FAKULTÄT FÜR ELEKTROTECHNIK UND INFORMATIONSTECHNIK
Faculty of Electrical Engineering and Information Technology

System Level Abstraction and Evaluation of Non-Orthogonal Multiple Access Transmissions

MASTER THESIS

to achieve the university degree of

Diplom-Ingenieurin

in the master's degree program

Telecommunications

submitted by

Agnes Fastenbauer

Immatriculation number: 01025266

Fakultät für Elektrotechnik und Informationstechnik
Institute of Telecommunications
at TU Wien

Supervision:
Asst. Prof. Dipl.-Ing. Dr.techn.
Projektass. Dipl.-Ing. Bashar Tahir BSc

Vienna, September 2021

Agnes Fastenbauer

Statement on Academic Integrity

Hiermit erkläre ich, dass die vorliegende Arbeit gemäß dem Code of Conduct Regeln zur Sicherung guter wissenschaftlicher Praxis (in der aktuellen Fassung des jeweiligen Mitteilungsblattes der TU Wien), insbesondere ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel, angefertigt wurde. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.
Die Arbeit wurde bisher weder im In noch im Ausland in gleicher oder in ähnlicher Form in anderen Prüfungsverfahren vorgelegt.

Vienna, September 2021

Agnes Fastenbauer

Abstract

The increasing demand for higher data rates and lower transmission latency requires new transmission techniques to be considered for future wireless networks. Non-orthogonal multiple access (NOMA) techniques promise higher data rates, lower latency and massive user deployments. To analyze the effects of NOMA techniques in networks, a system level abstraction is necessary to perform system level simulations with low complexity. This work introduces an efficient system level abstraction for power domain NOMA and provides an analysis of the effects of the deployment of NOMA techniques in wireless networks.

Contents

1 Introduction

Motivation

The steady increase in demand of mobile data access [1] requires an ever more efficient use of the limited spectrum available for mobile transmission. While orthogonal multiple access (OMA) methods have been exploited in the past, starting with frequency division multiple access (FDMA) followed by time division multiple access (TDMA) and code division multiple access (CDMA), and with orthogonal frequency division multiple access (OFDMA) being introduced in the long term evolution (LTE) standardization [2], NOMA methods have only recently found their way into standardization. In the 3rd generation partnership project (3GPP) standardization, NOMA is introduced as multiuser superposition transmission (MUST) [3], and in advanced television systems committee (ATSC) 3.0, NOMA is introduced as layered-division multiplexing (LDM) [4]. With promising theoretical results of higher system throughput [5] and lower latency [6, 7], NOMA is a candidate for beyond fifth generation (5G) and sixth generation (6G) wireless networks [8].
Simulations are a popular tool for the evaluation of new transmission techniques. As opposed to measurements, simulations allow the investigation of techniques that have not yet been deployed and also allow for the comparison of different techniques under controlled and repeatable conditions [2]. Thus, a number of system level simulators (SLSs) have been developed to examine the performance of wireless networks [9, 10, 11, 12, 13, 14, 15, 16].
With NOMA being a candidate technology for future wireless network generations and system level (SL) simulations being an essential tool to evaluate network performance, the need for a SL abstraction for NOMA transmissions arises. This need is addressed in this work.

Objective

The objective of this work is to develop an efficient SL abstraction for NOMA transmissions in wireless networks allowing to evaluate the behavior of large networks serving NOMA and OMA users.
In section 2, NOMA techniques are put into the historical context of OMA techniques, the current state of the art of different types of NOMA techniques is presented and an overview of the potential advantages of NOMA over OMA is provided. Then, the state of the art of SL simulations and openly available SL simulators is described. In section 3, terms used frequently in this work are defined. The flow of a SL simulation with OMA in the Vienna cellular communications simulators (VCCS) 5G SLS is described in section 4. In section 5, the extensions of the VCCS 5G SLS necessary for NOMA transmissions are described starting with the NOMA user pairing, followed by the scheduling of the

additional NOMA user, including the NOMA power allocation. Then, a SL abstraction for the NOMA transmission with a successive interference cancellation (SIC) receiver is developed, that extends the link quality model (LQM) used in the VCCS 5G SLS. This abstraction is validated for single input single output (SISO), multiple input single output (MISO) and multiple input multiple output (MIMO) transmissions through a comparison with the VCCS 5G link level simulator (LLS) [17]. With the extended version of the VCCS 5G SLS, the effects of the deployment of NOMA techniques in a wireless network are analyzed in section 6. The increase in cell throughput compared to OMA is evaluated for two different scheduling strategies, round robin and best channel quality indicator (CQI) scheduling. Next, the increase in the number of served users and the packet latency is analyzed. In section 6.5, the findings of this work are discussed. Finally, this work is concluded in section 7 and an outlook to future work is given.

2 State of the Art

In this section the state of the art of NOMA techniques and of SL simulations is described. After putting the development of NOMA in the historical context of the development of different OMA techniques, the approaches to NOMA, that have been investigated so far, are described. The potential benefits of NOMA, that are promised by theoretical results are explored before giving an overview of the advantages of SL simulations and the currently openly available SL simulators.

2.1 Non-Orthogonal Multiple Access

History

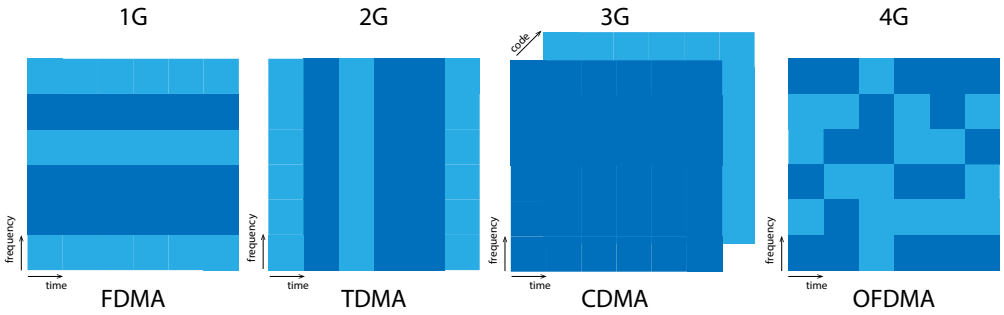


Figure 1: Evolution of orthogonal access methods over wireless generations.

Over the generations of wireless transmission networks, ever more OMA techniques have been introduced to wide spread use in mobile communications. The evolution of OMA techniques is depicted in figure 1, where each color represents a user transmitting over the time frequency resource grid. While the use of FDMA was already introduced in the first generation (1G) of mobile communication networks, TDMA and CDMA were introduced step wise in the subsequent second generation (2G) and third generation (3G) respectively [2]. In the fourth generation (4G), OFDMA was introduced. With the ever increasing demand for data, orthogonal multiplexing techniques are reaching their full potential and non-orthogonal techniques are envisioned to fulfill the demands of future wireless communication networks. The use of NOMA techniques is envisioned for 5G and 6G [18]. For specific applications, NOMA has already found its way into the standardization of networks. In 3GPP release 14 [19], NOMA techniques are specified under the name of MUST for downlink (DL) transmissions and in digital TV NOMA techniques have been introduced as LDM [4].

Different Types of NOMA

Different types of NOMA techniques have been proposed. They can be broadly categorized into power domain NOMA and code domain NOMA. Other types of NOMA that do not fit these two categories exist, one example is spatial division multiple access (SDMA) [20]. Code domain NOMA is inspired by CDMA with non orthogonal spreading sequences. Several different forms of code domain NOMA have been developed, among them are low density signature (LDS) CDMA [21, 22, 23], LDS OFDMA [24, 25] and pattern division multiple access (PDMA) [26].

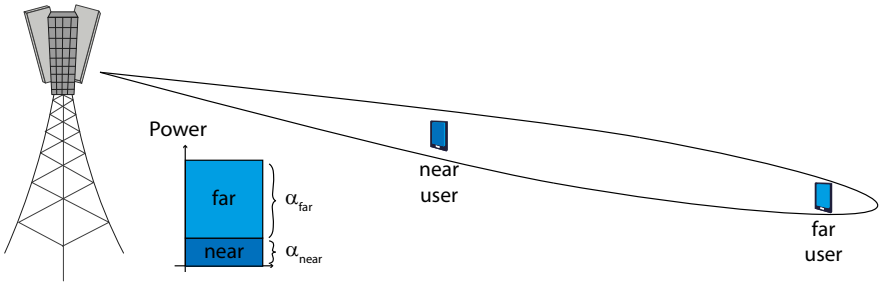


Figure 2: Example of a DL power domain NOMA transmission with two users sharing resources in the power domain.

In power domain NOMA two or more users share the same resources in the power domain. An example for two users in a DL transmission is shown in figure 2. The near user is shown in dark blue. This user will be allocated a small share of the transmit power α_{near} since this user has better channel conditions than the far user shown in light blue. The far user has bad channel conditions and will use the larger share of the transmit power α_{far}. For the far user, the signal of the near user can be treated as additional interference. The near user will have to perform SIC to retrieve its signal. For SIC, the near user first detects the signal of the far user, subtracts it from the received signal and then decodes its own data.
The detection of the far user signal at the near user can represent a security vulnerability [27, 28]. The security of NOMA systems has been addressed in literature, where [27, 29, 30, 31] optimize the security of a NOMA system by adapting the NOMA power allocation to security constraints. In [32], the influence of user mobility on security in NOMA scenarios is investigated. Contrary to the security concerns, [33] finds that NOMA systems can be more secure than OMA systems.
In uplink (UL) power domain NOMA, several users can transmit to the base station (BS) using the same resources and the BS starts the SIC with the signal of the nearest user with the strongest signal and then decodes the signals of the remaining users in descending order of channel quality.
Challenges for the adaptation of NOMA are to find an optimal user pairing for DL [5]

or an optimal decoding order for UL [34]. The mitigation of error propagation in SIC decoding has been addressed in [35]. In case of SIC, if the signal of the first user is detected incorrectly, the detection of all other users will fail too. The problem of grouping users to share transmission resources is aggravated by the lack of accurate channel state information (CSI) in practical wireless networks [36].

Potential Benefits of NOMA

Investigations have shown that the achievable rate region can be enlarged with NOMA [37] and that this enlargement increases the more distinct the channels of the two users are [5]. It is shown in [38], that the achievable rate region cannot be increased through NOMA in comparison to OMA for two users with the same channel conditions, while NOMA has the potential to increase the achievable rate region for a user pair with distinct channel qualities. In terms of quality of service (QoS), the NOMA users can experience lower latency [6, 7] and higher throughput [39, 40]. Since more users can be scheduled [41] at the same time with NOMA, the number of users served can be increased, which leads to higher connectivity. The fairness in the network can also be increased [40] since cell edge users can be scheduled with cell center users avoiding the allocation of resources to cell edge users that cannot exploit the resources effectively, while at the same time avoiding the preferential treatment of cell center users that would be prioritized if only the maximization of the network throughput is considered. In [42], it is shown that the spectral efficiency can be improved by more than 20% with the use of MUST and up to 24% for cell edge users. The energy efficiency can be increased with the deployment of NOMA small cells in a network [43].
In summary, NOMA techniques have the potential to increase network throughput, cell edge throughput, user fairness, spectral efficiency, energy efficiency and user connectivity as well as to reduce latency.

2.2 System Level Simulations

The effects of new transmission techniques in networks can be difficult to evaluate. Deploying new techniques in the network can be prohibitively expensive and can lower the performance of the network, if the new techniques are not as efficient as expected. Additionally, it is often difficult to measure the network performance adequately, since network measurements are expensive and the network traffic that is to be evaluated might not be present in the network. This is especially true, when techniques for future applications are developed, such as data traffic caused by autonomous vehicles. An alternative approach that would allow the evaluation of more experimental techniques would be to develop a testbed that recreates a network. This approach can also be prohibitively expensive. SL simulations on the other hand offer a comparatively cheap

and quick evaluation of new transmission techniques [44]. An even more theoretical approach is the analytical evaluation of the network. Stochastic geometry offers a framework for such analytical evaluations [45]. Due to the heterogeneity of modern wireless networks, analytical approaches might not be tractable and might only be practicable as numerical evaluations for complex scenarios. For these reasons, SL simulations are a popular approach to network evaluations and a number of SL simulation tools have been developed to evaluate wireless networks.

A number of SL simulation tools are commercially available and a smaller number are openly available, such as [10, 12, 13, 14, 16, 46, 47]. [10] offers a 3GPP standard compliant framework for the SL evaluation of massive MIMO techniques. [46] expands the existing ns-3 network simulator [47] for mmWave transmission. Another event driven simulator is presented in [11]. Simulators for vehicular communication are developed in [12] and [16].

The VCCS offer a flexible simulation suite that is available freely for academic use. The modular implementation allows the addition of new features like NOMA transmissions. The simulators are implemented in MATLAB using object-oriented programming (OOP) paradigms. The VCCS offer a 5G link level (LL) simulator [17] and a 5G SL simulator [9], where the 5G LL simulator supports NOMA transmissions according to 3GPP release 14 [19], whereas the 5G SL simulator does not support NOMA transmissions as of yet.

To the best of our knowledge none of the available SL simulation tools support NOMA transmission as of the writing of this thesis.

3 Terminology

In this section, terms that are used throughout this thesis are defined and explained. This section can be used as a reference throughout the reading of this thesis. The terms are listed in alphabetical order.

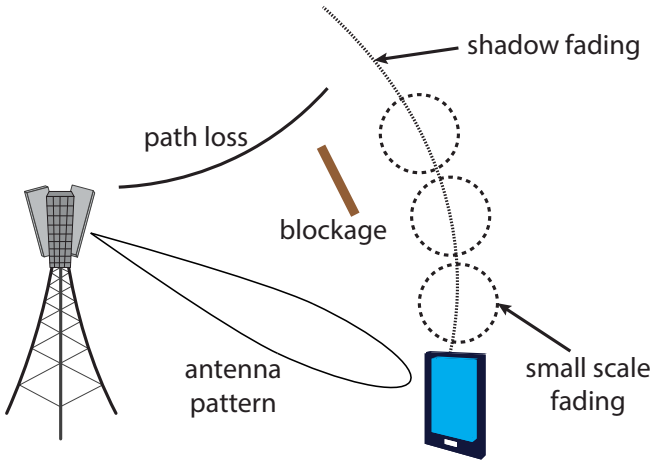


Figure 3: Fading in a wireless channel. The figure illustrates large scale and small scale fading. The path loss describes the distance dependent loss. The wall loss due to blockages and the shadow fading describe the fading due to the large scale geometry of the environment. The antenna gain describes the directivity of the antenna pattern. The small scale fading describes effects due to multi path propagation.

Large Scale Fading

The channel fading, i.e. the attenuation of the signal on a link, can be modeled as a combination of several different fading mechanisms [48]. A coarse classification in large scale and small scale fading can be made. There, the large scale fading models the position dependent fading mechanisms, while the small scale fading models the fading mechanisms that are caused by multi path propagation of wireless signals. The different fading mechanisms are illustrated in figure 3. The **large scale fading**, or **macroscopic fading**, can further be split up in the path loss, antenna gain, and the shadow fading and wall loss. The path loss describes the signal attenuation that is caused by the distance to the transmitter. The shadow fading and wall loss describe the signal attenuation that is caused by the geographical constellation, e.g. buildings, mountains, street canyons. In this work, the antenna gain pattern is also included in the large scale fading, since the attenuation depends on the positions of the transmitter, the receiver and blockages.

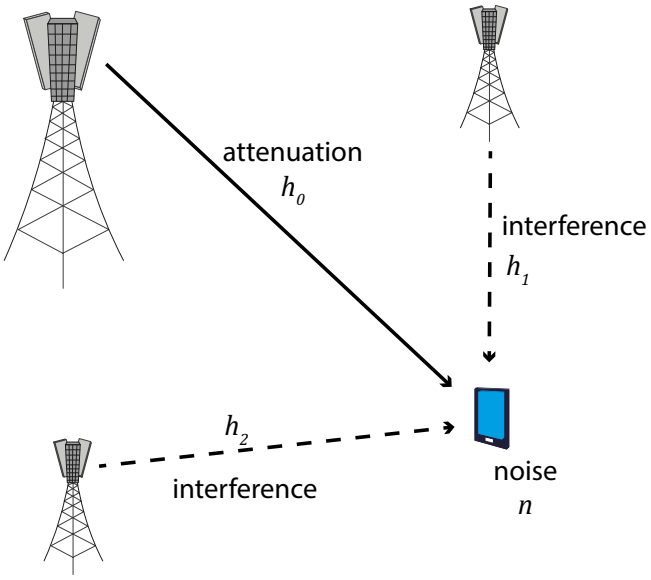


Figure 4: Illustration of a link. The desired link between a BS and a user is affected by an attenuation factor and the interference from other BSs.

