D1.4

RADIO NETWORK DEPLOYMENT OPTIONS FOR SMART MANUFACTURING

The 5G-SMART project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 857008.

Radio network deployment options for smart manufacturing

Grant agreement number: Project title: Project acronym: Project website: Programme:	857008 5G Smart Manufacturing 5G-SMART www.5gsmart.eu H2020-ICT-2018-3
Deliverable type: Deliverable reference number: Contributing workpackages: Dissemination level: Due date: Actual submission date:	Report D4 WP1 Public 2020-11-30 2020-11-30
Responsible organization: Editor(s): Version number: Status:	ERI-SE (ERI-FI) Kimmo Hiltunen V1.0 Final
Short abstract:	This report discusses the different radio network deployment options for smart manufacturing. It provides an overview of the different options and describes the input data necessary to select the most feasible deployment options for the desired industrial 5G scenarios and services. Furthermore, the report discusses and analyzes the feasibility of the radio network options for the different NPN architectures and the impact of the stakeholder, e.g., the mobile network operator or the industrial party, deploying and operating the non-public network. Finally, the system-level simulation results presented in this report demonstrate how the various radio network deployment options, such as the chosen frequency band, the applied TDD pattern and the type of the base station antenna, can significantly impact the performance of an industrial URLLC network, defining the type of URLLC services that the network can support.
Keywords:	Radio network deployment, radio network performance, spectrum, co-existence, non-public network

Contributor(s):	Kimmo Hiltunen (ERI-FI) Joachim Sachs (ERI-SE) Meriem Mhedhbi (Orange) Stefan Cerovic (Orange)

Disclaimer

Executive summary

This report provides an overview of the different radio network deployment options for smart manufacturing and the input data necessary to select the most feasible deployment options for the desired industrial 5G scenarios. Furthermore, the report discusses and analyzes the feasibility of the radio network deployment options for the different non-public network (NPN) architectures specified within the 3GPP and the impact of spectrum options available for the stakeholder deploying and operating the non-public network.

The analysis presented in this report suggests that the various radio network deployment options and features can significantly impact the performance of a non-public factory network supporting ultra-reliable low-latency communication (URLLC) services. To start with, the chosen frequency band has an impact on the achievable latency and the maximum system capacity defining the type of URLLC services that the non-public network can support. Furthermore, the applied TDD downlink-uplink pattern defines a lower bound for the achievable latency, has a clear impact on the maximum system capacity, and is also one of the key factors affecting the co-existence performance between the non-public network and the neighboring TDD networks operating in the same band. Finally, since the latency-critical URLLC services are sensitive to inter-cell interference, the type of the base station antenna has a large impact on the overall radio network performance.

The feasibility of the different radio network deployment options, as well as the corresponding radio network performance depends both on the chosen network architecture and the spectrum options available for the stakeholder deploying and operating the non-public network. For example, an industrial party has considerably less spectrum options available compared to a mobile network operator (MNO): it is very unlikely that the industrial party has access to low- or mid-band FDD spectrum, which makes it more difficult for the non-public network to support M-MTC services. Furthermore, it can be challenging for an industrial party having access to only a single mid-band TDD carrier to resolve all co-existence problems between the public network and the non-public network without collaborating with the MNO.

An MNO has an option to provide the non-public network “as a service,” integrated with the public network, in which case the MNO can utilize all its spectrum assets, possibly combined with the local spectrum when available, to provide all the required communication services. It also becomes straightforward to design, combine or coordinate the overlaid public network and the non-public network to resolve most of the co-existence problems. However, there could still be a need for the MNO to agree and coordinate with the neighboring MNOs operating in the same band to secure a sufficiently low level of inter-network interference to guarantee the desired URLLC performance.

Finally, if neighboring non-public networks are operating on the same frequency channel or in particular if the networks contain outdoor small cells, there could be a need to mitigate the interference between them, e.g., by synchronizing the applied TDD downlink-uplink patterns. However, if service-optimized (and NPN-specific) TDD patterns are preferred instead, some other means, such as a careful planning of the radio network deployments and agreeing on appropriate emission limits, should be applied to control the level of the inter-network interference.

Contents

1 Introduction

Wireless connectivity is increasingly becoming a necessity for business-critical services in industrial processes, such as those related to assembly lines and other modes of production. However, the specific needs and requirements can differ greatly between the different industries.

Perhaps the most common, and often the most crucial, requirements are related to high network availability and reliability. However, the requirements differ between the different manufacturers or industries and even within an industrial site. For example, an electronic component factory might realistically need to power thousands of simple sensors in an energy-efficient way. At the same time it may require low-latency and cloud-based steering of robotic arms. Hence, a connectivity solution would need to cater for various network needs simultaneously as well as cost-efficiently fulfill demanding use cases and services normally part of a public network, such as voice services, access to internet, and track and trace services. Figure 1 shows an example of a smart manufacturing site with diverse wireless devices and a wide range of connectivity requirements [NHB+20].

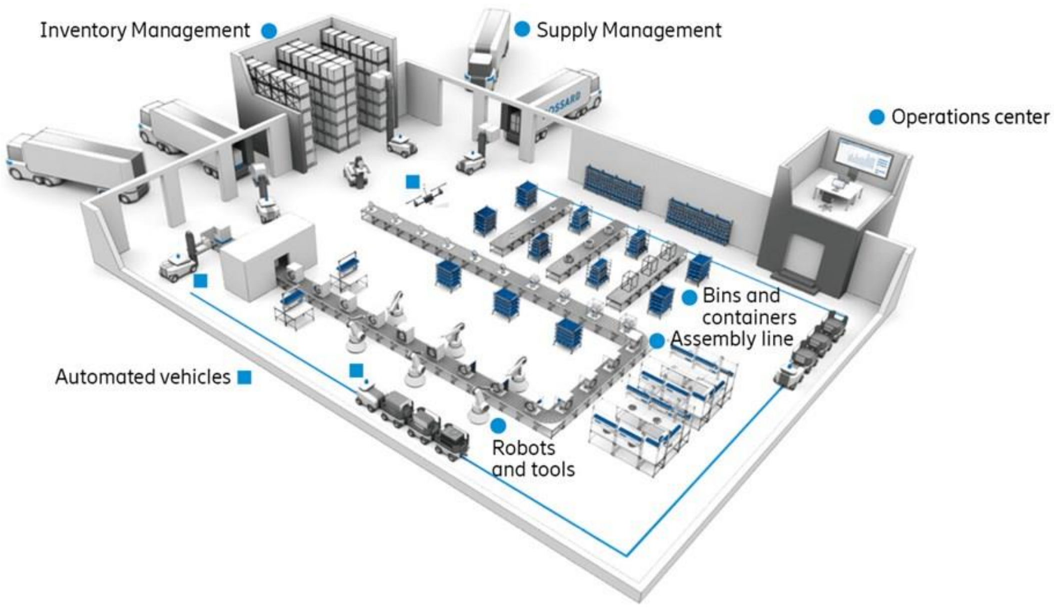

Figure 1. A smart factory with diverse use cases (source: Ericsson [NHB+20])

For some industries, service-level agreements (SLAs) will satisfy the needs for guaranteed network uptime and quality. However, for most of the manufacturing industries, the requirement for high network availability is critical for their operations. Therefore, they would either need to demand legally binding liabilities from external service and spectrum providers or get access to their own licensed spectrum, depending on the willingness of the company to build up in-house expertise and the trade-off between the cost and the reward of doing that.

The connectivity available at a given manufacturing site might not be enough to cover the complete set of requirements. For example, manufacturers might want the ability to upgrade and track products in the field to enable cost-efficient upgrades in the aftermarket area and improve customer experience. Hence, they would need access to certain wide-area communication services in addition to the local connectivity.

A production facility is normally a 15-20 years' lifecycle investment, and manufacturers will likely seek the availability and reliability of their connection over this period. Considering that businesses tend to prefer the freedom of choice when it comes to suppliers, the request would likely be to guarantee uninterrupted service for these 15-20 years and, at the same time, maintain flexibility in the supplier dimension. Another point manufacturers would like to consider in this circumstance would be how to handle commercial agreements for equipment for such a long time.

Finally, different industries and companies can have different strategies regarding what operations are core to their business and should be kept in-house (instead of those bought as a service). This will likely be reflected in the way they address connectivity. Consequently, there is a need to cater for industries that would like to own and operate equipment themselves as well as those of the opposite inclination, whose services can be outsourced and provided by either their own private networks or from shared public networks.

Considering the above, it becomes obvious that the communication network offering the wireless connectivity has to be tailored for each particular deployment scenario. While doing so, a large variety of different aspects has to be considered, e.g., characteristics of the use case requiring wireless connectivity, the overall business case, spectrum regulations, characteristics of the industrial site, and so on. All in all, finding the most appropriate combination of the different radio network deployment options to satisfy the desired communication service requirements will typically be a complex exercise, which will eventually influence all the stakeholders: industrial parties, mobile network operators (MNOs), service providers and the network vendors.

1.1 Objective of the report

The objective of this report is to identify different radio network deployment options for non-public industrial 5G networks to provide the required communication services for smart manufacturing. This report briefly describes the different radio network deployment options and how they are linked with each other. Furthermore, the required inputs that should be considered during the selection of the applied radio network deployment option are discussed as well. Finally, this report provides results from some initial performance evaluations to demonstrate the impact of a few selected radio network deployment options on the performance of the standalone non-public factory network.

The overall goal of this report is to provide a structured view on the problem of selecting the most appropriate radio network deployment options for the desired use case and the deployment scenario. A more detailed evaluation of both the various radio network deployment options and the corresponding radio network performance will then be provided in the next deliverable (D1.5) due at the end of the 5G-SMART project.

1.2 Relation to the other reports

This report is closely related to the other work performed within 5G-SMART Work Package (WP) 1 ("Use Cases, Business Models and Network Design"). This includes the 5G-SMART Deliverable 1.1 [5GS20-D11] discussing the forward-looking smart manufacturing use cases, requirements and key performance indicators, and the work on the common 5G terminology [5GS20-Term]. Furthermore, it is expected that the findings and results presented in this report will act as an input for the future work on the business aspects.

This report is also related to the work performed within 5G-SMART WP5 ("5G Optimization and Design for Manufacturing") when it comes to the different architecture options for the non-public networks and the descriptions of the different operation models (i.e., roles and stakeholders). In particular, some of the definitions presented in 5G-SMART Deliverable 5.2 [5GS20-D52] discussing about 5G network architecture options and assessments are reused within this report. However, any lower-level details regarding the network architecture and the operation model are omitted from this report and can be found in [5GS20-D52].

1.3 Structure of the report

The remainder of the report is organized as follows. Chapter 2 provides a definition of a radio network deployment option. It starts with an overview of a deployment option and a brief definition of the different roles and stakeholders. Then, the required inputs for selecting the most feasible deployment option, namely the use case and the deployment scenario, are described. Finally, the various radio network deployment options and features are briefly introduced and discussed.

Chapter 3 discusses the feasibility of the different radio network deployment options for an independent standalone non-public network (NPN). The chapter starts with an introduction of the architecture option and what it means from the radio network deployment options' point of view. Then, the available spectrum options and the different co-existence scenarios are discussed in more detail.

Chapter 4 discusses the feasibility of the different radio network deployment options when the non-public network is deployed in conjunction with a public network, i.e., in the case of shared radio access network (RAN), shared RAN and control plane, and NPN hosted by the public network. Otherwise, the structure of the chapter is the same as in Chapter 3, with the difference that only the things related to spectrum and co-existence that differ from the independent standalone deployment are discussed.

Chapter 5 presents and discusses results from a number of performance evaluations considering a standalone factory network. Two different types of evaluations are discussed: a) performance of an isolated factory network supporting ultra-reliable low-latency communication (URLLC) services and b) co-existence between an industrial URLLC network and neighboring (public or non-public) 5G networks.

Finally, conclusions are drawn in Chapter 6.

2 Definition of a radio network deployment option

2.1 Overview

Figure 2 presents an overview of a deployment option and the required inputs. The overall topic of deployment options can be divided into different sub-areas, including the 5G network architecture and network management options (discussed in more detail in *5G-SMART Work Package 5 – 5G Optimization and Design for Manufacturing*), as well as the radio network deployment options and the business aspects (discussed in *5G-SMART Work Package 1 – Use Cases, Business Models and Network Design*). The required input to the selection of a deployment option consists of the use case description and the different aspects describing the deployment scenario.

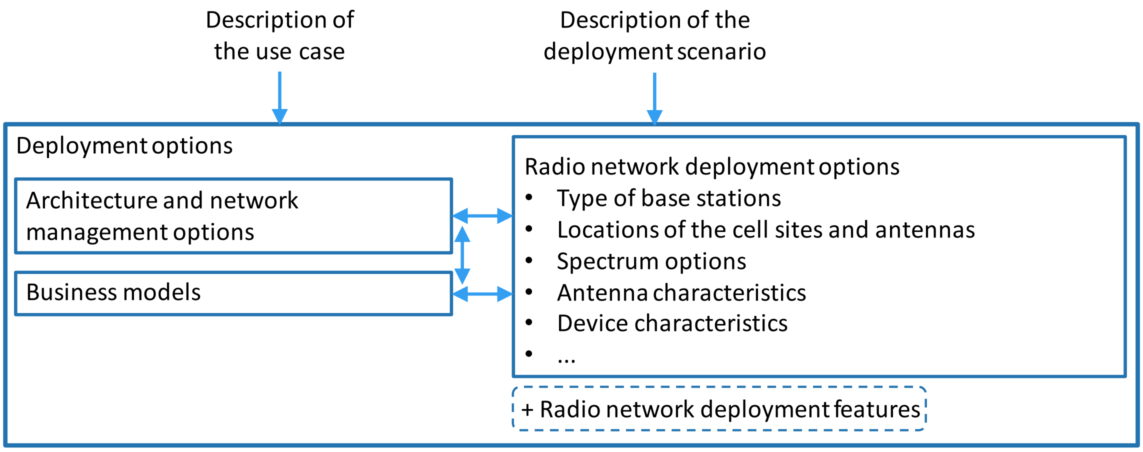

Figure 2. Overview of a deployment option

There is a clear connection between the business aspects, feasible architecture options and the radio network deployment options. For example, depending on the stakeholder (e.g., MNO or an industrial party) deploying the non-public network, the available architecture options will look different. Similarly, the availability of spectrum, co-existence scenarios as well as the overall radio network performance will depend on the stakeholder deploying the non-public network.

Next, the key aspects of Figure 2 are discussed in more detail from the radio network deployment options' point of view, starting with a brief introduction of the different roles and stakeholders.

2.2 Description of the different roles and stakeholders

When it comes to the ownership, deployment, operation, and management of non-public networks, the following roles can be defined:

- NPN Owner is the role of owning the NPN infrastructure (including both the hardware and the software components).
- Spectrum Owner is the role of having the right to transmit radio signals at a certain frequency band.

- NPN Integrator is the role of deploying and configuring the NPN according to a chosen architecture and making it ready to use.
- NPN Operator is the role of operating and managing the NPN on a day-to-day basis.
- NPN User is the role of using the services offered by the NPN.

Furthermore, three different stakeholders can be identified:

- Mobile Network Operator (MNO) is the stakeholder which owns and manages a public land mobile network (PLMN).
- Industrial Party is the stakeholder which requests NPN services for performing an industrial task or a group of industrial tasks.
- Third Party is a stakeholder which cannot be categorized as an MNO or an industrial party, e.g., a network vendor, system integrator or other third-party supplier. A third party can provide the NPN user with services such as the deployment, integration and management of the non-public network.

When it comes to assigning the roles to the different stakeholders, the role of the NPN user is exclusively assigned to the industrial party, while all the other roles can be taken by any of the three stakeholders, as discussed in more detail in 5G-SMART WP5 (see e.g., [5GS20-D52]). Hence, depending on the chosen deployment option, there can be numerous feasible combinations of the different stakeholders and roles. Therefore, a simplified approach has been chosen for this report, particularly since the topic of network operation and management will not be discussed in detail.

The more detailed discussion and analysis in this report are based on the following *NPN operation models* (note that the different spectrum options will be discussed in more detail in Section 2.5.4):

- The non-public network is *provided by the MNO*, utilizing its own national spectrum assets and potentially also the local spectrum. Industrial party has the role as the NPN user, while all the other roles are assigned to the MNO. This alternative corresponds to the NPN operation model 3 in [5GS20-D52].
- The non-public network is *deployed by the industrial party*, utilizing local spectrum assets. In this case all the roles are directly taken by the industrial party. This alternative corresponds to the NPN operation model 1 in [5GS20-D52].

2.3 Description of the use case

The use case will describe the application and communication service(s) for the desired deployment, the service characteristics, and the corresponding communication service requirements.

Figure 3 presents the anatomy of a use case, as defined by [5GS20-D11] and [5GS20-Term]. A use case (e.g., 5G-connected robot, cloud-based mobile robotics) consists of one or more application services (e.g., real-time control, ultra-high definition video), and each application service can be mapped to a corresponding communication service with a certain set of performance requirements and characteristics. A communication service category (e.g., enhanced mobile broadband (eMBB), massive machine-type communications (M-MTC) and URLLC) represents a set of communication services that share some common characteristics in terms of connectivity. Each communication service has its own requirements and key performance indicators (KPI) (e.g., latency, reliability, and system capacity), and characteristics, such as traffic model and communication area. These requirements and characteristics will then act as an input to the selection of the most feasible deployment options.

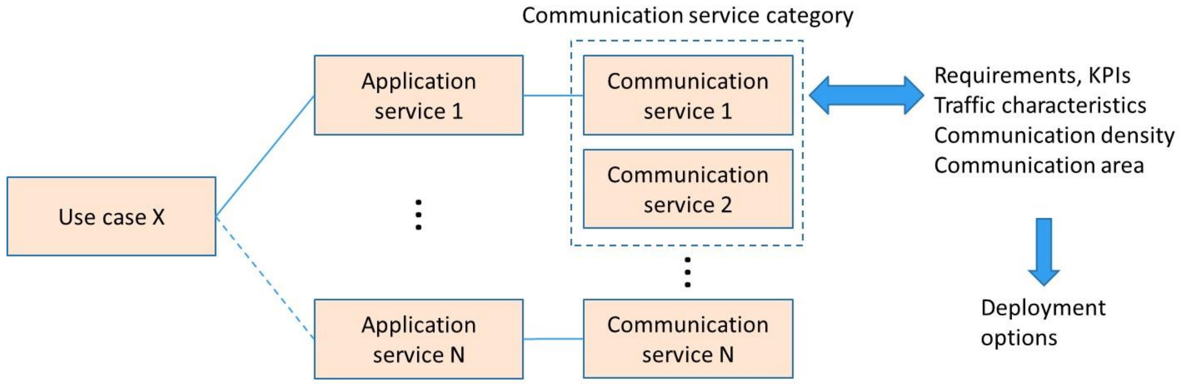

Figure 3. Anatomy of a use case

2.4 Description of the deployment scenario

The different aspects of the deployment scenario describe the environment where the desired use case will be offered. The required input consists of the following aspects:

- Deployment layout. Is the deployment within a single building or within an industry campus consisting of multiple buildings and outdoor areas?
- Coverage needs. Is the coverage needed only within indoor areas, only in outdoor areas, or both?
- Factory layout. How does the factory building look like? Is an actual blueprint available, or should the factory layout model be more generic, like the one related to the 3GPP Indoor Factory channel model [3GPP19-38901]. In case of the generic 3GPP model, the input consists of the following: average clutter size, average clutter density, average clutter height, inner wall type or penetration loss, and the outer wall type or penetration loss.
- Neighborhood. Are there any neighboring public or non-public networks, or is the deployment isolated? How does the potential neighbor look like: is the network deployed within a single building or within an industry campus?
- Macro network layout. The industrial site will typically be overlaid by a macro network or multiple macro networks. What is the cell size or the inter-site distance of the overlaid macro network? What is the separation distance from the industrial site towards the closest macro site? Which frequencies are being used or are available?
- Environment. How does the environment around the industrial site look like? Is it urban, suburban, or rural?

2.5 Radio network deployment options and features

This section will briefly introduce the different radio network deployment options and features available for an industrial 5G network. In general, the term "radio network deployment options" refers to the fixed characteristics of the network (e.g., the type, number and the locations of the base stations and antennas, the used spectrum and the devices) that should be determined at the deployment stage. Furthermore, a radio network deployment option includes different kinds of features that can be (re-)configured once the network starts its operation. Examples of features include, for example, the different scheduling options and the applied numerology (i.e., sub-carrier spacing (SCS) and slot length).

2.5.1 Type of base stations

Depending on the deployment, the communication service can be provided with different kinds of base stations (BS), which in the case of 3GPP NR are referred to as gNodeBs (gNB). For example, the following base station types could be used for non-public network deployments:

- Macro base stations include a large variety of different types of products with different characteristics (both in terms of bandwidth and transmission power), for example, 200 W/100 MHz. Macro base stations are typically used to provide wide-area coverage within the low- and mid-band spectrum (see Section 2.5.4).
- Outdoor small cell base stations ("micro base stations") could be assumed to correspond to the medium range base station class defined in 3GPP [3GPP20-38104]. The maximum allowed output power is 38 dBm/antenna connector, or the maximum allowed total radiated power (TRP) is 47 dBm [3GPP20-38104]. Small cell base stations are typically used to enhance coverage or capacity within a limited geographical area, and they are normally deployed within the mid- or high-band spectrum.
- Indoor (industrial) small cell base stations. Due to the sufficiently large indoor areas, and thus, sufficiently large minimum coupling losses, the indoor industrial small cell base stations can be assumed to correspond to the medium range base station class defined in 3GPP [3GPP20-38104].
- Indoor (enterprise) small cell base stations ("pico base stations") could be assumed to correspond to the local area base station class defined in 3GPP [3GPP20-38104]. Maximum allowed output power is 24 dBm/antenna connector, or the maximum allowed total radiated power is 33 dBm [3GPP20-38104].
- Active distributed antenna system (DAS), for example, the Ericsson radio dot system (RDS) [ERI20-RDS]. An active DAS can be an efficient way to secure coverage throughout the desired communication area instead of deploying multiple base stations. The maximum transmit power per antenna will typically depend on the bandwidth and the number of multiple-input and multiple-output (MIMO) streams, but it can be assumed to be in the order of 250 mW – 1 W per antenna. Active distributed antenna systems are often used for indoor deployments at the mid-band, and in the future, also for the high-band.

2.5.2 Antenna characteristics

A proper selection of both the base station and the device antenna can have a large impact on the achieved radio network performance. This is due to the fact that the deployed antennas have an impact on both the received signal power, as well as the level of the inter-cell interference transmitted to, or received from the neighboring base stations and devices. The base station antennas can be divided into the following groups:

- Isotropic and omnidirectional antennas are simple antennas with low directivity. Isotropic antennas are radiating with equal gain towards each three-dimensional (3D) direction, while omnidirectional antennas have roughly a constant antenna gain in the horizontal domain, but a “beam-shape” in the vertical domain. The directivity of the omnidirectional antenna will then depend on the half-power beamwidth of the vertical antenna pattern. Furthermore, the omnidirectional antenna pattern can be made to look more “umbrella-like” with the help of an electrical downtilt. These kinds of antennas can be used for all frequency bands, although they are perhaps less feasible for high-bands due to coverage reasons.
- Directional antennas are typically used for sectorized deployments, and they can be mounted either on the wall, pillar, or a pole (pointing sideways), or on the ceiling (pointing downwards). The cell coverage areas and antenna patterns can also be modified with the help of mechanical or electrical downtilting. The directivity of the antenna will depend on the half-power beamwidths of both the horizontal and vertical antenna radiation patterns. Directional antennas are useful for all frequency bands to enhance the coverage and increase the capacity by limiting the cell overlap.
- Beamformed antennas. While the directional antennas have one fixed “beam” serving all the users within the whole sector, the beamformed antennas offer multiple narrow beams to serve different users in different parts of the cell coverage area. There are three different ways to realize beamforming [MMM16-D51][MMM17-D52]: analog beamforming, digital beamforming, and hybrid beamforming.
    - In the case of analog beamforming, a number of narrow candidate beams pointing in different directions can be created, but usually, only one beam per polarization can be transmitted at a time. Since the number of radio chains is small, also the required number of active components (e.g., power amplifiers and digital-to-analog (DA) and analog-to-digital (AD) converters) is small, and hence, the cost and energy consumption of the antenna array becomes low. Another benefit is that analog beamforming is feasible for both mid- and high-band since it is not sensitive to the number of antenna elements or the channel bandwidth. The obvious downside is that frequency-selective beamforming is not possible. Since only one beam per polarization can be transmitted at a time, the scheduling of users has to be based on time domain multiplexing (TDM), while frequency domain multiplexing (FDM) of users would be possible only within the transmitted beam. In general, TDM is perfectly feasible for the mobile broadband type of traffic, which is not latency-critical and where the users typically have large volumes of data to transmit. However, in the case of machine-type communications (MTC), the traffic volumes are often small, and the traffic can also be latency-critical. If the number of served MTC users is small, TDM

would still be a feasible option, but in order to maximize the system capacity, FDM should be applied instead.
  - In the case of digital beamforming, the antenna array contains a radio chain per antenna element, enabling the transmission of multiple beams (pointing in different directions) at the same time ("user equipment (UE) -specific beamforming"). Hence, frequency-selective beamforming is possible, which means that the users can also be scheduled in the frequency domain while still benefiting from the high antenna gains. The downside of digital beamforming is that the cost and energy consumption can become high as the number of required radio chains increases, making the digital beamforming often unfeasible for deployments at the high-band, where large antenna arrays may be desired for coverage reasons. Another fact making the use of digital beamforming more difficult at the high-band is the considerably wider channel bandwidths compared to the mid-band, which increases the challenges related to the DA/AD converters and signal processing (e.g., fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT)).
  - Hybrid beamforming combines both analog and digital beamforming methods to enable a more flexible utilization of narrow beams, while still keeping the cost and power consumption under control. Hybrid beamforming allows a limited support of frequency-selective beamforming. However, an open question to be investigated is how many simultaneous beams would be sufficient to enable an efficient support of different ultra-reliable low-latency communication (URLLC) services.

