Master´s thesis

Fakultät für
Informations-, Medien-
und Elektrotechnik

Technology
Arts Sciences
TH Köln

Remediation of High-Load Scenarios Using a Network Digital Twin

Name: Kathrin Pauline Retz

Student ID: 11198015

Program: Master Communication Systems and Networks

Name 1st examiner: Prof. Dr. Andreas Grebe

Name 2nd examiner: Dr. Heiko Lehmann

Deadline: 03.11.2025

Declaration:
I hereby confirm on my honour that this thesis is the result of my independent work.
All sources and auxiliary material used in this thesis are cited completely.

Place, date, signature: Köln, 03.11.2025, 

Abstract

Mobile networks frequently face high-load situations, especially during temporary events or local demand peaks, which can push the network into congestion. Defining when such load becomes critical and finding efficient approaches for maintaining the Quality of Service (QoS) for users are important challenges for network operators. This thesis examines and compares several remediation strategies for alleviating resource-limited high-load scenarios in 5G networks, using a digital twin-based simulation environment.

This thesis employs the ns-3 Playground, a toolkit that combines ns-3, SUMO, and Sionna to recreate realistic radio propagation (raytracing), user mobility patterns, and packet-based network dynamics. The core scenario models a resource-limited overload with one macro gNodeB and 60 user equipments (UEs), reflecting real-world congestion rather than interference-limited cases. Three remediation methods are evaluated: (1) adding an additional small cell (with various positioning strategies), (2) increasing the gNodeB's transmit power, and (3) scaling channel bandwidth.

Results show that placing a small cell at the demand hotspot (Hotspot overlay strategy) produced the strongest load reduction and sustained good throughput and SINR, making it the most effective remediation. Bandwidth scaling to 15 MHz emerged as the next-best option, but requires regulatory access to spectrum. Increasing transmit power did not measurably improve congestion in resource-limited conditions. In terms of implementation effort, bandwidth scaling is less demanding than additional cell deployment, but not always feasible due to spectrum constraints. The ns-3 Playground proved to be a flexible and precise simulation platform, allowing detailed and reproducible evaluation of different strategies and surpassing the capabilities of conventional simulators for spatially-aware, event-driven analysis.

These findings provide guidance for network planners addressing temporary or localized capacity problems and demonstrate the benefits of using digital twin principles for data-driven network evaluation.

Keywords: Network High-Load, Remediation Methods, 5G, Network Digital Twin, ns-3, Traffic Modeling

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AM Acknowledged Mode.

AMRs Autonomous Mobile Robots.

BLER Block Error Rate.

CBR Constant Bit Rate.

COWs Cells on Wheels.

CQI Channel Quality Indicator.

EIRP Equivalent Isotropic Radiated Power.

ETSI European Telecommunications Standards Institute.

gNodeB, gNB Next Generation Node B.

HARQ Hybrid Automatic Repeat reQuest.

ITU International Telecommunication Union.

MAC Medium Access Control.

MCS Modulation and Coding Schemes.

MIMO Multiple Input Multiple Output.

OFDM Orthogonal Frequency-Division Multiplexing.

PDCP Packet Data Convergence Protocol.

PHY Physical Layer.

PRB Physical Resource Block.

QoS Quality of Service.

RAN Radio Access Network.

RLC Radio Link Control.

RSRP Reference Signal Received Power.

SINR Signal-to-Interference-plus-Noise Ratio.

SISO Single Input Single Output.

SNR Signal-to-Noise Ratio.

UE User Equipment.

1. Introduction

1.1. Background

The number of mobile network users has increased rapidly over the past years, driven by the continuous emergence of new applications, technologies, and data-intensive services that rely on mobile connectivity. As a result, network operators must continuously adapt and expand their infrastructure to meet the growing demand for reliable connectivity and high-data rates. This ongoing evolution requires strategic optimization and careful planning to ensure that networks can provide sufficient capacity and maintain the desired level of user experience.

Analyzing recurring patterns in network load data enables network operators to understand user activity trends and network behavior, which is fundamental for effective performance optimization. However, these patterns are not always consistent. Certain circumstances, such as large public events, concerts, or sports gatherings, can lead to temporary but extreme deviations from the usual traffic profile. Similarly, sudden changes in user density or demand within a single cell area can cause short-term traffic surges that push the network close to its operational limits.