Link

A **link** is a radio connection between a transmitter and a receiver. Since the discussion is limited to DL transmissions, the transmitter is a BS and the receiver is a mobile user equipment, as is illustrated in figure 4. The connection between the BS and the user is affected by an attenuation. Additionally, a link can be affected by interference from other transmitters. The quality of a link can be described through different interconnected measures. These measures include the received power, the signal to interference and noise ratio (SINR) or signal to noise ratio (SNR) that the user experiences. The fading parameters which play into the SINR and SNR can also be of interest. Examples for large scale fading parameters are the path loss, the antenna pattern, whether the user is indoor or outdoor, and whether a line of sight (LOS) connection exists between the user and the BS or not (non line of sight (NLOS)). Examples for small scale fading parameters are the speed at which the user moves and the power delay profiles.

NOMA User Pair

A NOMA user pair consists of two users, a **near user** and a **far user**. The near user is illustrated in dark blue in figure 2 and the far user is illustrated in light blue in figure 2. The near user has very good channel conditions and can perform SIC to retrieve the transmitted signal from a NOMA transmission. The far user has bad channel conditions and requires high transmission power to be able to decode the transmitted signal. The

link of the far user suffers additional interference from the transmission of the near user.
In DL power domain NOMA, the transmit power of the BS is divided between the near and the far user. The far user obtains the larger share of the transmit signal power $\alpha_{far} > 0.5$ and the near user obtains the remaining power

$$\alpha_{near} = 1 - \alpha_{far}, \tag{1}$$

with $\alpha_{near} < 0.5$. The power allocation factors α_{near} and α_{far} are illustrated in figure 2.

Small Scale Fading

The **small scale fading** describes the time dependent fading of a link, that is caused by multi path propagation of wireless signals. It is illustrated in figure 3. The models describing the small scale fading are often referred to as **channel models**. While small scale fading channel models can consider or model parts of the network geometry, the small scale fading in this work is modeled through position independent power delay profiles (PDPs) [49, 50, 51]. The effects of the small scale fading are modeled as a time dependent channel matrix $\mathbf{H} \in \mathbb{C}^{n_{RX} \times n_{TX}}$, where n_{TX} and n_{RX} are the number of transmit and receive antennas. Thus the channel matrix contains a complex fading factor modeling the attenuation for each transmit and receive antenna pair.

4 Flow of the System Level Simulation

In this section, the simulation of single user multiple input multiple output (SU-MIMO) OMA transmissions in the VCCS 5G SL simulator is described. The SL abstraction developed in section 5 will refer to this description.

4.1 Simulation Time Line

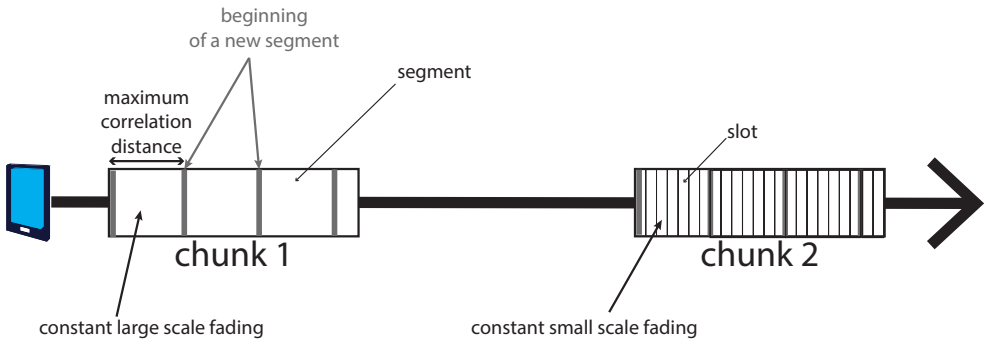


Figure 5: Timeline of a simulation.

The simulation time line is defined according to the coherence time of different fading parameters. The smallest time unit in a simulation is a slot. The duration of a slot is also the coherence time of the small scale fading, i.e., the small scale fading is assumed to be constant for the duration of a slot. The next larger time unit is the segment. The macroscopic fading values are assumed to be constant for the duration of a segment. The duration of a segment is defined through the maximum correlation distance, which is the maximum distance a network element can travel before the macroscopic fading values change.
To allow for parallelization of the simulation, another time unit is defined in the simulation: the chunk. The time line of the simulation can be divided into chunks, for which the simulation can run independently from the other chunks and thus in a parallel loop. The definition of chunks also allows the simulation of specific time windows in which burst traffic might occur, while skipping the time in between the traffic bursts if they are not of interest.
Figure 5 shows a user moving in a simulation. The direction of the movement is indicated by the large arrow. Along the user's path, the positions of the user are used in a simulation. If the user moves further than the maximum correlation distance, a new segment is started in the simulation. Each of these segments consists of several slots, which correspond to a small scale fading channel realization.

4.2 Network Setup

The first step in a VCCS 5G SL simulation is to create the network geometry. The network geometry includes buildings, streets, walls and the positioning of transmitters and receivers. The simulation of buildings allows the analysis of specific and realistic network setups.

Buildings, Walls and Streets

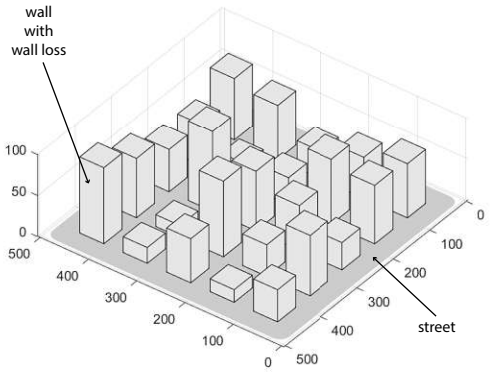


Figure 6: Placement of walls and streets shown with the example of a Manhattan city.

The first step in a simulation is the placement of walls and streets that define the geometric layout of the simulation region. Figure 6 shows an exemplary Manhattan city that is built with random building heights and a predefined street width.
Each wall in the simulation has a wall loss that describes the attenuation a link suffers, when passing through that wall. Buildings also determine whether a network element is indoors or outdoors. This attribute will later be used to choose the path loss model applied to a link. The choice of the path loss model for a link is described in section 4.3.3.
The streets have no influence on the properties of links, they enable realistic user positioning and allow the creation of users randomly positioned on streets.

Users

After the setup of the geography of the simulation scenario, users can be placed in the simulation region. The users can be placed according to stochastic processes, such as a Poisson point process (PPP), or deterministically. An arbitrary number of user groups can be created in a simulation, where each user group has a set of common properties. The common properties for a user group include the number of receive

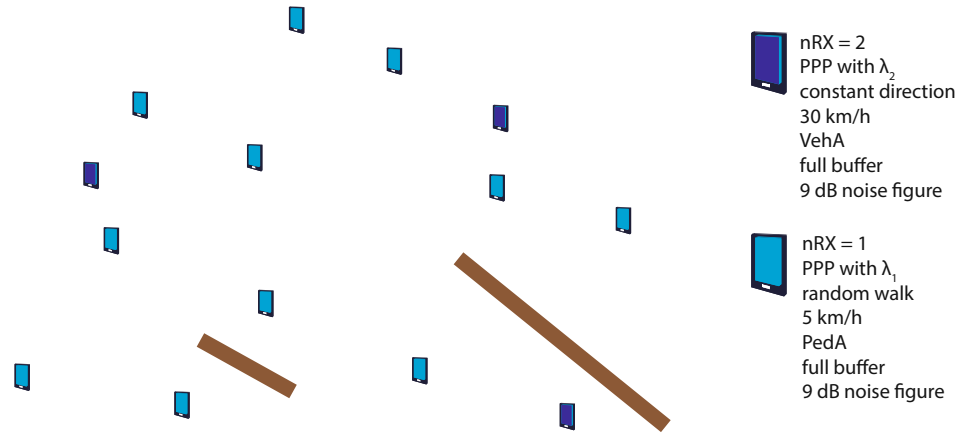


Figure 7: Example of user positioning. Two user groups with different properties are placed according to a Poisson point process (PPP) with density λ. The brown lines represent wall blockages.

antenna elements, the user movement, the receiver noise figure, the small scale fading channel model type and the traffic model. An exemplary setting for two user groups is shown in figure 7.
Figure 7 shows an exemplary user positioning with the placement of vehicular and pedestrian users in a simulation region with two blockages.

Base Stations

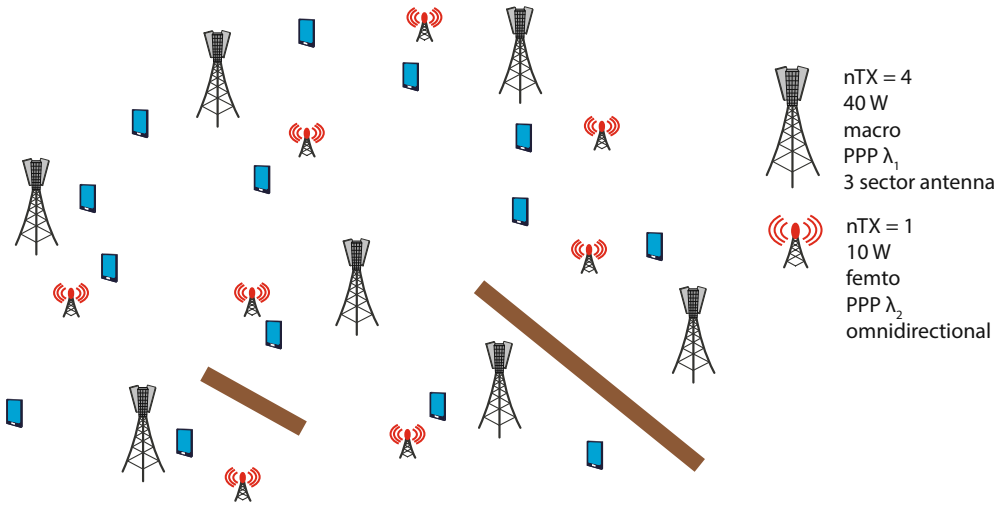


Figure 8: Example of BS positioning. Two BS types are placed according to a PPP.

The creation of BSs is similar to the creation of users: an arbitrary number of BS

groups can be simulated, with each group having a set of common parameters. Like the users, the BSs can be placed according to the network geography, for example on top of buildings or at the center of a cluster of users. The common parameters shared by a type of BSs in a simulation include the transmit power, the antenna gain pattern and the number of transmit antenna elements. Two exemplary BS settings are shown in figure 8 with a complete simulation region with all network elements and wall blockages.

Additional Interfering Base Stations

To mitigate border effects or to assure a realistic amount of interference in small simulation scenarios, additional interfering BSs can be placed outside of the region of interest (ROI) in an interference region. These BSs are simulated with simplified transmission schemes to save on computational complexity. If the ROI is large enough, a wraparound can be performed alternatively and the BSs in the ROI can create additional interference for users in the border region [52].

4.3 Cell Association

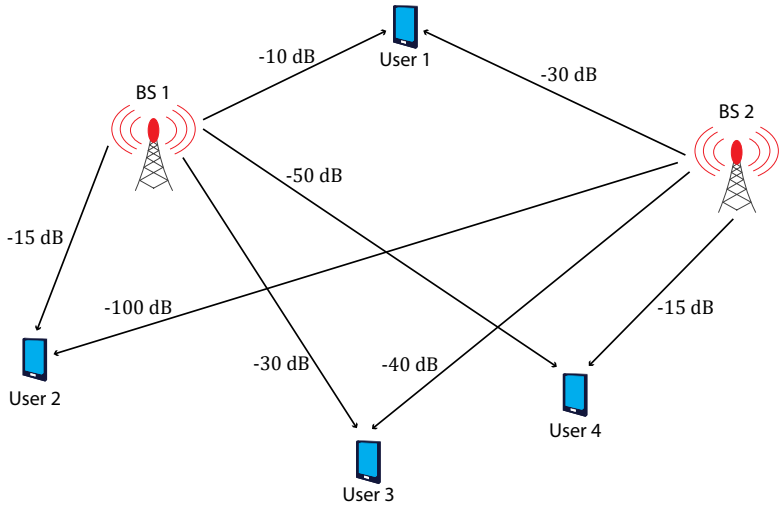


Figure 9: The cell association procedure. The cell association metric is evaluated for each user and BS pair. The user then chooses the BS with the best connection, i.e. the highest cell association metric.

With the network setup completed, the simulation can start. The first step is to attach users to their serving base station in the cell association. To determine which users are served by which base stations, the large scale fading parameters for the simulated network geometry are examined.

Figure 9 shows an exemplary network setup with the cell association metric evaluated for each user and BS pair. For cell association, each user chooses the BS with the highest cell association metric. In this example, user 1 would be attached to BS 1, as well as user 2 and 3. User 4 would be attached to BS 2. The cell association metric can be the macroscopic received power or the macroscopic SINR, as is explained below.
We now examine the different components of the large scale fading that play into the cell association metrics and describe the models that are used in this work. Then, the metrics considered for the cell association decision are defined and discussed.
In a simulation, each of the components described in the following is calculated for each possible link in the simulation. A possible link is a pair of a user and a BS. The calculated large scale fading values are saved and used for the evaluation of links in the LQM.

4.3.1 Transmit Power

Each BS has a transmit power which is the maximum power the BS can use to transmit a signal to a user. The transmit power can depend on the BS type. While macro BSs have a higher transmit power, small cells, like femto BSs have a smaller transmit power and smaller cell sizes. In general SL simulations consider interference limited networks, which means that an increase in the transmit power in general does not result in a better link quality.

4.3.2 Antenna Gain

The antenna gain pattern models the directivity of the antenna. Depending on the position of the user with respect to the base station, a larger or smaller amount of the transmit power is received by the user. The antenna gain depends only on the angle at which the user is positioned. The attenuation caused by the distance to the antenna is modeled by the path loss.
For example, a three sector BS radiates most of its transmit power in a 120° range, while an omnidirectional antenna radiates its available transmit power evenly in all directions.

4.3.3 Path Loss

The path loss models the attenuation of a signal that is caused by the distance between the transmitter and the receiver. As such it depends solely on the position of the network elements and the carrier frequency which is used for transmission. Two different path loss models are used for simulations in this work, these are the classical free space path loss model [48] and the urban macro (UMa) three dimensional (3D) model specified in [53].

Which path loss model is chosen for a link depends on the characteristics of the link. Different path loss models can be defined for different BS types, for the LOS/NLOS characteristic and the indoor/outdoor characteristic. During the simulation, the characteristics of a link are evaluated and the appropriate path loss model is chosen to determine the path loss.

4.3.4 Shadowing and Wall Loss

The strength of the received signal is influenced by the physical environment through which the transmitted signal passes. Blockages can attenuate the signal strength, while large reflective surfaces can create a transmission path for an otherwise blocked link. These effects are either modeled statistically with a shadow fading model [54] or are determined from the geometry of the network. In the later case, the signal strength is attenuated by the wall loss defined for each blockage in the line of sight of a link.

4.3.5 Cell Association Metrics

With the large scale fading parameters defined and calculated for each link in the simulation, each user can be associated to a serving cell. The cell association attaches each user to the BS that is closest to the user, where closest means that the cell association metric is the highest and the signal received from this BS is stronger than the signal received from all other BSs. This is shown in figure 9.
One of two cell association metrics can be used for cell association in the VCCS 5G SL simulator. These are the macroscopic received power and the macroscopic SINR. Both metrics are defined below.

Macroscopic Received Power

The macroscopic received power P_{rx} of a user is defined as

$$P_{\mathrm{rx}}|_{\mathrm{dB}} = P_{\mathrm{tx}}|_{\mathrm{dB}} + G|_{\mathrm{dB}} - PL|_{\mathrm{dB}} - S|_{\mathrm{dB}}, \tag{2}$$

where $P_{\mathrm{tx}}|_{\mathrm{dB}}$ represents the transmit power radiated by the BS, $G|_{\mathrm{dB}}$ represents the antenna gain in the direction of the user, $PL|_{\mathrm{dB}}$ represents the path loss between the user and the BS and $S|_{\mathrm{dB}}$ represents the shadow fading and wall loss experienced by the link between the user and the BS. Equation (2) can be written in non-logarithmic form as

$$P_{\mathrm{rx}} = \frac{P_{\mathrm{tx}} \cdot G}{PL \cdot S}. \tag{3}$$

Macroscopic SINR

The second metric to evaluate the strength of a link is the macroscopic SINR defined as

$$\mathrm{SINR}_i = \frac{P_{\mathrm{rx}}^{(i)}}{\sum_{j \neq i} P_{\mathrm{rx}}^{(j)} + N}, \tag{4}$$

where N represents the noise power at the receiver, $P_{\mathrm{rx}}^{(j)}$ represents the macroscopic receive power from BS j at the user and $j = 1, 2, ..., n_{\mathrm{BS}}$ with n_{BS} represents the total number of BSs in the simulation.
The choice of the cell association depends on the network constellation. While the macroscopic receive power is less computationally expensive to calculate, it can lead to undesired cell associations in networks with small cells that have small transmit powers. In homogeneous networks, both metrics yield roughly the same cell association and the simpler macroscopic received power metric can be chosen.

