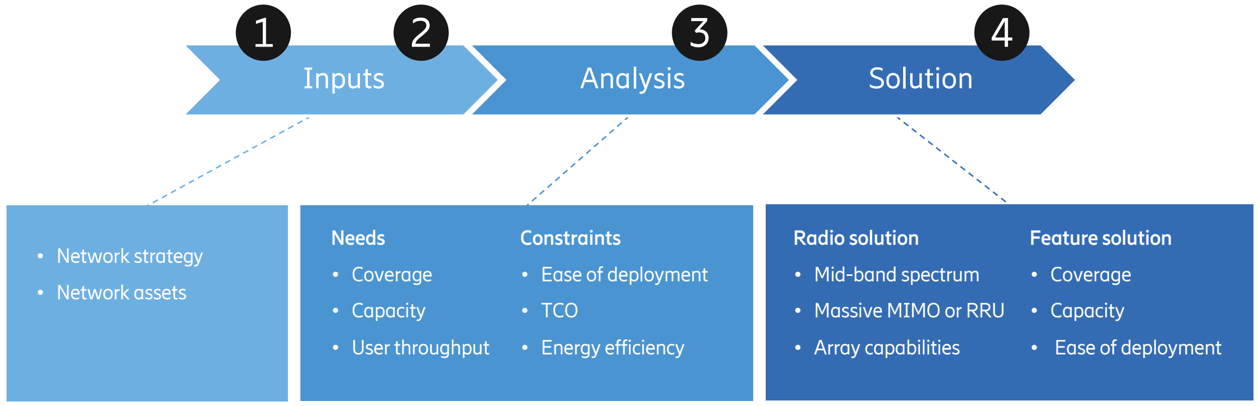
The last step in the process is to choose the site solution from the available product portfolio* that meets the requirements and constraints over the investment cycle following the analysis in step 3.

The radio portfolio (2021H2) is illustrated in [Ch. 3.2, p. 40] and it contains various Massive MIMO and RRU products from the segments, capacity, coverage and compact.

The software portfolio, illustrated in [Ch. 3.3, p. 45], contains a rich set of features addressing the different requirements and constraints derived in step 3 of the process. In contrast to the radio solution, feature solutions can be implemented more gradually as the requirements emerge. A radio product that is installed year one can thereby be improve in terms of capacity, coverage and quality some years later when additional software features are released, and new functionality has become available in the handsets. The important message here is to see the demand of certain functionality in relation to the network demand. As an example, the initial deployment will not need all software features for capacity enhancements since the pure addition of new bandwidth will be sufficient in the initial phase. However, after some years it will become very important to have the ability to add more capacity by simply activate new features to add capacity when needed.

*) Availability of different products on different bands will change over time, as market requirements develop. Since the evolution of the portfolio is going very quickly, it is however recommended to consider the products mentioned in this handbook as examples only and always consult the latest available portfolio.

The steps towards a radio solution



The steps to decide on a network solution are here summarized before giving some examples on how to apply the process. The steps outlined in the previous slides should only be seen as a 'guide'. The actual process used is of course decided by the network owner, usually the service provider.

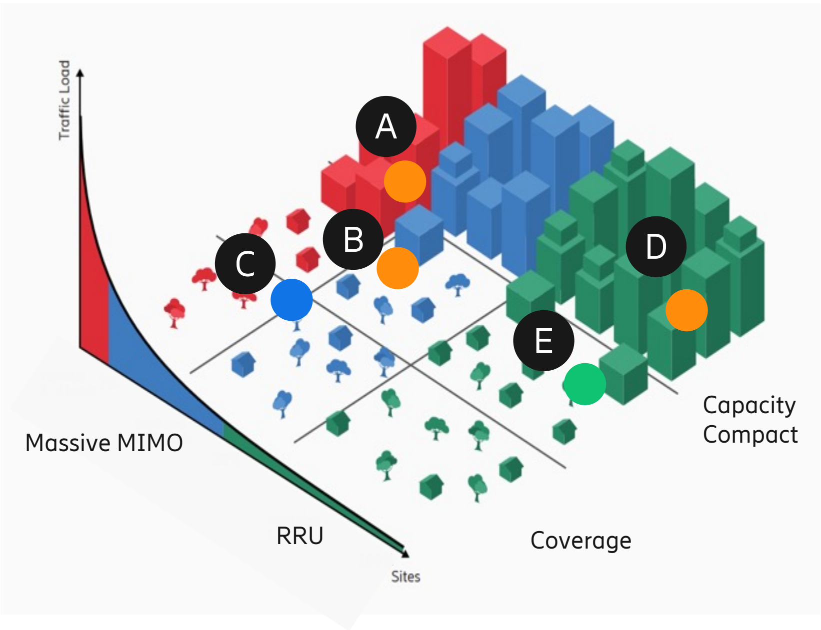
Potential solutions for how to meet the outlined network requirements are firstly evaluated. For a certain radio site, the network strategy, step 1, provides input concerning what services to deliver and what service quality to offer. These service requirements are mapped to requirements on coverage, capacity and user throughput as outlined in step 3. This mapping needs to be done on a site level considering aspects such as the actual traffic load, expectations on the traffic growth over time, and the deployment scenario. In this process, all means to meet the requirements should be considered. Thus, the use of spectrum and the possibility to add sites may also be considered. Generally, it is most cost-efficient to fully utilize the available sites before considering site densification. Therefore, site densification is not further discussed here. Assuming that the strategy is to deliver a 5G user experience to a major part of the subscribers, the most cost-efficient way to meet the corresponding requirements of higher user throughput and capacity is to make use of all available 5G spectrum on mid-band and make the product selection for this band according to the actual site requirements as a second step.

The next step is to select products depending on the relevant requirements and constraints. Having established a short list of product candidates, considerations must be made with respect to ease of deployment and cost. For example, size, volume, wind load and visual impact of the product need to be considered. If the preferred products cannot fit on site, a smaller product must be considered.

A TCO analysis should also be done to evaluate which product is most cost efficient. A recommended measure for evaluation is TCO per capacity. The outcome from such an analysis may support the preferred choice from performance perspective or, possibly, may suggest a different choice.

From feature perspective, it is important to make all features available that benefit the deployment conditions. Hence, an appropriate selection of features as outlined in [Ch. 3.3] should be done. As feature content can be updated at any point in time, the choices made are not as critical as those relating to hardware which will be physically implemented on site for a long period of time and are costly to change.

Radio solution selection – Examples 1(3)



Over the following pages, the process outlined above is illustrated with examples. To apply the process, certain assumptions concerning the different steps are made. The solutions are chosen in relation to these assumptions.

Assumptions:

1. Strategy

- The service provider offers a 5G user experience in a certain area, viz. urban and suburban areas with more than 1,000 subscribers/m2.
- The service provider has limited CAPEX and OPEX and intends to make TCO-per-bit-efficient product decisions for site expan sions.
- A selected number (200) of sites are deemed important, e.g. government buildings, stadiums, important businesses. For these sites, performance has highest priority.
- The service provider has an agreement with another service provider to share RAN equipment where possible.

2. Network assets

- The service provider has acquired 100 MHz mid-band spec trum on 3.5 GHz.
- The considered network contains all deployment scenario types, dense urban, urban, suburban and rural and is covered by 2G and 4G with 2000 sites.

3. Deployment considerations

- Roughly 1000 sites, spread over several areas, have restric tions on deployment.
- The distribution of load and the classification into high- (RED), mid- (BLUE) and low-traffic (GREEN) sites have been estab lished.
- Predictions of YoY traffic growth have been made.

4. Product portfolio

- The Ericsson radio and feature portfolio as of 2021H2.

Site A: High-traffic (RED), Dense Urban. Priority: High. ISD: ~200m. Distribution of building heights: High.

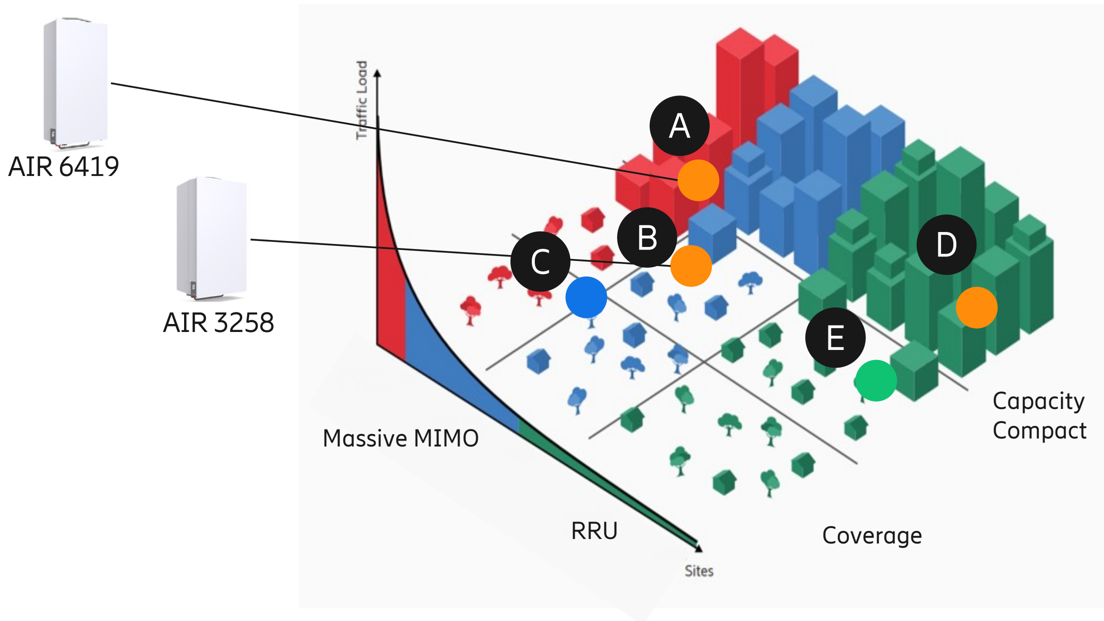
Site B: Medium-traffic (BLUE) Urban. Priority: Medium. ISD: ~350m. Distribution of building heights: Low. Candidate site for sharing equipment with other service provider.

Site C: High-traffic (RED) Suburban. Priority: High. ISD: ~800m. Distribution of building heights: Low.

Site D: Low-traffic (GREEN) Dense Urban. Priority: Medium. ISD: ~300m. Distribution of building heights: Medium.

Site E: Low-traffic (GREEN), Suburban. Priority: Low. ISD: ~600m. Distribution of building heights: Low.

Radio solution selection – Examples 2(3)

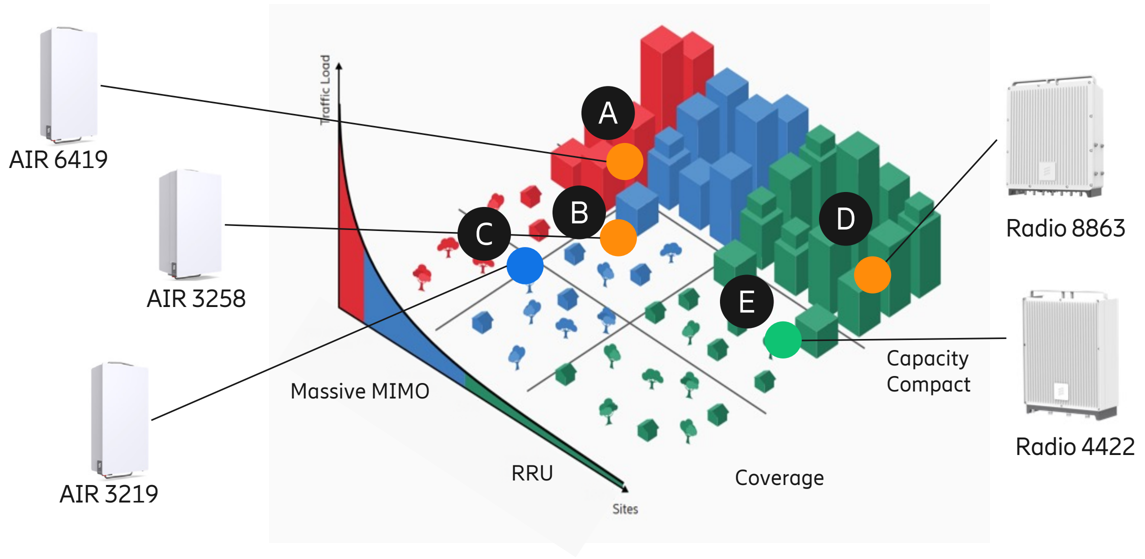


The most effective way to offer 5G user experiences with corresponding user throughput and capacity requirements in the whole 5G area is to make use of the available 5G spectrum: 100 MHz at 3.5 GHz on all sites. 100 MHz is a significant chunk of spectrum that will enable a significantly higher user throughput and capacity compared to the available spectrum for 4G.

Site A: The location is in an area with high traffic load and high traffic growth. This strongly suggests a product in the Massive MIMO Capacity segment. Since the priority is high, it is recommended to unlock the full potential of the mid-band by using superior Massive MIMO products and software features. Since the distribution of building heights is high, high steerability in the vertical dimension is needed, thus a 64T64R model. For this specific site, it is assumed that there are no deployment restrictions with respect to size, weight, wind load or visual impact. The preferred product choice (2021H2) is the AIR 6419. From feature perspective, the available capacity enhancing features should be deployed.

Site B: Similar conditions as for site A in terms of traffic and traffic growth. The service provider has an agreement to build a shared network with another service provider, hence the Massive MIMO product needs to support larger instantaneous bandwidths than in other areas. Therefore, another product in the Capacity segment with a large IBW is preferred, so that it can accommodate both service providers. As the building height distribution is much smaller, the benefits of a 64T64R product is not as significant, and hence a 32T32R product will have similar capacity performance but lower TCO. This motivates the selection (2021H2) of AIR 3258.

Radio solution selection – Examples 3(3)



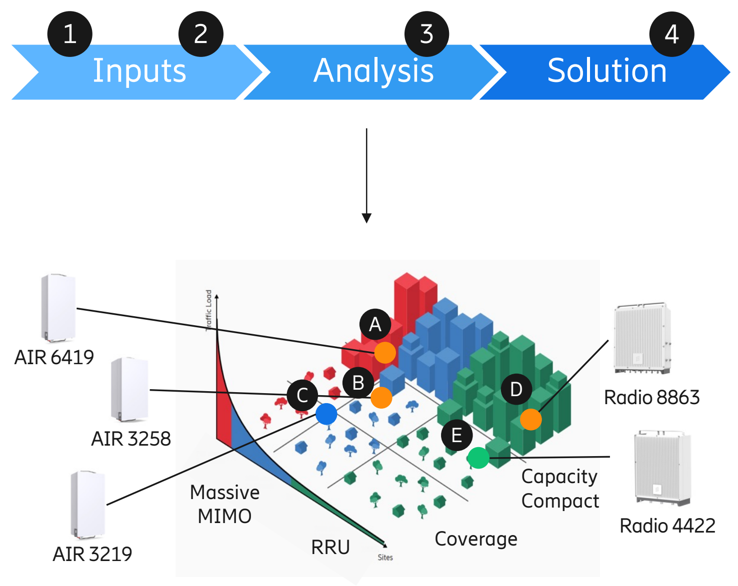
Site C: The location is in a suburban area with rather large ISD, high traffic load and high expectations on traffic growth. Therefore, products in both the Capacity and the Coverage segment can be considered. The distribution of building heights is low and therefore 32T32R models are expected to provide similar performance as 64T64R. The TCO analysis will therefore be useful to guide further concerning what choice to make. Considering the service provider strategy to prioritize dense urban and urban deployments, a product in the Coverage segment is chosen as the TCO is lower. Among the available products (2021H2) an AIR 3219 is an adequate choice.

Site D: This site is located in an urban area with low ISD. The actual traffic and the expected traffic growth is low. Therefore, the TCO-per-bit analysis indicates a benefit for a conventional solution based on RRU. The capacity from an RRU is considered sufficient. The priority of the site is considered high due to its location. Therefore, a Capacity RRU using 8T8R is preferred. There are no deployment restrictions. In this case a reasonable choice (2021H2) is Radio 8863.

Site E: The location is in a low-traffic area and the traffic growth is low. The site location is on rooftop in sub-urban region and the priority is to expand the site with as little site visual impact as possible (small antenna and small RRU). This suggests that a product in the Compact segment would be suitable. Since the load and the expected traffic growth are low, an RRU would provide sufficient capacity for the whole investment period (5-7 years). From a TCO perspective, a lower priced solution is preferred (lower price on RRU, less capacity demand from baseband licenses, less energy consumption, less complex and cheaper antenna). In this example, the Radio 4422 is a suitable choice (2021H2).

Summary – Guiding principles on how to select a radio solution

- To find cost-efficient radio solutions and to find where Massive MIMO provides superior performance, considerations concerning network strategy, assets, site-based analysis and available radio solutions are suggested
- Applying the guideline demonstrates what product solutions provide good performance and cost-efficiency



The service provider decides what their network shall deliver to the end users and a strategy for how to fulfill these objectives. The service provider then needs to find radio solutions for each site that meet the network requirements that follow from their strategy with respect to capacity, user throughput and coverage under the constraints of ease of deployment, cost efficiency and energy efficiency. The solution may involve any means available, e.g. taking new spectrum into use, finding a suitable product or site expansion. As radio sites are expensive to acquire and to operate, the preferred solutions build on the existing site grid. Once these are fully utilized in terms of use of available spectrum, equipment and functionality, site expansion can be considered.

In order to deliver a 5G experience, the new 5G mid-band spectrum is the solution to improve user experience with improved throughput and capacity. What products to use for mid-band expansion is then up to the service provider to decide based on the guidance in this handbook.

The main steps are:

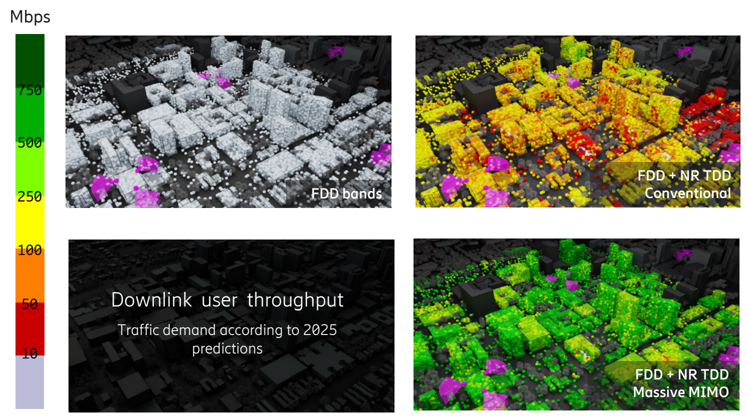
1. Specify the deployment objectives in different areas based on the network strategy.
2. What spectrum shall be deployed.
3. What are the deployment conditions at each individual site.
4. What products are available.

Selecting a set of products that meets the performance requirements, considerations need to be taken to deployment conditions and the requirements on cost efficiency.

5. Summary

Summary

- Massive MIMO technology improves network performance by making use of spatial properties of the radio channel and antenna array capabilities.
- 5G networks require cost-efficient solutions that provide significantly improved coverage, capacity, and user throughput to support higher capacity and enhanced user experience.
- All sites are unique with different performance requirements and deployment related constraints.
- Massive MIMO can effectively provide the required increase of coverage, capacity and user throughput.



Massive MIMO is a very powerful technology that can enhance network coverage, capacity and user throughput by making use of multi-antenna techniques, such as beamforming, null forming and spatial multiplexing (MIMO), and antenna array properties.

In an evolving 5G network, there is a need for increased coverage, capacity and user throughput to support increasing network traffic load and expectations on user experience. In addition to these performance requirements, there are also deployment related constraints for each radio site, e.g. ease of deployment, cost- and energy efficiency. All sites are unique, with different requirements and constraints. Massive MIMO is a versatile technology addressing many requirements.

Massive MIMO solutions are useful tools for providing the network performance improvements, viz. coverage, capacity and user throughput, required as the network evolves. The ability to increase coverage on higher bands is particularly important since this enables the full potential of the new mid-band spectrum on the existing site grid.

Massive MIMO radio- and feature solutions can be designed to meet certain performance requirements and deployment related constraints to improve the performance aspects most relevant and to increase cost efficiency. Ericsson has designed the radio portfolio into three segments, viz. capacity, coverage and compact to address the specific requirements for high capacity, coverage in wide areas and ease of deployment. The Ericsson features are segmented into feature classes, capacity, coverage, advanced performance and ease of deployment, cost- and energy efficiency.

6. Abbreviations

2D	two dimensional
3GPP	3rd generation partnership project
AAS	advanced antenna systems; note: synonymous to Massive MIMO
AIR	antenna integrated radio
AOSA	array of subarrays
AR	augmented reality
AS	antenna switching
BF	beamforming
BB	baseband
BW	bandwidth
CA	carrier aggregation
CAPEX	capital expenditures
CO2	carbondioxide
CSP	communication service provider
CSI	channel state information
DL	downlink
EIRP	equivalent isotropic radiated power
EM	electro magnetic
EMF	electro magnetic field
ERS	Ericsson radio system
FDD	frequency division duplex
GHz	giga Hertz
HW	hardware
Hz	hertz
IBW	instantaneous bandwidth
IRC	interference rejection combining
ISD	inter-site distance
MBB	mobile broadband
Mbps	mega bit per second
MHz	mega Hertz
MIMO	multiple input multiple output
mmWave	milimeter wave
MR	mixed reality
MU-MIMO	multi-user MIMO
NR	new radio
OPEX	operational expenditures
Qx	quarter x; x=1,2,3,4
R	receiver radio chain
RAN	radio access network
RDS	Ericsson radio dot system
RRC	radio resource control
RF	radio frequency
SINR	signal-to-noise-and-interference ratio
SNR	signal-to-noise ratio
SON	self-organizing network
SU-MIMO	single-user MIMO
SW	software
T	transmitter radio chain
TCBW	total configured bandwidth
TCO	total cost of ownership
TDD	time division duplex
UE	user equipment
UL	uplink
VR	virtual reality
XR	extended reality
YoY	year-on-year

7. References

1. Advanced Antenna Systems for 5G Network Deployments
Bridging the Gap Between Theory and Practice 1st Edition, Elsevier 2020, ISBN: 978-0-12-820046-9, Advanced Antenna Systems for 5G Network Deployments - 1st Edition (elsevier.com)

2. Advanced Antenna Systems for 5G networks (WP)
ericsson.com/en/reports-and-papers/white-papers/advanced-antenna-systems-for-5g-networks

3. Fixed Wireless Access handbook, 4 ed., EN/LZT 1/28701-FGD 101 449 Uen Rev G, Ericsson AB 2021
internal.ericsson.com/book-pages/82042/fixed-wireless-access-handbook

4. Ericsson Mobility Report June 2021
ericsson.com/en/mobility-report

5. Blog Ericsson Uplink Booster
ericsson.com/en/blog/2020/7/uplink-booster-significant-to-the-new-normal

6. Start simple with Ericsson Radio System Portfolio - 3Cs

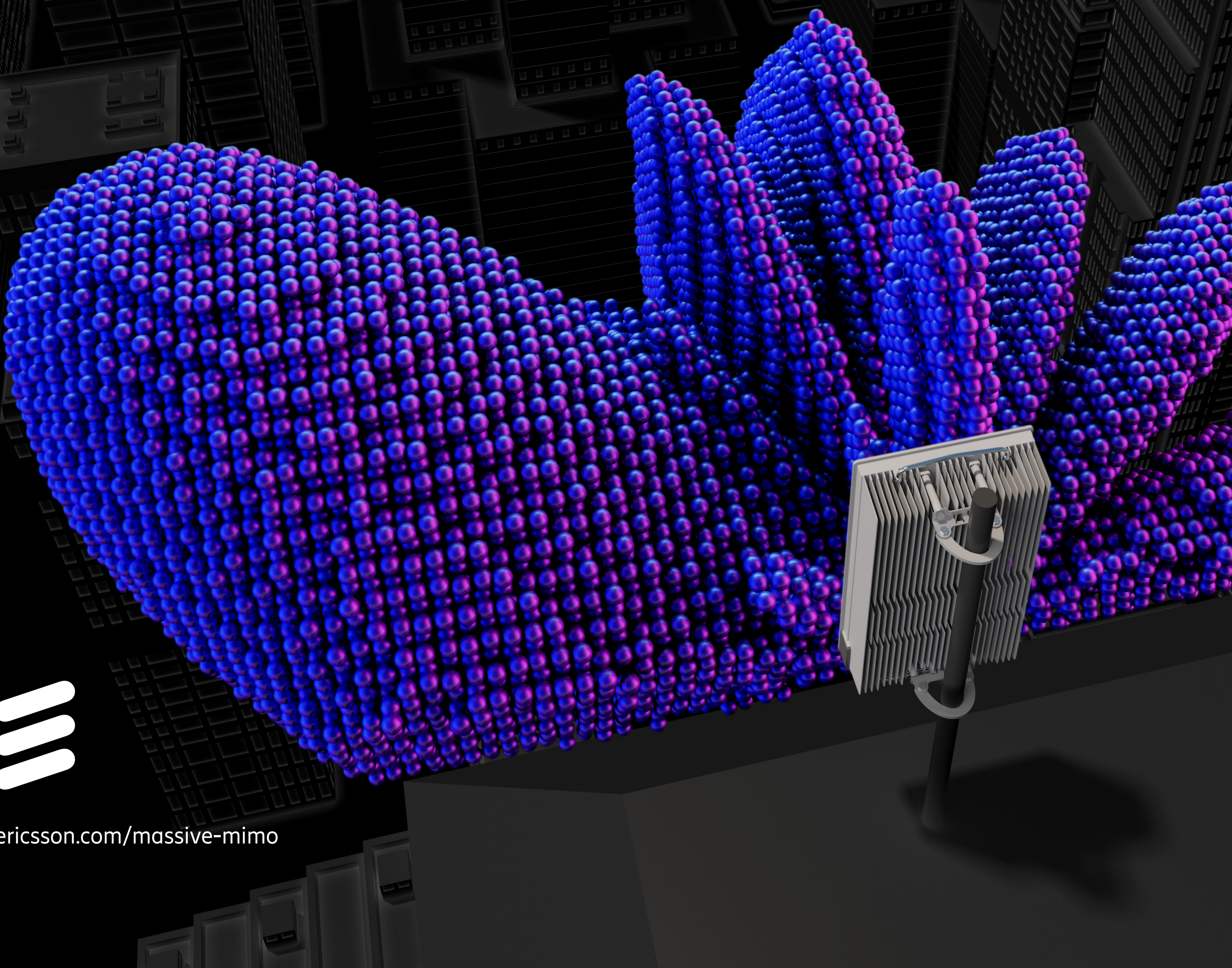
7. Blog How to build high-performing radios
ericsson.com/en/blog/2021/1/how-to-build-high-performing-radios

8. Blog How to build high-performing Massive MIMO system
ericsson.com/en/blog/2021/2/how-to-build-high-performing-massive-mimo-systems

Massive MIMO Handbook

Technology Primer

1st edition



ericsson.com/massive-mimo

Introduction

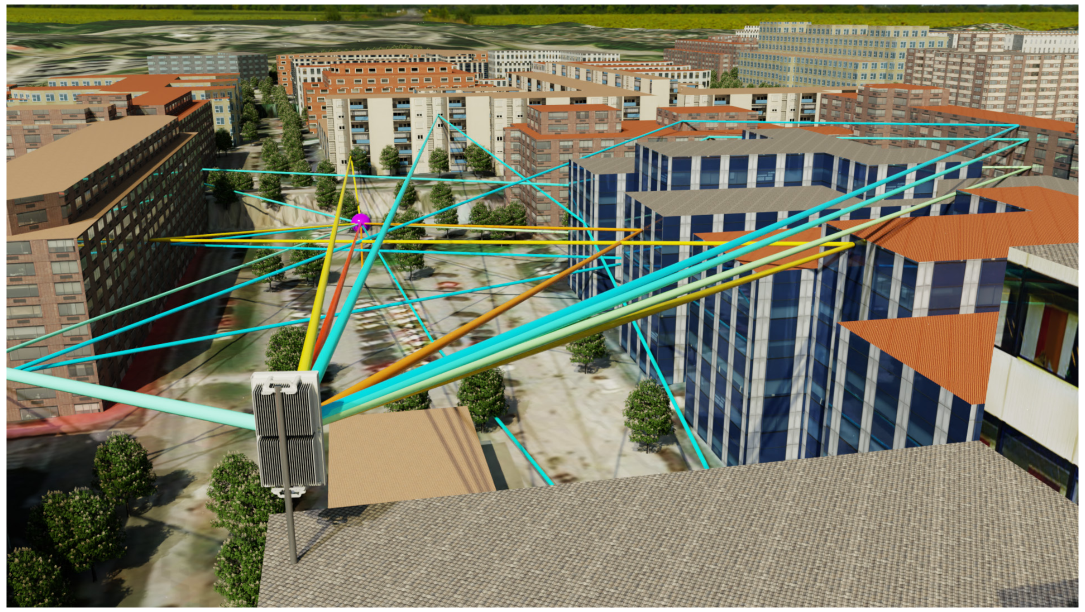
The purpose of this Technology Primer is to provide a deeper understanding to how Massive MIMO works, why it works and what performance is achievable in a real network deployment. Many related topics that provide additional insights to the background of Massive MIMO, e.g. antennas and wave propagation, the implications of Massive MIMO, e.g. architecture and implementation and radio requirements are also covered.

The different chapters of the Technology Primer can be read selectively and standalone to deepen knowledge where the reader chooses. The chapters are however organized in a way that they best are read in succession. For example, the chapters: antennas and wave propagation, antenna arrays, multi-antenna technologies, 3GPP solutions, network performance and Massive MIMO features will be better understood if read in a sequence. If readers has a reasonably good understanding of an area from start, they do not need to read everything in these chapters, and rather selectively read what is important to them.

Table of Contents

1. Antennas and wave propagation

Antennas and wave propagation – Introduction



The basis of wireless communication is the transmission and reception of radio waves using antennas. Traditionally, the way antennas have directed their transmission or reception has been fixed and decided during manufacture or installation. However, the introduction of antenna arrays and Massive MIMO allows the transmission and reception capabilities to be varied dynamically in response to the spatial distribution of user and traffic and the multiple ways that the radio waves propagate between the transmitters and receivers.

The properties of antennas and wave propagation are therefore fundamental to the design and operation of Massive MIMO systems, which attempt to exploit these properties and overcome some of the challenges to reliable communication that they present. This chapter introduces some key concepts of antenna theory and wave propagation, such as:

Wireless communication utilizes electromagnetic waves that are polarized and can be superimposed on each other.

Antennas are directive and can focus transmissions of waves in specific directions. The larger the antenna is with respect to the wavelength, the more focusing and hence higher antenna gain it can achieve.

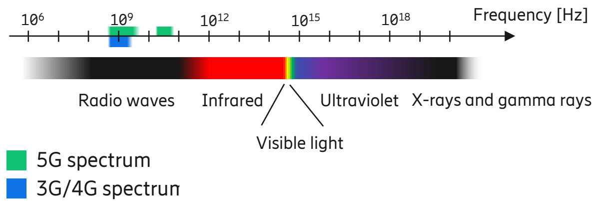
Waves travelling from a transmitter to a receiver experience path loss that tends to increase with increasing distance and frequency. This drives the need for more directive antennas at higher frequencies, preferably realized using Massive MIMO.

Multipath propagation and the associated delay and angular spread affects numerous aspects of wireless communication, including Massive MIMO algorithm and hardware design and performance.

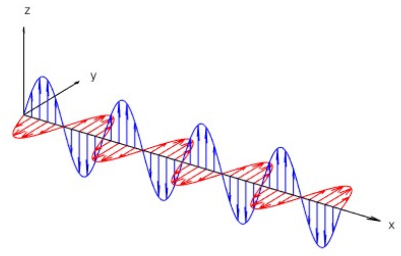
These concepts will be used in the chapters [Ch. 2], [Ch. 3], [Ch. 4], [Ch. 6].

Properties of electromagnetic waves

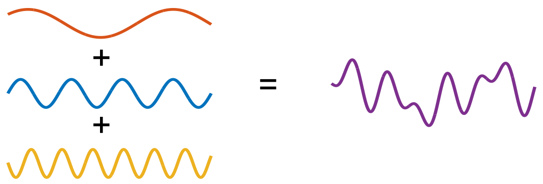
Electromagnetic waves come in a large range of frequencies and wavelengths



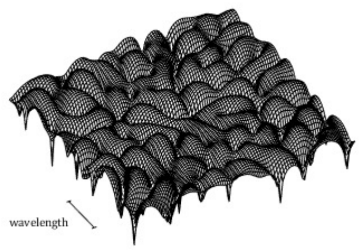
Electromagnetic waves are polarized



Many waves can be added together, "superimposed"



Superposition creates standing wave patterns, "fading"



The basis for wireless communication is the existence and use of electromagnetic fields. More specifically, by creating waves in these fields that travel through space, it is possible to transmit energy and information from one location to another location. The information is encoded into the amplitude, phase, and frequency of the waves. The range of frequencies and wavelengths, the spectrum, of naturally occurring or articificially produced electromagnetic waves span many orders of magnitude. Mobile wireless communication systems up to 4G typically use waves with frequencies from a few hundred megahertz (MHz) up to several gigahertz (GHz), while 5G can also utilize frequencies of several tens of GHz.

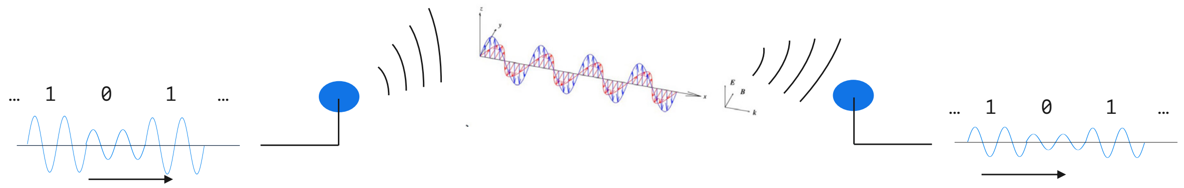
An electromagnetic (EM) wave is transversal, meaning that the electric field and the magnetic field are orthogonal to the direction of propagation and to each other. The polarization of the wave describes how the electric field is oriented. A wave can be linearly polarized if the field at a point in space is oscillating in a fixed direction, circularly polarized if it is rotating with constant amplitude, and elliptically polarized if it is both rotating and oscillating.

A very important characteristic of electromagnetic waves is that they can be can be added together, "superimposed". The field vectors in a point in space is the sum of the field vectors from each individual wave. Superimposing two or more waves with the same propagation direction can therefore create a new polarization. Any arbitrary polarization can be described as the superposition of two waves with orthogonal polarizations.

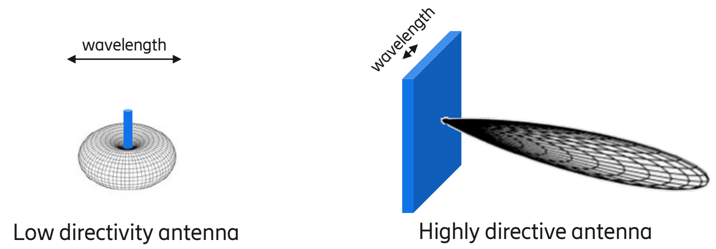
Superposition of waves with different propagation directions creates standing wave patterns where the average field strength is lower or even zero in some locations. This is referred to as fading and has traditionally been a challenge to the reliability of wireless communications. However, the increasing use of directional antennas that can distinguish the waves based on their propagation directions is turning this into an opportunity instead.

Antennas

Antennas act as the interfaces between conducted and radiated signals



The properties of the antenna (or array of antennas) determine how waves are radiated or received in different directions



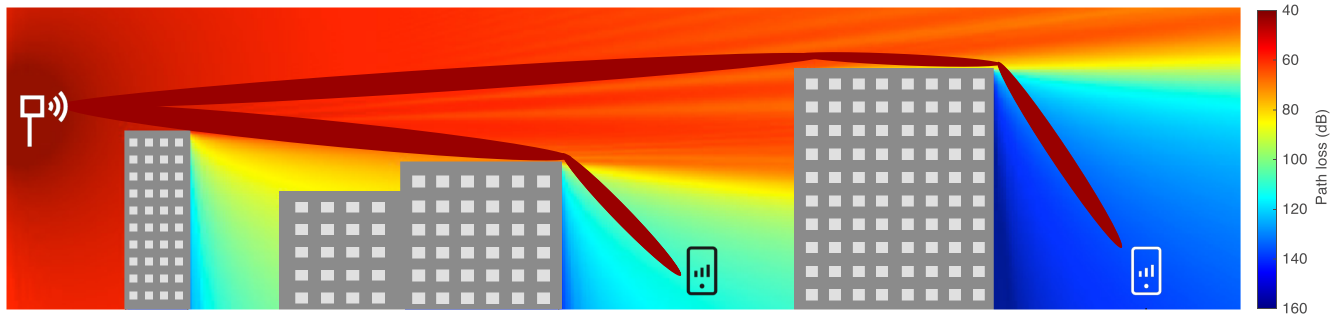
In wireless communication systems, antennas are used to radiate electromagnetic waves from one location and capture some fraction of them in another location. Once the wave has been generated it will continue to propagate outwards, radiate, into the surrounding space. When the radiated wave encounters a receiving antenna the time-varying fields will drive a voltage or current on the antenna terminals. Hence, the receiving antenna captures some of the power that was transmitted from the other antenna. If the captured waves have enough power then the receiver is able to decode the information carried by the waves.

The effectiveness with which antennas can radiate and capture waves is fundamental to the wireless communication quality. All antennas are directive to some extent, meaning that they are more effective in radiating waves in certain directions and with certain polarizations, as illustrated by the radiation pattern. The effectiveness or antenna gain is often given relative to that of an ideal isotropic antenna that radiates equally in all directions. Antennas are usually reciprocal, meaning that an antenna has the same radiation pattern when receiving as when transmitting. Pairs of antennas with orthogonal polarizations are commonly used in wireless communications to be able to transmit or receive waves with arbitrary polarizations.

Antennas that are small in relation to the wavelength usually have low directivity and therefore spread their radiated energy in many directions. Antennas that are large in relation to the wavelength have more freedom in shaping the radiation pattern, such as creating high directivity by focusing the transmitted energy in a narrow range of directions. High directivity is very useful for improving the communication quality but also presents a challenge since the transmission and reception directions need to be carefully aligned. Massive MIMO has evolved to solve this challenge in a dynamic environment with multiple moving users and complex wave propagation.

Antennas of a given physical size tend to become more directive and have more degrees of freedom for shaping the radiation pattern as the frequency increases. Alternatively, for a given directivity the antenna becomes physically smaller at higher frequencies. A consequence of this is that the ability of an antenna to capture energy from a passing wave, its effective antenna area, reduces with increasing frequency, assuming the same antenna gain at all frequencies. Therefore the use of more directive antennas at higher frequencies becomes necessary which is another reason for introducing Massive MIMO.

Path loss



Only a tiny fraction of the transmitted power reaches the receive antenna

Antenna directivity can help increase this power

Massive MIMO ensures that the transmissions occur in the best directions

When an antenna radiates electromagnetic waves with a certain power into free space, these waves will expand spherically. As the distance to the transmitter increases the power gets more spread out and hence the local field strength decreases inversely proportional to the distance. The path loss (basic transmission loss) is the ratio between the power transmitted by one isotropic antenna and that received by another isotropic antenna. In free space propagation, the path loss is proportional to the square of the distance. The free space path loss is also proportional to the square of the frequency, although this is not a "propagation loss" as such but rather a consequence of the reduced effective antenna area of the receiving isotropic antenna.

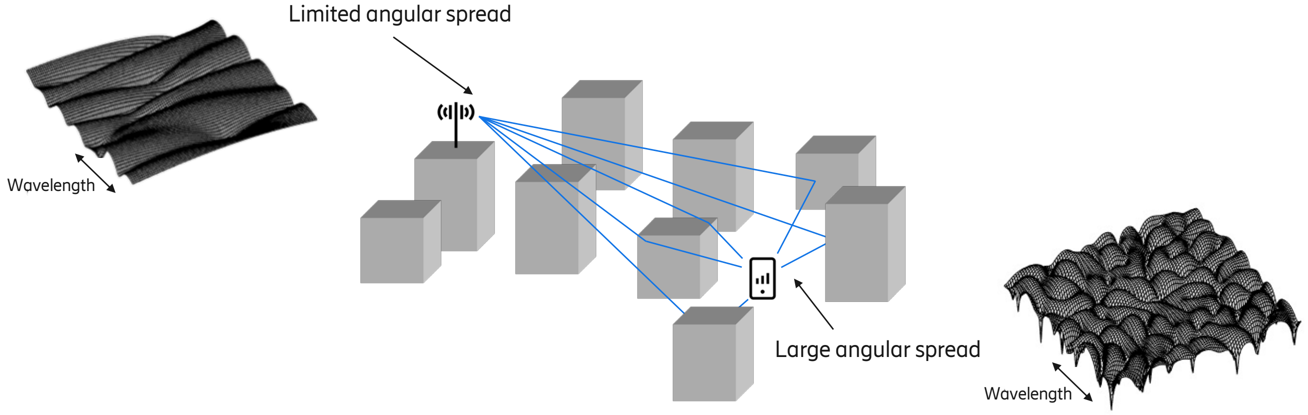
When there are obstacles such as terrain or buildings that may scatter or block some of the electromagnetic waves, the path loss can increase quicker with distance, typically by the distance to the fourth power for outdoor non-line of sight communications.

Many materials strongly attenuate radio waves passing through them. Therefore, outdoor to indoor propagation losses can be very high and also increase with frequency.

Directive antennas are regularly used to partially compensate for the path loss and thereby increase the communication range. This becomes particularly important at higher frequencies where the reduced effective area of the receive antenna and the frequency-dependent losses in non-line of sight need to be mitigated. Massive MIMO ensures that the high directivity is efficiently utilized by steering the radiation in the most beneficial directions.

Multipath propagation creates fast fading

Massive MIMO mitigates and exploits fast fading



As the electromagnetic waves travel along multiple propagation paths from the transmitter to the receiver we speak about multipath propagation. The superposition of many waves give rise to fading, i.e. variations in time, space, and frequency of the average field strength and polarization. The larger the range of directions in which propagation paths occur, the angular spread, the more rapid the variations are. A typical user often experiences multipath from many directions causing the distance between peaks and nulls in the fading pattern to be less than half of a wavelength. At a base station which is often elevated above its surroundings to provide large area coverage, the angular spread is typically confined to a more narrow interval and therefore the peaks and nulls gets spread out more.

The multipath and angular spread strongly impacts how antenna directivity can be used and whether simple beam shapes are adequate or more irregular radiation patterns are needed to optimize the communication. Massive MIMO contains methods and algorithms for acquiring information about complex and rapidly changing multipath conditions and using the degrees of freedom of electrically large antennas with dual-polarized elements to phase shift different multipath components such that they add constructively and enhance the communication quality.

A moving transmitter or receiver will experience rapid time variations of the signal strength due to the fading, often changing completely within a fraction of a second. This kind of fading is often referred to as fast fading, and has always been a challenge to reliable and efficient communication. Modern wireless communication uses opportunistic scheduling and link adaptation to ensure that the fading channel is utilized to the greatest extent possible.

The fading variations are not only confined to time and space. The different path lengths of the different multipath will cause a delay spread of any transmitted signal. The different delays will cause the superposition of the waves to change if the frequency is adjusted slightly, leading to fading variations also in within the frequency band in which communication is occurring.

Fast fading in space, time and frequency has always been one of the main challenges to maintaining reliable communication quality, particularly for high bitrates with low latency. Multi-antenna techniques started out as methods to mitigate the fading by adding diversity but has subsequently evolved to more directly take advantage of the multipath channel by potentially using the different propagation paths and polarizations as separate and parallel communication channels through MIMO schemes.

2. Antenna Arrays

Introduction to antenna arrays

Massive MIMO commonly uses planar arrays of dual polarized elements divided into subarrays

In this chapter, a basic description of antenna arrays and what can be accomplished with them is presented. In the figure above, an example with a 128-element uniform planar array with eight rows and eight columns of dual polarized element pairs. Such as uniform and planar structure is the most common structure in Massive MIMO deployments, and the description in this chapter is limited to this array type.

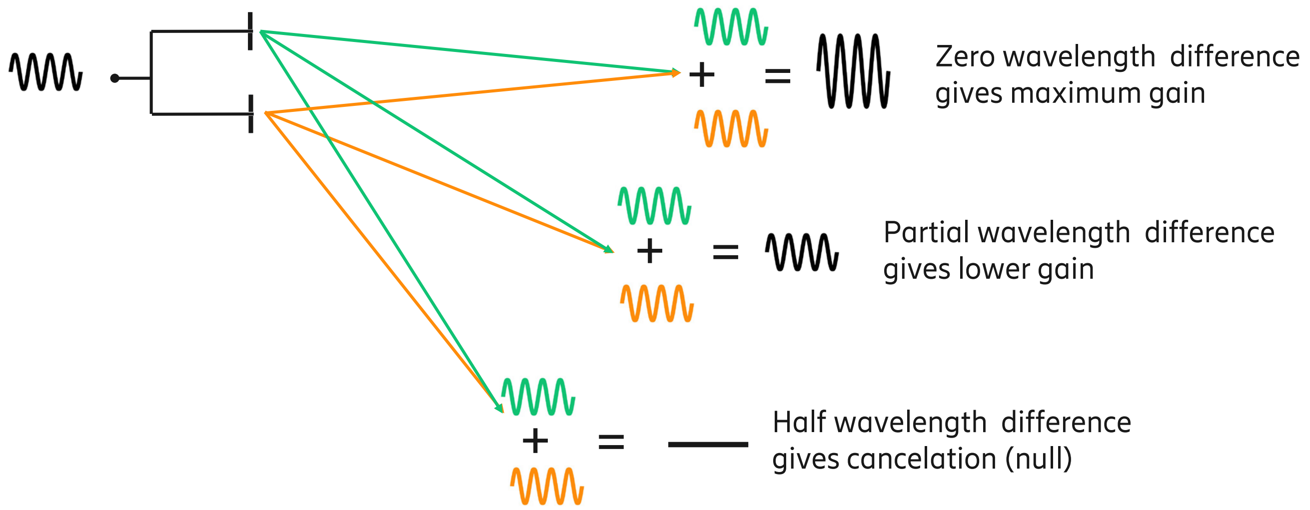
The basic principle for beamforming with an antenna array and a beamforming weight vector, which can be used to form a beam, are introduced. More specifically, so-called classical beamforming is taken as a baseline in this chapter.