When such dynamic shifts in traffic intensity bring the network near or beyond its capacity boundaries, service degradation may occur. These periods are known as high-load scenarios when the system still provides acceptable Quality of Service, and as overload or congestion scenarios when performance degradation becomes observable. Managing and mitigating such high-load conditions before they evolve into overload is a key challenge in mobile network engineering.

A variety of remediation methods exist to alleviate the effects of network high-load, including the deployment of additional small cells to offload traffic, adjustment of transmission power to balance cell coverage, and dynamic scaling of bandwidth resources to increase throughput. To apply these methods effectively, it is first necessary to identify when high load occurs by defining appropriate thresholds. Based on these insights, suitable remediation strategies can then be selected and adapted to specific network conditions.

This thesis investigates these remediation strategies through rigorous simulation, identifying which are most practical for rapid deployment during temporary high-load

scenarios, particularly hotspot-targeted small cell placement. From an operational perspective, rapid remediation is critical: service interruptions during peak events directly impact operator revenue and customer satisfaction, making rapid deployment a practical necessity.

1.2. Problem Definition

Mobile networks frequently encounter temporary high-load situations that can lead to overload and congestion, and thus, degradation of user experience. To ensure reliable service quality, it is essential to identify measures that effectively alleviate high-load conditions and maintain network performance under stress.

This thesis examines different remediation strategies that can be applied during high-load scenarios to improve the overall system capacity and quality of service. The methods considered include the deployment of additional small cells, adjustment of transmission power, and scaling of available bandwidth. Each approach presents unique benefits and trade-offs in terms of implementation complexity, network coverage, and efficiency.

Evaluating these methods in operational networks is challenging due to the risks and costs associated with real-world testing. Therefore, a Network Digital Twin, the ns-3 Playground environment, is employed to simulate realistic high-load conditions and remediation effects. The integrated framework combines ns-3 for packet-level network simulation, Sionna for physical-layer and propagation modeling, and SUMO for mobility simulation, allowing detailed and reproducible analyzes of complex network behaviors.

Compared to traditional network simulators, the ns-3 Playground provides a more flexible and precise environment through its multi-tool integration, enhanced environmental realism, and support for coupled radio and mobility modeling. This makes it particularly suitable for evaluating the performance and effectiveness of remediation techniques in dynamic high-load mobile network scenarios.

1.3. Structure of the Thesis

The remainder of this thesis is structured as follows.

Chapter 2 introduces the theoretical background and fundamental concepts of mobile networks, network high-loads, and remediation approaches. It also explains the Network Digital Twin concept and the multi-layer simulation environment used in this work.

Chapter 3 describes the simulation setup in detail, including the use case, network configuration, simulator environment, and the KPI extraction workflow.

Chapter 4 presents the implementation and evaluation of the baseline scenario and three remediation methods: an additional small cell, transmission-power adjustment, and bandwidth scaling.

Chapter 5 provides the evaluation, comparison, and discussion of the obtained results with respect to their efficiency and practical feasibility.

Finally, **Chapter 6** summarizes the main findings, draws conclusions, and outlines potential directions for future research.

2. Fundamentals of Networks

In this chapter, some fundamental knowledge will be introduced to support a better understanding of the content of this thesis. Additionally, this chapter describes the multi-layer simulator, which is employed for the subsequent simulation study.

2.1. Mobile Network Architecture

For this thesis, we focus on private 5G campus networks deployed in controlled environments such as university campuses, hospitals, industrial sites, or event venues. A 5G network consists of two areas: the Radio Access Network (RAN) and the Core Network [4]. To use the network, a User Equipment (UE) is required; the UE is the end device that is connected to the (5G) network [2]. The base station of a 5G network is also referred to as Next Generation Node B (gNodeB, gNB) [4]. The 5G RAN includes one or more gNodeBs providing the radio interface towards the UEs. The core network is responsible for functions that are not responsible for radio access, but for providing a network connection. For the fifth mobile communications standard, this core network is also referred to as 5GC [4]. In a 5G network, one or more UEs can therefore connect to the core network via the RAN [2].