When it comes to devices, the lack of physical space and the desire to simplify the signal processing will typically result in less complex antenna solutions compared to base stations. While the devices operating at the low- and mid-band typically have isotropic or omnidirectional antennas, at the higher frequency bands the device antennas could become more directional, or even utilize (analog) beamforming techniques. However, beamforming will potentially introduce some new challenges related to beam selection or tracking, since for example the rotation of the device may happen quickly and radically change the performance of the different beams.

2.5.3 Locations of the cell sites and antennas

One of the key outputs of a radio network planning process is to define the required number and types of the base stations, as well as the locations of the base station and device antennas to fulfill the required communication service requirements within the desired communication area. Aspects that could be considered include, for example, the following:

- Can the service be offered from the existing macro sites, or is a deployment of dedicated base stations needed?
- If dedicated base stations are needed, should they be located outdoor or indoors?
- How many base stations are needed, i.e., what is the required inter-site distance?
- Where should the indoor antennas be mounted? On the walls, ceiling, pillars/poles, or clutter? Above or below the average clutter height? Are there some special details in the factory building layout that should be taken into account (obstacles, walls, wall material, …) when

deciding the locations of the antennas? Are the devices uniformly distributed around the desired coverage area, or are there some traffic or device hotspots?
- How should the antennas be oriented to optimize the radio network performance, i.e., to find the proper balance between the achieved coverage and the level of the inter-cell and inter-network interference?

2.5.4 Spectrum options

Any deployment of industrial 5G radio networks requires access to spectrum, either unlicensed, shared, or exclusively licensed. Access to licensed spectrum can be provided in the following ways:

- Service Level Agreements (SLA) between the industrial parties and the MNOs.
- Spectrum leasing, where the MNO is acting as a lessor towards the industrial parties.
- Local licensing, where the national regulator licenses spectrum directly to industrial parties over a limited geographical deployment, typically associated with property rights for the covered area.

Referring to the discussion about roles and stakeholders in Section 2.2, the first alternative corresponds to a scenario where the non-public network is provided by the MNO, while the second and the third alternative correspond to a scenario where the non-public network is deployed by the industrial party.

Usage of an unlicensed spectrum for 5G New Radio (NR) is possible through 5G NR for unlicensed spectrum (NR-U) and may be an alternative solution for scenarios where a licensed spectrum is not available or a complement to using local spectrum only. 5G NR-U can operate either standalone in unlicensed spectrum or together with 5G NR in carrier aggregation or dual connectivity mode. However, for NR-U operating standalone in unlicensed spectrum, the quality of service cannot be guaranteed in the same way as with licensed spectrum. Using unlicensed spectrum for 5G NR-U carrier aggregation with licensed spectrum provides great flexibility between licensed and unlicensed spectrum use and provides possibilities to scale depending on load in the factory deployment. This report will focus only on radio network deployments on a licensed spectrum, and thus, the topic of using unlicensed spectrum for industrial internet of things (IIoT) will not be discussed further.

In general, 5G NR deployment has been defined for a large number of candidate spectrum bands, see [3GPP20-38101], where the availability of the specific bands depends on the country of the deployment. 5G NR allows the simultaneous usage of multiple bands by means of carrier aggregation or dual connectivity. Furthermore, by means of dynamic spectrum sharing [KAF+18][ERI19-ESS], 5G NR can be deployed in the same carrier that is already used for 4G Long Term Evolution (LTE) by dynamically sharing the carrier bandwidth between LTE and NR. Thus, spectrum bands defined for LTE can be opened up for NR, allowing a smooth migration from 4G to 5G.

The characteristics of the different licensed spectrum bands relevant for industrial 5G deployments can be described as follows[1]:

- Low-band (below 1 GHz, e.g., 700, 800 and 900 MHz) offers good wide-area coverage and indoor penetration for nationwide deployments by the MNOs. The duplexing scheme is frequency division duplex (FDD), and the networks will typically apply a low 5G numerology (SCS equal to 15 kHz and the slot length equal to 1 ms). The available bandwidth is typically quite limited[2], and the industrial 5G deployments would have to share the resources with public eMBB services resulting in a relatively low capacity. Therefore, the low-band will mainly be an option for low-capacity URLLC services, offered by the MNOs towards small enterprises. In addition, outdoor low-band deployments can be useful for providing coverage for logistics use cases along transport paths. A potential topic to be studied further is if it would be feasible to reuse a low-band FDD carrier within a factory and utilize carrier aggregation between low-band and mid/high-band to achieve both 100% service coverage and decent capacity. This reuse would, unfortunately, result in inter-network interference between the macro and the factory network (as will be discussed in Section 3.3), but it could potentially be mitigated by moving the visiting or close-by macro users to other macro frequencies or to LTE.
- Mid-band (between 1 GHz and 10 GHz, e.g., 1800, 2100, 2300, 2600 and 3500 MHz) can typically be used to provide coverage and capacity for wide-area macro deployments but also for local micro or pico deployments. The mid-band offers decent bandwidth[3] in total over a few different carriers, and the duplexing mode is either FDD (1800, 2100, 2600 MHz) or time division duplex (TDD) (2300, 2600, 3500 MHz). The applied numerology will typically be based on SCS equal to 30 kHz and the slot length equal to 0.5 ms. The frequencies have generally been assigned to the MNOs; however, special local licenses have been introduced in some countries (e.g., Finland (20 MHz at the 2300 MHz band), France (50 MHz at the 2600 MHz band and Germany (100 MHz at the 3500 MHz band)). Due to the decent amount of available mid-band spectrum, MNOs may have the possibility to reserve some of the carriers for industrial 5G deployments, particularly in rural locations where the capacity needs for the macro network may not be as high as in urban environments. In general, the mid-band will be very suitable for high-quality local industrial 5G networks offered either by the MNOs or the industrial parties. However, in the case of the TDD bands, there may be problems related to latency or capacity, if the mismatch between the applied TDD pattern and the traffic characteristics (downlink (DL) -heavy, uplink (UL) -heavy, or balanced) is too large. Furthermore, since the non-public industrial networks will have to share the band with neighboring public macro networks, special attention has to be paid to the potential co-

[1] Spectrum allocations in Finland [TRAFICOM20] are provided as an example, the corresponding information regarding Sweden and Germany can be found e.g., in [PTS20] and [BNetza20], respectively).
[2] As an example, in Finland the MNOs have been allocated 2x10 MHz (700 MHz), 2x10 MHz (800 MHz) and 2x11.4 MHz (900 MHz) of low-band FDD spectrum.
[3] As an example, in Finland the MNOs have been allocated 2x24.8 MHz (1800 MHz), 2x19.8 MHz (2100 MHz) and 2x20 MHz or 2x25 MHz (2600 MHz) of mid-band FDD spectrum. In addition, each MNO has been allocated 130 MHz (3500 MHz) of mid-band TDD spectrum. Furthermore, one MNO (Elisa) has an additional 50 MHz (2600 MHz) of mid-band TDD spectrum.

existence and inter-network interference problems that may arise, in particular for the TDD bands.

- High-band (above 10 GHz, e.g., 26 GHz) offers a lot of bandwidth, but the coverage areas become quite limited due to the more challenging propagation conditions. Therefore, the high-band will mostly be useful for local indoor or outdoor deployments offering extreme capacity and very low latency. The duplexing mode is TDD, and a high numerology (e.g., with SCS equal to 120 kHz and slot length equal to 0.125 ms) will be applied. Only a few countries have auctioned the 26 GHz band so far, but based on the results and the communicated plans, many countries will split the 26 GHz band into both MNO licenses and local licenses[4]. Due to the high propagation losses, or more specifically, due to the high wall penetration losses [SEF+14][ITU15-P2040], the high-band could be expected to be more easily shared on a geographic base compared to the low- and mid-band. It is also worth highlighting that due to the high numerology, the slot duration is short which in practice means that it becomes more feasible to schedule URLLC users also in the time domain.

Local licenses allow the usage of spectrum in limited small areas, e.g., individual premises. These local licenses can be tied to certain land ownership or land utilization rights and are typically limited to a closed private user group and are not intended for public communication services. However, the terminology and details of regulation for local licensing of the spectrum differ from country to country. The general technical usage conditions for the respective frequency range apply, but some additional measures (e.g., related to maximum emission limits or synchronization) may be needed to facilitate coexistence between these local networks and the neighboring regional and/or national licensed networks. Perhaps the best-known example of local licensing is the 3700-3800 MHz in Germany [BNetza20], see Figure 4, where the national regulator (German Federal Network Agency, Bundesnetzagentur) has decided to reserve 100 MHz of spectrum for local non-public networks. Furthermore, the regulator requires that a local license holder negotiates the maximum allowed emission levels at the edge of the coverage area with its neighboring operators. If an agreement cannot be reached, the maximum allowed limit for the measured field strength at the border of the area covered by the local spectrum license is set to 32 dBμV/m/5 MHz at the height of 3 meters [BNetza19]. For 3700 MHz, this limit translates to a received power of -116.6 dBm/5 MHz, if an isotropic antenna with an antenna gain of 0 dBi is assumed.

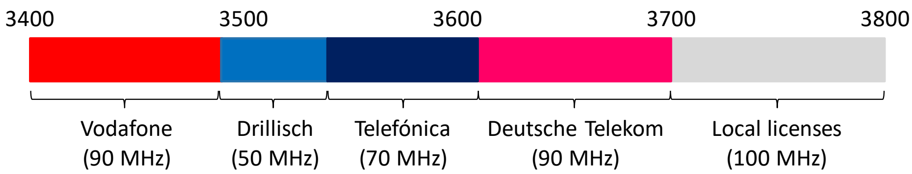

Figure 4. Spectrum allocation at 3400-3800 MHz in Germany.

[4] As an example, in Finland the MNOs have been allocated 800 MHz of high-band TDD spectrum at the 26 GHz band. In addition, the national regulator is planning to reserve up to 850 MHz of high-band TDD spectrum for local licenses at the 26 GHz band.

A summary of the available spectrum options is provided in Figure 5 [NHB+20]. One clear benefit of the sub-6 GHz FDD bands is that they typically support all radio access technologies. This will be a major advantage for scenarios where the same industrial 5G network will be used to offer different kinds of services: URLLC, eMBB, and M-MTC. While the 5G NR-based URLLC addresses the critical communication needs, M-MTC based for example on the Narrowband Internet of Things (NB-IoT) and LTE machine-type communication (LTE-M) is ideal for sensor communication. Finally, LTE or 5G NR-based eMBB provides shop-floor connectivity required for example by smartphones and tablets. Beyond factories, there are also wide-area use cases like smart logistics that will rely on the eMBB and M-MTC services supplied by the MNO networks [SWA+19].

Figure 5 demonstrates also the benefits of leveraging the flexible spectrum assets of the MNOs to deliver optimal results in terms of network performance, diverse use cases, and indoor/outdoor coverage, with or without the availability of local spectrum. In most regions, locally licensed spectrum, when available, is in mid- and/or high-band TDD spectrum. Leveraging the MNOs’ spectrum assets with complementary characteristics can provide major benefits, including improved coverage and availability, support for M-MTC services and low latency. A combination of the local spectrum in the mid-band and the high-band MNO spectrum can potentially boost the system capacity and reduce latency. Furthermore, the MNOs can leverage their public spectrum assets to provide premium eMBB and voice services to the industries. Finally, the 5G inter-band carrier aggregation can also be employed as a powerful tool by dynamically routing traffic through different carriers, achieving the best trade-offs in terms of coverage, reliability, latency, spectral efficiency, and capacity [NHB+20]. Carrier aggregation can be beneficial for both the MNO (to combine different MNO carriers as well as MNO spectrum and local spectrum) and the NPN operator (to combine a mid-band carrier with a high-band carrier).

Frequency band	Key characteristics	Spectrum assets
High-band TDD	High capacity Limited coverage	Local spectrum MNO spectrum
Mid-band TDD	Decent coverage and capacity Latency penalized if DL- or UL-heavy TDD pattern TDD co-existence requirements	Local spectrum MNO spectrum
Low and mid-band FDD	Widest coverage All radio access technologies supported Limited capacity in low FDD bands	MNO spectrum

NR LTE NB-IoT & LTE-M

Figure 5. Summary of the available spectrum options for industrial 5G deployments

2.5.5 Device characteristics

Devices are an important part of the overall deployment as well. Furthermore, in many cases, the selection of the radio network deployment options is dependent on the device characteristics, and vice versa. Among other things, the following device characteristics should be considered, when selecting the most appropriate radio network deployment options:

- Device availability or the existing device ecosystem
- Device type and form factor (external power supply or battery, module, dongle or smartphone)
- Supported frequency bands
- Support for carrier aggregation
- Supported architecture options (non-standalone, standalone)
- Radio frequency (RF) characteristics of the devices (a type of antenna, output power)
- Supported 3GPP release and features

3 Radio network deployment options for an independent standalone non-public network

3.1 Overview

When the non-public network (NPN) is deployed as an independent standalone network, all network functions are located inside the logical perimeter of the defined premises and the non-public network is separate from the public network (PN), as shown in Figure 6 [5GA19-NPN][5GA20-Web][NHB+20]. Furthermore, all control and data traffic stays within the defined premises, which is favorable not only from the confidentiality and security perspective but also from the latency point of view. The only communication path between the public and the non-public networks is via a firewall, e.g., if remote access to the non-public network should be desired via the public network. In all, it is possible to render the availability of the connectivity solution independent from external factors. For example, the connectivity can continue uninterrupted within the factory even when the connectivity to the manufacturing plant is down.

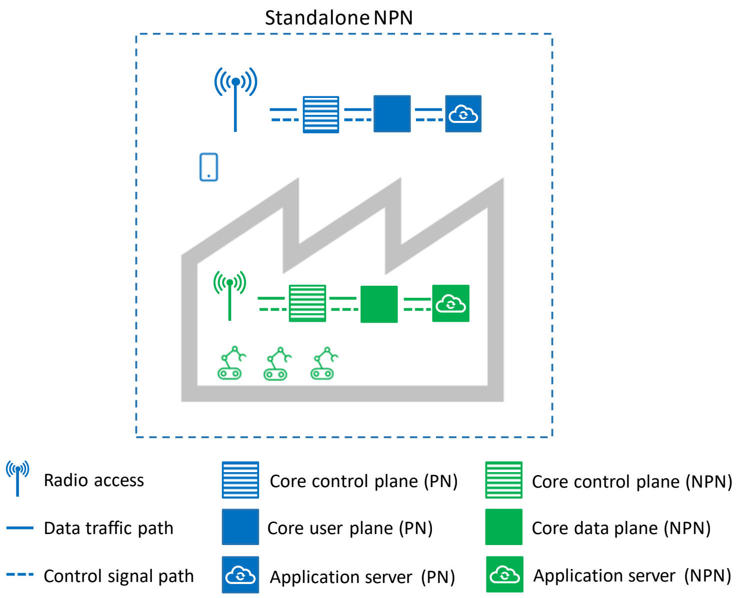

Figure 6. Deployment of the non-public network as an independent standalone non-public network (source: Ericsson)

The independent standalone non-public network is based on 3GPP-defined technologies and is entirely independent with its own NPN identity (ID). NPN devices can access the public network services via a second subscription (dual subscription) if allowed by both the connectivity policy of the NPN administrator and the security requirements of the public network. A more detailed description of the independent standalone non-public network is not part of this report, but can be found e.g., in [5GS20-D52].

The independent standalone non-public network can be provided by the MNO utilizing its national spectrum assets or deployed by the industrial party with the spectrum assets it has acquired through leasing or local licensing. Finally, it is also likely that larger enterprises will prefer the independent standalone deployment over the public network-integrated deployments described in Chapter 4.

3.2 Spectrum options

3.2.1 General

The available spectrum options depend on the stakeholder deploying and operating the independent standalone non-public network. If the industrial party deploys the non-public network, it can get access to licensed spectrum either by leasing spectrum from an MNO (in the countries where leasing is allowed), or by applying a local license from the national regulator (in the countries where local licensing has been implemented). In the countries where neither leasing, nor local licenses are available, the industrial party has to cooperate and negotiate with an MNO to get access to the licensed spectrum. If the MNO provides the non-public network, the MNO can always use its own spectrum assets to offer the required communication services towards the industrial party. Based on an agreement with the industrial party, it may also use the local spectrum licensed to the industrial party, or it may also pool both the MNO spectrum and the local spectrum if allowed by the regulation.

When it comes to independent standalone non-public networks operating on MNO spectrum (either operated by the MNO, or using spectrum leased from an MNO), allocating low-band spectrum for the non-public network may not be a feasible option due to the fact that the low-band spectrum is very valuable for wide-area coverage, and will therefore be widely utilized to offer nationwide voice and eMBB services. Hence, a more reasonable option is to allocate either mid-band, or high-band MNO spectrum for non-public networks. However, an arrangement where a low-band macro carrier is reused within the factory in combination with mid/high-band spectrum to enable 100% service availability and decent capacity could potentially be feasible, despite the possible co-channel inter-network interference between the macro and the factory network, and should be studied further.

The MNO spectrum within the high-band is completely TDD-based, while the mid-band includes both FDD and TDD spectrum. It is unclear how willing an MNO would be to lease part of its FDD spectrum to the industrial party, even if that would be allowed by the regulation. On the other hand, if the MNO provides the non-public network, it can decide to (locally) reserve part of its FDD spectrum, e.g., the 1800 MHz, for non-public use, in particular within areas with less demanding capacity needs within the macro network. Since the available bandwidth per a single FDD carrier is rather limited, the corresponding network capacity becomes quite limited as well. However, a combination of multiple FDD carriers via e.g., carrier aggregation can provide larger capacity. To achieve an even higher capacity for non-public deployments, also the MNO would need to rely on TDD spectrum on mid- or high band, or a combination of both the FDD spectrum and the TDD spectrum.

Local licensing has so far been implemented only in a few countries [NHB+20]. The allocated local spectrum has been either in the mid-band (2300 MHz, or 3.5-4 GHz) or in the high-band (26 GHz). It is also worth highlighting that the local licenses have typically been implemented within the TDD bands, which means that the opportunities for an industrial party to deploy a non-public network in FDD spectrum are very limited.

3.2.2 Spectrum options to support mixed IIoT services

It is highly important for an industrial 5G communication system to support mixed services, e.g., critical machine-type communications (C-MTC) requiring URLLC communication services, industrial eMBB communication services, and industrial wireless sensor services via, e.g., NB-IoT or LTE-M for M-MTC. Since it is expected to be quite straightforward to combine eMBB traffic with either URLLC or M-MTC, as demonstrated by the evaluation results in Section 5.2.3, the discussion here focuses on the more challenging task to combine URLLC traffic with M-MTC.

On a high level, there are two types of spectrum-related alternatives available to deploy an industrial 5G communication system with mixed services: a single-band IIoT system and a multi-band IIoT system. In the case of a single-band IIoT system all services are offered on the same frequency band. If the IIoT system is FDD-based (i.e., operating on low- or mid-band MNO spectrum) the services can co-exist without any major limitations. The drawback of the FDD-based deployment is that the total bandwidth of a single FDD carrier, and thus also the network capacity shared between the different services is typically limited. The more advanced IIoT systems with higher capacity and/or quality-of-service (QoS) requirements would therefore have to rely on combining multiple carriers, either FDD or TDD.

If the single-band IIoT deployment is TDD-based, one of the key aspects for ensuring the co-existence of different services is the synchronization of the TDD pattern both in terms of the frame start and the transmission direction. An important evolution of the NR standard is the introduction of a new device type with *reduced capability* (NR RedCap) that is being specified in 3GPP NR Release 17 and can operate in all NR-supported frequency bands [3GPP20-RP201677][3GPP20-38875]. NR RedCap enables devices with reduced complexity and cost, as well as extended battery lifetime for infrequent transmission of small amounts of data [ZBN+20]. NR RedCap devices can operate in all NR TDD configurations and thus can share the same carrier with NR URLLC devices. NR RedCap is intended for devices such as industrial wireless sensors and wearables, and it will provide higher achievable data rates and better latency than the M-MTC devices supported by LTE-M and NB-IoT. LTE-M and NB-IoT can support the other M-MTC services. However, if the TDD configurations of NR and NB-IoT/LTE-M are not synchronized, cross-link interference makes the deployment of these two systems on a single TDD carrier impossible. Thus, the combined support for the LTE-based M-MTC and the NR-based URLLC on a single TDD carrier is only possible with the same LTE-aligned TDD pattern for both services. In practice, this results in a limited URLLC performance with respect to both the latency and the capacity. A challenge with the current TDD-based M-MTC technology is a lack of a device ecosystem: there are no known end-user devices or chipsets for NB-IoT or LTE-M in the frequency region above 3 GHz, even if this is in principle supported by the standard. One remedy for the latter can be to rely on LTE devices that are commonly available for those bands. LTE devices of UE category 1 or 4 are also devices with limited complexity that have somewhat higher capabilities than LTE-M or NB-IoT in terms of data rate; by applying the same power saving features[5] significantly extended battery lifetimes are also achievable. Still, the problem pertains that a common LTE-compatible TDD configuration is required for the carrier, which limits the achievable URLLC performance. If URLLC performance with a

[5] Power saving features used by LTE-M and NB-IoT are the power saving mode (PSM) and extended discontinuous reception (eDRX) see [LSE+19]; those are also available for LTE devices in general.

latency bound in the single-digit millisecond range is required, an LTE-based M-MTC solution cannot be used on the same TDD carrier.

A multi-band IIoT system can be realized in two different ways: by complementing the existing NR IIoT carrier with an M-MTC radio technology utilizing unlicensed spectrum, or by using multiple licensed spectrum carriers. In the latter case, one carrier could be used for URLLC and eMBB services, while the other carrier is used for M-MTC and eMBB services (and possibly URLLC services with higher acceptable latency bound). The following options describe how an industrial party can make use of multiple spectrum carriers:

1. Two mid-band TDD carriers are obtained via local licensing or leasing. As discussed, M-MTC on mid-band TDD is, in general, possible with some restrictions on the available device ecosystem. Furthermore, the industrial party might need to reach an agreement with an MNO to lease the second TDD carrier.
2. One mid-band TDD carrier and one high-band TDD carrier. The industrial party could get access to both carriers either via local licensing and/or leasing. However, similar to the first option, some limitation remains with regard to the available M-MTC device ecosystem for mid-band TDD.
3. One FDD carrier in low- or mid-band and one TDD carrier in mid- or high-band. This option would be most feasible from the technology and the device availability point of view. However, the industrial party would need to reach an agreement with an MNO to be able to lease the FDD carrier, which could be difficult.

The last option can also be addressed by an industrial party with a single mid- or high-band TDD carrier in another way: the industrial party can co-operate with an MNO for providing M-MTC services via the MNO network. However, in that case the MNO may need to deploy additional sites, or share non-public network gNodeBs, unless it can offer the required coverage and service availability with the already existing macro sites. Furthermore, the M-MTC traffic can be terminated locally in an MNO-provided local core network or be routed through the MNO network and reconnected to the industrial 5G network via the central operator core network. It should be noted that with this kind of "MNO-provided M-MTC component," the IIoT network is not anymore independent, but will be at least partially integrated with the public network, as addressed later in Chapter 4.

As a summary, the MNOs are in a good position to support local standalone 5G IIoT solutions for all service categories due to their broader licensed spectrum assets. They can easily support two different mid-band carriers: one configured for M-MTC (LTE-M/NB-IoT/RedCap) and the other configured for eMBB and URLLC. The industrial parties relying solely on the local spectrum on mid- and/or high-band will face challenges due to requirements related to synchronized TDD patterns and due to the lack of the M-MTC device ecosystem for the mid-band TDD. To be able to efficiently support all IIoT services, they would need to co-operate with an MNO to lease an additional spectrum or request the MNO to provide the required M-MTC services to them. Alternatively, they could utilize unlicensed technologies for M-MTC services. A further option for connecting industrial sensors, namely NR RedCap, will become available with 3GPP NR Release 17.

3.3 Co-existence scenarios

3.3.1 General

In general, when evaluating the co-existence performance between two networks operating within the same geographical area, the following four different inter-network interference scenarios should be taken into account (see also the description in Figure 7):

1. Downlink-to-downlink ("near-far") interference, where the downlink transmissions from an aggressor gNodeB are interfering the downlink reception at a victim UE. This inter-network interference scenario is valid for FDD, synchronized TDD (sTDD) and unsynchronized TDD (uTDD) deployments.
2. Uplink-to-uplink ("near-far") interference, where the uplink transmissions from an aggressor UE are interfering the uplink reception at a victim gNodeB. This inter-network interference scenario is valid for FDD, sTDD, and uTDD deployments.
3. Downlink-to-uplink ("cross-link") interference, where the downlink transmissions from an aggressor gNodeB are interfering the uplink reception at a victim gNodeB. This inter-network interference scenario is valid for uTDD deployments.
4. Uplink-to-downlink ("cross-link") interference, where the uplink transmissions from an aggressor UE are interfering the downlink reception at a victim UE. This inter-network interference scenario is valid for uTDD deployments.