4.4 Small Scale Fading

To save computational effort within the simulation loop over all slots, the small scale fading is precalculated before the simulation loop and the PDP channel trace saved to be read out during the simulation.
To generate the channel traces, all possible channel constellations are evaluated and a channel trace is calculated for each channel constellation. The channel constellation considers, among other parameters, the number of transmit and receive antennas, the channel model defining the PDP and the maximum user speed.
From the saved channel traces, channel realizations will be read out in the simulation loop. Within the simulation loop, a random position in the channel trace is chosen for each user and the channel realizations are then read out from this position in the channel trace. The random choice of the starting point allows to read uncorrelated channel realizations for each user from the same channel trace.

4.5 Simulation Loop

With the network set up, the users associated with their desired BS and the fading parameters calculated, a transmission can take place. The following steps are executed for each slot in the simulation. The description of the simulation is limited to one user, the same procedure is applied for all users in the simulation.

4.5.1 Traffic Generation

The first step for a transmission is to generate data to transmit. For this, the traffic model fills the transmission buffer of the user according to the chosen traffic model and traffic model parameters. Traffic can either be generated in constant packet sizes at a constant rate or a full buffer model is assumed. The full buffer traffic model assumes that all users have data to transmit at all times.
The traffic model used in this work, is the constant rate traffic model, that produces packets of a specified size in a specified time interval. The traffic model allows to simulate different types of users. For example, an internet of things (IoT) user would generate small packets over large time intervals, while a user that streams content would generate large packets over small time intervals.

4.5.2 Scheduling

Now that the traffic to transmit has been generated, each BS can schedule its available resources to its associated users, if they have traffic to transmit. Two different scheduling strategies are considered in this work: the round robin scheduler and the best CQI scheduler.
After determining the allocation of the resource blocks (RBs) to the users, the scheduler sets the power allocation and chooses the precoder and the CQI. The CQI determines the modulation and coding scheme (MCS) with which the data is transmitted in each RB.

The Resource Grid

The granularity of the scheduling is in RBs. In a slot, each BS has physical resources available on which the BS can transmit data. These resources are organized in a resource grid that is composed of a number of subcarriers in frequency and a number of OFDMA symbols in time. All subcarriers for which the small scale fading is assumed constant are combined into an RB. Additionally, the resources can be divided into an arbitrary number of RBs in time. These RBs are then the smallest unit of physical resources considered in the simulation. An example of a resource grid is shown in figure 10.
Each RB can have an individual resource allocation and digital precoding and for each RB all simulation steps are performed. These simulation steps include calculation of a post equalization SINR in the LQM and the calculation of feedback values.

Round Robin Scheduler

The round robin scheduler sequentially allocates RBs to all users that want to transmit data. An example of a round robin scheduling is shown in figure 11. The first RB in

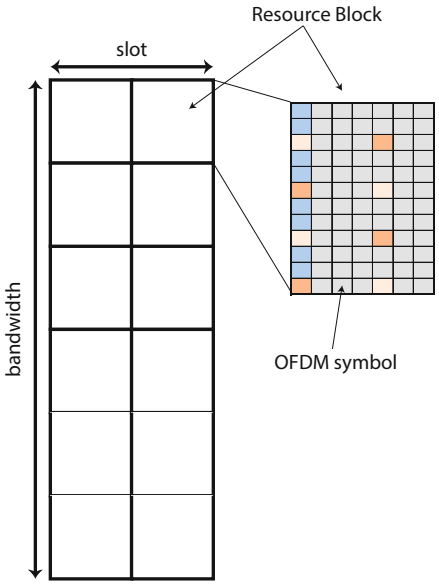


Figure 10: The resource grid. On the left, the resource grid that a BS has as available resources in each slot is illustrated. Each square box represents a resource block. On the right, one resource block is shown in more detail. The resource block consists of several OFDMA symbols, of which some can be used for signaling. In this example the orange symbols could be used for the transmission of reference symbols, the blue symbols could be used for a control channel and the gray symbols could be used for the transmission of user data.

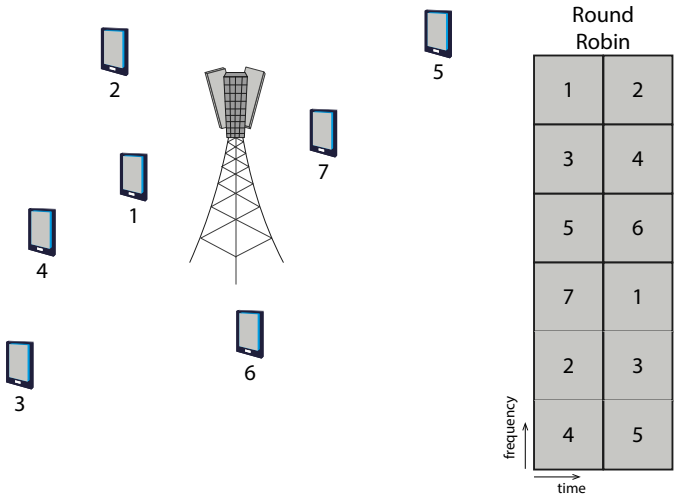


Figure 11: Example of a resource allocation performed by a round robin scheduler.

the resource grid is allocated to user 1, the second RB to user 2 and so on until all users have been allocated resources. If all users are served and RBs remain available, the scheduling starts again at user 1 and allocates an RB to each user until no more

RBs are available. Thus user 6 and 7 get one RB for data transmission and users 1 to 5 get two RBs each. In the next slot, user 6 will be the first user to be served.
The advantage of the round robin scheduler is that it is a very simple scheduler that distributes the resources between all users in a fair manner, i.e. all users get resources allocated. The downside to this fairness is that the overall network throughput is not maximized, since resources are blocked by users that only generate a low throughput due to bad channel conditions.

Best CQI Scheduler

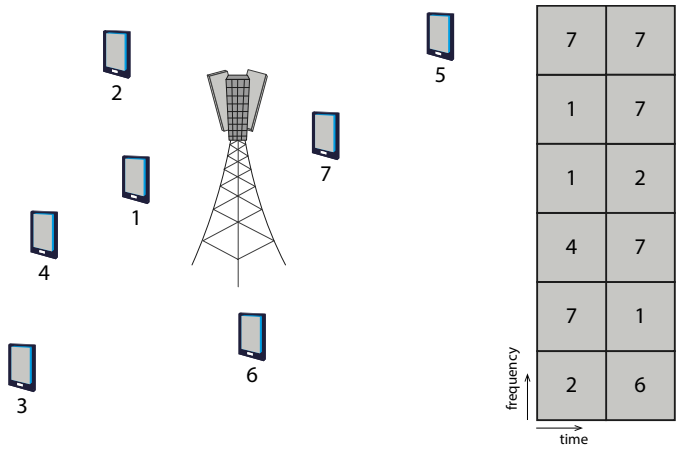


Figure 12: Example of resource allocation performed by a best CQI scheduler.

The best CQI scheduler allocates each RB to the user with the highest CQI value. If several users have the same (and highest) CQI value in an RB, then the best CQI scheduler, schedules these users in a round robin fashion. An example for a resource allocation according to a best CQI scheduler is shown in figure 12. The estimated CQI values for each RB for each user are provided by the feedback. In the example in figure 12, user 7 has the highest CQI value in the first and second RB. Due to frequency selective fading, user 1 has a higher estimated CQI in the third RB. Each RB is allocated to the user with the highest estimated CQI.
The best CQI scheduler maximizes the overall network throughput, since the users with the best channel conditions are transmitting and generating the highest throughput possible [55] per RB. However, users with poor channel conditions at the cell edge get starved and are not allocated any resources to transmit their data. The best CQI scheduler leads to an unfair distribution of resources.

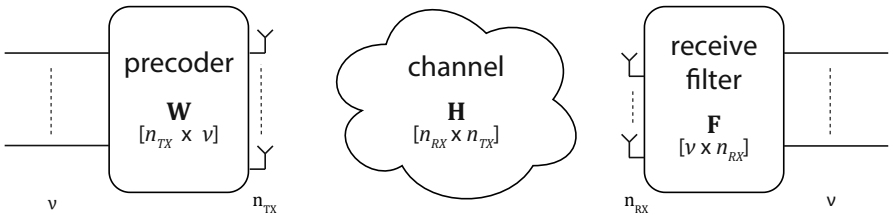


Figure 13: Transmit chain considered in the LQM.

4.5.3 Link Quality Model

The LQM determines the quality of the link for each RB on which a user transmits data. The measure for the quality of the link is the post equalization SINR γ_l that is calculated for each layer $l = 1, 2, \ldots, \nu_0$, where ν_0 is the number of layers used for transmission by the desired BS.
To evaluate the quality of the link, the LQM considers the precoder $\mathbf{W}$, the small scale fading characteristic of the wireless channel $\mathbf{H}$ and the receive filter $\mathbf{F}$ as shown in figure 13.

Precoder

The precoder $\mathbf{W} \in \mathbb{C}^{n_{TX} \times \nu_0}$ maps the signals from the ν_0 layers to the n_{TX} transmit antennas and can perform beam forming in the process. In the following, the precoder used by the desired BS will be denoted as $\mathbf{W}_0$ and the precoders used by the interfering BSs will be denoted as $\mathbf{W}_i \in \mathbb{C}^{n_{TX_i} \times \nu_i}$, where n_{TX_i} denotes the number of transmit antennas of the ith BS and ν_i denotes the number of layers used for transmission by the ith BS. The precoder can be set individually for each RB. In this thesis, the precoder is chosen from a set of possible precoders (codebook) as defined in [19].

Small Scale Fading

The channel matrix representing the small scale fading is read out of a precalculated channel trace. In the following, the small scale fading of the desired channel, i.e. the link between the desired BS and the user, will be denoted by the matrix $\mathbf{H}_0 \in \mathbb{C}^{n_{RX} \times n_{TX}}$, where n_{RX} is the number of receive antennas at the user and n_{TX} is the number of transmit antennas at the desired BS. The channel between the user and the interfering BSs will be denoted as $\mathbf{H}_i \in \mathbb{C}^{n_{RX} \times n_{TX_i}}$, with $i = 1, 2, \ldots, N_{int}$, where N_{int} is the total number of interfering BSs.

Receive Filter

The user filters the received signal to eliminate inter layer interference and, if possible, reduce the undesirable effects of the channel fading. In the VCCS 5G SLS a zero forcing (ZF) receive filter is used [56]. The receive filter assumes perfect channel knowledge and inverts the effective channel $\mathbf{H}_0\mathbf{W}_0$. The receive filter $\mathbf{F} \in \mathbb{C}^{\nu_0 \times n_{RX}}$ is calculated as the pseudo-inverse of the effective channel, given by:

$$\mathbf{F} = [(\mathbf{H}_0\mathbf{W}_0)^{\mathrm{H}}\mathbf{H}_0\mathbf{W}_0]^{-1}(\mathbf{H}_0\mathbf{W}_0)^{\mathrm{H}}, \tag{5}$$

where $\mathbf{A}^{\mathrm{H}}$ denotes the Hermitian transpose of $\mathbf{A}$.

Post Equalization SINR

The post equalization SINR for layer l in an RB can be expressed as:

$$\gamma_l = \frac{\zeta_l P^{(l,0)}}{\sum_{i=1}^{N_{int}} \theta_{i,l} P^{(l,i)} + \psi_l \sigma^2}, \tag{6}$$

where we introduced the terms $P^{(l,i)}$, ζ_l, $\theta_{i,l}$ and ψ_l to simplify the notation. Each of these terms is defined next.
The term $P^{(l,i)}$ represents the received power on the current resource block for layer l from transmitter i, where the transmitter with index 0 is the desired BS and the other transmitters are interfering BSs. The transmit power P_{rx} from equation (3) is assumed to be evenly distributed over the bandwidth and evenly distributed over the layers, which results in:

$$P^{(l,i)} = \frac{P_{rx}^{(i)}}{n_{RB_{Freq}} \nu_0}, \tag{7}$$

where $n_{RB_{Freq}}$ is the number of RBs in the frequency domain.
For the signal power fraction ζ_l on layer l, we first introduce the signal power fraction matrix $\mathbf{Z} \in \mathbb{C}^{\nu_0 \times \nu_0}$ as

$$\mathbf{Z} = \mathbf{F}\mathbf{H}_0\mathbf{W}_0. \tag{8}$$

The matrix elements of $\mathbf{Z}$ will be denoted as $Z_{i,j}$ with $i, j = 1, 2, \ldots, \nu_0$. With the use of a ZF receive filter, the signal power fraction matrix results in an identity matrix $\mathbf{Z} = \mathbf{I}_{\nu_0 \times \nu_0}$.
The signal power fraction ζ_l for layer l is the fraction of the signal power that is received on layer l and can be expressed as:

$$\zeta_l = |Z_{l,l}|^2, \tag{9}$$

with $\zeta_l = 1$ on all layers for a ZF receive filter.

For the interference enhancement $\theta_{i,l}$ of interferer i on layer l, we introduce the interference enhancement matrix $\mathbf{\Theta}_i \in \mathbb{C}^{\nu_0 \times \nu_i}$ for each interfering BS:

$$\mathbf{\Theta}_i = \mathbf{F}\mathbf{H}_i\mathbf{W}_i. \tag{10}$$

The elements of the interference enhancement matrix $\mathbf{\Theta}_i$ are denoted as $\Theta_{i,l,m}$ with $l = 1, 2, \ldots, \nu_0$ and $m = 1, 2, \ldots, \nu_i$ The interference enhancement $\theta_{i,l}$ represents the interference suffered on layer l from the transmission on all layers of the interfering BS i. It can be calculated from the interference enhancement matrix as follows:

$$\theta_{i,l} = \sum_{m=1}^{\nu_i} |\Theta_{i,l,m}|^2. \tag{11}$$

Finally, the noise enhancement ψ_l on layer l can be directly calculated from the receive filter $\mathbf{F}$ as

$$\psi_l = \sum_{k=1}^{n_{RX}} |F_{l,k}|^2, \tag{12}$$

where $F_{l,k}$ is the element in row l and column k of the receive filter $\mathbf{F}$.
To simplify the notation, we introduce the term N_l for the noise power and the term I for the interference from other cells as

$$N_l = \psi_l \sigma^2 \tag{13}$$

and

$$I_l = \sum_{i=1}^{N} \theta_{i,l} P^{(l,i)}. \tag{14}$$

The interference enhancement $\theta_{i,l}$ and the noise enhancement ψ_l are used as defined in equation (11) and equation (12) and the macroscopic receive power per layer is defined in equation (7).
This and the use of a ZF receive filter, results in the very concise expression for the post equalization SINR on layer l:

$$\gamma_l = \frac{P^{(l,0)}}{I_l + N_l}. \tag{15}$$

4.5.4 Link Performance Model

The link performance model (LPM) evaluates the throughput of a given link based on the post equalization SINR calculated in the LQM. Figure 14 shows a block diagram of the LPM. In a first step, the LPM collects the post equalization SINR values from all layers from all RBs on which data for this user was transmitted. Then, the post

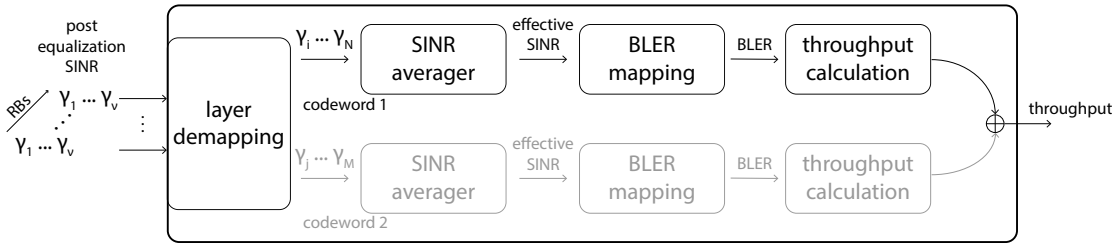


Figure 14: Link performance model.

equalization SINRs are separated on a codeword by codeword basis. This is done in the layer demapping shown in figure 14. The post equalization SINR values are then averaged for each codeword through a mutual information effective signal to interference and noise ratio mapping (MIESM) in the SINR averager. This averaged SINR value is called the effective SINR and can then be mapped to a block error ratio (BLER) value. Then, a random experiment with the failure probability corresponding to the BLER value is conducted and determines whether the transmission of this codeword was successful or the transmission has failed.
If the transmission was successful the number of transmitted bits is calculated based on the MCS chosen by the scheduler and the size of the RBs.
Additionally, the throughput that would have been achieved if the optimum MCS would have been used is calculated. This best CQI throughput can be used to evaluate the performance of the feedback algorithm.
Besides the throughput and the best CQI throughput, the effective SINR and the BLER are saved as simulation results.