Then terms such as main lobe and side lobes, which are often used when describing radiation patterns, or beam patterns, are defined, and it is explained that the patterns can be expressed as the product of an element factor and an array factor. Some properties of the beams as a function of for example the number of array elements are then stated.

By partitioning the array into subarrays, as is also illustrated in the figure above, it is possible to reduce the number of radio chains and the chapter also gives some background for how subarrays are often used in practice.

Finally, antenna arrays can however not only be used to adapt the beam shapes, but also for spatial multiplexing, denoted MIMO (multiple-input multiple-output). There is both spatial multiplexing to a single user, SU-MIMO, and spatial multiplexing to multiple users, MU-MIMO.

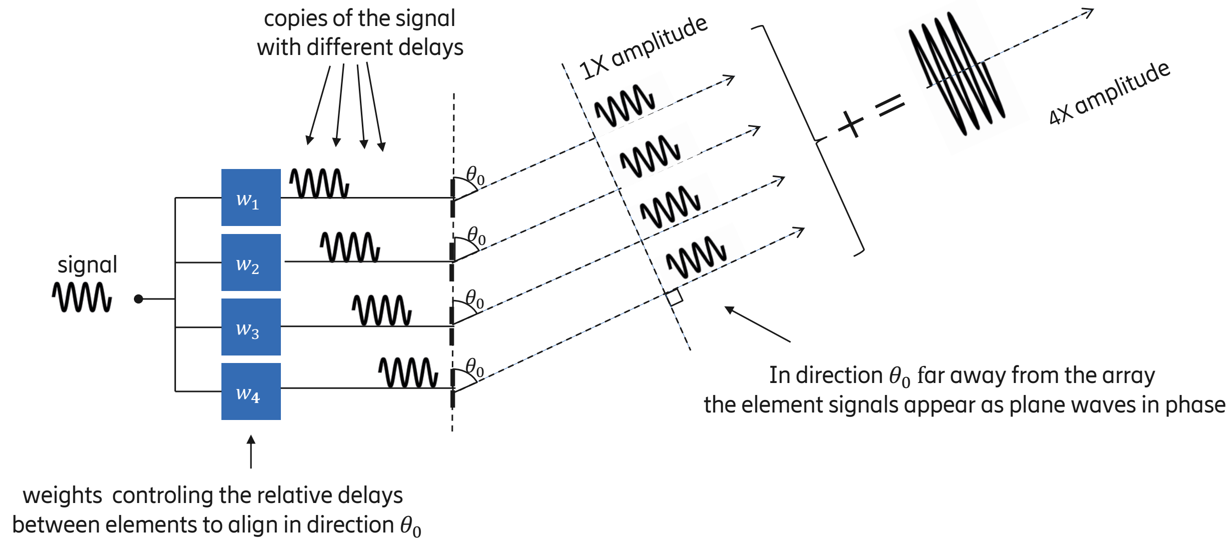
The basic beamforming principle is to transmit the same signal from multiple antennas



The electromagnetic field generated by an array is a superposition, a sum, of the contributions from the individual elements, [Ch. 1, p. 6]. Put another way, this means that a beam may be formed by transmitting the same signal from multiple antennas. It takes time for the signals to propagate to the receiver which thus sees a delayed signal. The signals from the antennas add up over-the-air and appear as a sum signal on the receiver side. This is illustrated above assuming a free-space channel and the simplest case with two antennas above.

If all the signals have identical time delay, or phase, on the receiver side, completely constructive addition is achieved and maximum gain in signal strength is obtained. In the example above, this happens right in front of the array (top). Conversely, if the signals are completely out of phase, they cancel in the summation and a zero signal is obtained which corresponds to zero gain (bottom). For directions in-between these two extremes, the phase difference is somewhere in-between and so is the resulting beamforming gain (middle). At such angles, the maximum gain in signal strength is between zero and the maximum gain, which in this case is two.

Classical beamforming directs power in desireddirections by adjusting relative delays



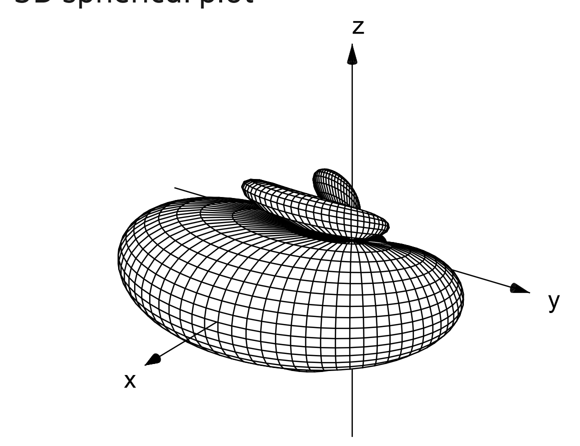
By transmitting copies of the same signal and appropriately adjusting the different amplitudes and delays from all the elements it is possible to control the radiation pattern. In fact, since the total delay for the signal from each antenna is the sum of the adjusted delay and the propagation delay it is possible to control the direction with maximum gain stemming from constructive addition of the signals from all antennas. In other directions, the contributions add destructively and when they cancel out there will be nulls.

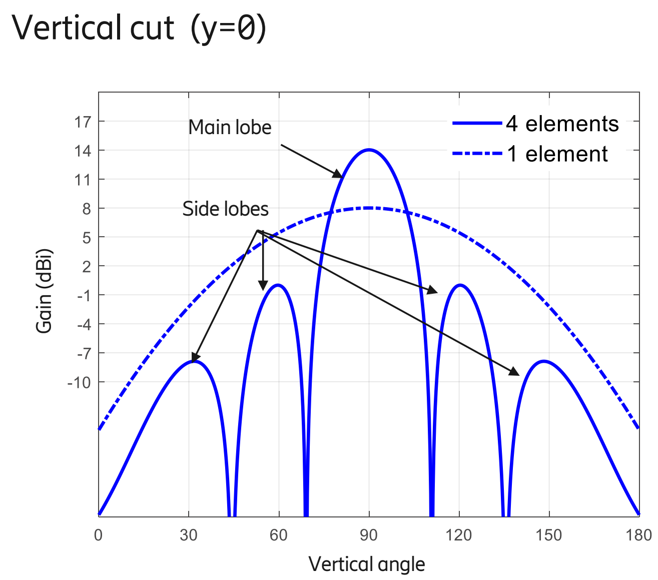
The delay is equivalent to the phase of the signal. The amplitude and the delay adjustment of a signal may be represented by a complex number, or weight. The adjustment can be implemented by a complex multiplication of the signal to be transmitted. This since the communication signals transmitted at radio frequency can be represented with equivalent complex valued baseband signals. The set of weights for all the antennas is often collected in a beamforming weight vector, where each element of this vector represents the delay and amplitude of a specific element. In the Figure above w_1, w_2, w_3 and w_4 are the element adjustments and the beamforming weight vector w would be a vector defined from these elements.

For classical beamforming, the beamforming weight vector is chosen to maximize the transmitted signal power in a desired direction. Such a choice is optimal in free-space single-path propagation with LOS between to the receiver. Classical beamforming is illustrated above for transmission of a single frequency sinusoid in a direction θ_0, where the contributions from all the elements appear as plane waves in phase. Again, the superposition is a sinusoid with the same frequency but with an amplitude which is the sum of the amplitudes i.e., four times the amplitude as compared to a single element for an array with four elements.

Above, transmission was described. The array may also, just as any antenna in general, be used for reception, [Ch. 1, p.6], and it is then possible to amplify desired signals received from certain directions and suppress, or null, interfering signals in other directions.

Beam patterns typically have lobes





A radiation pattern describes properties of the electromagnetic field as a function of direction, and a beam is often described by its radiation pattern. Most often, the power of the field is illustrated in the form of gain relative an isotropic radiator, and this referred to herein as a beam pattern.

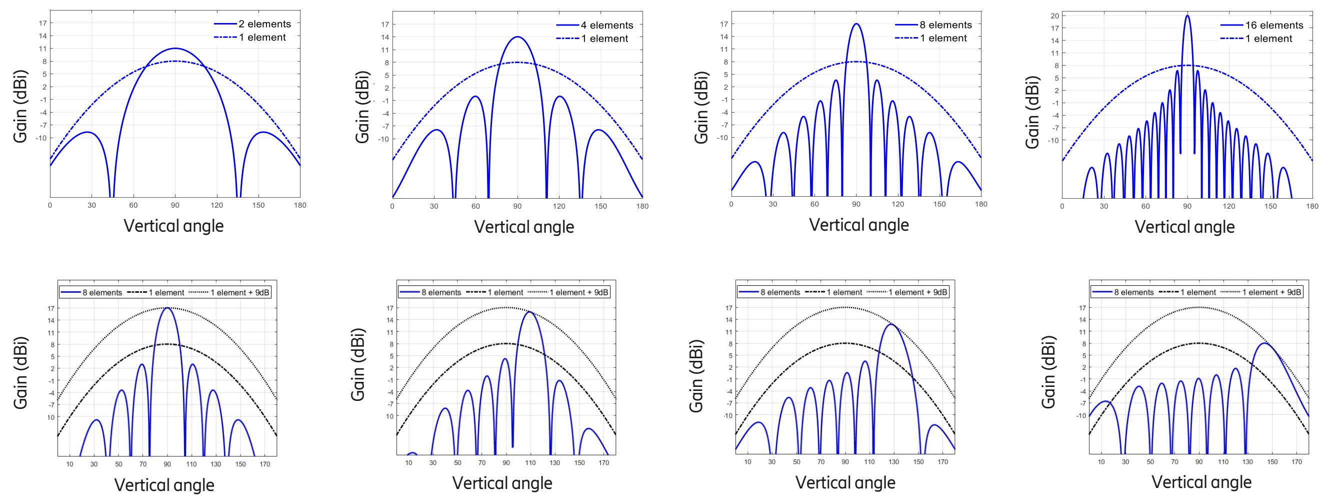
To the left above, a 3D spherical plot of such a gain pattern is exemplified. Often, cuts of such a pattern are also illustrated and to the right, a vertical cut in the *xz*-plane *y*=0 of the same pattern as function of the angle θ is shown. In here, θ is the angle to the positive z-axis, also referred to as vertical angle, and the beam pattern illustrates a beam steered to direction θ_0=90°, i.e., the direction forward, perpendicular to the antenna array.

Beam patterns typically have multiple lobes. Lobes are different regions of the beam pattern separated by nulls, and a null is a direction in which the contributions from the different antennas add destructively. A main lobe is a lobe in which the gain attains its maximum. Other lobes, where the gain does not attain it maximum, are referred to as side lobes.

For classical beamforming, introduced on the previous slide, the beam looks like a 'beam' in the intuitive sense of a 'directed entity'. By this we mean that the beam pattern has a single well-defined main lobe with maximum gain and sidelobes with significantly lower gain levels, similar to a flashlight beam.

As a sidenote, the term "beam" of antenna was originally defined as the main lobe of the beam pattern of an antenna. With generalized beamforming in multipath propagation channels, however, not only the main lobe but the entire beam pattern is of interest. This will be further elaborated in [Ch. 3, p. 24]. The reader is referred to [1, Ch. 5] for a more thorough discussion.

Beam patterns depend on the number of elments and the element pattern



For an array, the total electrical field is the sum of the contributions from the individual elements. This means it depends not only on the relative amplitude and delays between the different copies from the individual antennas but also on the radiation patterns of the elements. On a logarithmic scale, the total gain is the sum of the elements' gain and the array gain,

Total gain pattern [dBi] = element gain pattern [dBi] + array gain pattern [dB].

Here we refer to the gain patterns rather than just gain to highlight that the equation holds in all directions and not only in a single direction. It is furthermore assumed that the all the elements of the array have identical radiation patterns so that there is a single element pattern. The array gain pattern on the other hand captures how the contributions combine in different directions for a given beamforming weight vector. In directions where the contributions add constructively, the array gain is high and in directions with nulls the array gain is zero (on linear scale).

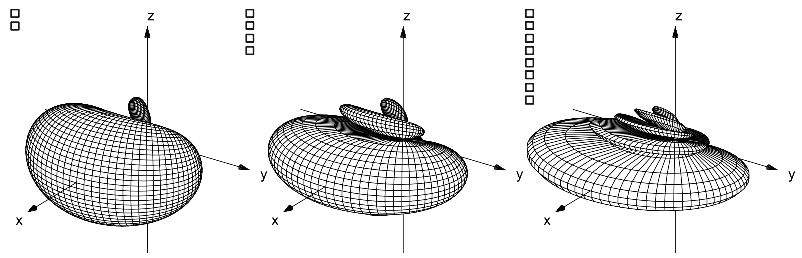
For classical beamforming with an *N*-element uniform linear array, the maximum gain is *N* and the width of the main lobe is proportional to the inverse of the number of elements, ~*1/N*. This means that as the size of the array grows in terms of the number of elements, the maximum gain increases but the main lobe becomes narrower. This is illustrated in the top figure above, and the reader is referred to [1, Ch. 4] for more details and derivations.

Since the main lobe becomes narrower with increasing number of antennas, dynamic beamforming or UE specific beamforming, as discussed in [Ch. 3, p. 23] becomes necessary. This mean that the direction of the main lobe can be pointed to where the UE is when it is scheduled. Furthermore, since the direction of the main lobe in which the array gain pattern has its maximum *N* can be steered by changing the beamforming weight vector, it follows from the relation above that the element gain pattern shapes the envelope over the set of beams, or maximum gain, that can be generated.

Vertical or horizontal linear arrays – same maximum gain but different beam patterns

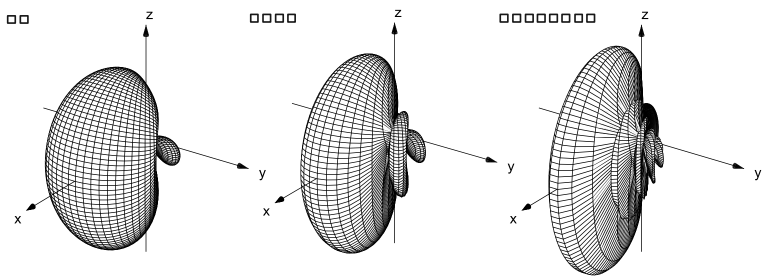
Vertical arrays

The higher the array, the narrower the beam in the vertical dimension



Horizontal arrays

The wider the array, the narrower the beam in the horizontal plane



It is possible to form both vertical arrays as well as horizontal arrays by stacking elements on top of each other and next to each other respectively.

Assuming again classical beamforming and uniform linear arrays, the width of the main lobe will change in both cases:

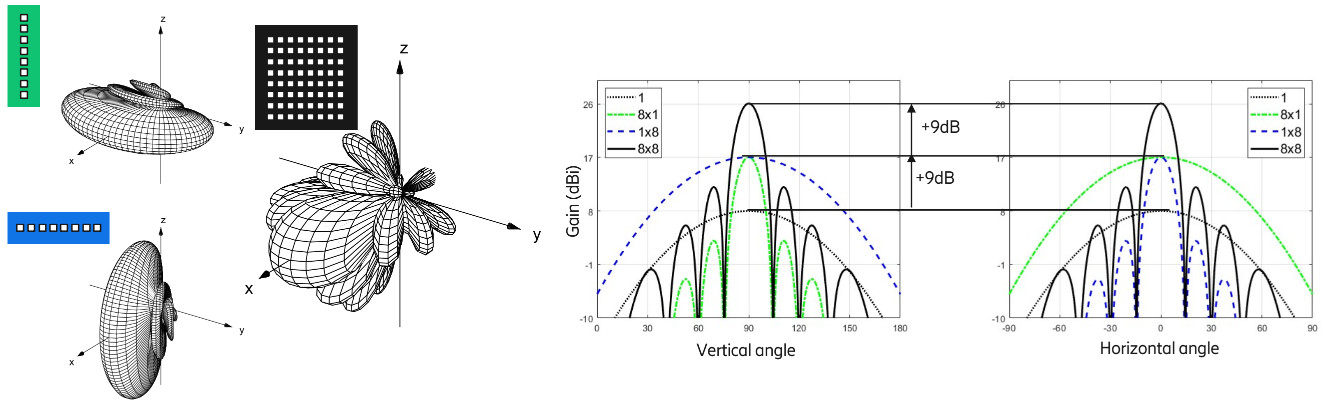
For a vertical array, the vertical main lobe width, which is the width of the main lobe in the vertical cut (*y=0* above) scales with the height of the array. The higher the array, the narrower the main lobe in the vertical dimension.

For a horizontal array, the horizontal main lobe width (in the cut *z=0* above) scales with the width of the array. The wider the array, the narrower the beam becomes in the horizontal plane.

Recall from the previous slide that the maximum gain as compared to a single element depends on the number of antennas. In the illustrations above, the maximum gains are obtained along the x-axis and are the same for the same number of elements for both the vertical and horizontal arrays. Their shapes are however different.

Note that conventional sector antennas are implemented as vertical arrays, for example as a single column of dual polarized elements. Such a single column antenna could have two connectors, one for each polarization, feeding all elements with the same polarization through a feeder network. Since the feeder network is not dynamically changed one can think of this as static beamforming and the maximum gain of the sector antenna as compared to an individual element is then given by the number of elements in the column. So, a tall antenna with many elements has a high gain, but at the same time, the width of the main lobe in a vertical cut is narrow. This follows from the previous slide.

Planar arrays enjoy the benefits of both vertical and horizontal arrays



More antenna elements lead to narrower main lobe in both vertical and horizontal dimensions

A linear array can be either vertical or horizontal, and by increasing the sizes of such arrays the main lobe width in the vertical or horizonal dimension will decrease while the gain grows. A uniform planar array consist of elements arranged in a uniform grid with N_v rows and N_h columns. As mentioned in the introduction, uniform planar arrays are the most common configurations used in practice for Massive MIMO.

Two ways to view such an array include:

- A vertical array where each element is a horizontal array
- A horizontal array where each element is a vertical array

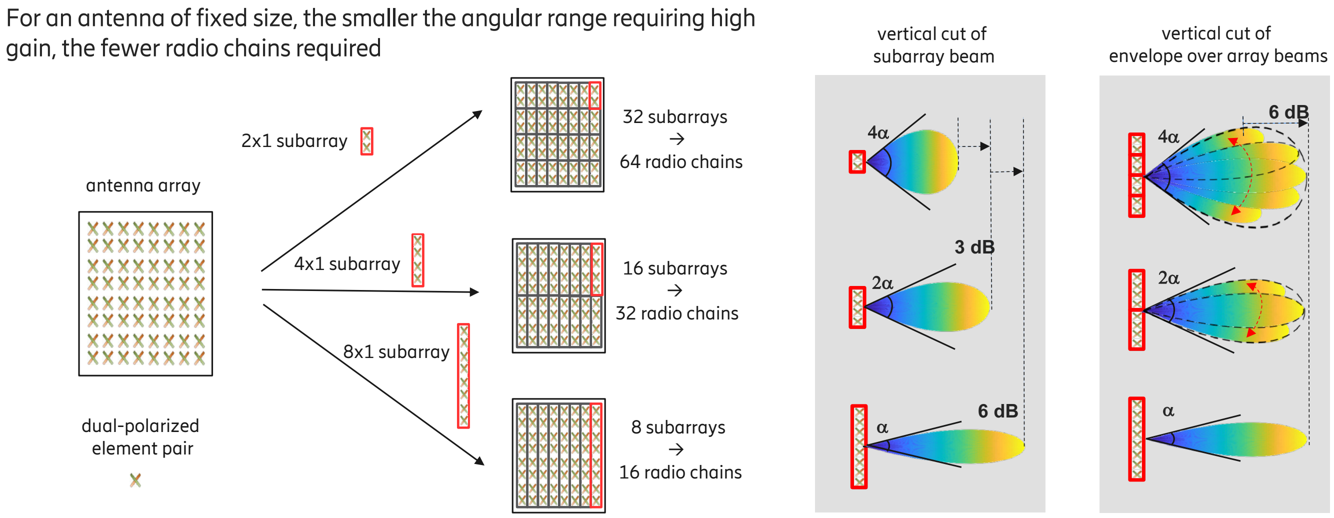
Recall that the total gain pattern (on a logarithmic scale) is the sum of the element gain pattern and the array gain pattern. For classical beamforming, the maximum of the array gain pattern will be given by the total number of elements,

$$10\log_{10}N_v N_h = 10\log_{10}N_v + 10\log_{10}N_h$$

which can be recognized as the sum of the gains for the corresponding vertical and horizontal arrays. Thus, classical beamforming with an 8x8 planar array offers an array gain of 18 dB which can be thought of, for example, as 9 dB from beamforming with an 8x1 vertical array of elements which all enjoy a 9 dB array gain using 1x8 horizontal arrays.

Furthermore, the width of the horizontal main lobe scales with $\sim 1/N_h$, whereas the vertical scales with $\sim 1/N_v$. Thus, in the case with a planar array the main lobe widths decrease in both dimensions when the number of elements increases in both dimensions.

Partition the array into an array of subarrays to reduce the number of radio chains



Uniform planar arrays of dual polarized antenna pairs are commonly used. The two polarizations can be used for spatial multiplexing, for diversity and to mitigate losses due to polarization mismatch.

A uniform planar array can be partitioned into an array of subarrays (AOSA). A key benefit of such a partitioning is that the number radio chains needed is reduced compared with one radio chain for every element). Instead of one pair of radio chains for each dual-polarized element pair, only one pair of radio chains is needed per subarray. The disadvantage of the grouping of elements into subarrays is that the range of angles within which the beamforming can be done without significant gain drop becomes narrower.

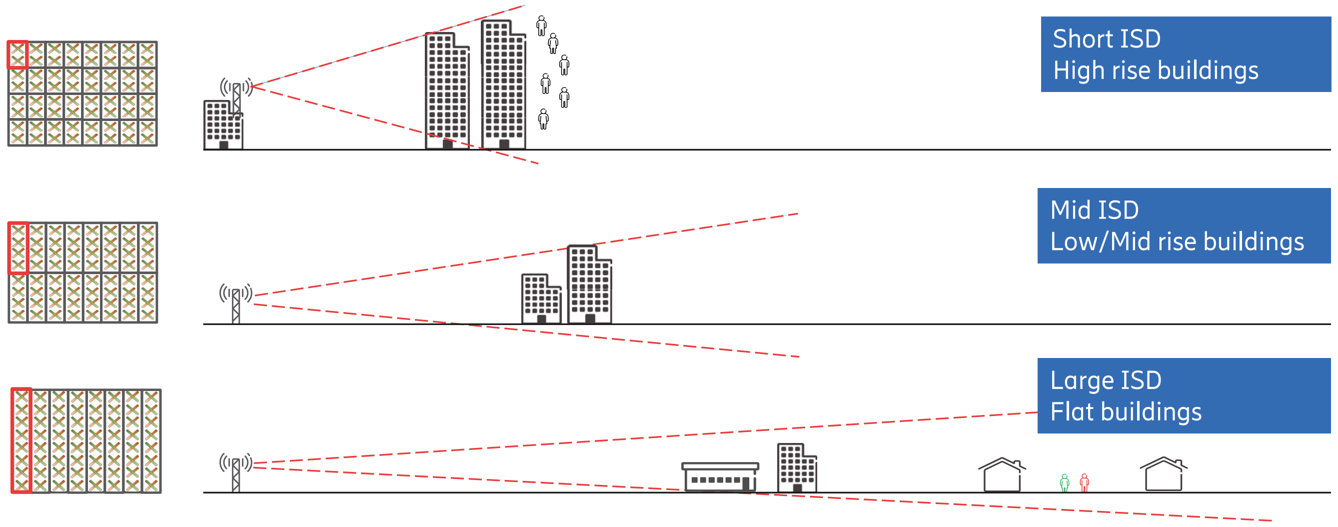
To see this, first note that the maximum gain is proportional to the antenna size in terms of the number of elements. So, partitioning the array into subarrays does not reduce the maximum gain achievable.

However, the larger the subarrays, the narrower its half power beamwidth and since the total gain pattern is given as the product of the array gain pattern and the subarray pattern (which is the element in the array of subarrays) it follows that the envelope of the beams, which describes the achievable total gain, also becomes narrower.

In the figure to the right above, partitioning of an eight-element array into subarrays of different sizes is illustrated, two elements at the top, four elements in the middle and eight elements at the bottom.

More reading can be found in Advanced antenna systems for 5G network deployments, [1], and Advanced Antenna Systems for 5G networks, [2].

Choose subarray size depending on deployment

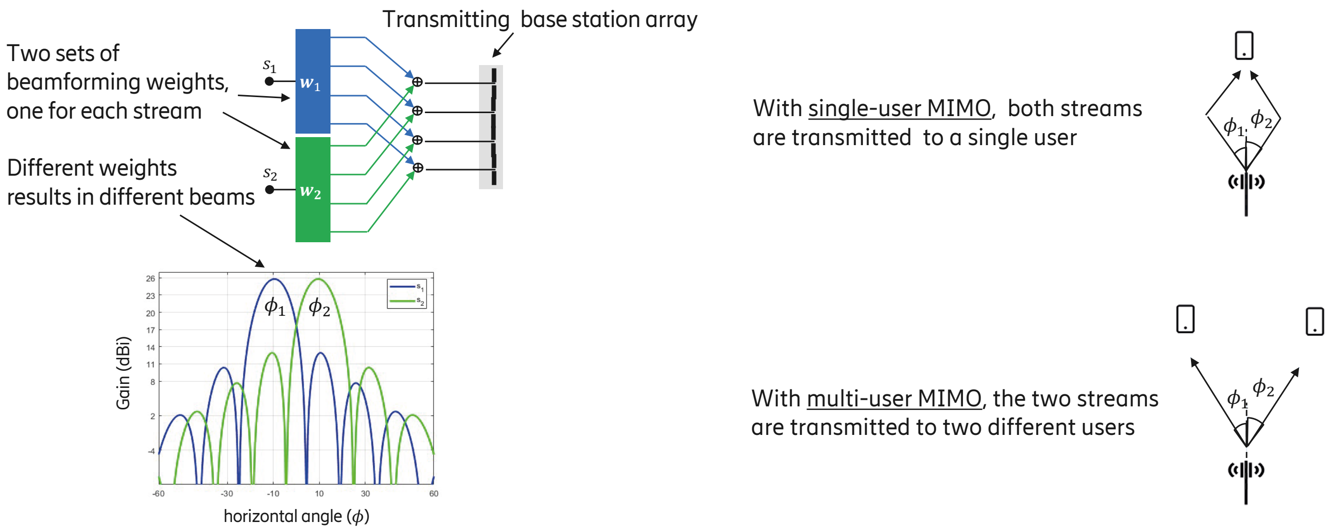


In practice it is of interest to have as large subarrays as possible while ensuring that most of the UEs are within the main lobe of the subarray. This offers high gain while not requiring too many radio chains. In practice, vertical subarrays are used, and the choice of subarray size depends on the UE and channel angular distribution.

In dense urban high-rise scenario, a small subarray such as 2x1 is preferable. This is so since the angular spread [Ch. 1, p. 8] of users in the vertical domain is large.

In suburban rural scenario with more sparse deployment of base stations, larger subarrays such as 8x1 can be used due to the small angular spread of users in vertical domain.

Antenna arrays can be used for spatial multiplexing, both single-user MIMO and multi-user MIMO



Antenna arrays can also be used to transmit multiple streams, or layers, simultaneously with different beams on the same time-frequency resource. This is referred to as spatial multiplexing and the streams of data symbols multiplexed are referred to as layers.

There are two basic use cases:

- Single user-MIMO, (SU-MIMO) where multiple layers are transmitted to a single user terminal. This requires a multipath propagation channel as well as a receiver with multiple receiver antennas.
- Multi-user-MIMO (MU-MIMO) where multiple layers are transmitted to different user terminals in different directions.

Suppose that two signals s_1 and s_2 are to be transmitted and that each of them has a beamforming weight vector to generate the corresponding beam. The signals transmitted from each antenna is then taken as the sum of contributions for the two signals and the signals transmitted from all the antennas can then be expressed as:

$s = w_1 s_1 + w_2 s_s$

Here w_1 and w_2 are the beamforming weights for the two signals and the vector s represent the signals transmitted from all the antennas. Put another way, to transmit multiple signals, a sum of signals is transmitted from each antenna rather than a single signal. This most often means that contributions from all signals are transmitted from all the antennas.

Spatial multiplexing is further discussed in [Ch. 3, p. 27-33].

Summary

- Classical beamforming with *N* elements has a main lobe with maximum gain over a single element equal to *N* and width proportional to 1/*N*.
- An array can be partitioned into an array of subarrays, with short subarrays in small cells and tall subarrays in large cells.
- An array can be used to transmit multiple streams at the same time.

This chapter introduced concepts such as main lobe and side lobe which are used when discussing radiation patterns which is referred to as beams.

For classical beamforming,

- The maximum gain in the main lobe as compared to a single element is proportional to the number of elements in the array.
- The width of the main lobe is proportional to the inverse of the number of elements.

An array can be partitioned into subarrays to reduce the number of needed radio chains since high beamforming gain most often is only needed in a limited angular range.

- The larger the subarray, the narrower is the range of angles with high gain.
- In practice , vertical subarrays are used, and in large cells larger subarrays can be used and in small cells small subarrays can be used.

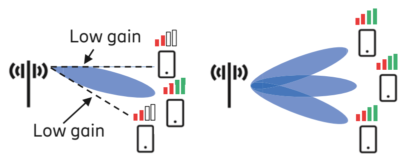
An array can be used to transmit several streams with different beams at the same time.

- Typically, all streams are transmitted from all antennas.
- Both to a single user as well as to multiple users (SU-MIMO and MU-MIMO).

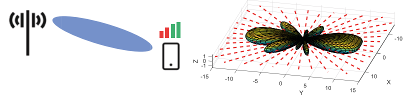
3. Multi-antenna technologies

Introduction

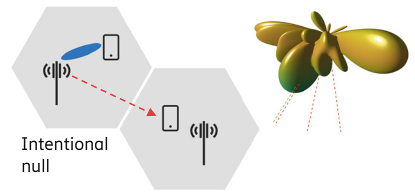
Static versus UE-specific beamforming



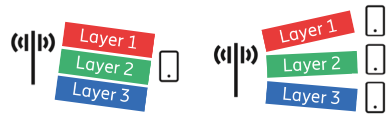
Beam-pattern based analysis and pitfalls



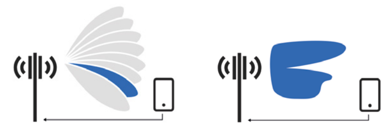
Null forming and beam pattern



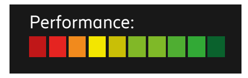
Spatial multiplexing for SU- and MU-MIMO



Channel info needed for beamforming



MU-MIMO – realistic expectations?



This chapter deals with miscellaneous topics concerning Massive MIMO technologies with the intent to give a more detailed understanding of selected topics than the more high-level exposition in the main story.

Static versus UE-specific beamforming: A key feature of a Massive MIMO system is its ability to tailor the beam pattern for each UE instead of as in classic systems where many UEs share the same beam. The difference between such UE-specific and static beamforming needs to be understood.

Beam-pattern based analysis & pitfalls: Beam-patterns are commonly used to describe transmission and reception techniques in Massive MIMO systems. It is a simple and intuitive tool but a source of much confusion if its limits are not fully understood.

Null forming & beam-pattern: Null forming is a key beamforming technique especially important for MU-MIMO to suppress interference towards victim co-scheduled UEs. The properties of the multi-path environment affect where co-scheduled UEs can be located relative to each other for efficient MU-MIMO operation. The quite intricate interaction between beam pattern, multi-path, null forming and MU-MIMO therefore deserves a more detailed treatment.

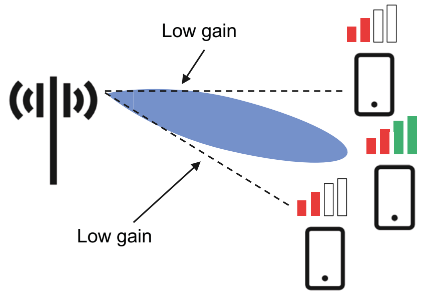
Spatial multiplexing for SU- & MU-MIMO: Spatial multiplexing can be achieved in several ways depending on the antenna setup and UE locations. Understanding how it works is important for making appropriate choices of technologies and deployment.

Channel info needed for beamforming: Information about the channel between the base station and the UE is needed to know where to beamform. Collecting such information can be done in various ways, largely categorized as codebook or reciprocity based. The two categories have different strengths and weaknesses.

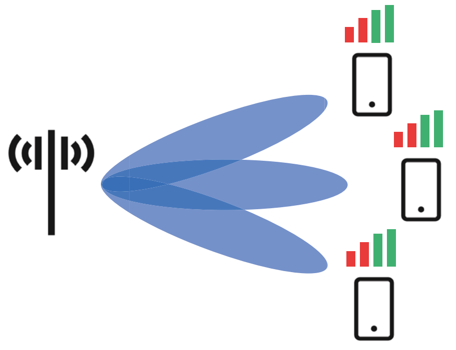
MU-MIMO – realistic expectations?: Marketing of MU-MIMO with many layers have at times been very strong. The use of demo setups in the industry, as well as analysis in the literature, easily give the impression that MU-MIMO is a ubiquitous technique with extraordinary benefits. Get to understand why industry demos and the literature tend to over-estimate the benefits due to assumptions which are rather far from realistic commercial network deployments. Setting realistic expectations and understanding in what subset of cells MU-MIMO makes a real impact is an important take-away.

Static vs. UE-specific beamforming

(Semi) Static beamforming
- The same fixed beam is shared by many users
  - All UEs in cell ➔ cell-specific beamforming
- Users often miss the beam peak – gain suffers
- Example: A fixed sector beam with some down-tilt



UE-specific beamforming
- Each UE gets a beam of its own
- No sharing so beam can be tailored for each UE
- Beamforming gain can be maximized



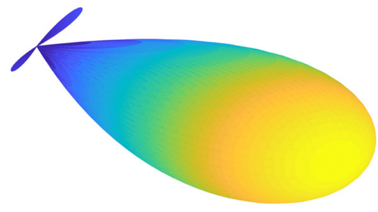
As mentioned in [Ch. 2, p. 14], the beam pattern can be static or vary more or less dynamically. Beamforming that dynamically changes beam pattern in time and frequency is typically exploited to provide so-called UE-specific beams that are tailored for each UE. In contrast, static beamforming has a beam pattern that is constant over time and is a reasonable simplification when all UEs in a cell can share the same beam. Semi-static beamforming allows the beam pattern to change over time, but substantially more slowly than dynamic beamforming so that it essentially is constant over long periods of time. Automatic adaptation of beam pattern for cell shaping
(c.f. [Ch. 5, p. 73]) is an example of the latter.

Static beamforming
The use of a static beam pattern in elevation is common in cellular networks. For example, a single-column antenna whose beam pattern is fixed and with a narrow main lobe in the vertical domain and wide sector covering in the horizontal domain is often used to serve all the UEs in a cell. To obtain good beamforming gains with a fixed beam system it is important that all the UEs that are intended to be served by the fixed beam are located in a direction which is close to a strong peak of the beam, otherwise beamforming gain quickly drops. With above-rooftop deployments and large cells this is often approximately fulfilled in the vertical domain for a relatively large portion of the UEs in a cell; as seen from the base station many UEs tend to be close to the horizon so the vertical UE angular spread is relatively small. The horizontal spatial UE distribution is however typically entirely different with often a wider and rather uniform UE angular spread over the sector. This means that in horizontal direction beamforming gain is lost with a static beam approach, either since the static beam must be made wide to cover all UEs or because a narrow beam misses most of the users.

UE-specific beamforming
A key benefit of Massive MIMO systems is their ability to form high gain UE-specific beams towards UEs of interest. Thus, there is possibility to tailor the beam pattern to suit a particular UE at each moment in time. This allows the beam to be narrow and still track movement of the intended UE. Base station side beamforming allows each UE in the cell to enjoy a specific beam pattern with a strong peak instead of being forced to share the same beam over many geographically dispersed UEs as in the static beamforming case. This leads to higher SINR levels and thus better overall system performance.

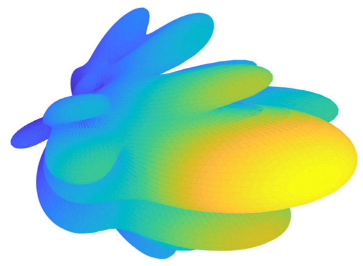
What is a beam? Two beamforming types

Classical beamforming



- Well-defined main lobe – single pointing direction
- Good for low angular spread – optimal for free-space

Generalized beamforming



- Does not look like a "beam" – may have many lobes
- Beamforms into all the propagation directions of the channel
- Optimal for multi-path channels with angular spread

Beamforming may be classified into the two categories of classical beamforming and generalized beamforming. Such a classification is useful to help resolve common misunderstandings concerning radiation/beam pattern and what constitutes a "beam".

Classical beamforming
Classical beamforming is easy to understand and agrees well with intuition. The beam pattern has a well-defined and often a relatively narrow dominating main lobe which can be pointed into a desired physical direction, for example from a base station towards a UE. Classical beamforming is the focus of traditional antenna theory, which is covered in more detail in [Ch. 2]. In free-space or line-of-sight propagation, classical beamforming is, or is near, optimal in the sense of maximizing signal strength achieving maximum gain. Such channels have (essentially) a single propagation path, i.e., zero channel angular spread, and no or little multi-path fading as discussed in [Ch. 1, p. 8)]. Thus, a beam with a single direction matches the channel very well.

Generalized beamforming
In contrast, generalized beamforming includes totally arbitrary beam patterns with potentially no obvious connection to propagation path directions in the channel. These beam patterns typically vary also over frequency to track the frequency-variations of multi-path channels, leading to so-called frequency-selective beamforming. There can be many lobes and it can be hard to identify a single dominating main lobe. A somewhat simplified explanation for this is that the beamformer now needs to emit power in all of the plethora of propagations paths and hence the beam pattern resembles the distribution of "channel directions". To be more precise is difficult, since the relative phases of the radio waves on the various propagation paths and how they combine now start to matter. As a consequence, analysis via beam pattern provides limited understanding on how the communication
link will perform.

With significant channel angular spread, classical beamforming is no longer optimal and leads at least in theory with ideal channel knowledge to an SNR loss as it is unable to match the fast-fading properties of the channel.

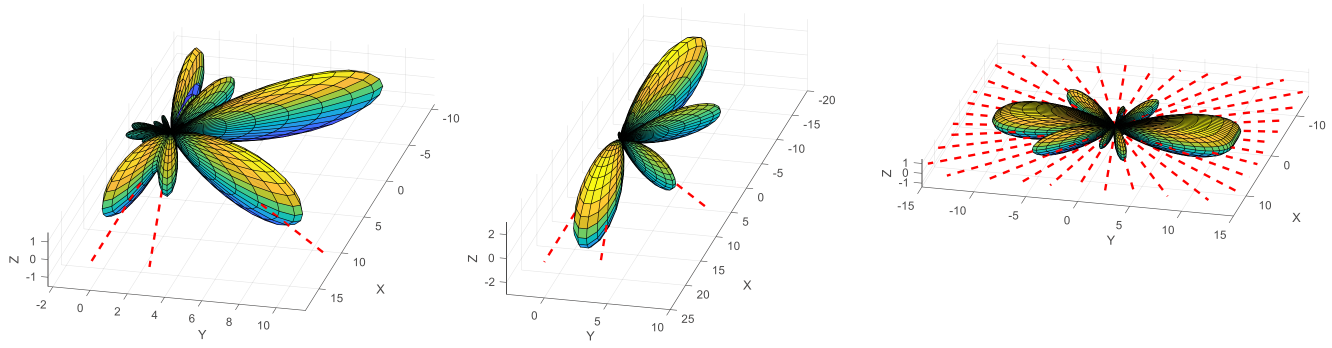
It is surprisingly difficult to give a good, relevant, and precise definition of what a beam is. An attempt to nevertheless do so is found in Section 5.5 of [1]. It suffices to say here, and roughly speaking, that a beam is used for transmitting/receiving a stream of symbols, such as a layer. So a layer is transmitted/received on a beam of it own. Note in particular that the shape of the beam, i.e., the characteristics of its radiation pattern, as is also evident from above not part of the definition, although the particular type of beams illustrated in this handbook generally have a corresponding radiation pattern.

Optimal beam patterns may have weak correspondence with propagation directions

Propagation directions: 0, 10, 40 degrees
1:st set of random propagation phases
Ok correspondence with beam peaks

Propagation directions: 0, 10, 40 degrees
2:nd set of random propagation phases
Bad correspondence with beam peaks

Propagation directions: uniform 360 degrees
3:rd set of random propagation phases
No correspondence with beam peaks at all



Phases matter but not visible in beam patterns!

It is reasonable to think that a beam optimized for a certain channel would clearly show the propagation directions as corresponding peaks in its beam pattern, in-line with classical antenna theory as in [Ch. 2, p. 12)]. This is however far from always the case as will be demonstrated next.

Consider a multi-path scenario with three distinct propagation directions and the use of generalized beamforming. In the left figure, the three lobes of the beam are seen to point roughly in those three propagation directions. This confirms the intuition that an optimal beam pattern should match the distribution of channel directions. Clearly, the use of classical beamforming in this case would not be optimal and lead to a lower SNR as it is unable to match multiple propagation directions at once.

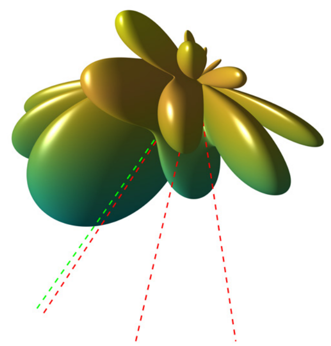
Unlike in a single-propagation path case, the lobe peaks do not exactly match the propagation directions. This is because the array has a finite resolution so is unable to completely resolve individual channel directions and in particular is unable to resolve directions that are very close to each other. Nevertheless, there is a fairly good and easy to understand correspondence between channel directions and beam shape.

However, consider now the middle figure, where again the same three propagation directions as in the left figure are used. Despite identical propagation directions, the beam pattern looks vastly different and no longer at all matches for two of the propagation directions where only a single lobe peak is present instead of two distinctly different peaks as in the left figure. This qualitatively substantial difference in optimal beam pattern is only due to the difference in the relative phases of the three propagation paths between the two figures. The relative phases between paths determine whether signals of those paths combine constructively or destructively, which may have a major impact on the beam pattern. But the phase characteristic is not observable in a beam pattern analysis, which greatly diminishes the value of such an analysis in cases of significant channel angular spread.

This analysis method completely breaks down in the extreme case of 360 degree uniformly distributed propagation paths as shown in the right figure. The beam pattern then has absolutely nothing to do with propagation directions of the channel.

What is a beam? Two beamforming types

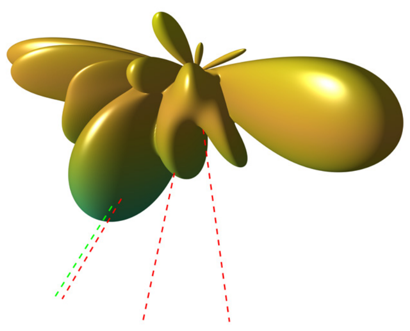
Classical null forming



- Angular spread is small
- A beam pattern null points towards a victim UE

Beam pattern completely informative!

Generalized null forming



- Angular spread not small, hence phases of paths matters
- Direction to a victim UE may not correspond to a beam pattern null

Beam pattern not informative!

Null forming is a technique for on purpose reducing the beamforming gain in certain "directions" so as to reduce interference at intended points in space. For example, reducing unwanted inter-cell interference stemming from transmissions from a base station that hits a cell-edge UE in a neighboring cell. Receive side null forming is very common both in UEs and in base stations while transmit side null forming is lately becoming increasingly popular in the base station primarily due to the introduction of MU-MIMO. Similarly, as for beamforming, null forming can be categorized into a classical and a generalized flavor.

Classical null forming
For classical null forming, the physical directions of the nulls in the beam pattern each point towards some victim UE. This make intuitive sense and corresponds to a scenario with no or very little channel angular spread. In such a scenario, null forming also ensures there is a zero or low signal level at the location of the UE. The left figure depicts the beam pattern of such classical null forming and it is seen how the red dashed lines, showing directions of three different victim UEs each align perfectly with a corresponding null in the beam pattern. This particular beam is targeting a UE in a direction along the green dashed line. It is seen that although the present scenario assumes a free-space channel, the direction of the UE of interest is not along the peak of the lobe, illustrating how null forming may decrease the beamforming gain for the served UE by offsetting the peak.

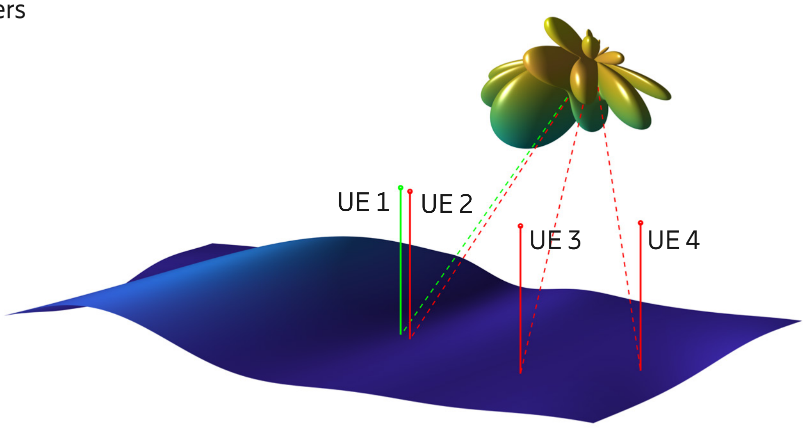
Generalized null forming
Generalized null forming is a potential component of generalized beamforming. Just like the latter, generalized null forming targets and supports also the multi-path propagation scenario. As expected, the beam pattern using generalized null forming may look rather arbitrary and non-intuitive. In fact, it may be even more chaotic than for generalized beamforming without null forming. Since there are several propagation paths, lobes in different parts of the beam pattern may form signals that, by carefully crafted phase characteristic, add destructively at a specific location in space. This relies on fast fading properties and therefore a null at one location in space may turn into a strong signal at another location just some fraction of a wave-length away. Null forming is therefore a technique which is sensitive to errors in the Channel State Information (CSI), including due to noise and doppler, and thus needs good coverage locations and low mobility.