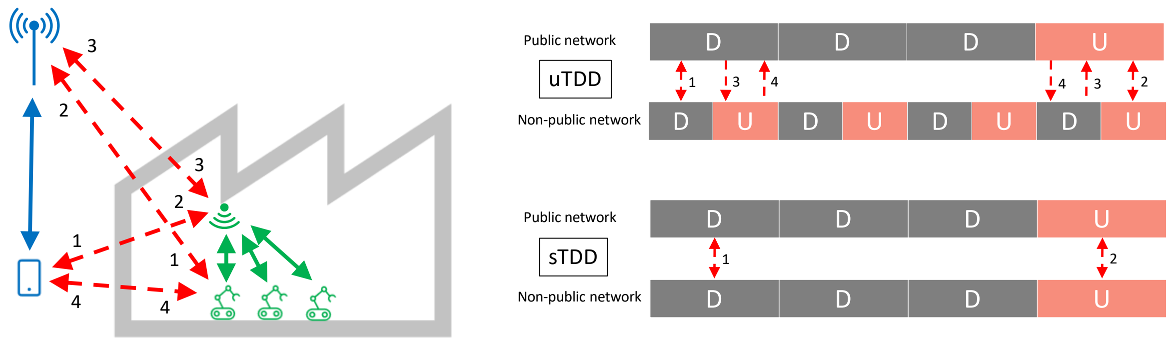

Figure 7. Description of the inter-network interference scenarios between a public macro network and a non-public factory network (source: Ericsson)

A synchronized TDD deployment refers to a scenario, where the networks apply the same TDD downlink-uplink pattern with aligned slot borders. Hence, there will never be time instants with colliding transmissions between the uplink and the downlink. For all the other cases, the TDD deployment is unsynchronized: there will be at least some time instants with colliding transmissions between the uplink and the downlink.

In the case of an independent standalone non-public network, two different co-existence scenarios have been considered within 5G-SMART:

- Co-existence between a public network and a non-public network.
- Co-existence between two neighboring non-public networks.

Depending on the frequency allocation, the co-existence scenario can be classified either as a co-channel deployment (i.e., both networks are operating on the same frequency channel), adjacent channel deployment (i.e., networks are operating on adjacent frequency channels), or a deployment on isolated frequency channels (i.e., networks are operating on isolated frequency bands). In the case of an adjacent channel deployment, part of the isolation between the networks is offered by the adjacent channel attenuation both in the transmitter (Adjacent Channel Leakage power Ratio, ACLR) and in the receiver (Adjacent Channel Selectivity, ACS). As an example, looking at the minimum requirements for the 5G NR (low- and mid-band) specified by the 3GPP in [3GPP20-38101] and [3GPP20-38104], the following values can be found: $ACLR_{gNodeB}$ = 45 dB, $ACLR_{UE}$ = 30 dB, ACS_{gNodeB} = 40 dB and ACS_{UE} = 33 dB. The minimum required Adjacent Channel Interference Ratio (ACIR) values for each of the four inter-network interference scenarios can be derived by combining the corresponding ACLR and ACS values: $ACIR_{DL\text{-}DL}$ = 32.7 dB, $ACIR_{UL\text{-}UL}$ = 29.6 dB, $ACIR_{DL\text{-}UL}$ = 38.8 dB, and $ACIR_{UL\text{-}DL}$ = 28.2 dB. In the case of a co-channel deployment, ACIR = 0 dB, while in the case of an isolated frequency deployment the ACIR can be assumed to be sufficiently large, e.g., due to additional filters, so that the inter-network interference can be ignored.

In addition to the frequency, the isolation between the networks can be secured, for example, with the help of separation distance, exclusion zone (see Figure 8), wall penetration losses and antennas. Furthermore, more dynamic solutions based on various radio resource management (RRM) mechanisms can be applied as well. However, since the networks are assumed to be isolated, tight coordination mechanisms operating on the slot level (e.g., coordinated scheduling) will not be feasible.

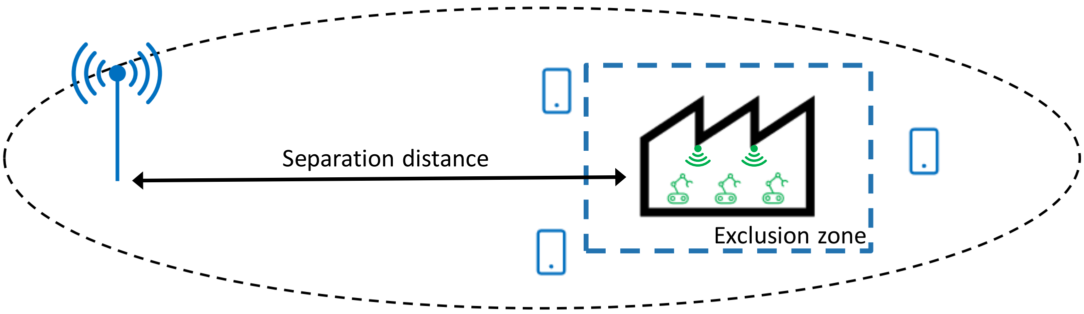

Figure 8. Description of a separation distance and an exclusion zone (source: Ericsson)

3.3.2 Co-existence between a public and a non-public network

This co-existence scenario is an example of an overlapping coverage between the networks: it is likely that the public network will have coverage and will potentially also serve users inside the coverage area of the non-public network. The public network can be either a macro network, a small cell outdoor network, a small cell indoor network, or a heterogeneous deployment consisting of both macro and small cells. Similarly, the non-public network can be either a small cell outdoor network, a small cell indoor network, or a mixture of both.

This co-existence scenario can be a co-channel deployment if the same MNO operates both networks. It can be an adjacent channel deployment if the networks are operated by the same MNO, different MNOs, or by an MNO (public network) and an NPN operator (non-public network). Finally, it can also be a deployment on isolated frequency channels if the networks are operated by the same MNO, different MNOs, or by an MNO and an NPN operator.

An example of this co-existence scenario, namely a co-existence scenario between a public macro and a non-public factory network, is shown in Figure 7. The macro network is assumed to be offering wide-area eMBB services, while the factory network is assumed to be offering local URLLC services. Since the service requirements, both in terms of latency, reliability and capacity are very different for eMBB and URLLC, also the service-optimized TDD patterns will look quite different: eMBB is often downlink-heavy and does not have strict latency and reliability requirements, which leads to downlink-heavy TDD patterns (e.g., DDSU, or even DDDSU, where the “S” slot has a configuration of 10:2:2 between the downlink symbols, guard period and the uplink symbols). At the same time, the URLLC service can be more balanced between the downlink and the uplink, or even uplink-heavy, which would be the case e.g., for remote-controlled vehicles where a high-quality video stream is transmitted in the uplink while only some low bitrate control data is transmitted in the downlink. Furthermore, the URLLC service will typically have strict latency and reliability requirements, resulting in TDD patterns with faster switches between the downlink and the uplink. In the example shown in Figure 7, a balanced DUDU pattern, based on sub-slots with 7 symbols, is assumed to be the optimum for the URLLC service provided by the factory network.

The service-optimized TDD patterns would typically lead to an unsynchronized TDD deployment between the networks. Unfortunately, that could also result in severe inter-network interference problems with contributions from both the near-far and the cross-link interference scenarios. The problems related to the cross-link interference can be avoided by synchronizing the TDD networks (both networks apply the same TDD pattern with aligned slot borders), but that will not solve the problems related to the near-far interference, and what is more, synchronized TDD may lead to problems related to either the eMBB capacity (eMBB follows the URLLC-optimized pattern) or the URLLC uplink latency and capacity (URLLC follows the eMBB-optimized pattern).

If both networks apply FDD as the duplexing method, there will not be any cross-link interference between the networks, but the same near-far problems as in the case of TDD will remain. However, as already discussed in Section 3.2, the use of FDD will mainly be an option for the MNO-provided non-public networks.

Figure 7 represents a scenario where the factory does not contain any UEs served by the overlaid public network. In this particular case, the inter-network interference is always attenuated by the wall penetration loss, which makes it quite straightforward to isolate the networks. Furthermore, the impact of the inter-network interference on the performance of the public network can be assumed to be quite limited. The co-existence scenario becomes much more challenging if some active (co- or adjacent channel) UEs served by the overlaid public network are located inside the factory, as illustrated in Figure 9. This is due to a few main differences compared to the scenario in Figure 7. First and most importantly, the wall penetration loss is no longer helping to mitigate the inter-network interference from the public network UEs towards both the factory gNodeBs (“near-far interference”) and UEs ("cross-link interference"). In fact, the situation becomes even worse if the wall penetration loss is increased, since as a result the transmission powers will be increased for the public network UEs inside the factory. Secondly, the increased path losses will contribute to an increased average resource utilization within the public network cells, resulting further in a higher level of inter-network interference from the public network gNodeBs towards both the factory UEs and gNodeBs, which otherwise would not be affected by moving some of the public network UEs inside the factory. Finally, the public network UEs and the factory UEs get closer to each other which increases the level of the cross-link interference as well. From the public network point of view, the users entering the factory are expected to experience greatly reduced throughputs. This is mostly due to the impact of the wall penetration loss, which reduces the received signal power levels from the serving public network gNodeBs and at the same time increases the level of the received inter-network interference from the factory gNodeBs ("near-far interference") and UEs ("cross-link interference").

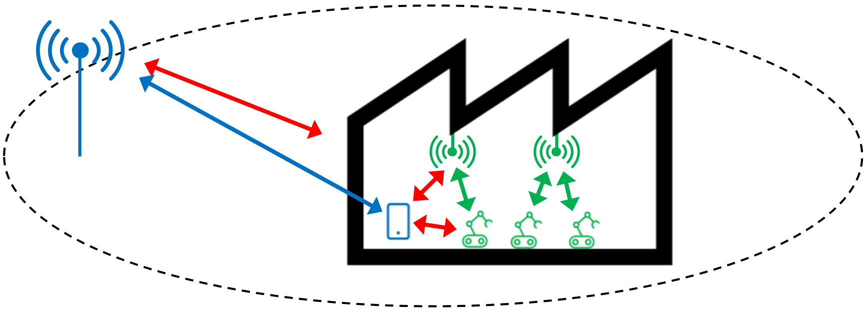

Figure 9. Description of a co-existence scenario between a public macro network and a non-public factory network, when the factory contains users served by the overlaid public network (source: Ericsson)

The performance of this kind of co-existence scenario between a public macro TDD network and an independent standalone non-public factory TDD network is evaluated in Section 5.3.1. As a summary, the following conclusions can be highlighted:

- Frequency. An adjacent channel deployment between uncoordinated networks is a feasible option, if the factory does not contain any adjacent channel macro users. However, the adjacent channel attenuation will not be able to provide a sufficiently high level of isolation, if the adjacent channel macro users are allowed to enter the factory. In that kind of a scenario, the use of an isolated frequency, or some other means to mitigate the inter-network interference, is recommended. Finally, if the non-public network is operating on an MNO spectrum, the question about the feasibility of reusing the NPN carrier in the overlaid macro cells can be raised. There, the general recommendation is to avoid the co-channel deployment between the factory network and the overlaid macro cells. However, the feasibility of reusing a low-band FDD carrier as a complement to mid- or high-band TDD carrier to achieve 100% service coverage and decent capacity should be evaluated further. That evaluation should also consider if the normal inter-frequency and inter-radio access technology (RAT) handover mechanisms would be able to mitigate the resulting co-channel interference between the macro and the factory network, in particular the near-far uplink interference towards the factory network, or if some other means to mitigate the interference would be needed.

- Synchronization. Synchronization will avoid the cross-link interference, but it will not solve the problems related to the near-far interference. Furthermore, it can lead to additional problems related to strict latency requirements and network capacity, typically within the factory network if it has to be aligned with the eMBB-optimized TDD pattern of the macro network.
- Slot blanking. A potential solution to enable different TDD patterns while avoiding the cross-link interference of unaligned downlink and uplink transmissions is to introduce a "slot blanking" mechanism into the macro network, in which case the macro network could be silent during some of the downlink slots, allowing the factory network to increase the amount of uplink slots. This will lead to a reduced maximum downlink capacity within the macro cells where such mechanism is applied. Slot blanking would always require a coordination agreement between the NPN operator and the MNO.
- Shared RAN or a separate public factory network. Near-far interference can be avoided if the macro UEs located inside the factory can connect to the factory gNodeBs either with the help of a shared RAN (as discussed in more detail in Chapter 4) or by deploying a separate public factory network. However, in that case the public and the non-public factory networks would have to be synchronized, or some other means, e.g., slot blanking, should be applied to avoid cross-link interference between the networks.
- Separation distance. If the factory is located close to a macro site, the (co- or adjacent channel) inter-network interference caused by the high-power macro gNodeB will be significant for both the downlink-to-downlink interference towards the factory UEs and the downlink-to-uplink interference towards the factory gNodeBs [CHT19]. With a larger separation distance, the interference from the macro gNodeB will become lower, but at the same time the interference from the macro UEs will become higher. One special form of the separation distance is the so-called "virtual fencing", in which case the macro network avoids the use of the interfering carrier within the closest macro cells surrounding the factory, based on an agreement between the NPN operator and the MNO. From the factory network point of view this would then look like a deployment on an isolated frequency, while from the macro network point of view it would lead to a reduced maximum network capacity within the affected macro cells.
- Wall penetration loss. An effective way to reduce the leakage from the macro network towards the factory UEs and gNodeBs, or vice versa, is to deploy the non-public factory networks within buildings with high wall penetration losses, e.g., as a result of metallic wall structures. However, this will become a downside if the desire is to have public macro network coverage inside the factory since the transmission powers of the macro UEs located inside the factory will become higher, resulting in a higher level of uplink near-far and cross-link interference towards the factory gNodeBs and UEs. In general, the required wall penetration loss value to secure the desired performance of both the macro and the factory network will be highly scenario-dependent, and will depend on the characteristics of both the aggressor and the victim network.

- Exclusion zone. The use of a physical exclusion zone around the factory, e.g., with the help of a fence, will reduce the uplink interference from the macro UEs towards the factory gNodeBs and UEs, as well as the downlink interference from the factory gNodeBs and UEs towards the macro UEs. However, in many cases it may not be possible, or even desirable, to forbid the use of public mobiles within the factory area. Furthermore, the factory wall penetration loss will often provide a sufficient level of protection even without any additional exclusion zone.
- Transmission power. Securing a sufficient dominance of the factory network over the macro network, e.g., by increasing the downlink and uplink transmission powers, is important since it will reduce the required level of isolation between the macro network and the factory network. By doing so, the external interference will have only a minor impact on the signal-to-interference-plus-noise ratios (SINR) of the factory users, and hence, on the factory network performance. In downlink, the 3GPP specifications define an upper limit for the gNodeB transmission power as described in Section 2.5.1. In uplink, the transmitted UE powers can be increased by making the power control algorithm more aggressive. However, that will help to combat the external interference only as long as the UEs are not transmitting at their maximum allowed power, because after that, the uplink performance will become limited by the inter-cell interference within the factory.
- Network densification. The received signal power levels can be improved also by making the factory cells smaller, i.e., by densifying the factory network. Due to the reduced path losses, the received downlink signal powers will be directly improved. In the uplink, the network densification makes it possible to apply an even more aggressive power control before reaching the maximum allowed UE transmission power, and in that way to increase the received uplink signal powers. The unfortunate downside of network densification is that the (intra-network) inter-cell interference will typically increase faster than the received signal power, at least when the gNodeBs do not have directional antennas or do not utilize beamforming, reducing the SINR for the users that are not limited by the noise and the inter-network interference. In the end, the overall URLLC network performance can in fact suffer as a result of the network densification, as demonstrated by the evaluation results in Section 5.2.2. If the network is coverage-limited, it is often more beneficial to improve the performance by adding more antennas into the existing cell (e.g., by having more antennas in an active distributed antenna system) instead of adding new gNodeBs with omnidirectional antennas.
- Antennas. The use of directional antennas or beamforming will be an efficient way to reduce the level of both the inter-cell and the inter-network interference, as well as to improve the level of the received signal power compared to omnidirectional or isotropic antennas. Furthermore, as demonstrated by the evaluation results in Section 5.2.2, the use of directional antennas or beamforming to suppress the inter-cell interference will be highly beneficial for network densification.

As a summary, in the case of a scenario where the factory does not contain any public macro users operating either on co- or adjacent channel, the main co-existence problems are related to the downlink interference from the high-power macro gNodeBs. An uncoordinated deployment (without synchronized TDD or some other form of coordination) between the public macro and the non-public factory network is possible if the isolation between the networks is high enough, either in form of a separation distance, wall penetration loss or a frequency separation. Hence, in this case it is possible for the NPN operator to resolve the co-existence problems without the help of the MNO, in particular if the NPN operator has access to an isolated frequency not used within the overlying macro cells.

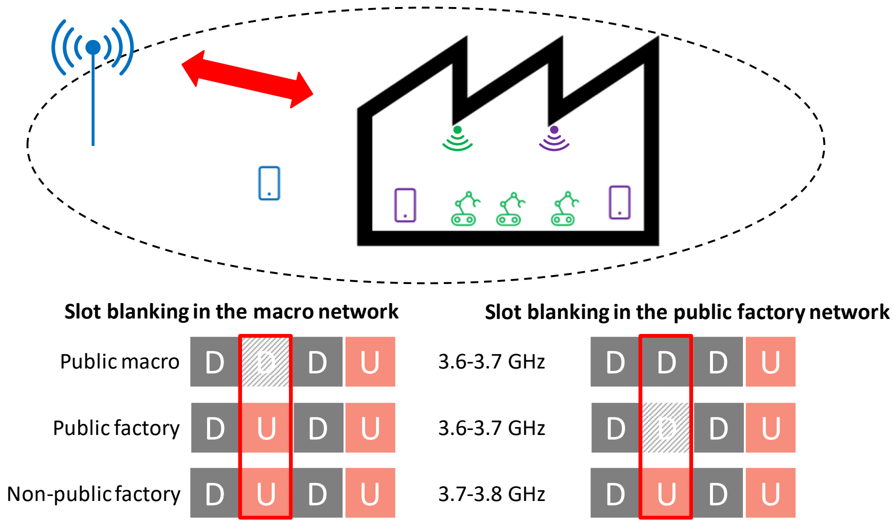

Figure 10. Adjacent channel co-existence scenario between the public macro and the non-public factory network when the factory contains a separate public factory network (source: Ericsson)

If it is desired to allow co- or adjacent channel public users to enter the factory, the co-existence situation becomes much more difficult. A co-channel deployment will lead to serious near-far problems between the factory network and the visiting macro users and is therefore not recommended. An adjacent channel deployment will reduce the level of the inter-network interference, but it will not always be sufficient. A high isolation in form of the wall penetration loss will not be helpful in this situation, since it will increase the uplink interference from the visiting macro users towards the factory network. Furthermore, applying a synchronized TDD between the networks will avoid the cross-link interference, but it will not solve the problems related to the near-far interference. One efficient way to resolve the near-far problems is to allow the visiting macro users to be served by the factory gNodeBs, either with the help of shared RAN (discussed in more detail in Chapter 4), or by deploying a separate public factory network. However, in that case the factory networks have to be synchronized, or some other mechanisms (e.g., slot blanking) should be utilized to avoid the cross-link interference between the public and the non-public factory gNodeBs and UEs. In the scenario depicted in Figure 10, the goal of applying slot blanking is to avoid both the co-channel cross-link interference between the public macro network and the public factory network, and at the same time to avoid the adjacent channel cross-link interference between the public and non-public factory networks. If slot blanking is applied in the macro network, the factory networks could be synchronized to avoid the cross-link interference. This will reduce the amount of available downlink radio resources within the affected macro cells and within the public factory network, but at the same time the macro network will benefit from the offloading of the “expensive” factory users, reducing the negative impact of the slot blanking on the overall macro network capacity. An alternative option is to blank the interfering downlink slot within the public factory network. This will reduce the maximum downlink capacity, but since the utilization of the public factory network would typically be low, it would most likely not have any visible impact on the actual network performance, keeping also in mind that at the same time the performance of the macro network would benefit from the traffic offloading. The remaining adjacent channel cross-link interference between the macro and the non-public factory network would be quite straightforward to manage, since the factory would not contain any public macro users anymore. Finally, if an uncoordinated deployment between the public macro network and the non-public factory network is desired instead, the recommendation is to deploy the factory network on an isolated frequency, e.g., on the high-band. Hence, it may be difficult for the NPN operator to resolve the co-existence problems without the help of the MNO, unless it has access to an isolated frequency not used within the overlying macro cells.

Until now, we have assumed a non-public network to be deployed indoors, e.g., in a factory. The co-existence scenario becomes considerably more difficult if the non-public network contains outdoor gNodeBs, for example in the case of an overlay deployment between a public macro network and a non-public small cell outdoor network, as shown in Figure 11. The performance of such a scenario has not been evaluated in detail in 5G-SMART, but the initial expectation is that the co-channel deployment would not be feasible, due to the very high level of both near-far and cross-link (with unsynchronized TDD) interference. Furthermore, an unsynchronized adjacent channel deployment would not be feasible either, as a very large separation distance would still be needed to limit the cross-link interference between the outdoor gNodeBs. However, a synchronized adjacent channel deployment could potentially be feasible. Again, a deployment with unsynchronized TDD would in practice require the use of an isolated frequency (or frequencies) within the non-public network.

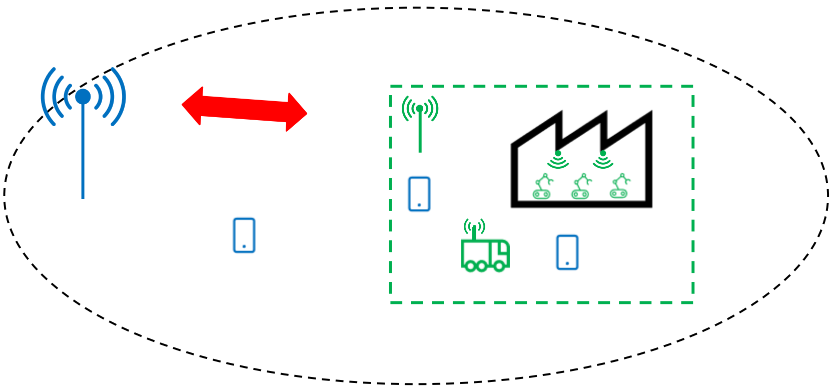

Figure 11. Co-existence between a public macro network and a non-public small cell outdoor network (source: Ericsson)

3.3.3 Co-existence between neighboring non-public networks

This co-existence scenario represents a non-overlapping coverage between the networks: the (local) non-public networks will typically be deployed on neighboring properties, and the users are not expected to be moving between them. As illustrated in Figure 12, the non-public network can be either a small cell outdoor network, a small cell indoor network, or a mixture of both.

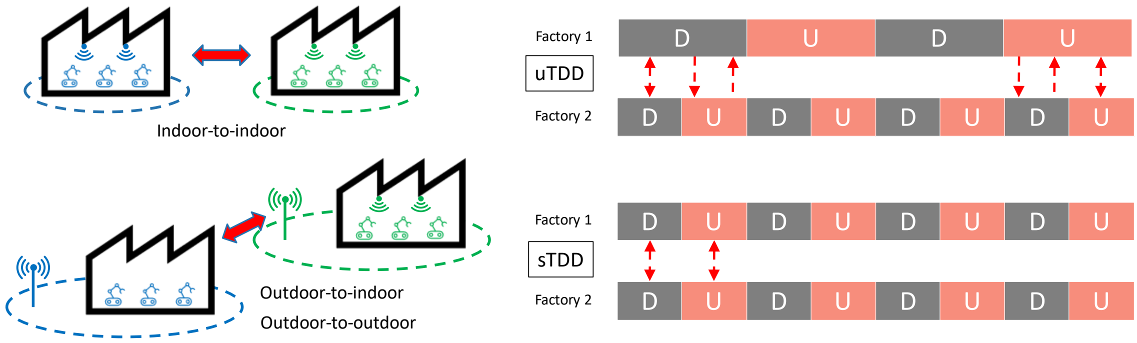

Figure 12. Co-existence between neighboring non-public networks (source: Ericsson)

This co-existence scenario can be a co-channel deployment if both networks are operated by the same MNO, or by different NPN operators. It can be an adjacent channel deployment or a deployment on isolated frequency channels if the networks are operated by the same MNO, different MNOs, an MNO and an NPN operator, or by different NPN operators.

It is quite straightforward to assume that the neighboring non-public networks would desire to apply a TDD pattern that has the best match to their respective service and capacity needs: the most suitable TDD pattern could in practice vary from the more eMBB optimized downlink-heavy pattern (e.g., DDDU), to a more balanced pattern (e.g., DUDU), or even to an uplink-heavy pattern (e.g., UUUD). Furthermore, the non-public networks may be configured with sub-slots instead of full slots for latency reasons. In the example of Figure 12, both factories apply a balanced DUDU pattern. However, since the TDD pattern in factory 1 is based on full slots (14 symbols) and in factory 2 on half-slots (7 symbols), the corresponding co-existence scenario is "unsynchronized" (uTDD) including all the four different inter-network interference scenarios as described in Section 3.3.1. For this particular scenario with balanced TDD patterns, a synchronized TDD deployment based on half-slots would most likely be a simple solution, but the situation would look considerably different if the required services would have large differences between the neighboring factories. For such scenarios, the question of how to make an unsynchronized TDD deployment feasible becomes highly important.

Depending on the characteristics of the neighboring non-public networks, there are three different high-level interference scenarios that could be considered: indoor-to-indoor interference, outdoor-to-indoor interference and outdoor-to-outdoor interference between the networks. When the gNodeBs and the UEs are located inside the factory buildings (indoor-to-indoor interference), the networks are isolated by two outer walls and the separation distance between the factories. As a result, the near-far interference will typically not be any bigger issue, and can be easily handled with the help of a proper network planning securing a sufficient dominance of the serving factory network within the desired service area. However, as demonstrated by the evaluation results in Section 5.3.2, the cross-link interference between the gNodeBs can cause problems in mid-band requiring special attention (e.g., with respect to the type, locations and orientations of gNodeB antennas). At high-band, the wall penetration losses are expected to be so high [SEF+14][ITU15-P2040], that an uncoordinated deployment between the networks will typically be feasible.