Figure 2.1 illustrates a simplified 5G Network architecture with exemplary UEs and a 5G Core [2, 4]. Both gNodeBs operate as monolithic units implementing the complete base station functionality. They communicate via the Xn interface and are connected to the 5G Core via the NG interface for control (N2) and user data (N3) [4]. The radio access network comprises the functions that are responsible for a wireless connection to the end devices and provide access to the core network.

A cell in mobile network terminology refers to the geographical area served by a single base station. Macro cells provide wide-area coverage with typical transmission power of 40–46 dBm [13], while small cells encompass lower-power base stations (23–30 dBm for picocells, 10–23 dBm for femtocells) that serve localized areas [9][34]. Small cell deployment is essential for network densification in high-demand regions [11][19].

To configure these simulations appropriately, the spectrum and signaling characteristics used in this thesis must be described. 5G operates across multiple frequency bands defined by the 3GPP standards. This thesis uses the n78 band (3.7–3.8 GHz),

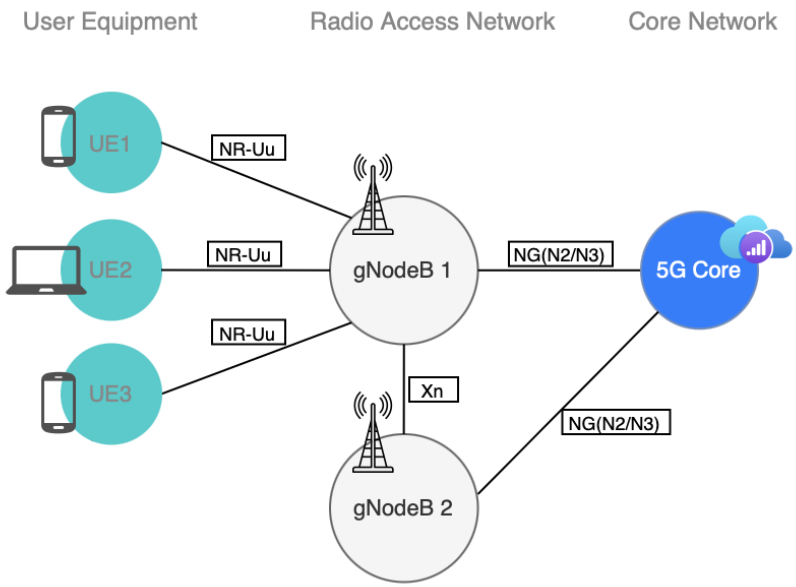

Figure 2.1.: Exemplary 5G Radio Access Network (RAN) Architecture

a mid-band spectrum allocation commonly used for private 5G deployments in industrial and campus environments [15, Chapters 2-3]. Mid-band frequencies offer a balance between coverage and capacity, making them well-suited for localized high-load scenarios [15][13].

In 5G NR, a Physical Resource Block (PRB) is the smallest allocatable unit in the frequency domain. As shown in Figure 2.2, each PRB occupies 12 consecutive subcarriers for one slot in the time domain [6][43]. For the 60 kHz subcarrier spacing ($\mu = 2$) used in this thesis, each PRB spans $12 \times 60\,\text{kHz} = 720\,\text{kHz}$ in frequency [14] and comprises 14 Orthogonal Frequency-Division Multiplexing (OFDM) symbols in time [6]. One OFDM symbol at $\mu = 2$ has a duration of approximately $18\,\mu s$, making one slot equal to $14 \times 18\,\mu s \approx 250\,\mu s$ [14]. Each 10 ms frame is divided into 10 subframes of 1 ms duration, and at $\mu = 2$, each subframe contains exactly 4 slots [14]. PRB utilization, expressed as a percentage of total available PRBs occupied by active downlink traffic, serves as a direct indicator of cell capacity saturation and represents the fundamental metric for identifying resource-limited high-load conditions [1].