4.5.5 Feedback

The feedback estimates the CQI for each resource block and codeword for each user based on the post equalization SINR in the current slot. Additionally, the feedback can set the rank indicator (RI) and the precoding matrix indicator (PMI) [57]. The feedback values are used to choose the precoder and they define the used CQI, which in turn determines the MCS used for a transmission.

4.5.6 Traffic Transmission

After the data transmission is completed, the traffic buffer is updated and successfully transmitted packets are removed from the transmit buffer. The slot in which the packet transmission was successfully completed is saved and is used to determine the latency of the network.

4.6 Simulation Results

Once the previous steps have been repeated for all users and all slots of a simulation, the results of all chunks in the simulation are combined to show the simulation results. The network performance is characterized with the empirical cumulative distribution functions (ECDFs) of different link properties such as the user throughput per slot, the packet latency, the BLER or the SINR. Additionally, other link properties can be saved and used to evaluate the network performance. These link properties include the feedback and scheduling, the path loss and macroscopic received power and whether a link has a LOS connection or whether a user is indoors.

5 NOMA System Level Abstraction

In this section a SL abstraction for a DL two user power domain NOMA transmission with SIC at the near user is developed. The implementation of the NOMA user pairing and the scheduling for NOMA in the VCCS 5G SL simulator are first presented in this section. The NOMA SL abstraction is then separated in two parts. In the first part a simple SISO transmission is considered; in the second part, the SISO model is extended for MIMO transmissions.

5.1 User Pairing

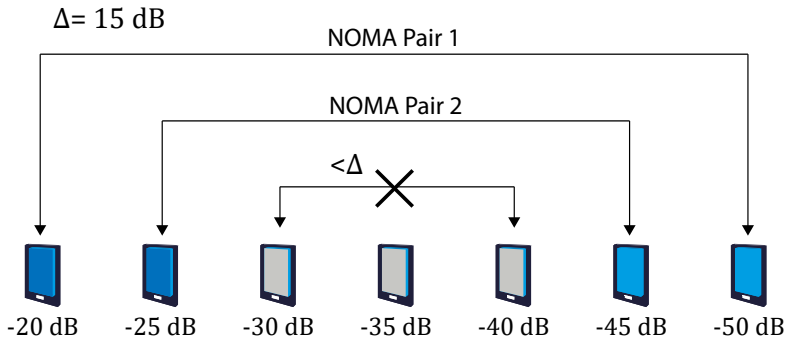


Figure 15: Pairing of NOMA users. Δ is the minimum distance between users for NOMA pairing.

NOMA is more effective for larger differences in link strength between the paired users, thus the user pairing strategy is of importance to achieve the promised gains. The user pairing chooses which users will be paired for combined NOMA transmissions by pairing a far user with poor channel conditions to a near user with strong channel conditions. The NOMA user pairing is built on top of the cell association described in section 4.3. It uses the same metric as that chosen for the cell association, which can be either the macroscopic received power or the macroscopic SINR.
An example for a NOMA user pairing is shown in figure 15, which corresponds to the user pairing strategy suggested in [5]. The users in each cell are sorted by decreasing link strength according to the cell association metric. Then, the user with the strongest and the user with the weakest link strength are paired if the difference between their cell association metric is higher than the user pairing threshold Δ. This process is repeated for the remaining users until either no users are left, or a pair of users has a difference in link strength lower than the threshold. The user with high link strength is categorized as near user and the user with low link strength is categorized as far user. The unpaired users perform classical OMA transmissions.

5.2 Scheduling

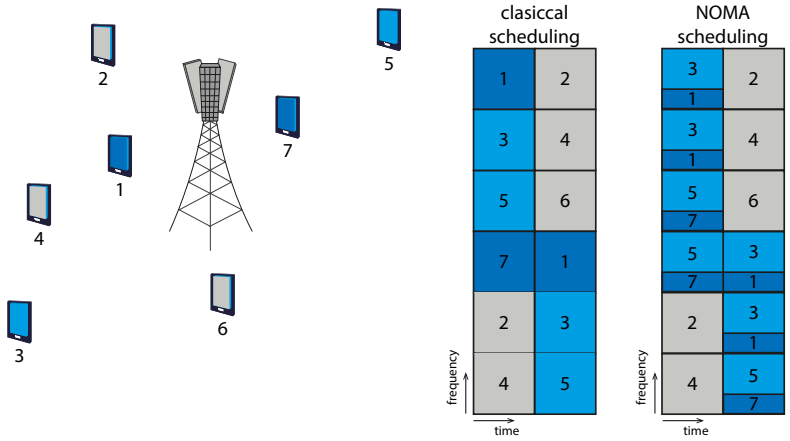


Figure 16: NOMA scheduling: the scheduling from the classical simulation is overwritten to schedule both users in a NOMA pair to the assigned resources.

After the scheduling is performed as described in section 4.5.2, the additional NOMA users are added to the BS scheduling. For this, all RBs assigned to a NOMA user pair are scheduled to the NOMA pair with the far user obtaining the majority of the power allocation share and the near user obtaining the remaining power share.
In figure 16, an example of a BS scheduling is shown. In this example, two NOMA user pairs are present in the cell. User 1 and user 3 form a NOMA pair with user 1 being the near user and user 3 the far user. The other NOMA pair is formed by near user 7 and far user 5. The classical scheduling assigns the RBs in a round robin fashion in this example, such that all users have been assigned one or more RBs, as can be seen in the left resource grid in figure 16. For the NOMA scheduling, the RBs assigned to any user in a NOMA pair are reassigned to both users of the NOMA pair. For the first NOMA pair with near user 1 and far user 3, that are the first two RBs in frequency and two more RBs in the second half of the time slot. These RBs are now reassigned to both users in the NOMA pair, as can be seen in the second resource grid in figure 16. The same procedure is repeated for the second NOMA pair consisting of users 7 and 5.
Aside from the user allocation, the scheduler also chooses the precoder and CQI used for transmission based on the feedback described in section 4.5.5. Figure 17 shows an example of the BS signaling set by the scheduler in a NOMA transmission. The signaling concerning NOMA near users is marked in dark blue, signaling for NOMA far users is shown in light blue and signaling for OMA users is shown in gray. The signaling that is identical for near and far users is shown in white boxes. The grid represents the resource grid divided into RBs. In case of a NOMA transmission, the far user allocation shows the allocated far and OMA users, while the near user allocation in an additional

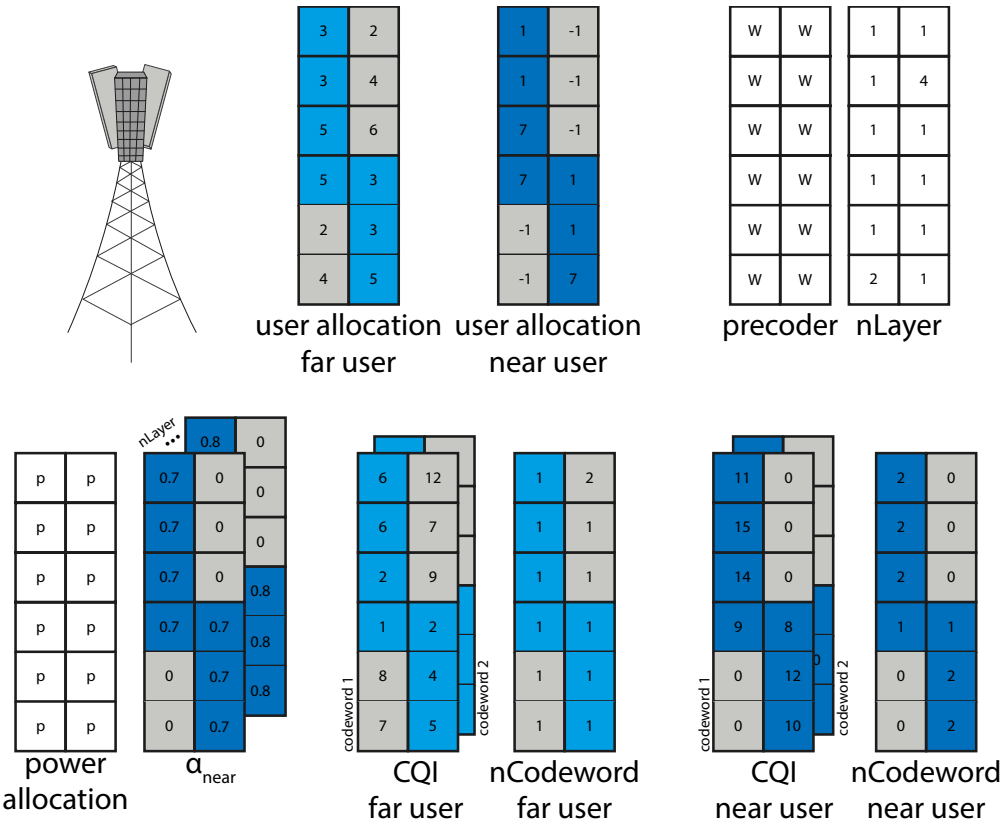


Figure 17: Scheduler information saved at the BS, where nLayer is the number of layers used for transmission and nCodeword is the number of codewords transmitted. A user allocation of -1 indicates that no user is scheduled on this resource and the CQI is set to 0 for resources and codewords that are not allocated to a user.

grid. The precoder used by the BS is chosen based on the feedback of the far user to assure that the number of layers used does not exceed the number of layers available in the far user channel. While the power is assumed to be evenly distributed over all resources, as can be seen from the power allocation, the NOMA power allocation factor is set for each layer as described in section 5.2.1. The used CQI is saved individually for the near and far users and is set for each codeword according to the feedback.

5.2.1 NOMA Power Allocation

The NOMA power allocation factor α is set according to [19], depending on the modulation order used by the near user as indicated in table 1.
Three different MUST modes are defined in the 3GPP standard, that define the different power allocation factors α_{far} for the NOMA transmission.

MUST index	Modulation Order QPSK	16QAM	64 QAM
01	0.8	$\frac{32}{42}$	$\frac{128}{170}$
10	$\frac{50}{58}$	$\frac{144.5}{167}$	$\frac{40.5}{51}$
11	$\frac{264.5}{289}$	$\frac{128}{138}$	$\frac{288}{330}$

Table 1: Power allocation factors α_{far} [19].

The power allocation factor of the near user α_{near} can be derived from α_{far} as

$$\alpha_{near} = 1 - \alpha_{far}. \tag{16}$$

5.3 NOMA System Model

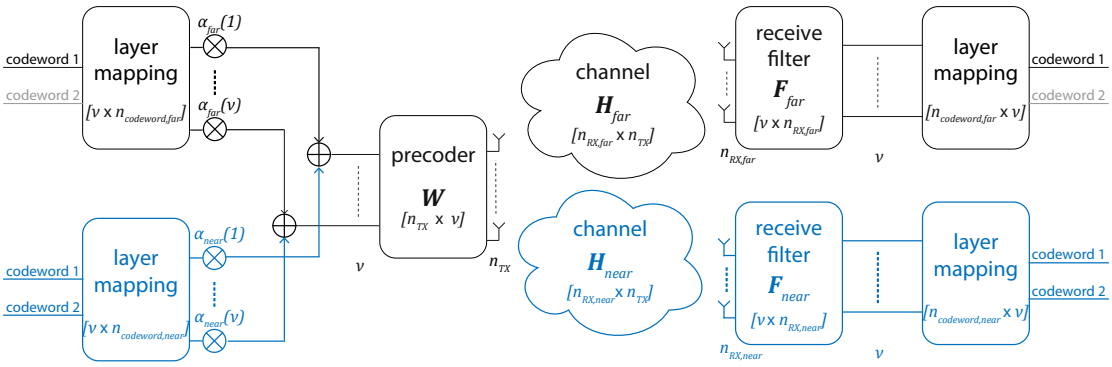


Figure 18: Block diagram of the transmit and receive chain of a NOMA transmission. The additional blocks of the transmission of the NOMA near user are shown in blue.

Figure 18 gives an overview of the transmit and receive chain of a NOMA transmission. The modulation and coding are abstracted through BLER curves in a SL simulation. Thus, the transmission starts with the codewords of data to be transmitted. Both the near and the far user have data to be transmitted. The second codeword of the far user is grayed out to indicate that it is possible, that a user only transmits one codeword. The upper and lower part of figure 18 correspond to a classical OMA transmission. Since the near and the far user are located at different positions, their received signal is affected by different channel fading coefficients indicated as $\mathbf{H}_{far}$ and $\mathbf{H}_{near}$ in the channel. The difference to a classical OMA transmission is the power allocation factor α with which the signals on each layer are multiplied and the combination of the near and far user's signal before transmission. These steps are shown in figure 18 with the multiplication of the signal on each layer by a power allocation factor $\alpha(l)$ defined for each layer.The addition of the near and the far user signals is shown where the blue

part of the near user transmission is merged with the far user transmission before the signals reach the precoder.
At the receiver side, the combined transmission of the near and far user signal reduces the post equalization SINR of the far user, since the signal of the near user has to be treated as interference. The near user performs SIC to decode its signal. A SL abstraction for the SIC procedure and the additional interference at the far user is developed in the following.

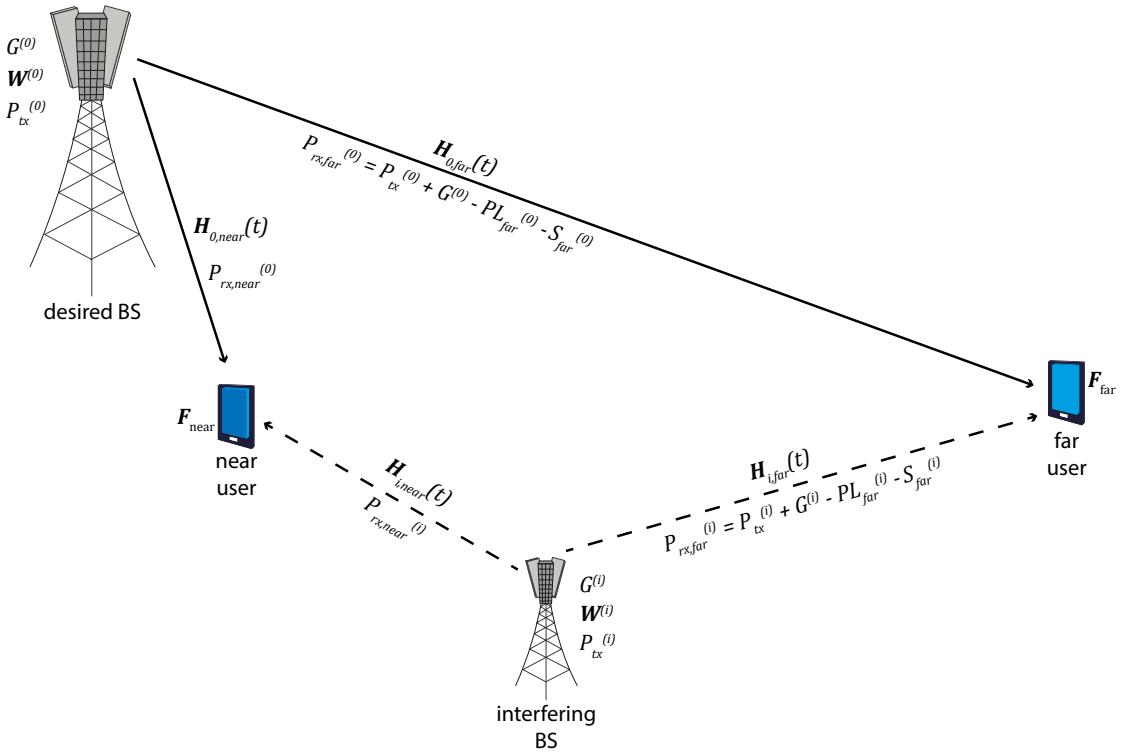


Figure 19: Schematic of the network elements involved in a NOMA transmission.