The impact of the outdoor-to-indoor and outdoor-to-outdoor interference has not been evaluated so far within the 5G-SMART project. However, since the isolation between the networks will be lower compared to the indoor-to-indoor interference, an initial estimate is that there will be larger challenges related to the co-existence. It can be very challenging to assume unsynchronized TDD for the outdoor-to-outdoor case, even for the high-band, since that would in practice require large separation distances between the networks and a careful planning of the locations, antennas and orientations of the local small cell outdoor sites. In more detail, it is very likely that an unsynchronized co-channel deployment will not be feasible. At the same time, a synchronized adjacent channel deployment is estimated to be feasible. However, it is currently unclear how large separation distances would be needed between two non-overlapping synchronized co-channel or unsynchronized adjacent channel outdoor networks.

As a summary, applying a synchronized TDD pattern will be a quite effective way to mitigate the inter-network interference in this particular co-existence scenario with non-overlapping communication service areas. However, if service-optimized (and NPN-specific) TDD patterns are preferred instead, some other means, such as a careful planning of the radio network deployments and agreeing on appropriate emission limits, should be applied to control the level of the inter-network interference. The national regulator could also define an emission limit as part of the local license conditions, to be valid for the situations when the neighboring operators cannot reach an agreement by themselves.

3.4 Summary

The available spectrum options for an independent standalone non-public network depend on whether the non-public network is provided by the MNO utilizing its national spectrum assets or deployed by the industrial party with the spectrum assets it has acquired through leasing or local licensing. In general, the MNOs are in a good position to support independent standalone non-public networks for all service categories due to their broader licensed spectrum assets. They have the possibility to utilize FDD spectrum and can easily support deployments with multiple carriers. The industrial parties relying solely on the local spectrum on mid- and/or high-band will face challenges due to requirements related to synchronized TDD patterns and due to the lack of the M-MTC device ecosystem for the mid-band TDD. To be able to efficiently support all IIoT services, they would need to co-operate with an MNO to lease an additional spectrum or request the MNO to provide the required M-MTC services to them. Alternatively, they could utilize unlicensed technologies for M-MTC services. A further option for connecting industrial sensors, namely NR RedCap, will become available with 3GPP NR Release 17.

If the factory does not contain any visiting macro users operating either on the co- or adjacent channel, the main co-existence problems between a public macro network and a non-public factory network are related to the downlink interference from the high-power macro gNodeBs. An uncoordinated deployment (without synchronized TDD or some other form of coordination) is possible if the isolation between the networks is high enough, either in form of a separation distance, wall penetration loss or a frequency separation. Hence, in this case it is possible for the NPN operator to resolve the co-existence problems without the help of the MNO, in particular if the NPN operator has access to an isolated frequency not used within the overlying macro cells.

If it is desired to allow co- or adjacent channel public users to enter the factory, the co-existence situation becomes much more difficult. A co-channel deployment will lead to serious near-far problems between the factory network and the visiting macro users and is therefore not recommended. An adjacent channel deployment will reduce the level of the inter-network interference, but it will not always be sufficient. A high isolation in form of the wall penetration loss will not be helpful in this situation, since it will increase the uplink interference from the visiting macro users towards the factory network. Furthermore, applying a synchronized TDD between the networks will avoid the cross-link interference, but it will not solve the problems related to the near-far interference. One efficient way to resolve the near-far problems is to allow the visiting macro users to be served by the factory gNodeBs, either with the help of RAN sharing, or by deploying a separate public factory network. However, in that case the factory networks have to be synchronized, or some other mechanisms (e.g., slot blanking) should be utilized to avoid the cross-link interference between the public and the non-public factory gNodeBs and UEs. If an uncoordinated deployment between the public macro network and the non-public factory network is desired instead, the recommendation is to deploy the factory network on an isolated frequency. Hence, it may be difficult for the NPN operator to resolve the co-existence problems without the help of the MNO, unless it has access to an isolated frequency not used within the overlying macro cells.

Finally, when it comes to the co-existence between neighboring non-public factory networks with non-overlapping communication service areas, applying a synchronized TDD pattern will be a quite effective way to mitigate the inter-network interference. However, if service-optimized (and NPN-specific) TDD patterns are preferred instead, some other means, such as a careful planning of the radio network deployments and agreeing on appropriate emission limits, should be applied to control the level of the inter-network interference. The national regulator could also define an emission limit as part of the local license conditions, to be valid for the situations when the neighboring operators cannot reach an agreement by themselves.

4 Radio network deployment options when the non-public network is deployed in conjunction with public networks

4.1 Overview

Deploying the non-public network in conjunction with a public network allows reuse of network infrastructure, efficient utilization of spectrum and seamless mobility. The network infrastructure can be deployed inside or outside the enterprise's premises in part or in its entirety and can be shared between the public and non-public users. There are three ways of realizing this [5GA19-NPN][5GA20-Web][NHB+20]:

- Shared RAN
- Shared RAN and control plane
- Non-public network hosted by a public network

The two last alternatives can also be referred to as public network integrated non-public networks (PNI-NPN) in 3GPP terminology [3GPP20-23501]. While the independent standalone deployment described in the previous chapter is likely to appeal to large enterprises, smaller enterprises will likely prefer an MNO-provided non-public network ("NPN as a service", NPN hosted by the public network) instead.

When the same network infrastructure (i.e., RAN) is shared between the public and the non-public services, the topic of RAN slicing becomes important. In general, RAN slicing will define how the radio resources within the shared gNodeBs will be dynamically allocated between the different services with different QoS and capacity needs. However, it should be noted that the details of the different RAN slicing methods and aspects will not be discussed in this report. Furthermore, it should be highlighted that a more detailed discussion of the different NPN architecture options listed above is omitted from this report and can be found for example in [5GS20-D52].

4.1.1 Shared RAN

Based on a RAN sharing agreement between the NPN operator and the MNO, the public and non-public users share the radio access network (i.e., gNodeBs) while the rest of the network components are kept segregated, see Figure 13. In other words, both public and non-public users connect to their separate networks only, by using the same gNodeB. Similar to the independent standalone non-public network, all non-public user data and control traffic stays within the enterprise's logical premise. The non-public network is based on 3GPP-defined technologies and has its own NPN ID. The non-public users can access the public services with the help of a dual subscription.

The RAN sharing can be realized with the help of a solution based on either a multi-operator core network (MOCN) or multi-operator RAN (MORAN). In case of MOCN, the MNO and the NPN operator are sharing both the RAN (gNodeB) and the spectrum resources, while in case of MORAN the operators are sharing the gNodeB, but they have their own non-shared spectrum resources

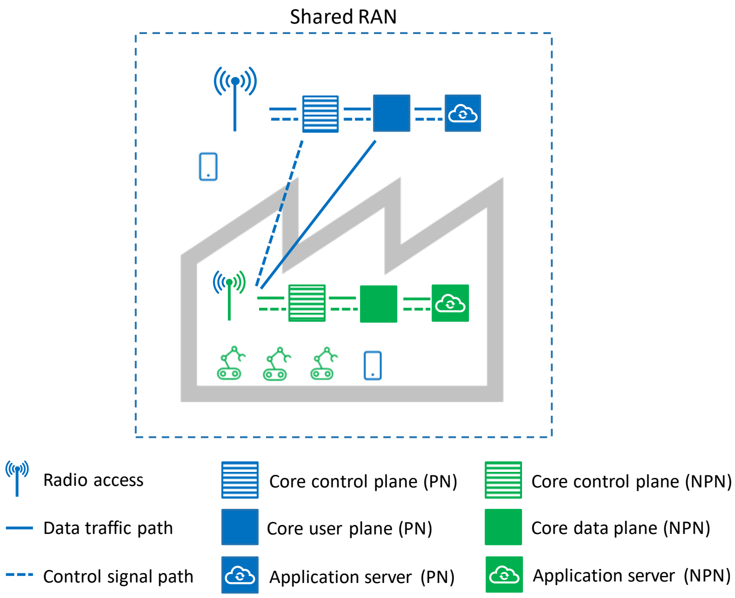

Figure 13. Description of shared RAN (source: Ericsson)

Shared RAN can be either provided by the MNO or deployed by the industrial party. The main motivations for shared RAN include for example: keeping the costs down, reducing the installation and maintenance work and sharing the gNodeB antenna locations that are favorably located from the radio network perspective. With the RAN sharing agreement, the industrial party can provide a free local site for the MNO, while the MNO may provide its spectrum resources for the network. Since the same gNodeBs provide both public and non-public services, the co-existence situation between the public and the non-public network can be improved. Furthermore, a shared network may be motivated by different services, e.g., the MNO may provide conventional enterprise services on the industrial site, while the non-public network is used for local IIoT connectivity.

4.1.2 Shared RAN and control plane

In this scenario, the NPN and the public network share the radio access network for the defined premises, while the network control tasks are always performed in the public network, see Figure 14. All NPN traffic flows remain within the logical perimeter of the defined premises, while the public network traffic portion is transferred to the public network. Since all the data flows related to the non-public network are local, there will be no extra penalties with respect to the achievable latencies. In addition to the shared gNodeBs, the non-public network can also contain gNodeBs that are accessible for the non-public users only.

This kind of solution can be implemented by means of network slicing, i.e., the creation of logically independent networks within a single, shared physical infrastructure. Segregation of the public and the non-public networks is achieved by employing different network slice identifiers. This scenario can also be implemented by means of a 3GPP-defined feature called access point name (APN). The APN denotes the final target network where to route the traffic, allowing differentiation between traffic portions.

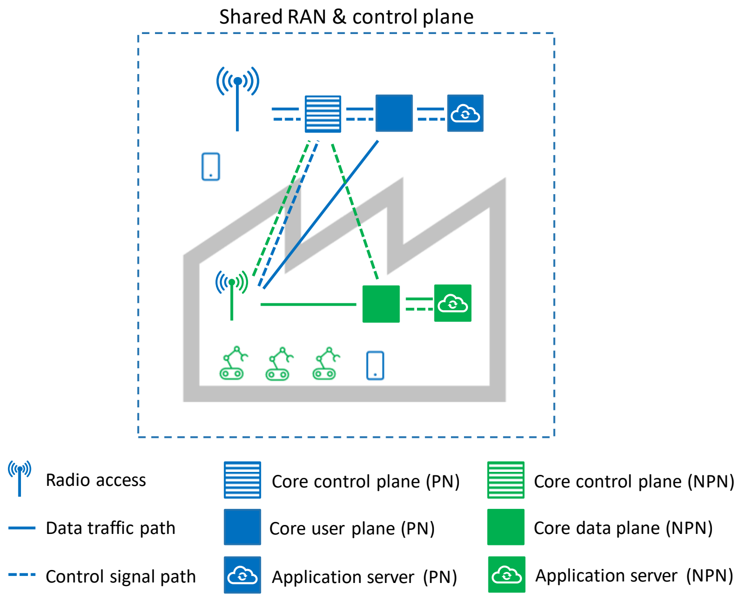

Figure 14. Description of a shared RAN and control plane (source: Ericsson)

In this scenario, the control plane of the non-public network is hosted by the public network and the non-public devices are public network subscribers. This allows the non-public devices to connect directly to the public network and its services. Furthermore, the non-public devices can connect to the non-public services via the public network, even when they are outside the coverage area of the non-public network. The public and non-public users being served by the shared gNodeBs can be operating either on a shared spectrum, or have their own, non-shared frequencies. Finally, since in this case the non-public network will be integrated with the public network, it will probably be likely that the MNO will deploy and operate the non-public network.

4.1.3 Non-public network hosted by a public network

When the non-public network is hosted by the public network, or in other words when the non-public network is provided as a service from the public network, as shown in Figure 15, the non-public user data leaves the enterprise's premises while still allowing the enterprise to obtain dedicated resources from the MNO's infrastructure (for example, through end-to-end dedicated network resources across radio, transport and core networks) with a service-level agreement. This scenario can be implemented by means of network slicing or APN functionality. It should be noted that since the user plane is terminated in the MNO's network, the additional delays caused by the core and transport network may have a negative impact on the achievable end-to-end latencies [ABS+20]. In addition to providing the requested communication service from the existing macro sites, the MNO can also deploy additional gNodeBs inside the enterprise's premises for coverage and performance reasons. In addition to the shared gNodeBs, the non-public network can also contain gNodeBs that are accessible only for the non-public users. Finally, the MNO has the sole responsibility over the network infrastructure and its operation.

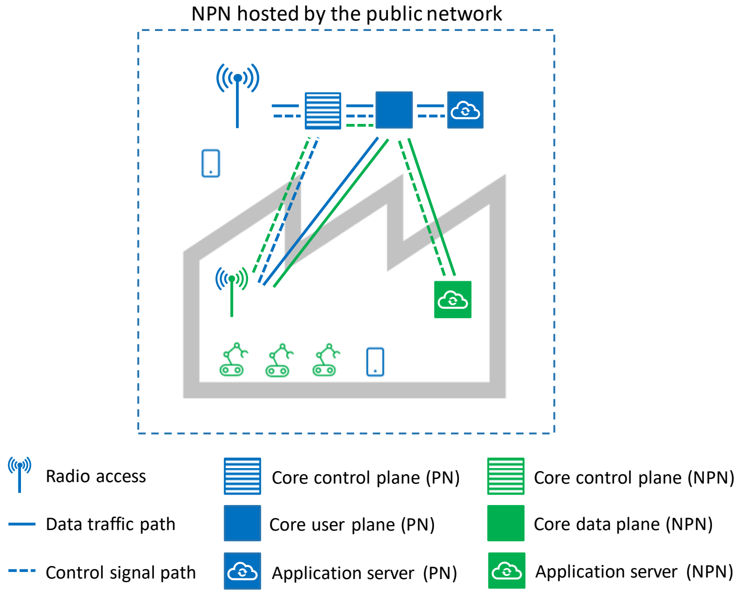

Figure 15. Description of the alternative where the non-public network hosted by a public network (source: Ericsson)

In this scenario the non-public subscribers are also public network subscribers. Since all data is routed via the public network, access to public network services and the ability to roam can be implemented easily in accordance with the agreement between the industrial party and the MNO.

4.2 Spectrum options

Similar to the independent standalone deployment, the spectrum options will depend on whether the non-public network is deployed by the industrial party or provided by the MNO. Hence, most of the discussion under Section 3.2 is valid for these scenarios as well, where the non-public network is deployed in conjunction with a public network.

A non-public network based on shared RAN can be deployed by the industrial party or provided by the MNO. From the spectrum point of view, the operators can be either sharing the same frequency resources (MOCN) or operate on their own non-shared frequencies (MORAN). The utilized spectrum resources can either come from the industrial party, MNO, or both. For example, in the case of an MNO-provided non-public network, the MNO can use its own spectrum resources, or it can also agree with the industrial party to combine MNO spectrum and local spectrum, when available, to improve the performance of the non-public network. For the usage of spectrum obtained by local licensing, consideration needs to be put on the licensing conditions, if e.g., restrictions are provided on the usage of the spectrum for public services.

If the non-public network is integrated with the public network, the non-public network will typically be deployed and operated by the MNO. Therefore, the MNO has the possibility to utilize all its spectrum assets, possibly combined with the local spectrum if available, to provide all the required non-public IIoT services.

Since the MNO is always involved, it becomes easier for the non-public network to support different types of IIoT services. Thus, the MNO can provide access to low- or mid-band FDD spectrum for the M-MTC services, while the mid- or high-band TDD spectrum, provided either by the MNO or the industrial party, can be utilized for the industrial eMBB and URLLC services.

4.3 Co-existence scenarios

Since the non-public network can offer both non-public and public services, co-existence problems related to the near-far interference can be considerably reduced. However, the potential cross-link interference problems described in Section 3.3 will still remain, as well as the near-far problems between the non-public network and the other public networks operating on the same frequency band as the non-public network. Furthermore, deployment of gNodeBs accessible only for the non-public users should be handled with care, since that may lead to new co-existence problems towards both the non-public and the public factory network, as shown in Figure 16.

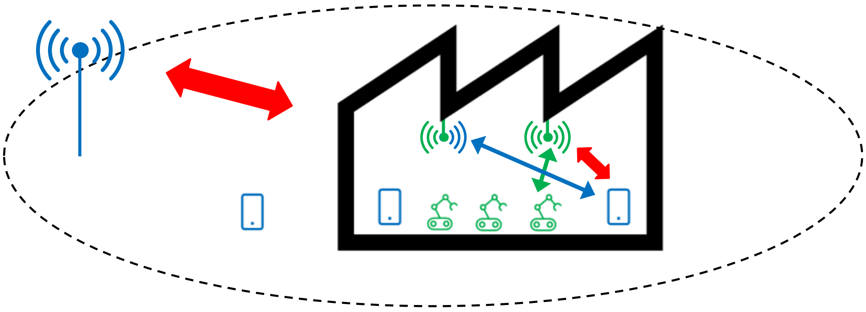

Figure 16. Co-existence between the public macro and the non-public factory network when the factory network contains both shared and non-public gNodeBs (source: Ericsson)

In the case of shared RAN, or when the non-public network is hosted by the public network, the gNodeBs providing the non-public network may be utilizing multiple frequencies either for capacity reasons or to be able to support multiple services in a more efficient way. The UE can then combine these frequency carriers for example by means of carrier aggregation. In order to avoid any harmful "cross-carrier interference" within the gNodeB and the UE, special attention can be required. For example, considering a few different spectrum combinations within the same gNodeB:

- FDD + FDD (i.e., a combination of two MNO carriers). This will be feasible without any special attention to avoid cross-carrier interference.
- FDD + TDD (i.e., a combination of two MNO carriers, or a combination of an MNO carrier and a local NPN carrier). This will most likely be feasible without any special attention, in particular if the carriers are not right next to each other, since the cross-carrier interference between the transmitter and the receiver within the gNodeB can be filtered away.
- Mid-band TDD + high-band TDD (i.e., a combination of two MNO carriers, a combination of an MNO carrier and a local NPN carrier, or a combination of two local NPN carriers). This

will most likely be feasible without any special attention, even for the unsynchronized TDD, since the cross-carrier interference can be filtered away.

- Mid-band TDD + mid-band TDD, or high-band TDD + high-band TDD (i.e., a combination of an MNO carrier and a local NPN carrier in the same band). This kind of deployment would have to be synchronized as a default due to the co-existence problems related to the cross-carrier interference (i.e., adjacent channel cross-link interference) within the gNodeB. However, an open question is if there would be any commercially viable ways to enable an unsynchronized deployment as well, e.g., by leaving a guard band between the carriers, by additional carrier-specific filters within the gNodeBs, or in scenarios, where the two frequency carriers are in the same band, but they are not right next to each other.

As already discussed in Section 3.3.2, shared RAN is a good way to resolve the near-far problems between public macro users and the non-public network. However, since in adjacent channel scenarios the TDD carriers sharing the gNodeB have to be synchronized, that will potentially create problems related to the co-channel cross-link interference between the macro network and the public users of the shared factory network, see Figure 17. The cross-link interference can be avoided by synchronizing all three networks to the eMBB-optimized TDD pattern. Alternatively, if a balanced TDD pattern is desired for the factory network, slot blanking can be applied either in the overlaid macro network or in the public factory network, as demonstrated in Figure 17. If the slot blanking is applied in the macro network, it reduces the maximum downlink capacity in the affected macro cells. However, at the same time the macro network will benefit from the offloading of the users located inside the factory, which will reduce the overall negative impact of the slot blanking on the capacity of the macro network. If the slot blanking is applied in the public factory network, the maximum uplink capacity will be reduced. However, since the utilization of the public factory network will typically be low, the negative impact of slot blanking will be rather modest. In addition to that, the performance of the macro network will benefit from the traffic offloading. Finally, as was already discussed in Section 3.3.2, the remaining adjacent channel cross-link interference between the public macro network and the non-public factory network will be straightforward to manage. An open question that remains for both slot blanking options is if the non-public network would be able to co-exist with the other MNOs operating in the same frequency band (and not being part of the shared RAN) while assuming an unsynchronized TDD configuration between the networks. Similarly, even though there would not be any cross-carrier interference requiring synchronized TDD within the shared gNodeBs, there could still be a requirement to synchronize one or both of the TDD carriers with the neighboring TDD networks or to apply some other means to mitigate the cross-link interference.

Finally, in the case of scenarios where the public and the non-public services are sharing the same TDD frequency carrier, both of them have to follow the same TDD pattern, which will typically lead to reduced system performance compared to scenarios with service-optimized TDD patterns. In all, similar to the independent standalone deployment, if a completely uncoordinated deployment between the public and the non-public network is desired, the non-public network should be deployed on an isolated frequency.

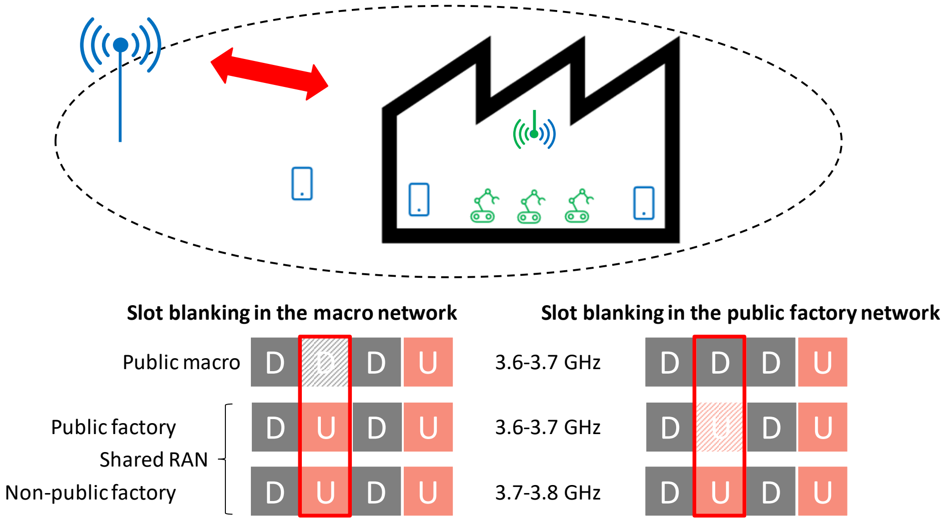

Figure 17. Description of an exemplary co-existence scenario between a public macro network and a non-public factory network with a shared RAN (source: Ericsson)

4.4 Summary

Deploying the non-public network in conjunction with a public network allows reuse of network infrastructure, efficient utilization of spectrum and seamless mobility. The network infrastructure can be deployed inside or outside the enterprise's premises in part or in its entirety and can be shared between the public and non-public users. There are three ways of realizing this: shared RAN, shared RAN and control plane, and non-public network hosted by a public network. Shared RAN can be provided by the MNO or deployed by the industrial party based on a RAN sharing agreement with the MNO. In the case of the other two alternatives the non-public network will be integrated with the public network, which means that the MNO will likely deploy and operate the non-public network.

Similar to the independent standalone deployment, the spectrum options will depend on the stakeholder deploying and operating the non-public network. In the case of shared RAN, the utilized spectrum resources can either come from the industrial party, MNO or both, depending on the chosen operation model and the regulations. If the non-public network is integrated with the public network, the non-public network will typically be deployed and operated by the MNO. Therefore, the MNO has the possibility to utilize all its spectrum assets, possibly combined with the local spectrum if available, to provide all the required non-public IIoT services. In general, since the MNO is always involved, it becomes easier for the non-public network to support different types of services. Thus, the MNO can provide access to low- or mid-band FDD spectrum for the M-MTC services, while the mid- or high-band TDD spectrum, provided either by the MNO or the industrial party, can be utilized for the industrial eMBB and URLLC services.

Since the non-public network can offer both non-public and public services, co-existence problems related to the near-far interference can be considerably reduced. However, the potential cross-link interference problems will still remain, as well as the near-far problems between the non-public network and the other public networks operating on the same frequency band as the non-public network. Furthermore, RAN sharing will potentially introduce new “cross-carrier” interference scenarios between the networks sharing the same gNodeBs. Finally, deployment of gNodeBs accessible only for the non-public users should be handled with care, since that may lead to new co-existence problems towards both the non-public and the public factory network.

5 Performance of a standalone factory network

This chapter will present results from a number of performance evaluations considering an independent standalone factory network. These evaluations do not specifically focus on the 5G-SMART use cases described in [5GS20-D11], but on more generic use cases, e.g., the ones defined within 3GPP [3GPP20-22104][3GPP20-22804] or by 5G-ACIA [5GA19-Aut][5GA20-UC]. Two types of evaluations are discussed: a) performance of an isolated industrial URLLC network, and b) co-existence between an industrial URLLC network and neighboring 5G networks.

As a default, the assumed deployment scenarios, simulation models and parameters have not been harmonized between the different studies presented in this chapter. However, they are in most of the cases aligned, or at least inspired by the models and parameters in [3GPP17-38802], [ITU17-M2412], [3GPP19-38824] and [3GPP19-38901].