More specific examples of classical and generalized null forming will be given on the two following pages.

Null forming for MU-MIMO – Free-space

- Want strong signal for UE 1 & nulls for the others
- Free-space propagation
- Channel does not distort signals
- Nulls on UE side and on BS side now coincide
  - Peaks as well
- Beam pattern based analysis totally relevant
- Sufficiently low angular spread required

Low angular spread makes things intuitive!



The figure considers a free-space scenario where MU-MIMO is used to transmit four layers, each to one of four different UEs 1 – 4. The blue surface represents the received power level, where the height of the surface is proportional to the received power level at the corresponding position in the horizontal plane. Although the received signal is a superposition of the four different transmitted layer signals, the figure only shows the transmitted beam pattern and received signal that is intended for UE 1, the UE marked in green. The transmitted signal spreads all over the horizontal plane as shown by the blue surface, potentially creating interference to the three other victim UEs 2 –4.

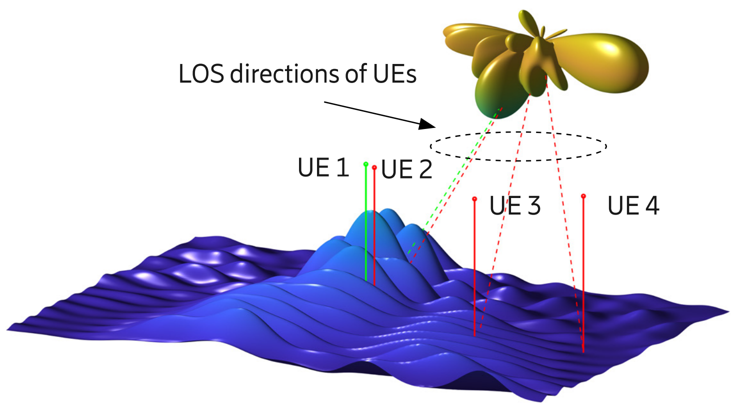
The interfering received signal contributions at the three victim UEs is however zero since null forming is used to form a beam pattern so that nothing of the transmitted signal intended for UE 1 leaks to the victim UEs. Due to the free space assumption, there is a perfect correspondence between the beam pattern and the resulting received signals and it is seen how each victim UE direction has a null in the beam pattern. Although the beam diagram looks much more complicated then for classical beamforming, it still makes intuitive sense.

The intention of null forming is to create a strong signal for the served UE 1 with small leakage to the victim UEs. However, it is here seen that UE 1 has a close to zero received signal. This is due to that UE1 is very close to UE2, and the latter has by the null forming design a zero received signal. It is in the present case physically impossible to have a strong signal at the UE 1 location when the requirement is to have a zero signal level just next to it at the UE 2 location.

Null forming for MU-MIMO – Multi-path

- Nulls on UE side but not always on BS side
- Thinking in terms of beam pattern is a useful tool
  - When its limitations are understood
  - Otherwise a source of much confusion
- No multi-path: beam pattern is intuitive
- Multi-path: beam pattern non-intuitive & arbitrary

Multi-path with MU-MIMO leads to real mess!



This figure illustrates exactly the same setup as in the previous page, but now in a scenario involving significant multi-path. Thanks to the null forming, the three victim UEs 2– 4 still have zero received interfering signal. However, the beam pattern is no longer a good predictor for what really matters – the received signal levels. Although UE 4 has indeed a null in the beam pattern, UE 2 and UE 3 instead have peaks! This may seem confusing, but due to the multi-path several strong lobes on the transmit side may now convey the signals that after propagating over the channel combine at the UE location to cancel each other. The phase characteristic of the beam now plays a crucial role to achieve cancellation but is not observable from the diagram.

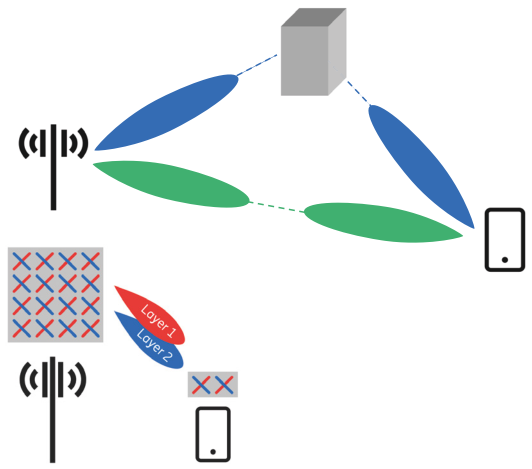
Although hard to discern in the figure, the served UE 1 has now in fact a strong received signal, despite the proximity to the null at the UE 2 location. The presence of multi-path again provides the explanation. Just as signals from different lobes may now follow different propagation paths and still reach the same victim UE and cancel each other, with proper phase characteristics signals can now constructively combine at UE 1 location. Roughly speaking, since a propagation path reflection point is typically very far away from the transmitting antenna array, from the perspective of reception the reflection point acts as creating a very large transmit antenna array and such a large array can produce very narrow lobes (c.f [Ch. 2, p. 14]) which avoid the problem of a nearby null necessarily limiting the signal strength.

Clearly, the beam pattern is in this multi-path case quite misleading. The use of MU-MIMO involving null forming tends to result in messy beam patterns to an even larger extent than what's caused by the multi-path alone. Although a beam pattern based analysis is an often used and useful tool, its limits need to be understood. Otherwise, it is a source for much confusion and possible harm. As an example, there is a risk of over-emphasizing the need of requirements on various beam pattern properties, with suboptimal designs for reality as a potential negative consequence.

SU-MIMO spatial multiplexing

How does it work?

- Several layers to the same UE
- "One layer/beam per significant propagation path"
- Channel angular spread needed for more than 1 layer per polarization
- Dual-layer transmission in LOS on separate polarizations important use case
- Needs at least as many tx/rx antennas as layers
- Fast fading based null forming can reduce inter-layer interference



Ubiquitous benefits for both coverage and capacity!

Spatial multiplexing for SU-MIMO is a way of increasing system capacity via improved link performance by multiplexing several layers of data streams on the same time/frequency resources.

To roughly understand the underlying technical principle, for simplicity, temporarily assume single-polarized antenna setups. A possible mental model of spatial multiplexing where all the layers are transmitted to a single user is that each layer is transmitted on a separate beam and the beam is matched towards a particular and significantly strong propagation path. This is illustrated in the top figure and it is easy to understand that transmitted layers can be separated out on the receive side by means of receive beamforming (spatial filtering) if those layer-carrying paths have different arrival angles. Thus, one layer/beam per propagation path and you need multiple such paths to be able to convey multiple layers – channel angular spread needs to be sufficiently large. The maximum number of layers is ultimately limited by the number of base station/UE antennas, whichever is smallest. Note that the beams here, and in most other places in the handbook, are drawn as classical beams, even though they often may be of generalized nature.

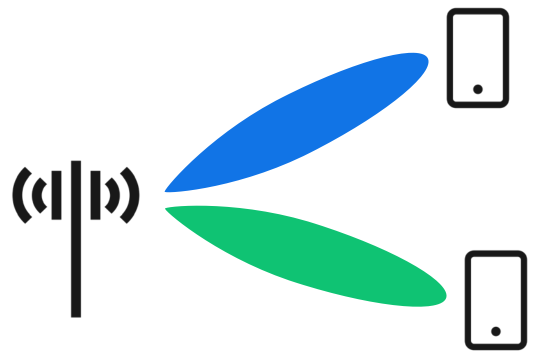
The above mental model holds well if directions of the propagation paths are very well-separated relative to the beam widths. This is however rarely the case and hence the beams can overlap in arbitrary ways and the separation of the layers to a large degree instead hinges on fast fading properties of the channel where phase relations are crucial but hard to understand via a physical direction-based analysis (c.f. the limitations of beam pattern analysis for generalized beamforming/null forming). Generalized null forming is then a key tool in separating the layers so as to reduce inter-layer interference. Receive side such null forming is basically always used while in more advanced cases generalized beamforming/null forming is also used for transmission, e.g. in the downlink from the base station with reciprocity based TDD or with advanced codebook-based feedback [Ch. 4, p. 48].

In practice, it is common to use dual-polarized antenna setups at both the base station and UE sides, see the bottom figure which shows dual-polarized antenna pairs with ±45 degree polarizations. Two layers can then be efficiently multiplexed even in the single propagation path case such as with free-space channels, one on each polarization. Since signals with orthogonal polarizations are well-isolated from each other in a near free-space setting, there is little inter-layer interference and SINR levels may be high. There is often significant isolation between orthogonal polarizations also in non-line-of-sight channels. Spatial multiplexing via exploitation of dual polarizations is thus a key technique of great importance and applicability. It also means that for each significant propagation path in the simplified mental model, up to two layers can be transmitted efficiently if dual-polarized antennas are employed on both sides of the link.

MU-MIMO spatial multiplexing

How does it work?

- Several layers/beams to multiple UEs
- Different layers may go to different UEs
- UEs directionally well-separated: even classical beams may have limited cross-talk
- UEs closely located: rely on base station side generalized null forming
- Several layers within same polarization possible even with low angular spread



Capacity benefits in cells having high load with high SINR levels!

Spatial multiplexing for MU-MIMO is a way of increasing system capacity by multiplexing several layers of data streams on the same time/frequency resources serving multiple users.

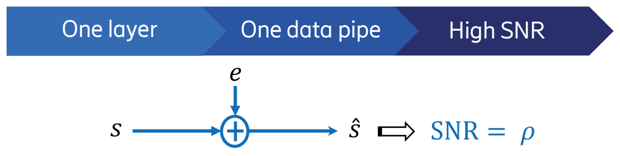
MU-MIMO is easier to understand than SU-MIMO in the most common case when multiplexed UEs are sufficiently far apart for them to be directionally well separated as seen from the base station. The depicted obvious and intuitive mental model of how the layers/beams towards different UEs have little cross-talk is then applicable. The mental model holds if the channel angular spread is sufficiently low. Even classical beamforming without any base station side null forming may under such favorable conditions be successfully used.

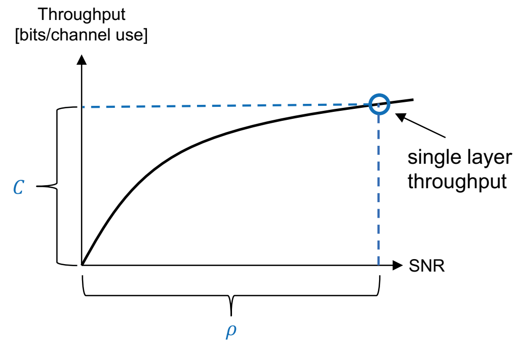
As multiplexed UEs come closer to each other it becomes more challenging to support multiplexing of several layers. The situation starts to resemble the SU-MIMO case with a UE with as many antennas as the sum of antennas over all multiplexed UEs, but with the crucial limitation that the multiplexed UEs cannot process the received signals jointly across their combined set of antennas. This severely limits the null forming capabilities on the UE side. However, spatial multiplexing can nevertheless be supported also in this case if there is sufficient multi-path and the base station uses generalized null forming on its side to reduce the inter-layer interference between the UEs.

Similarly, as for SU-MIMO, MU-MIMO requires sufficiently high SINR. However, an additional requirement is that the traffic load within the cell needs to be very high so that it is likely that at least two UEs in the same time slot wants to be served with data. Since high network load often means SINRs become lower due to more inter-cell interference, the two requirements are somewhat in conflict with each other. This makes MU-MIMO a capacity enhancing feature for a subset of the cells – those that simultaneously have high load and many high SINR UEs.

No spatial multiplexing may lead to inefficiency

Single layer





- All transmit power allocated to single layer
- High SNR level for that single layer

- Throughput saturates for high SNR (flat throughput slope at high SNR)

An increase in already high SNR only gives a small increase in throughput!

To understand the basic principle behind the gain of spatial multiplexing and the need for high SINR levels, it is instructive to consider a comparison between single and dual layer transmission. First, consider as depicted at upper right a link where a single layer is transmitted creating a single data pipe. The modulation symbol *s* is transmitted over an additive white Gaussian noise (AWGN) channel to produce a noise impaired received symbol $\hat{s}$ that experiences an SNR level of ρ. The total transmitted power is *P* and will subsequently be kept constant regardless of the number of layers. This is a highly simplified model of a communication link but will nevertheless serve the purpose of illustrating the principal differences between single-layer and multi-layer spatial multiplexing.

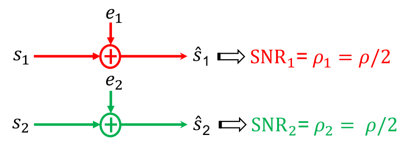
The upper right figure shows how the data rate throughput increases with an increasing SNR level ρ. It is observed how the slope of the throughput curve changes from being steep at low SNR levels while flattening out at high SNR levels. In particular, the throughput saturates for high SNR meaning that even if the SNR would be increased further from an already high SNR level, there is very little additional throughput gain. Beamforming increases the SNR and hence as the beamforming gain improves of a Massive MIMO system with larger and larger antenna arrays, the pay-off in terms of increased throughput becomes smaller and smaller.

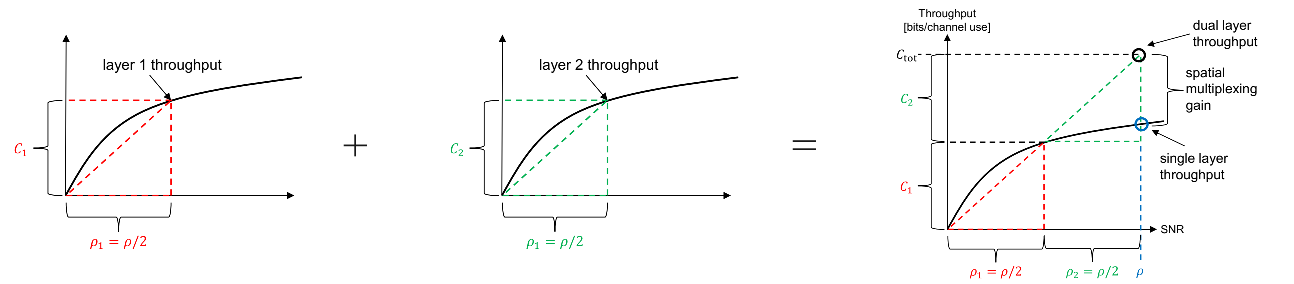
It would be unfortunate to not be able to benefit from high SNR levels to substantially improve user experience via better data rate throughput. As will be demonstrated on the next page, spatial multiplexing is a way of exploiting high SNR levels for obtaining data rate improvements.

Why spatial multiplexing provides gain at high SNR

Example: single versus dual layers

Transmit power shared by two layers





Net gain: More data pipes compensate for lower SNR per data pipe!

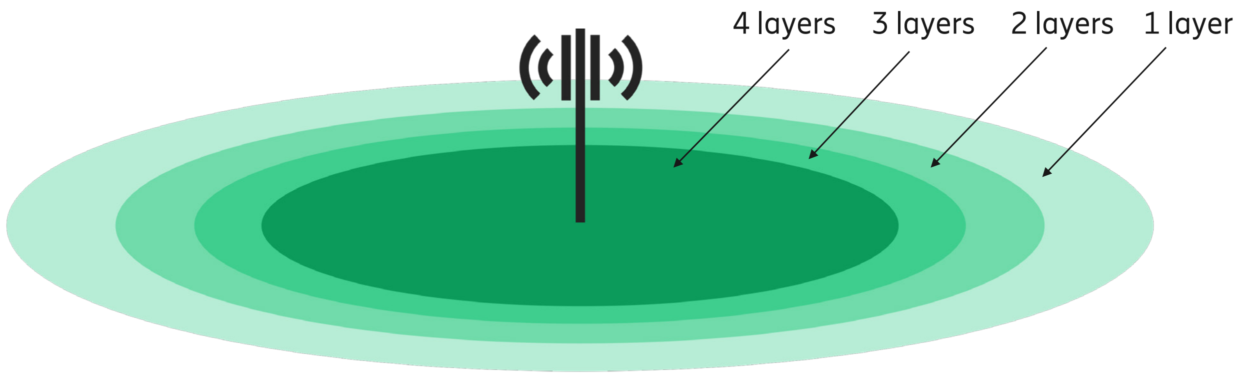
Consider as depicted a link with two layers, where each layer *k* transmits a modulation symbol s_k over an additive white Gaussian noise (AWGN) channel to produce a noise impaired received symbol $\hat{s}_k$ that experiences an SNR of ρ_k. For simplicity, cross-talk between the two parallel AWGN channels is neglected and thus the two layers are transmitted on two independent "data pipes". The total transmitted power *P* is the same as in the single layer case on the preceding page. Hence, the transmit power for each layer is *P/r*, where *r* is the number of transmitted layers. In other words, the transmit power per layer decreases as the number of layers increases. In the above example it is seen how the single layer SNR of ρ is halved to $\rho_k=\rho/2$ for each layer in the two-layer case. Thus, the SNR, and hence the data-rate, for each layer is lower than in the single layer case. There is however two layers instead of only one, so it is conceivable that the sum data rate over the two layers could be larger than the data-rate in the single-layer case.

To see that indeed the reduction in per layer data-rate is more than compensated for by the increase in number of parallel layers/data-pipes, consider the bottom graphs. The two left graphs shows how data-rate depends on the SNR for each of the two layers separately. The throughput for layer 1 and layer 2 are identical and are summed together to produce the dual layer throughput as illustrated in the right-most graph. As previously explained, the data-rate for a single-layer (black curve) is seen to saturate as its SNR=ρ increases – trying to increase an already high SNR even further does not pay off much in term of data-rate increase. For the two-layer case it is seen how the red and green layers each get half the SNR=ρ_k=ρ/2 via power sharing and as expected the throughput C_k for an individual layer is less than in the single layer case. However, the sum data rate C_{tot}=C_1+C_2 is larger! This since the individual layers operate on a steeper lower SNR part of the throughput curve. In essence, a spatial multiplexing gain is achieved by sharing a "too high SNR" among more layers.

This simple special case illustrates the basic principle explaining how spatial multiplexing achieves its benefits and why it thrives on high SNR. The AWGN model of the link is obviously over-simplified and idealistic, but the main qualitative conclusion still applies to more realistic models.

In a more realistic setting inter-layer interference would act to decrease the per layer S(I)NR compared with the analysis above. This means that using too many layers in relation to the SNR may incur a loss in sum data-rate, instead of as above always a gain. Carefully adapting the number of layers to the channel conditions at hand thus becomes important.

Number of layers decrease as cell edge is approached



The previous two pages explained why spatial multiplexing can give a throughput gain. For specific channel conditions, there is an optimal number of layers to use, and that number tends to increase with the SINR level.

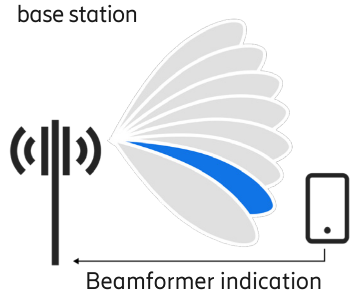
In a cell, SINR tends to decrease with distance to the base station. Consequently, the optimal number of layers is decreasing as the cell edge is approached. Using many layers is beneficial close to cell center while fewer, or only one layer, is used at cell-edge.

Efficient wireless systems perform dynamic adaption of the number of layers depending on many factors, including the channel conditions and number of UEs that have data in the buffer at a particular time instant. Inter-cell interference acts as adding noise and hence the SINR levels in a cell tends to vary with the load of the neighboring cells. As the load in neighboring cells increases, the optimal number of layers to use in the cell at a particular location is reduced.

Determining transmit beamformer/precoder from channel state information (CSI)

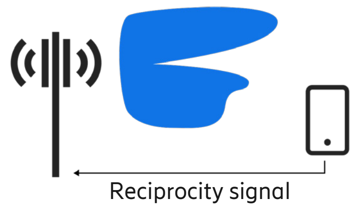
Codebook based feedback

- UE measures the channel based on a base station transmission
- Measurements ➔ UE selects beamformer from a codebook
- Indication of recommended beamformer sent to base station



Reciprocity

- Base station measures uplink channel based on a UE transmission
- DL channel properties inferred from uplink channel
  - TDD ➔ UL & DL channels are the same (reciprocity)
- Base station uses DL channel for computing a beamformer



Channel knowledge, or so-called channel state information (CSI), is needed to know where to beamform. There are two different approaches to CSI acquisition for transmission purposes.

Codebook based feedback
In codebook based feedback, the codebook corresponds to a finite and countable set of predetermined precoders/beamformers. Precoding is a popular and more specific term for beamforming in the transmit direction. It is the UE that based on channel estimates determines a precoder from a codebook and thereafter feeds that precoder back to the base station so that the base station can use that precoder recommendation when determining how to precode its transmitted signals towards the UE:

1. UE estimates the channel based on receiving known reference signals from the base station
2. UE selects a precoder matching the estimated channel properties from a codebook
3. UE feeds back the selected precoder(s) as a recommendation to the base station
4. The base station uses the recommended precoder in its transmission towards the UE, or uses it as one source of input in proprietarily determining a precoder for the transmission.

Codebooks are often based on grid of beams (GoB) precoders and as such codebook based feedback tends to be related to classical beamforming.

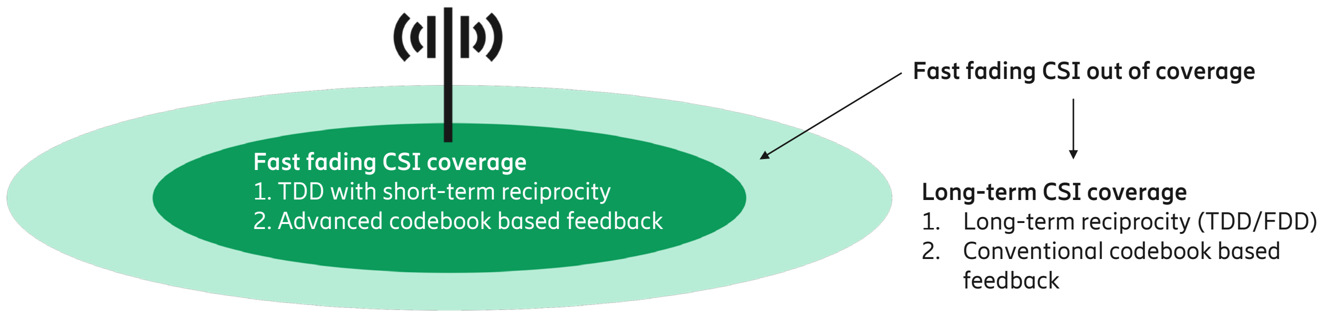
Reciprocity
The reciprocity approach exploits that the channels in both the forward and reverse directions are more or less similar. The transmitting base station can therefore do channel estimation from reverse link measurements and use that to determine a precoder in the forward link towards the UE:

1. Base station estimates the uplink channel based on a UE transmission of known reference signals
2. Base station uses uplink channel estimate to infer properties of the downlink channel
3. Base station uses downlink channel to determine a precoder to use in its transmission towards the UE

Reciprocity based techniques tends to lead to beam patterns similar to generalized beamforming.

For TDD, uplink and downlink are on the same frequency and hence the corresponding channels are identical. The inferred and actual downlink channels thus match very well, including their fast fading properties. Reciprocity also works for FDD but in a more limited way. Uplink and downlink are then on different frequencies so the inferred downlink channel may capture reciprocal large-scale (/long-term) channel properties such as propagation directions and path loss, but not the fast fading. In either case, reciprocity requires base station tx/rx antenna calibration.

Where to use different CSI acquisition methods



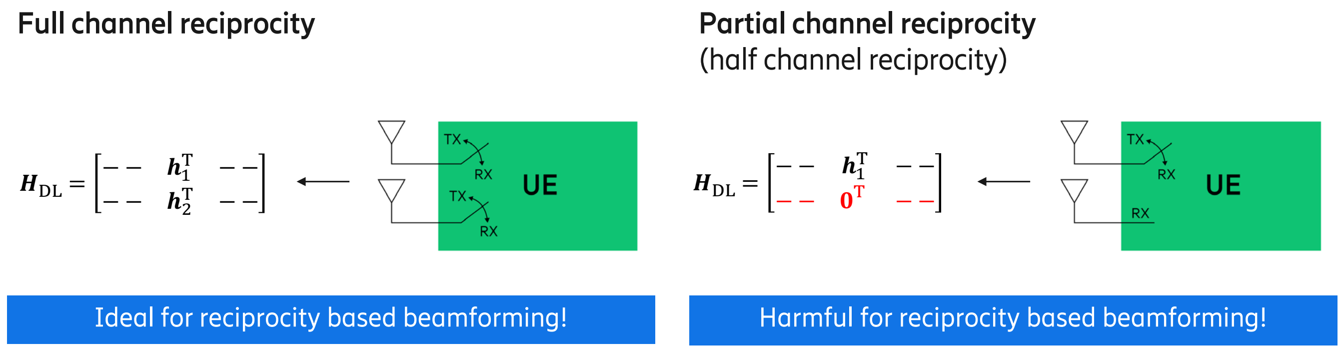
CSI can come in different detail levels. In particular, there can be CSI that includes short-term (/small-scale) channel properties or CSI that only contains long-term (/large-scale) channel properties. Reciprocity for TDD capturing fast fading properties is an example of the former while reciprocity for FDD is an example of the latter.

Including short-term properties provides more details (constituting more "information") and hence puts stronger demands on a good signal level to be able to reliable extract a larger amount of CSI. Thus, schemes relying on short-term CSI usually have smaller coverage than schemes that only rely on long-term CSI. On the other hand, performance is usually better for short-term based schemes when coverage is sufficient to support them.

Reciprocity for TDD is one way of obtaining short-term CSI. Advanced codebook-based feedback as discussed in Chapter 4 [Ch. 4, p. 48] is another. More conventional codebook-based feedback primarily focuses on providing long-term CSI. Similarly, reciprocity for FDD is also limited to providing long-term CSI. However, for TDD, using long-term CSI is more of a voluntary option that may be beneficial where coverage is bad.

These observations lead to the conclusion that it can be beneficial for the base station to switch CSI acquisition scheme depending on where in the cell the UE of interest is. The farther out in the cell a UE is, the more likely it becomes that a simpler, but more robust, long-term CSI scheme is preferable.

Partial channel reciprocity limits performance



In NR and in LTE, UEs may have more receive antennas than transmit antennas. This presents a problem for reciprocity because uplink channel measurements will then only provide part of the whole downlink channel, so-called partial channel reciprocity. The figure above illustrates the difference between full and partial (half) channel reciprocity. Having only parts of the channel for determining downlink precoders in the base station significantly impairs the performance since the unknown part of the channel prevents an effective precoder to be formed for all of the UE receive antennas. It may even result in reciprocity ceasing to be a competitive technology.

While it may be prohibitive for many UEs to have the same number of radio chains for transmit as for receive, some UEs support transmit antenna switching so that which of its antennas are used for transmission can be dynamically switched from one time instance to another. This allows the sounding reference signals commonly used for reciprocity measurements to alternate between transmitting on the various UE antennas. Full channel reciprocity can in this way be achieved despite having fewer radio chains (even only one) in the transmit direction. Potential drawbacks include increased sounding overhead or time to acquire all channel measurements as well as some insertion loss due to the switch.

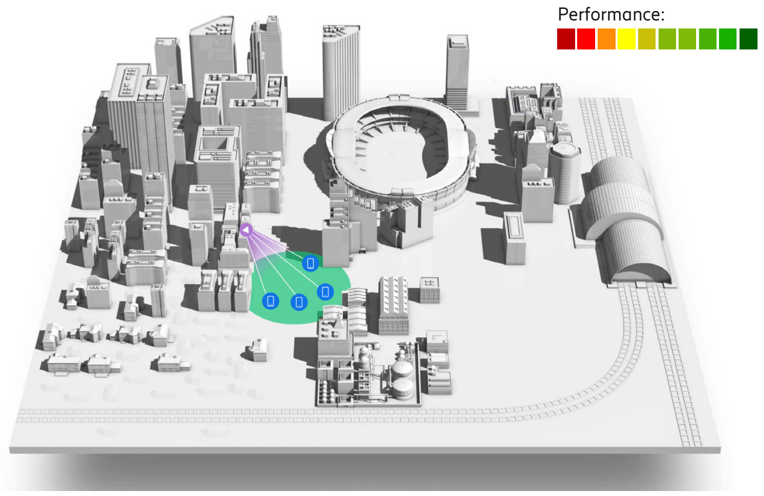
For LTE, the UE norm is to have two receive antennas and one non-switchable transmit antenna. Suffering from half channel reciprocity is thus common. In NR, the situation is less clear and it will take time to be able to judge what the majority of UEs support. In the present still early phase of NR, it seems UEs commonly have four receive antennas and possibility to switch transmit antenna(s) for sounding reference signal purposes. This avoids the partial channel reciprocity problem for the present UE population, but this may change as the UE population inevitably shifts from high end to lower end UEs.

MU-MIMO – realistic expectations?

Easy to get too high expectations...

Industry demos

- Single cell
- Full buffer traffic
- All connected UEs assumed dynamically active
- Real-life practical issues often neglected!



MU-MIMO is clearly an important technique for boosting the capacity in highly loaded cells. The technique has been seen as the essence of Massive MIMO ever since its introduction in the literature. Extraordinary performance gains over SU-MIMO have been demonstrated both in the academic literature as well as in industry demos. Expectations have therefore often been set very high and it is easy to get the impression that adding more and more layers is always key to good performance and would offer a factor of 10 or more in performance boost. In contrast, realistic simulations and trial results show good gains of MU-MIMO in certain scenarios, but the gains are more on the order of a factor of 1.5 or somewhat more (c.f. generalized vs. classic beamforming for 64 T in [Ch. 6, p. 85]).

There is thus a risk that the expectations of MU-MIMO are unrealistically high. There is in fact an abundance of differences between real-life commercial wireless networks and what is typically being assumed in industry demos and academic literature. These differences unfortunately tend to all artificially boost the relative gain of MU-MIMO over SU-MIMO compared with what will be the case in realistic settings. This and the next page will provide more details on why there is a tendency to get unrealistic MU-MIMO gains in industry demos as well as in the academic literature.

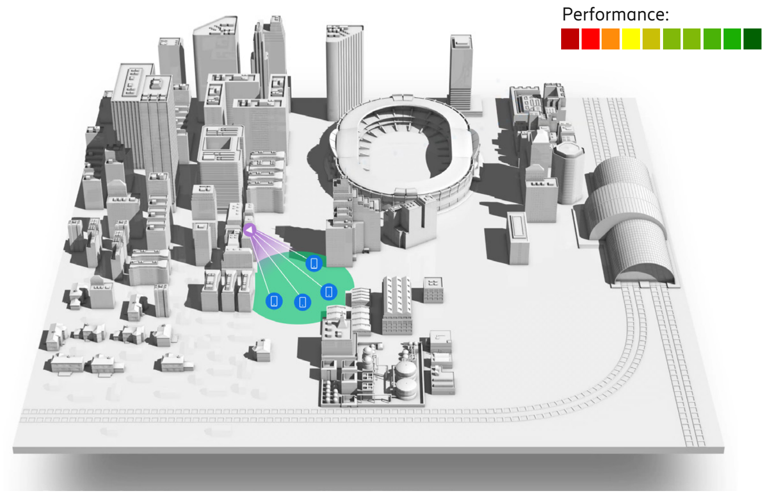
In industry demos, it is very common to use only a single cell and therein manually place, often in MU-MIMO favorable high SNR locations, a bunch of UEs. This is illustrated above and neglects that in real networks there is often substantial inter-cell interference that will lower the SINR levels and thus the gains of MU-MIMO with many layers over fewer layers SU-MIMO. Also, the data traffic is assumed to be full-buffer, meaning that all UEs want infinite amount of data in every time slot. In reality, the data is packet based and often very bursty with significant silent periods between the packets. This reduces the likelihood that many UEs want data in the same time instance and thus limiting the amount of MU-MIMO layers needed compared with the unrealistic full buffer assumption. Also, in real networks there is often a strong tendency that the set of UEs that want data in the same time slot is dynamically changing over time, making it hard to provide the necessary accurate channel knowledge in advance of the scheduling of the UEs.

MU-MIMO – realistic expectations?

Easy to get too high expectations...

Academia

- An extreme channel model: IID Rayleigh fading
- SU-MIMO crippled UE: single- antenna or layer UE
- Full channel reciprocity: #TX = #RX antennas for UE
- Optimistic quality of reciprocity based CSI
  - Narrowband/SINR/frequency
  - Relying on UE Tx inter-slot phase coherency
- Real-life practical issues often neglected!



The academic literature in general strives for closed-form analysis. As such, several simplifying assumptions are often made as discussed below.

It is very common to use a Rayleigh fading channel model with spatially uncorrelated fading. This model has no notion of "direction", all channel directions are equally likely. In contrast, real channels have a significant spatial correlation and have widely different characteristics vertically and horizontally. Many UEs tend to be confined in a small angular range in the vertical domain with high spatial correlation and this seriously limits the multiplexing potential for MU-MIMO.

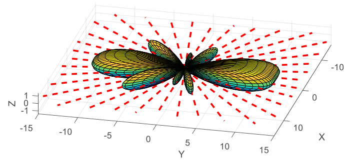
Academic studies commonly assume all UEs have only a single receive/transmit antenna, which restricts SU-MIMO to only a single layer and removes possibility to do inter-cell interference rejection in the UE. This cripples in particular the SU-MIMO performance, thereby artificially increasing MU-MIMO over SU-MIMO gains.

The assumption of a single receive/transmit antenna at the UE also means that the so-called partial channel reciprocity problem does not exist. This greatly benefits reciprocity based schemes since no channel information is missing and MU-MIMO in particular relies on accurate channel knowledge, while SU-MIMO is less demanding when it comes to quality of channel knowledge.

The quality of reciprocity based channel estimates may also be too optimistic as often the assumed bandwidth is small so uplink SINR per frequency becomes high. There may also be filtering of channel estimates across time-slots, something that requires that UEs support phase coherency over time, which current commercial UEs do not.

Summary

- A beam may not look like a beam



- Beamforming for improving signal levels is the main component in massive MIMO performance

- MU-MIMO important capacity enhancer but its gains are often exaggerated

This chapter addressed miscellaneous technology components essential to Massive MIMO in greater detail than in the main story. There are many conclusions and findings in the chapter. The three most important take-aways can be summarized as:

- A beam may not look like a beam, i.e., like a single directed entity as implied by intuition. It may in fact look like an arbitrary blob when there is significant angular spread in the channel and/or MU-MIMO is being used. Yet for simplicity, and to focus on the concept of beamforming, illustrations typically show very well-behaved beams with a strong dominating main lobe (also this handbook does that). Confusion arises when this leads to a belief that this is how a beam is supposed to look like while in reality the exact pattern of the beam is often much less relevant.
- Beamforming for improving signal levels is the main component in Massive MIMO performance. The large beamforming gain from large antenna arrays and use of UE specific beamforming ensures great coverage and facilitating reuse of a site grid for higher frequency bands. The resulting increase in SNR levels is also a prerequisite for techniques such as MU-MIMO to thrive.
- MU-MIMO is surely an important capacity enhancement technique for those cells having very high load. It is possible to demonstrate extraordinary performance gains on the order of a magnitude or so under very specific artificial conditions. Academic literature and artificial industry demos have consequently set very high expectations and Massive MIMO has for this reason been thought to be synonymous with MU-MIMO. The specific conditions needed for achieving these kind of gains are however rarely found in realistic network settings and thus MU-MIMO capacity gains in practice are considerably less, although still good and often around a respectable 30-50%.

4. 3GPP – Physical layer support for Massive MIMO

3GPP Technology Primer

The purpose of the following pages is to give some background of relevant parts of the 3GPP physical layer specifications

The purpose of this chapter is to:

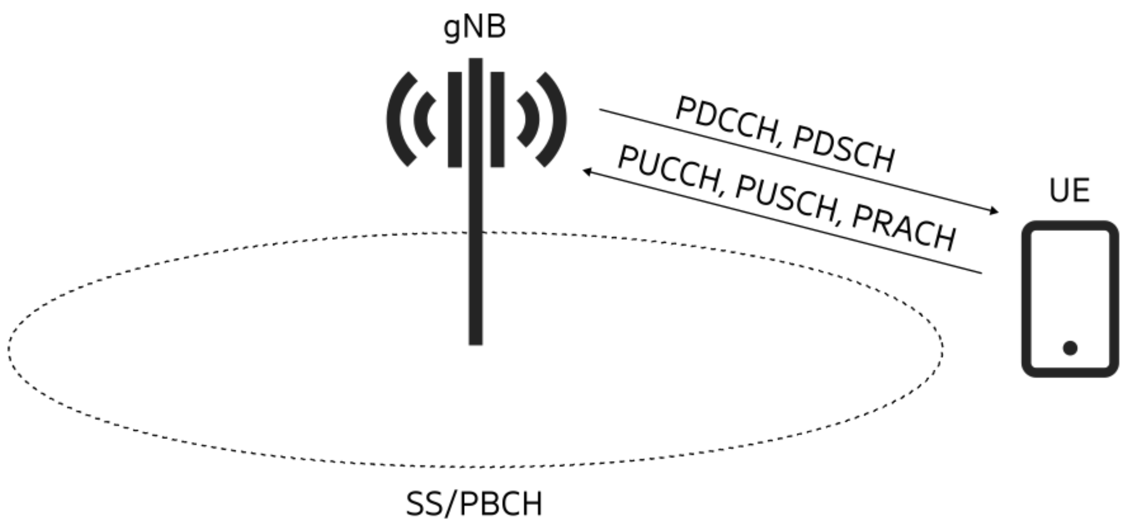
- Give some background of relevant parts of the 3GPP physical layer specifications.
- Assist the reader to understand terms and concepts for the sections feature solutions and network performance.

The content and motivation/relevance for Massive MIMO:

- 3GPP physical channels..
  - Key to understand the physical layer of NR.
- Carrier aggregation (CA) and dual connectivity (DC).
  - The combination of Massive MIMO and carrier aggregation is common, and it is important to at least grasp the fundamentals of carrier aggregation.
- Cell defining signals and control channels.
  - This part is done differently in NR compared to LTE due to introduction of beam based initial access by Massive MIMO.
- Channel State Information (CSI) feedback principles and types including MIMO codebooks.
  - This is a key area for Massive MIMO performance and understanding the different features.
- Sounding Reference Signal (SRS) and antenna switching.
  - For reciprocity based Massive MIMO, the uplink RS is important, and the basics is covered here.
- Periodic, semi-persistent and aperiodic behavior of CSI reporting, SRS, CSI-RS.
  - The 3GPP specifications allow a large flexibility to configure time behavior of measurements and reporting, and this provides a brief overview.

For a more extensive description of the 3GPP physical layer standard related to AAS (Massive MIMO), refer to [1].

Summary of the physical channels in NR



In order to understand the different physical channels in NR, which are combined with the Massive MIMO technology components in [Ch. 3], an overview is given here. See also the 3GPP specifications TS 38.211, TS 38.213 and TS 38.214 (3gpp.org).

The three downlink channels in NR are:

- Physical broadcast channel (PBCH) is a broadcasted channel (potentially received by all UEs) which carry small but essential part of system information needed to access the cell.
- Physical downlink control channel (PDCCH) carries the downlink control information (DCI), used for scheduling PDSCH and PUSCH etc.
- Physical downlink shared channel (PDSCH) carries the downlink data.

The three uplink channels in NR are:

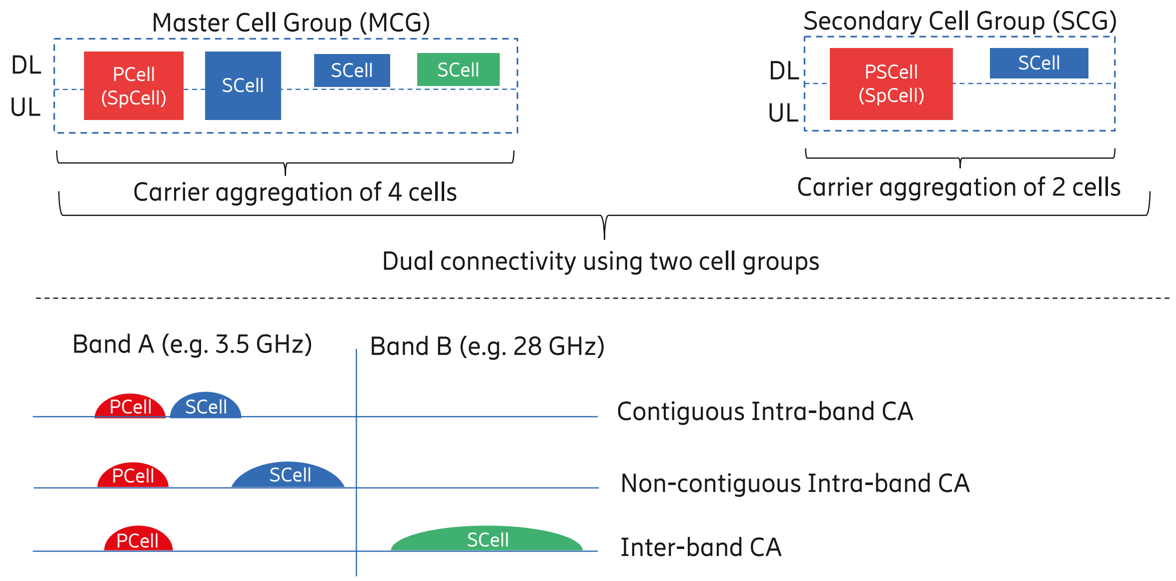
- Physical uplink shared channel (PUSCH) carries the UE uplink data and/or aperiodic CSI feedback.
- Physical uplink control channel (PUCCH) carries the uplink control Information (UCI).
- Physical random access channel (PRACH) which is used for random access and to adjust uplink timing.

The PBCH is transmitted adjacent to the synchronization signals (SS) and carries a few bits with info on how to access the cell at initial access. The PBCH+SS (known as an SSB) covers the whole cell, or a portion of the cell, in case beamformed SS/PBCH is used.

The channels have different coverage (in some cases depending on the information payload and used code rate) and these differences needs to be taken into account when planning the deployment of the network.

Note that NR also support side-link and other recent features, for which additional physical channel has been added in later NR specification releases.

Used terminology in carrier aggregation



Carrier aggregation (CA) is an important feature for Massive MIMO as CSI reporting can be configured to carrier with better uplink coverage. In Chapter 3, Solutions, the Beamforming/carrier aggregation mode switching is one example where carrier aggregation is combined with Massive MIMO. Here, the basics of carrier aggregation is introduced, see TS 38.300 and TS 38.331 (3gpp.org) for the relevant specifications.

In NR, "cell" is used to denote a carrier in carrier aggregation and up to 16 such serving cells can be configured to the UE using non-overlapping frequency bandwidths. A serving cell can either be a special cell (SpCell) or a secondary cell (SCell). Furthermore, the SpCell is either a Primary Cell (PCell) or a Primary SCG Cell (PSCell). Frequency bands are defined and there is a distinction of intra-band CA (e.g. among carriers at 3.5 GHz carrier frequency) and inter-band CA (e.g. between carriers at 3.5 GHz and 28 GHz carrier frequencies).

In addition, in intra-band CA, there is a distinction between contiguous and non-contiguous CA, where in the latter, the carriers are not adjacent in frequency.

A Primary Cell (PCell), is a SpCell/PSCell and it always has an uplink which also carries uplink control information (UCI) feedback (e.g. HARQ-ACK,CSI) for the Pcell and by default, also UCI for all SCells unless configured differently.

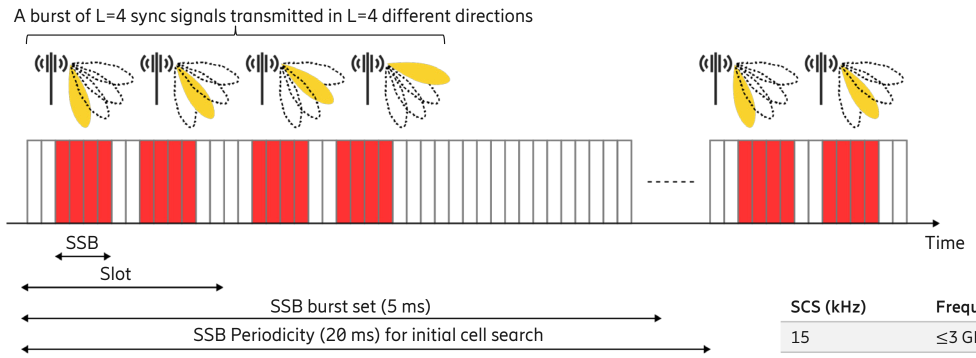
- PCell handles initial access/receives the Random Access Channel, handle Radio Link Monitoring (RLM).
- The cell that is the PCell can be reconfigured for the UE, using the handover procedure which takes about 40 ms.

A Secondary Cell (SCell) may not have an uplink (for a given UE) and may or may not be configured with SS/PBCH. If not configured, the UE use SSB of another cell.

- May have a different SCS and duplex mode compared to the Pcell.
- For example, in TDD, an SCell can be downlink only and the UE cannot transmit SRS for reciprocity-based operation. If supported by the UE, SRS Carrier Switching can be used in this case.
- An SCell can rapidly be activated or deactivated depending on the need (latency is a few ms if MAC CS is used (Rel.15) or less if DCI is used (Rel.16)).