5.3.1 Single Input Single Output

We start with a SISO transmission, where each BS has one transmit antenna and each user has one receive antenna. To model the behavior of a network, additional BSs are present that create interference for the transmission of the desired signal. The interfering BSs also have one transmit antenna. The BS the user is associated to will be referenced to with the index 0, while the other BSs will be referred to with the indices $i = 1, 2, \ldots, N_{int}$, where N_{int} is the total number of interfering BSs.
A SISO transmission allows the simplification of the notation in section 4.5.3 to scalar channel fading coefficients and scalar precoders that are equal to 1. Thus, we will

denote the precoders as $w_0 = 1$ and $w_i = 1 \; \forall i = 1, 2, \ldots, N_{int}$ and the channel coefficients to the desired and interfering BSs as h_0 and $h_i \; \forall i = 1, 2, \ldots, N_{int}$. To discern the channels of the near and far users, we introduce the additional subscripts $near$ and far for the channel coefficients, which leads to the notations $h_{0,near}$ and $h_{0,far}$ for the channels to the desired BS of the near and the far user respectively and $h_{i,near}$ and $h_{i,far}$ for the corresponding channels of the interfering BSs. From this follows the notation of the receive filters of the near and far user $f_{near} = h_{0,near}^{-1}$ and $f_{far} = h_{0,far}^{-1}$.
The signal power fraction ζ defined in equation (9) can be reduced to

$$\zeta_{near} = |f_{near} h_{0,near} w_0| = |(h_{0,near} \cdot 1)^{-1} \cdot h_{0,near} \cdot 1|^2 \tag{17}$$
$$\zeta_{far} = |f_{far} h_{0,far} w_0| = |(h_{0,far} \cdot 1)^{-1} \cdot h_{0,far} \cdot 1|^2 \tag{18}$$

for a NOMA SISO transmission with a ZF receive filter. This leads to $\zeta_{near} = \zeta_{far} = 1$. The noise enhancement defined in equation (12) can be denoted as $\psi_{near} = |h_{0,near}^{-1}|^2$ and $\psi_{far} = |h_{0,far}^{-1}|^2$ for the near and the far user. The interference enhancement defined in equation (11) of a SISO transmission can be expressed as

$$\theta_{i,near} = |f_{near} \cdot h_{i,near}|^2 = |h_{0,near}^{-1} \cdot h_{i,near}|^2 \tag{19}$$
$$\theta_{i,far} = |f_{far} \cdot h_{i,far}|^2 = |h_{0,far}^{-1} \cdot h_{i,far}|^2. \tag{20}$$

The user receives a signal with a transmit power $P_{tx}^{(0)}$ from the associated BS. In case of a NOMA transmission, the signal from the desired BS contains the signal for the near user and for the far user, where the transmitted power is separated between the signal for the near user and the signal for the far user according to the power allocation factor from equation (1):

$$P_{tx}^{(0)} = \alpha_{far} P_{tx}^{(0)} + \alpha_{near} P_{tx}^{(0)}. \tag{21}$$

The transmitted signal is affected by large scale fading, that reduces the transmitted power $P_{tx}^{(0)}$ as described in equation (2). The macroscopic received power is distributed between the signal of the near user and the signal of the far user in the same fashion as the transmit power

$$P_{rx}^{(0)} = \alpha_{far} P_{rx}^{(0)} + \alpha_{near} P_{rx}^{(0)}. \tag{22}$$

This received power is assumed to be distributed evenly to layers and RBs as is shown in equation (7) and the power received from the desired BS on each layer and RB is denoted as $P_{far}^{(l,0)}$ and $P_{near}^{(l,0)}$ for the far and near user respectively. For the power received from the interfering BSs, the notation $P_{far}^{(l,i)}$ and $P_{near}^{(l,i)}$ is used. For the SISO case with only one layer, the shortened notations $P_{far}^{(0)}$, $P_{near}^{(0)}$, $P_{far}^{(i)}$ and $P_{near}^{(i)}$ are used. The terms affecting the channels in a NOMA transmission and their notation for the MIMO case are shown in figure 19, where bold letters are used for matrices. In a SISO transmission, the matrices reduce to scalars.

Far User

With the introduced notation, the post equalization SINR of the far user for a SISO NOMA transmission can be calculated as:

$$\gamma_{far} = \frac{\alpha_{far}\zeta_{far}P_{far}^{(0)}}{\alpha_{near}\zeta_{far}P_{far}^{(0)} + \sum_{i=1}^{N}\theta_{i,far}P_{far}^{(i)} + \psi_{far}\sigma^2}. \tag{23}$$

In equation (23), the numerator $\alpha_{far}\zeta_{far}P_{far}^{(0)}$ expresses the power with which the far user's signal is received by the far user. The term consists of the NOMA power allocation factor α_{far} of transmit power that is dedicated to the far user, the macroscopic received power $P_{far}^{(0)}$ described in equation (7) for a single layer and the received signal power fraction ζ_{far} described in equation (9) for a single layer in a SISO transmission. The term $\alpha_{near}\zeta_{far}P_{far}^{(0)}$ in the denominator of equation (23) represents the additional interference from the transmission of the near user experienced by the far user. The factor α_{near} is the share of the transmit power dedicated to the near user.
The second and third term in the denominator $\sum_{i=1}^{N}\theta_{i,far}P_{far}^{(i)}$ and $\psi_{far}\sigma^2$ remain unchanged from the expression for the post equalization SINR in the OMA case in equation (6) for a single layer.
With $\zeta_{far} = 1$, the expression for the post equalization SINR from equation (23) can be simplified to

$$\gamma_{far} = \frac{\alpha_{far}P_{far}^{(0)}}{\alpha_{near}P_{far}^{(0)} + \sum_{i=1}^{N}\theta_{i,far}P_{far}^{(i)} + \psi_{far}\sigma^2}. \tag{24}$$

For a characterization in terms of channel fading coefficients of the post equalization SINR at the far user in case of a SISO transmission, the terms for the interference and noise enhancement can be expanded to obtain

$$\gamma_{far} = \frac{\alpha_{far}P_{far}^{(0)}}{\alpha_{near}P_{far}^{(0)} + \sum_{i=1}^{N}|h_{0,far}^{-1}h_{i,far}|^2P_{far}^{(i)} + |h_{0,far}^{-1}|^2\sigma^2}. \tag{25}$$

Equation (25) is the expression that is evaluated by the LQM for each RB. The set of post equalization SINRs for all RBs will be handed over to the LPM to determine the achieved throughput over a link.

Near User - SIC

Now the model for the near user is derived. The near user performs SIC and first detects the signal of the far user, to then subtract the far user signal from the received signal

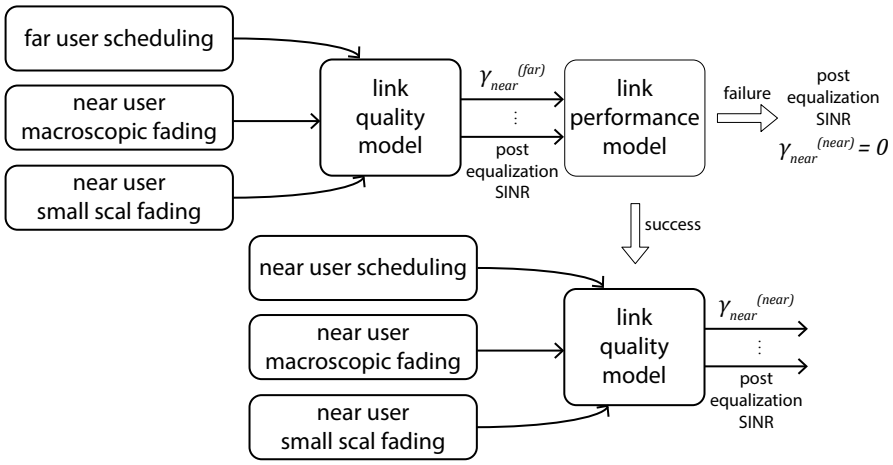


Figure 20: Block diagram of the LQM for a transmission with SIC at the near user.

and, in a second step, detect the near user signal. A block diagram of the LQM of the near user with SIC is shown in figure 20.
In a first step the far user signal is detected at the near user as shown in the upper part of figure 20. This is modeled analogous to the detection of the far user signal at the far user in equation (23). Only that here the channel and receiver of the near user are employed:

$$\gamma_{near}^{(far)} = \frac{\alpha_{far}\zeta_{near}P_{near}^{(0)}}{\alpha_{near}\zeta_{near}P_{near}^{(0)} + \sum_{i=1}^{N}\theta_{i,near}P_{near}^{(i)} + \psi_{near}\sigma^2}. \tag{26}$$

$\gamma_{near}^{(far)}$ is the post equalization SINR that the near user experiences when detecting the far user signal. This post equalization SINR is then passed to the LPM, which determines if the far user signal can be decoded successfully.
If the far user signal cannot be decoded the transmission fails and the near user cannot detect its own signal. Then, the post equalization SINR is set to zero for the signal at the near user, as is shown in figure 20 for the case of a transmission failure.
If the far user signal can be decoded, then the near user can attempt to decode its own signal. For this, the post equalization SINR of the signal of the near user is calculated as follows:

$$\gamma_{near}^{(near)} = \frac{\alpha_{near}\zeta_{near}P_{near}^{(0)}}{\varepsilon\alpha_{far}\zeta_{l}P_{near}^{(0)} + \sum_{i=1}^{N}\theta_{i,near}P_{near}^{(i)} + \psi_{near}\sigma^2}. \tag{27}$$

This expression for the post equalization SINR again follows the same structure as equation (23) with the introduction of the additional term ε, which is the interference cancellation factor.

The interference cancellation factor ε has a value between 0 and 1 and describes how successful the SIC was at removing the interference from the transmission of the far user. An interference cancellation factor of $\varepsilon = 0$ indicates full interference cancellation, where the far user signal is completely removed, while an interference cancellation factor $\varepsilon = 1$ indicates that no interference from the far user has been removed and all signal power allocated to the far user has to be treated as interference for the near user. When SIC is not be able to remove all interference from the far user, a share ε of the far user signal remains as interference at the near user.
The expression in equation (27) is used to determine the post equalization SINR for each RB assigned to the NOMA pair and the set of post equalization SINRs is then handed over to the LPM to determine the throughput that has been achieved in this transmission.

5.3.2 Multiple Input Multiple Output

In the case of a MIMO transmission, the precoder, the channel and the receive filter are denoted as matrices as described in section 4.5.3 and shown in figure 19. The post equalization SINR is calculated for each layer on each RB assigned to the user. We again denote the number of layers used as ν_0, the number of transmit antennas at the BS as n_{TX} and the number of receive antennas at the user as n_{RX}.

Far User

Following the notation from section 4.5.3 used in equation (6), the post equalization SINR on layer l for one RB for the far user can be calculated as:

$$\gamma_{l,far} = \frac{\alpha_{far}\zeta_{l,far}P_{far}^{(l,0)}}{\alpha_{near}\zeta_{l,far}P_{far}^{(l,0)} + \sum_{i=1}^{N}\theta_{i,l,far}P_{far}^{(l,i)} + \psi_{l,far}\sigma^2}. \tag{28}$$

The signal power fraction $\zeta_{l,far}$ for layer l is defined as in equation (9) with the signal power fraction matrix $\mathbf{Z}$ as defined in equation (8). In the case of a ZF receiver the signal power matrix reduces to an identity matrix $\mathbf{Z} = \mathbf{I}_{\nu_0 \times \nu_0}$ and the signal power fraction on each layer is $\zeta_{l,far} = 1$. Thus, for a ZF receive filter, the post equalization SINR expression of the far user on layer l can be reduced to:

$$\gamma_{l,far} = \frac{\alpha_{far}P_{far}^{(l,0)}}{\alpha_{near}P_{far}^{(l,0)} + \sum_{i=1}^{N}\theta_{i,l,far}P_{far}^{(l,i)} + \psi_{l,far}\sigma^2}. \tag{29}$$

The interference from other cells is represented by the second term in the denominator. It remains unchanged compared to the OMA post equalization SINR in equation (6).

With the notation from equation (14) and equation (13), we can shorten the notation of the post equalization SINR for the far user on layer l to

$$\gamma_{l,far} = \frac{\alpha_{far} P_{far}^{(l,0)}}{\alpha_{near} P_{far}^{(l,0)} + I_{l,far} + N_{l,far}}. \tag{30}$$

Near User - SIC

As in the SISO case, the near user performs SIC to decode its signal. In a first step, the near user decodes the signal of the far user received at the near user. To model this transmission in a SL simulation, the post equalization SINR $\gamma_{l,near}^{(far)}$ of the signal of the far user received by the near user on each layer l is calculated as:

$$\gamma_{l,near}^{(far)} = \frac{\alpha_{far} \zeta_{l,near} P_{near}^{(l,0)}}{\alpha_{near} \zeta_{l,near} P_{near}^{(l,0)} + \sum_{i=1}^{N} \theta_{i,l,near} P_{near}^{(l,i)} + \psi_{l,near} \sigma^2} \tag{31}$$

$$= \frac{\alpha_{far} P_{near}^{(l,0)}}{\alpha_{near} P_{near}^{(l,0)} + I_{l,near} + N_{l,near}}. \tag{32}$$

Where we used the simpler notation for the interference $I_{l,near}$ and noise $N_{l,near}$, that has been introduced in equations (14) and (13), as well as the assumption of a ZF receive filter, that simplifies the signal power fraction $\zeta_{l,near}$ to $\zeta_{l,near} = 1$.
This post equalization SINR $\gamma_{l,near}^{(far)}$ of the signal of the far user received by the near user is again fed into the LPM to determine if the SIC has been successful. Then, if the signal of the far user can be decoded, the near user attempts the detection of its own signal, which is modeled in the SL simulation through the post equalization SINR $\gamma_{l,near}^{(near)}$ of the near user that can be expressed as

$$\gamma_{l,near}^{(near)} = \frac{\alpha_{near} \zeta_{l,near} P_{near}^{(l,0)}}{\varepsilon \alpha_{far} \zeta_{l,near} P_{near}^{(l,0)} + \sum_{i=1}^{N} \theta_{i,l,near} P_{near}^{(l,i)} + \psi_{l,near} \sigma^2} \tag{33}$$

$$= \frac{\alpha_{near} P_{near}^{(l,0)}}{\varepsilon \alpha_{far} P_{near}^{(l,0)} + I + N_l}. \tag{34}$$

Where again, the interference cancellation factor ε models the degree to which the far user signal has been canceled out by SIC.

5.4 Verification and Validation

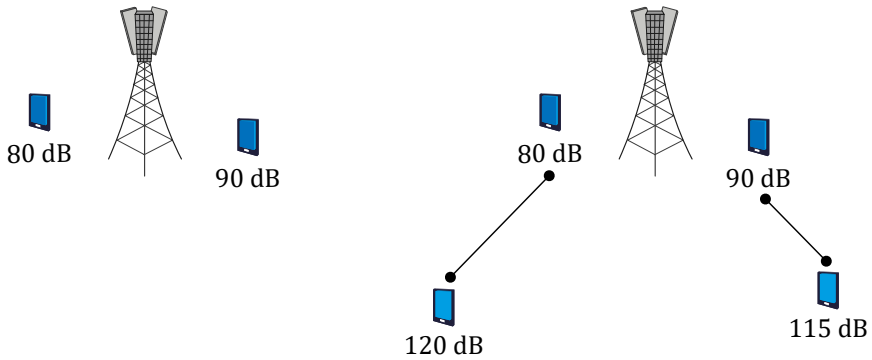
Verification

The implementation of NOMA transmissions in the VCCS 5G SLS has been verified through unit tests, that check the expected output of each function affected by the NOMA transmission, as well as integration tests in the form of NOMA scenario tests, that verify that the implemented features behave as expected in combination with each other and all other components of the simulator.
The MATLAB unit test framework [58] has been used for the implementation of these unit tests.

Validation

To validate the NOMA SL abstraction developed in the previous section, the simulation results is compared to the VCCS 5G LLS [17]. The following section describes the validation simulations and presents the simulation results. The plots in figures 23 to 25 validate the accuracy of the NOMA SL abstraction developed in this work for SISO, MISO and MIMO transmissions.

5.4.1 Simulation Scenario



(a) OMA (b) NOMA

Figure 21: Simulation scenario for validation by comparison with LL simulation.