Understanding capacity constraints requires defining the key performance metrics used throughout this thesis. *Data rate* refers to the speed at which data are transmitted over the wireless channel, typically measured in megabits per second (Mbps) or gigabits per second (Gbps) [15]. Peak data rates represent the theoretical maximum achievable

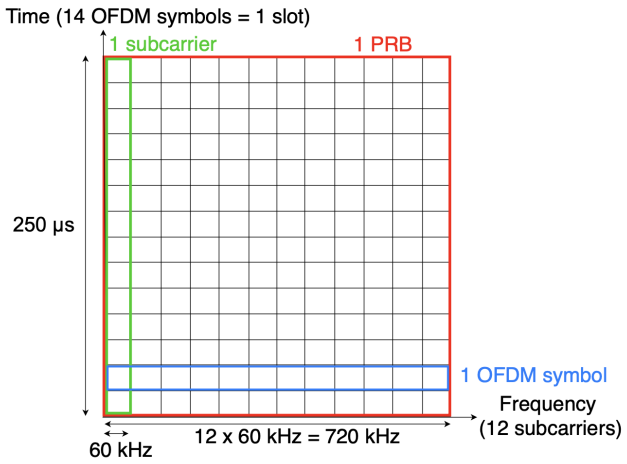

Figure 2.2.: Physical Resource Block (PRB) Structure in 5G NR at Numerology $\mu = 2$

under ideal conditions: 5G NR can theoretically reach up to 20 Gbps in downlink under optimal configurations with high bandwidth and Multiple Input Multiple Output (MIMO) [15] . However, user-experienced data rates are typically much lower, ranging from 50–300 Mbps in typical 5G deployments, depending on network load, signal quality, and available spectrum [18][7].

Throughput measures the actual volume of data successfully transmitted per unit time, accounting for protocol overhead, retransmissions, and radio conditions [15]. Unlike peak data rate, throughput reflects real-world performance, including the effects of scheduling, interference, and congestion [15]. In congested cells, even with good signal quality, per-user throughput may degrade significantly as radio resources must be shared among many active users [15][20, Chapter 9].

The *Signal-to-Interference-plus-Noise Ratio (SINR)* measures the received signal strength relative to interference and noise, typically expressed in decibels (dB) [20, Chapter 6]. Higher SINR values enable the selection of higher-order modulation and coding schemes, directly affecting achievable data rates and spectral efficiency [20, Chapter 6, Section 6.1, p. 159–160][6, Section 5.1]. In this thesis, SINR serves as a key quality indicator to distinguish between resource-limited and interference-limited congestion scenarios [1]. A resource-limited scenario, which is the focus of this work, is characterized by high *cell load* despite consistently high SINR values which indicates that capacity constraints, not signal quality degradation, are the primary limitation [1]. In the context of this thesis, Physical Resource Block (PRB) utilization is directly mapped to the metric termed DL Cell Load, where both quantities represent the scheduler's radio-layer occupancy and are treated equivalently [1].

Quality of Service (QoS) involves a set of performance metrics that characterize the user experience in mobile networks, including throughput (data rate), latency (packet delay), *packet loss rate*, and reliability Block Error Rate (BLER) [15]. *BLER* measures the fraction of erroneous blocks after error correction [27]. Maintaining acceptable QoS during high-load scenarios is the primary objective of the remediation strategies evaluated in this thesis. QoS requirements vary by application type: real-time services such as voice or video calls demand low latency (typically $<$100 ms), while elastic services such as file downloads prioritize throughput [15][32]. During congestion, the challenge for network operators is to maintain QoS commitments for as many users as possible while managing finite radio resources, a challenge that motivates the remediation methods studied here [15] .

Traffic models describe the statistical patterns and temporal characteristics of user data demand. Common traffic models include Constant Bit Rate (CBR) for voice or streaming applications, Poisson arrivals for bursty web traffic, and on-off models where users alternate between active transmission periods and idle intervals [32]. Accurate traffic modeling is essential for capacity planning and for reproducing realistic load conditions in network simulations [32]. In this thesis, UDP-based on-off traffic models are employed to generate controlled, reproducible load patterns that emulate real-world user behavior during high-demand scenarios.

2.2. Network Congestion and High-Load Scenarios

To establish foundational terminology, the following definitions distinguish between high-load, overload, and congestion.

Network performance can deviate from design expectations when user density or data demand rises sharply in parts of the system. Such events lead to high load in specific cells, where resource usage approaches network capacity [19][15].