There is also a possibility to configure a supplementary uplink (SUL), where a downlink carrier have two uplink carriers. If dual connectivity is configured, a master cell group (MCG) and a secondary cell group (SCG) is defined. Both the MCG and SCG has a "PCell", and it is defined as PSCell for the SCG. The PSCell handles random access and handle RLM just as the PCell in the MCG.

Overview of NR synchronization signals



SCS (kHz)	Frequency range	L_{max}
15	≤3 GHz	4
15	3 GHz to 7.125 GHz	8
30	≤3 GHz (paired spectrum)	8
30	≤2.4 GHz (unpaired spectrum)	4
30	2.4 GHz to 7.125 GHz (unpaired spectrum)	8
120 or 240	24.25 GHz to 52.6 GHz	64
120, 480 or 960	52.6 GHz to 71 GHz	64

L_{max} is the maximum number of supported SSB beams per frequency range and subcarrier spacing

The cell-defining signal is the signal or set of signals the UE detects from an SpCell and use as the basis to begin access procedures to establish a connection to a cell. If these cannot be reliably received, then the UE is outside that cell area.

In NR, the cell defining signals are the synchronization signals (SS) together with the broadcast channel (PBCH) and the system information block 1 (SIB1) which is carried on PDSCH.

The cell defining signals is specified differently in NR compared to LTE since a cell can be configured with multiple cell defining signals, which are time multiplexed and transmitted in different beam directions. The maximum number of such cell defining signals Lmax depends on the subcarrier spacing (SCS) and frequency range, see the table.

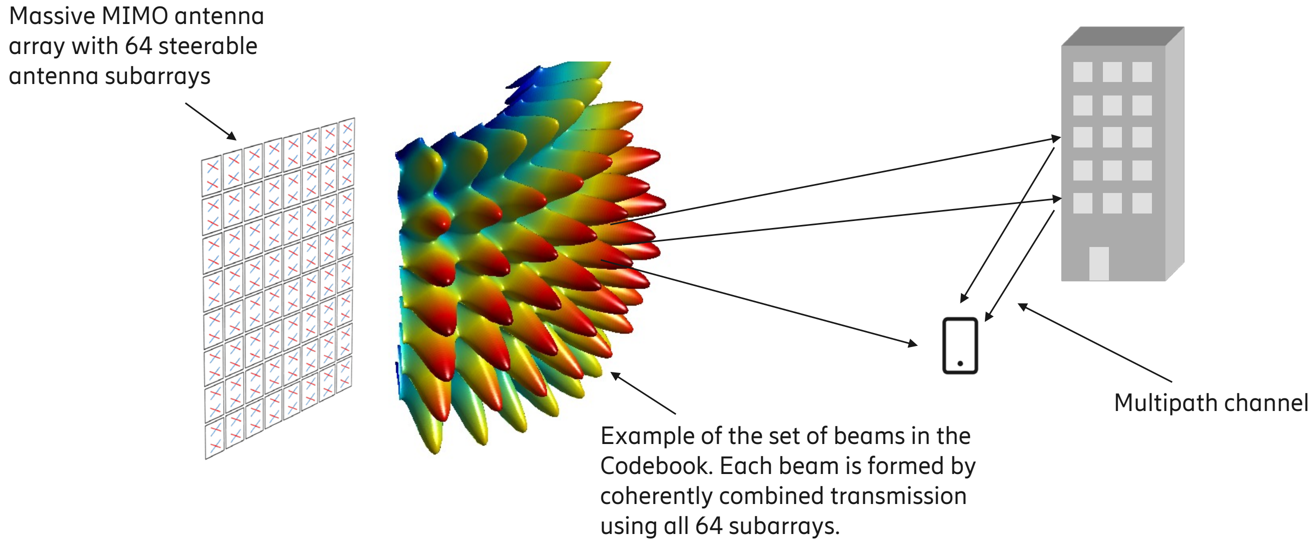
The SS/PBCH (SSB) duration is 4 OFDM symbols and 20 RB wide and it contains the primary and secondary synchronization sequence (PSS/SSS) and the physical broadcast channel (PBCH). The SSB period is 20 ms and are localized in a burst set of duration 5 ms in the beginning of the 20 ms period, see the example figure of L=4 SSB and thus four different beam directions. Note that the burst set length and periodicity (5ms and 20 ms) are independent of the subcarrier spacing. With increased subcarrier spacing, there are more slots within the burst set duration and thus room for a larger L.

A SIB1 is transmitted in each beam as the associated SSB, but not necessarily adjacent in time to the SSB but at specified point in time relative to an SSB.

Note that cell defining beams and data beams are not necessarily the same. Typically, data beams are narrower and have higher antenna gain, while the cell defining beams are bounded by the fact that they need to jointly cover the whole cell area. With such divergence in beam shapes there is a risk that the selected cell is not the strongest for the data transmissions and hence performance may suffer. Ideally, the cell should have been selected based on the strongest data beam.

To conclude, initial access signals can flexible be configured in NR, as opposed to LTE (which would correspond to L_{max} =1, i.e. a non-beam based initial access system). Different operators or network vendor may therefore approach this differently, e.g. the selection of number of cell defining signals L. There are pros and cons to select a large or a small (e.g. L=1) value such as larger L implies more SSB overhead but better match between data beam and cell defining beams.

What is Channel State Information (CSI)?



The information of the multipath propagation channel between gNB transmitter and UE receiver is in 3GPP called channel state information (CSI). It is conveyed to the gNB in a CSI report and used by gNB to determine how the Massive MIMO antenna array should transmit to maximize e.g. the received signal power at UE. For Massive MIMO solutions discussed in Chapter 3, the concept of CSI is central, as it is crucial for good MU-MIMO performance but also for SU-MIMO. See TS 38.214 (3gpp.org) for the related specification.

To acquire CSI at the gNB for a given UE, the gNB can alternatively use reciprocity, i.e. measure on an uplink transmission from the UE such as SRS or DMRS transmission.

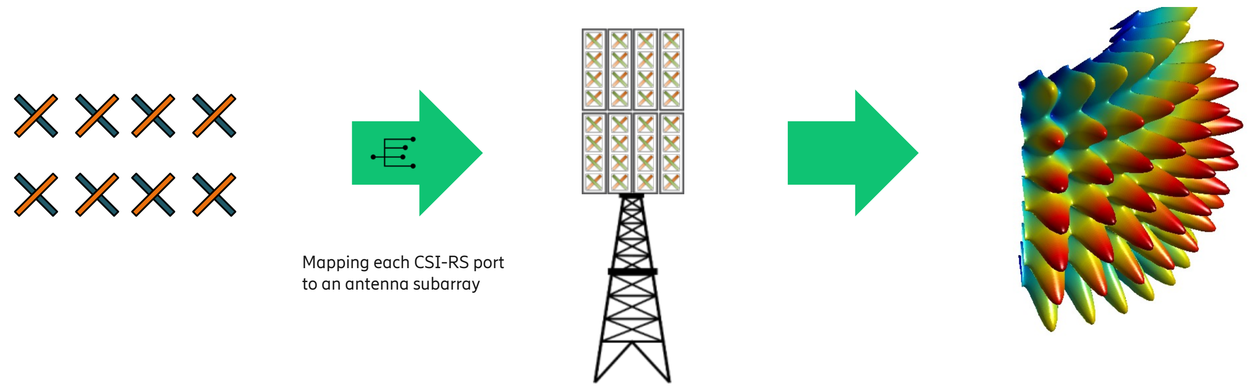
The Massive MIMO antenna array use a large number of controllable antenna elements, using phase shifters (and sometimes also amplitude shifters). The objective of acquiring CSI at the transmitter, is to improve the performance of the downlink transmission by utilizing information about the channel between each antenna element (e.g. each subarray or antenna port) of the gNB and each receiver antenna element of the UE.

To reduce CSI reporting overhead and complexity, a codebook has been defined in 3GPP which specifies a small but highly efficient subset of possible phase settings per antenna subarray. 3GPP has defined such codebooks under some assumption of typical Massive MIMO antenna array structures (such as a planar array with element spacing in the order of a wavelength), and typical mobile communication radio channels,

Note that a beam is the results of transmission from all subarrays, with carefully adjusted phase shifters to coherently combine the signals from these subarrays in a certain desired direction.

The takeaway is that the channel is multi-dimensional and to quantize it for reporting would cost too much in CSI reporting overhead. Instead, a carefully designed codebook of a set of likely transmission hypotheses (e.g. beams) is used which reduces the overhead. Alternatively, the gNB can use uplink measurements and rely on reciprocity to acquire CSI.

The UE measures on an equivalent, "logical" antenna array



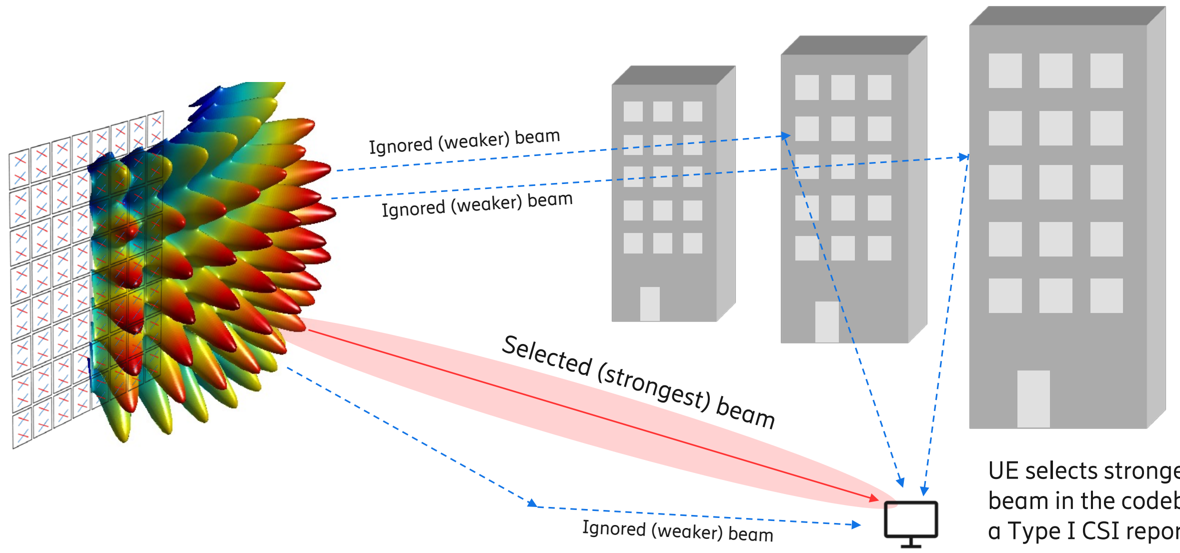
In codebook based operation, See Chapter 3 and TS 38.214 (3gpp.org), the UE measures the downlink channel using CSI-RS and finds the best possible precoder W and transmission rank for a hypothetical PDSCH transmission. It should also report a channel quality index (CQI), which is the modulation and coding scheme that gives the highest throughput given a target transmission error rate (e.g. 10%). Note that in codebook based operation, the beams from which a selection is done are virtual and exists only in the UE based band processing, the actual CSI-RS transmission does not represent multiple beams. This has the advantage that a large number of beam hypotheses can be evaluated from a single CSI-RS measurement.

Assume the AAS consists of 8 subarrays which are arranged as 2x4 (see the middle figure). Each subarray has 4x1 column of dual polarized antenna elements, hence in total 2x4x4x2 = 64 "antennas". In the implementation it is chosen to use two CSI-RS ports per subarray, one per polarization. Hence, the UE is configured a MIMO codebook with N1=2 rows and N2=4 columns in the antenna port layout (left figure). The UE is thus configured a CSI-RS resource with 2*N1*N2=16 CSI-RS ports. Therefore, from the UE perspective, the AAS array appears as a 2x4 antenna array of dual polarized antenna elements and the implementation using subarrays is not "visible" to the UE.

From the N CSI-RS ports, the codebook defines a 2D grid of beams (GoB), where 4*N1 beams are in the first dimension and 4*N2 beams are defined in the second dimension, in total 16*N1*N2 beams. See the figure to the right. The use of 4 times the number of beams compared to the number of ports of that dimension, is denoted spatial oversampling, as more beams than just the set of orthogonal beams can be selected. This reduces the losses as beams in between orthogonal beams can be selected.

A beam is created using a linear phase progression across the spatially separated antenna ports, per polarization, and is in 3GPP defined using the discrete Fourier transform (DFT), which has exactly this property. The CSI report also contains co-phasing factors, for example if rank 1 is selected, the UE reports how to co-phase the two polarizations within a selected beam. The co-phasing can change rapidly across the bandwidth and can thus be configured to be reported per subband, while the beam selection is the same across the whole bandwidth.

Type I CSI (coarse channel report)



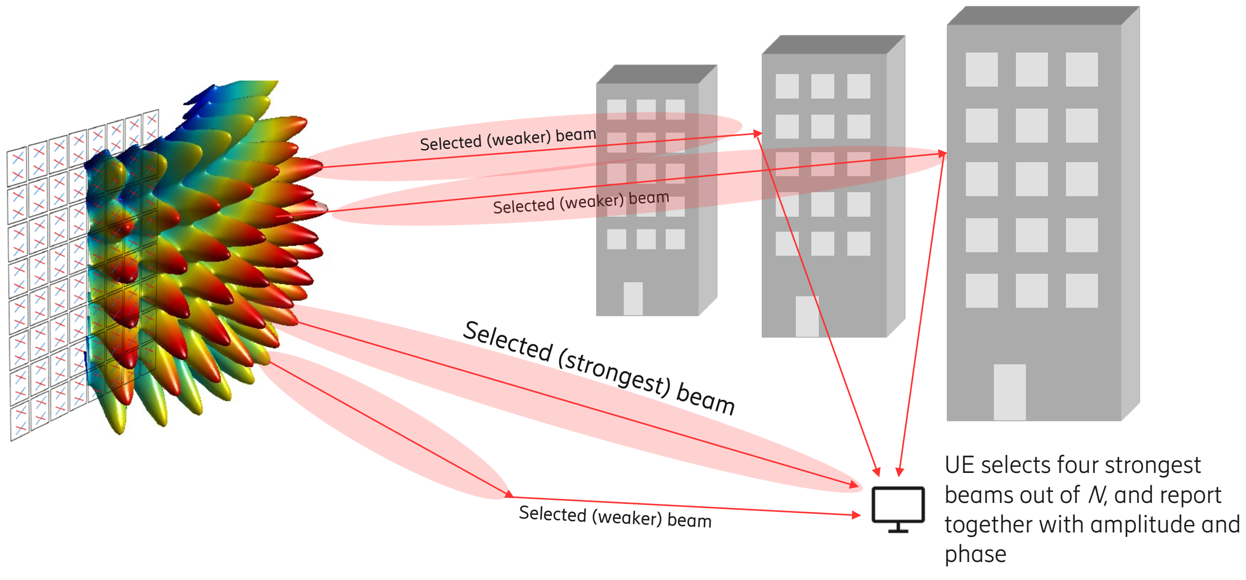
For CSI reporting, either Type I or Type II codebook based reporting can be configured to the UE. See Chapter 3 and TS 38.214 (3gpp.org). The figure shows an illustration for the example of rank 1 CSI reporting for Type I (and only one of the polarizations is shown for simplicity). In Type I, the main principle is that only the strongest reflection in the channel is selected and reported as CSI to the gNB, the weaker ones are ignored. This type of CSI is often sufficient for low rank SU-MIMO transmission, since the gNB only need to know where to focus all it's transmission power to reach the UE. The Type I CSI report has low payload, compared to the richer Type II report.

In both Type I and II cases, a basis set of dual polarized beams mapped to vertical and horizontal, are used for feedback reporting. Either the gNB is creating the set of beams by actually transmitting multiple CSI-RS by using antenna virtualization V. One CSI-RS of 2 ports (one port per polarization) is in this case transmitted in each beam and the UE is selecting the preferred beam for reporting. This is denoted beam-based reporting.

Alternatively, the gNB is transmitting a single CSI-RS resource of N ports, where each such port covers the whole sector, and the multiple beams are the created virtually in the UE by using a codebook of pre-defined beams as described in previous pages. This is denoted antenna-based CSI reporting.

In Type I reporting, the antenna-based reporting is used, and the UE evaluate the whole set of beams in the codebook and selects one preferred beam. For higher rank, multiple beams are selected. The selection thus corresponds to a preferred precoding matrix W that is feed back to the network using the PMI field in the report. The PMI field also contains co-phasing of the polarizations (note that the figure assumes a single polarization for simplicity of illustration) and this can be configured to be reported per subband. As this report only report the most preferred direction, it works well for SU-MIMO scheduling and transmissions. The CSI reporting overhead is low.

Type II CSI (rich channel report)



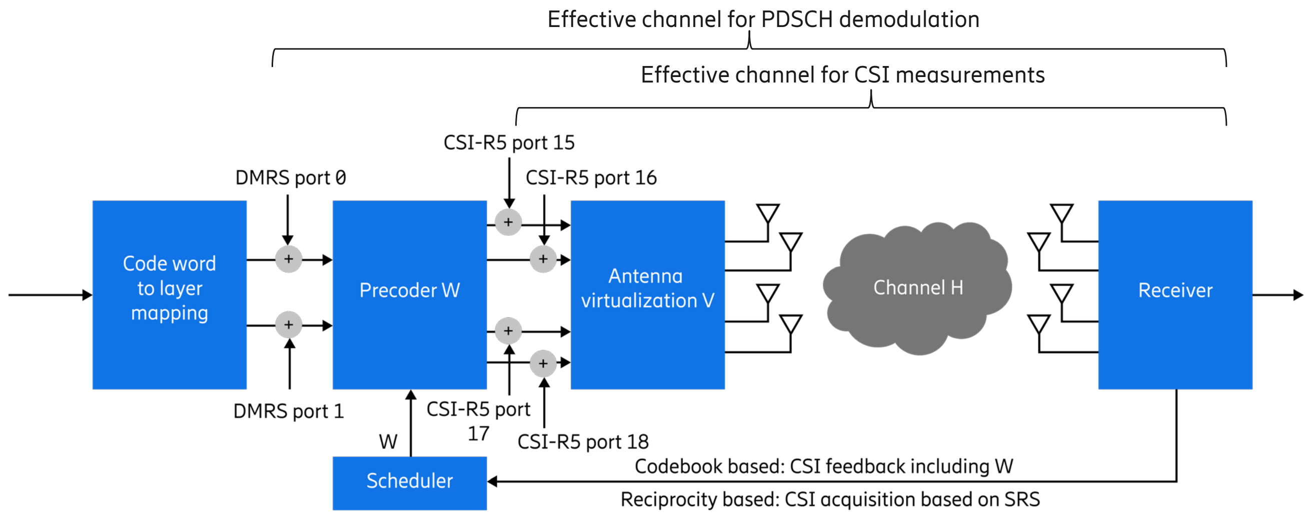
The figure shows an illustration for the example of rank 1 CSI reporting for Type II CSI (and only one of the polarizations is shown for simplicity). The principle is to feed back a rich channel report, aimed for MU-MIMO scheduling. It is important to know how a transmission towards one UE interferes another, co-scheduled, UE. Therefore, not only the strongest beam is reported but also weaker beams, as they are utilized to compute this interference. A drawback of Type II is the larger reporting overhead.

In Type II reporting, the UE evaluates the whole set of beams in the codebook and selects multiple preferred beams. It scales the power of each beam according to 4 quantization levels, based on the CSI-RS measurement and computes how these beams should be co-phased using a QPSK or an 8-PSK alphabet. The purpose is to report multiple beams and for each, both amplitude values and co-phasing. This gives a good representation of the whole channel, not only for the main direction as in the Type I reporting. The overhead depends on whether the amplitude report and co-phasing is subband or wideband. Also, the number of beams the UE shall select can be configured to the UE. When performing MU-MIMO scheduling, knowledge of the whole channel matters, not only the main direction, since this gives the network a possibility to control the interference that is created towards co-scheduled UEs.

In Rel.16, an enhancement was introduced to the Type II codebooks. The fact that the co-phasing per subband is correlated across adjacent subbands was utilized in order to reduce the CSI reporting overhead. Thereby, the performance improves, or the overhead reduce, depending on the chosen configuration.

In Rel.17, a further enhancement was introduced for the Type II codebook, targeting mainly FDD deployment. It is observed that due to the FDD duplex distance, the phase information is not reciprocal between uplink and carrier frequencies, but the angles and delays maintain reciprocity despite the duplex distance. Hence, the gNB can measure angles of arrivals and delays of the channel components and transmit beams in the downlink which are pre-compensated with respect to delays (and beam directions). This will reduce the UE complexity significantly, while more computations is needed on the gNB side..

Selection of precoder W utilizes the CSI



The figure illustrates a downlink transmission chain for a MIMO system that is the foundation of the solutions described in Chapter 3 and is the assumed model used in 3GPP specifications of single TRP transmission of data to a UE. The encoded set of bits (code word) is mapped to layers before precoding and transmission. The OFDM block that converts from frequency to time domain is placed before or after the antenna virtualization (not shown).

3GPP defines "antenna ports" by inserting reference signals at a point in the transmission chain. In this figure, four CSI-RS are inserted, but these pass through an antenna virtualization matrix V, hence, the UE measures the effective channel for CSI measurement HV using CSI-RS. It is therefore up to gNB implementation how to choose V, it can be used to beamform in some direction if desired. The UE is not aware of how V is selected. In mmwave operation, it is common that V is implemented using analog phase shifters, i.e. a time domain analog signal processing operation. In addition, each physical antenna shown here may actually represent the transmission from a subarray of physical antennas.

In case of codebook based operation, the UE estimates a rank and a preferred precoder W (PMI + rank feedback) in the CSI feedback report to the gNB. The network can configure the W to be a wideband report (one precoder for the whole bandwidth), or subband W reporting, which consumes more reporting overhead but gives increased channel information as the channel is frequency dependent, in general.

In case of reciprocity based operation, the UE transmits sounding reference signals (SRS) in the uplink, so that that the gNB can estimate the uplink channel. In TDD, the uplink and downlink channel H are reciprocal in case of full reciprocity. See the details on SRS antenna switching on a following page on how 3GPP have tried to address the partial channel reciprocity issue.

When transmitting PDSCH, the stream of modulated symbols (known as a code word), is mapped to layers (i.e. the rank of the transmission to the UE). One demodulation reference signal (DMRS port) is transmitted, per layer. The DMRS is added before the precoder, so that the transmission will pass through the combined channel HVW before being received in the UE receiver antennas. Hence, when UE measures the channel using DMRS, it measures the effect of the combined, effective channel for PDSCH demodulation, on the transmitted modulated symbols. The precoder W is thus also part of the effective channel for PDSCH demodulation. Therefore, for data transmission, gNB can choose arbitrary W, and what the UE feeds back in a CSI report is just a recommendation.

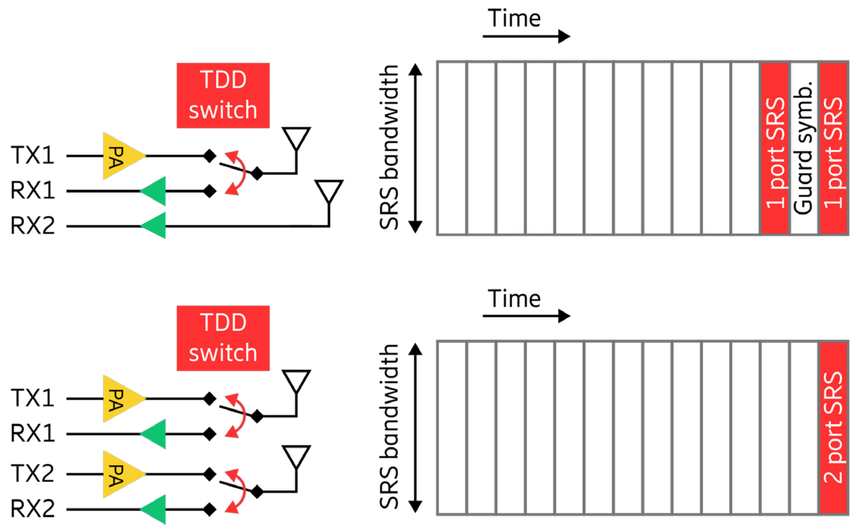
SRS antenna switching options for 2 RX UE

1T2R

UE implementation that requires antenna switching. Guard symbol needed for antenna switch.

2T2R

UE implementation doesn't require antenna switch



Sounding reference signals (SRS) can be configured to be transmitted from the UE and is important in reciprocity-based operation, particularly for MU-MIMO scheduling. A UE may have fewer TX antennas than RX antennas and in this case SRS antenna switching can be utilized so that the gNB acquires CSI for the full MIMO channel (full channel reciprocity) and not only partial channel.

A terminology for antenna switching is introduced where 1T2R means that the UE have 1 TX and 2 RX antennas and support antenna switching. The UE will report 1T2R as a capability to the network in this case. If the UE has full reciprocity capability the UE can in this case report 2T2R and there is no need to use antenna switching in this case. See the figures for examples where also the TDD switch is illustrated. It should also be noted that antenna switching is costly in overhead in that it (currently) requires an empty Guard Symbol to allow for the UE to perform the antenna switch. For UEs with 4, 6 or 8 RX antennas, there are similar switching schemes, e.g. 2T4R, 2T6R, 2T8R, 4T8R, ... and it could in those cases be even more guard symbols as there are more antenna switches needed.

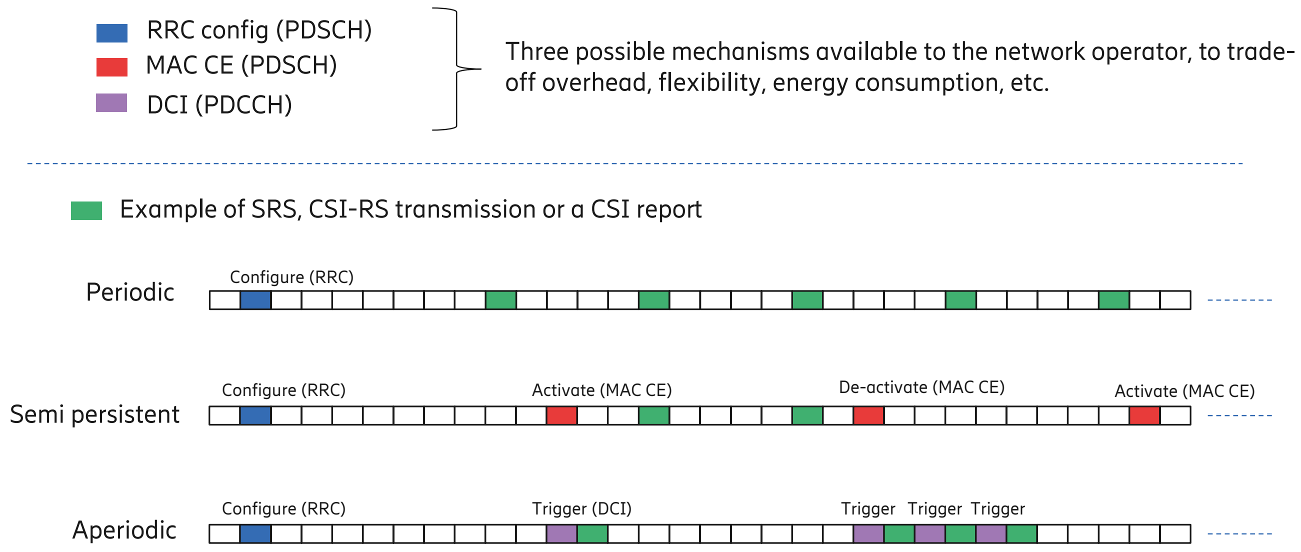
In carrier aggregation, not all downlink carriers have an associated uplink carrier, and to be able to transmit SRS also in such carrier without an uplink, SRS carrier switching has been introduced. The UE can be configured to intermittently transmit SRS in such uplink carrier (which is an uplink carrier that cannot be used for transmitting data or control) so that reciprocity transmission can be used for the corresponding downlink. However, when performing such temporary switch to another carrier, the transmission on the default carrier is paused.

In Rel.15, the SRS (including guard symbol) can be transmitted in one or more symbols in any of the last 6 symbols in the slot and the SRS bandwidth is configurable, while in Rel.16, the up to 6 SRS symbols can be transmitted in any set of adjacent symbols of the slot. Rel.17 further enhanced this by removing the restriction of 6 symbols.

As the total output power is limited, the wider the SRS bandwidth the worse is the coverage. Interference mitigating and coverage enhancing techniques such as frequency hopping (FH) and repetition with a factor of 2 or 4 can be also configured. In FH, the SRS is transmitted in a different part of the bandwidth in each symbol, while in repetition, the same part is used in each symbol.

The SRS can be configured to be transmitted on every 2nd or every 4th subcarrier (known as "combs") which defines up to four orthogonal resources for SRS. In Rel.17, every 8th subcarrier is additionally supported.

CSI measurements and reporting can be configured using three approaches



NR support three mechanisms to enable or disable measurement reference signals (SRS and CSI-RS) and to enable/disable CSI reporting. This allows the network operator to make a tradeoff between reference signal overhead (periodic vs aperiodic), triggering overhead etc.

Periodic SRS or CSI-RS

This is typically used for shared downlink signals among multiple UE. The UE is configured a periodic signal using RRC signaling and continues to receive (CSI-RS) or transmit (SRS) until RRC deactivates the signal.

Aperiodic SRS or CSI-RS

This is used for signals for one UE, e.g. when the network would like a UE to perform a CSI measurement or get a report or the network would like to perform the measurement on SRS for one UE.

The UE is configured an aperiodic signal using RRC signaling and DCI (i.e. PDCCH) is used to trigger the RS transmission/measurement. The benefit is low RS overhead as it is only transmitted when network needs information, but on the other hand it consumes PDCCH overhead as each UE must be triggered with an individual PDCCH. Medium access control channel element (MAC CE) can be used to change the configuration of aperiodic triggers. If aperiodic SRS is triggered, then the SRS transmission can be in the same slot as the DCI that carried the trigger (at least in FDD or in special slots in TDD).

Semi-persistent SRS or CSI-RS

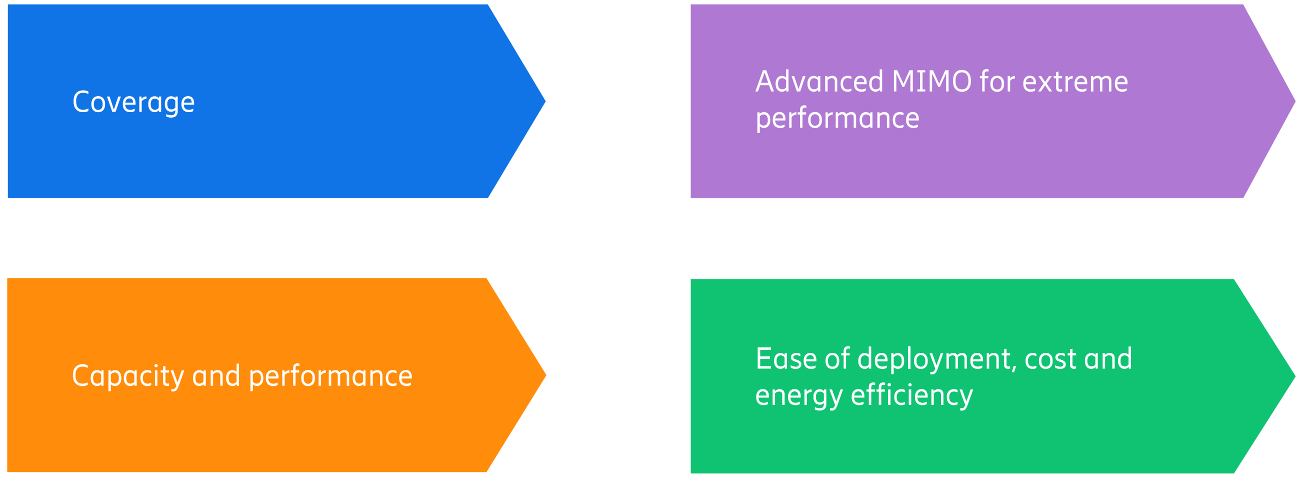
This option provides a trade-off between periodic and aperiodic where the network can quickly (without the need to use RRC) activate or deactivate the configured periodic signal using the MAC CE which is scheduled to the UE using PDCCH+PDSCH. In semi-persistent operation, the UE is configured a periodic signal using RRC signaling and starts to receive (CSI-RS) or transmit (SRS) the signal only when MAC CE has activated (or deactivated).

When it comes to CSI reporting, there are three possibilities to use time dynamics:

- Periodic CSI reporting is configured by RRC and can only use periodic CSI-RS for its measurements.
- Semi-persistent CSI reporting is configured by RRC and triggered by MAC CE (for PUCCH reporting) or DCI (for PUSCH reporting). It can use either periodic or semi-persistent CSI-RS for the measurements.
- Aperiodic CSI reporting is configured by RRC and triggered by DCI. The trigger information (states) can be reconfigured using MAC CE. The aperiodic CSI report can use either periodic, semi-persistent or aperiodic CSI-RS for the measurements. If aperiodic CSI-RS is used, then the CSI-RS can be in the same slot as the DCI that triggers the CSI report.

5. Massive MIMO features

Introduction



The purpose of this chapter is to elaborate on the Massive MIMO features introduced in [3; Ch. 3.3, p. 42-51]. We can categorize the possible Massive MIMO features in four different categories, which will be further elaborated on in the coming slides.

1. Coverage
2. Capacity and Performance
3. Advanced MIMO feature for extreme performance
4. Deployability, Cost & Energy Efficiency

Coverage features intend to improve either the access coverage or the data channel coverage typically by applying beamforming to boost the signal strength. This can be done either in uplink or downlink. Capacity and performance feature on the other hand typically aims to improve peak user throughput and/or capacity of the data channels and generally uses a combination of beamforming, null forming and spatial multiplexing. Advanced MIMO features targets more specific deployments or scenarios which may have an extreme capacity need, very high load or some other special characteristics. For these cases, more advanced e.g. inter-cell coordination features or higher-order spatial multiplexing can be applied (which would not be required or provide additional benefit for e.g. less capacity-demanding cells).

Features in the Deployability, Cost & Energy Efficiency category, on the other hand addresses non-performance related key performance indicators (KPI) [3; Ch. 2, p. 26] such as e.g. fulfilling regulatory requirements on emissions (EMF) which may require dynamically adapting the power level depending on the beamforming characteristics of the cell or reducing the amount of beamforming that is performed in order to conserve energy.

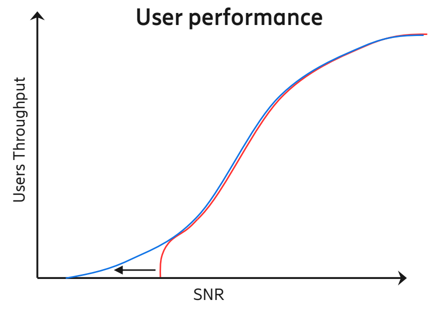
5.1. Coverage

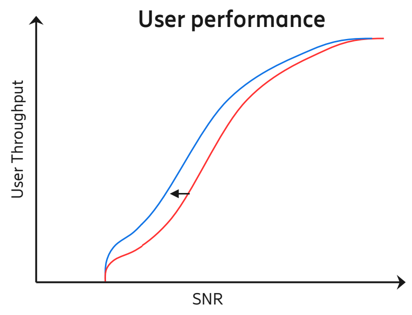
Types of coverage improvements

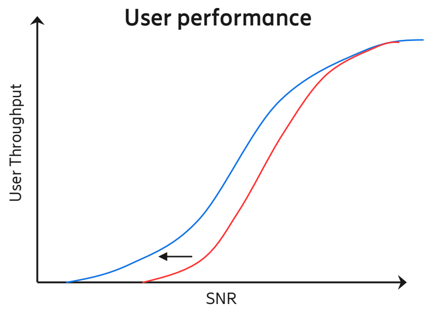
Access improvement
Feature which increases the link budget of a limiting control channel so that the UE can connect to or maintain a connection to the network in poorer coverage

Data channel improvement
Feature which improves the coverage/performance of the traffic channel ("App coverage")

General coverage improvement
Improves both access and data channel coverage







When addressing coverage enhancements, it's important to differentiate between access coverage and data channel coverage.

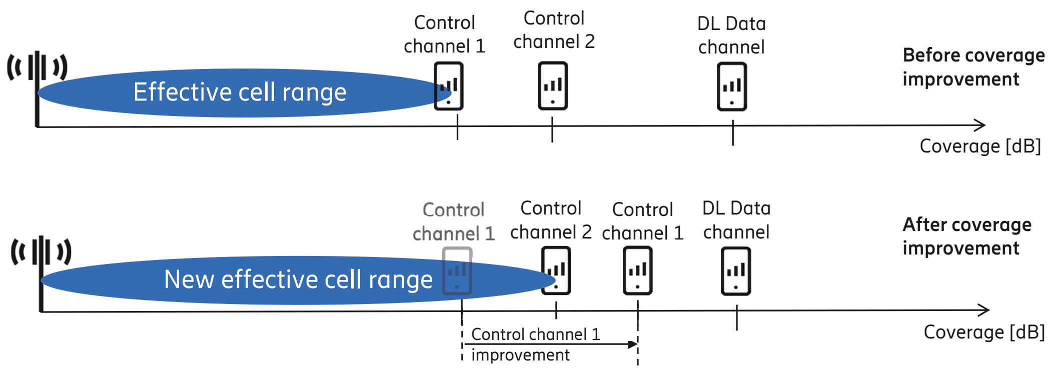
Access coverage can further be broken down into initial access coverage, which is the coverage of the cell-defining signals and messages in the inital access procedure. All of which needs to be in coverage for a UE to find and connect to the cell, and control channel coverage, which is the coverage of the supporting control channels needed for scheduling and providing feedback of/for the data channels. For example, an access coverage limitation manifests itself as a sharp drop to low user throughput when the SNR reaches a low enough value, as is illustrated on the red curve of the left picture. The interpretation is that the data channel could be sustained and provide service to the UE at even lower SNR levels (with smoothly and gradually decreasing user throughput), but one or more of the control channels are out of coverage and therefore not able to support the data channel. By improving the coverage of said control channels so that they are no longer limting data channel performance, it is possible to reach the performance of the blue curve.

A data channel coverage improvement on the other hand will manifest itself as a shift of the entire link curve to the left as is illustrated in the middle picture, meaning that a certain user throughput can be acheived at a lower SNR after the coverage improvement. This is also commonly referred to as improving the app coverage, considering that a certain application may have a requirement on X Mbps user throughput to properly function. After the data channel coverage improvement we can acheive this at a lower SNR level and thus in a larger part of the cell.

Then there are of course general coverage enhancements which can improve both access and data channel coverage (e.g. increasing the transmission power).

Improving access coverage

- Coverage can be defined as the ability to serve traffic at a certain rate
  - All procedures and physical channels required to serve the traffic need to have coverage, data channels, control messages, initial access messages
- Different messages have different baseline link budget
- Link budgets can be different in different implementations



The coverage of the supporting channels needs to be on the same level as the downlink/uplink data channel, otherwise there is a coverage limitation. However, larger "coverage" of the supporting channels does not necessarily result in actual improved downlink or uplink coverage. Ideally all supporting channels of a traffic channel should have the same coverage as the traffic channel itself. If any supporting channel has smaller coverage, that channel will be limiting and the traffic channel is not operational, but if a supporting channel has larger coverage than the traffic channels it supports it provide no additional benefit. That is, it is over-dimensioned which may have a determintal impact on system performance as the improving the coverage of a channel generally implies that more overhead is required (which means less resources for the data channel itself).

To support the data channels, a number of supporting channels needs to be in coverage as well, see [Ch. 4, p. 42] for further details.

PDSCH:

- Initial access coverage
- Downlink grant carried on PDCCH
- HARQ ACK/NACKs carried on PUCCH Data on PDSCH
- CSI carried on PUSCH/PUCCH
- RLC ACK/NACK carried on PUSCH

PUSCH:

- Initial access coverage
- Scheduling request (SR) using PUCCH
- Grant carried on PDCCH
- Data on PUSCH

Initial access:

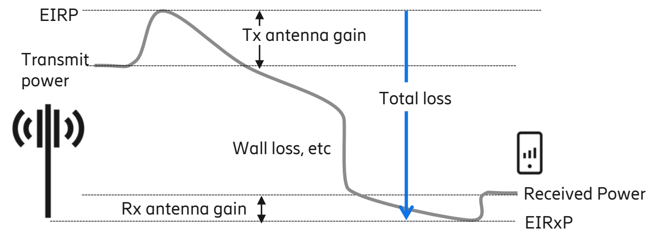
- Synchronization and SIB1
  - PSS/SSS/PBCH
- PRACH (MSG 1)
- MSG2 (RAR) carried on PDSCH with associated PDCCH
- MSG 3 carried on PUSCH
- MSG 4 carried on PDSCH with associated PDCCH (not used in non-standalone deployments)

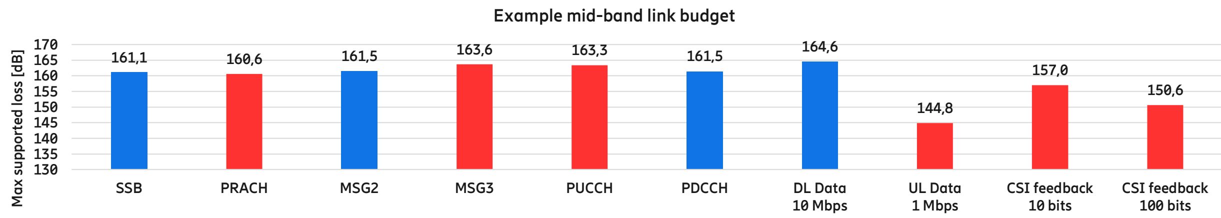
The effective cell range is thus determined by which of the supporting channels that break first. A coverage enhancement feature which provides a coverage boost of the most limiting channel will thus increase the effective cell range so that it is now determined by the next limiting control channel, as will be further discussed on the next page [Ch. 5, p. 57].

Link budgets further explained

Key parameters affecting coverage include:

- Massive MIMO and antenna gain
- Different physical channels have different SINR requirements
- Product impact: Transmit power, receiver sensitivity, bandwidth and duplex operation





Different products and physical channels have different link budgets (assumptions matter)
Important to improve coverage limiting channels

The Link budget is a useful tool to estimate coverage, measured as maximum isotropic loss that can be supported. Link budgets are useful both for uplink and downlink. Both base station and UE performance matter. Maximum supported isotropic loss depends on several parameters, viz. Tx power, Tx antenna gain, Rx antenna gain, Rx sensitivity and possibly also some other parameters, e.g. power booster factor. Different physical channels have different parameter values and hence different coverage.

Max isotropic loss is calculated as (simplest version):

Max_iso_loss (dB) = Tx power + Tx antenna gain + Rx antenna gain – Rx sensitivity

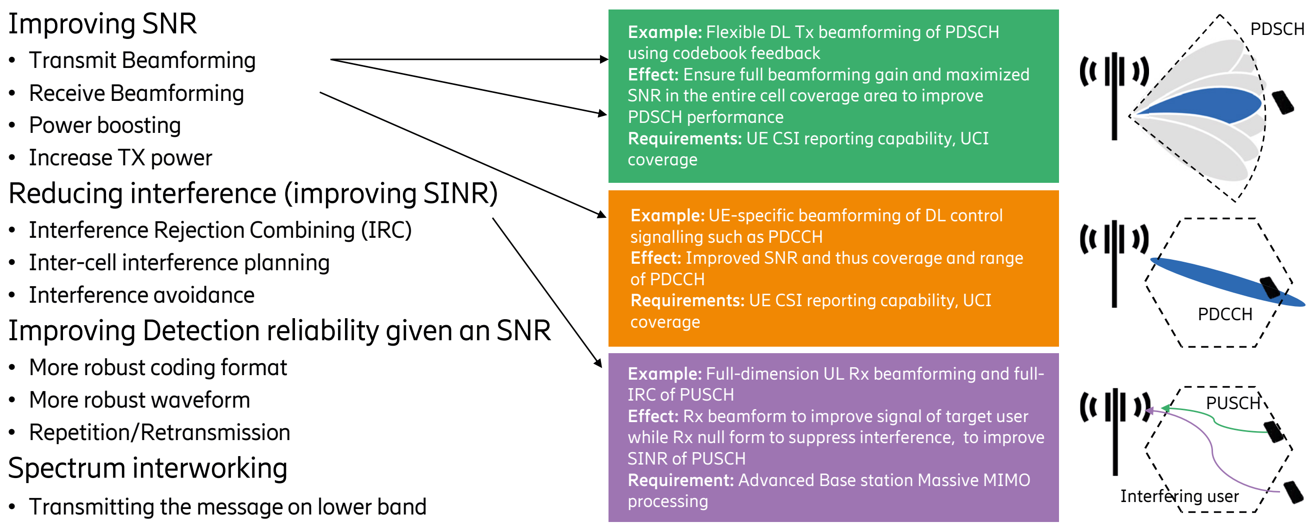
And is used to assess the maximum coverage for different physical channels.

Some aspects to consider:

- Data vs. control and access coverage and app coverage (what CSI is available; common beamforming with less gain vs UE-specific bf with more gain).
- Link budgets often consider Boresight, i.e. in the direction where the main lobe takes its maximum value. With a Massive MIMO you can realize the UE-specific gain over a larger area compared to a conventional antenna.
- Uplink vs. downlink Link budgets can be used for both uplink and ownlink. The parameters will of course be different, but the principles are the same. Often, the uplink is limiting, since the UE output power is limited according to standard, whereas the power from the RRU or Massive MIMO can be typically 30-50 dB higher.

Since all different channels have different coverage, it is important to identify which channel is limiting and improve coverage of these limiting channels first. The coverage enhancing features typically address this issue.