The simulation scenario used for validation of the NOMA SL abstraction is modeled after the scenario described in [17]. The network setup is shown in figure 21. Two cells are simulated separately. In the LL simulation the cells are separated through a large path loss of 1000 dB. In the SL simulation two distinct simulations are performed. Each cell has two near users with a strong connection to the BS that are affected by a low, fixed path loss. The near users are shown in figure 21 with path losses of 80 dB and

Time and Frequency	
slot duration	1 ms
frequency	2.5 GHz
bandwidth	1.4 MHz
subcarrier spacing	15 kHz
Link Properties	
transmit mode	{1x1 SISO, 2x1 MISO, 2x2 MIMO}
small scale fading	Pedestrian A
correlated fading	true
user speed	1 m/s
feedback delay	1
precoder	3GPP TS 36.211 [19]
Path Loss	
user 1	80 dB
user 2	90 dB
user 3	120 dB
user 4	115 dB
Coding and Decoding	
MUST index	01
maximum far user CQI	6
coding	Turbo
decoding	Linear Log-MAP
decoding iterations	8

Table 2: Simulation parameters for validation by comparison with LL simulation.

90 dB. One cell has two additional far users placed in the simulation whose connection to the BS is affected by a large path loss. The far users are shown in figure 21b with path losses of 115 dB and 120 dB. The cell with the additional users will perform NOMA transmissions.
The NOMA user pairing is shown in figure 21b, the users are paired as described in section 5.1. The strongest user in the NOMA cell with a path loss of 80 dB is paired with the weakest user with a path loss of 120 dB. The two remaining users in the NOMA cell also form a NOMA user pair.
Other macroscopic fading mechanisms are not included in the simulation, i.e. no shadow

fading and no walls and wall losses are used and an omnidirectional antenna is used at the BSs.

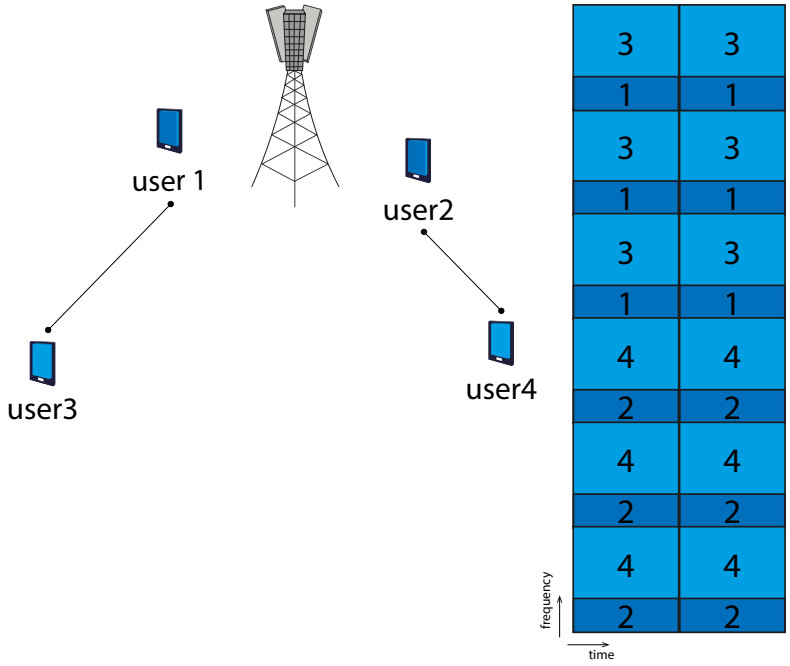


Figure 22: Scheduling in the NOMA cell for the LL validation simulation. In the OMA cell, users 3 and 4 are removed and the whole power is allocated to users 1 and 2.

To assure comparability the scheduling is fixed in each simulation. Each of the near users gets allocated half of the utilized system bandwidth. The scheduling in the NOMA cell is depicted in figure 22. The upper half of the bandwidth is allocated to the first NOMA pair consisting of users 1 and 3. The lower half of the bandwidth is allocated to the other NOMA pair, users 2 and 4. In the OMA cell, users 3 and 4 are removed and the bandwidth is divided between users 1 and 2.
The interference cancellation factor ε for NOMA transmissions is fixed to complete interference cancellation, i.e. $\varepsilon = 0$.
The power allocation factor α is set according to MUST mode 01. The modulation order of the far user is limited to quadrature phase shift keying (QPSK). In the SL simulation this is assured by limiting the CQI of the far user to a maximum value of 6.
The time and frequency parameters are chosen according to LTE standard settings with a carrier frequency of 2.5 GHz and a small bandwidth of 1.4 MHz. The small scale fading is modeled as a pedestrian A channel [49] with correlated fading and a user speed of 1 m/s.
The NOMA SL abstraction validation is performed for SISO, MISO and MIMO transmission modes, where the number of transmit and receive antennas is varied between 1 and 2. This means that for the SISO simulations all BSs have one transmit antenna

and all users have one receive antenna. For the 2x1 MISO simulation the number of transmit antennas at the BSs is increased to two and for the 2x2 MIMO simulations, the number of receive antennas at the users is also increased to two.
The transmit power of the BS is a sweep value and the simulations are performed for a range of transmit powers. An overview of the simulation parameters used is given in table 2.

5.4.2 Simulation Result

Calibration

Transmission Mode	Receiver Noise Figure
SISO	33 dB
MISO	33 dB
MIMO	31 dB

Table 3: Receiver noise figure set for calibrated SL simulation.

The OMA simulation is used to calibrate the SL simulator to match the results of the LL simulation. Without calibration, the throughput over BS transmit power curve of the SL simulator is shifted to the left. As a calibration parameter the receiver noise figure was chosen. With this parameter, the SNR in the SL simulation can be reduced and the throughput curve shifted to match the LL result. The calibrated noise figures for each transmission mode are shown in table 3. With these receiver noise figures, the result curves of the OMA simulation overlap, as can be seen below.

SISO

Figure 23 shows the cell throughput for each cell in the simulations. It can be seen that the results of the SL and the LL simulation overlap well. The maximum throughput for high transmit powers is slightly lower for OMA SL simulations than for the LL simulation. This small deviation is likely due to synchronization symbols that are included in the SL simulator, but not in the LL simulator and reduce the maximum throughput.

MISO

Figure 24 shows that the results of the LL and the SL evaluation overlap well. The discrepancy between the maximum throughput in the OMA scenario appears consistently for all transmission modes. For transmit powers between 0 dB and 30 dB the throughput increase is less steep in the SL simulation. This is likely due to the different

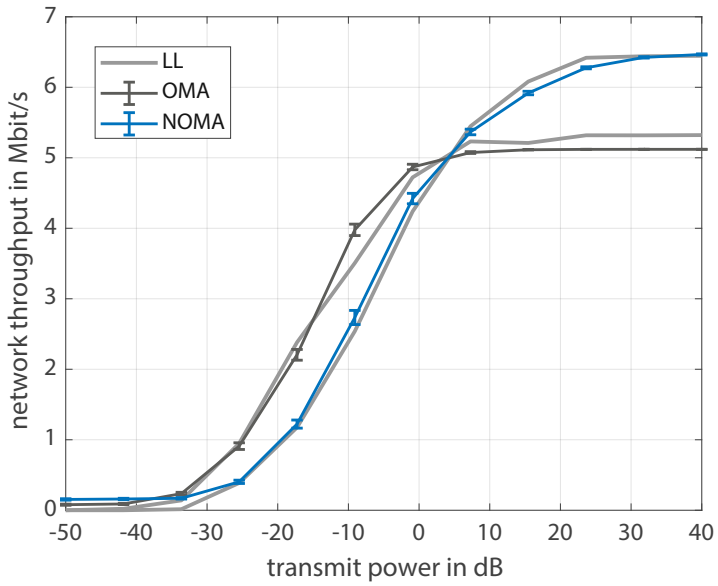


Figure 23: Cell throughput over BS transmit power for SISO transmission mode. The results of the LL simulation are shown in light gray and the results of the SL simulation are shown in dark gray and blue with 99% confidence intervals. The confidence intervals of the LL results are left out for better readability, they are in a similar range as those of the SL results.

implementations of the receiver in the LL and the SL simulator. In the LL simulation, a maximum likelihood (ML) receiver is used to detect the composite signal constellation of the combined near and far user signal. The SL simulation uses SIC as described above. The same behavior can be observed in figure 23, but is less pronounced in the SISO case.
The maximum achieved throughput of the MISO simulations is lower than that of the SISO simulations for NOMA and OMA in both the LL and the SL simulator. This can be explained by the increase in the number of reference symbols, which reduces the number of data symbols.

MIMO

In the results for MIMO transmission mode shown in figure 25, the throughput curves of the LL and the SL simulation overlap closely. The effects observed in the SISO and MISO results can be observed in the MIMO case as well. The maximum throughput of the SL OMA simulation is slightly lower than that of the LL simulation. The region in which the throughput increases more steeply in the LL simulation is smaller in the MIMO case and reduced to the transmit power range from 10 dB to 30 dB.

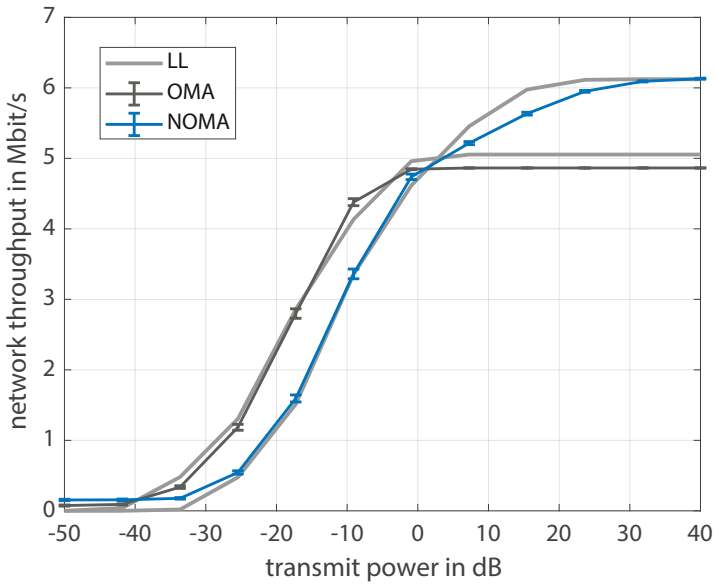


Figure 24: Cell throughput over BS transmit power for MISO transmission mode. The results of the LL simulation are shown in light gray and the results of the SL simulation are shown in dark gray and blue with 99% confidence intervals. The confidence intervals of the LL results are left out for better readability, they are in a similar range as those of the SL results.

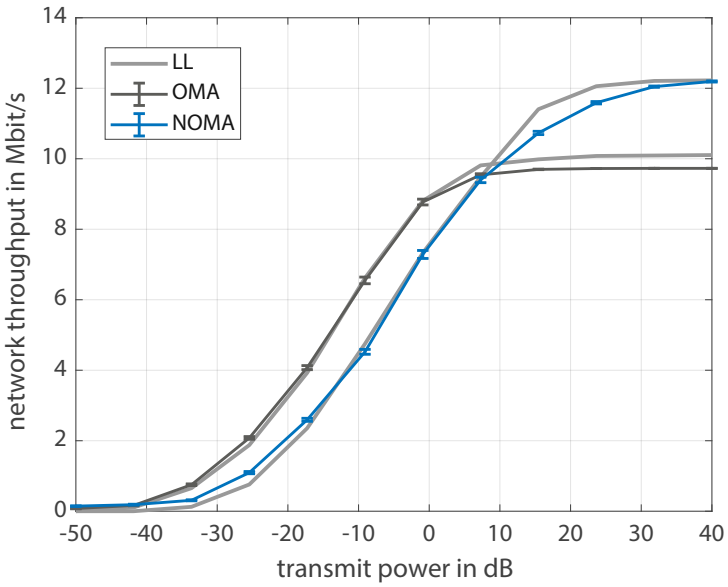


Figure 25: Cell throughput over BS transmit power for MIMO transmission mode. The results of the LL simulation are shown in light gray and the results of the SL simulation are shown in dark gray and blue with 99% confidence intervals. The confidence intervals of the LL results are left out for better readability, they are in a similar range as those of the SL results.

6 Evaluation of NOMA Performance

The theoretical evaluation of NOMA promises higher system throughput, lower latency and massive connectivity by allowing for a more efficient use of the resources and making them available to more users simultaneously. To evaluate the extent to which the theoretical gains are translated into practical networks, system level simulations are performed, where the network performance using NOMA transmissions is compared to a classical network without NOMA transmissions. The simulation scenario used for performance evaluation is described in section 6.1. The performance in terms of throughput, user connectivity and latency is evaluated in sections 6.2 to 6.4, respectively.

6.1 Simulation Scenario

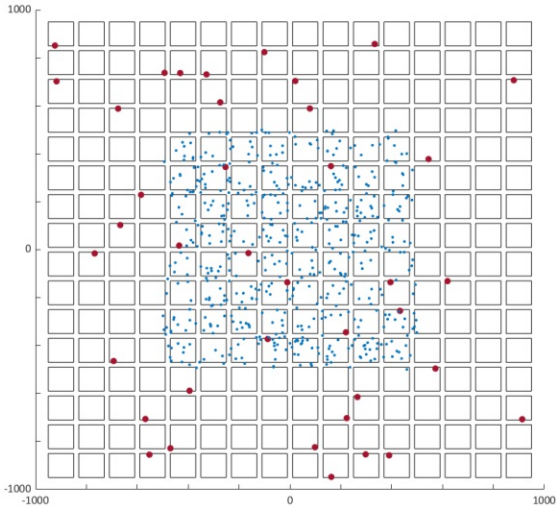


Figure 26: Example realization of a simulation scenario. The blue dots represent users and the red dots represent three sector BSs.

For the SL performance evaluation of NOMA, a Manhattan grid is chosen for the placement of buildings and BSs. The Manhattan grid is created with fixed building sizes and fixed street width, and randomly chosen building heights, that change for each realization. The BSs are placed randomly on an edge of a building with a probability of 0.2. The BSs antennas are placed 5 m above the building on which the BS is placed. Each BS is equipped with three 3 sector antennas, which are modeled as one transmit antenna element that radiates a power of up to 40 W with a sectorized radiation pattern. An interference region with additional BSs is simulated around the ROI. The users are placed in the simulation region through a PPP. This creates indoor users

Simulation Region	
length	1000 m
width	1000 m
interference region factor	2
Manhattan Grid	
building length	100 m
building width	100 m
minimum building height	40 m
maximum building height	80 m
wall loss	20 dB
street width	20 m
Base Station	
placement	on building
BS occupation probability	0.2
BS height	5 m
antenna	3 sector
n_{TX}	1
transmit power	40 W

Table 4: Network geometry parameters.

inside the buildings and outdoor users on the streets. The indoor users are separated from their serving BS through a wall with a wall loss of 20 dB. The wall loss experienced by the indoor users makes them viable candidates for paired NOMA transmissions with the outdoor users in the streets, that often have a LOS connection to the BS. The parameters of the network setup are collected in table 4.
The user parameters are shown in table 5. The users are placed through a PPP with a density of $600/\mathrm{km}^2$. Each user has one receive antenna and moves according to a random walk model with a speed of 4 km/h. The small scale fading is modeled as a pedestrian A channel. The traffic of each user is generated as a large amount of small packets. Each user has the maximum amount of data that can be transmitted in the simulation time with maximum CQI, as is elaborated below.
A carrier frequency of 1 GHz with a small bandwidth of 1.4 MHz has been chosen for the simulation. The small bandwidth has been chosen to simulate a crowded cell with more users than available resources. A short simulation duration of 11 slots is repeated over 100 realizations. An urban path loss model for macro BSs with an average building

User	
placement	PPP
λ	$600/\mathrm{km}^2$
n_{RX}	1
receiver noise figure	$9\ \mathrm{dB}$
channel model	Pedestrian A
speed	$4\ \mathrm{km/h}$
indoor decision	geometry
movement type	random walk
Traffic Model	
packet size	$1000\ \mathrm{bits}$
number of packets	59

Table 5: Parameters of the users.

height of $60\ \mathrm{m}$ is used. The macroscopic receive power is used for cell association and for NOMA user pairing. Two users are paired for NOMA transmission if the difference in their macroscopic receive powers Δ is larger than $15\ \mathrm{dB}$. The interference left from combined NOMA transmission after SIC is assumed to be zero. Further, MUST mode 01 is used for choosing the NOMA power allocation factor. The transmission parameters are listed in table 6.