Network high-load: A state in which a communication network or individual cell operates at a high level of resource utilization, close to its design or planning capacity, while service quality and user experience remain within acceptable limits is defined as network high-load [1][19]. High load conditions reflect normal but critical operational states in which system efficiency and stability must be maintained [15][19].

In this thesis, a high-load scenario refers to the operational condition of a 5G radio access cell where resource utilization (e.g., PRB or power allocation) approaches the cell's capacity threshold, yet the system still delivers acceptable QoS to users. This definition corresponds to [19] where Freeman writes "under heavy-traffic conditions

the network operates near its engineered limit but still provides an acceptable grade of service".

Network overload: Network overload is a condition where the offered traffic or workload exceeds the available network capacity, preventing proper allocation of resources to all users [24]. Overload leads to performance degradation manifested as failed transmissions, increased blocking probability, or impaired throughput and latency [15][8].

The network overload as defined in this thesis, represents a critical escalation beyond high-load conditions, where the RAN scheduler or resource manager can no longer satisfy QoS requirements for all users [24, 15]. The goal of the digital twin framework is to predict and prevent this transition from high load to overload. ITU-T E.417 [24] chapter 6.2.4 formally states that "a link may be said to be in overload whenever demand exceeds currently available capacity over a prolonged period."

Network congestion: Network congestion is the observable symptom of overload, where data queues and delays build up due to limited transmission or processing capacity [19]. It is typically recognized through increased latency, packet loss, or buffer overflow within network nodes and links [19][8]. The definition of congestion in [19] supports this explanation.

To determine when a high load in a network can be specified, network operators define thresholds [1]. One of these thresholds concerns the amount of UEs per channel bandwidth. Another threshold focuses on the number of active PRBs [1]. Chapter 3 goes into more details about thresholds. This thesis focuses on resource-limited overload, where PRB scarcity, not signal quality, is the dominant constraint [1].

2.3. Remediation Methods in Cellular Networks

Network overload has many different aspects[19]: what is the cause of the overload? Is it something that can be predicted or even happens on a regular basis? How can such situations be prevented? Does it require a long-term or a short-term solution? The question relevant for this work is how such a situation can be addressed; several established techniques are available to resolve instances of network high-load. Classical approaches include increasing the transmission power of a base station or deploying additional small cells to locally offload traffic [11][15]. Network-side measures such as enabling or disabling load balancing mechanisms can also be applied [15][32]. Setting up a sufficient bandwidth configuration is another important strategy [15][32]. Beyond these strategies, there are more advanced concepts such as Beamforming and Massive MIMO, Load Balancing, Carrier Aggregation or even including AI agents [15][22][47].

Remediation methods have the purpose of offering powerful ways to enhance capacity and user experience [15][11][19].

This master's thesis concentrates on three fundamental and widely applied strategies that can be directly implemented and evaluated within the chosen simulation framework:

- **Additional Small cell**
- **Transmission Power Adjustment**
- **Bandwidth Scaling**

These methods were selected because of their practical relevance and were tested under high-load conditions using a Network Digital Twin to assess their effectiveness.

The methods will be explained further in the following sections.

2.3.1. Additional Small Cell

The deployment of additional small cells represents one of the most effective strategies for alleviating network congestion in localized high-load scenarios [11]. By introducing a secondary base station within the coverage area of an overloaded macro cell, traffic can be offloaded, and radio resources can be distributed across multiple transmission points [11][15].

In the context of this thesis, the terms macro cell and small cell refer to the functional roles of the base stations rather than differences in hardware configuration [15][11]. Both are modeled as standard gNodeBs with identical transmission power (30 dBm) and hardware specifications. The macro cell provides the primary coverage for the entire simulation area, while the small cell is introduced as an additional node to share the load. This approach reflects realistic deployment scenarios where operators activate standby base stations or deploy mobile cells during temporary events such as concerts, sports matches, or festivals [11][17].

The effectiveness of small cell deployment depends critically on the placement strategy [11][34]. Three positioning methods are evaluated in this work:

1. **K1 - Geometric split:** Places the small cell to divide the coverage area roughly equally based on geographic distance, following a Voronoi-like partitioning principle. This method aims for balanced load distribution without requiring knowledge of traffic hotspots [28][45].