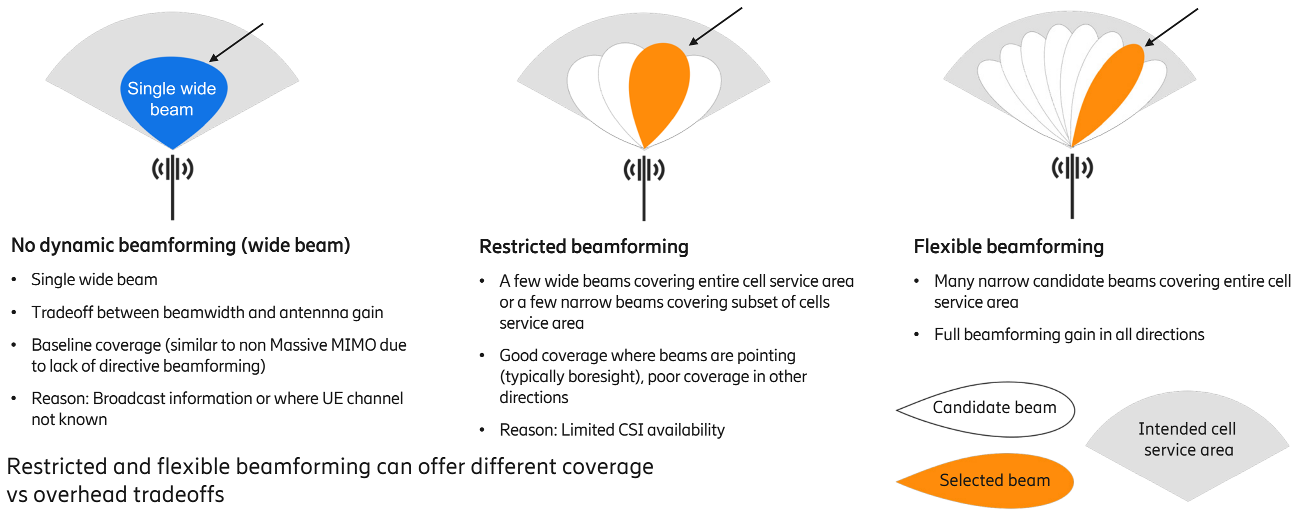
Ways to improve access and app coverage



There are many ways to improve access and app coverage of a mobile network, but on a high-level it is achieved by either boosting the signal strength of a signal/channel or reducing the interference level at the receiver. Boosting the signal strength (or SNR) can either be done directly by increasing the transmit power, maintaining the same power level but boosting the PSD of certain resource elements in frequency while de-boosting others, or performing beamforming on the transmit or receive side. Or, indirectly by improving the detection reliability for a given SNR (and thus increasing the energy per bit), such as by performing repetition or choosing a more robust coding/transmission format. Data channel coverage can also be improved by reducing the interference level (in interference-limited scenarios) by either inter-cell interference planning of the transmission, or by utilizing advanced receiver processing such as IRC. Finally, one can improve the coverage of a signal by choosing to transmit it on another channel with better propagation conditions such a lower frequency band with
lower path loss.

All of these methods are generally applicable to both conventional radios with few radio chains as well as Massive MIMO radios. However, with transmit/receive beamforming and IRC receiver processing, the benefit increases with the number of radio chains and so these methods can be considered as differentiators for Massive MIMO systems to further improve the coverage compared to conventional radio systems.

Flexible vs restricted beamforming



Restricted and flexible beamforming can offer different coverage vs overhead tradeoffs

Even though a Massive MIMO radio has the possibility to use classical beamforming to create narrow beams maximizing the beamforming gain in a certain direction, it also has the possibility to create beams with larger beamwidth (and thus lower antenna gain). The benefit with using wider beams is that less beams are required to cover the entire intended cell service area, which is some cases may be desirable, on the other hand the downside is that the experienced beamforming gain is lower.

In general, it is most beneficial for coverage to apply flexible beamforming, with many narrow candidate beams covering the entire cell service area, whenever possible in both downlink and uplink in order to maximize the beamforming gain and thus improving the coverage. However, there are some exceptions where restricted beamforming or even a single wide beam may be more suitable:

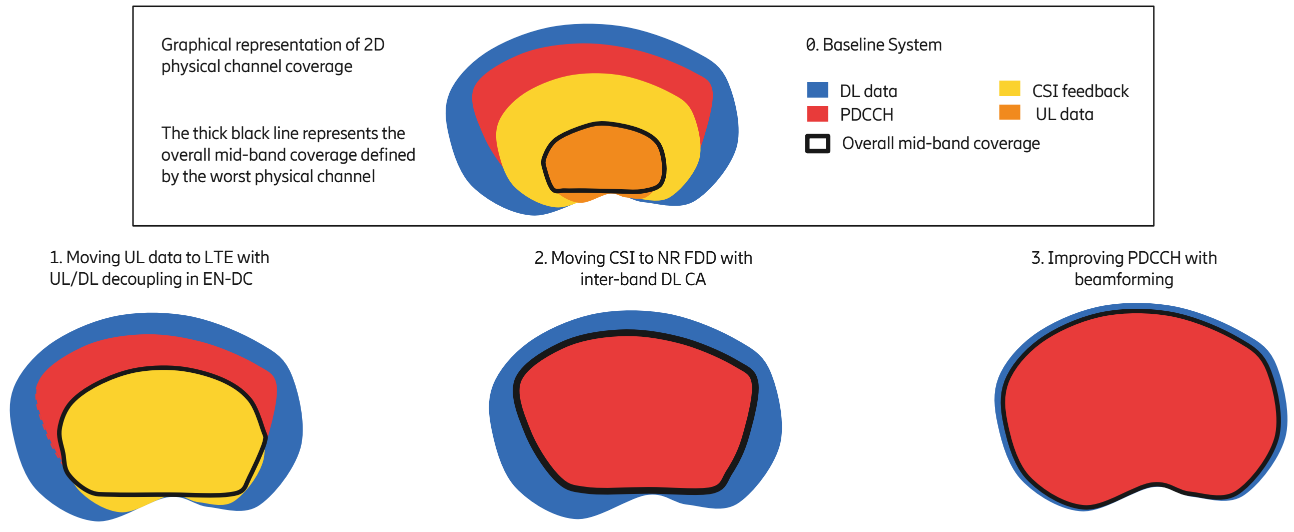
1. Additional coverage improvement is not necessary
2. Lack of CSI availability

For instance, in order to apply flexible beamforming of the PDSCH for a 32Tx Massive MIMO, CSI needs to be available either from uplink sounding on or 32 port CSI-RS feedback. However, some UEs may only support 8 CSI-RS ports and so a restricted beamforming with a fixed narrow beam in the vertical domain but many narrow beams in the horizontal domain can be applied for that UE. In this case, the restricted beamforming still results in narrow beams with maximum beamforming gain but does not cover the entire cell service area in the vertical domain and so a UE which is not in the vertical direction of the fixed vertical beam will see limited beamforming gain.

Another way to apply restricted beamforming is to instead create a set of wide beams which covers the entire intended cell service area, but with lower maximum beamforming gain per beam. In the extreme case of this variant of restricted beamforming, we end up with a single wide beam covering the entire cell service area. This way of beamforming may be suitable for broadcasting of cell-defining signals such as SSB and broadcasting of system information.

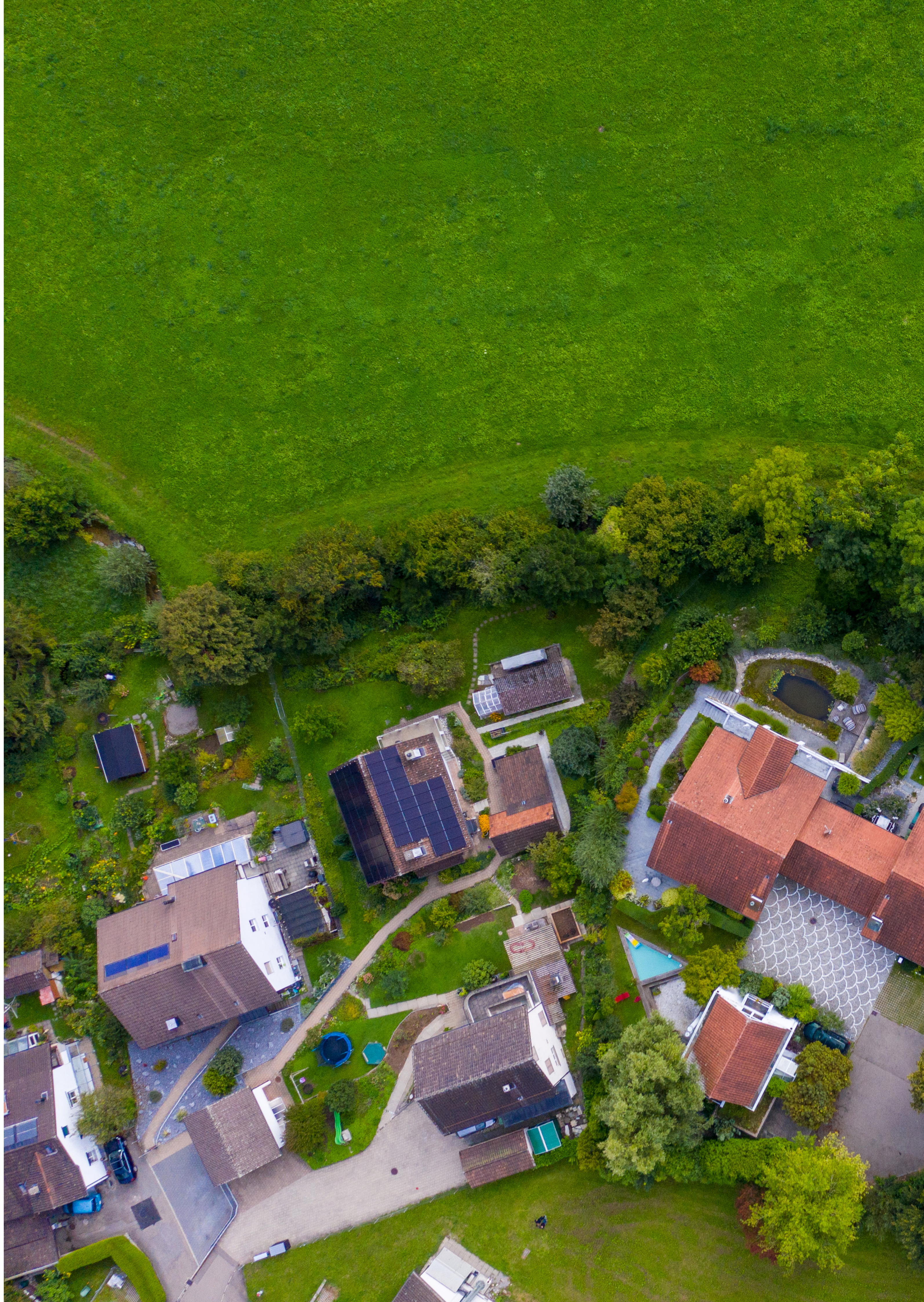
Note: The beams are for simplicity depicted as classical beams. In reality, however, they should viewed as generalized beams [Ch. 2, p.24].

Example: Access coverage improvement path



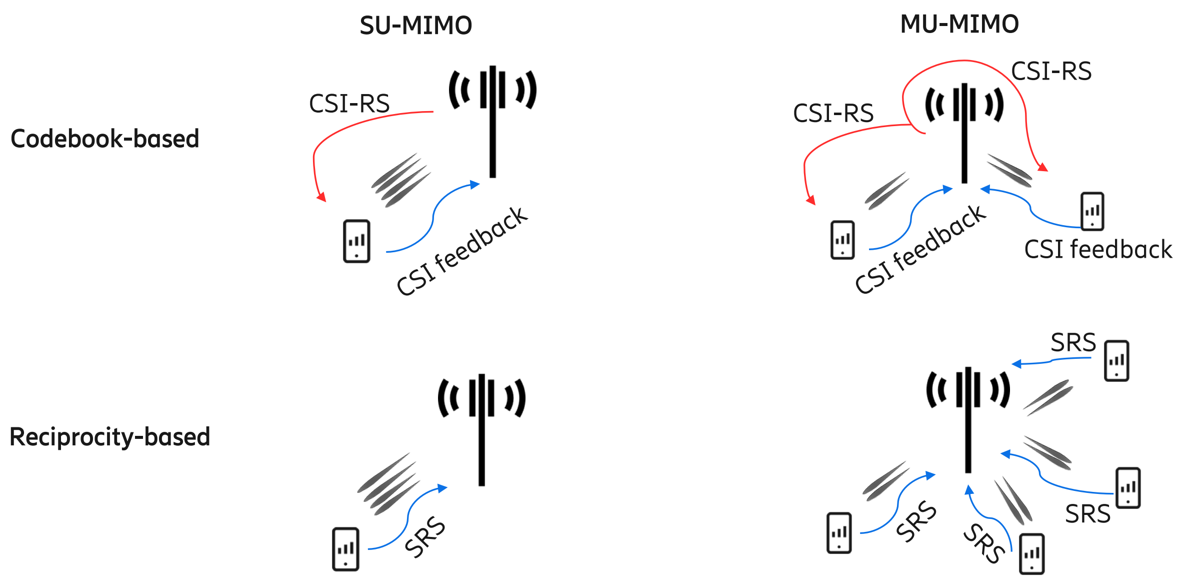
We here provide one illustrative practical example of how overall mid-band TDD coverage can be improved by a set of coverage enhancement features.

1. In the baseline system, no spectrum interworking is used, and all channels are transmitted on the mid-band TDD carrier. The effective cell range is here limited by the uplink data channel (PUSCH).
2. By using spectrum interworking with low-band FDD in EN-DC, the uplink leg can be moved to LTE on the lower FDD frequency band with better coverage properties, thereby removing the first coverage limitation. The effective cell coverage is now determined by the CSI feedback coverage of the UCI.
3. To further improve the coverage, inter-band downlink carrier aggregation can be activated to move the uplink control channel (carrying the CSI report) also to low-band FDD, thereby increasing the range where CSI reports can be correctly decoded and increasing the effective cell range further, which is now limited by PDCCH.
4. User-specific beamforming of PDCCH can be introduced to beamforming PDCCH in a similar fashion as PDSCH.



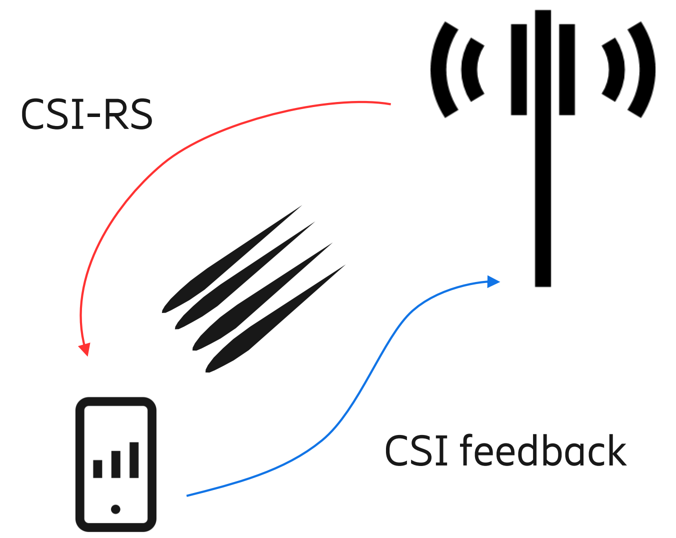
5.2. Capacity and Performance

Overview of Massive MIMO features for downlink data channel



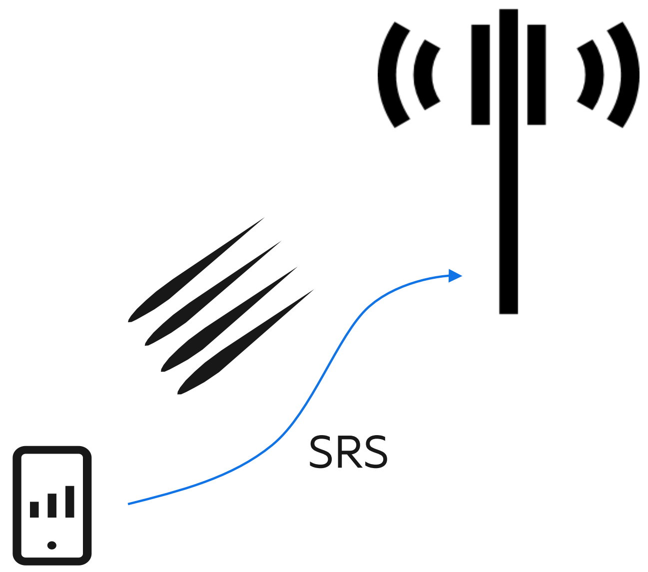
On a high level, the basic bread-and-butter features for downlink performance and capacity in a Massive MIMO system can be classified according to 1) if codebook-based or reciprocity-based channel information is used and 2) wheter single- or multi-user MIMO is applied. This results in four distinct high level downlink Massive MIMO features. Typically, all of these four features would be activated in a cell, but which transmission scheme is used at a certain point in time would depend e.g. traffic conditions and UE characteristics. Of course, there are various different possible options for how reciprocity-based and codebook-based SU/MU-MIMO can be implemented, both from what is available in the 3GPP standard as well as from proprietary algorithm perspective, however on a high-level the various options still share the same characteristics with respect to use case and performance.

Codebook-based beamforming



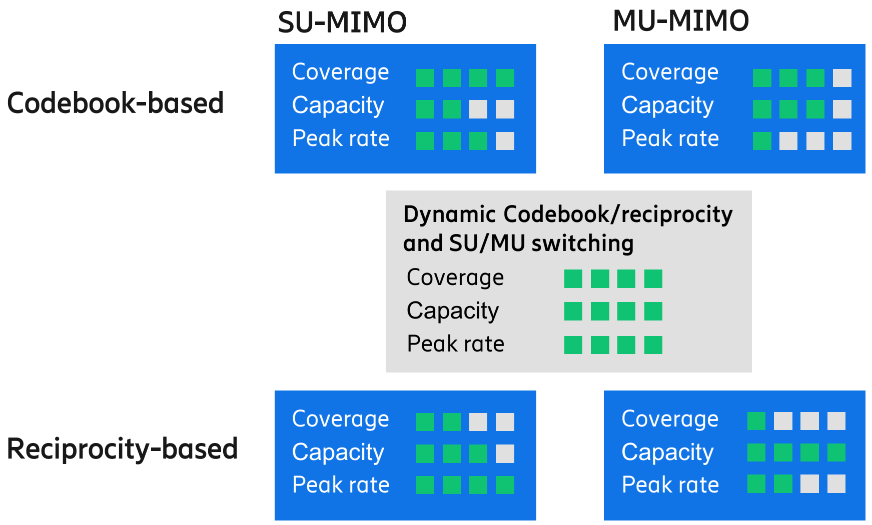
With codebook-based transmission, the downlink channel state is aquired by first sounding the channel in the downlink by the base station transmitting a Channel State Information Reference Signal (CSI-RS) which is measured on by the UE. The UE thus has full information about the downlink channel state, but needs to convey this information to the base station in order for it to determine in which direction to beamform. Theoretically this could be acheived by simply quantizing all the channel coefficient to a number bits and transmitting these in the uplink, so called explicit feedback, however this would result in massive overhead. Instead, a so called implicit feedback mechanism is used, where a number of candidate beamformers are defined in the 3GPP standard, a so called precoder codebook, where the UE recommends the most suitable beamformer for the base station to use (i.e. the one that best matches the measure downlink channel on the CSI-RS). The number of candidate beamformers defined in the 3GPP codebooks are relativly few, up to ~256, so the selection can easily be encoded into a few bits and sent with channel coding in the uplink control information, making the coverage of such feedback quite robust. However, since the feedback format is limited (the precoder codebook can be seen as a course spatial quantization only giving information about the strongest propagation path of the downlink channel), only simple classical beamforming, [Ch. 2, p. 12], can be used. Such beamforming works quite well for low-rank transmission in line-of-sight channels, but generalized beamforming requiring more detailed channel information is typically needed for good performance with high-rank (rank >2) transmisison or when the channel contains a lot of multi-path. As the codebook-based feedback only gives the base station information about the dominant downlink channel direction, it is more suitable for SU-MIMO transmission. MU-MIMO with codebook-based feedback is still possible and can give a benefit in some cases (especially with well-separated users or with sparse channels without a lot of multi-path), but will in the general case not give substantial gains since the interference between users is difficult to supress.

Reciprocity-based beamforming



With reciprocity-based beamforming, the downlink channel state is instead acquired by sounding the uplink channel by the UE transmitting a Sounding Reference Signal (SRS). By utilizing that the uplink propagation channel is reciprocal since the uplink and downlink is on the same frequency in a TDD system, the downlink channel state can be estimated from the uplink (while the propagation channel is reciprocal, the channel effects due to RX/TX processing is generally not, so calibration between the uplink and downlink chains is required for reciprocity-based beamforming to work). The benefit with reciprocity-based beamforming is that the complete channel information is avilable at the transmitter, including it's fast fading properties, which gives the base station very granular channel information in spatial and frequency domains enabling the use of more advanced generalized beamforming, [Ch. 3, p. 24]. The generalized beamforming can be used to increase both SU-MIMO performance by by cretaing high-rank beams which correpsond better to the actual propagation channel (no longer limited by the fixed feedback format of the 3GPP codebooks), and particularly to boost MU-MIMO performance since the more detailed channel information can be used for null forming to reduce interference between UEs. However, in order to use the fast-fading properties of the channel for precise null forming and beamforming, the channel estimates needs to be of sufficently good quality. This means that coverage range where such transmission can be performed is limited. Trying to conduct precise null forming/beamforming based on too noisy channel estimates generally results in worse performance than what can be acheived with classical beamforming using codebook-based feedback. Another limitation with reciprocity-based beamforming is that the entire bandwidth to be scheduled in the downlink needs to be sounded in the uplink, for each Rx antenna of the UE. This means that the UE needs to have an uplink allocated on the TDD cell (and so reciprocity-based beamforming may not be available in a downlink carrier aggregation scenario). And since mid-band NR UEs typically only are equipped with 1 or 2 Tx chains but have 4 Rx antennas, the proper SRS antenna switching functionality needs to be supported by the UE so that all Rx antennas can be sounded. Each user also needs to get allocated a dedicated SRS resource. Since the SRS resources are limited from an air interface perspective, not all users may have the possibility to have an SRS resource assigned if many users are connected to a cell and hence SRS reosurces should be prioritized to the users that would benefit from reciprocity-based beamforming the most at a certain point in time.

Overview of features and benefit for downlink data channel



If the four Massive MIMO features are compared with respect to our network key performance indicators of interest: coverage, capacity and single-user peak rate we can see that they have different strengths and weaknesses. Codebook-based beamforming has an advantage in coverage over reciprocity-based beamforming as was discussed on the previous slide. Similarly, SU-MIMO has a coverage advantage over MU-MIMO as the TX power needs to be split between multiple users in the latter case. When it comes to capacity, reciprocity-based MU-MIMO can provide the most (but since it is coverage is limited, the maximum benefit can only be seen in dense deployments with small cell sizes so that most of the users in the cell is within coverage).

To fully utilize to potential of a Massive MIMO system, it is thus necessary to dynamically switch between all of the four above schemes so that coverage capacity and peak rate jointly can be maximized, which is also how the typical Massive MIMO systems are designed.

Codebook-based beamforming

+ Superior coverage of the CSI acquisition, since only wideband channel information needed (e.g. ~10 bits PMI feedback)

+ Coverage extension possible with inter-band carrier aggregation by sending CSI report on low-band FDD

+ MU-MIMO for well separated users

+ Good low-rank (1-2) SU-MIMO performance

– Limited high-rank (3-4) SU-MIMO performance

– Poor MU-MIMO performance when channels are overlapped

Reciprocity-based beamforming

+ Better high-rank SU-MIMO and MU-MIMO due to improved inter-layer interference suppression capability due to spatially rich channel information

– SRS resource limitations

• Each user needs dedicated SRS resource(s)

• Need to sound the entire bandwidth to be scheduled

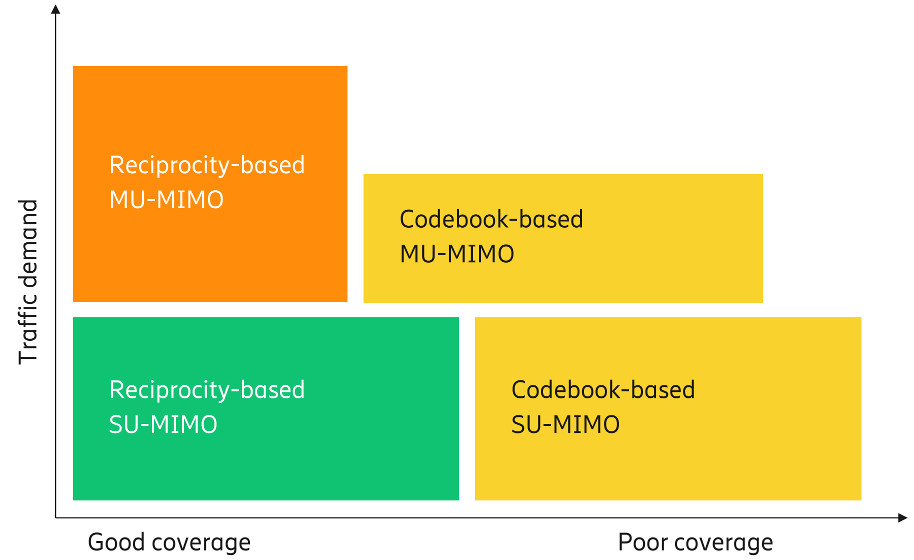
• Allocated uplink needed, either as PCell or using SRS carrier switching

• Each Rx antenna needs to be sounded (e.g. 1T4R SRS Tx switching capability needed)

– SRS coverage limitations

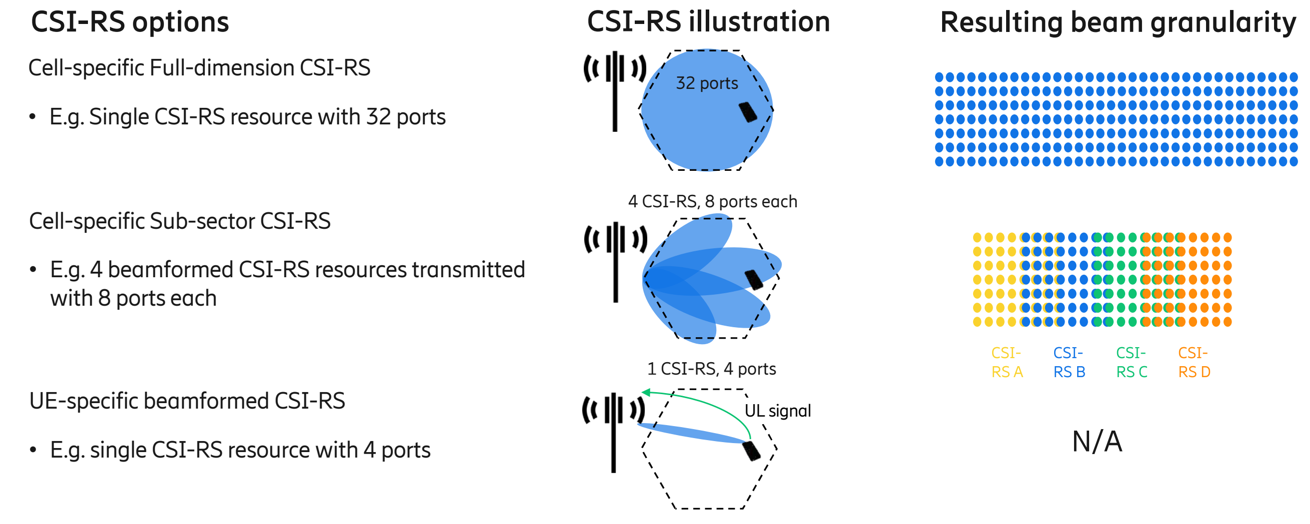
• Good uplink channel estimation quality needed → limited coverage

In which scenario is each scheme most useful?



We can visually map out where the different schemes would typically be applied on a two-dimensional grid with the traffic demand (e.g PRB utilization) of the cell on the y-axis and coverage on the x-axis.

Codebook-based beamforming options

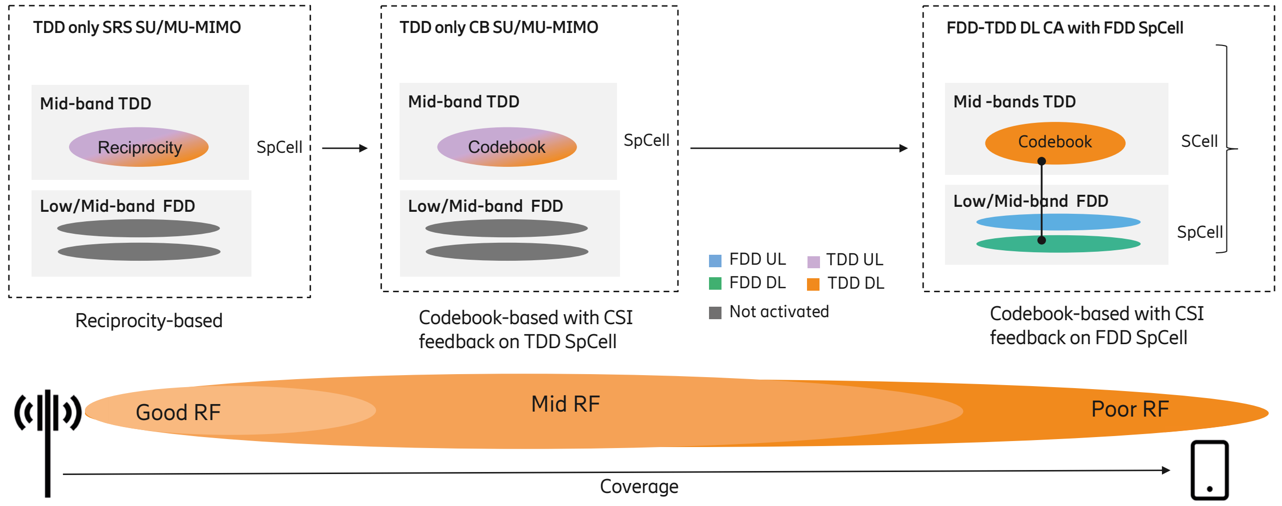


The 5G standard is flexible and generally offers multiple implementation choices. Specifically for codebook-based beamforming, there is a variety of different ways it may be realized. Three common implementation choices are illustrated here. The CSI-RS transmission can be either cell-specific or UE-specific.

Cell-specific and UE-specific CSI-RS
With cell-specific CSI-RS, a single CSI-RS resource or set of resources are broadcasted to all the UE in the cell, i.e. the UE's share a common set of CSI-RS resources, whereas with UE-specific CSI-RS each UE gets its own CSI-RS resource, beamformed towards that UE. UE-specific beamformed CSI-RS is typically used when the base station knows the beam direction towards the UE already based on some other information, e.g. uplink sounding, and the base station just wants CQI and RI information to aid the link adaptation. With cell-specific CSI-RS on the other hand, the intention is for the base station to also determine the beam direction. The RS overhead for cell-specific CSI-RS is static and does not scale with the number of UEs connected to the cell, since all UE share the same CSI_Rs resource, however for UE-specific CSI-RS the beam direction for each UE needs to be tracked and each UE requires it's own CSI-RS resource.

Full-dimension and sub-sectorized CSI-RS
Cell-specific CSI-RS transmission can further be broken down into full-dimension or sub-sector CSI-RS transmission. With full-dimension CSI-RS, one CSI-RS resource with many ports, e.g. 32, are used. For a 32 Tx Massive MIMO radio, one CSI-RS port can thus be transmitted from each Tx chain, enabling sounding of the full channel in downlink. The UE is thus aware of the full dimensional propagation channel and can select a precoder from the 32-port codebook ,which corresponds to 256 hypothetical beam directions covering the entire channel space. With a sub-sector based CSI-RS transmission strategy on the other hand, multiple beamformed CSI-RS resources, each with a smaller number of CSI-RS ports, such as 8 ports, is used. For instance, 4 CSI-RS resources, each with a different beamforming can be used to cover a different vertical or horizontal sub-sector of the cell. The UE may be dynamically or semi-statically allocated to a CSI-RS resource and the UE would report a PMI from the 8-port codebook, corresponding to a selection from 64 hypothetical beams within the beamforming pattern for the sub-sector. Each sub-sector thus corresponds to a subset of the full channel space. Depending on how the base station selects the beamforming weights of the sectors, the sub-sector may be overlapping or not.

Example: Beamforming/Carrier Aggregation (CA) mode switching



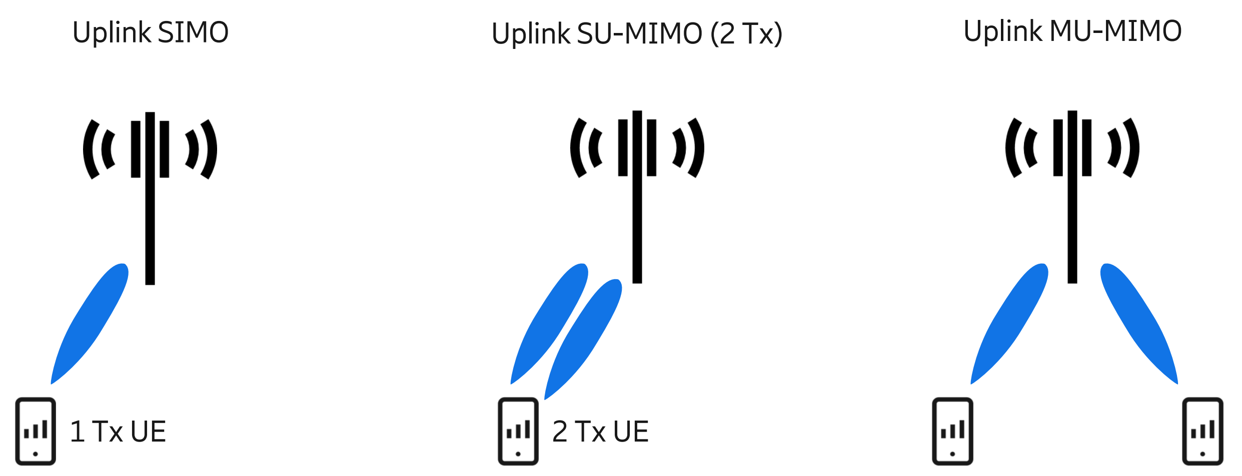
As what kind of beamforming can be used depends on what kind of CSI is available to the base station, it is important to consider the interaction with Massive MIMO feature together with other RAN features such as Carrier Aggregation. Different combinations of Massive MIMO and carrier aggregation features may be activated in different parts of the cell depending on e.g. coverage and other factors.

Consider the above example with one TDD carrier and one FDD carrier. Close to the cell center when the UE is in good RF conditions, it is beneficial to set the TDD cell as a SpCell so that the uplink is allocated to the TDD carrier (the FDD cell may then be inactivated as in this example or set as an SCell). Since the TDD uplink is in good coverage, SRS can be received with good quality and the UE can use reciprocity-based SU/MU-MIMO.

When the UE moves out towards the cell edge into the mid RF condition region, the uplink SNR may be so poor that the quality of the SRS channel reception is not sufficient to provide good reciprocity-based beamforming performance. Instead, the base station falls back to codebook-based beamforming for this UE and may also deallocate this UE's SRS resource and give it over to another UE in better RF condition.

As the UE moves further out towards the cell edge into the poor RF condition region, the TDD uplink quality may be too poor to even carry the CSI report in the uplink control information. At this point, the base station would activate downlink carrier aggregation and configure the FDD carrier as the SpCell (or simply make a SpCell change if downlink carrier aggregation is already activated with a TDD SpCell), thereby moving the uplink to the lower band with better propagation conditions, leaving the TDD carrier as a downlink only cell. The UE can still measure the CSI-RS on the TDD Scell and determine the preferred codebook beam, but it sends the CSI report with the indication on the uplink of the FDD SpCell, thereby increasing the coverage range.

MIMO features for uplink data channel

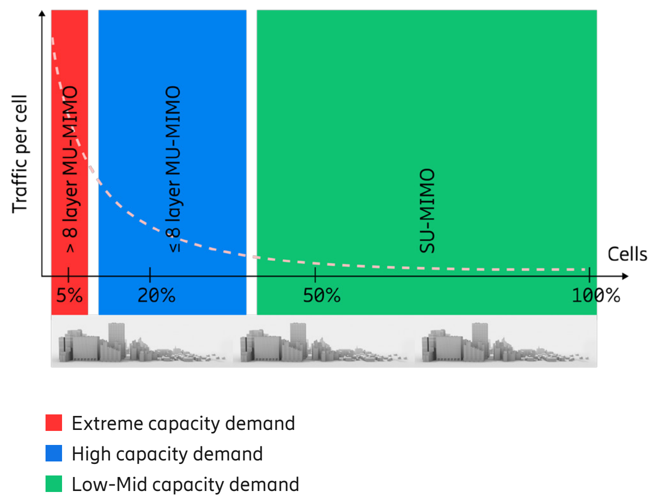


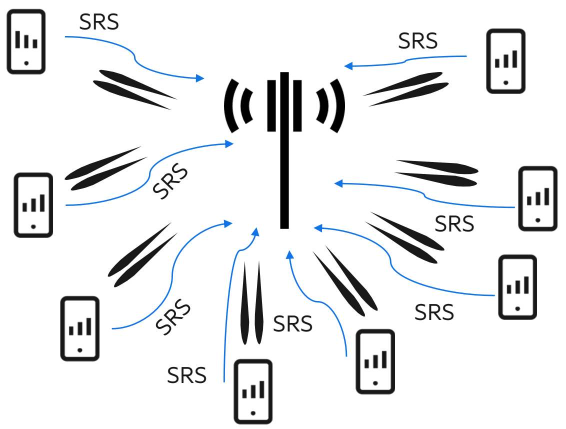
Similar to the Massive MIMO features for downlink data channel, SU- and MU-MIMO can be applied on the uplink data channel as well. However, there are some fundamental differences compared to the DL.

1. Since the Massive MIMO processing is performed at the reciever side after transmission (i.e. not before transmission as for downlink MIMO), the base station does not need to acquire CSI but can directly base the Rx beamforming/null forming processing on the DMRS transmitted with the PUSCH.
2. Commerical UEs are typically only equipped with either one or two (non-coherent) Tx chains and correpsonding antennas, which means that the precoding operation on the UE is is quite simple: transmit one layer from one antenna or two layers, each from a separte antenna / Tx chain. This means that the base station does not need to provide any CSI to the UE either, but can simply indicate transmisison rank and possibly antenna selection (although the 3GPP standard defines more advanced uplink transmission modes, these are not expected to be implemented on a larger scale).
3. Resulting uplink performance depends mainly on the reciver processing at the base station.

5.3. Advanced MIMO for extreme performance

High-order MU-MIMO





While MU-MIMO is a key feature to provide capacity in a Massive MIMO system, the number of MU-MIMO layers required to fulfill the capacity demand is generally quite few in practice. In fact, in a large portion of the network the capacity demand can be fulfilled with only SU-MIMO. Even for cells with high capacity demand, it is generally sufficient with 8 downlink layers and 4 uplink layers to reap most of the MU-MIMO gains with MBB traffic. However, the high-capacity Massive MIMO radios are dimensioned to support 16 downlink layers and 8 uplink layers, since this could be required in certain cell sites with extreme capacity demands, such as traffic hotspots or stadium/arena deployments with a lot of people in a small area, or non-MBB use cases such as FWA. Higher-order MU-MIMO can also be useful for scenarios with different traffic characteristics than bursty packet-based MBB traffic. MU-MIMO is further elaborated on in [Ch. 3, p. 27 - 33, p. 37- 38] and [Ch. 6, p. 101].

Coordinated multi-point (CoMP)

Coordinated beamforming (CBF)
Avoids inter-cell interference

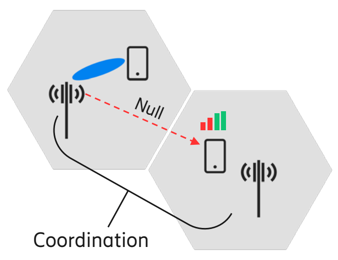
Coordinate nulls towards victim UEs

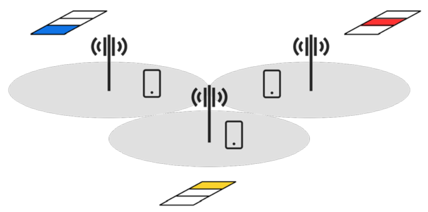
Coordinated scheduling (CS)
Avoids inter-cell interference

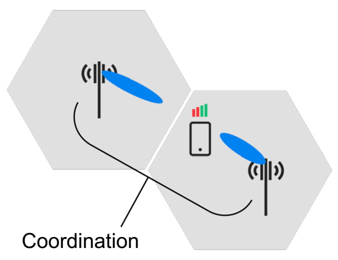
Coordinate resource allocation
Coordinate link adaption

Joint transmission/reception (JT/JR)
Increases signal level
May increase or decrease interference

JT: Multiple points/cells transmit to a UE
JR: Multiple points receive from a UE







Coordinated multi-point transmission/reception, or CoMP, utilizes cooperation and coordination between multiple cells (or transmission/reception points, TRPs, of the same cell) to improve network performance, primarily by either aiming to reduce inter-cell interference or utilizing the transmission/reception resources of multiple TRPs to boost the signal energy of a user. Many different flavors of CoMP exists, with different pros and cons, different requirements on backhaul/coordination and applicability to different scenarios.

Coordinated scheduling
One set of CoMP features utilizes coordination on the scheduler level to for instance adapt the resource allocation between TRPs to reduce inter-cell interference. For instance, the scheduler can make a dynamic decision to not schedule any UEs from one or more TRPs, thereby reducing interference to the UEs served by the remaining TRPs. Another flavor of coordinated scheudling is coordinated link adaptation, CoLA, where each TRP makes scheduling decisions independently, but exchange information between one another so that the link adaptation can take information about interference from the neighboring TRPs into account.

Coordinated beamforming
With coordinated beamforming, the intention is for reduce the generated inter interference by adapting the beamforming pattern so that neighboring cell UEs are not struck by interference. Typically, this can be achieved by generating nulls towards the victim UEs in the neighboring cells, similar to how nulls are generated towards the other layers/UEs in a MU-MIMO scheduling. Hence, coordinated beamforming can eb seen as an extension of MU-MIMO null forming also considering neighboring cell UEs.

Joint transmission/reception
For coordinated scheduling and beamforming, the UE is only receiving transmissions from a single TRP at a given time, in joint transmission (JT), the UE can receive transmissions from multiple TRPs simultaneously. JT can further be broken down into two flavors, non-coherent JT (NC-JT) and coherent JT (C-JT). For C-JT, the same precoding layer(s) are transmitted from the multiple cooperating TRPs with the intention for the transmissions to coherently add up at the receiving UE. In order to accomplish this, tight synchronization is required between the TRPs and a high CSI accuracy is needed in order to design the precoding weights, which makes C-JT difficult to practically implement and requires large CSI feedback overhead. In NC-JT, on the other hand, the requirement for synchronization and CSI accuracy is comparably lower, since each TRP is transmitting different layers.

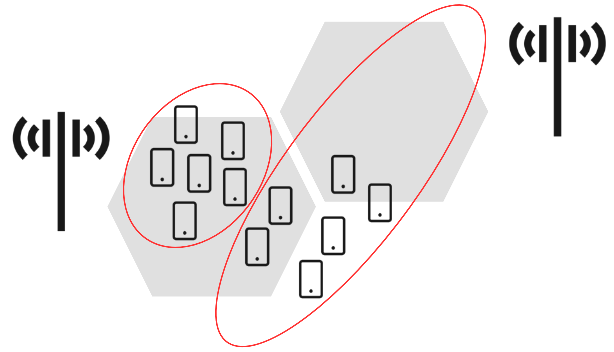
Analogously, with joint reception (JR) the received signal (either before or after demodulation and mapping to soft bits) from multiple TRPs can be accumulated, hence boosting the signal level and increasing decoding performance.

(Automated) Cell-shaping

- Shape the beam(s) of cell defining signals
  - Broadcast signals such as SSB in NR
- Slow update rate – static or semi-static

Possibility to tailor the cell shape
  - Suiting the actual spatial UE distribution

- Lower inter-cell interference – avoid cell-edge in hotspot
- Raise the signal level – point where the UEs are
- Load balancing – move UEs from one cell to another
- **Combined with machine learning makes it a self optimizing network (SON)**



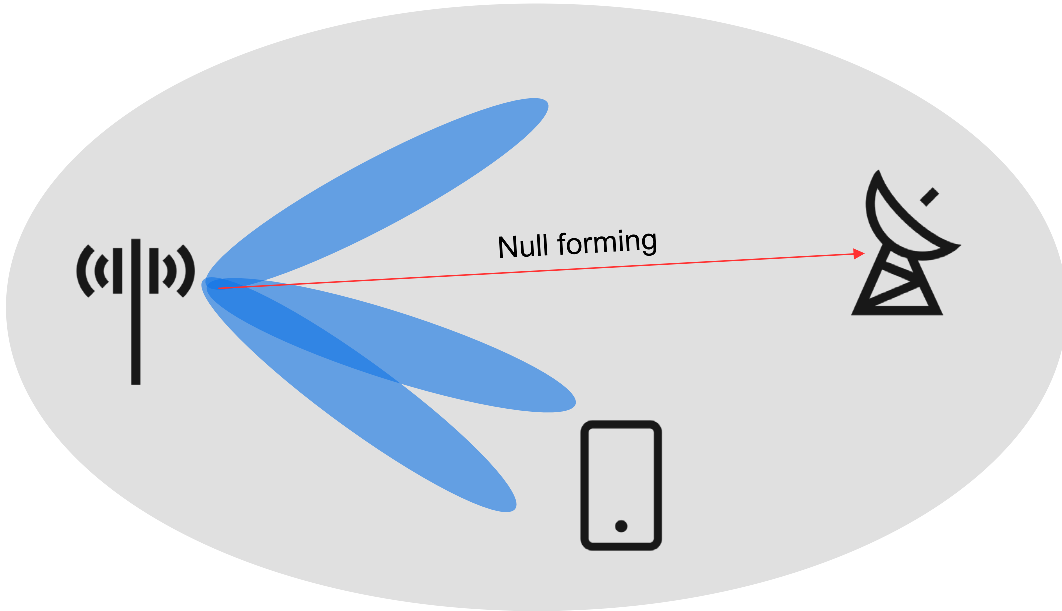
The cell shape depends on the beam shape of the cell-defining signals, such as the SSB in NR. This provides the possibility for the network to tailor the cell shape for some specific purpose by changing the beam shape and slowly update the cell shape instead of striving for fixed cell areas with potentially hexagonal structure. The beam shape of a cell defining signal is controlled by the use of a beamforming vector (1) that maps the signal onto the antennas, just as in the dynamic beamforming case. By measuring various parameters in the network, suitable beam shapes can be determined. The technique of letting the system adapt and update the beam shape of cell-defining signals is called cell shaping. Cell shaping can be used together with dynamic and UE specific beamforming to further improve performance or it can be used stand alone.