Maximum Achievable Throughput

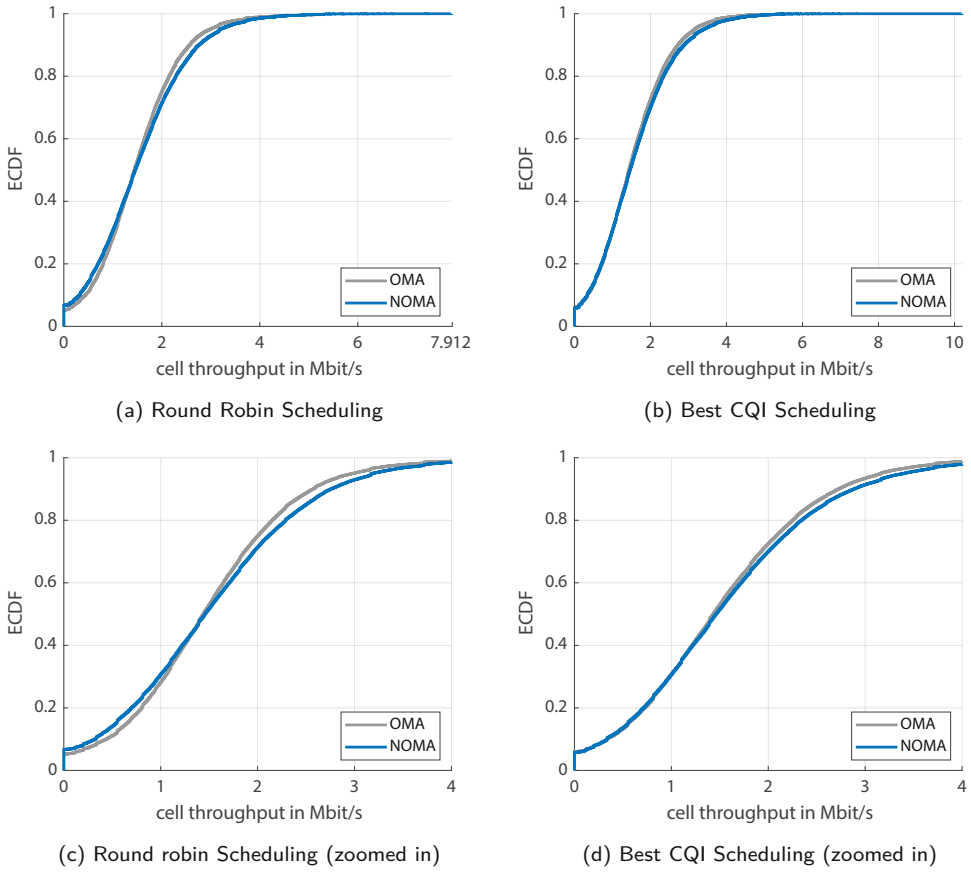
With a bandwidth of $1.4\ \mathrm{MHz}$, a total of 12 RBs is used and 80 data symbols are transmitted per RB. The maximum throughput is achieved with the maximum CQI of 15 for which 64-QAM and a coding rate of $\frac{948}{1024}\mathrm{data\ bit/coded\ bit}$ are used. Thus, the maximum number of data bits transmitted in an RB with OMA is 444 and the maximum achievable throughput in the simulated OMA network is $5328\ \mathrm{bits/slot}$.
In a NOMA network, the throughput per slot can be doubled since the near and far user can each transmit the maximum amount of bits. Thus, the maximum throughput is $10.656\ \mathrm{Mbit/s}$, which corresponds to $10656\ \mathrm{bits/slot}$. However, the far user usually does not achieve the maximum CQI of 15 and the achieved throughput is lower than the maximum throughput.

Time and Frequency	
frequency	2 GHz
bandwidth	1.4 MHz
slot duration	1 ms
feedback delay	1 slot
number of slots	11
realizations	100
Macroscopic Fading	
model	Urban Macro [53]
average building height	60 m
minimum coupling loss	70 dB
Cell Association and Scheduling	
cell association	macroscopic receive power
scheduler	round robin, best CQI
NOMA	
Δ	15 dB
MUST index	01
ε	0

Table 6: Transmission parameters.

6.2 Throughput Gain

One of the benefits promised by NOMA simulations is an increase in the network throughput, since NOMA allows for an efficient use of the available resources. To examine the extent to which additional throughput can be achieved through NOMA transmissions, we compare the cell throughput in a network with NOMA transmissions with the cell throughput in a network without NOMA. The ECDF of the cell throughput is shown for two different scheduling strategies in figure 27. It can be seen that the network with NOMA transmissions achieves much larger maximum cell throughput, while the OMA network is limited to the maximum throughput of 5.328 Mbit/s. Figure 27a shows that the maximum throughput with a round robin scheduler is increased to 7.912 Mbit/s and figure 27b shows that with the best CQI scheduler a maximum throughput of 10.2 Mbit/s is achieved, which is close to the maximum achievable throughput of 10.656 Mbit/s.
In figures 27a and 27c, it can be seen that more cells have no throughput in the NOMA



(a) Round Robin Scheduling

(b) Best CQI Scheduling

(c) Round robin Scheduling (zoomed in)

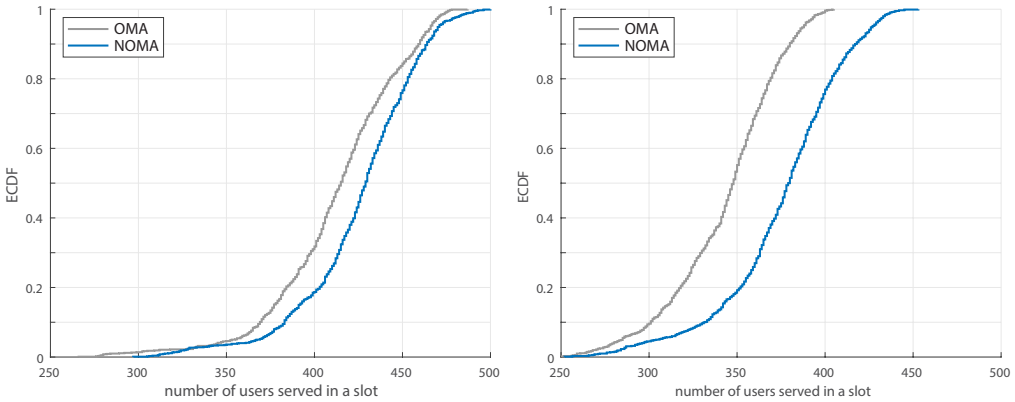
(d) Best CQI Scheduling (zoomed in)

Figure 27: ECDF of cell throughput.

case than in the OMA case. This can be explained by the reduction in SINR experienced by NOMA users. The additional interference and reduced transmit power for NOMA users leads to more users falling below the SINR threshold for possible transmissions and thus reduces the cell throughput. This behavior is not observed for the best CQI scheduler in figures 27b and 27d, since the best CQI scheduler only schedules near users whose channel quality is high enough to allow successful transmissions. The majority of cells with no throughput are cells without users in them.

6.3 Massive Connectivity

Another benefit of NOMA is that the number of users served in a cell can be increased. In theory, the number of users can be doubled. The increase in number of users served is analyzed by comparing the number of users served per slot in a network with NOMA to the number of users served in an OMA network. Figure 28 shows the ECDF of the number of users that receive data in a slot for two scheduling strategies. For the round

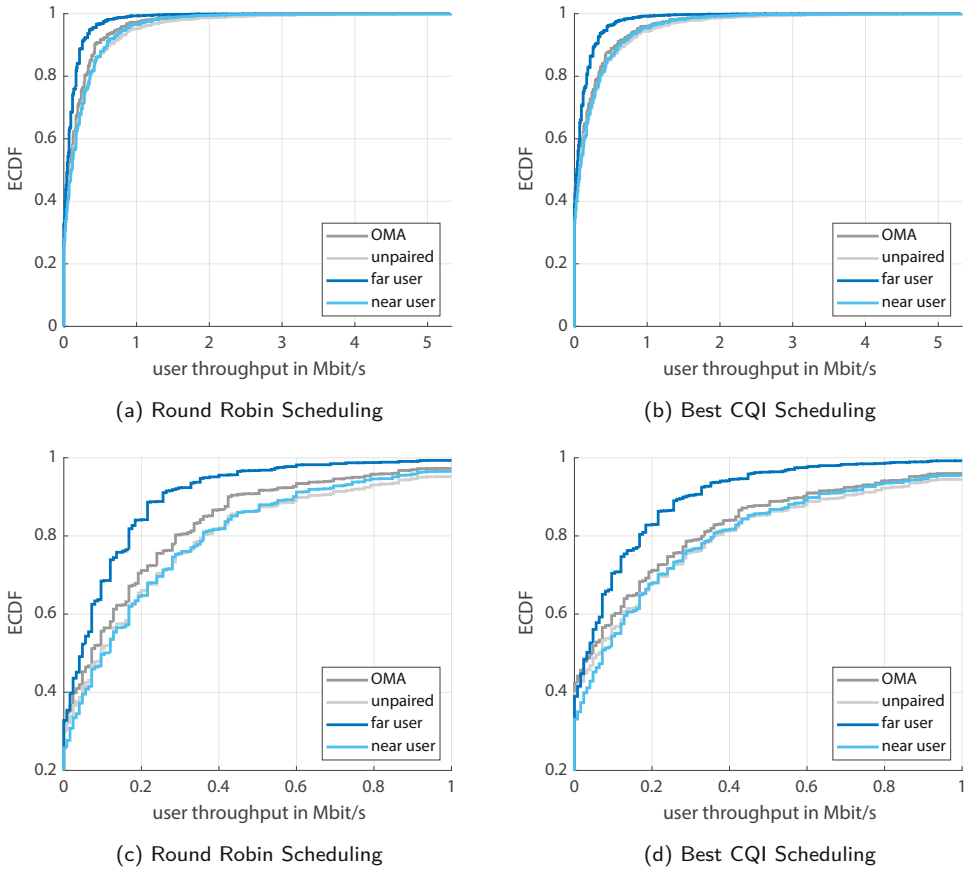


(a) Round Robin Scheduling (b) Best CQI Scheduling

Figure 28: ECDF of number of users served per slot for different scheduling algorithms.

robin scheduler in figure 28a, on average 406 of the 600 users are served in a slot in the OMA network, while 422 users are served on average in the NOMA network, this is an increase of 4%. In case of the best CQI scheduler, the average number of users served per slot increases from 345 to 373, which is an increase of 8%. The increase is more pronounced for the best CQI scheduler, since, with this scheduling strategy, users with bad channel conditions are not served at all in the OMA network. The ECDFs in figure 28 show that more users can be served in a network with NOMA independent of the scheduling strategy.

Figure 29 shows the ECDF of the user throughput for each user category. For the NOMA network, the ECDF is shown for the near users, the far users and the users that have not been paired for NOMA transmission and use OMA techniques. For the OMA network, only one type of users exists that use OMA techniques for transmission. Figure 29 shows the trade off made with NOMA transmission. The far users achieve lower throughput than the other user groups since they suffer from poor channel conditions. The performance of the OMA users lies between that of the near and the far users, since the group of OMA users contains all users, those with good and those with poor channel conditions. The unpaired users outperform the far users in the high throughput region, since their transmission is not affected by the transmission of the far user data. In the low throughput region, there are fewer near users, since near users have good channel conditions and can achieve high throughput, while the group of unpaired users contains both users with good and poor channel conditions.



(a) Round Robin Scheduling (b) Best CQI Scheduling

(c) Round Robin Scheduling (d) Best CQI Scheduling

Figure 29: ECDF of user throughput.

Fairness

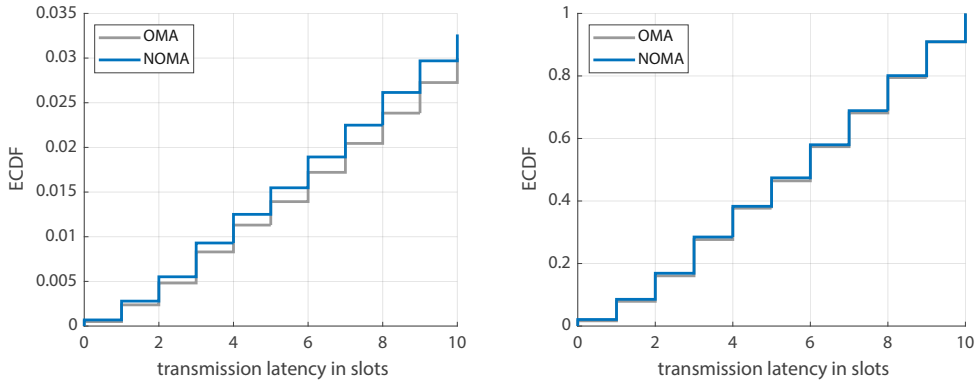
A measure for fairness is Jain's fairness index [59]

$$J(T) = \frac{\left(\sum_{k=1}^{K} T(k)\right)^2}{K \sum_{k=1}^{K} T(k)^2}, \tag{35}$$

where K is the number of users and $T(k)$ is the user throughput.
Considering the fairness over all network realizations, Jain's fairness index increases by 1% when introducing NOMA in the network. In the OMA system $J(T) = 25.8\%$ and in the NOMA system $J(T) = 26.7\%$.
In conclusion, NOMA allows to increase the number of users served in a network, at the cost of slightly reducing the throughput achieved by users with good channel conditions. Thus NOMA increases fairness and enables massive connectivity.

6.4 Latency Gain

Since more users are served at the same time with NOMA, the network latency can be reduced. The latency reduction is analyzed for a round robin scheduling strategy.



(a) ECDF of latency of all generated packets.

(b) ECDF of latency of transmitted packets.

Figure 30: ECDF of transmit latency in slots.

Figure 30 shows the ECDF of packet latency in an OMA and a NOMA network. Since the traffic to transmit is chosen to be the maximum amount of data that a user can transmit in the simulation duration, a large amount of packets is left in the transmit buffer at the end of a simulation. In figure 30a, it can be seen that only about 3% of the packets are transmitted. It can also be seen that more packets are transmitted in a NOMA simulation than in a OMA simulation, which reduces the network latency.
Figure 30b shows the ECDF of the transmitted packets in an OMA and NOMA network. This shows that even without taking into consideration that more packets are transmitted in the NOMA network, the packet latency in the NOMA network is slightly lower than that in the OMA network.

6.5 Discussion

Section 5.4 shows that the SL abstraction model for NOMA transmission developed in section 5 is a valid SL abstraction. However, the discrepancy between LL and SL result that is circumvented through the calibration of the receiver noise figure remains unexplained. The difference in maximum throughput achieved in OMA SL and OMA LL simulations also remains unexplained. The assumption of perfect interference cancellation modeled through a choice of $\varepsilon = 0$ used for validation and performance evaluation is not realistic. The interference cancellation factor ε should increase with the BLER experienced for the decoding of the far user signal at the near user. Further, the comparison between LL and SL simulation is flawed, since the decoding of the NOMA signals is performed through SIC in the SL simulation, but through ML detection of the combined signal constellation on the LL. Since both detection methods are optimal, the comparison is still viable, especially at sufficiently high SINR, where near-perfect interference cancellation can be guaranteed.

The performance comparison between NOMA and OMA in section 6 shows that the promised theoretical gains of NOMA are small in a practical environment. Gains in terms of throughput are only achieved in very few cases and the effectiveness of NOMA relies on the existence of viable user pairs with high differences in channel quality. Thus, NOMA gains can only be achieved in highly loaded cells with specific user constellations. In future radio networks with highly directive antennas, the occurrence of viable NOMA pairs might not be given. However, in IoT scenarios, with very large numbers of users, the advantage of NOMA allowing massive connectivity might outweigh the downsides and the occurrence of viable NOMA user pairs might be increased.

NOMA is a viable technique to increase fairness in the network, since it can increase the user experience for the far user while only slightly impacting the QoS of near user. However, if no users with good channel conditions exist in the network, NOMA can reduce the number of successful transmissions. A possible benefit of NOMA that has not been investigated in this work is that it allows the pairing of users with high traffic volume to users with low traffic volume with low impact on the QoS of the high throughput user.

The performance comparison should be extended to explore questions that are left open after this analysis. The network setup for performance comparison of NOMA and OMA was designed to create users that are viable for NOMA pairing by placing users indoor of building with large wall losses. It should be investigated what channel conditions are present in practical networks to evaluate the usefulness of NOMA techniques. The investigation of the effects of NOMA on the packet latency is limited to a short simulation duration and thus can not explore all effects. Longer periods should be observed to explore the effects of NOMA. The effect of different user pairing strategies has not been taken into consideration. In this work, the maximum difference in channel quality was maximized. Maximizing the average difference in channel quality of all

NOMA pairs might lead to better performance. Another factor that has not been investigated is the necessary difference in channel conditions. The optimal difference in channel quality Δ can increase the number of NOMA pairs and reduce the number of failed transmissions, increasing the throughput. Further, the user pairing should not only rely on the difference Δ, but also consider the absolute channel quality, since pairing a user with low channel quality with a user with very low channel quality will lead to failed transmissions instead of increasing the throughput.