5.1 Performance KPIs for the industrial 5G services

When it comes to URLLC services with strict QoS requirements, the evaluation methodology and the most relevant performance KPIs will typically look somewhat different compared to the traditional best-effort services such as eMBB. A URLLC service can be characterized for example with the following metrics and KPIs [5GS20-Term]:

- Traffic characteristics: user data rate, average/peak data rate, network layer packet sizes, packet transmission rates, traffic periodicity.
- (End-to-end) Latency: the time that it takes to transfer application data of a given size from a source to a destination, from the moment it is transmitted by the source to the moment it is successfully received at the destination (one-way latency).
- Network packet transmission reliability (or in short: reliability): the percentage of sent network layer packets successfully delivered to a given system entity within the time constraint required by the targeted service. Instead of specifying the network packet transmission reliability, one can refer also to requirements on survival time, communication service availability and communication service reliability (or mean time between failures), as discussed in detail in [5GS20-Term].
- Communication area: the given area where a communication service should operate, under given conditions and requirements.
- Communication density, or an area capacity: the maximum number of served devices, or the maximum served traffic volume within the desired communication area.

In addition to the discussion in [5GS20-D11] and [5GS20-Term], more detailed information on the industrial use cases, requirements and the KPIs can be found in [3GPP20-22104], [3GPP20-22261], [3GPP20-22804], [5GA19-Aut] and [5GA20-UC].

The URLLC coverage can be evaluated with the help of a metric called service availability, which is defined as the probability (measured over time and space) that the URLLC UEs can fulfil both the latency and reliability requirement. The maximum URLLC capacity $C(L,R)$ is in [3GPP19-38824] defined as the maximum offered cell or system load under which Y% of the URLLC UEs operate with target reliability *R* under latency bound *L*. In other words, *Y*% denotes the service availability, or alternatively *X* = (100-*Y*)% denotes the amount of UEs that are in outage, i.e., the UEs that cannot meet both the latency and the reliability requirement. Furthermore, it should be highlighted that the capacity model in [3GPP19-38824] assumes periodic URLLC data traffic, and that the data for all users arrives at the same time.

5.2 URLLC performance evaluations

5.2.1 Impact of the applied TDD pattern on the URLLC performance

When it comes to providing URLLC services on a TDD spectrum, one of the most critical aspects to consider is the applied TDD pattern, and how that will affect the URLLC performance. In general, the applied TDD pattern will define a lower bound for the achievable latency and it will have a clear the maximum system capacity. In case of URLLC services, the connection between the applied TDD pattern and the maximum system capacity is two-fold: On one hand, the applied TDD pattern defines the amount of time-domain resources that are available for downlink and uplink. On the other hand, the applied TDD pattern affects the general trade-off between latency, reliability and capacity. If the given latency bound allows only one transmission attempt, the packet has to be encoded with a very low and robust code rate, which will typically not be spectrally efficient. However, if the round-trip time (RTT) is shorter than the application latency constraint, it can be more efficient to use a higher, less robust initial code rate and perform retransmissions based on feedback in case the initial transmission fails. Thus, the shorter the RTT is compared to the application latency constraint, the higher spectral efficiency (i.e., capacity) may be achieved.

When evaluating the one-way latency, the downlink transmission consists of: the gNodeB transmitting a physical downlink control channel (PDCCH) to schedule a physical downlink shared channel (PDSCH), and the UE decoding the PDSCH and sending a corresponding hybrid automatic repeat request acknowledgement (HARQ-ACK) feedback which may trigger a HARQ retransmission of the same transport block. For configured grant[6] (CG) uplink transmission, initial transmission of the physical uplink shared channel (PUSCH) is according to the configured uplink grant without the need of sending a scheduling request, while the further PUSCH retransmission is based on dynamic scheduling from the gNodeB. Latency evaluation is done based on the time involved in the transmission process described above including all relevant processing delays in the gNodeB and the UE, and the alignment delay with respect to the transmission opportunities. A more detailed description of the assumed latency evaluation methodology and principles can be found in [SWD+18], [SKA+18] and [3GPP19-R11903446].

For URLLC performance, the focus is on what latency can be provided with high reliability. This means that the achievable latency is largely determined by the worst-case timing, e.g., when the downlink packet arrives exactly at the beginning of an uplink slot or vice versa (resulting in the largest possible alignment delay). An example of how the applied TDD pattern, together with the numerology and the choice between a slot-based or a sub-slot-based scheduling, will affect the achievable latency bound

[6] In URLLC, the UE can be configured to transmit in configured grant resources for uplink transmission that does not require the UE to transmit a scheduling request and receive uplink grant to reduce latency.

(i.e., the worst-case latencies) is shown in Table 1 (mid-band) and Table 2 (high-band). The calculations are assuming a system based on 3GPP NR Release 15, transmission time interval (TTI) of either 14 symbols (TDD pattern based on full slots) or 7 symbols (TDD pattern based on half-slots) and a CG-based scheduling for the uplink. For the mid-band with sub-carrier spacing (SCS) equal to 30 kHz, this means that the TTI is either equal to 0.5 ms or 0.25 ms. Furthermore, for the high-band with SCS equal to 120 kHz, the TTI is equal to 125 μs or 63 μs.

	DDDDU 14		DDDU 14		DUDU 14		DUDU 7		UUUD 14	
	DL	**UL**	**DL**	**UL**	**DL**	**UL**	**DL**	**UL**	**DL**	**UL**
1st Tx	1.68	3.18	1.68	2.68	1.68	1.68	0.93	0.93	2.68	1.68
2nd Tx	4.18	5.68	3.68	4.68	2.68	2.68	1.93	1.93	4.68	3.68
3rd Tx	6.68	8.18	5.68	6.68	3.68	3.68	2.93	2.93	6.68	5.68
4th Tx	9.18	10.68	7.68	8.68	4.68	4.68	3.93	3.93	8.68	7.68
5th Tx	11.68	13.18	9.68	10.68	5.68	5.68	4.93	4.93	10.68	9.68

Table 1. Downlink and uplink worst-case latencies (expressed in ms), assuming different TDD patterns, 1-5 transmission attempts, and a sub-carrier spacing equal to 30 kHz.

	DDDDU 14		DDDU 14		DUDU 14		DUDU 7		UUUD 14	
	DL	**UL**	**DL**	**UL**	**DL**	**UL**	**DL**	**UL**	**DL**	**UL**
1st Tx	0.63	1.00	0.63	0.88	0.63	0.63	0.44	0.44	0.88	0.63
2nd Tx	1.63	2.25	1.50	1.88	1.38	1.38	1.19	1.19	1.88	1.50
3rd Tx	2.38	3.50	2.50	2.88	2.13	2.13	1.94	1.94	2.88	2.50
4th Tx	3.50	4.75	3.50	3.88	2.88	2.88	2.69	2.69	3.88	3.50
5th Tx	4.25	6.00	4.50	4.88	3.63	3.63	3.44	3.44	4.88	4.50

Table 2. Downlink and uplink worst-case latencies (expressed in ms), assuming different TDD patterns, 1-5 transmission attempts, and a sub-carrier spacing equal to 120 kHz.

In addition to the processing delays in both the gNodeB and the UE, the worst-case latency values for the initial transmission attempt depend on the assumed TDD pattern and the TTI value. For example, the downlink values are the same for “DDDDU 14”, “DDDU 14” and “DUDU 14”, because all of them have the same TTI and a maximum of one uplink slot between two consecutive downlink slots. “DUDU 7” has a shorter TTI compared to the other TDD patterns, and thus, a shorter maximum delay before the initial transmission attempt. Finally, “UUUD 14” has the worst downlink delay due to the fact that two consecutive downlink slots are always separated by three uplink slots. The retransmissions will experience different delays depending on the assumed TDD pattern, TTI value, and whether a “scheduling request” or a “configured grant” is assumed for the uplink. Again, the TDD pattern and the TTI value define the maximum delay for a certain transmission direction, but the difference compared to the initial transmission is that due to the required hybrid automatic repeat request (HARQ) feedback, the total delay will have to consider the delays related to both the downlink and the uplink transmission.

Looking at the values in Table 1 and Table 2, can be seen that if a very stringent URLLC service with a 1 ms latency bound is assumed for both the downlink and the uplink, the mid-band can support it only with the DUDU pattern based on half-slots, i.e., when the TTI is equal to 0.25 ms. At the same time, the high-band with a shorter TTI can support it with all of the evaluated TDD patterns. However, for both bands, only one transmission attempt is possible within the given latency bound, and hence, achieving the high reliability would be spectrally inefficient. If a less stringent URLLC service with latency bound of 5 ms is assumed instead, mid-band can support it even with the more eMBB-optimized TDD patterns. However, the maximum number of transmission attempts, and hence, the spectral efficiency will vary between the TDD patterns: the downlink- or uplink-heavy patterns allow a maximum 1 or 2 transmission attempts, while the balanced patterns allow a maximum of 4 or 5 transmission attempts, making them more spectrally efficient. Hence, for services with unbalanced traffic requirements, e.g., requiring 75% uplink or downlink data, with strict requirements on latency and reliability, there is a trade-off between selecting a TDD pattern that matches the offered traffic or a TDD pattern that reduces the latency.

In order to evaluate the impact of the TDD pattern, the sub-carrier spacing (i.e., frequency band) and the scheduling choices on the maximum URLLC system capacity, a number of system-level simulations have been run. The evaluations assume a single-floor factory building with the size of 120x60x10 m^3, containing three omnidirectional and ceiling-mounted gNodeBs, see Figure 18. The propagation model is based on the 3GPP model for Indoor Factory with Dense clutter and High base station height (InF-DH) [3GPP19-38901]. Furthermore, the factory layout has been defined with the following parameter values: average clutter density equal to 60%, average clutter height equal to 6 m, average clutter size equal to 2 m, gNodeB antenna height equal to 8 m and UE antenna height equal to 1.5 m.

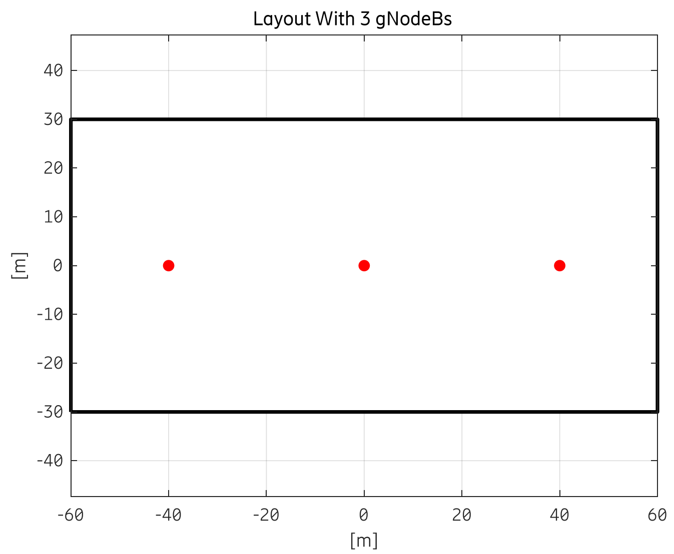

Figure 18. Assumed factory layout with three ceiling-mounted gNodeBs.

The URLLC packet size is assumed to be equal to 32 bytes, the latency bound is assumed to be equal to {1, 2, 3, 5} ms and the reliability requirement is assumed to be equal to {99.9, 99.999}%. The main system-level simulation assumptions are listed in Table 3, and a more detailed description of both the link-level simulation assumptions and the reliability model (considering the received quality of both the control and the shared channels) can be found in [SKA+18] and [AHK+19].

Parameter	Mid-Band	High-Band
Frequency [GHz]	3.6	26
Channel bandwidth [MHz]	100	400
gNodeB transmission power [dBm]	30	30
UE transmission power [dBm]	23	23
gNodeB antenna gain (omni) [dBi]	2	2
UE antenna gain [dBi]	0	9
gNodeB receiver noise figure [dB]	5	7
UE receiver noise figure [dB]	9	10
Uplink power control setup	SNR target = 10 dB α = 0.8	SNR target = 10 dB α = 0.8

Table 3. Main system-level simulation parameters

The obtained maximum URLLC downlink (DL) and uplink (UL) system capacity values for the mid-band and high-band are presented in Table 4 and Table 5, respectively. Here, the maximum URLLC system capacity is defined as the maximum level of the offered traffic that can be served with a service availability within the factory floor equal to 100%. Furthermore, a user is assumed to be served if both the latency and the reliability requirements can be met. In practice, the desired QoS cannot be guaranteed if a) the maximum achievable user bit rate is less than what would be required to transmit the message payload during one TTI, or b) the system does not have enough radio resources to successfully serve the total network offered load.

Latency	Reliability	DDDDU 14		DDDU 14		DUDU 14		DUDU 7		UUUD 14	
		DL	UL	DL	UL	DL	UL	DL	UL	DL	UL
1 ms	99.999%	-	-	-	-	-	-	20	61	-	-
	99.9%	-	-	-	-	-	-	99	110	-	-
2 ms	99.999%	-	-	-	-	22	60	130	145	-	-
	99.9%	-	-	-	-	105	105	200	230	-	-
3 ms	99.999%	-	-	33	30	145	150	190	220	11	90
	99.9%	-	-	160	54	220	230	265	310	54	160
5 ms	99.999%	235	24	220	74	265	280	265	310	73	225
	99.9%	355	43	335	115	320	330	300	345	110	340

Table 4. Maximum URLLC downlink and uplink system capacity (in Mbps) for the mid-band.

Looking at the results, it becomes clear that the maximum system capacity is very low for the situations when the latency budget allows only one transmission attempt. As already discussed, the packet has to be encoded with a very low and robust code rate to be able to guarantee the desired QoS with a single transmission attempt, which reduces the spectral efficiency. Another thing affecting the maximum system capacity is the fact that in order to secure a sufficiently high (momentary, or per TTI) SINR even for the worst users (so that the system would have enough radio resources to serve them), the level of the average cell resource utilization has to kept at a low level to limit the probability and the level of the inter-cell interference peaks. If the QoS requirement is relaxed, either in terms of reliability or latency, the service becomes more tolerant to inter-cell interference, allowing the use of a less robust code rate and also a higher level of the cell resource utilization while still being able to guarantee the desired QoS. As a result, the maximum system capacity is increased.

Latency	Reliability	DDDDU 14		DDDU 14		DUDU 14		DUDU 7		UUUD 14	
		DL	UL	DL	UL	DL	UL	DL	UL	DL	UL
1 ms	99.999%	375	92	350	115	235	230	185	230	100	350
	99.9%	640	155	600	195	400	385	360	385	200	580
2 ms	99.999%	815	92	765	240	505	485	620	670	250	730
	99.9%	1250	155	1170	375	775	760	915	1005	390	1140
3 ms	99.999%	1090	190	1025	330	725	745	660	745	340	1010
	99.9%	1615	300	1515	505	1080	1085	985	1080	500	1510
5 ms	99.999%	1325	300	1240	435	985	1015	895	1035	410	1295
	99.9%	1800	425	1690	560	1170	1195	1070	1190	560	1695

Table 5. Maximum URLLC downlink and uplink system capacity (in Mbps) for the high-band.

In order to visualize the impact of the mismatch between the offered traffic and the applied TDD pattern, the values in Table 4 can be further processed by assuming that the offered traffic distribution between downlink and uplink (DL:UL) is equal to {1:1, 3:1, 1:3}. Again, the maximum URLLC system capacity is defined as the maximum level of the offered traffic that can be served with a service availability within the factory floor equal to 100%. However, now a user is assumed to be served if both the latency and the reliability requirements can be met simultaneously for both the downlink and the uplink.

The corresponding maximum URLLC system capacity values are presented in Table 6 – Table 8 for the different types of the offered traffic (i.e., the different DL:UL traffic ratios) at the mid-band and in in Table 9 – Table 11 for the high-band.

Latency	Reliability	DDDDU 14		DDDU 14		DUDU 14		DUDU 7		UUUD 14	
		DL	UL	DL	UL	DL	UL	DL	UL	DL	UL
1 ms	99.999%	-	-	-	-	-	-	20	20	-	-
	99.9%	-	-	-	-	-	-	99	99	-	-
2 ms	99.999%	-	-	-	-	22	22	130	130	-	-
	99.9%	-	-	-	-	105	105	200	200	-	-
3 ms	99.999%	-	-	30	30	145	145	190	190	11	11
	99.9%	-	-	54	54	220	220	265	265	54	54
5 ms	99.999%	24	24	74	74	265	265	265	265	73	73
	99.9%	43	43	115	115	320	320	300	300	110	110

Table 6. Maximum URLLC system capacity (in Mbps) for the mid-band, when the offered traffic has a DL:UL ratio of 1:1.

Latency	Reliability	DDDDU 14		DDDU 14		DUDU 14		DUDU 7		UUUD 14	
		DL	UL	DL	UL	DL	UL	DL	UL	DL	UL
1 ms	99.999%	-	-	-	-	-	-	20	7	-	-
	99.9%	-	-	-	-	-	-	99	33	-	-
2 ms	99.999%	-	-	-	-	22	7	130	43	-	-
	99.9%	-	-	-	-	105	35	200	67	-	-
3 ms	99.999%	-	-	33	11	145	48	190	63	11	4
	99.9%	-	-	160	53	220	73	265	88	54	18
5 ms	99.999%	72	24	220	73	265	88	265	88	73	24
	99.9%	129	43	335	112	300	100	300	100	110	37

Table 7. Maximum URLLC system capacity (in Mbps) for the mid-band, when the offered traffic has a DL:UL ratio of 3:1.

Latency	Reliability	DDDDU 14		DDDU 14		DUDU 14		DUDU 7		UUUD 14	
		DL	UL	DL	UL	DL	UL	DL	UL	DL	UL
1 ms	99.999%	-	-	-	-	-	-	20	60	-	-
	99.9%	-	-	-	-	-	-	37	110	-	-
2 ms	99.999%	-	-	-	-	20	60	48	145	-	-
	99.9%	-	-	-	-	35	105	77	230	-	-
3 ms	99.999%	-	-	10	30	50	150	73	220	11	33
	99.9%	-	-	18	54	77	230	103	310	53	160
5 ms	99.999%	8	24	25	74	93	280	103	310	73	219
	99.9%	14	43	38	115	110	330	115	345	110	330

Table 8. Maximum URLLC system capacity (in Mbps) for the mid-band, when the offered traffic has a DL:UL ratio of 1:3.

What is clear for both the mid-band and the high-band is that if an eMBB-optimized downlink-heavy TDD pattern is assumed, maximum capacity of the more balanced or even uplink-heavy URLLC services will greatly suffer. At the same time, a balanced TDD pattern offering lower latency bounds may in some cases offer a higher system capacity, in particular at the mid-band, even though the offered traffic is not balanced. This is due to the fact that the shorter latencies allow more transmission attempts, and hence, the use of higher modulation and coding schemes (MCS), which will be more spectrally efficient. Furthermore, a higher level of the average cell resource utilization can be allowed, which increases the overall system capacity. In the case of the latency-critical URLLC services, this can then turn out to be more important than the fact that the DL:UL ratio of the applied TDD pattern does not match the DL:UL ratio of the offered traffic. However, the results in Table 6 and Table 9 demonstrate that if the TDD pattern matches the offered traffic (i.e., looking at the results for "DUDU 14" and "DUDU 7") it is not always beneficial to select a pattern that allows more transmission attempts. Patterns with shorter TTIs (e.g., "DUDU 7") allow lower latencies (and hence, more transmission attempts) but with the cost of an increased overhead. Looking at the capacity numbers, it becomes clear that after a certain point the gain of being able to perform more retransmissions with a less robust code rate is no longer able to outweigh the capacity losses caused by the increased overhead. A possible way to improve the efficiency of the "DUDU 7" pattern could be to limit the number of retransmissions by selecting a more robust code rate (i.e., a lower block error rate (BLER) target for the link adaptation) and use the remaining latency budget to time-multiplex users instead.

Latency	Reliability	DDDDU 14		DDDU 14		DUDU 14		DUDU 7		UUUD 14	
		DL	UL	DL	UL	DL	UL	DL	UL	DL	UL
1 ms	99.999%	92	92	115	115	230	230	185	185	100	100
	99.9%	155	155	195	195	385	385	360	360	200	200
2 ms	99.999%	92	92	240	240	485	485	620	620	250	250
	99.9%	155	155	375	375	760	760	915	915	390	390
3 ms	99.999%	190	190	330	330	725	725	660	660	340	340
	99.9%	300	300	505	505	1080	1080	985	985	500	500
5 ms	99.999%	300	300	435	435	985	985	895	895	410	410
	99.9%	425	425	560	560	1170	1170	1070	1070	560	560

Table 9. Maximum URLLC system capacity (in Mbps) for the high-band, when the offered traffic has a DL:UL ratio of 1:1.

Latency	Reliability	DDDDU 14		DDDU 14		DUDU 14		DUDU 7		UUUD 14	
		DL	UL	DL	UL	DL	UL	DL	UL	DL	UL
1 ms	99.999%	276	92	345	115	235	78	185	62	100	33
	99.9%	465	155	585	195	400	133	360	120	200	67
2 ms	99.999%	276	92	720	240	505	168	620	207	250	83
	99.9%	465	155	1125	375	775	258	915	305	390	130
3 ms	99.999%	570	190	990	330	725	242	660	220	340	113
	99.9%	900	300	1515	505	1080	360	985	328	500	167
5 ms	99.999%	900	300	1240	413	985	328	895	298	410	137
	99.9%	1275	425	1680	560	1170	390	1070	357	560	187

Table 10. Maximum URLLC system capacity (in Mbps) for the high-band, when the offered traffic has a DL:UL ratio of 3:1.

Latency	Reliability	DDDDU 14		DDDU 14		DUDU 14		DUDU 7		UUUD 14	
		DL	UL	DL	UL	DL	UL	DL	UL	DL	UL
1 ms	99.999%	31	92	38	115	77	230	77	230	100	300
	99.9%	52	155	65	195	128	385	128	385	193	580
2 ms	99.999%	31	92	80	240	162	485	223	670	243	730
	99.9%	52	155	125	375	253	760	335	1005	380	1140
3 ms	99.999%	63	190	110	330	248	745	248	745	337	1010
	99.9%	100	300	168	505	362	1085	360	1080	500	1500
5 ms	99.999%	100	300	145	435	338	1015	345	1035	410	1230
	99.9%	142	425	187	560	398	1195	397	1190	560	1680

Table 11. Maximum URLLC system capacity (in Mbps) for the high-band, when the offered traffic has a DL:UL ratio of 1:3.

5.2.2 Impact of network densification, base station antenna and frequency band on the URLLC performance

Densification of the network deployment, for example by adding new sectors or sites within an existing communication network, has traditionally been an efficient way to increase the system capacity. That will be the case even for URLLC deployments, but the overall situation changes a bit compared to e.g., best-effort eMBB deployments.

It is known that network densification can often reduce the maximum cell capacity, but as the number of cells serving a certain geographical area is increased, the overall maximum system capacity increases. From the maximum cell capacity point of view, network densification has both positive and negative impacts. To start with, network densification has often a negative impact on the observed geometries[7], since the reduced distances between the base stations and the users will increase the received inter-cell interference faster than the received signal from the serving base station [ZA15][DWL16]. However, assuming a fixed level of offered area traffic (Mbps/m^2), network densification will reduce the level of the offered traffic per cell, which can potentially lead to a lower level of average resource utilization, and a lower level of the average inter-cell interference. When it comes to the impact of network densification on the received SINR, it will benefit from the improved received signal power from the serving cell and the potentially reduced utilizations of the neighboring cells. However, the SINR will suffer from the fact that the maximum number of interfering neighbors is increasing and that the interferers are moving closer. Hence, the SINR may well get worse as a result of the network densification as long as the impact of traffic offloading is not large enough to compensate for the increased inter-cell interference peaks (i.e., the impact of the reduced geometry). Finally, the traffic offloading within the serving cell will allow the served users to get more radio resources, which has a positive impact on the user throughputs, and reduces the negative impact of the worse SINRs.

While the overall eMBB system performance is typically measured by looking at the average and the "cell-edge" (defined often as the worst 5th percentile) user throughputs, the URLLC system performance is typically measured by looking at the performance of the worst users, which in practice means that the impact of network densification on the worst SINRs becomes highly important. If the worst users cannot reach a good enough SINR to guarantee the desired level of reliability, the overall URLLC system cannot reach the desired service availability target, and as a result, the maximum URLLC system capacity will degrade. Therefore, it will be important to design the network densification so that the new sectors or sites will not degrade the worst SINRs in any significant way.

To demonstrate this, a simple system-level performance evaluation has been performed. The default system setup, models and assumptions are the same as in Section 5.2.1. What is different compared to the study in Section 5.2.1 is that a) the number of gNodeBs is varied between {1, 2, 3, 6, 8, 12, 18} (see Figure 19), b) a DUDU TDD pattern based on half-slots (7 symbols) is always assumed, c) the offered traffic is assumed to be balanced (DL:UL = 1:1), and d) the performance evaluations are

[7] Geometry = downlink signal-to-interference ratio (SIR) for a fully-loaded network, i.e., when all base stations are continuously transmitting at the maximum power.

repeated for a few different gNodeB antenna options (see below). Finally, two different latency requirements (1 ms and 3 ms) and a reliability requirement of 99.999% are assumed.