Cell shaping can be used to solve various problems. For example, the cell-border with its interference impairments can be moved to avoid being in the middle of a location where typically there are many simultaneously connected UEs, known as a hotspot. An example of such a hotspot is a train station. Using cell shaping to avoid a cell border within a hotspot is illustrated in the figure. Cell shaping also offers the possibility for load balancing among cells by moving UEs from one cell to another when altering the uptake areas of cells.

Cell shaping can also reduce interference in unwanted directions. Such a reduction is especially important in transmit directions that negatively impact particularly many UEs in neighboring cells.

5.4. Deployability, cost and energy efficiency

Directional interference mitigation



Some mid-band TDD bands in certain regions have spectrum coexistence requirements with regulatory limitations on generated interference towards spectrum incumbents, such as satellite earth station receivers, which may be operating either in-band or on adjacent band. The Directional Interference Mitigation feature enables limiting the radiated power of a Massive MIMO in certain sensitive directions, such as the direction of an earth station receiver, by utilizing spatial null forming. This enables the radiated power to be reduced only in those directions required , but full power to be utilized in other non-sensitive directions, to maximize the coverage and capacity of the cell.

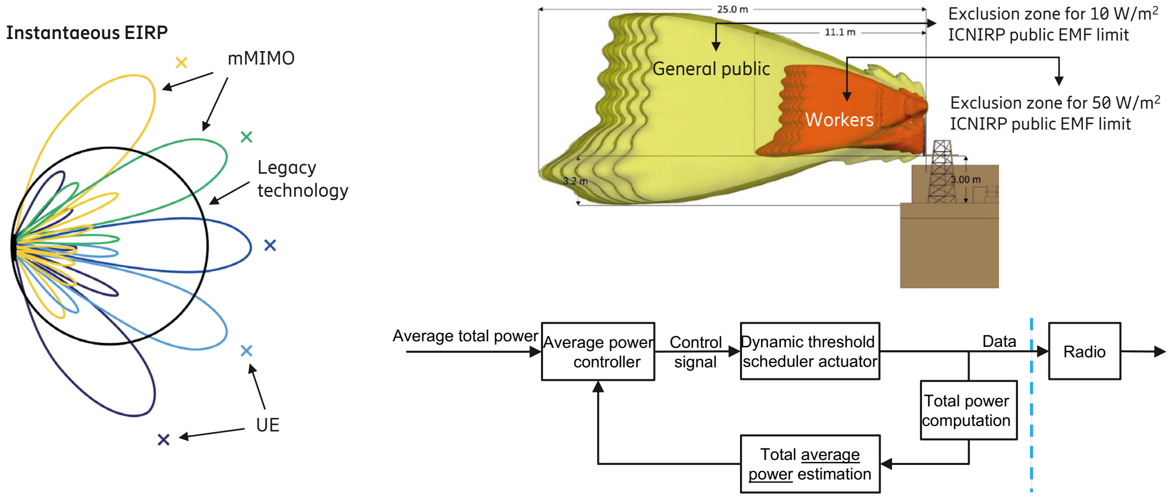
Spatial nulls towards the sensitive directions can be inserted in both the broadcast beam, CSI-RS beam as well as PDSCH beam.

The interference generated towards particular sensitive directions can be reduced, in order to comply with regulatory requirements, without reducing the signal level in other directions.

Broadcast beam pattern can be adopted to change the cell uptake area to not attach users in the sensitive directions.

Electromagnetic Field (EMF) Power lock protection

On Massive MIMO solutions, the instantaneous EIRP and the resulting peak RF EMF levels can be higher than those of traditional BS antennas

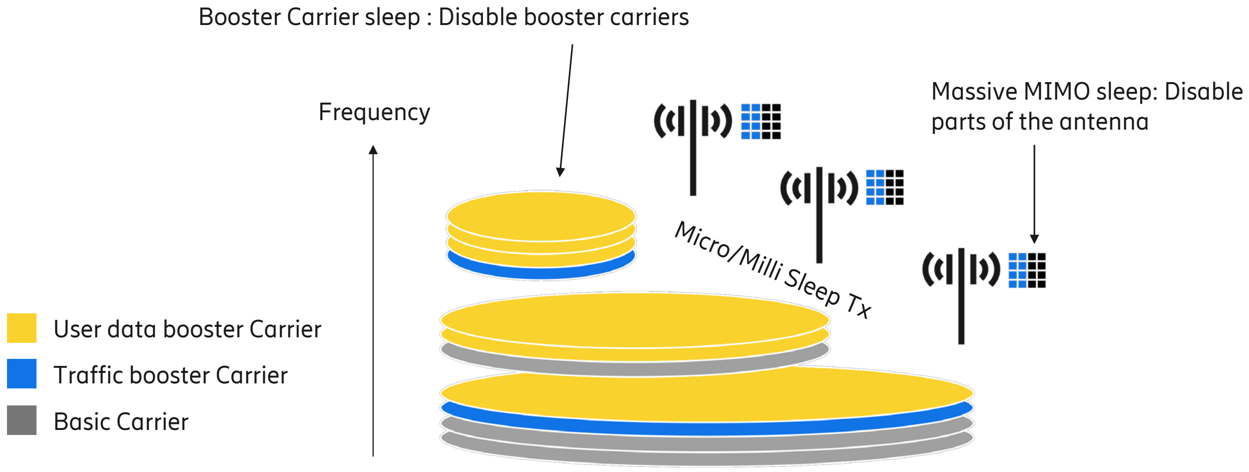


The use of Massive MIMO solutions means the instantaneous EIRP and the resulting peak (non time-averaged) radio frequency (RF) electromagnetic field (EMF) levels can be higher than those for traditional base-station antennas. The RF EMF exposure is regulated in most markets, the RF EMF exposure limits are typically defined as time-averaged values, e.g. over 6 minutes.

If time-averaging is not considered, the size of the RF EMF exclusion zone increases when the BF gain increases. Such an increased exclusion zone could make deployment challenging in e.g. dense urban environments. IEC TR 62669 - is recommending monitoring and control of the time-averaged power, to a pre-determined level, to obtain a certain exclusion zone, with a minimum impact on the capacity and coverage of the served cell.

EMF Power Lock controls the average transmit power per sector carrier allows the configuration of a power back-off threshold ranging btw 20.6% - 100% of the max available total power by controlling the scheduled number of PRBs for PDSCH.

Energy efficiency Massive MIMO features



Massive MIMO solution provides high capacity and performance which is useful for handling high traffic load scenarios such as repeating busy hours and also occasional peak loads.

In most traffic load scenarios, even in the most loaded areas, there are still periods with low traffic load where the available capacity is not required. During these periods power can be saved by deactivating not required capacity. By switching off radio unit components power consumption is reduced.

Different Massive MIMO radios have different capability in terms of what HW components that can be switched off. The RAN SW part of the solution is agnostic to different radio units. This enables a feature such as Massive MIMO Sleep to work with any Massive MIMO radio.

Apart from Massive MIMO specific energy saving solutions there are also RAN level solutions that for each coverage area adapts provided capacity to the actual needed by sensing and predicting traffic.

Different features, handling the different perspectives, collaborates to provide the required service level of capacity and performance. Three types of features are in the solution:

- Micro/milli sleep Tx, autonomous discontinuous transmission features, act to set different components in radio PA in sleep based on empty slots in transmitted down link. No impact on any key performance indicator.
- Frequency domain-based Booster Carrier Sleep that deactivates not required carriers. All available carriers are tagged as Basic or Traffic Booster or User Data Booster carrier. Basic carrier is the carrier that will transmit necessary cell info such as system information, thus providing the cell coverage. It will always be on. The two Booster carriers will be set to sleep based on need.
- Massive MIMO Sleep is a Massive MIMO radio specific feature that will deactivate (set to sleep) not required segments. Coverage is maintained.

All features can be activated and they will inter-work to find the best power savings. In addition to have even more savings Energy Performance Optimizer (EPO), a central AI based tool, can orchestrate the different energy savings features to find even more savings. By using advanced traffic sensing and prediction functionalities EPO can find the periods for saving and also decide the most efficient sleep level for the equipment. The deeper sleep the longer activation time. Deeper sleep means more parts of the radio unit will be deactivated.

6. Network Performance

Network performance – Outline

- Radio network performance evaluations
- Key aspects impacting M-MIMO performance
- Capacity dimensioning
- TCO analysis
- Number of layers
- Frequency interworking
- Feedback vs reciprocity
- Ericsson M-MIMO architecture

Performance evaluations are vital in the research and development (R&D) flow. It provides an understanding of how different deployments, configurations and algorithms perform before making costly R&D product investments. It can also guide our customers when it comes to network evolutions and product needs.

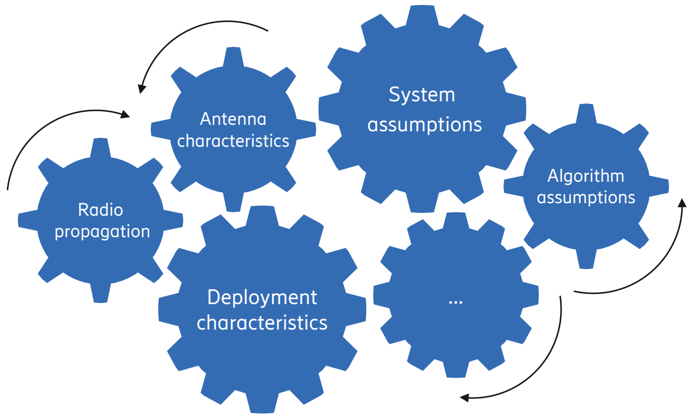
The radio network performance depends on many aspects that interact in an intricate manner, and it is important to understand the driving mechanisms and having models that capture the reality sufficiently well while still being computationally efficient. Ericsson are in the fore-front in the area of radio network performance evaluations using state-of-the-art knowledge and models. The predictions based on simulations have proven to match real network performance on countless occasions.

This chapter will give some insights into the area of radio network performance for Massive MIMO. Predictions based on simulations will also be compared with measurements in real networks. The chapter is divided into three parts:

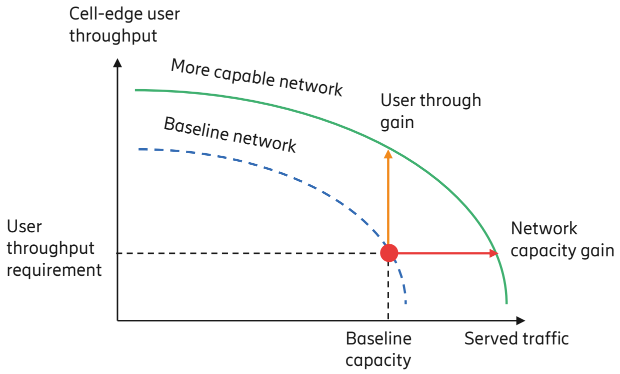
- Very short introduction to radio network evaluations.
- An overview of key aspects impacting Massive MIMO performance will be given, followed by some illustrations of capacity dimensioning with TDD mid-band and Massive MIMO, and finally a total cost of ownership analysis showing how Massive MIMO often provide superior cost per capacity.
- Deep-dives into a few selected areas including the impact of the number of MU-MIMO layers, benefits of frequency band interworking, impact of type of channel state acquisition approach (feedback-based vs reciprocity-based), and effects of the Ericsson Massive-MIMO architecture.

Radio network performance evaluations

Network characteristics matter!
Need to capture a diverse and complex reality



Key performance metrics
Coverage, capacity and end-user throughput



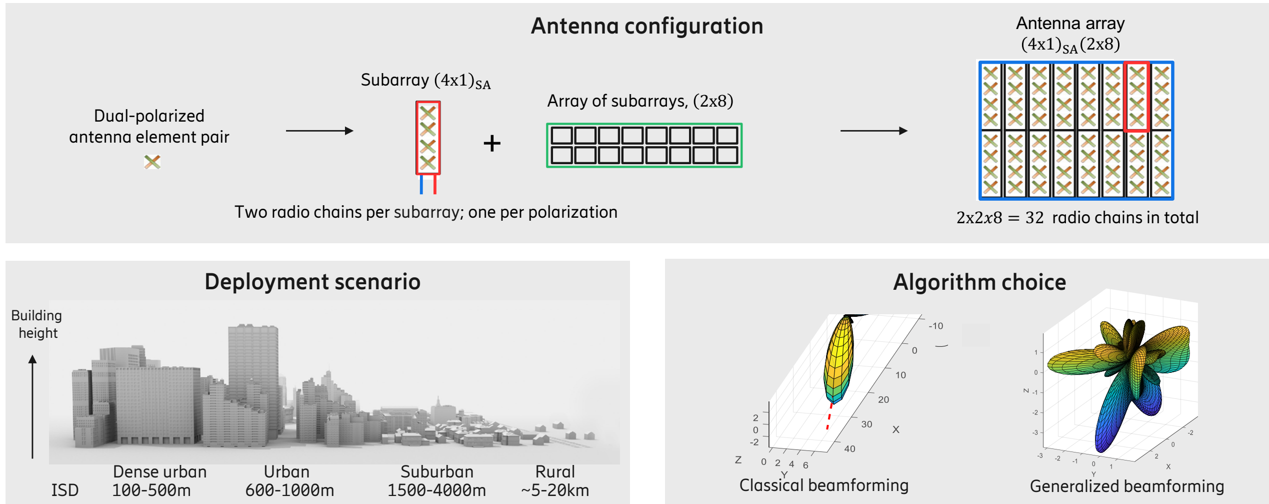
Performance on radio network level depends on many aspects that interact in an intricate manner, and it is important that there are simulator models that capture the essential characteristics of the reality and of the radio access network features to be evaluated. Depending on the feature to evaluate and the results of interest, certain aspects are often more important than others, and for massive-MIMO features, spatial characteristics are obviously very important. Some key aspects impacting Massive MIMO performance include deployment scenario and radio propagation, antenna characteristics, and algorithm assumptions. These aspects will be further described in the following slides. System assumptions, such as frequency band, output power and bandwidths can also have a significant impact on the performance.

Some key performance metrics reflecting needs in the network are coverage, capacity, and user throughput [3; Ch. 1, p. 11]:

- Different metrics reflecting coverage can be envisioned, but they all try to quantify how reliably user data and control information can be transmitted. A common measure of coverage is to look at the achievable user throughput for the worst users in the network, the so-called cell-edge users (often defined as the 5th percentile worst users in the network). Coverage is a key performance metric when dimensioning a network as it relates to the number of required sites and therefore also directly affects the total network roll-out cost.
- For packet-based MBB services, capacity is often expressed as the average served traffic (bits per second) per cell or per area unit given a certain resource utilization or user throughput requirement at a specific UE percentile, often the cell-edge user percentile. Capacity is a key metric to decide how many users or how much traffic the network can sustain with maintained coverage. The capacity is often of high interest for operators as it reflects the production cost (cost per bit).
- User throughput reflects the user-experience and is closely related to coverage and capacity. Throughput will differ between users and depends on their respective link quality, total cell or network traffic load, scheduling, system assumptions, etc. It is common to assess user throughput gains for different user percentiles, for example, the 5th (cell-edge), 50th and 95th percentiles, at specific network loads.

6.1. Key aspects impacting Massive MIMO performance

Key aspects impacting Massive MIMO performance



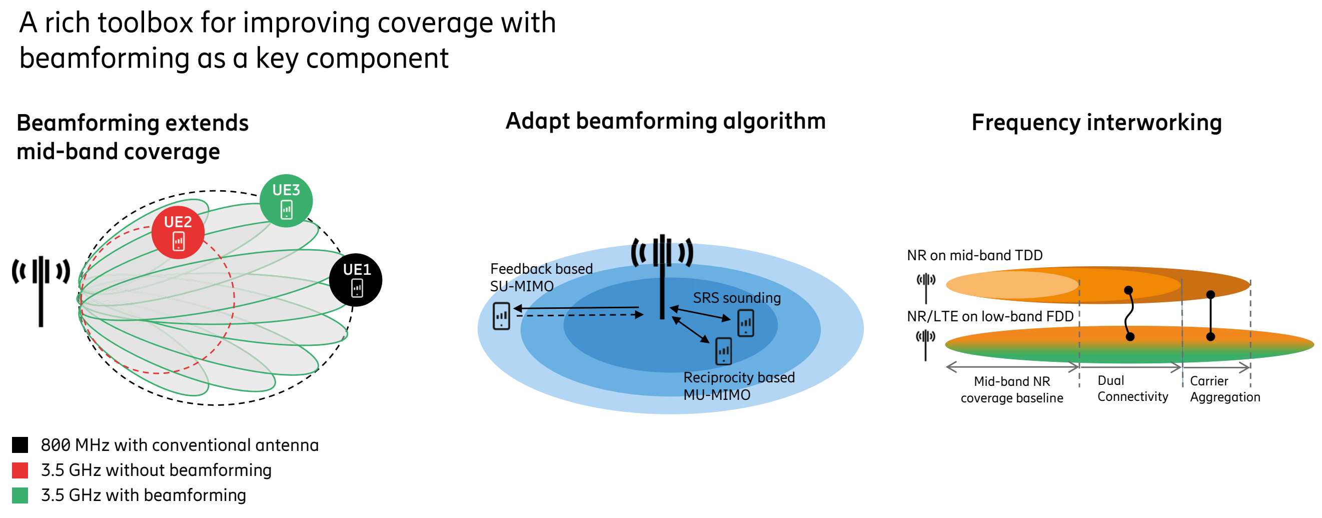
This slide highlights three factors that have a substantial impact on the performance of Massive MIMO evaluations: the antenna configuration [3; Ch. 1, p. 14], [3; Ch. 2, p. 17], the deployment scenario [3; Ch. 2, p. 24-25], and the Massive MIMO algorithm choice [3; Ch. 1, p. 12], [Ch. 3].

The antenna configuration (upper figure) is central for the performance. The antenna area determines the maximum antenna gain (coverage). The subarray consists of a number of semi-statically combined dual-polarized antenna element pairs and is the smallest dynamically controllable entity (two radio chains per subarray, one per polarization). It determines the angular coverage area (in the vertical domain), i.e. the area where beamforming can be done without significant gain drop. A small subarray (few vertical elements) gives a wider coverage area but also less gain, whereas a larger subarray (more elements) gives a narrower coverage area but also higher gain. The number of radio chains (Ts and Rs) determines the degrees of freedom (or steerability) for beamforming. Stacking subarrays horizontally (radio chains in the horizontal domain) provides horizontal domain beamforming and stacking subarrays vertically facilitates vertical domain beamforming. Stacking subarrays of an appropriate size (increasing the number of radio chains) enables the use of a larger antenna, hence more gain that can be steered over the entire service area.

The deployment scenario (lower left figure) including inter site distance, the radio environment, user distribution in both horizontal and vertical domains, etc. has a substantial impact on the performance and how different Massive MIMO configurations perform. In particular, the user distribution in the vertical domain impacts what subarray size that is feasible and the usefulness of vertical domain beamforming, and the inter site distance affects the coverage.

The Massive MIMO algorithm choice (lower right figure) is yet another important factor affecting the performance. Two main classes of beamforming are so-called classical beamforming and generalized beamforming [Ch. 3, p. 24]. Two other central aspects are how the channel state information needed for beamforming is acquired and what granularity of channel state information that is needed. See also [Ch. 3, p. 34 - 36] and [Ch. 5, p. 62].

Coverage of 5G networks



Coverage is a key performance metric when dimensioning a network as it relates to the number of required sites and therefore also directly affects the total network roll-out cost. Coverage matters from day one when deploying a new 5G network, whereas capacity needs grow gradually.

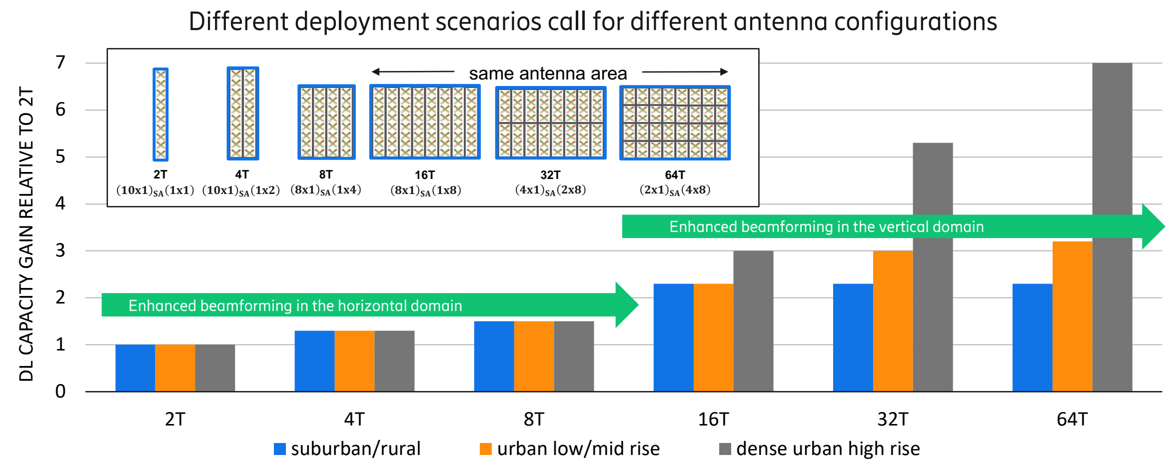
Beamforming is a key feature for increasing coverage. Maximum antenna gain is determined by the antenna size, hence from a coverage perspective, both uplink and downlink, it is beneficial to have as large antenna as possible. One potential challenge with a large antenna is that the beamwidth is inversely proportional to the gain. Hence, if you want gain, the corresponding beams will be narrower. One solution to this is to employ UE-specific beamforming, where subarrays of a size that fits the deployment scenario are stacked horizontally and/or vertically. Stacking more subarrays renders a larger antenna, hence more gain, and by using UE-specific beamforming, the gain can be directed over the entire coverage area given by the subarray size.

The left-most figure illustrates that by employing beamforming, the coverage of the 3.5 GHz deployment can match the corresponding 800 MHz deployment when both deployments use the same site grid.

Other solutions to the coverage problem include (see [Ch. 5, p. 62] for details):

- Algorithm adaptation (middle figure): dynamically adapt to the coverage situation, e.g. use more advanced beamforming algorithms that rely on better channel state information in areas with good coverage, and more basic beamforming algorithms requiring less accurate channel state information when coverage is worse.
- Frequency band interworking (right figure): employ frequency band interworking with dual connectivity or carrier aggregation. By using carrier aggregation with cross carrier scheduling, one can maximize the area where downlink data can be sent on mid-band by sending coverage limited, but essential data and control info (often uplink related) on a lower band with better coverage.
- Increased output power (including power boosting), better receiver sensitivity, more robust coding or repetition are examples of other common ways of increasing the coverage.

Capacity simulations for Massive MIMO



Capacity is a key metric to decide how many users or how much traffic the network can sustain with maintained coverage. Capacity is here defined as traffic per area unit (bits per second per area unit) given a user throughput requirement of 10 Mbps for the 5th user percentile (so-called cell-edge users).

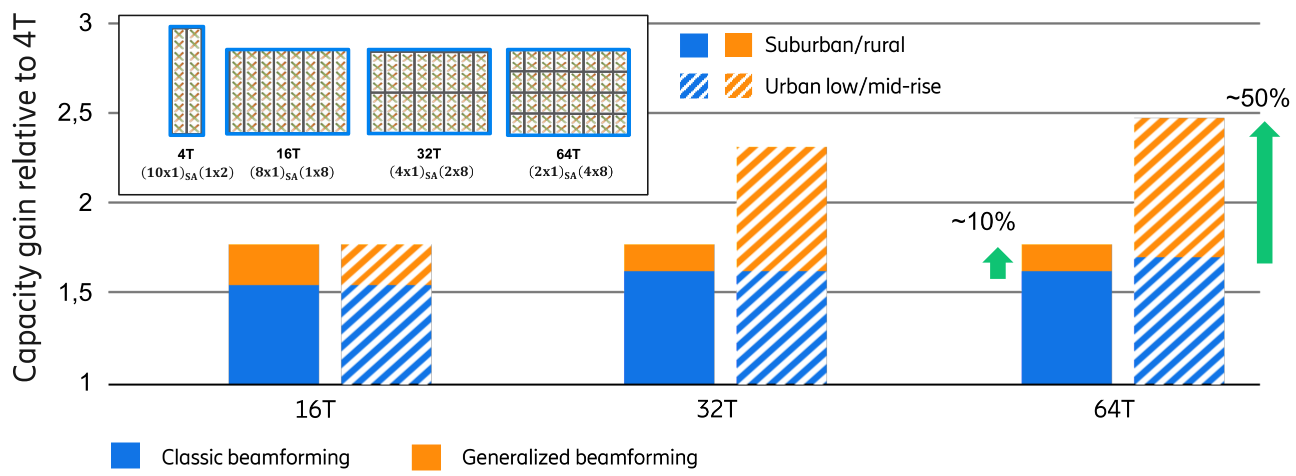
The figure shows capacity gain relative to 2T classical beamforming for different antenna configurations [Ch. 2, p. 17] in different deployments. The antenna configurations range from 2T to 64T, and generalized beamforming using a mixture of feedback-based and reciprocity-based channel state information (referred to as Reci-GoB in [Ch. 6, p. 105]) is used.

The key takeaway is that the effectiveness of a specific antenna configuration is deployment dependent. More specifically, horizontal domain beamforming (2T to 16T) is useful in all deployments, and a common rule-of-thumb is that capacity increases by 30-50% by doubling both the antenna area and the number of radio chains in the horizontal domain, e.g. going from 2T to 4T or from 8T to 16T. Gains of vertical domain beamforming (16T to 64T) depends on the deployment, and a key parameter is the angular distribution of the users in relation to the size of the subarrays.

- In the rural/suburban deployment (~1500 m inter-site distance), the gains of going from 16T to 32T or 64T are rather small. The main reason is that the 8x1 subarray used in 16T fits well the narrow angular spread of users in the vertical domain.
- In the urban low/mid-rise deployment (~500 m inter-site distance), the gains of going from 16T to 32T is significant, but the additional gain by going from 32T to 64T is small.
- Results in the dense urban high-rise deployment stand out with substantial gains by going from 16T to 32T to 64T. Compared to the other deployments, this scenario has smaller inter-site distance and tall high-rise buildings, hence better coverage, substantially more spread of users in the vertical domain and more channel angular spread. Also, the inter-cell interference becomes problematic at high load. All this combined makes Massive MIMO superior to other solutions.

Capacity simulations for Massive MIMO

Opportunistic gains from advanced beamforming with MU-MIMO



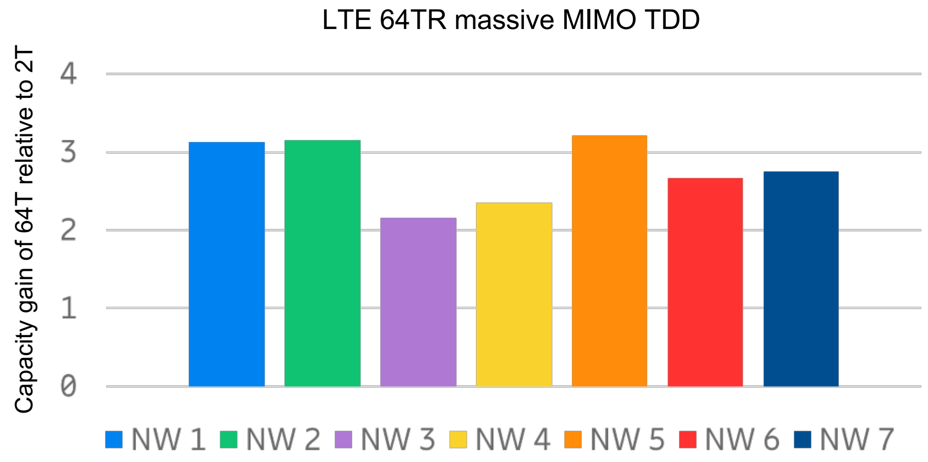
The previous slide illustrated that the performance gains of deploying Massive MIMO depends on the deployment scenario. The results in this slides illustrates the performance impact of different beamforming algorithms.

The figure shows capacity gains relative to 4T classical beamforming for different antenna configurations [Ch. 2, p. 17] using classical or generalized beamforming [Ch. 3, p. 24] in rural/suburban and urban low/mid-rise deployment scenarios. Classical beamforming uses SU-MIMO feedback-based channel state information with a grid-of-beams type of codebook, whereas generalized beamforming uses a mixture of SU/MU-MIMO and feedback-based and reciprocity-based channel state information (referred to as Reci-GoB in [Ch. 6, p. 105]).

There is a deployment and antenna configuration dependency on the gain of generalized beamforming over classical beamforming. For example, for the 16T antenna configuration, the gain of generalized beamforming compared to classical beamforming is similar for both deployment scenarios, whereas the corresponding gain for the 64T antenna configuration is ~10% and ~50% in suburban/rural and urban low/mid-rise deployment scenarios, respectively.

Capacity results from initial trials of Massive MIMO

Results match our state-of-the-art radio network predictions

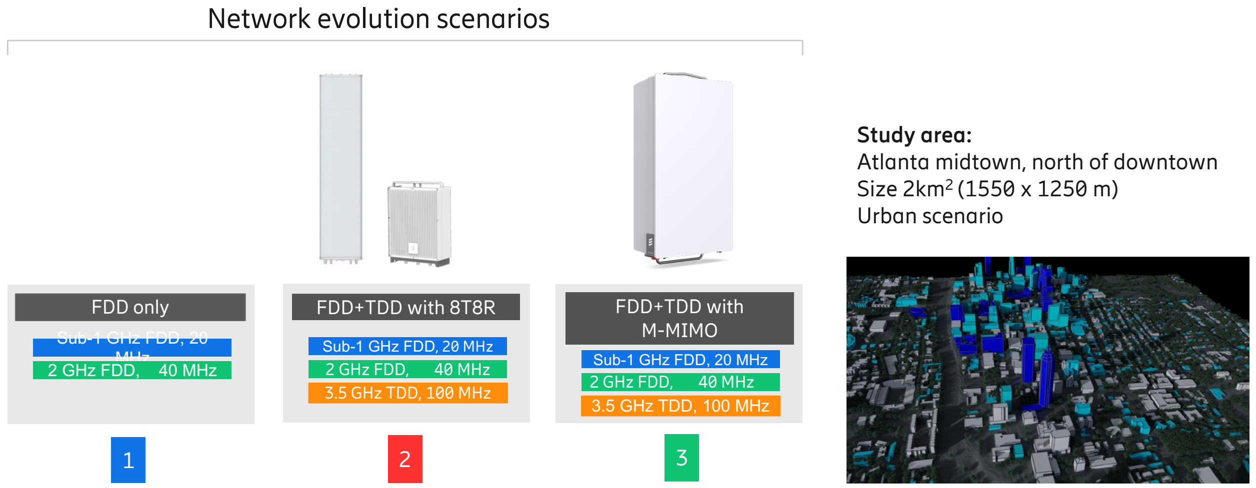


Early trial and deployment result from LTE deployments of AIR6468 64T64R radio in different networks (NW) around the world.

The observed results of ~2-3x capacity of 64TR compared to 2TR in different networks coincide well with the predictions made by our simulations, see e.g. [Ch. 6, p. 84].

6.2. Illustration of capacity dimensioning

Quantifying the benefit with mid-band TDD and Massive MIMO



The following slides will illustrate how different radio deployment options will handle the expected traffic growth over the coming years.

The deployment scenario is Atlanta midtown (US dense urban scenario), and results are based on multi-band simulations for three different radio deployment options:

1. FDD bands only: FDD 700 MHz with 20 MHz bandwidth and 2T2R radio, and FDD 2 GHz with 40 MHz bandwidth and 4T4R radio
2. FDD bands plus 3.5 GHz TDD with 100 MHz and 8T8R radios
3. FDD bands plus 3.5 GHz TDD with 100 MHz and Massive MIMO (64T64R)

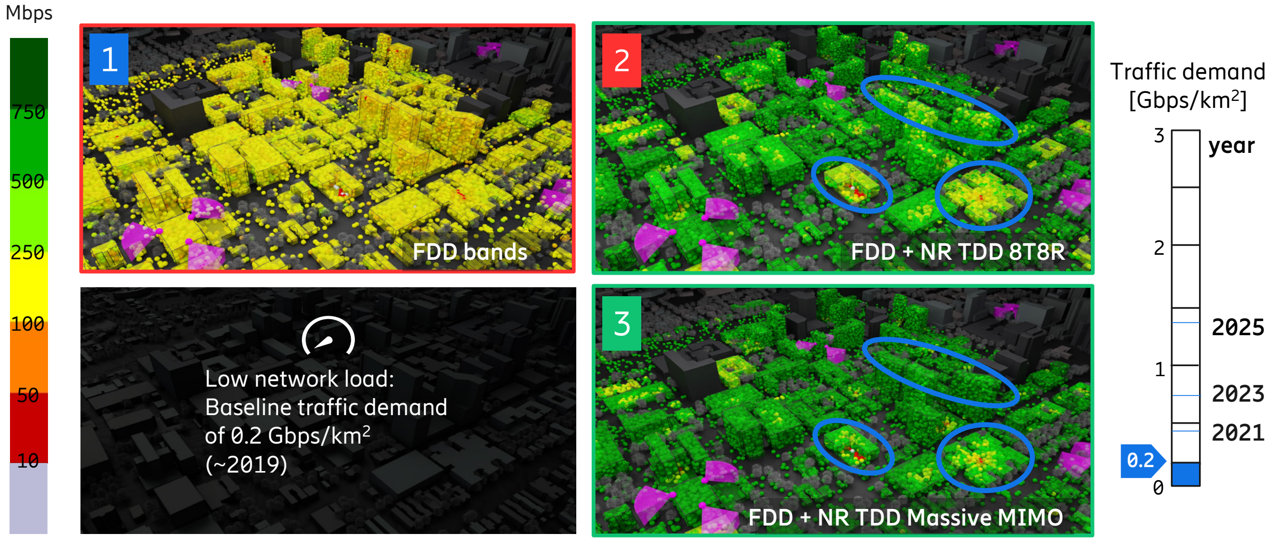
The coming slides will show user throughput (represented by different colors) for the individual users in the city map at different traffic loads. The first four slides present downlink results, followed by three slides showing uplink results

The main takeaways from this example are

- adding NR TDD mid-band with Massive MIMO provides a future proof investment that can handle the anticipated traffic growth for many years
- low-band frequency deployments are essential to ensure uplink coverage and superior uplink user experience

Downlink user experience visualized

Baseline load: NR TDD provides massive boost in user experience



This slide shows downlink user throughput (represented by different colors) for the individual users in the Atlanta city map at a baseline traffic load of 0.2 Gbps/km2, corresponding to the traffic demand ~2019.

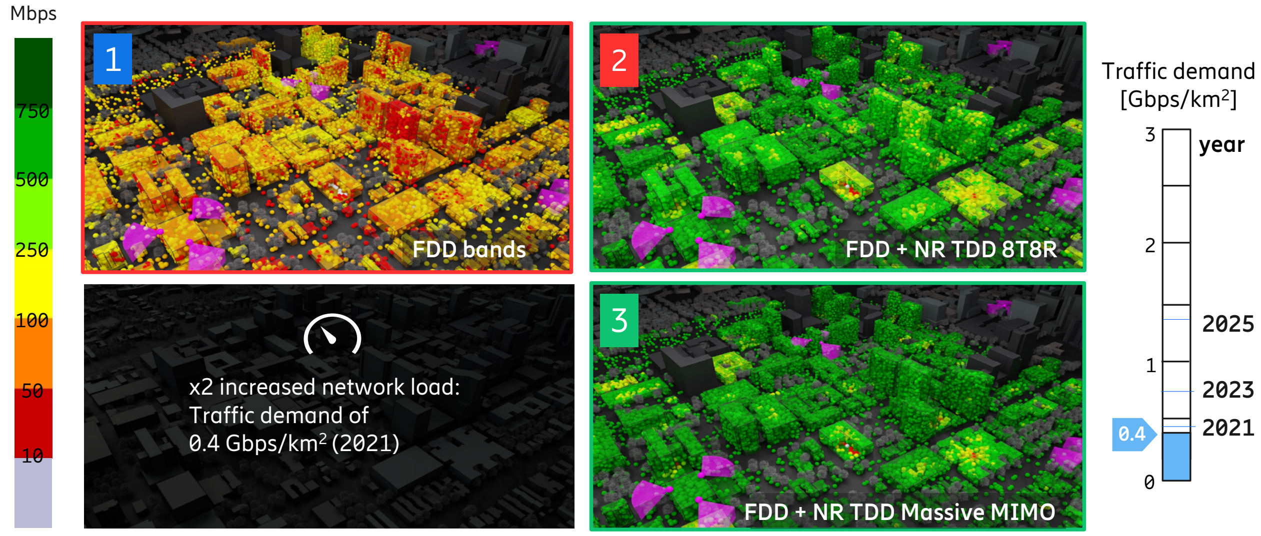
Results for the three different radio deployment options described in the previous slide are marked with 1, 2 and 3, respectively.

It is clear from the results that adding NR TDD mid-band to existing FDD bands provides a substantial boost in performance. Only FDD bands (1) provides downlink speeds in the range of 100-250 Mbps but by adding TDD mid-band, speeds exceed 500 Mbps.

It is also seen that differences between the 8T8R radio deployment and Massive MIMO are only noticeable in a few regions marked with blue circles. These areas contain high-loss buildings with high path loss, hence the extra beamforming gain offered by Massive MIMO is beneficial.

Downlink user experience visualized

x2 load: FDD bands become gradually over-loaded



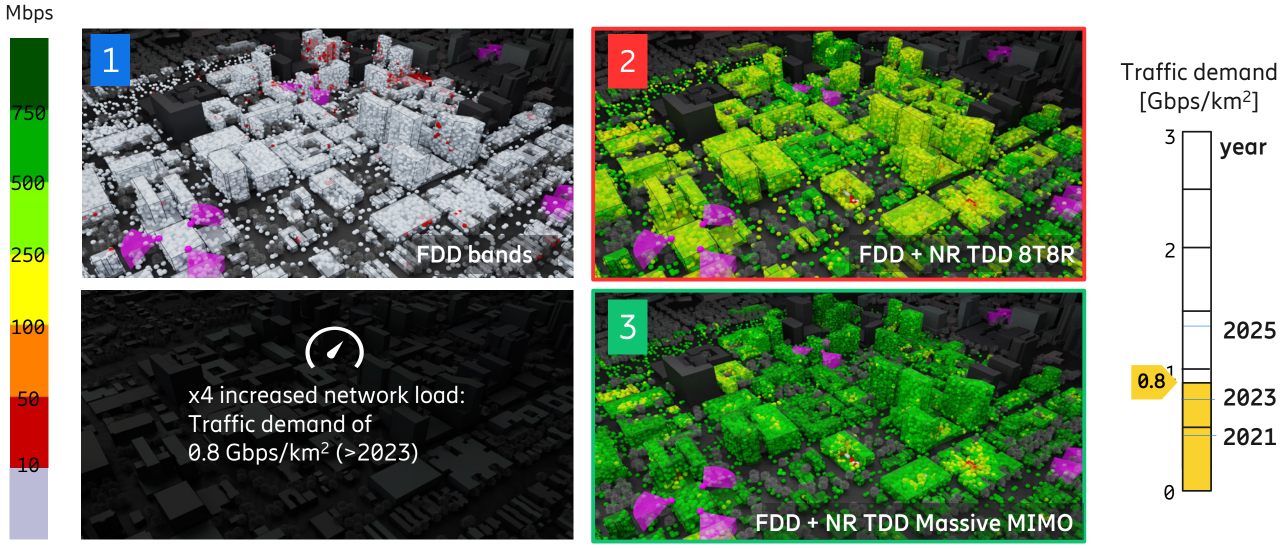
This slide shows results after doubling the baseline traffic demand, now reaching 0.4 Gbps/km2, corresponding to the traffic demand in ~2021.

The FDD bands alone are not sufficient to deliver expected downlink speeds. There are many orange and red areas in the city model indicating downlink speeds that do not meet the expectations from a 5G experience.

By adding the TDD band, the user experience is mainly on green levels throughout the network, but compared to the previous slide, some more yellow areas are appearing, in particular when employing 8T8R.

Downlink user experience visualized

x4 load: emerging benefits of Massive MIMO vs conventional 8T8R

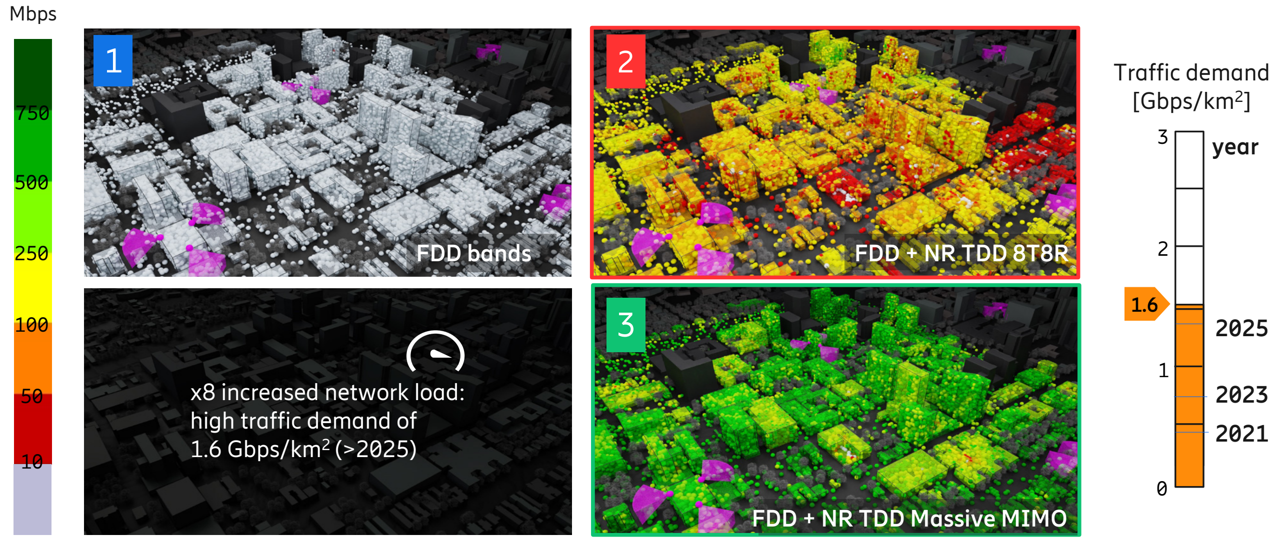


This slide shows results after yet another doubling of the traffic demand, now reaching 0.8 Gbps/km2, corresponding to an expected traffic demand in 2023+

Clearly, only using the FDD bands cannot support the traffic demand, and adding mid-band becomes necessary. It is also clear that Massive MIMO starts to provide substantial gains compared to 8T8R. Massive MIMO provides downlink user throughputs in the order of 250 Mbps in the network while for the 8T8R radio solution, there are many users experience 100 Mbps or lower.

Downlink user experience visualized

x8 load: substantial gains of Massive MIMO over conventional 8T8R

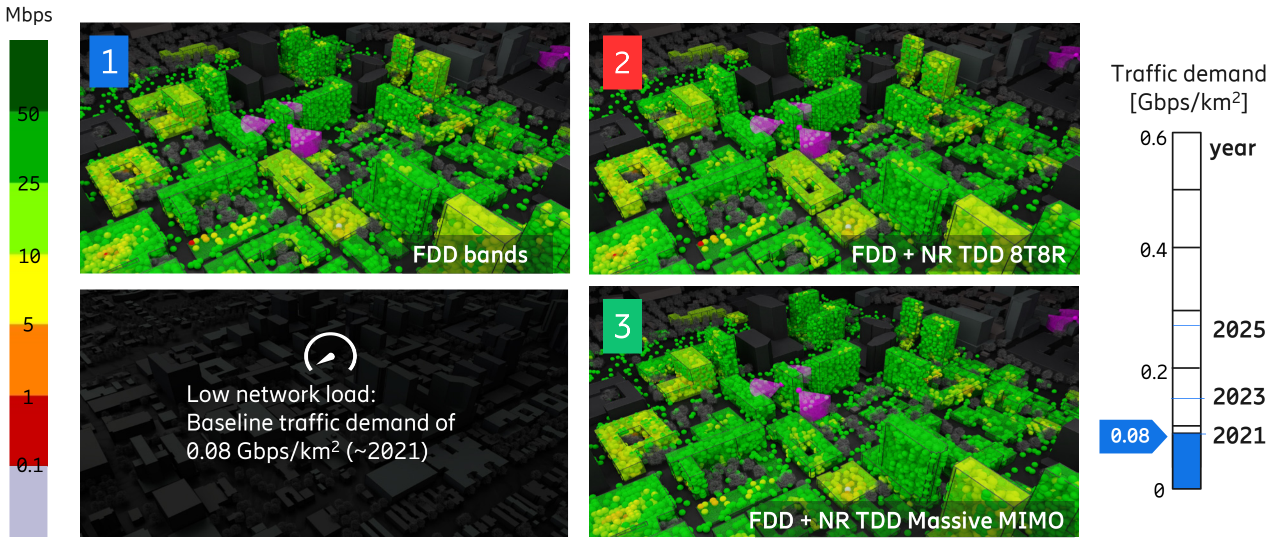


As a final illustration, results with yet another doubling of the traffic demand, reaching 1.6 Gbps/Km2 corresponding to the expected traffic demand in 2025+, are shown.