2. **K2 - Hotspot overlay:** Positions the small cell directly at identified traffic concentration points, such as areas with high user density or data demand. This strategy maximizes local capacity where it is most needed, reflecting common operator practice at stadiums, transit hubs, or event venues [31][26].
3. **K3 - Intermediate split:** A compromise approach that places the small cell between the macro site and the hotspot, balancing macro relief with targeted capacity enhancement.

These three positioning strategies were derived from principles identified in network planning literature and operator practices [11], rather than being standardized, formally named methods. The naming conventions used here were chosen for clarity and do not correspond to formal 3GPP or ITU terminology. The two fixed strategies (K1 and K2) employ either geographic partitioning or hotspot targeting, whereas the intermediate split (K3) was introduced in this thesis as a practical compromise [11].

The deployment of a small cell also includes **practical limitations**. There are deployment costs as hardware acquisition, site installation, and backhaul connectivity require capital investment [9]. Site acquisition involves physical space, permissions and regulatory approvals that may delay deployment [41]. Energy consumption is another concern, as operating an additional base station increases operational energy costs [9]. Finally, interference management requires coordination between macro and small cell to avoid inter-cell interference and optimize handovers [9].

Small cell deployment is widely adopted in practice due to its flexibility and scalability [11]. Temporary deployments using mobile base stations Cells on Wheels (COWs) [40] are routinely used during large public events to prevent congestion and maintain quality of service [11][17].

2.3.2. Transmission Power Adjustment

Adjusting the transmission power of the macro cell gNodeB is a straightforward remediation approach based on the principle that higher transmit power increases the received signal strength at User Equipments, potentially improving SINR and channel quality [15][20].

According to Goldsmith [20] the received signal Power P_{RX} at a UE can be expressed as:

$$P_{RX} = P_{TX} + G_{TX} + G_{RX} - PL$$

where P_{TX} is the base station transmission power in dBm, G_{RX} and G_{RX} are the transmitter and receiver antenna gains in dBi, and PL is the path loss in dB. Here,

dBm is an absolute power unit referenced to 1 milliwatt, dBi quantifies antenna gain relative to an isotropic radiator, and dB represents relative power ratios such as path loss [19, p. 609–630]. Increasing P_{TX} directly raises P_{RX}, which in turn improves the SINR [20]. In the context of this thesis, P_{signal} is the received power from the serving base station, $P_{interference}$ comprises signals from neighboring cells, and P_{noise} is thermal noise at the UE receiver [20, Chapter 6]. This relationship is expressed as:

$$\text{SINR} = \frac{P_{signal}}{P_{interference} + P_{noise}} \tag{2.1}$$

Higher SINR values enable the selection of higher-order Modulation and Coding Schemes (MCS), increasing the number of bits transmitted per symbol and thus improving per-user throughput [6].

The simulations conducted as part of this thesis increase the transmission power incrementally in steps of 3 dB (from 30 dBm to 33 dBm, 36 dBm, and 39 dBm). The choice of 3 dB steps is motivated by practical considerations—each 3 dB increase corresponds to a doubling of linear transmit power [19, p. 609–630], representing a meaningful incremental change while remaining within regulatory and hardware constraints [20].

Transmission power adjustment has **practical limitations**. First, regulatory constraints must be considered: in Germany, maximum transmission power for 5G gNodeBs in the n78 band (3.4–3.8 GHz) is limited to 43 dBm EIRP as per Bundesnetzagentur regulations [13, p. 12–14]. Equivalent Isotropic Radiated Power (EIRP) is the product of base station transmission power and antenna gain, representing the effective radiated power in the direction of maximum antenna gain. The typical operating range for macro base stations is 40–43 dBm, while small cells operate at 20–30 dBm [13, p. 13–14]. Beyond regulatory constraints, hardware limits present a fundamental barrier: power amplifiers have maximum rated output specifications, and exceeding these causes distortion and hardware degradation. Additionally, increased interference becomes problematic in dense deployments—higher transmit power increases inter-cell interference, which may degrade performance for neighboring cells despite improving coverage in the target cell [11]. Finally, energy consumption is a practical concern, as higher transmit power raises operational costs and environmental impact [9].