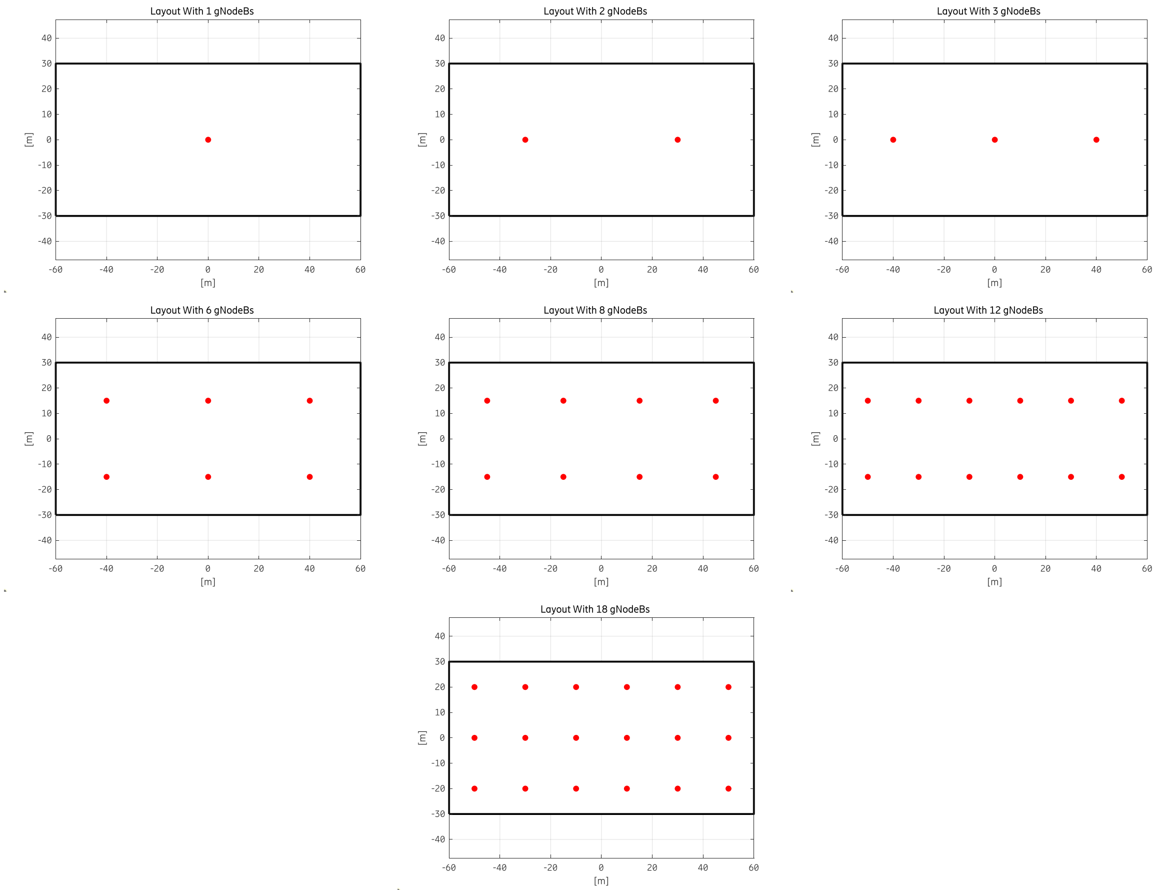

Figure 19. Assumed factory layout with {1, 2, 3, 6, 8, 12, 18} ceiling-mounted gNodeBs

The evaluations have been done for the following gNodeB antenna options:

- Omnidirectional ("dipole") antenna. Ceiling-mounted, no downtilt
- Directional antenna. Ceiling-mounted, 90° half-power beamwidth both in horizontal and vertical direction, 90° mechanical downtilt (i.e., the antenna is pointing downwards)
- Beamformed antenna. Ceiling-mounted, 4x4x2 antenna array (i.e., 16 antenna elements per polarization), digital UE-specific beamforming, 90° mechanical downtilt
- Active distributed antenna system (DAS) consisting of 8 omnidirectional, ceiling-mounted antennas with 35° electrical downtilt. Note that in this case the network consists of only one cell (i.e., one gNodeB). The transmission power per antenna is assumed to be equal to 30 dBm, and the uplink receiver noise figure is increased to 19 dB. It should be noted that due to the assumed uplink power control, the received uplink signal-to-noise ratio (SNR) will be roughly the same no matter if the network consists of 8 gNodeBs, or an active DAS with 8 antennas.

During these evaluations, the maximum system capacity is defined as the maximum level of the offered traffic that can be served with a service availability within the factory floor equal to 100%. Furthermore, a user is assumed to be served if both the latency and the reliability requirements can be met simultaneously for both the downlink and the uplink. A summary of the mid-band results is provided in Figure 20. As can be seen, performance of the most stringent URLLC services is very sensitive to the inter-cell interference, since the assumed latency bound of 1 ms allows only one transmission attempt, requiring in practice a fairly high (momentary) SINR to guarantee the desired level of reliability. In other words, the maximum URLLC capacity is very sensitive to the occasional inter-cell interference peaks, in particular since a 100% service availability is required[8]. As can be noticed, network densification with omnidirectional or directional gNodeB antennas will reduce the overall system capacity, unless the network is made dense enough. In fact, in most of the cases, the multi-cell performance is worse than the single cell performance, with either one gNodeB deployed in the middle of the factory, or an active DAS with 8 distributed antennas. This then means that if the main reason to densify the network is to improve the coverage throughout the factory floor, and not to achieve extreme system capacity, it would be better to rely on an active distributed antenna system instead of deploying additional gNodeBs. The situation becomes different if the gNodeBs have beamformed antennas. For that kind of a deployment, network densification by adding new gNodeBs is clearly beneficial, and improves the overall URLLC system capacity. The main reason why the beamforming is so beneficial is that the narrow UE-specific beams are very efficient in limiting the level of the inter-cell interference, even when the cells become smaller.

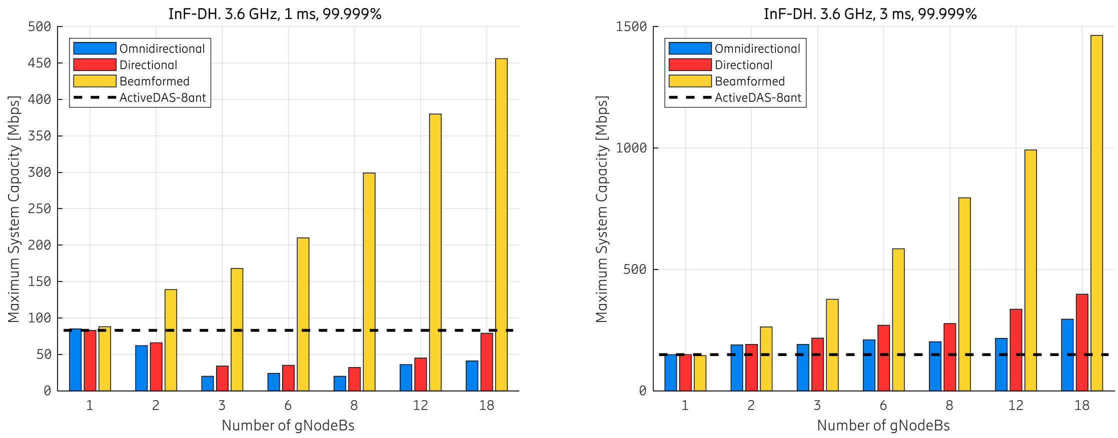

Figure 20. Maximum URLLC system capacity at mid-band.

As demonstrated by the results in Figure 20, the maximum system capacity of the more relaxed URLLC services, allowing multiple transmission attempts, can be improved also with omnidirectional or directional gNodeB antennas. With multiple transmission attempts, a maximum of three in this case

[8] This would in practice require the use of a large interference (or SINR) margin when designing the network, i.e., the URLLC network should be planned and dimensioned to cope with the worst-case inter-cell interference peaks. Alternatively (or in addition), the average cell load (resource utilization) should be kept at a low level to limit the probability of excessive inter-cell interference peaks.

(latency bound of 3 ms), the desired level of reliability can be guaranteed with a lower SINR than in the case of the single transmission attempt, which makes the system much more tolerant to the occasional inter-cell interference peaks. As a result, higher levels of the average cell resource utilization can be allowed, as demonstrated by the results in Figure 21, leading to a higher maximum system capacity. It can also be highlighted that the performance of the directional antennas could potentially be improved by utilizing narrower beams for the denser deployments to minimize the cell overlap. However, the sparse network deployments would still require wider beams for coverage reasons.

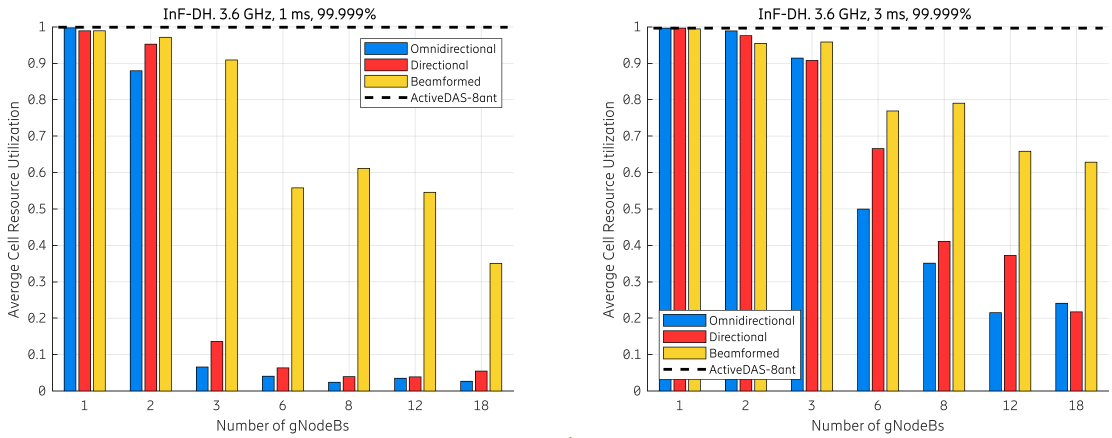

Figure 21. Average cell resource utilization when the offered traffic is equal to the maximum URLLC system capacity

A summary of the high-band results is presented in Figure 22. In general, the conclusions are similar to the mid-band. However, it is clear that due to the more challenging propagation conditions, one gNodeB deployed in the middle of the factory is not able to provide good coverage throughout the entire factory floor. In the case of the most stringent URLLC services, it would be better to improve the coverage by adding more antennas into an active distributed antenna system, or to densify the network with beamformed gNodeBs. For the more relaxed URLLC services, the system performance can be enhanced also by adding gNodeBs with omnidirectional or directional antennas.

As discussed in Section 2.5.2, it might not be realistic to assume a fully digital UE-specific beamforming for the high-band deployment. If an analog or hybrid beamforming would be assumed instead, only one or a few users could be simultaneously scheduled at the same time, which would sacrifice the spectral efficiency compared to the digital UE-specific beamforming for serving latency-critical traffic based on small packets. The situation would be slightly better for the relaxed URLLC services, having a latency requirement which is much greater than the slot duration, because then it would be possible to time-multiplex the URLLC users while still fulfilling the latency requirement. But even then, the scheduled users might not have enough data to occupy all the available frequency-domain resources. In order to rely on frequency-domain multiplexing of users to satisfy the most stringent latency requirements, it might in fact be necessary to assume directional antennas for the high-band deployments instead of beamforming. However, by doing so, there will be a loss in the link budget due to the reduced antenna gains, as well as an increased level of inter-cell interference. In the end, the maximum achievable system capacity can become considerably lower, as demonstrated by the results in Figure 22. An open question to be investigated further is the feasibility of fully digital beamforming for the more reasonably sized antenna arrays, e.g., the 4x4x2 array assumed in these simulations. The obvious downside of such small arrays is the worse beamforming gain, and hence, worse coverage compared to the large antenna arrays. However, as indicated by the results in Figure 22, the use of such antennas can provide a clear performance improvement compared to omnidirectional or directional gNodeB antennas and could in fact be sufficient for most of the industrial 5G deployments.

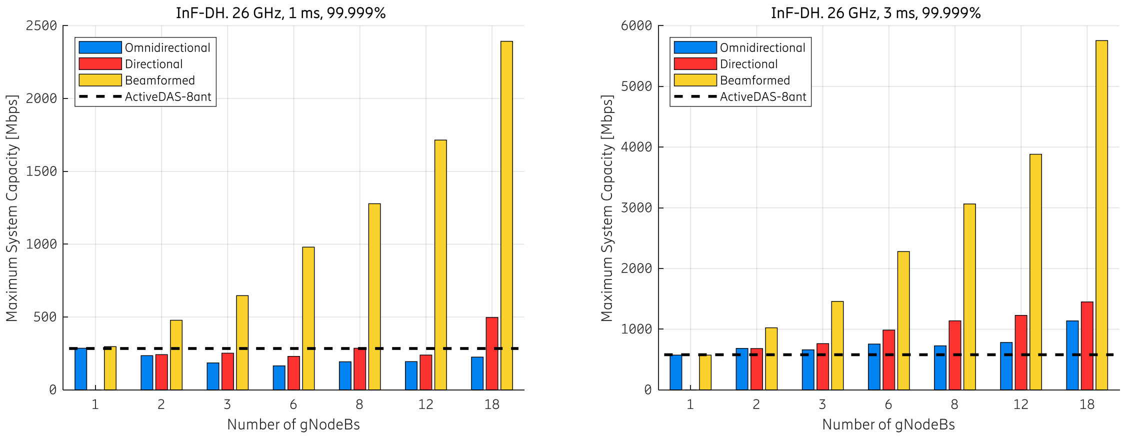

Figure 22. Maximum URLLC system capacity at high-band.

The high-band deployment is in this study assumed to have four times as much bandwidth compared to the mid-band deployment, which means that with the same spectral efficiency the maximum system capacity will become four-fold. When the maximum system capacities in Figure 20 are compared with the corresponding maximum system capacities in Figure 22, it becomes clear that the performance of the high-band deployment suffers greatly from the more challenging propagation conditions for the sparse network deployments, i.e., with either one or two gNodeBs. However, once the network is dense enough, i.e., with three gNodeBs or more, the frequency bands have quite similar spectral efficiencies. The only cases, when the high-band deployment offers a clearly better spectral efficiency, are related to the URLLC service with the most stringent QoS requirements and when the number of gNodeBs with omnidirectional or directional antennas is between 3 and 18, i.e., when the performance of the mid-band deployment is heavily limited by the inter-cell interference.

5.2.3 Downlink resource sharing between URLLC and eMBB traffic

The deployment of a non-public 5G network within a smart factory aims at providing communication services for smart manufacturing use cases and offers on-premises industrial communication services including URLLC, but also eMBB. The results presented in this section assume a scenario where the URLLC and the eMBB services are sharing the factory network resources, but only the downlink has been evaluated. On the one hand, the factory workers use eMBB services in the scenario where the overlaid public macro network is not available. On the other hand, the factory robots communicate with the central controller using the URLLC services.

Downlink system-level simulations considering different deployment scenarios and configurations have been conducted in order to evaluate both the URLLC performance in terms of the latency and the packet loss probability and the eMBB performance in terms of the user throughput. It should be highlighted that the assumed simulations do not provide a complete picture of the system performance since only the downlink is considered. It is well known that due to the typically used TDD configurations, the uplink is very challenging in terms of both resource allocation and latency and is to be considered in the future studies. The aim of these downlink simulations is to evaluate the delay that a robot needs to execute and perform the tasks indicated by the messages sent by the controller, and to estimate the impact of the deployment options and configurations on the downlink performance.

The simulation environment is a factory including a set of wirelessly connected assembly (production) line robots. Each robot is equipped with a URLLC transceiver that communicates with a central controller via a set of gNodeBs deployed within the factory. As presented in Figure 23 the size of the communication service area is equal to 160 m × 160 m where 280 robots are distributed in five production lines. In addition, five eMBB users are randomly located within the communication service area in order to study the feasibility and the performance of the multi-service scenario. The simulations consider three different deployment options with one, two, or four ceiling-mounted gNodeBs. The gNodeBs are located at a height of 10 m and they are equipped with two directional transmit antennas. The assumed channel model is based on the 3GPP Indoor Factory Dense clutter High BS (InF-DH) model defined in [3GPP19-38901].

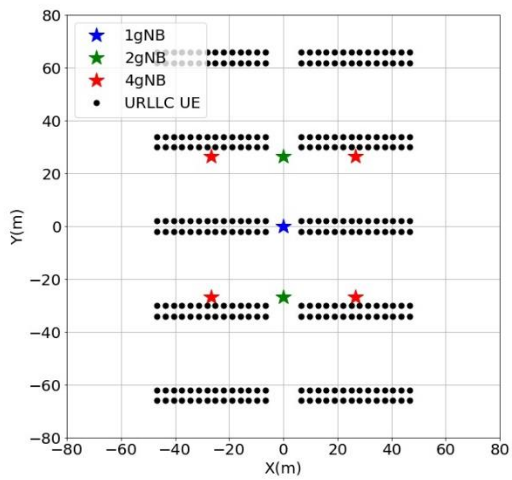

Figure 23. Factory layout for the different deployment options with one, two or four gNodeBs

The URLLC traffic model is periodic with a 1 ms interval between two consecutive packets. The size of the URLLC packet is 96 bits (12 bytes) and the block error rate (BLER) target is 10^{-5}, while the eMBB follows a full buffer traffic model where the transport block size depends on the number of users and the available resource blocks. A target BLER of 10^{-1} is assumed for the eMBB traffic. As the URLLC and the eMBB users share the resources to satisfy the requirements of both traffic types, the adopted scheduling algorithm performs bandwidth (B) sharing with the goal to satisfy the stringent QoS requirements of the URLLC users while at the same time guaranteeing a minimum QoS for the eMBB traffic. This is achieved by prioritizing the URLLC traffic as follows:

- The scheduler manages a queue where the generated packets are stored until they are correctly delivered.
- At each time slot, the scheduler checks if there are any URLLC packets awaiting service in the buffer. If so, the resources are allocated to those packets starting from the oldest one until either all URLLC packets are served or the maximum amount of reserved resources αB is attained, where the preconfigured parameter α represents the fraction of the bandwidth reserved for the URLLC service. The eMBB packets are then served on the rest of the bandwidth, $(1-\alpha)B$ which is exclusively dedicated to the eMBB traffic, following a round robin scheduling scheme.
- When serving a packet, the number of required resource blocks (RB) is determined for each MCS satisfying the target BLER, according to the modulation order, the coding rate, the RB size and the packet size. If several MCSs can satisfy these conditions, the MCS that offers the best spectral efficiency is selected.
- In case none of the MCSs can achieve the target BLER requirement with the available RBs, which means that either the channel quality is not good enough or the amount of available RBs is not sufficient, the packet is delayed to a later TTI until the channel quality becomes better or more resources are available.

In the present evaluations α is assumed to be equal to 0.7, which means that 70% of the bandwidth is reserved for the URLLC traffic and 30% for the eMBB traffic. Furthermore, it should be noted that for implementing this scheduling scheme in the system-level simulator, lookup tables providing a mapping between the BLER and the SINR in an additive white Gaussian noise (AWGN) channel are used for each MCS. Classical LTE-like turbo coding is implemented for the eMBB, while enhanced turbo coding is used for the URLLC in order to allow a very low BLER.

When evaluating the delay, the focus is particularly on the physical (PHY) and medium access control (MAC) layer user plane latency T, which is defined by the following equation:

$$T = T_{Tx} + T_{Alg} + T_{OT} + T_{Rx}$$

Parameter T_{Tx} represents the transmitter processing time which in our case corresponds to the gNodeB processing delay equal to one TTI. T_{Alg} is the alignment delay and T_{OT} corresponds to the over-the-air transmission. Finally, T_{Rx} is the UE processing delay which depends on the UE receiver type, which is assumed as advanced in the present evaluations and the configurations.

The simulations are performed for both the mid-band (3.5 GHz) and the high-band (26 GHz). Some additional details on the assumed spectrum options, as well as information on the assumed gNodeB and UE antenna types are presented in Table 12.

Parameters	**Frequency band**	
	26 GHz	**3.5 GHz**
Number of gNB Tx antennas	2	2
Number of UE Rx antennas	2	2
gNodeB antenna type	Directional (tri-sector)	Directional (tri-sector)
UE antenna type	Omnidirectional	Omnidirectional
gNodeB transmit power (dBm)	27	27
TDD pattern	DUDU	P1: DDDDDDDSUU P2: DDDSU
Bandwidth (MHz)	100	50
Sub-carrier spacing (kHz)	60 and 120	30
Mini slot size (symbols)	7	2, 4, 7 and 14

Table 12. Parameters for the assumed frequency bands

The slot size is always equal to 14 orthogonal frequency division multiplexing (OFDM) symbols for the eMBB traffic while the mini-slot based transmission is considered for the URLLC traffic. Regarding the TDD configurations, the TDD patterns at 3.5 GHz are based on assumption of SCS equal to 30 kHz and a slot size equal to 14 symbols, resulting in a slot duration of 0.5 ms. For example, the TDD pattern P1 corresponds to a period of 5 ms and if the mini-slot of 7 symbols is considered, the number of DL transmission opportunities (slots) becomes equal to 14 during this period (i.e., two transmission opportunities per each "D"-slot). The high band TDD configuration is more flexible where no reference SCS is considered and where the TDD pattern consists of 7 symbols long downlink and uplink slots.

Table 13 presents the obtained latency results for the high-band (26 GHz). Values for both the average latency, corresponding to the average value of the packet transmission delay *T* across the URLLC users, the minimum occurred latency and the maximum occurred latency are shown. It should be noted that the minimum and maximum latency values are extracted from the instantaneous latency distribution as presented in Figures 24 and 25. The time axis is divided into short intervals of 0.05 ms duration and the number of received packets during each interval is evaluated. The minimum latency corresponds to the lower bound of the time interval relative to the first occurred bar in the distribution while the maximum latency is the upper bound of the time interval of the last occurred bar.

Number of gNodeBs	**SCS (kHz)**	**Minimum latency (ms)**	**Average latency (ms)**	**Maximum latency (ms)**
1	60	0.4	0.45	1.2
	120	0.2	0.3	1.1
2	60	0.4	0.43	1.2
	120	0.2	0.25	1.1
4	60	0.4	0.47	1.2
	120	0.2	0.27	1.1

Table 13. Average and maximum downlink latency at 26 GHz

The main conclusions that can be drawn from the latency results at 26 GHz are:

- The average latency values differ only slightly between the different deployments. The difference is mainly related to the available resources and the radio conditions. For example, the average latency decreases when deploying two gNodeBs instead of one because there are more scheduling occasions (for SCS of 60 kHz the average latency decreases from 0.45 ms to 0.43 ms and for SCS of 120 kHz it decreases from 0.3 ms to 0.25 ms). However, if the number of gNodeBs is increased to four, the average latency increases compared to a deployment with two gNodeBs due to the higher level of inter-cell interference (for SCS of 60 kHz the average latency increases from 0.43 ms to 0.47 ms, while for SCS of 120 kHz it increases from 0.25 ms to 0.27 ms). This is better highlighted by the latency distributions shown in Figure 24, where SCS of 60 kHz is assumed for the figure on the left and SCS of 120 kHz is assumed for the figure on the right. There are two main things impacting the performance: the lack of resources and the level of the inter-cell interference. In the case with only one gNodeB, the lack of resources is explaining the decreased number of received packets with very low latencies (compared to deployments with two and four gNodeBs). The number of received packets with high latencies increases when two or four gNodeBs are deployed because of the higher level of inter-cell interference.
- From the latency point of view the use of SCS of 120 kHz and mini-slot size of 7 OFDM symbols (OS) might be relevant for the scenarios with very stringent latency budgets, i.e., smaller than 1 ms. For instance, if the latency budget is equal to 1 ms, the difference between SCS equal to 60 kHz and 120 kHz does not seem to be that large.
- A significant gap is observed between the average and the maximum latency values which suggests that a very limited number of users are suffering from a lack of resources and/or bad radio conditions, resulting in increased waiting times and transmission durations. This is confirmed also by the results for the packet loss probability.

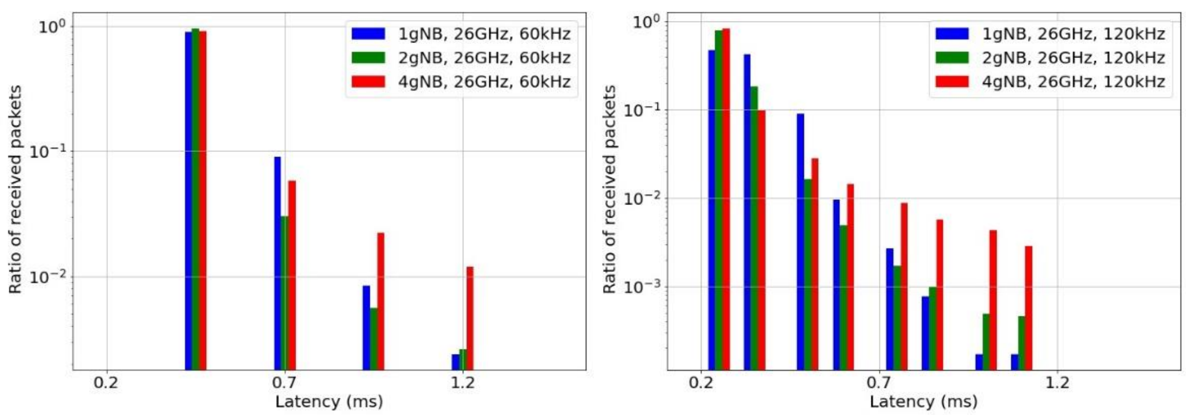

Figure 24. Downlink latency distribution of the received packets at 26 GHz

The packet loss probability is defined as the ratio between the number of generated packets and the number of dropped packets considering two different packet loss possibilities:

- Latency drop, which occurs when a packet transmission delay is higher than the predefined latency budget. The presented results are estimated for latency budget of 1 ms.

- Buffer drop, which occurs when the waiting time of a given packet in the buffer, corresponding to the gNodeB processing delay plus the alignment delay, exceeds the latency budget.