These results clearly illustrate the superior performance offered by adding TDD mid-band with Massive MIMO to the existing site solution. The radio solution with FDD bands plus mid-band with conventional 8T8R starts to become over-loaded, while the corresponding radio solution using Massive MIMO on mid-band still provides a superior user-experience with downlink speeds well above 250 Mbps.

Uplink user experience visualized

Baseline load: lower FDD bands are key for uplink coverage



To complement the downlink results shown in the previous slides, the following slides will show uplink user throughput. The baseline traffic load is 0.08 Gbps/km2, corresponding to the traffic demand ~2021.

Results for the three different radio deployment options described in [Ch. 6, p. 88] are marked with 1, 2 and 3, respectively.

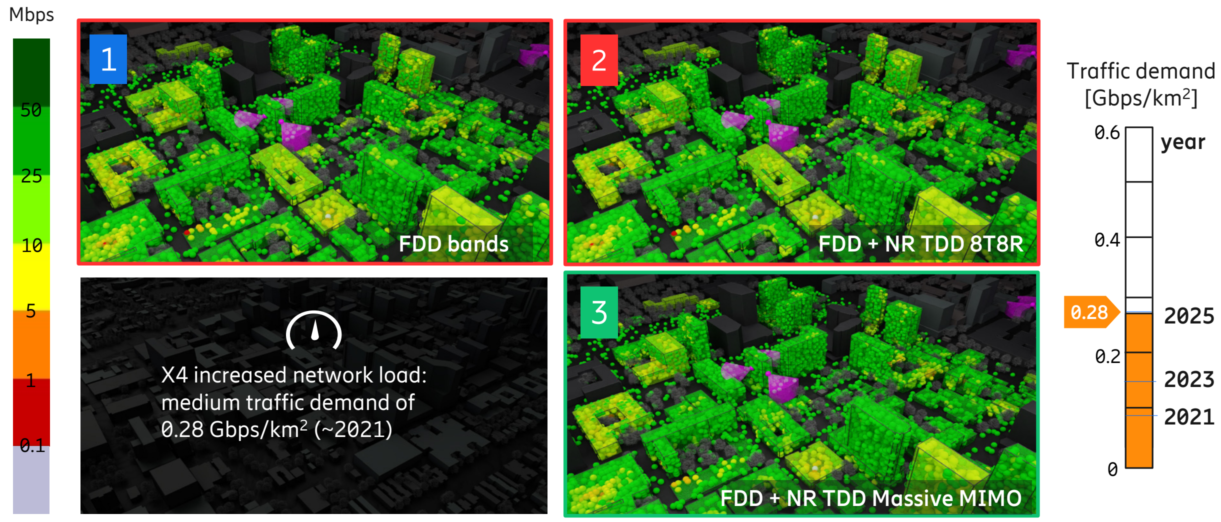
The results show only small differences in user experience for the three different radio solutions. As opposed to the downlink results, the addition of mid-band does not bring any significant performance gains compared to using only FDD bands. There are, however, some users that get a better experience with the addition of mid-band, especially with Massive MIMO.

These results illustrate that low-band deployments are essential to reach superior uplink performance. A main reason being that uplink coverage is challenging on higher frequency bands, and it is to a large extent the users with worst coverage that dictate the overall network performance as those users consume a lot of the radio network resources. Letting a low-band with better coverage serving the worst users, will not only help the coverage-limited users, but also boost the entire network performance.

Further discussions of the benefits with frequency band interworking and carrier aggregation with cross-carrier scheduling can also be found in [Ch. 5, p. 68] and [Ch. 6, p. 103] .

Uplink user experience visualized

x4 load: modest throughput gains by NR TDD



This slide shows results after quadrupling the baseline traffic demand, now reaching 0.28 Gbps/km2, corresponding to the traffic demand in ~2025.

The results are still rather similar for all three radio deployment solutions, although using FDD bands only shows more yellow regions than FDD bands together with mid-band, and FDD plus mid-band with Massive MIMO is slightly greener than the other two solutions.

Uplink user experience visualized

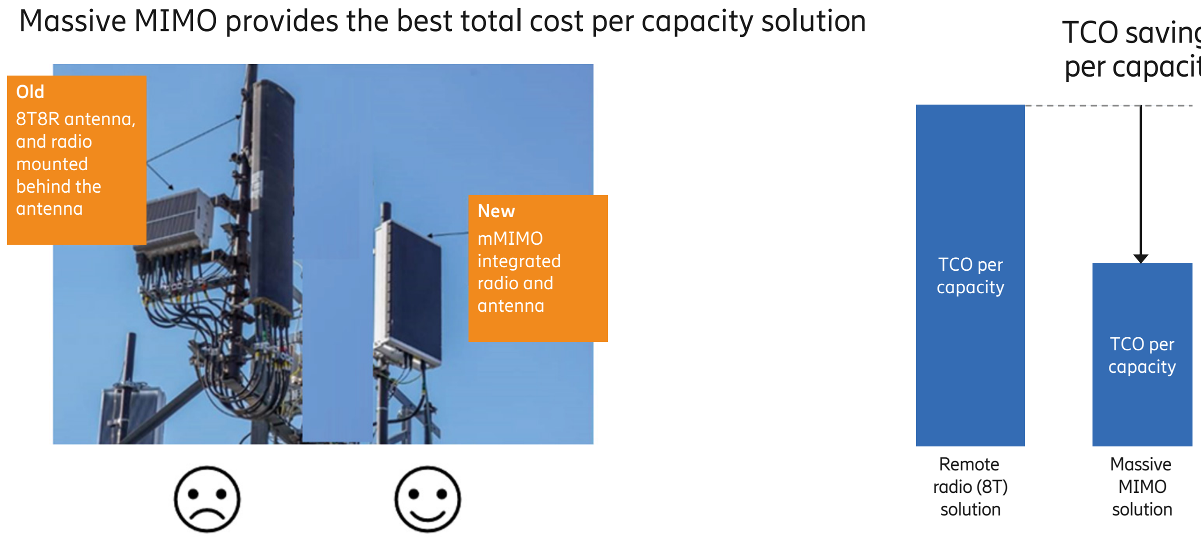
This slide shows results with eight times the baseline traffic, now reaching a high traffic demand of 0.6 Gbps/km2.

This extreme traffic demand cannot be supported by using FDD bands only and many users get very poor user experience. The addition of mid-band to the FDD bands makes the situation much better with most users reaching above, or even well above 5 Mbps. It is also seen that mid-band with Massive MIMO provides a much more consistent user experience in the network, that is, less red or yellow areas compared to mid-band with 8T8R.

The main reason why the addition of mid-band, and especially mid-band with Massive-MIMO, boosts the performance compared to using only FDD bands in this case is that the mid-band can off-load the valuable low-bands. The capacity of the low-bands is not enough to support the traffic demand. The mid-band can serve many of the users with good experience, allowing the low-bands to focus more on serving the users with poor coverage.

6.3. TCO analysis

TCO considerations for mid-band deployments



Total cost of ownership (TCO) calculations depend on many factors that are specific for each service provider and may differ from site to site. The following three slides aim to illustrate some generic key aspects of TCO; see also [Ch 2, p. 27].

Cost can be measured in different ways, but the proposed metric is TCO per Capacity. Using TCO as a cost measure over the investment cycle is reasonable as it takes both capital expenditures (CAPEX) and operational expenses (OPEX) into consideration. Capacity is a measure of the benefits that the product brings and is therefore a good measure of value.

When making TCO comparisons, all important cost drivers should be considered, including radio & antenna equipment, RAN compute capacity, site material, rollout & installation, site rental and energy consumption. Service providers do usually rent the sites, and, for example, the site rental may be different for the case of adding conventional remote radios with passive antennas compared to adding one Massive MIMO product.

The latest Ericsson Massive MIMO product advancements have significantly improved the TCO. In fact, the TCO for Massive MIMO is now comparable with the TCO for a conventional radio and antenna solution.

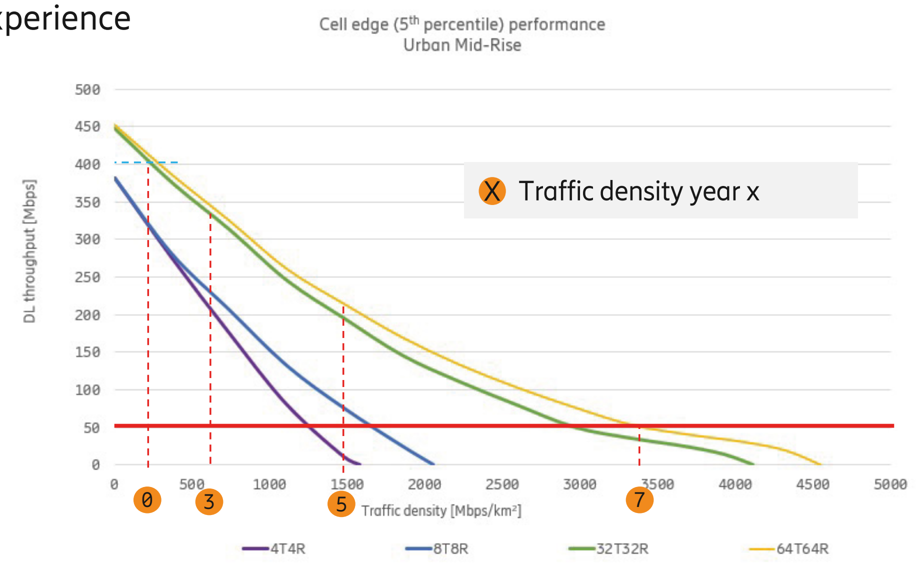
The capacity benefits of Massive MIMO compared to conventional antenna solutions depend to a large extent on the deployment scenario [Ch. 6, p. 84]. However, in many deployment scenarios, the capacity gain of Massive MIMO over a conventional 8T solution is in the order of two to three times.

Combining the TCO analysis and the predicted capacity benefits, results in that Massive MIMO provides a much better total cost per capacity solution in many deployments compared to conventional radio + antenna solutions.

The details of the TCO per capacity process is illustrated on the following slides.

Downlink capacity dimensioning

Massive MIMO facilitates a future proof superior mid-band experience



Traffic growth year-over-year according to Ericsson mobility report ~50%

A key radio network dimensioning criterion is capacity that determines how much traffic the network can sustain with maintained user experience (see [Ch. 6, p. 80]). Capacity is defined as traffic per area unit (bits per second per area unit) given a minimum user throughput requirement for the so-called cell-edge users (typically 5th user percentile). Here a throughput requirement of 50 Mbps is assumed to reach a superior 5G experience.

The figure illustrates the downlink user throughput for cell-edge users as a function of the traffic density for different radio products in an urban mid-rise deployment scenario. The four different radio products are RRU 4T4R, RRU 8T8R, Massive MIMO 32T32R and Massive MIMO 64T64R.

A network capacity dimensioning analysis needs to consider the traffic increase over time. Given today's assumed traffic density represented by the point 0, the different points 3, 5 and 7 in the figure represent the predicted capacity demand after 3, 5, and 7, years, respectively.

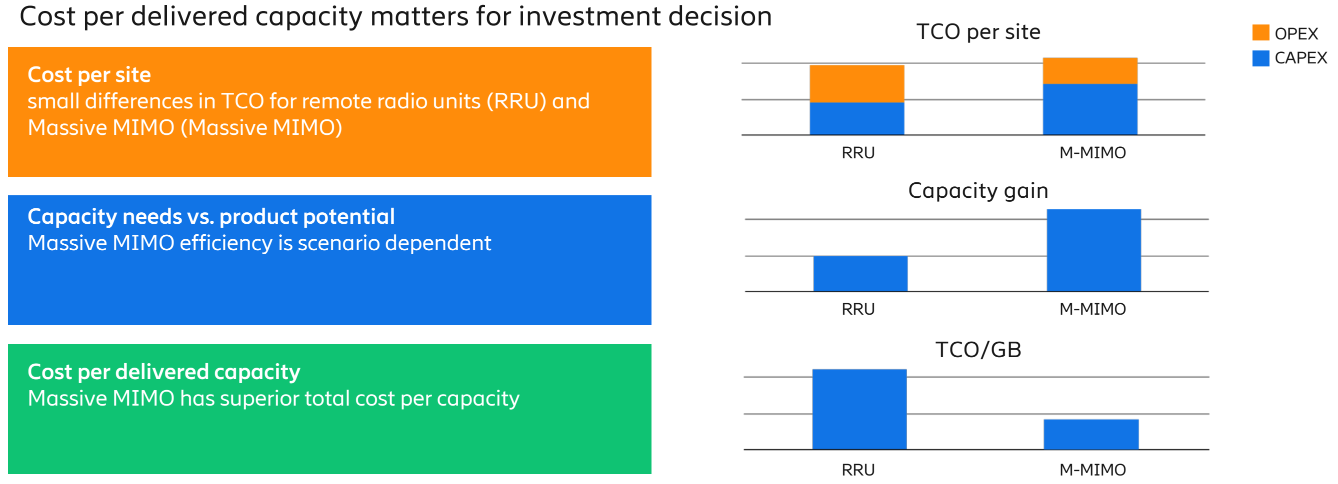
Some reflections from the results shown in the figure are as follows:

- Both Massive MIMO variants deliver more than twice the capacity compared to the RRU solutions
- It is seen that all radio solutions provide good downlink throughputs well above the stipulated downlink throughput requirement of 50 Mbps in today's network (represented by the point 0), as well as after 3 years (point 3)
- After 5 years expansion (point 5), there is, however, a dramatic difference between RRU solutions and Massive MIMO solutions. In particular, the 4T4R solution will no longer fulfill the throughput requirement and the network needs to be upgraded. The 8T8R solution is still meeting the requirement but delivers significant reduced quality compared to the two Massive MIMO solutions
- The Massive MIMO solutions, at least the 64T64R, can sustain the capacity demand for 7 years before a network capacity upgrade is needed.

From this capacity dimensioning example, it is clear that the capability of different radio solutions to handle the expected capacity demand of the future can be very different. The example here indicates that Massive MIMO could last up to 7 years while the RRU radio solutions last 4-5 years.

Structure of TCO analysis

Cost per delivered capacity matters for investment decision



A per-site total cost of ownership (TCO) analysis is an important step in the process of deciding a recommended radio solution. TCO calculations depend on many local and CSP specific factors. This slides illustrates three steps of the TCO analysis that highlights that what matters most for investments decisions is TCO per delivered capacity.

First step is to run a traditional TCO analysis of the cost parts, where both capital expenditures (CAPEX) and operational expenses (OPEX) are evaluated over a typical depreciation time. Although the hardware cost for Massive MIMO is higher than for conventional solutions, this step often demonstrates that the TCO difference between deploying mid-band with a conventional solution and a more advanced Massive MIMO product is rather small. This is mainly because the conventional solution has higher cost related to baseband capacity, network roll out and site rental.

Second step is to determine the offered capacity of the products. As illustrated in [3; Ch. 2, p. 27], the capacity benefits of Massive MIMO over a conventional solution depend on many aspects. Still, in many deployment scenarios, see previous slide, it is reasonable to assume ~2-3 times more capacity from Massive MIMO than a conventional system. The figure illustrates that 64T Massive MIMO provides twice the capacity compared to 8T RRU in an urban mid-rise deployment scenario.

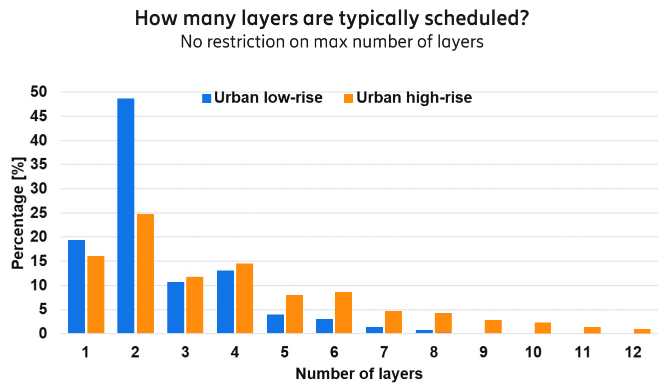
Third step is then to normalize the TCO results from the first step with the predicted capacity gain from the second step. As the TCO for Massive MIMO is only slightly higher than for RRU, but Massive MIMO provides roughly twice the capacity compared to the RRU, the "TCO per capacity" is reduced with more than 40% if Massive MIMO is used.

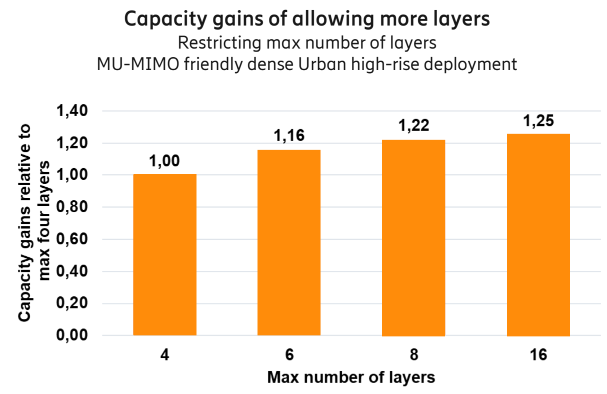
Section [Ch. 6, p. 87] illustrated that adding mid-band with Massive MIMO provides a future proof deployment with superior performance, and here it is further illustrated that Massive MIMO often has superior total cost per capacity compared to other solutions. Hence, in many cases Massive MIMO is a cost-efficient solution for mid-band deployments in many cases.

6.4. Deep dives

MU-MIMO and number of layers from simulations

Opportunistic gains from MU-MIMO – rather few layers are typically needed





Chapter 3 [Ch. 3, p. 27-28, p. 37 - 38] gave some insights into realistic expectations on MU-MIMO, that will be further illustrated here.

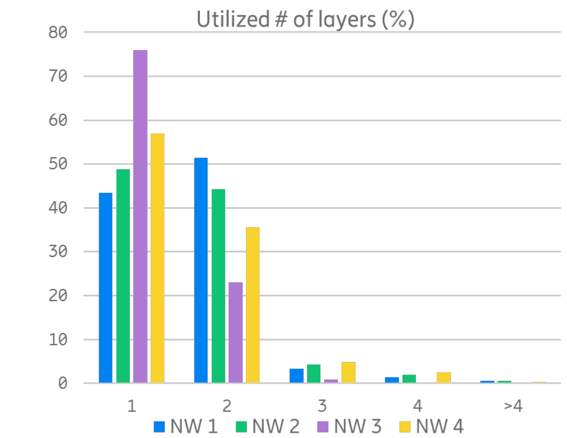
Downlink results for two MIMO favorable deployment scenarios are shown; dense urban high-rise and dense urban low/mid-rise. Both deployments have small inter-site distances and large channel and user spread. The setup is TDD 3.5 GHz with 100 MHz bandwidth, 200 W output power and 64T 8x8 antenna array using generalized beamforming. Furthermore, it is assumed that each UE can have max two SU-MIMO layers. As an example, when four layers are scheduled, it can be one layer to four different UEs, two layers to two different UEs, or one layer to two different UEs and two layers to one UE.

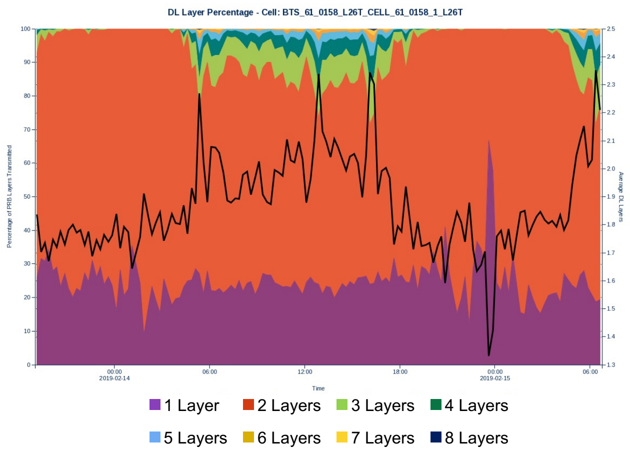
Some observations are as follows:

- The left figure shows the distribution of the number of scheduled layers as seen from the schedulers without any restriction on max number of layers. It is seen that 2 layers clearly dominates, and it is rare to schedule more than 6 layers. Slightly more layers are beneficial in the dense urban high-rise deployment (no indoor systems are assumed, i.e. all indoor users are served by outdoor Macros).
- The right figure shows the capacity gain relative to max four layers for different restrictions on the maximum number of layers in the MU-MIMO friendly high-rise deployment. It is seen that max 16 layers provides a capacity gain of 25%, but most of the gain is achieved already with max 6 or 8 layers.

Number of layers from initial trials of Massive MIMO

Few layers typically used which is in line with our predictions from simulations





This slide shows MU-MIMO layer statistics from real networks and results matches well our simulation predictions (se previous slide) that rather few layers are typically needed.

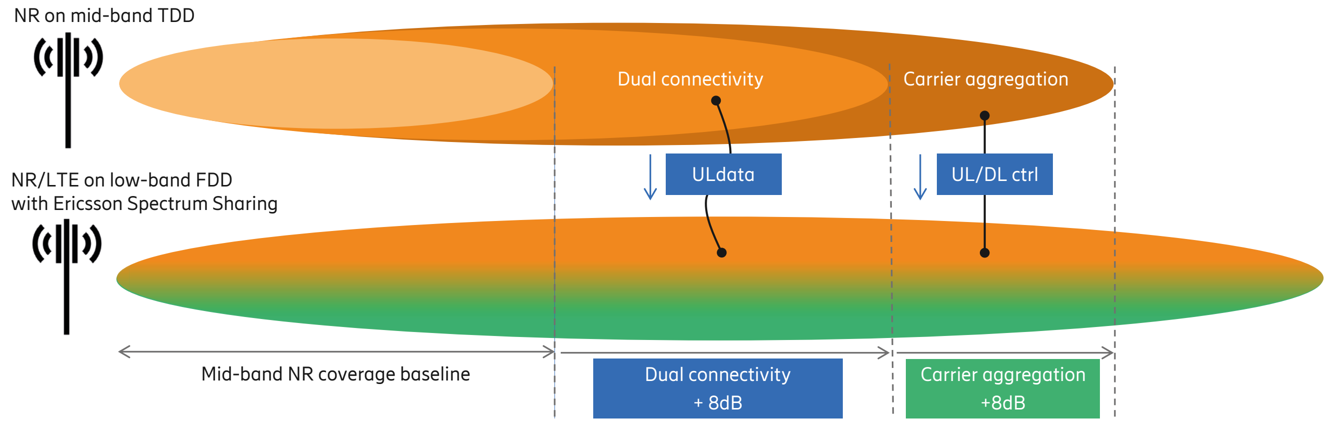
The left figure illustrates the distribution of the number of layers in four different networks, and it is seen that one or two layers dominate.

The right figure shows downlink layer statistics as a function of time in one cell. The left y-axis shows the percentage of number of used layers for all PRBs, where different colors correspond to different number of layers. The right y-axis with the associated solid black curve shows the average number of layers. It is seen that one or two layers clearly dominate, and more layers are only used at peak hours.

Note: These results are for an early LTE deployment and may look different for more advanced NR deployments with better MU-MIMO support.

Frequency band interworking

Maximize mid-band coverage with low-band and carrier aggregation



Note: Gain numbers should be seen as illustrative. Exact gain depends on frequency band combinations and product assumptions.

The capacity dimensioning examples [Ch. 6, p. 97] showed that adding mid-band gives a substantial downlink capacity and end-user throughput improvement, but mid-band together with Massive MIMO is needed to reach superior performance. It was also seen that frequency interworking with a lower band is essential when deploying TDD mid-bands from an uplink perspective. This slides illustrates another benefit of frequency band interworking, namely the offered coverage extension of mid-band.

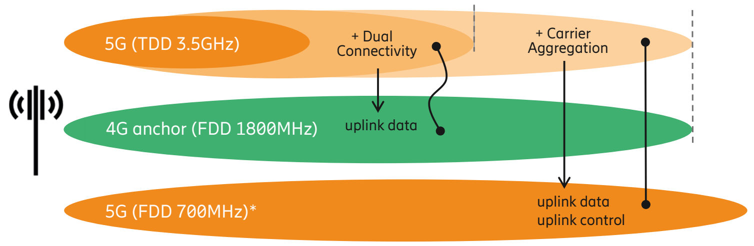
One major benefit of frequency band interworking is the offered coverage extension of mid-bands. By noticing that different physical channels experience different coverage (have different link budgets), weak channels can be moved to a lower band with better coverage, hence being able to use strong channels in mid-band in a larger area. It is important to notice that the link budgets of the different channels depend on many factors, such as, output powers, bandwidths, receiver sensitivities, antenna gains, and required operating points (SINRs), but in many cases downlink channels have better coverage than uplink channels.

The slide illustrates an example with an NR mid-band deployment with a supportive NR/LTE low-band deployment via Ericsson Spectrum Sharing (ESS). In this example, the most limiting physical channel on mid-band is uplink data (PUSCH). Given an App coverage requirement of 1 Mbps in the uplink, the link budget of PUSCH is 8 dB worse than the second worst physical channel which is channel state information feedback via PUSCH. Channel state information feedback via PUSCH has in this example 8 dB worse link budget than downlink synchronization (SSB).

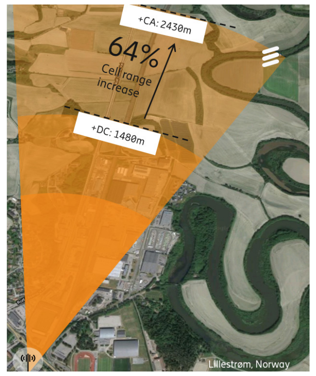
By using dual connectivity or carrier aggregation, the limiting uplink data (PUSCH) can be moved to the low-band with better coverage as soon as PUSCH looses coverage on mid-band. This step increases the overall mid-band link budget by 8 dB. To further increase the mid-band coverage, the cross-carrier scheduling functionality offered via carrier aggregation can be utilized, hence transmitting also L1 control channel information on the low-band. Moving channel state information feedback related to mid-band PDSCH to low-band gives in this example an additional 8 dB improved mid-band link budget. As SSB is typically needed on mid-band, the mid-band cannot be extended by moving/removing SSB. There are, however, other means to enhance the mid-band link budget, for example, employ power boosting for SSB.

Frequency band interworking

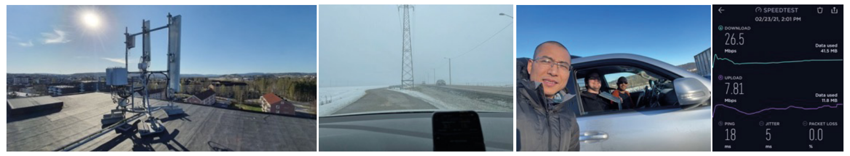
Trial with Telia Norway gives 64% cell range increase



*) 5G on FDD 700MHz enabled with ESS (4G part not shown, since not involved in CA call)



Trial performed in Lillestrøm, Norway by Telia and Ericsson on February 23rd, 2021



Our mobile and FWA customers will experience enhanced capacity, coverage and speed... We will also be able to considerably improve indoor coverage

– Espen Weum, Acting Head of Infra, Telia Norway

This slide illustrates the coverage extension gains offered by frequency band interworking in a real trial (5G FDD-TDD Carrier Aggregation Trial)

In this trial it turns out that carrier aggregation (CA) provides 64% increased cell range, which matches well our link budget predictions. Note, though, that other assumptions may lead to other conclusions, e.g. less gain of CA.

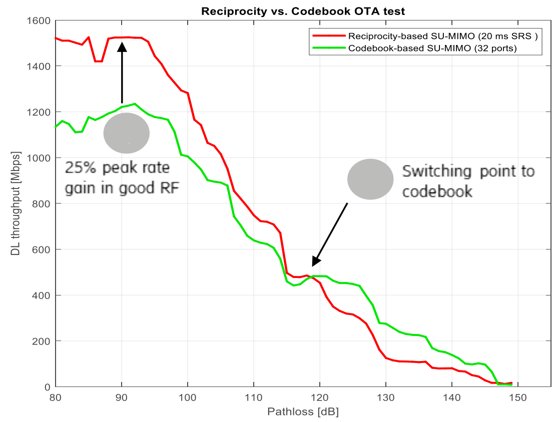
5G (TDD 3.5GHz, n78)

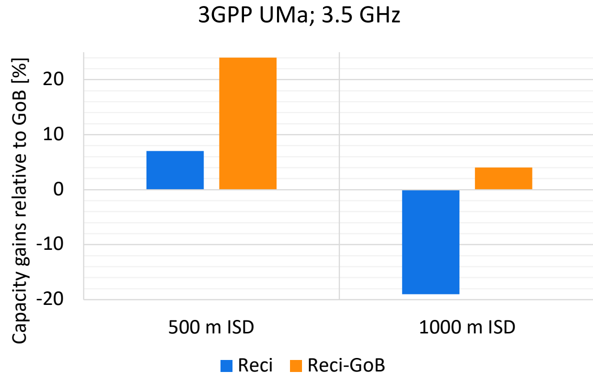
4G/5G with Ericsson Spectrum Sharing (FDD 700MHz, n28)

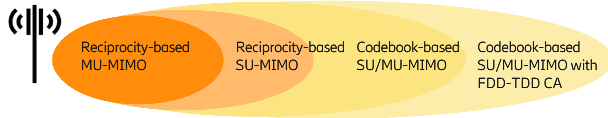
(4G anchor band 1800 MHz, b3)

Feedback-based vs reciprocity-based beamforming

Switch between feedback and reciprocity for superior performance







Fundamentally the performance potential of feedback-based and reciprocity-based beamforming is similar. What matters is to get sufficient channel state information for the purpose of the transmission (for example, wideband SU-MIMO requires much less channel state information compared to frequency-selective MU-MIMO). There are, however, many ifs and buts that impact which beamforming type that works best in practice, see [Ch. 3, p. 34 - 36] and [Ch. 5, p. 62] for further details.

This slide illustrates that with current standard and algorithm product choices, reciprocity-based beamforming excels in deployments with sufficient good coverage for uplink sounding, otherwise feedback-based beamforming performs better. This is the reason why Ericsson promote a solution where the beamforming algorithm is adapted to the situation. The beamforming algorithm will essentially be switching between reciprocity-based MU-MIMO, reciprocity-based SU-MIMO, feedback-based SU/MU-MIMO, and finally feedback-based SU/MU-MIMO with feedback on a lower frequency band, depending on the coverage situation (illustrated by the lower right figure).

Left upper figure illustrates OTA results comparing downlink throughput of reciprocity-based and feedback-based beamforming as a function of path loss and shows that there is a switching point where feedback-based performs better than reciprocity-based beamforming.

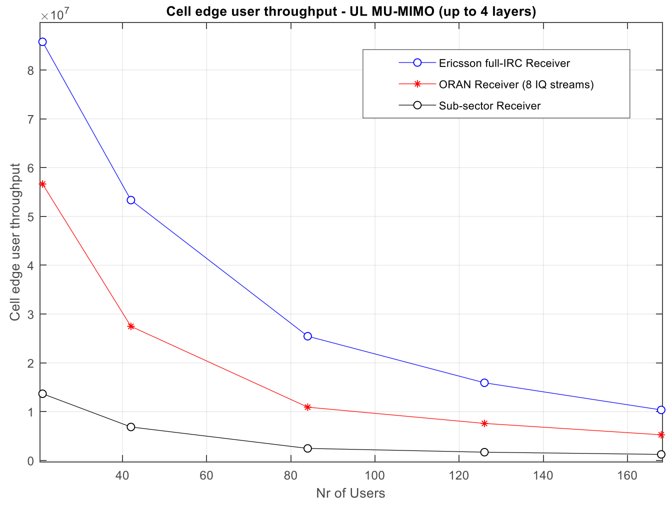
Right upper figure shows simulation results comparing gains of different beamforming algorithms relative to feedback-based GoB in 3GPP Urban Macro with 500 m inter-site distance and 1000 m inter-site distance, respectively. It is seen that reciprocity-based beamforming (Reci) performs better than feedback-based beamforming (GoB) for 500 m inter-site distance, whereas for 1000 m inter-site distance, Reci shows a loss. Best performance in all deployments and for all key performance indicators is given by the so-called Reci-GoB algorithm, which switches between reciprocity-based and feedback-based beamforming based on the uplink sounding signal quality. Hence, UEs that have a sounding SNR better than a threshold are using reciprocity-based beamforming, otherwise feedback-based beamforming.

Ericsson superior Massive MIMO architecture

Provides significant uplink performance benefits

Full-IRC receiver is superior

- >4x cell capacity vs. Sub-sector receiver
- 1.6x capacity vs. ORAN receiver
- Poor cell edge performance with sub-sector



The Ericsson Massive MIMO architecture and Open RAN (ORAN) are discussed in [Ch. 8, p. 124-127]. The figure illustrates the good uplink performance of the Ericsson Massive MIMO architecture by considering cell-edge user throughput as a function of network load (number of users) in a dense urban deployment using a 64TR Massive MIMO. The sub-sector receiver uses four 'beams' (eight I/Q streams) for receiver processing, the ORAN receiver uses eight I/Q streams for receiver processing and Ericsson's full IRC receiver uses all 64 receiver signals for receiver processing. It is seen that the full-IRC scheme outperforms the other architecture choices in this setup.

It should be noted that the ORAN performance can be improved by 'better' standardization or more investments compared to Ericsson's full-IRC. For example, doubling the front-haul capacity of the ORAN deployment compared to the E/// deployment would render similar performance between ORAN and Ericsson's full-IRC, but at a higher cost.

Summary of radio network performance

- Good match between our simulation results and real network performance
- Deployment environment, array structure and algorithm choices are key components affecting the performance of Massive MIMO
- Deploying mid-band with Massive MIMO enables superior performance in a cost-effcient way

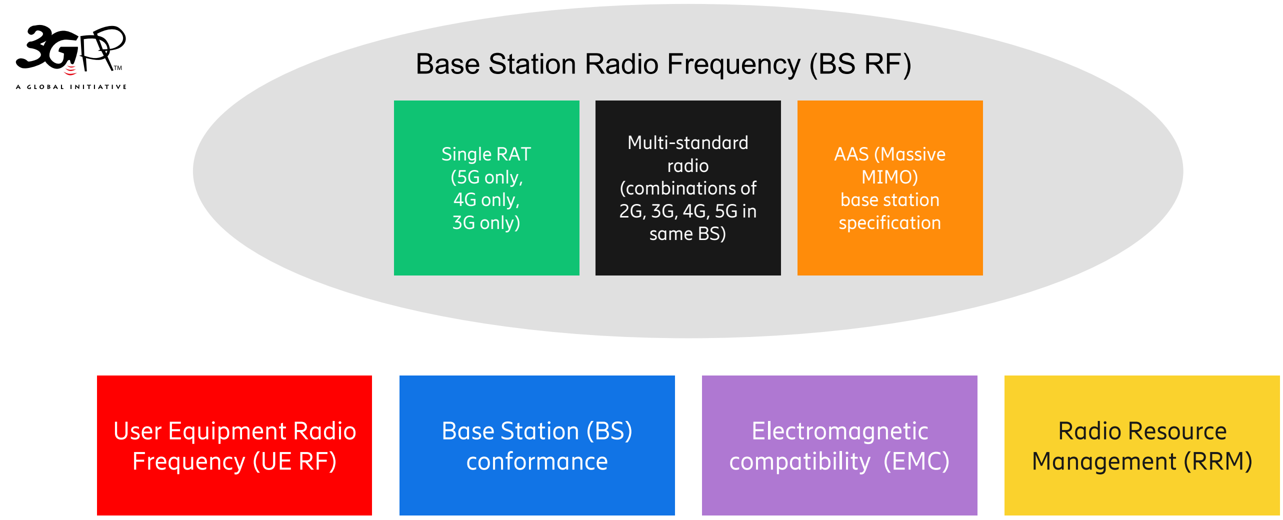
Performance evaluations are vital in the R&D flow. It provides an understanding of how different deployments, configurations and algorithms perform before making costly R&D product investments. It can also guide our customers when it comes to their network evolution and product needs. Radio network performance depends on many aspects that interact in an intricate manner, and it is important to understand the driving mechanisms and having models that capture the reality sufficiently well while still being computationally efficient. Ericsson is in the forefront in the area of radio network performance evaluations using state-of-the-art knowledge and models. The predictions based on simulations have proven to match real network performance on countless occasions.

Several illustrations of that massive MIMO deployments facilitates superior radio network performance have been given. In particular, key aspects impacting Massive MIMO performance, such as antenna configuration, deployment scenario and beamforming algorithm choice, have been discussed and illustrated. A detailed illustration of capacity dimensioning benefits of Massive MIMO over other radio solutions has been provided. An illustration of key aspects for TCO analysis has been given showing that Massive MIMO is the best solution in many deployments by using TCO per capacity as main metric.

Deep-dives into a few selected areas that are subject to frequency asked questions were given, including effects of the number of MU-MIMO layers, benefits of frequency band interworking, impact of type of channel state acquisition approach (feedback-based vs reciprocity-based), and effects of the Ericsson Massive MIMO architecture.

7. 3GPP – Radio requirements

Why radio requirements



In RAN4 and radio regulations, the term AAS is used to refer to Massive MIMO base stations.

Spectrum is an expensive and scarce resource and so efficient compatibility and co-existence in a multi-operator environment is essential. This is achieved by defining radio requirements in 3GPP RAN4 and in regulatory specifications. The requirements consider co-existence between operators in mobile bands and in the case of regulation, between different types of system. In addition to enabling efficient spectrum usage, the RAN4 specification also ensures consistent and predictable behaviour from UEs, which is essential for building high performance and reliable networks. Separate specifications are defined for base stations (BS) and User Equipment (UE)s. In addition to spectrum requirements, the BS specifications contain other requirements such as transmitter signal quality or receiver behaviour to provide for good minimum network performance and coverage.

The UE RF specification covers similar requirements as BS but is more extensive to ensure interoperability and minimum performance with different UE vendors. The UE specification contains RF requirements for an extensive number of Carrier Aggregation (CA) and Dual Connectivity (DC) combinations because, differently to the BS, the hardware requirements for supporting different combinations differs in the UE. The specification needs to contain suitable combinations for all operators needs.

Another important 3GPP RAN4 specification is the Radio Resource Management (RRM specification) where UE behaviour and performance in relation to measurements and procedures that support network operation are captured. The RRM specification has a significant impact on, amongst other things network timing and synchronization.

EMC (ElectroMagnetic Compatibility) is yet another RAN4 specification covering both BS and UE and describing the requirements for radiated emission and immunity aspects. The EMC specifications embedds many of existing regulatory specifications.

In addition to requirements, 3GPP also specifies the configurations, test set-ups and accuracy to be used for assessing compliance to the requirements. The BS test specification is a part of 3GPP RAN4 responsibility.

In summary, 3GPP radio requirements provide a means to secure co-existence and compatibility, cover some regulatory requirements and ensure good performance and predictability in respect to UE behaviour in the network. In addition, the 3GPP RAN4 specifications provide a bench-marking tool for operators to enable fair comparasion of performance between different vendors under similar conditions.

Electro magnetic field (EMF) exposure

Strict limits on EMF must be considered for the whole site

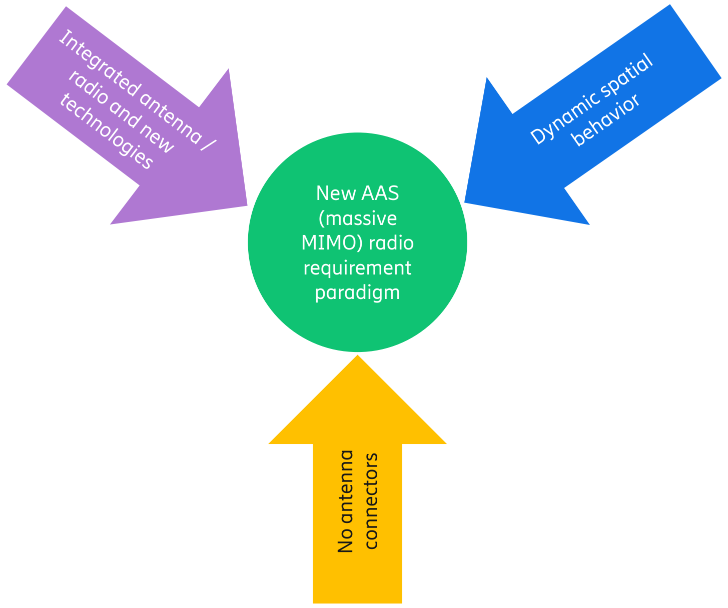
All radio equipment including Massive MIMO (AAS) base stations need to comply to regulation on EMF (Electro-Magnetic Field) exposure. EMF requirements are specified by international bodies such as ICNIRP or IEEE or national administrations such as FCC (EMF is not, however in the scope of 3GPP). The EMF exposure limit is frequency dependent and is expressed as incidence quivalent planewave power density (W/m2) averaged over time (6 minutes for frequencies up to 10 GHz).

The recommended EMF exposure limits provide a good margin of protection against all known health effects. Some countries require even stricter levels than the recommended ones.

Active Antenna Systems (AAS) / Massive MIMO BS have the capability to do flexible beamforming both in azimuth and elevation, which is specifically intended to obtain higher directional radiated power. The assessment of EMF compliance for Massive MIMO (AAS) is more complex than for fixed antenna systems. The emissions depend on type of traffic and also radiation pattern depending on chosen multi-antenna transission schemes for any particular user.

With increased market demand for increased radiated power from Massive MIMO (AAS) BS, incorporating advanced algorithms to restrict the radiated power and ensure compliance to EMF exposure limits becomes a necessity.

Challenge for AAS conformance requirements



Compared to traditional base stations, for Massive MIMO (AAS) there are new and additional challenges in definining suitable requirements, designing and building to meet the requirements, and in measuring compliance.

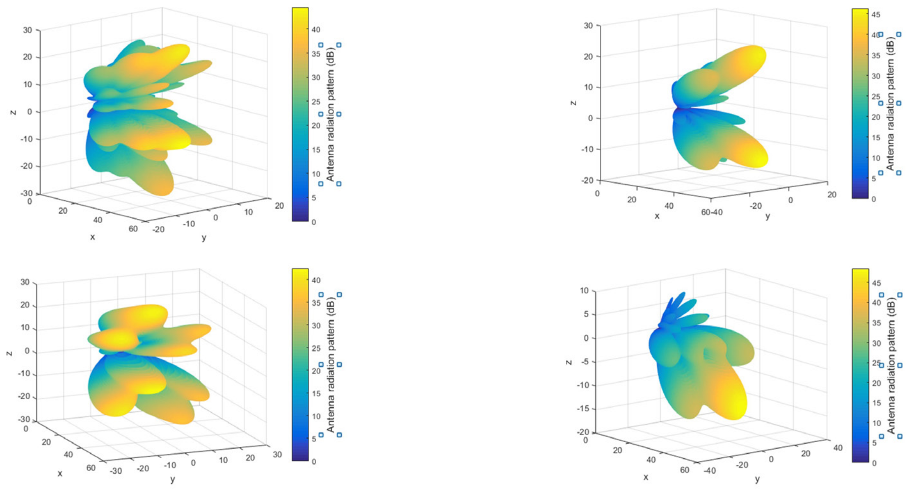
Three key aspects of AAS impact the definition of requirements. The first is that the antenna elements are integrated into the base station and antenna performance is part of the overall radio performance. This differs from traditional base stations, for which the antenna was a separate component independent from the radio operation such that requirements could be set on the radio only. The second is that due to dynamic beamforming, the spatial behaviour of the wanted signal, unwanted emissions and receiver is complex and time varying. For non-AAS base stations, spatial behaviour is both static and quite similar between different base stations. The third is that direct connectors to the radio output may not be available for AAS. This is because efficient antenna array designs are not possible whilst providing connectors to each individual radio. The lack of connectors makes traditional testing methods, in which test equipment is connected directly to the radio output, infeasible.

When it comes to meeting requirements, AAS presents new challenges. AAS arrays consist of a large number of individual low power transmitters, and the performance and efficiencies achieved with a single high power transmitter do not easily scale. In addition, the very limited space in a tightly integrated array places extreme constraints on the possibilities for e.g. filtering, which is key to meeting unwanted emissions requirements. These aspects and a number of other technolgy considerations are explored on the next page.

Achieving all radio requirements, including unwanted emissions and receiver selectivity, drives the complexity and cost of AAS solutions. In some cases, beamforming physics are able to compensate radio RF performance for the wanted signal itself. However, beamforming cannot comensate radio performance for unwanted emisisons and receiver effects in the same way.

To capture the whole effect of the antenna array including spatial properties and to enable assessment of the radio performance for AAS base stations that do not have connectors, over-the-air (OTA) testing is needed. Prior to 5G, OTA testing of base stations was an unknown paradigm and a large amount of effort has been expended in 3GPP to define requirements in such a manner that OTA testing is possible and to investigate the practical possibilities for OTA testing. The NR specification contains OTA requirements for all parameters and test tolerances based on realistic expectations of what can be achieved in test environments.

3GPP Over-the-air specification scope



Illustrative examples of beam patterns of advanced Massive MIMO systems

The purpose of 3GPP radio specifications is to set minimum requirements to enable

- Efficient sharing of spectrum between operators and services.
- Predictable UE behaviours .
- A minimum signal fidelity.

Requirements are generally absolute and must be met by all implementations.

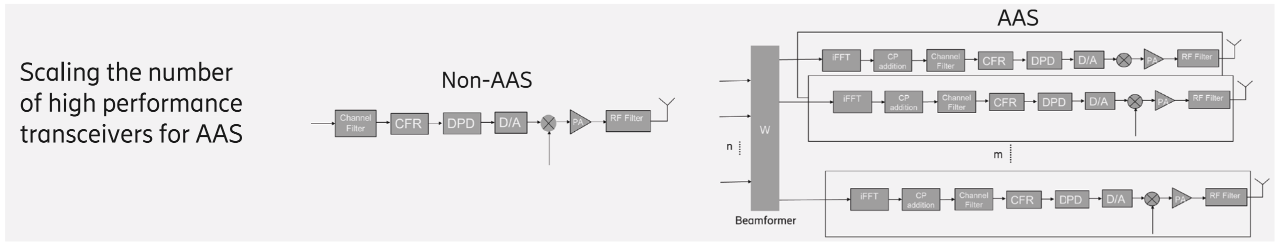
What the radio specifications do not do is set requirements relating to the "quality" of the beamforming achieved.