7 Conclusion

In this work, a SL model for two user power domain NOMA transmission with SIC was developed and validated through comparison with LL simulation. In the performance evaluation of NOMA transmissions, the improvements projected by theoretical evaluation of NOMA were found to be reduced in practical systems. NOMA only slightly increases the network throughput, but achieves good improvements in terms of user connectivity and latency. The number of users served in a cell and the throughput of users with poor channel conditions can be improved at the cost of increased receiver complexity and reduced throughput at the near NOMA user. However, the effectiveness of NOMA is tied to the presence of viable user pairs for NOMA transmission. For NOMA to reach its full potential a near user with good channel conditions and a far user with poor channel conditions need to be available for NOMA pairing. This might not be satisfied in large scale practical networks.

7.1 Outlook

This work is limited to two user NOMA. In a future work, NOMA transmission for more than two users can be considered. Further, a more realistic model for the interference cancellation factor ε needs to be developed, that considers the achieved BLER of the detection of the far user signal at the near user. Then, the influence of the interference cancellation factor ε can be evaluated. The combination of NOMA with other future radio access technologies such as beamforming should also be investigated. In this work, the optimal precoder for the far user is chosen for NOMA transmission. A precoder designed for combined NOMA transmission could yield better results.
The power allocation in NOMA transmissions was not analyzed. The influence of the power allocation factor α should be evaluated and the choice of the power allocation factor α should be optimized to increase the efficiency of NOMA.
The user pairing in this work is based on a simple user pairing strategy that maximizes the difference in channel condition Δ of the first NOMA user pair. This strategy can be improved to maximize the average difference in channel condition for all NOMA pairs. Another approach could be to draw far users for user pairing from neighboring cells, which could further increase the user experience of cell edge users and allow managing cell load more flexibly.

8 References

[1] Ericsson, "Average monthly smartphone traffic worldwide from 2015 to 2026 (in exabytes).." https://www.statista.com/statistics/739002/worldwide-smartphones-monthly-data-traffic/, Jan. 2021. Chart.

[2] E. Dahlman, *4G LTE/LTE-Advanced for mobile broadband*. Amsterdam Boston: Elsevier/Academic Press, 2011.

[3] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network; Study on Downlink Multiuser Superposition Transmission (MUST) for LTE," TR 36.859, 3rd Generation Partnership Project (3GPP), Dec. 2015.

[4] L. Zhang, W. Li, Y. Wu, X. Wang, S.-I. Park, H. M. Kim, J.-Y. Lee, P. Angueira, and J. Montalban, "Layered-Division-Multiplexing: Theory and Practice," *IEEE Transactions on Broadcasting*, vol. 62, pp. 216–232, Mar. 2016.

[5] Z. Ding, P. Fan, and H. V. Poor, "Impact of User Pairing on 5G Nonorthogonal Multiple-Access Downlink Transmissions," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 6010–6023, Aug. 2016.

[6] C. Xiao, J. Zeng, W. Ni, X. Su, R. P. Liu, T. Lv, and J. Wang, "Downlink MIMO-NOMA for Ultra-Reliable Low-Latency Communications," *IEEE Journal on Selected Areas in Communications*, vol. 37, pp. 780–794, Apr. 2019.

[7] M. Amjad and L. Musavian, "Performance Analysis of NOMA for Ultra-Reliable and Low-Latency Communications," in *2018 IEEE Globecom Workshops (GC Wkshps)*, IEEE, Dec. 2018.

[8] A. Akbar, S. Jangsher, and F. A. Bhatti, "NOMA and 5G emerging technologies: A survey on issues and solution techniques," *Computer Networks*, vol. 190, p. 107950, May 2021.

[9] M. K. Müller, F. Ademaj, T. Dittrich, A. Fastenbauer, B. R. Elbal, A. Nabavi, L. Nagel, S. Schwarz, and M. Rupp, "Flexible multi-node simulation of cellular mobile communications: the Vienna 5G System Level Simulator," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, p. 17, Sept. 2018.

[10] J. Lee, M. Han, M. Rim, and C. G. Kang, "5G K-SimSys for Open/Modular/Flexible System-Level Simulation: Overview and its Application to Evaluation of 5G Massive MIMO," *IEEE Access*, vol. 9, pp. 94017–94032, 2021.

[11] S. Martiradonna, A. Grassi, G. Piro, and G. Boggia, "5G-air-simulator: An open-source tool modeling the 5G air interface," *Computer Networks*, vol. 173, p. 107151, May 2020.

[12] T. Deinlein, R. German, and A. Djanatliev, "5G-Sim-V2I/N: Towards a Simulation Framework for the Evaluation of 5G V2I/V2N Use Cases," in *2020 European Conference on Networks and Communications (EuCNC)*, IEEE, Jan. 2020.

[13] C.-K. Jao, C.-Y. Wang, T.-Y. Yeh, C.-C. Tsai, L.-C. Lo, J.-H. Chen, W.-C. Pao, and W.-H. Sheen, "WiSE: A System-Level Simulator for 5G Mobile Networks," *IEEE Wireless Communications*, vol. 25, pp. 4–7, Apr. 2018.

[14] M. Liu, P. Ren, Q. Du, W. Ou, X. Xiong, and G. Li, "Design of system-level simulation platform for 5G networks," in *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, IEEE, July 2016.

[15] Y. Wang, J. Xu, and L. Jiang, "Challenges of System-Level Simulations and Performance Evaluation for 5G Wireless Networks," *IEEE Access*, vol. 2, pp. 1553–1561, 2014.

[16] D. Wang, R. R. Sattiraju, A. Weinand, and H. D. Schotten, "System-Level Simulator of LTE Sidelink C-V2X Communication for 5G," in *Mobile Communication - Technologies and Applications; 24. ITG-Symposium*, pp. 1–5, 2019.

[17] S. Pratschner, B. Tahir, L. Marijanovic, M. Mussbah, K. Kirev, R. Nissel, S. Schwarz, and M. Rupp, "Versatile mobile communications simulation: the Vienna 5G Link Level Simulator," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, p. 226, Sept. 2018.

[18] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network; Study on Non-Orthogonal Multiple Access (NOMA) for NR," TR 38.812, 3rd Generation Partnership Project (3GPP), Dec. 2018.

[19] 3rd Generation Partnership Project (3GPP), "Evolved Universal Terrestrial Radio Access (E-UTRA) physical channels and modulation," TS 36.211, 3rd Generation Partnership Project (3GPP), Jan. 2015.

[20] M. Y. Alias, S. Chen, and L. Hanzo, "Multiple-Antenna-Aided OFDM Employing Genetic-Algorithm-Assisted Minimum Bit Error Rate Multiuser Detection," *IEEE Transactions on Vehicular Technology*, vol. 54, pp. 1713–1721, Sept. 2005.

[21] R. Razavi, R. Hoshyar, M. A. Imran, and Y. Wang, "Information Theoretic Analysis of LDS Scheme," *IEEE Communications Letters*, vol. 15, pp. 798–800, Aug. 2011.

[22] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "CTH06-4: Novel Low-Density Signature Structure for Synchronous DS-CDMA systems," in *IEEE Globecom 2006*, IEEE, Nov. 2006.

[23] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel Low-Density Signature for Synchronous CDMA Systems Over AWGN Channels," *IEEE Transactions on Signal Processing*, vol. 56, pp. 1616–1626, Apr. 2008.

[24] R. Hoshyar, R. Razavi, and M. Al-Imari, "LDS-OFDM an Efficient Multiple Access Technique," in *2010 IEEE 71st Vehicular Technology Conference*, IEEE, 2010.

[25] M. Al-Imari, M. A. Imran, and R. Tafazolli, "Low Density Spreading for Next Generation Multicarrier Cellular Systems," in *2012 International Conference on Future Communication Networks*, IEEE, Apr. 2012.

[26] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern Division Multiple Access—A Novel Nonorthogonal Multiple Access for Fifth-Generation Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 3185–3196, Apr. 2017.

[27] B. M. ElHalawany and K. Wu, "Physical-Layer Security of NOMA Systems Under Untrusted Users," in *2018 IEEE Global Communications Conference (GLOBECOM)*, IEEE, Dec. 2018.

[28] G. B. Satrya and S. Y. Shin, "Security enhancement to successive interference cancellation algorithm for non-orthogonal multiple access (NOMA)," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, IEEE, Oct. 2017.

[29] B. He, A. Liu, N. Yang, and V. K. N. Lau, "On the Design of Secure Non-Orthogonal Multiple Access Systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 2196–2206, Oct. 2017.

[30] Y. Sun, D. W. K. Ng, J. Zhu, and R. Schober, "Robust and Secure Resource Allocation for Full-Duplex MISO Multicarrier NOMA Systems," *IEEE Transactions on Communications*, vol. 66, pp. 4119–4137, Sept. 2018.

[31] Z. Qin, Y. Liu, Z. Ding, Y. Gao, and M. Elkashlan, "Physical layer security for 5G non-orthogonal multiple access in large-scale Networks," in *2016 IEEE International Conference on Communications (ICC)*, IEEE, May 2016.

[32] J. Tang, L. Jiao, N. Wang, P. Wang, K. Zeng, and H. Wen, "Mobility Improves NOMA Physical Layer Security," in *2018 IEEE Global Communications Conference (GLOBECOM)*, IEEE, Dec. 2018.

[33] Y. Zhang, H.-M. Wang, Q. Yang, and Z. Ding, "Secrecy Sum Rate Maximization in Non-orthogonal Multiple Access," *IEEE Communications Letters*, vol. 20, pp. 930–933, May 2016.

[34] S. Thapar, D. Mishra, and R. Saini, "Decoding Orders for Securing Untrusted NOMA," *IEEE Networking Letters*, vol. 3, pp. 27–30, Mar. 2021.

[35] H. Sun, B. Xie, R. Q. Hu, and G. Wu, "Non-Orthogonal Multiple Access with SIC Error Propagation in Downlink Wireless MIMO Networks," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, IEEE, Sept. 2016.

[36] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the Performance of Non-orthogonal Multiple Access Systems With Partial Channel Information," *IEEE Transactions on Communications*, vol. 64, pp. 654–667, Feb. 2016.

[37] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, IEEE, June 2013.

[38] K. Higuchi and A. Benjebbour, "Non-orthogonal Multiple Access (NOMA) with Successive Interference Cancellation for Future Radio Access," *IEICE Transactions on Communications*, vol. E98.B, no. 3, pp. 403–414, 2015.

[39] K. Higuchi and Y. Kishiyama, "Non-Orthogonal Access with Random Beamforming and Intra-Beam SIC for Cellular MIMO Downlink," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, IEEE, Sept. 2013.

[40] N. Nonaka, Y. Kishiyama, and K. Higuchi, "Non-Orthogonal Multiple Access Using Intra-Beam Superposition Coding and SIC in Base Station Cooperative MIMO Cellular Downlink," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, IEEE, Sept. 2014.

[41] S. Mishra, L. Salaun, C. S. Chen, and K. Giridhar, "Analysis of Downlink Connectivity in NB-IoT Networks Employing NOMA with Imperfect SIC," in *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, IEEE, June 2021.

[42] H. Lee, S. Kim, and J.-H. Lim, "Multiuser Superposition Transmission (MUST) for LTE-A systems," in *2016 IEEE International Conference on Communications (ICC)*, IEEE, May 2016.

[43] Y. Liu, Z. Qin, M. Elkashlan, A. Nallanathan, and J. A. McCann, "Non-Orthogonal Multiple Access in Large-Scale Heterogeneous Networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 2667–2680, Dec. 2017.

[44] M. Rupp, S. Schwarz, and M. Taranetz, *The Vienna LTE-Advanced Simulators: Up and Downlink, Link and System Level Simulation*. Signals and Communication Technology, Singapore: Springer, 2016.

[45] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and Analysis of Cellular Networks Using Stochastic Geometry: A Tutorial," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 167–203, 2017.

[46] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi, "End-to-End Simulation of 5G mmWave Networks," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2237–2263, 2018.

[47] T. R. Henderson, M. Lacage, G. F. Riley, C. Dowell, and J. Kopena, "Network simulations with the ns-3 simulator," *SIGCOMM demonstration*, vol. 14, no. 14, p. 527, 2008.

[48] A. F. Molisch, *Wireless Communications*. John Wiley & Sons, Nov. 2010.

[49] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network; High Speed Downlink Packet Access: UE Radio Transmission and Reception," TR 25.890, 3rd Generation Partnership Project (3GPP), May 2002.

[50] 3rd Generation Partnership Project (3GPP), "Radio transmission and reception, annex c.3 propagation models," TS 38.101, 3rd Generation Partnership Project (3GPP), 2009.

[51] 3rd Generation Partnership Project (3GPP), "Universal Mobile Telecommunications System (UMTS) Deployment aspects," TR 25.943, 3rd Generation Partnership Project (3GPP), Feb. 2010.

[52] A. Fastenbauer, M. K. Müller, and M. Rupp, "Investigation of Wraparound Techniques for the Simulation of Wireless Cellular Networks," in *WSA 2019; 23rd International ITG Workshop on Smart Antennas*, 2019.

[53] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network; Study on channel model for frequencies from 0.5 to 100GHz," TR 38.901, 3rd Generation Partnership Project (3GPP), Dec. 2017.

[54] T. Dittrich, M. Taranetz, and M. Rupp, "An Efficient Method for Avoiding Shadow Fading Maps in System Level Simulations," in *WSA 2017; 21st International ITG Workshop on Smart Antennas*, pp. 1–8, 2017.

[55] M. H. Habaebi, J. Chebil, A. G. Al-Sakkaf, and T. H. Dahawi, "Comparison between Scheduling Techniques in Long Term Evolution," *IIUM Engineering Journal*, vol. 14, Mar. 2013.

[56] J. C. Ikuno, S. Pendl, M. Simko, and M. Rupp, "Accurate SINR estimation model for system level simulation of LTE networks," in *2012 IEEE International Conference on Communications (ICC)*, IEEE, June 2012.

[57] S. Schwarz, C. Mehlführer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE," in *6th conference on Wireless advanced (WiAD)*, pp. 1–6, IEEE, 2010.

[58] MATLAB, "Class-Based Unit Tests." https://de.mathworks.com/help/matlab/class-based-unit-tests.html, 2017. R2017b.

[59] R. Jain, D.-M. Chiu, and W. R. Hawe, "A Quantitative Measure Of Fairness and Discrimination for Resource Allocation in Shared Computer Systems," *Tech. Rep. TR-301, DEC*, Sept. 1984.

Table of Abbreviations

3D	three dimensional
3GPP	3rd generation partnership project
1G	first generation
2G	second generation
3G	third generation
4G	fourth generation
5G	fifth generation
6G	sixth generation
ATSC	advanced television systems committee
BLER	block error ratio
BS	base station
CDMA	code division multiple access
CQI	channel quality indicator
CSI	channel state information
DL	downlink
ECDF	empirical cumulative distribution function
FDMA	frequency division multiple access
IoT	internet of things
LDM	layered-division multiplexing
LDS	low density signature
LL	link level
LLS	link level simulator
LOS	line of sight
LPM	link performance model
LQM	link quality model
LTE	long term evolution
MCS	modulation and coding scheme
MIESM	mutual information effective signal to interference and noise ratio mapping
MIMO	multiple input multiple output
MISO	multiple input single output
ML	maximum likelihood
MUST	multiuser superposition transmission
NLOS	non line of sight
NOMA	non-orthogonal multiple access
OFDMA	orthogonal frequency division multiple access
OMA	orthogonal multiple access
OOP	object-oriented programming
PDMA	pattern division multiple access
PDP	power delay profile

PMI	precoding matrix indicator
PPP	Poisson point process
QoS	quality of service
QPSK	quadrature phase shift keying
RB	resource block
RI	rank indicator
ROI	region of interest
SDMA	spatial division multiple access
SIC	successive interference cancellation
SINR	signal to interference and noise ratio
SISO	single input single output
SL	system level
SLS	system level simulator
SNR	signal to noise ratio
SU-MIMO	single user multiple input multiple output
TDMA	time division multiple access
UL	uplink
UMa	urban macro
VCCS	Vienna cellular communications simulators
ZF	zero forcing