The latency drop and the buffer drop are quite related because if the buffer constraint is not implemented in the simulator and packets with buffer delays exceeding the latency budget would be transmitted, they would in any case be dropped due to the latency constraint. Hence, the packets dropped in the buffer can be considered as a subgroup of the packets dropped due to the latency. In both cases, the packet drop is impacted by the waiting delay in the buffer which occurs as a result of two factors. The first one is related to the capacity where the packet is dropped if the number of available RBs is not sufficient. This corresponds to the case where the number of required RBs given by the selected MCS achieving the target BLER is higher than the number of available RBs, and which can be avoided by increasing the bandwidth. The second one is related to the channel quality (or coverage), where the target BLER cannot be achieved with any of the MCSs, forcing the packets to wait in the buffer until the channel quality becomes good enough, which is not guaranteed in the case of a high level of inter-cell interference.

Table 14 presents the average packet loss probability calculated across all 280 URLLC users and the maximum packet loss probability, which corresponds to the packet loss probability of the worst URLLC user. Based on the latency distribution and the average packet loss probabilities, it can be concluded that the significant percentage of the lost packets are dropped in the buffer. The average values increase with respect to the number of deployed gNodeBs since the level of the inter-cell interference as well as the waiting time in the buffer are increased. However, there is a large gap between the average and the maximum packet loss probabilities. For example, with one gNodeB and SCS of 120 kHz the maximum packet loss probability is equal to 0.024, which means that at least one URLLC user has lost 2.4% of the generated packets. Since the corresponding average packet loss probability is as low as 10^{-4}, it can be concluded that only a very limited number of URLLC users are suffering from coverage problems. By increasing the number of gNodeBs, the overall coverage is improved but at the same time the number of users experiencing high levels of the inter-cell interference is increased. This becomes even more evident when the network contains four gNodeBs. In that scenario, some users drop more than 80% of generated packets due to the high level of inter-cell interference. The results in Table 14 indicate also that the use of SCS of 120 kHz leads to a lower packet loss probability compared to SCS of 60 kHz. This is due to the larger number of transmission opportunities offered by the higher SCS, and hence, a shorter TTI.

Number of gNodeBs	SCS (kHz)	Average packet loss probability	Maximum packet loss probability
1	60	0.0035	0.2
	120	0.0001	0.024
2	60	0.0056	0.3
	120	0.001	0.072
4	60	0.058	0.96
	120	0.03	0.86

Table 14. Average and maximum downlink packet loss probabilities at 26 GHz

As a summary for the high-band, the results for the latency and the packet loss probability indicate that even though the applied TDD pattern would in theory be able to support the downlink latency requirement of 1 ms, some individual users are in outage, i.e., not able to fulfill the requirements for both the latency and the reliability. This can be either due to insufficient coverage, lack of resources (i.e., too high load) within the serving cell, or too high level of the inter-cell interference. Potential solutions could include a relaxed latency requirement, the use of beamformed gNodeB antennas, or some other means to enhance the coverage and/or capacity without increasing the level of the inter-cell interference too much, see also the related discussion in the previous sections (Section 5.2.1 and Section 5.2.2).

Regarding the mid-band, only the deployments with two or four gNodeBs have been considered because of the limited bandwidth of 50 MHz. Table 15 presents the latency results for 3.5 GHz while Figure 25 shows an example of the latency distribution (assuming 2 OS mini-slots). Looking at the results, the following conclusions can be drawn:

- The latency values differ slightly between the deployments with two and four gNodeBs. This is due to the differences in the inter-cell interference levels, which affect also the tail of the latency distribution.
- There is a significant difference in the results for the different TDD patterns as they impact directly the alignment delay and the transmission opportunities as shown in the latency distribution.
- Unlike for high-band, the latency budget is always exceeded not because of the inter-cell interference or the bad radio conditions, but because of the used TDD pattern, where some packets have to wait during the special and the uplink symbols for the first downlink transmission opportunity. The latency results show that for both TDD pattern P1 and P2, the maximum latency values are always the same because the worst case alignment delay that can be encountered is the same for both.

Number of gNodeBs	Mini slot size (symbols)	TDD Pattern	Minimum latency (ms)	Average latency (ms)	Maximum latency (ms)
2	2	P1	0.2	0.35	1.15
		P2	0.2	0.44	1.15
	4	P1	0.35	0.52	1.25
		P2	0.35	0.6	1.25
	7	P1	0.55	0.75	1.35
		P2	0.55	0.77	1.35
	14	P1	1.05	1.19	1.6
		P2	1.05	1.19	1.6
4	2	P1	0.2	0.38	1.15
	4	P1	0.35	0.53	1.25
	7	P1	0.55	0.77	1.35
	14	P1	1.05	1.2	1.6

Table 15. Average and maximum downlink latency at 3.5 GHz

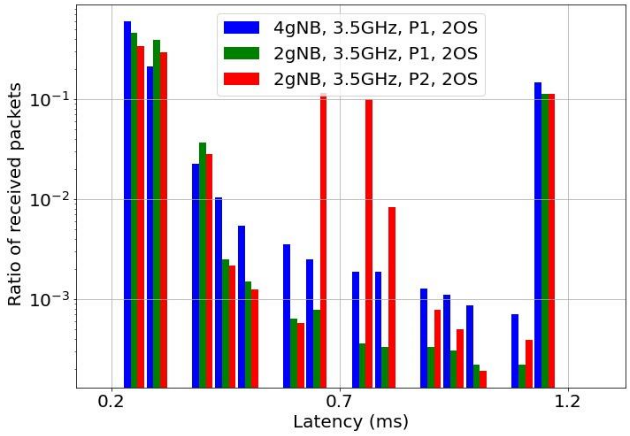

Figure 25. Downlink latency distribution of the received packets for 2 OS mini-slot at 3.5 GHz

The results for the packet loss probability presented in Table 16 confirm the impact that the high level of the inter-cell interference has on the maximum packet loss probability in the case of the deployment with four gNodeBs. In general, this suggests that to make the network densification efficient, it should be based on elaborated techniques (e.g., beamforming) or other antenna types to limit the level of the inter-cell interference. Table 16 shows that the maximum packet loss probabilities obtained with TDD pattern P2 are higher than the ones obtained with TDD pattern P1 which is due to the different number of available downlink slots during the same time period. Also, the packet loss probability is always very high which suggests that the mid-band TDD is better suited for use cases with relaxed latency requirements. Furthermore, the TDD pattern should be adapted according to the uplink and downlink load and the latency budget.

Number of gNodeBs	Mini slot size (symbols)	Pattern	Average packet loss probability	Maximum packet loss probability
2	2	P1	0.19	0.34
		P2	0.19	0.39
	4	P1	0.20	0.39
		P2	0.20	0.48
	7	P1	0.4	0.6
		P2	0.41	0.65
	14	P1	1	1
		P2	1	1
4	2	P1	0.22	0.99
	4	P1	0.24	0.99
	7	P1	0.44	0.99
	14	P1	1	1

Table 16. Average and maximum downlink packet loss probabilities at 3.5 GHz

In general, the results for the mid-band in Table 15, Table 16 and Figure 25 are well in line with the findings in Section 5.2.1, indicating that the stringent downlink latency requirement of 1 ms cannot be supported with the eMBB-optimized TDD patterns, such as the patterns P1 and P2 assumed in this study. However, it is likely that URLLC services with relaxed latency requirements could be supported, as suggested by the results in Section 5.2.1.

As mentioned, a part of the bandwidth is reserved for the eMBB traffic in order to study the feasibility of combining eMBB and URLLC traffic and to evaluate the capacity that can be offered to the eMBB users. Figure 26 presents the average and the maximum throughput (considering all the eMBB users) for five different configurations. Just as the URLLC performance, also the eMBB capacity is improved when deploying two gNodeBs and it is affected by the TDD pattern because of the number of downlink transmission opportunities. Thus, the eMBB traffic can be handled with a minimum of quality of service with respect to the URLLC service requirements.

As a summary, when the results for the different network configurations are compared with each other, the main conclusion is that the high-band spectrum is very suitable for resource sharing between eMBB and URLLC services with stringent latency requirements since it can offer both more bandwidth and more frequent transmission opportunities. The mid-band deployment may be a good option for use cases with relaxed latency requirements.

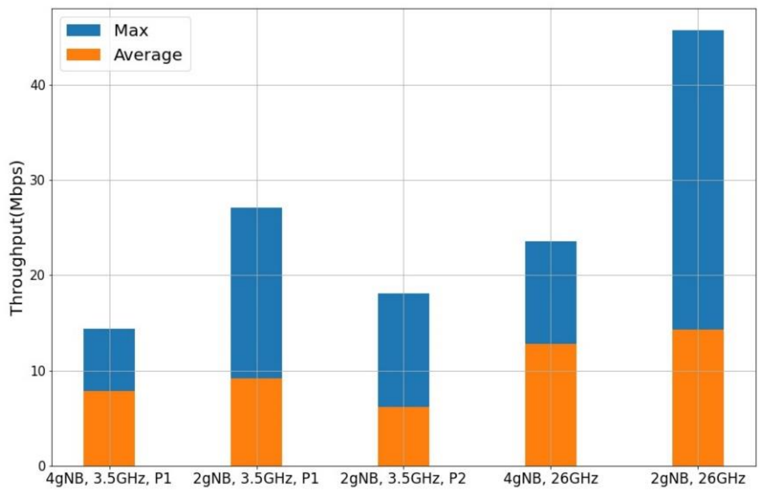

Figure 26. The average and the maximum downlink throughput for the eMBB users

5.3 Co-existence evaluations

5.3.1 Co-existence between a public macro eMBB network and a non-public factory URLLC network

In order to better understand the feasibility of the co-existence scenario between a public eMBB macro network and a non-public URLLC factory network, the impact of the inter-network interference on both the URLLC network and the eMBB network has been evaluated with system-level simulations. The evaluations have considered different kinds of radio network deployment options and features to find out which of them would be feasible for each scenario.

The evaluations have been conducted in an area of 1500 m by 1500 m, which consists of a two-tier overlaid macro network and a factory network, as shown in Figure 27. The macro network consists of seven tri-sector sites with beamformed antennas, providing eMBB service to the UEs. The macro sites are 25 m high and the inter-site distance is equal to 500 m. The factory, with a size of 120x50x10 m^3, is located close to the macro cell border, at a distance of 200 m from the closest macro site. By default, the factory network contains three omnidirectional ceiling-mounted gNodeBs offering a URLLC service to the UEs. The factory is surrounded by an eMBB traffic hotspot ("impact area") to ensure that the factory network is able to cope with the worst-case scenarios. Due to the eMBB traffic hotspot, the macro gNodeBs close to the factory will operate with a higher level of resource utilization compared to the surrounding macro cells and will, therefore, generate a higher level of inter-network interference towards the factory.

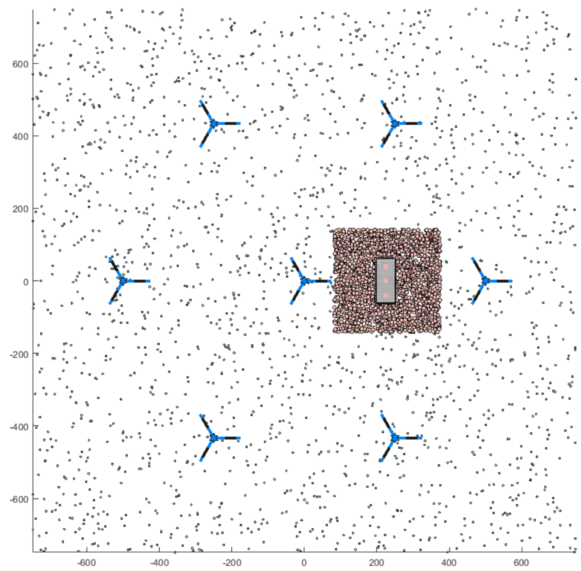

Figure 27. Layout of the co-existence scenario between a public macro network with 7 tri-sector sites and a non-public factory network. The factory is located at a distance of 200 m from the closest macro site and it is surrounded by an eMBB traffic hotspot

The 3GPP Urban Macro (UMa) propagation model is assumed for the links between the macro gNodeBs and the outdoor eMBB UEs, and the 3GPP Indoor Hotspot (InH) model for the links between the factory gNodeBs and the URLLC UEs [3GPP19-38901]. Furthermore, the path losses between the macro gNodeBs and the UEs or gNodeBs inside the factory are calculated as a combination of the UMa propagation model, wall penetration loss and an indoor loss. Finally, the path losses between the factory gNodeBs or UEs and the outdoor eMBB UEs are calculated as a combination of the 3GPP Urban Micro (UMi) propagation model [3GPP19-38901], wall penetration loss and an indoor loss. The wall penetration loss consists of two parts: a constant value for a perpendicular penetration, assumed to be equal to 14 dB[9], and an additional loss depending on the grazing angle [SEF+14]. Finally, the indoor

[9] A wall penetration loss of 14 dB corresponds to an 18 cm thick slab of concrete [ITU15-P2040]. At the same time it is clearly lower than the loss for a solid concrete wall given in [SEF+14], equal to 21 dB for 4 GHz.

loss is expressed as $D \cdot d_{in}$, where D is 0.5 dB/m as in [3GPP19-38901] and d_{in} is the travelled indoor distance.

The main system-level simulation parameters are listed in Table 17. Details of the assumed URLLC link-level and reliability models can be found in [SKA+18] and [AHK+19]. The URLLC traffic consists of packets with a size of 32 bytes, and during the simulations the offered URLLC traffic within the factory has been fixed to 40 Mbps for both the downlink and the uplink. The URLLC users have been assumed to be successfully served if they can fulfill the reliability requirement of 99.999% within a one-way latency bound of 1 ms. Furthermore, the target for the service availability has been set to 100%, which means that the service requirements have to be fulfilled for each location within the factory floor. The total offered eMBB downlink traffic volume has been assumed to be equal to 150 Mbps within the impact area surrounding the factory, and 200 Mbps for the rest of the system area. Assuming the 3:1 traffic split, the corresponding uplink traffic volumes are equal to 50 Mbps and 67 Mbps, respectively.

Parameter	eMBB network	URLLC network
Frequency [GHz]	4	4
Channel bandwidth [MHz]	50	50
Total offered traffic within the system area [Mbps]	350 (DL) 117 (UL)	40 (DL) 40 (UL)
Traffic ratio DL:UL	3:1	1:1
Sectors per site	3	1
gNodeB transmit power [dBm]	50	30
UE transmit power [dBm]	23	23
gNodeB noise figure [dB]	5	5
UE noise figure [dB]	9	9
Max gNodeB antenna element gain [dBi]	8	2
gNodeB antenna array (V x H x (Vs x Hs x Ps)	8x8x(1x1x2)	Omnidirectional
UE antenna	Isotropic (0 dBi)	Isotropic (0 dBi)
Beamforming scheme	Long-term wideband eigen beamforming	No beamforming
Uplink power control setup	Target SNR = 10 dB α = 0.8	Target SNR = 10 dB α = 0.8

Table 17. Main system-level simulation parameters

This study evaluates the performance of two different TDD co-existence scenarios:

- Unsynchronized TDD (uTDD), where the macro network follows a slot-based DDDU pattern, while the factory network follows a half-slot-based DUDU pattern.
- Synchronized TDD (sTDD), where both networks follow a half-slot-based DUDU pattern.

The slot borders have been assumed to be aligned for both TDD configurations. It should be highlighted that from the wide-area eMBB performance point of view, a more realistic alternative would have been to synchronize the networks to a common eMBB-optimized pattern, i.e., the slot-based DDDU pattern, as depicted in Figure 7. However, as already discussed in Section 5.2.1 and demonstrated by the results in Table 18 for the isolated factory network assumed for this study, the downside of assuming the DDDU pattern for the URLLC network is that it is not able to satisfy the stringent latency requirement of 1 ms and would result in considerably worse maximum system capacity values even for the more relaxed latency requirements, assuming a balanced offered traffic between the downlink and the uplink.

Latency	Reliability	DUDU 7	DDDU 14
1 ms	99.999%	44 Mbps	-
	99.9%	84 Mbps	-
3 ms	99.999%	146 Mbps	28 Mbps
	99.9%	177 Mbps	46 Mbps
5 ms	99.999%	173 Mbps	63 Mbps
	99.9%	188 Mbps	84 Mbps

Table 18. Maximum URLLC system capacity in terms of the offered traffic volume for two different TDD patterns, a half-slot based DUDU and a full-slot based DDDU, assuming an isolated factory deployment, different URLLC service requirements and a balanced traffic between the uplink and the downlink.

In general, this co-existence scenario can be divided into two different sub-scenarios: a) factory does not contain any co- or adjacent channel macro users, and b) factory can contain co- or adjacent channel macro users. When the factory does not contain any co- or adjacent channel macro users, the situation is quite straightforward; an increased isolation between the networks, e.g., in form of a frequency separation, a separation distance or a wall penetration loss, will improve the co-existence performance. This kind of co-existence deployment has been discussed and evaluated in more detail for example in [CHT19], where it has been demonstrated how the required wall penetration loss will depend on both the separation distance, and the spectrum allocation between the networks. However, there was only a minor positive impact of synchronized TDD, since for both the unsynchronized and the synchronized TDD the main source of inter-network interference was the downlink transmissions from high-power macro gNodeBs interfering either the factory gNodeBs (uTDD) or the factory UEs (sTDD).

Considering the overlaid two-tier deployment described above, the results in Figure 28(a) demonstrate clearly that the assumed wall penetration loss of 14 dB is not able to provide a sufficient level of isolation if the unsynchronized networks are operating on the same channel. In order to reach the desired service availability of 100%, the wall penetration loss should be equal to 35 dB or more. However, since the ACIR will attenuate the inter-network interference by approximately 30 dB (see the discussion in Section 3.3.1), it can be stated that the wall penetration loss of 14 dB is clearly sufficient for an adjacent channel deployment between the networks.

The required level of isolation can be reduced by making the factory network more dominant within the factory floor. This can be achieved for example by increasing the downlink and uplink transmission powers, or by densifying the factory network. The impact of the more aggressive uplink power control, i.e., an increased SNR target, is shown in Figure 28(a). As can be seen, if the SNR target is increased from 10 dB to 20 dB, the required wall penetration loss is reduced from 35 dB to 29 dB. In general, an increased SNR target will improve the uplink co-existence performance as long as the transmission powers are below the maximum UE transmission power. However, when more and more users become power-limited, the uplink performance becomes limited by the (intra-network) inter-cell interference instead of the inter-network interference. When it comes to the impact of factory network densification, the results in Figure 28(b) demonstrate that the network densification does not improve the overall situation. This is due to the fact that even though the factory network becomes more dominating over the macro network, the overall URLLC performance becomes limited by the (intra-network) inter-cell interference instead. As discussed in Section 5.2.2, a solution could be to deploy gNodeBs with directional or beamformed antennas, or rely on an active distributed antenna system.

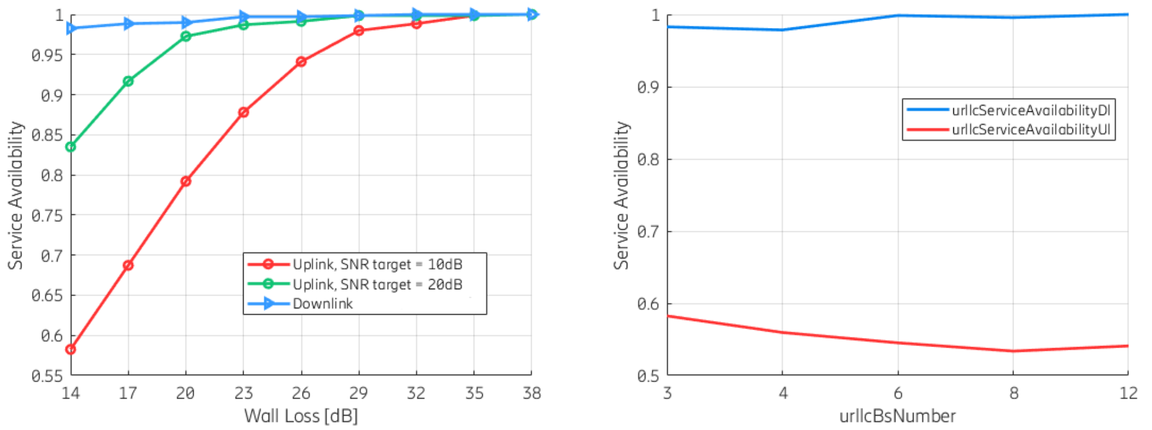

Figure 28. The impact of a) wall penetration loss and the increased uplink transmission power, and b) factory network densification on the URLLC service availability, assuming a co-channel deployment with unsynchronized TDD and no eMBB traffic inside the factory

Since the overall URLLC capacity is in this case limited by the uplink, one possible way to improve the co-existence situation would be to synchronize the networks. The level of the uplink inter-network interference from the macro network towards the factory gNodeBs becomes very low with synchronized TDD since it originates from the outdoor macro UEs. Furthermore, keeping in mind that in this study the synchronized TDD is realized by changing the TDD pattern from DDDU to DUDU in the macro network, the average uplink utilization of the macro network is reduced due to the increased amount of allocated time domain resources, resulting in an even lower level of the uplink inter-network interference towards the factory gNodeBs. At the same time, the (macro) downlink gets less time domain resources resulting in a higher resource utilization and thus, a higher level of inter-network interference towards the factory, making the URLLC downlink performance a bit worse compared to the unsynchronized TDD. As a summary, the URLLC performance becomes downlink-limited, and a wall penetration loss of at least 30 dB is still required to reach the desired URLLC service availability of 100%. Finally, from the macro network point of view, synchronized TDD (realized as the changed TDD pattern within the macro cells) will have a negative impact on the downlink performance but will considerably improve the uplink performance.

The co-existence scenario becomes very challenging if active co- or adjacent channel users served by the overlaid macro cells can be located inside the factory. The main difference is that the wall penetration loss is no longer helping to mitigate the uplink inter-network interference from the visiting macro users towards the factory gNodeBs and UEs. From the macro network point of view, the users entering the factory are expected to experience greatly reduced throughputs. This is mostly due to the impact of the wall penetration loss, which will reduce the received signal power levels and at the same time increase the level of the downlink inter-network interference caused by the factory gNodeBs and UEs. In all, it will be quite safe to assume that a co-channel deployment between the networks will not be feasible due to the very high level of near-far interference between the factory gNodeBs and the macro UEs located inside the factory. Hence, the remaining question is then, if an adjacent channel deployment between the networks would be feasible under some conditions. In order to evaluate that, a set of system-level simulations have been performed, where a part of the eMBB traffic, 1 Mbps or 10 Mbps, has been moved from the impact area into the factory.

Evaluation results for the URLLC users are shown in Figure 29. As can be seen, both the URLLC downlink and uplink can reach 100% service availability when the factory does not contain any public macro traffic (baseline deployment). With a low level of macro traffic inside the factory, URLLC downlink can still reach 100% service availability with both unsynchronized and synchronized TDD. However, with a higher level of factory eMBB traffic, synchronized TDD can still reach 100% service availability, but unsynchronized TDD cannot. In the case of URLLC uplink, the desired 100% service availability is not reached even with the low level of factory eMBB traffic. Hence, it is clear that the co-existence problems are to a large extent related to the inter-network interference caused by the uplink transmissions of the macro users located inside the factory.

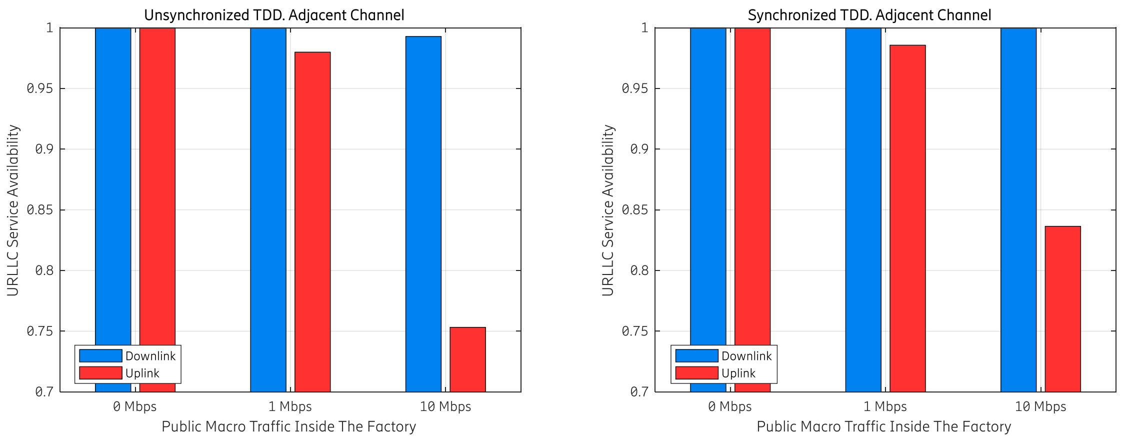

Figure 29. URLLC service availability, assuming an adjacent channel deployment and different levels of public macro traffic inside the factory

The results in Figure 30 demonstrate that the performance of the macro users located inside the factory is affected for both the unsynchronized and synchronized TDD. Compared to the baseline deployment (without any macro users inside the factory), the downlink is affected mainly by the inter-network interference from the factory gNodeBs and in the case of unsynchronized TDD also from the

URLLC UEs. At the same time the uplink is mainly affected by the wall penetration loss, while the inter-network interference from the adjacent channel URLLC factory network does not have any noticeable impact on the uplink performance of the macro users.