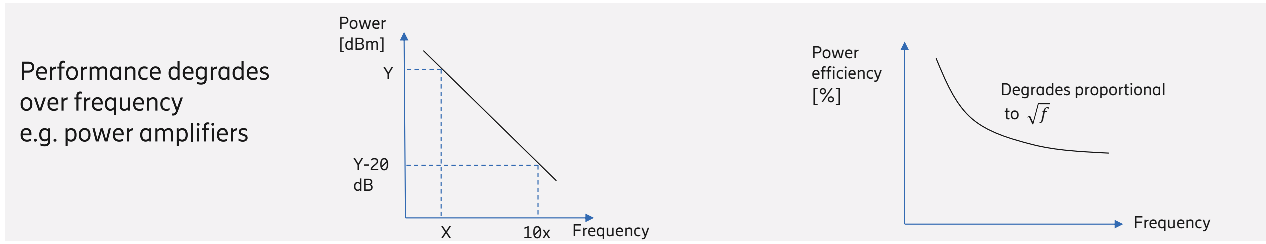
Traditional fixed passive antenna performance was assessed by measuring well-known parameters such as sidelobe levels, beamwidth, ripple of gain within the beam, front to back ratio and others. For an AAS performing dynamic beamforming, the traditional parameters have less meaning as they vary depending on the instantaneous beamforming. Furthermore, for advanced MIMO systems the concept of a "beam" is not applicable, see [Ch. 3, p.24]. Energy is radiated in a complex pattern according to the channel state, as shown in the figures. In the figures, the multiple lobes of energy are deliberately created due to multi-path propagation and it is not possible to differentiate "useful" and "not useful" radiated energy.

Beamforming and MIMO performance depends on baseband, radio and the channel statistics. Current OTA test chamber technology does not enable re-creating of fading channel statistics in a controlled environment. For these reasons, determination and measurement of appropriate metrics for determining the "quality" of beamforming and MIMO performance is non-trivial and needs to be the subject of research.

Even if such metrics would exist, it is not likely that it would be suitable to set minimum requirements. To do so would indirectly standardize the type of MIMO operation and risk to restrict innovation. A potential future aim could be to determine a toolbox of industry wide standardized metric descriptions whose implications are understood, but which are not subject to minimum levels. From these a suitable sub-set could be selected and assessed depending on beamforming implementation.

Why is AAS challenging





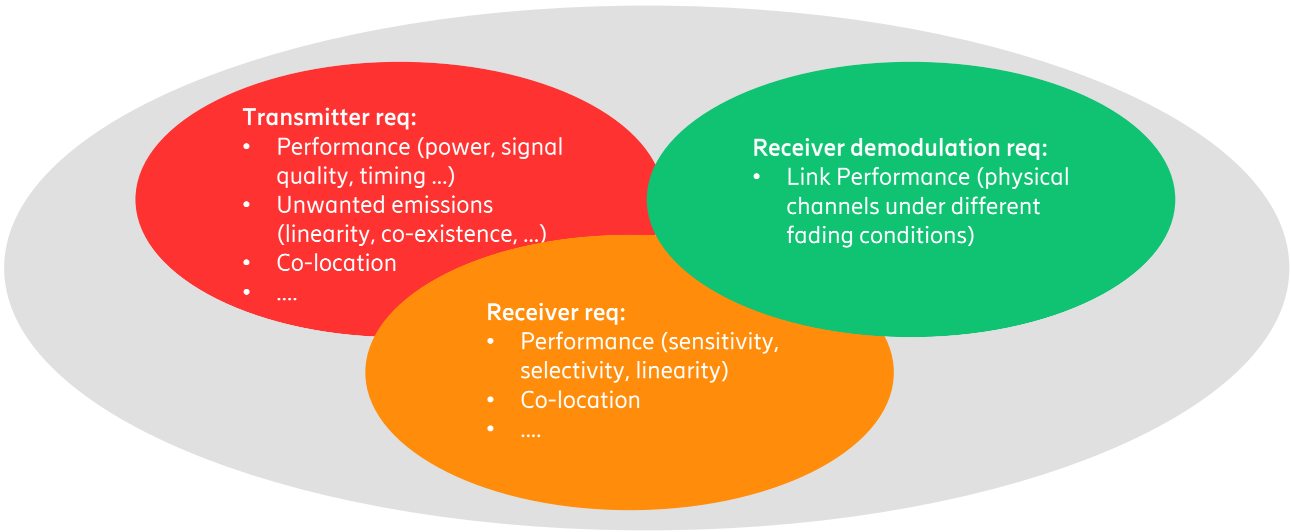
A fundamental aspect that makes Massive MIMO (AAS) BS implementation different compared to non-AAS is the scaling of the number of transceivers. In general, each transceiver in AAS needs to maintain similar RF performance as a non-AAS BS transceiver for the same frequency, but clearly the complexity and cost per transceiver needs to be reasonable.

In general, the RF performance degrades over frequency. Often, AAS BS are intended to combat the propagation conditions at higher frequencies, which leads to the need for acommodating large number of transceivers into a small space.

In addition to these general aspects, at high frequencies a number of additional technology challenges arise, for example:

- The power amplifier output power decays by around 20 dB over a decade of frequency, while the power efficiency degrades proportial to $\sqrt{f}$. For AAS BS when at higher frequency, in addition to reduced size, the degraded output power and efficiency make the thermal design quite complex, which results in a delicate balance between linearity, available power and power efficiency.
- The phase noise due frequency generation in AAS also degrades over frequency with 10*Log (f_1/f_2) affecting the signal quality as well as some fundamental receiver aspects.
- The receiver Noise Figure, which directly relates to uplink coverage also degrades over frequency. As an example, the noise figure for bands ~3 GHz is typically ~5 dB better compared to bands around 30 GHz. In addition, there is a complex dependency between Noise Figure, linearity of receiver as well as dynamic range of the receiver.
- At higher frequencies, the available bandwidths become larger. Within the radio, digital algorithms such as PA Linearization become more sophisticated at larger bandwidths. Also, larger volumes of data must be processed, which puts more stringent requirements on interfaces. Data handling is exacerbated due to the large number of transceivers for AAS BS.

What are the radio requirements



BS radio requirements in general consist of 6 groups. The first of the groups is Transmit power related requirements. The power related requirements cover the accuracy of the BS power, the transmitter dynamic range requirements etc.

The second group, Unwanted emission requirements comprises:

- In-band requirements on absolute operating band unwanted emission masks and ACLR (Adjacent Channel Leakage Ratio), which is a power ratio, regulating emissions towards other operators in the same 3GPP band.
- Out-of-band requirements comprising general spurious emission requirements (in general set by regulators in different regions), which regulate co-existence to other types of system and and co-existence and co-location requirements.

The third group is transmit signal quality requirements, which consists of Error Vector Magnitude (EVM), frequency error and Time Alignment Error (TAE). EVM is a measure of all transmitter inpairments that degrade the signal quality. The frequency error captures the frequency variation in relation to assigned frequency and TAE captures the time difference betwen different MIMO layers or carriers when using carrier aggregation (CA).

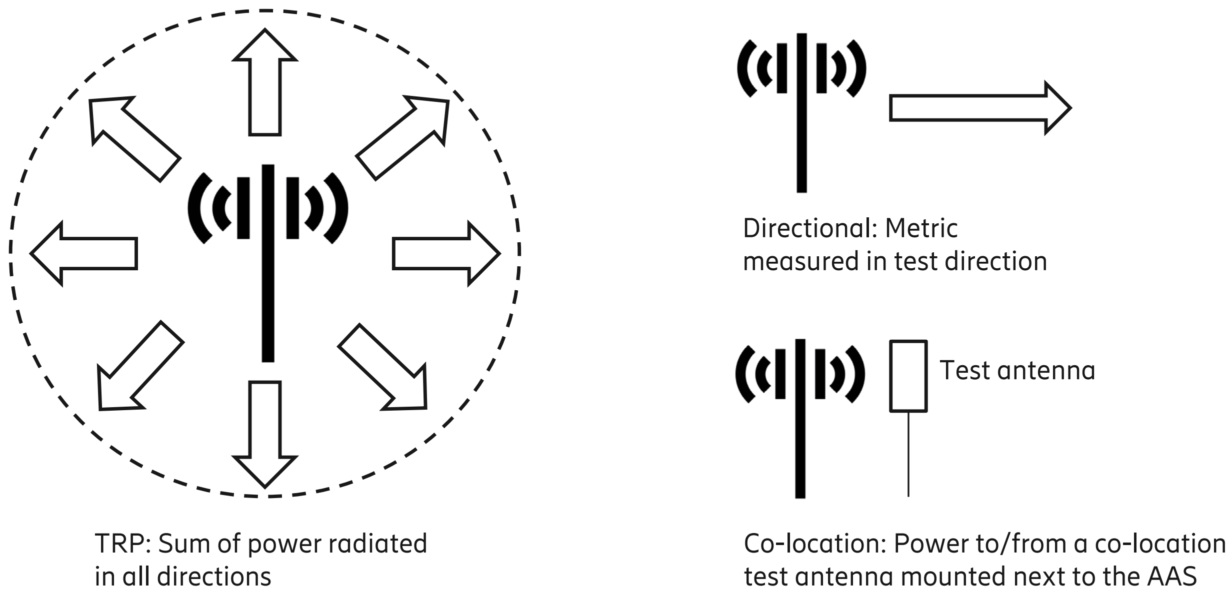
On the receiver side, the next group is receiver sensitivity.

Receiver sensitivity is a defined as a minimum level of received signal that can be demodulated properly. RX sensitivty relates to the receiver noise figure, which is directly related to uplink coverage.

Apart from the receiver sensitivity, requirements are defined on Receiver selectivity and blocking. The receiver blocking is a measure of receiver´s capability to receive a weak wanted signal in the presense of a strong interferer in the adjacent channels. The receiver blocking thus defines the needed receiver in-band selectivity. Receiver blocking can arise in the real world when another transmitter, such as another operator's UE is close to the base station. In addition there are co-location blocking requirements specified considering very high interferer level in other bands.

A final set of requirements are on receiver demodulation performance. Receiver demodulation performance requirements capture the performance of receiver baseband algorithms for different physical channels.

Classes of OTA requirement



To define OTA requirements with the same scope as the conducted requirements, it is useful firstly to differentiate

- requirements based on the wanted signal (such as TX power and dynamics, TX signal quality, RX sensitivity) and
- requirements relating to other frequencies.

For the wanted signal, there is always a single direction of interest, which is towards the intended UE.

For other frequencies, requirements are further differentiated into those related to impacts on other radio receivers that are not at the same site and those related to avoiding interference towards co-located equipment.

Most unwanted emissions requirements are specified base on the so-called "Total Radiated Power" metric, which is the average power radiated in all directions. The reasons for selecting TRP are that:

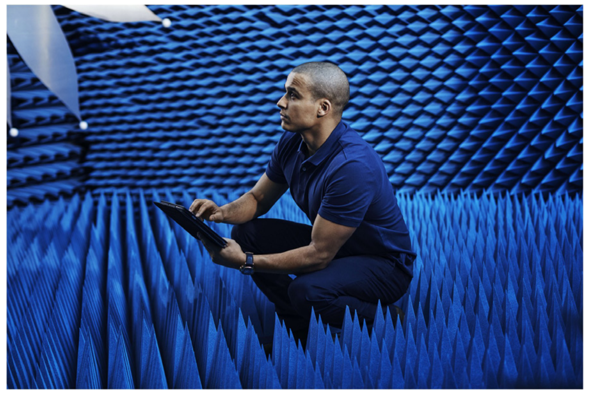
- the location of victims of emissions are not correlated to the beamforming pattern
- the unwanted emissions are often not beamformed
- the impact of unwanted emissions to other systems is often statistical in nature (such that average interference relates to average impact towards other systems).

For emissions and RX requirements related to co-located equipment, requirements are defined based on a co-location concept. A test antenna is placed 10 cm from the BS under test such that it experiences or causes interference in a similar manner to co-located equipment.

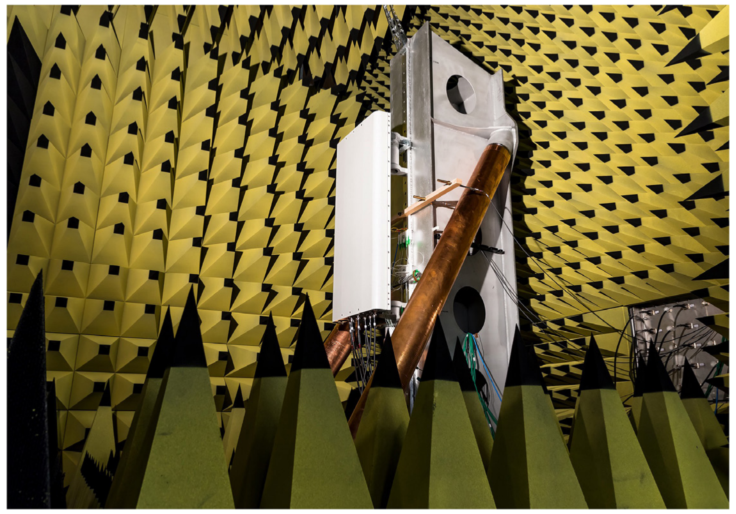
Some receiver requirements include both a wanted signals and interferers, such as blocker interferers on other carriers. These requirements are also defined as directional with the interferer directions the same as the wanted signal directions. Although aligning the wanted signal and interferer directions does not exactly corresponding to real-life situations, the interferer level is determined on a statistical basis such that the impact to the radio electronics is similar to real life. Aligning the signals enables a significant reduction in test complexity compared to a set-up with signals coming from different directions.

The scope for spatial testing of baseband performance requirements is limited; a test setup is defined in 3GPP but is capable of testing at most polarization diversity.

OTA testing



Examples of OTA test facilities, depicting absorbant material and an example of mounting AAS (Massive MIMO) base stations for testing



Measurement of OTA requirements with a reasonable degree of accuracy is complex and specialized. A basic assumption is that measurements are carried out in an RF shielded environment such that testing does not interfere with other nearby systems.

Conceptually, testing of directional requirements is not difficult; the BS is illuminated from a particular direction or receives from a particular direction. Design of testing is, however complex:

- The test facility needs to be designed such that a so-called "quiet zone", in which only planar waves occur exists around the BS under test.
- Careful calibration is needed to manage measurement uncertainties.
- The mechanical design of the test setup must also be sufficiently exact and carefully designed such that the AAS can be mechanically rotated, in order that the measurements can be made from multiple directions, without leading to degradation of accuracy. It should be noted that as long as the BS can be rotated, the IAC, CATR, PWC and NFTC type chambers are all suitable.

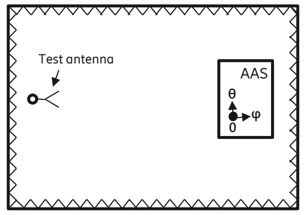
TRP requirements can be tested by means of making directional measurements in a large number of directions around the AAS such that an average interference power can be calculated. Design of the measurement grid needs to be considered carefully and depends on the geometry of the BS and the expected beam patterns. If the test grid is not properly designed then the measurement result may contain systematic and/or random error factors.

An alternative to testing TRP by measuring in multiple directions is the use of a reverberation chamber. A reverberation chamber moves metal sheets within the chamber to randomize a reflection pattern and enable average TRP to be measured. In this case, care is needed to take care of calibration and of Passive Intermodulation effects.

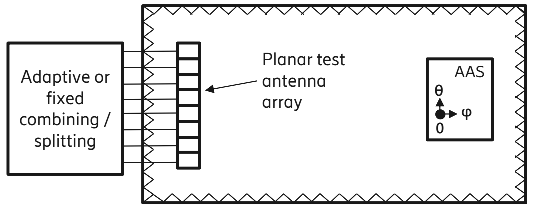
Co-location testing requires the BS to be mounted together with the co-location test antenna. For some types of co-location requirement (in particular, transmitter intermodulation) it is necessary to rotate the combination of BS and test antenna such that measurements of emissions can be made in multiple directions. For most types of co-location measurement, the measurement is made in the co-location test antenna or in the BS, and the only requirement for the test chamber is that it is shielded and anechoic.

Types of OTA testing chamber (I)

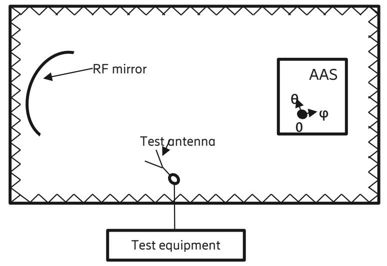
Example schematic diagrams



Indoor Anechoic Chamber (IAC)



Planar Wave Chamber (PWC)



Compact Antenna Test Range (CATR)

There are several kinds of OTA test facility that are relevant for testing of BS. Not all of the facilities are suitable for every type of requirement. For testing the full set of requirements, it is likely that more than one type of test facility is needed.

Indoor Anechoic Chamber (IAC): The indoor anechoic chamber is a shielded chamber with absorbtion such that a quiet zone is achieved around the BS. The BS is placed in the far field of the test antenna. IACs operate within a range of frequencies they are defined for. A disadvantage with an IAC is that it may be large in size.

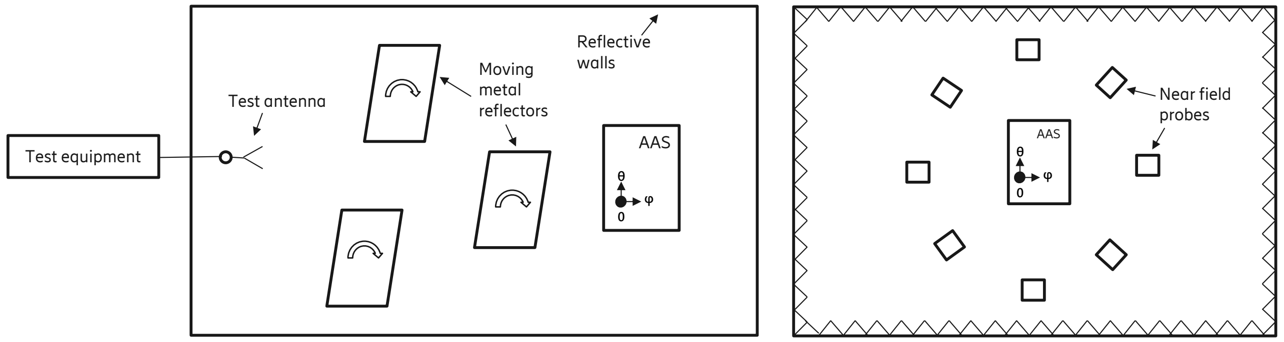
Compact Antenna Test Range (CATR): A CATR is a shielded, anechoic chamber with a specially designed reflector (RF mirror) between the AAS and the test antenna. A planar wave and quiet zone is achieved at the BS under test, and the dimensions of the chamber can be reduced significantly.

Planar Wave Chamber (PWC): In a PWC, a steerable array is used to transmit to / measure from the BS. The beamforming weights at the array are set such that the BS experiences a planar wave as if from a source in the far field. The PWC is similar in concept to a CATR, except that the RF mirror is replaced by the electronic array, and is much smaller in size than an IAC.

IAC, CATR and PWC are generally suitable for all types of in-band measurements and some out of band measurements that are close to the band in frequency.

Types of OTA testing chamber (II)

Example schematic diagrams



Reverberation Chamber (RC) Near Field Test Chamber (NFTC)

Near Field Test Chamber (NFTC): A near field test facility is a type of chamber in which test probes are placed within the near field of the AAS under test. The size is considerably reduced. NFTC is suitable for assessment of in-band TRP and some of the TX directional requirements, but not RX requirements.

Reverberation Chamber (RC): A reverberation chamber is shielded, but deliberately not anechoic. The inside walls are matallic, and metallic moving "stirrers" cause the profile of reflections to continuously shift. The shifting reflection patterns mean that it is possible to measure Total Radiated Power (TRP). RC are suitable for measuring TRP based emissions requirements for in band and parts of out of band, but not the entire out of band range.

Electromagnetic Compatibility (EMC) chambers: EMC chambers are shielded and somewhat anechoic and cover a large frequency range. The uncertainties associated with EMC chamber measurements are larger than for other chambers. EMC chambers are useful for measuring out of band emissions because they cover a wide frequency range and often the increased uncertainty does not matter because the emissions far from the carrier frequency passes the requirement with a large margin for most BS.

Summary

- 3GPP and regulatory radio requirements enable efficient usage of the spectrum with predictable performance and behavior in networks
- Accommodating Massive MIMO/AAS BS necessitated totally new approaches to define and measure the radio and baseband requirements over the air
- Designing a Massive MIMO/AAS BS to meet the OTA requirements is complex and challenging
- It is critical to be able to validate that a Massive MIMO/AAS BS meets all relevant requirements
- Proper OTA testing is specialized and costly. However, if OTA testing is not done properly, the Massive MIMO/AAS BS behavior and performance may be poor and endanger the eco-system

8. Architecture and implementation

Introduction

The commercialization of Massive MIMO radio units has led to some key architecture choices in order to realize cost-efficient hardware implementations.

Key concepts :

Fronthaul evolution

- eCPRI
- The O-RAN Alliance
- Lower Layer Split

Beamforming

- Analog vs. digital beamforming
- Frequency domain vs. time domain beamforming

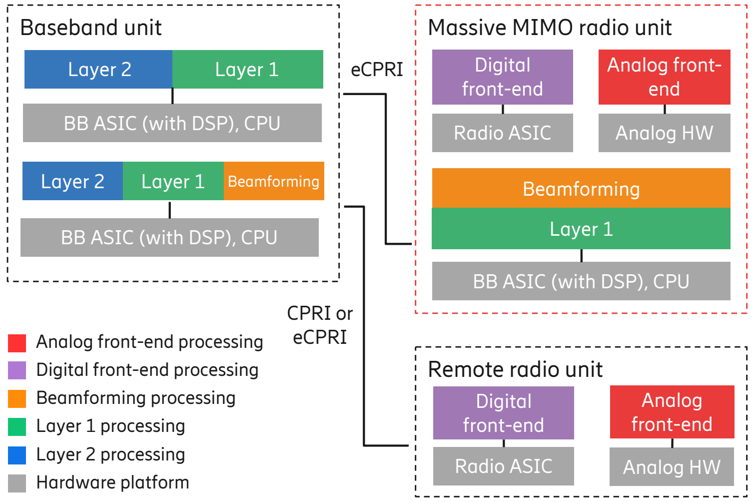
The commercialization of Massive MIMO radio units has led to some key architecture choices in order to realize cost-efficient hardware implementations.

To support a large number of radio chains, the traditional CPRI interface has evolved in the eCPRI interface, which support the so-called lower layer split, where the some of the traditional baseband functions have been moved to the radio unit. Depending on which functions that are move, different lower layer splits can be defined.

Beamforming, which is the function that maps information signals on multiple antennas, can be implemented in different ways. One choice is if it should be implemented with analog or digital components. For the choice of digital beamforming, there is also a choice between frequency domain or time domain beamforming.

Hardware architecture

- Massive MIMO radios have more steerable radio chains than conventional radio units. To handle this in a cost efficient way, there are hardware impacts/choices on the system architecture.
- The location of the beamforming function.
- The choice of digital beamforming or analog beamforming.



Massive MIMO radio units typically has many more radio chains than conventional radio units, e.g., 64 radio chains instead of 4 radio chains. The traditional CPRI fronthaul interface protocol transfer data per radio chain. For Massive MIMO radio units this would mean very large bitrates over the fronthaul interface. To handle this in a cost-efficient way, there are hardware impacts/choices on the system architecture.

In order to lower these bitrates, the beamforming function has been moved from the RAN Compute (e.g baseband) to the radio unit. This makes the fronthaul bitrate scale with the user information bitrate, rather than with the number of radio chains. And since the user information bitrate typically is much lower than the combined bitrates for all radio chains, this makes a large HW saving.

Beamforming can be either implemented using digital components or analog components (or a combination). Typically, digital beamforming provides higher flexibility and higher performance. However, for cases with very many radio chains and/or very large frequency bandwidth, the complexity and cost can become large, and in those cases analog beamforming is an alternative.

Fronthaul interface-standardization

CPRI
- Proven interface for classic radio units with few antennas.
- Bit rates scales with the number of antennas.

eCPRI
- Ethernet packet-based.
- Enables new functional splits between baseband and radio units that can reduce the bit rate over the fronthaul for Massive MIMO radio units.

Standardization of new functional splits
- Studied in 3GPP.
- Standardized by the O-RAN Alliance.

The Common Public Radio Interface (CPRI) and eCPRI are standardized by the CPRI Forum.

CPRI is a proven interface for Classic Radio units with few radio chains. The connection is point-to-point, as a circuit switched network. The bitrates scales with the number of radio chains.

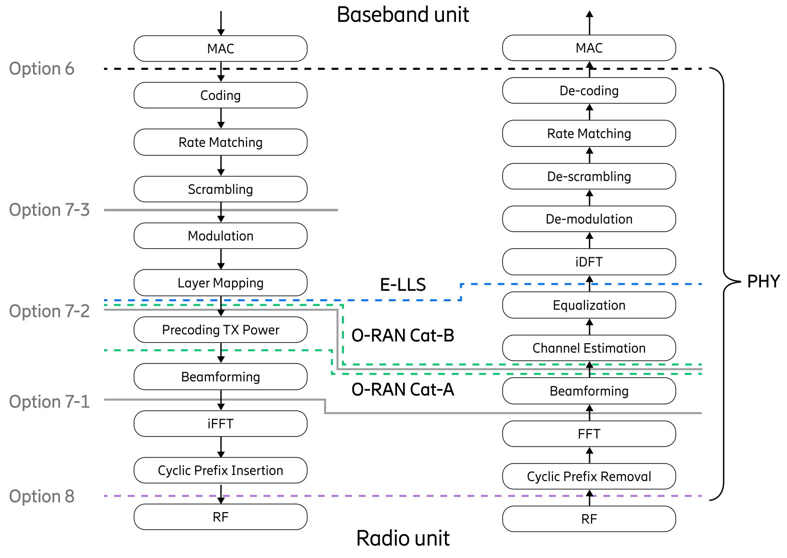
The purpose of creating a new fronthaul interface, eCPRI, was to enable more efficient implementations of Massive MIMO radio units with many radio chains, and to introduce Ethernet based transport. An Ethernet based transport enables switched front-haul.

The CPRI Forum did not standardized the functional split between the baseband and the radio unit. 3GPP did a study on different functional splits, however decided to not standardize which split to use. The O-RAN Alliance started in 2018 with the aim to "Leading the industry towards open, interoperable interfaces and RAN virtualization". To standardize functional splits between the baseband and the radio units is the task of one of the working groups.

As of today (2022Q1), the O-RAN Alliance has standardized two different categories of the so-called lower layer split, and several extensions and options.

Lower-layer split options

- **3GPP option 6**: MAC – PHY split
- E-LLS: Ericsson LLS split for Massive MIMO
- O-RAN split:
  - Category A: precoding in baseband unit
  - Category B: precoding in radio unit
- 3GPP option 8: CPRI split



The so-called lower-layer split (LLS), is a functional split between a baseband unit and a radio unit. The 3GPP study (3GPP TR 38.816) on functional splits included the "Option"-alternatives in the picture.

Option 6 is a split between the media access control (MAC) layer and the physical (PHY) layer. The payload information conveyed over that split consists of transport blocks per layer.

Options 7-1, 7-2 and 7-3 are different splits within the PHY layer.

Option 8 is a split between the PHY layer and radio processing (which can include both digital and analog processing). The payload information conveyed over that split consists of time-domain IQ samples per radio chain.

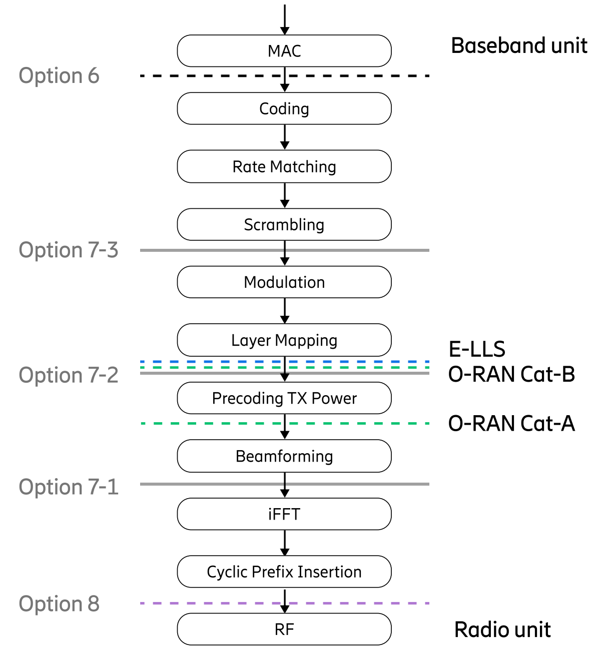
Ericsson Massive MIMO radio units for mid-band uses the split E-LLS, that for the downlink has precoding, beamforming, iFFT and cyclic prefix insertion in the radio unit, and for the uplink has cyclic prefix removal, FFT, beamforming, channel estimation and equalization in the radio unit. The payload information conveyed over this split consists of modulated bits per layer in downlink and equalized IQ samples per layer in uplink.

The O-RAN Alliance has standardized two variants of the LLS. The Category A and Category B splits, which are close the 3GPP Option 7-2. Category A is typically used for Remote radio units, and Category B for Massive MIMO radio units.

Performance difference – Downlink

O-RAN LLS vs. E-LLS

- No time-critical control loops that impacts the basic throughput performance in the downlink.
- Therefore, no basic beamforming performance difference between different splits.
- Differences in fronthaul bitrates.



For the downlink, there are no time-critical control loops that impacts the basic throughput performance, and the user data functionality is the same for all splits. Therefore, the user throughput performance is basically the same for the different splits.

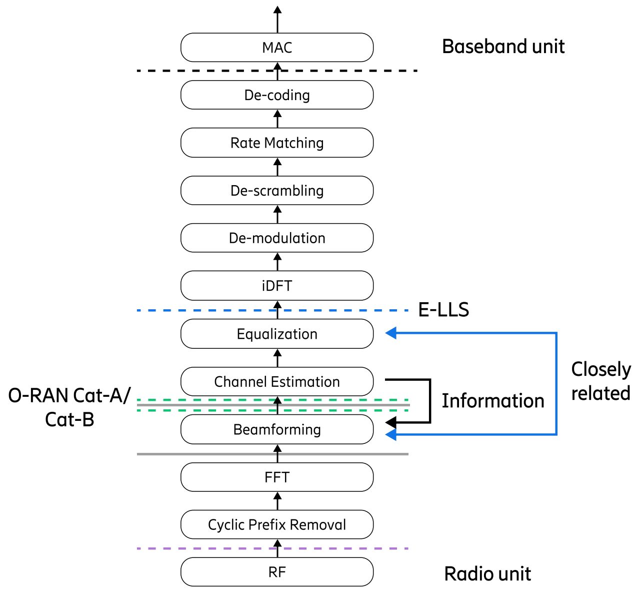
However, the different flavors of splits can give different front-haul bitrates of the user data, and they will also have different control signaling, which indirectly could lead to difference in user throughput.

For example, both E-LLS and O-RAN Cat B supports efficient so-called "modulation compression" of the user data (not shown in the picture), which O-RAN Cat A does not support. And therefore E-LLS and O-RAN Cat B will have lower fronthaul bitrates compared with O-RAN Cat A.

Performance difference – Uplink

O-RAN LLS vs. E-LLS

- Time-critical control loop that affects the O-RAN split!
  - It takes too long time to transfer the channel information back and forth between the radio and baseband units.
  - Therefore, the beamforming operation in the radio unit will be based on channel estimates from previous slots.
- With same uplink fronthaul bitrates, the O-RAN LLS Cat A/B is much worse than E-LLS.



For the uplink, there are tight relations between some of the functional blocks, the beamforming, channel estimation and equalization. In particular, the channel estimation has time-critical dependency on the beamforming.

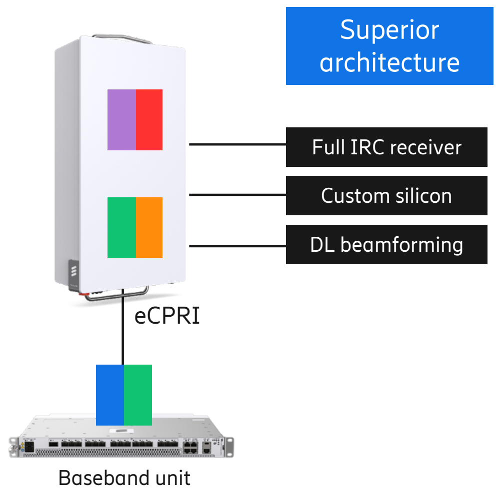
With the channel estimation in the baseband unit, the beam weights applied in the radio unit will be based on old channel estimates, due to the processing and transfer delay of sending the information back and forth between the radio unit and the baseband unit. And even if the pure transmission delay not that large, it's still too much to get the information back in time to be able to apply the beam weights based on the received reference signals in the same slot. Instead, the beam weights needs to be calculated on received SRS or DMRS from previous slots.

This means delayed information of at least several milliseconds, even 10s of milliseconds. Which gives performance issues both for the case with high velocity UEs and for the case with significant inter-cell interference.

The O-RAN split is in-between the channel estimation and the beamforming. This leads to large performance losses compared with the E-LLS split if the fronthaul bitrate is chosen to be the same for both. The performance difference can partly be compensated by increasing the fronthaul bitrates for the O-RAN split. However, even with a doubling of the fronthaul bitrate, there is still a performance gap.

The E-LLS advantage

- Ericsson is technical leader in Lower-layer split (LLS)
- The maturity of E-LLS gives performance advantages
- E-LLS has better support for various radio features, e.g., power saving features, PIM detection/avoidance.
- There is a clear uplink performance advantage of E-LLS over O-RAN LLS, which can either translate to higher user throughput or less fronthaul fibers.



Ericsson is technical leader in LLS. Ericsson was the driver of the eCPRI specification in the CPRI Forum and was first with eCPRI based products for Massive MIMO. Several years before the O-RAN Alliance was created.

Ericsson have worked on LLS for many years and now have ASICs that optimize the performance, costs and power consumptions.

The maturity of E-LLS gives us a performance advantages. And this is not only about the better uplink performance for Massive MIMO, but also that E-LLS has better support for various radio features, e.g., power saving features, PIM detection/avoidance.

Even if O-RAN LLS products eventually will reach the maturity of the E-LLS products, there are also improvements done for E-LLS that will keep the performance advantage in the future.

As mentioned before, the uplink performance issue for O-RAN LLS can partly be solved by doubling the amount of data streams transferred from the radio unit to the baseband unit. For some configurations, more fronthaul fibers are needed to allow more data streams for O-RAN LLS. For example, 100 MHz 64T64R needs one 25G fiber for E-LLS, but two fibers for O-RAN LLS.

Open interfaces does not equal multivendor

Multi-vendor interoperability is the major driving force for the O-RAN LLS from an operator perspective.

While multi-vendor interoperability over the fronthaul could offer the benefit of an increased ecosystem, it also introduce several challenges:

System integration

Life-cycle management

System performance

Assurance of KPIs & security

The major driving force for the O-RAN LLS from an operator perspective is the possibility to connect baseband and radio units from different vendors, multi-vendor interoperability.

However, open interfaces doesn't equal multivendor. Multivendor introduces several challenges.

System integration – Extensive integration project to verify an open interface between vendors adds TTM & cost to the solution

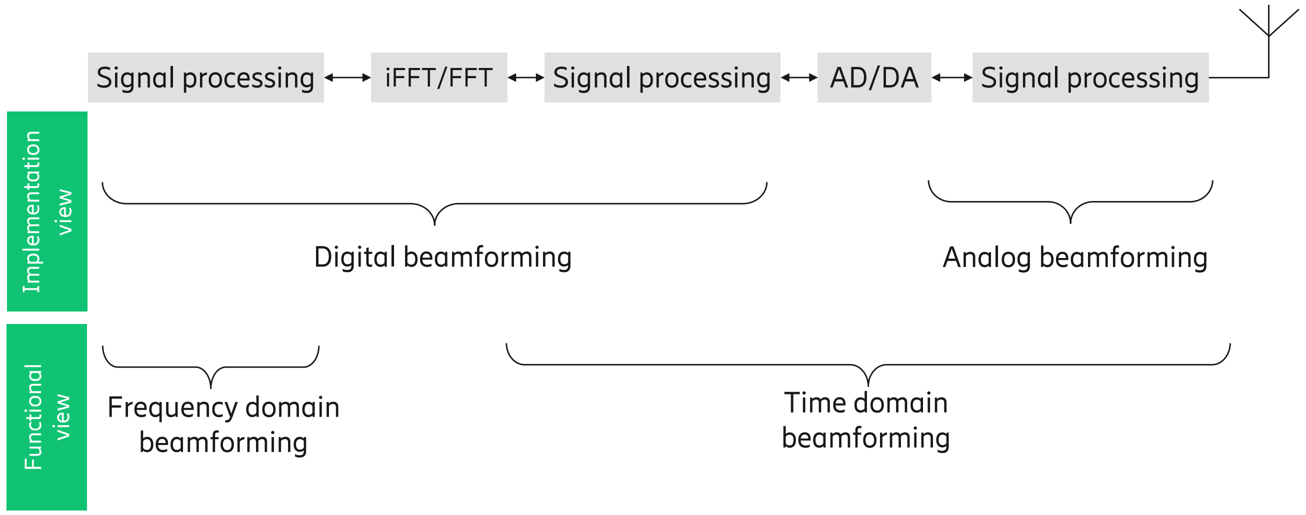
Life-cycle management – Software releases between vendors need to be coordinated, tested and verified to ensure interoperability is not broken.

System performance - Minimum common denominator dictates feature support by the vendors involved, resulting in performance limitations.

Assurance of key performance indicators & security – Challenging root cause analysis to identify vendor at fault and who is responsible for providing fixes.

Example: air interface is fully standardized in 3GPP, including detailed test specifications, but still requires extensive interoperability device testing (IODT) between the RAN vendors and the chipset vendors.

Analog vs. digital/time domain vs. frequency domain



This picture illustrates different beamforming types.

In the transmission direction, the signal flow starts from the left with a signal processing block, which then is followed by an iFFT/FFT that translates the signal from frequency domain to time domain. After another signal procssing block, the signal is transfered from the digital domin to the analog domain in the AD/DA block. And then finally some signal processing is performed before the signal reaches the air.

The reception can be descried in a similar way, but with a signal flow starting from the right.

Digital beamforming = beamforming in the digital domain

Analog beamforming = beamforming in the analog domain

Frequency domain beamforming is always performed by digital beamforming.

Time domain beamforming can be performed by digital beamforming, analog beamforming, or a combination of both digital and analog beamforming.

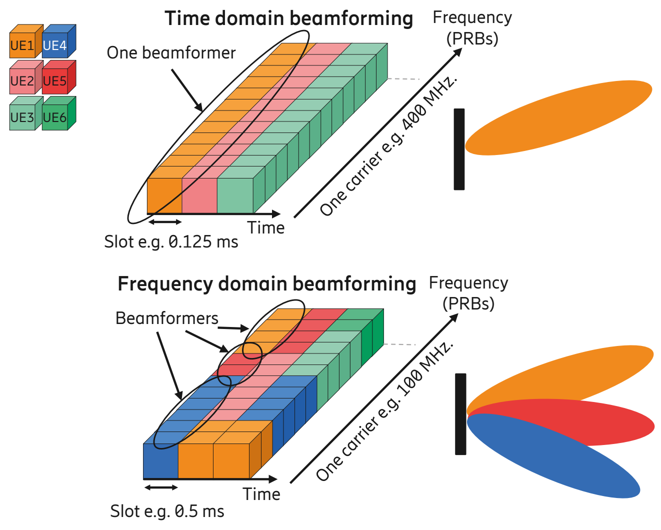
Beamforming implementations

Time domain beamforming vs. Frequency domain beamforming

- Time domain beamforming => Few beamformers
- Frequency domain beamforming => Many beamformers

Digital beamforming vs. Analog beamforming

- Digital beamforming enables frequency domain beamforming
- Total beamforming complexity: number of antennas x number of beamformers
- If the total complexity becomes to high, analog beamforming, using time domain beamforming, reduces the complexity, but lowers the throughput performance.



Beamforming can be implemented either in the frequency domain or in the time domain. The fundamental difference between these two implementations is the frequency resolution of the signal processing. Typically, frequency domain beamforming operates on a resource block level (typically a few 100 kHz), and time domain beamforming typically operates on a carrier level (typically in the order of 100 MHz).

A beamformer is an entity that operates on a frequency/time block with one set of beamforming weights. Frequency domain beamforming has more beamformers than time domain beamforming, due to its finer frequency granularity. This finer granularity gives a higher level of control, and therefore an expected higher throughput performance.

Frequency domain beamforming requires digtial beamforming. Time domain beamforming can be implemented either by digital beamforming or analog beamforming.

The total beamforming complexity depends both on the frequency resolution and the number of radio chains. If the total complexity becomes too large, using analog beamforming (and time domain beamforming) is a way to reduce the complexity, but at the expense of throughput performance.

Therefore, the following implementations are typical choices of beamforming.

- Mid-band 100 MHz bandwidth, 64 antennas => Digital frequency domain beamforming
- High-band 800 MHz bandwidth, 512 antennas => Analog time domain beamforming

Summary – Key points

The development of Massive MIMO radio units has led to two important architecture choices, the fronthaul solution and the beamforming type.

The eCPRI fronthaul interface:

- Enables lower fronthaul bitrates by supporting more efficient functional splits.
- Enable switched fronthaul solutions due to its Ethernet-based transport.

Beamforming types:

- **Digital beamforming** in the **frequency domain** is preferable for radio units with limited number of antennas and limited frequency bandwidth, typically for **mid-band products**.
- **Analog beamforming** in the **time domain** is preferable for radio units with large number of antennas and large frequency bandwidth, typically for **high-band products.**

The development of Massive MIMO radio units has led to two important architecture choices, the fronthaul solution and the beamforming type.

The eCPRI fronthaul interface enables lower fronthaul bitrates, by supporting more efficient functional splits, and enables switched fronthaul solutions due to its Ethernet-based transport.

For mid-band products, typically digital beamforming in the frequency domain is preferable since the radio units have both a limited number of antennas and a limited frequency bandwidth (compared with high-band products).

For high-band products, typically analog beamforming in the time domain is preferable since the radio units have both a large number of antennas and a large frequency bandwidth (compared with mid-band products).

9. Abbreviations

3GPP	3rd generation partnership project
2D	two dimensional
AAS	advanced antenna systems; note: synonymous to Massive MIMO
ACK	acknowledgement (positive)
AI	artificial intelligence
AIR	antenna integrated radio
AOSA	array of subarrays
AR	augmented reality
AS	antenna switching
BF	beamforming
BB	baseband
BW	bandwidth
CA	carrier aggregation
CAPEX	capital expenditures
CC	component carrier
CCH	control channel
CO2	carbondioxide
CoMP	coordinated multi-point
CQI	channel quality indicator
CSP	communication service provider
CSI	channel state information
CSI-RS	CSI-reference signal
DC	dual connectivity
DCI	downlink control information
DFT	discrete fourier transform
DL	downlink
DMRS	demodulation reference signal
EIRP	equivalent isotropic radiated power
EIS	equivalent isotropic sensitivity
EM	electro magnetic
EMF	electro magnetic field
EPO	energy performance optimizer
FCC	federal communications commission
FDD	frequency division duplex
FH	frequency hopping
GHz	giga Hertz
GP	guard period
gNB	node B (NR)
HARQ	hybrid automatic repeat request
HW	hardware
Hz	hertz
IBW	instantaneous bandwidth
IEEE	institute for electrical and electronics engineers
IRC	interference rejection combining
ISD	inter-site distance
MAC CE	medium access control channel element
MBB	mobile broadband
Mbps	mega bit per second
MIMO	multiple input multiple output
ML	machine learning
mmWave	milimeter wave
MCG	master cell group
MSGx	message x
MU-MIMO	multi-user MIMO
NACK	negative ACK
NR	new radio
OPEX	operational expenditures
ORAN	open RAN architecture
OSI	open systems interconnection
PCell	primary cell
PSCell	primary SCG cell
PBCH	physical broadcast channel

PDCCH	physical downlink control channel
PDSCH	physical downlink shared channel
PMI	precoding matrix indicator
PRACH	physical random access channel
PUCCH	physical uplink control channel
PUSCH	physical uplink shared channel
PSK	phase shift keying
PSS	primary synchronization signals
QAM	quadrature amplitude modulation
QPSK	quadrature PSK
QoS	quality-of-service
Qx	quarter x; x=1,2,3,4
R	receiver radio chain
RAN	radio access network
RAN1	3GPP working group for radio layer 1 (physical layer)
RAN2	3GPP working group for radio layer 2 and radio layer 3 radio resource control
RAN4	3GPP working group for UTRAN/E-UTRAN/NG-RAN architecture and related network interfaces
RAN4	3GPP working group for radio performance and protocol aspects
RAN5	3GPP working group for mobile terminal conformance testing
RAR	random access response
RMSI	remaining system information (SIB1)
RI	rank indicator
RRC	radio resource control
RF	radio frequency
SCell	secondary cell
SCG	secondary cell group
SCS	subcarrier spacing
SIB x	system information block x
SINR	signal-to-noise-and-interference ratio
SNR	signal-to-noise ratio
SON	self-organizing network
SUL	supplementary uplink
SpCell	special cell
SR	scheduling request
SS	synchronization signals
SRS	sounding reference signal
SSB	block containing SS and PBCH
SSS	secondary synchronization signal
SU-MIMO	single-user MIMO
SW	software
T	transmitter radio chain
TCBW	total configured bandwidth
TCO	total cost of ownership
TDD	time division duplex
TRP	transmission point
UCI	uplink control information
UE	user equipment
UL	uplink
VR	virtual reality
XR	extended reality
YoY	year-on-year

Ericsson.com

EN/LZT 4/28701-FGB1010987 Uen Rev A