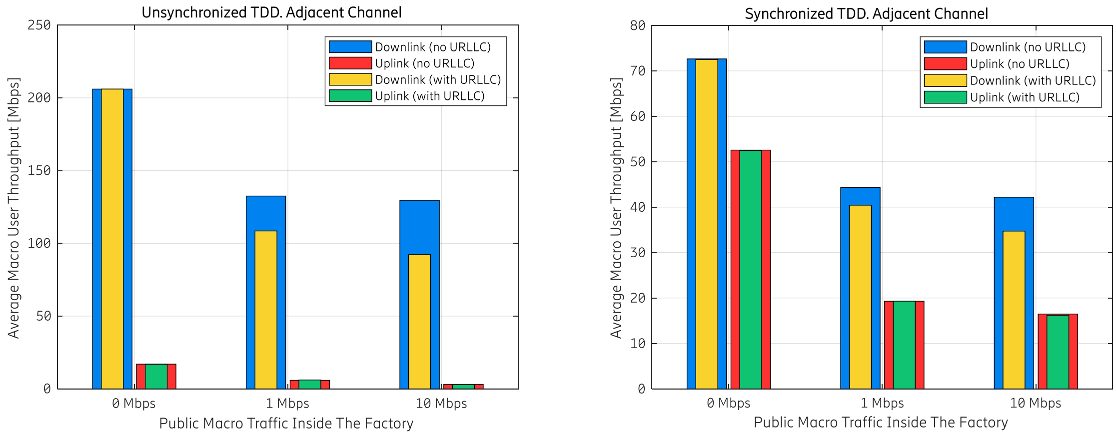

Figure 30. Average macro user throughput, assuming an adjacent channel deployment and different levels of public macro traffic inside the factory.

As a summary, the results in Figure 29 suggest that even with an adjacent channel deployment between the public macro network and the non-public URLLC network the level of the inter-network interference can be too high to guarantee an acceptable URLLC performance even with a synchronized TDD deployment. A solution to the problem could be to allow the visiting macro users to be served by factory gNodeBs, either with the help of a shared RAN, or by deploying a separate public factory network. However, in both cases the public and the non-public factory network would have to be synchronized, or some other actions should be taken to avoid the cross-link interference between the networks. Alternatively, if an unsynchronized TDD is desired instead, the networks should be deployed on isolated frequencies.

5.3.2 Co-existence between neighboring non-public factory URLLC networks

When it comes to the co-existence between neighboring non-public networks, one of the main questions is if an uncoordinated operation on the same frequency channel is feasible or not. To better understand the required conditions for such co-existence scenario, system-level simulations have been run to evaluate the impact of the inter-network interference between two neighboring factories, see Figure 31. The factories are assumed to have a size of 120x50x10 m^3 and they are separated by a distance *D*. Furthermore, the factories are assumed to contain randomly placed metallic blockers with a size of 3x3 m^2 and a height in between 2 and 7 m. The blockers are assumed to cover approximately 40% of the factory floor. Path losses both within a factory and between the factories are calculated with the help of a 3D ray-tracing tool taking into account the impact of both the 3D blockers and the outer walls, and considering the different propagation alternatives (direct path, reflection, diffraction and scattering) when calculating the overall path loss between two nodes. Finally, two different outer wall penetration losses (for perpendicular penetration) are considered: 14 dB (“concrete wall”) and 28 dB (“heavy concrete wall”).

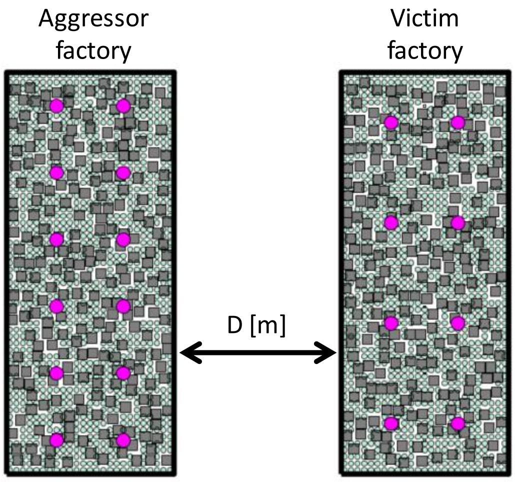

Figure 31. Assumed deployment scenario with two neighboring factories. The aggressor factory contains 12 omnidirectional gNodeBs, while the victim factory contains 8 omnidirectional gNodeBs.

A worst-case co-existence scenario is assumed, meaning that the aggressor factory is assumed to contain a dense factory network with 12 ceiling-mounted omnidirectional gNodeBs. Furthermore, the network is assumed to be almost fully-loaded: depending on the evaluated scenario, the average cell resource utilization is approximately equal to 76-81% for the downlink and 96-98% for the uplink. The aggressor network is assumed to follow a balanced TDD pattern (DUDU) based on full slots.

The evaluations focus on the URLLC performance within the victim factory where the URLLC service is assumed to have a latency requirement equal to 1 ms and a reliability requirement equal to 99.999%. Furthermore, the desired service availability is equal to 100% throughout the whole factory floor. In order to be able to satisfy the stringent QoS requirements, a half-slot-based TDD pattern (DUDU) is assumed. The victim factory network consists of 8 ceiling-mounted omnidirectional gNodeBs, which was found to be the minimum number of gNodeBs to be able to provide the desired service availability in the case of an isolated factory. A summary of the other simulation assumptions is provided in Table 19.

Parameter	Value
Frequency [GHz]	3.5
Channel bandwidth [MHz]	50
Duplexing mode, pattern	TDD, half-slot DUDU (victim) TDD, full-slot DUDU (aggressor)
gNodeB transmit power [dBm]	30
UE transmit power [dBm]	23
gNodeB noise figure [dB]	5
UE noise figure [dB]	7
gNodeB antenna (gain)	Omnidirectional (2 dBi)
UE antenna (gain)	Isotropic (-3 dBi)
Uplink power control setup	SNR target = 10 dB α = 0.8

Table 19. Main system-level simulation parameters

Results for the observed service availability and the relative system capacity as a function of the separation distance *D* are shown in Figure 32. As can be noticed, the required separation distance between the factories is less than 25 m for all the other cases except for the uplink with 14 dB wall penetration loss, which seems to require a very large separation distance of 1.5 km. However, if the wall penetration loss is increased to 28 dB, the performance of the victim network is no longer affected by the aggressor network. This also means that an uncoordinated adjacent channel deployment between the neighboring factory networks would be feasible even with the 14 dB wall penetration loss due to the additional isolation provided by the adjacent channel attenuation.

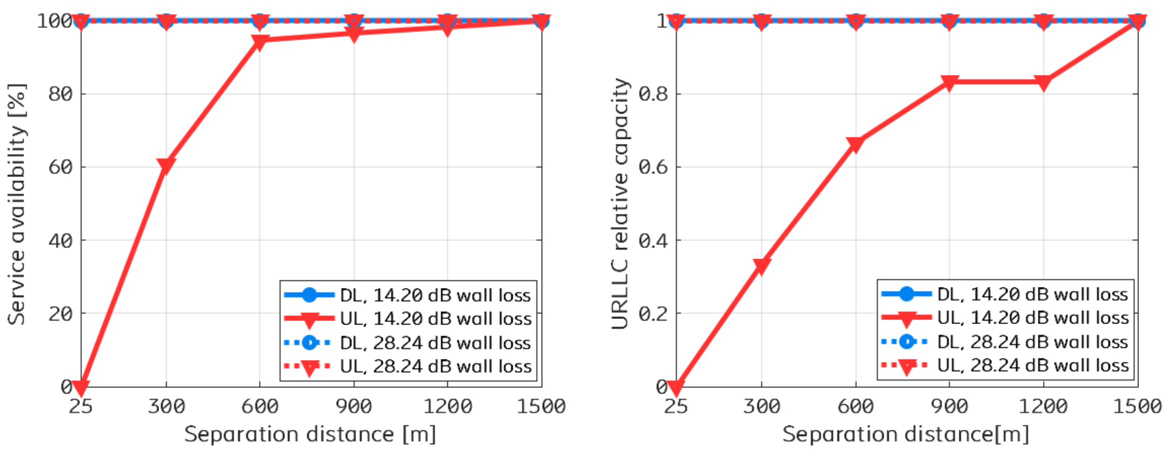

Figure 32. URLLC service availability and relative system capacity within the victim factory as a function of the separation distance between the factories.

There are a few main reasons why the uplink is much more sensitive to the inter-network interference than the downlink, even though the main source of interference is the same for both, i.e., the gNodeBs within the aggressor factory. To start with, the level of the received inter-network interference is higher at the victim gNodeBs compared to the victim UEs. This is due to the different propagation conditions: a line-of-sight propagation over the clutter for the gNodeB-to-gNodeB links and to a large extent non-line-of-sight propagation for the gNodeB-to-UE links. Furthermore, there is a 5 dB difference in the combined antenna gains. Secondly, there is also a difference when it comes to the impact of the received inter-network interference on the SINRs, and hence, the performance of the victim network. The downlink is to a large extent limited by the high level of intra-network interference, which means that the inter-network interference will have only a small impact on the SINR of the more “noise-limited” users, while the “interference-limited” users are more or less unaffected by the inter-network interference. However, the situation looks completely different for the uplink, since the applied power control is limiting the level of both the received signal power and the intra-network interference. Now, the level of the received inter-network interference is roughly at the same level, or even higher than the intra-network interference. Therefore, it has a significant impact on the uplink SINRs and the corresponding network performance.

The co-existence situation can be improved by making the uplink power control more aggressive within the victim factory. For example, if the SNR target is increased from 10 dB to 20 dB, the required separation distance can be reduced from 1.5 km to 600 m to secure the same victim network performance as an isolated factory would have with SNR target equal to 10 dB. Another simple way to improve the co-existence situation would be to utilize directional or beamformed gNodeB antennas in either one of the factories, or in both of them. By doing so, the level of the inter-network interference between the gNodeBs could be lowered, improving the uplink performance within the victim factory. Finally, a synchronized TDD deployment would also be an efficient way to improve the co-existence performance in the scenarios with non-overlapping coverage areas, where the near-far interference will not be causing any major problems. For example, if both factories assume a synchronized half-slot-based DUDU pattern, the obtained simulation results indicate that the required separation distance would become less than 25 m even with the 14 dB wall penetration loss. Assuming a synchronized TDD deployment would most likely be a simple solution for the specific co-existence scenario assumed in this study, where both factories desire to use the 5G NR network to support URLLC services with a balanced traffic between the downlink and the uplink. However, the situation would look quite different if the desired URLLC services would differ between the factories, e.g., a downlink-heavy or a balanced service versus an uplink-heavy service. There, some other means, such as a careful planning of the radio network deployments and agreeing on appropriate emission limits, should be applied to control the level of the inter-network interference.

As a summary, there could be a need to mitigate the inter-network interference between neighboring non-public networks operating on the same frequency channel. With non-overlapping coverage areas, the near-far interference will typically not be causing any major problems, but at the same time the cross-link interference between the gNodeBs can potentially require a large isolation between the networks, e.g., in form of a large separation distance or a high wall penetration loss. An effective way to mitigate the inter-network interference would be to synchronize the TDD patterns. However, if service-optimized (and NPN-specific) TDD patterns are preferred instead, some other means, such as a careful planning of the radio network deployments and agreeing on appropriate emission limits, should be applied to control the level of the inter-network interference.

6 Conclusion

5G NR is a prime enabling technology for facilitating the industrial transformation to Industry 4.0, providing wireless connectivity in and around the factory, based on a global standard with global economy of scale. It can connect a variety of industrial devices with different service needs, and it can also provide URLLC to bring wireless connectivity to demanding industrial equipment. A 5G-connected factory is based on a local 5G NR non-public network using licensed spectrum. The non-public network can be provided by an MNO, or it can be deployed and operated standalone by an industrial party in locally leased or licensed spectrum.

This report has discussed the different radio network deployment options for smart manufacturing. It has provided an overview of the different options and the input data necessary to select the most feasible deployment options for the desired industrial 5G scenarios and services. Furthermore, the report has discussed and analyzed the feasibility of the radio network deployment options for the different NPN architecture options and the impact of spectrum options available for the stakeholder deploying and operating the non-public network.

The analysis presented in this report has suggested that the various radio network deployment options and features can significantly impact the performance of an industrial 5G network. For example, the chosen frequency band has an impact on the achievable latency and the maximum system capacity defining the type of URLLC services that the non-public network can support. Low- and mid-band FDD spectrum offers very low latencies and good coverage but quite moderate system capacities due to the limited channel bandwidths. Furthermore, it is likely that the MNOs would like to prioritize those frequencies for wide-area eMBB services, instead of reserving some of them for local network deployments. Mid-band TDD spectrum offers a decent system capacity, but can face challenges with fulfilling the most stringent latency and reliability requirements due to the potential co-existence problems. Finally, high-band TDD spectrum offers low latencies and high system capacity, but typically only limited cell coverage areas.

This report has demonstrated how the applied TDD pattern defines a lower bound for the achievable latency and what kind of an impact it has on the maximum system capacity. The applied TDD pattern is also one of the key factors affecting the co-existence performance between the non-public network and the neighboring TDD networks operating in the same band. Finally, the fact that the latency-critical URLLC services are sensitive to inter-cell interference has a big impact on how the industrial URLLC networks should be densified to improve the coverage and to increase the system capacity. For example, the analysis presented in this report has indicated that the use of an active distributed antenna system can be an efficient way to enhance the coverage, while the system capacity can be enhanced by deploying gNodeBs with directional or beamformed antennas.

In general, the feasibility of the different radio network deployment options, as well as the corresponding radio network performance depends both on the chosen architecture and the stakeholder deploying and operating the non-public network. To start with, an independent standalone non-public network can be deployed and operated by both the MNO and the industrial party. However, an industrial party has considerably less spectrum options available compared to an MNO. For example, it is very likely that the industrial party does not have access to low- or mid-band FDD spectrum, which makes it more difficult for the non-public network to support M-MTC services.

Furthermore, it can be challenging for an industrial party having access to only a single mid-band TDD carrier to be able to resolve all co-existence problems between the overlaid public network and the non-public network without collaborating with the MNO. This collaboration between the MNO and the industrial party can be in form of e.g., a coordination agreement or a RAN sharing agreement. For a completely independent and uncoordinated operation of the non-public network with respect to the overlaid public network, the recommendation is to deploy the non-public network on an isolated frequency.

If the non-public network is deployed in conjunction with a public network, either as a shared RAN, or in particular as a public network integrated non-public network, it is likely that the non-public network is deployed and operated by an MNO. In that case, the MNO can utilize all its spectrum assets, possibly combined with the local spectrum when available, to provide all the required IIoT services. It becomes also straightforward to design, combine or coordinate the overlaid public network and the local non-public network to resolve most of the co-existence problems. However, there could still be a need for the MNO to agree and coordinate with the neighboring MNOs operating in the same band to secure a sufficiently low level of inter-network interference to guarantee the desired URLLC network performance.

Finally, independent of the stakeholder deploying and operating the non-public network, there could be a need to mitigate the interference between neighboring non-public networks. This will be the case if the networks are operating on the same frequency channel (e.g., utilizing a local license in the mid-band) or in particular if the networks contain outdoor small cells. Since in this particular co-existence scenario the communication service areas are non-overlapping, applying a synchronized TDD will be a quite effective way to mitigate the inter-network interference. However, if service-optimized (and NPN-specific) TDD patterns are preferred instead, some other means, such as a careful planning of the radio network deployments and agreeing on appropriate emission limits, should be applied to control the level of the inter-network interference. The national regulator could also define an emission limit as part of the local license conditions, to be valid for the situations when the neighboring NPN operators cannot reach an agreement by themselves.

7 References

[3GPP20-22104] 3GPP TS 22.104, "Service requirements for cyber-physical control applications in vertical domains; Stage 1 (Release 17)", July 2020.

[3GPP20-22261] 3GPP TS 22.261, "Service requirements for the 5G system; Stage 1 (Release 17)", July 2020.

[3GPP20-22804] 3GPP TR 22.804, "Study on Communication for Automation in Vertical Domains (Release 16)", July 2020.

[3GPP20-23501] 3GPP TS 23.501, "System architecture for the 5G System (5GS) (Release 16)", September 2020.

[3GPP20-38101] 3GPP TS 38.101-1, "User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone (Release 16)", June 2020.

[3GPP20-38104] 3GPP TS 38.104, "Base Station (BS) radio transmission and reception (Release 16)", June 2020.

[3GPP17-38802] 3GPP TR 38.802, "Study on New Radio Access Technology Physical Layer Aspects (Release 14)", September 2017.

[3GPP19-38824] 3GPP TR 38.824, "Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC) (Release 16)", March 2019.

[3GPP20-38875] 3GPP TR 38.875, "Study on support of reduced capability NR devices (Release 17)", November 2020.

[3GPP19-38901] 3GPP TR 38.901, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 16)", December 2019.

[3GPP19-R11903446] 3GPP Tdoc R1-1903446, "Scheduling/HARQ/CSI Processing Timeline Enhancements for NR URLLC (Source: Ericsson)", February 2019.

[3GPP20-RP201677] 3GPP Tdoc RP-201677, "Revised SID on Study on support of reduced capability NR devices (Source: Ericsson)", September 2020.

[5GA19-Aut] 5G-ACIA, "5G for Automation in Industry: Primary use cases, functions and service requirements", White Paper, July 2019.

[5GA19-NPN] 5G-ACIA, "5G Non-Public Networks for Industrial Scenarios", White Paper, July 2019.

[5GA20-UC] 5G-ACIA, "Key 5G Use Cases and Requirements From the Viewpoint of Operational Technology Providers", White Paper, May 2020.

[5GA20-Web] 5G-ACIA, "Non-Public 5G Networks for Manufacturing", Web Seminar, August 2020.

[5GS20-D11] 5G-SMART, Deliverable 1.1, "Forward looking smart manufacturing use cases, requirements and KPIs", June 2020.

[5GS20-D52] 5G-SMART, Deliverable 5.2, "5G network architecture options and assessments", November 2020.

[5GS20-Term] 5G-SMART, "5G common terminology", June 2020.

[ABS+20] F. Alriksson, L. Boström. J. Sachs, Y.-P. E. Wang and A. Zaidi, "Critical IoT connectivity: Ideal for time-critical communications", Ericsson Technology Review, June 2020.

[AHK+19] O. Al-Saadeh, K. Hiltunen, K. Kittichokechai, A. Shapin, M. Gerami, H. Asplund, E. Wang, G. Wikström and J. Sachs, "5G ultra-reliable low-latency communication for factory automation at millimeter wave bands", in Proc. IEEE Global Communications Conference (GLOBECOM) 2019, Waikoloa, HI, USA, December 2019.

[BNetza19] Bundesnetzagentur, "Verwaltungsvorschrift für Frequenzzuteilungen für locale Frequenznutzungen in Frequenzbereich 3.700-3.800 MHz", November 2019.

[BNetza20] Bundesnetzagentur, "Spectrum diagram in the areas 700 MHz to 3.8 GHz", January 2020.

[CHT19] U. Challita, K. Hiltunen and M. Tercero, "Performance evaluation for the co-existence of eMBB and URLLC networks: Synchronized versus unsynchronized TDD", in Proc. IEEE Vehicular Technology Conference (VTC) 2019 Fall, Honolulu, HI, USA, September 2019.

[DWL+16] M. Ding, P. Wang, D. López-Pérez, G. Mao and Z. Lin, "Performance impact of LoS and NLoS transmissions in dense cellular networks", IEEE Transactions on Wireless Communications, vol. 15, no. 3, pp.2365-2380, March 2016.

[ERI19-ESS] Ericsson, "Sharing for the best performance", 2019. [online]: https://www.ericsson.com/en/networks/offerings/5g/sharing-spectrum-with-ericsson-spectrum-sharing/download-form

[ERI20-RDS] Ericsson, Radio Dot System. [online]: https://www.ericsson.com/en/portfolio/networks/ericsson-radio-system/radio/indoor/radio-dot-system

[ITU17-M2412] ITU-R M.2412, "Guidelines for evaluation of radio interface technologies for IMT-2020", October 2017.

[ITU15-P2040] ITU-R P.2040, "Effects of building materials and structures on radiowave propagation above about 100 MHz", July 2015.

[KAF+18] F. Kronestedt, H. Asplund, A. Furuskär, D.H. Kang, M. Lundevall and K. Wallstedt. "The advances of combining 5G NR with LTE at existing sites", Ericsson Technology Review, October 2018.

[LSE+19] O. Liberg, M. Sundberg, Y.-P. E. Wang, J. Bergman, J. Sachs, G. Wikström, *Cellular Internet of Things - From Massive Deployments to Critical 5G Applications*, Academic Press, second edition, ISBN: 9780081029022, October 2019. https://www.elsevier.com/books/cellular-internet-of-things/liberg/978-0-08-102902-2

[MMM16-D51] mmMAGIC, Deliverable 5.1, "Initial multi-node and antenna transmitter and receiver architectures and schemes", March 2016.

[MMM17-D52] mmMAGIC, Deliverable 5.2, "Final multi-node and multi-antenna transmitter and receiver architectures and schemes", June 2017.

[NHB+20] M. Norin, R. Högman, M. Buchmayer, G. Lemne, F. Pedersen and A. Zaidi, “5G spectrum for local industrial networks”, Ericsson White Paper, June 2020.

[PTS20] Post- och telestyrelsen (PTS), “Blocktillstånd”, 2020 [online]: https://pts.se/sv/bransch/radio/blocktillstand/

[SEF+14] E. Semaan, F. Harrysson, A. Furuskär and H. Asplund, ”Outdoor-to-indoor coverage in high frequency bands”, in Proc. IEEE Globecom 2014 Workshop – Mobile Communications in Higher Frequency Bands, Austin, TX, USA, December 2014.

[SKA+18] A. Shapin, K. Kittichokechar, N. Andgart, M. Sundberg and G. Wikström, "Physical Layer Performance for Low Latency and High Reliability in 5G”, in Proc. International Symposium on Wireless Communication Systems (ISWCS) 2018, Lisbon, Portugal, August 2018.

[SWA+19] J. Sachs, K. Wallstedt, F. Alriksson and G. Eneroth, “Boosting smart manufacturing with 5G wireless connectivity”, Ericsson Technology Review, February 2019.

[SWD+18] J. Sachs, G. Wikstöm, T. Dudda, R. Baldemair and K. Kittichokechai, “5G radio network design for ultra-reliable low-latency communication”, IEEE Network, vol. 32, issue 2, March-April 2018.

[TRAFICOM20] Finnish Transport and Communications Agency (Traficom), “Frequencies and license holders of public mobile networks”, 2020 [online]: https://www.traficom.fi/en/communications/communications-networks/frequencies-and-license-holders-public-mobile-networks

[ZA15] X. Zhang and J.G. Andrews, “Downlink cellular network analysis with multi-slope path loss models”, IEEE Transactions on Communications, vol. 63, no. 5, pp. 1881-1894, May 2015.

[ZBN+20] A. Zaidi, A. Bränneby, A. Nazari, M. Hogan, C. Kuhlins, “Cellular IoT in the 5G era”, Ericsson White Paper, February 2020.

Appendix

List of abbreviations

3D	Three-Dimensional
3GPP	Third Generation Partnership Project
4G	Fourth Generation Mobile Network
5G	Fifth Generation Mobile Network
5G-ACIA	The 5G Alliance for Connected Industries and Automation
5G-SMART	5G for Smart Manufacturing
ACIR	Adjacent Channel Interference Ratio
ACLR	Adjacent Channel Leakage power Ratio
ACS	Adjacent Channel Selectivity
AD	Analog-to-Digital
APN	Access Point Name
AWGN	Additive White Gaussian Noise
BLER	Block Error Rate
BS	Base Station
CG	Configured Grant
C-MTC	Critical Machine-Type Communication
DA	Digital-to-Analog
DAS	Distributed Antenna System
DL	Downlink
eDRX	Extended Discontinuous reception
eMBB	Enhanced Mobile Broadband
FDM	Frequency Domain Multiplexing
FFT	Fast Fourier Transform
gNB	gNodeB (5G NR base station)
HARQ	Hybrid Automatic Repeat Request
HARQ-ACK	Hybrid Automatic Repeat Request Acknowledgement
ID	Identity
IFFT	Inverse Fast Fourier Transform
IIoT	Industrial Internet of Things
InF-DH	Indoor Factory with Dense clutter and High base station height
InH	Indoor Hotspot
KPI	Key Performance Indicator
LTE	3GPP Long Term Evolution
LTE-M	LTE Machine-type communication
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MIMO	Multiple-Input and Multiple-Output

M-MTC	Massive Machine-Type Communication
MNO	Mobile Network Operator
MOCN	Multi-Operator Core Network
MORAN	Multi-Operator Radio Access Network
MTC	Machine-Type Communication
NB-IoT	Narrowband Internet of Things
NPN	Non-Public Network
NR	3GPP New Radio
NR-U	3GPP New Radio for Unlicensed spectrum
OFDM	Orthogonal Frequency Division Multiplexing
OS	Orthogonal Frequency Division Multiplexing Symbol
PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channel
PHY	Physical layer
PLMN	Public Land Mobile Network
PN	Public Network
PNI-NPN	Public Network Integrated Non-Public Network
PSM	Power Saving Mode
PUSCH	Physical Uplink Shared Channel
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block
RDS	Radio Dot System
RF	Radio Frequency
RRM	Radio Resource Management
RTT	Round-Trip Time
SCS	Sub-Carrier Spacing
SINR	Signal-to-Interference-plus-Noise Ratio
SIR	Signal-to-Interference Ratio
SLA	Service-Level Agreement
SNR	Signal-to-Noise Ratio
sTDD	Synchronized TDD
TDM	Time Domain Multiplexing
TRP	Total Radiated Power
TTI	Transmission Time Interval
UE	User Equipment
UL	Uplink
UMa	Urban Macro
UMi	Urban Micro

URLLC	Ultra-Reliable Low-Latency Communication
uTDD	Unsynchronized TDD

Table 20. List of abbreviations