2.3.3. Bandwidth Scaling

Bandwidth scaling increases the channel bandwidth allocated to a cell, directly expanding the number of available PRBs and thus the cell's capacity to serve simultaneous

users and data streams [6]. Within the constraints of available spectrum and hardware capabilities, wider bandwidth allocations provide more radio resources.

The fundamental relationship governing wireless channel capacity is given by the Shannon-Hartley theorem [20]:

$$C = B \cdot \log_2(1 + \text{SNR}) \tag{2.2}$$

where C is the channel capacity in bits per seconds, B is the channel bandwidth in Hz, and the Signal-to-Noise Ratio (SNR) is in linear scale (not dB). Equation 2.2 reveals that the capacity increases linearly with bandwidth but only logarithmically with SNR [20]. Consequently, doubling the bandwidth from 10 MHz to 20 MHz doubles the theoretical capacity, whereas doubling the transmit power (increasing SNR by 3 dB) yields a much smaller capacity gain [20]. In 5G NR, bandwidth scaling is implemented by increasing the number of allocated Resource Blocks [6]. Each Physical Resource Block occupies 12 subcarriers x 1 slot in the time-frequency grid [6]. For a 10 MHz channel with 60 kHz subcarrier spacing (numerology μ=2), approximately 51 PRBs are available. Scaling to 15 MHz increases this to 79 PRBs, and 20 MHz provides 106 PRBs [43]. More PRBs enable the scheduler to assign resources to more UEs simultaneously, reducing queueing delays and increasing aggregate throughput [6].

Practical limitations of bandwidth scaling must be acknowledged. First, spectrum cost is a fundamental constraint: radio spectrum is scarce and expensive, with operators acquiring spectrum through government auctions that often cost billions of euros for nationwide licenses [13][41]. Beyond cost, regulatory allocation enforces strict boundaries—available spectrum is finite and must be shared among multiple operators and services, limiting the bandwidth available to any single operator [13]. Technical challenges also arise: interference and adjacent channel effects mean that wider channels may overlap or interfere with adjacent operators' spectrum, requiring careful frequency planning and guard bands to prevent cross-operator degradation [13]. Finally, hardware compatibility is often overlooked—not all base stations and user equipment support all bandwidth configurations, and achieving network-wide support may require expensive hardware upgrades [41].

Despite these challenges, bandwidth scaling is highly effective for resource-limited congestion scenarios. By expanding the pool of schedulable radio resources, it directly addresses the root cause of overload when PRB utilization saturates [1].

2.4. Network Digital Twin Concept

Digital Twin has been a trending topic in recent years across various fields, including manufacturing, urban planning and healthcare. It is known to be a virtual repre-

sentation of a physical object. It is often visualized as a digital model. A Network Digital Twin (NDT) [22] is a virtual replica of a real communication network. It may include data about the geographic information of a network and thus resembles a 3D digitalized model. Additionally, it may involve ray-tracing data, mobility patterns (e.g., if UE is moving), and packet-level communication data. Core network components and RAN functionalities are typically modeled as part of the NDT. (Nevertheless, what differentiates a NDT from other digital twins is the inclusion of detailed simulated versions of an existing network, but also real-time data from that specific network.) As a result, Network Digital Twins are designed to be more detailed and more accurate than traditional network simulators. Depending on the use case, an NDT may focus on specific network domains, such as the RAN, core network, or transport layer, or model end-to-end functionality. In this thesis, the NDT focuses on RAN-layer modeling and can be used as a testbed for safe and controlled experiments with remediation strategies and network optimizations before deployment in the actual network.

Traditional network simulators such as ns-2 [25], ns-3 [44] or OMNeT++ [44, 46] are powerful tools for modeling communication protocols and network behavior under controlled conditions. However, they typically operate on pre-configured, static parameter sets and do not incorporate live data from operational networks. Simulators are primarily used for what-if analysis and protocol validation during the design phase, before deployment.

In contrast, a Network Digital Twin integrates real-time or near-real-time data from the physical network into its virtual model. This unidirectional data flow enables the NDT to

- **Mirror current network state:** Reflect live traffic loads, user distributions, and radio conditions.
- **Enable predictive analysis:** Forecast network performance under future scenarios (e.g., upcoming events, traffic growth).

While the digital twin concept encompasses bi-directional closed-loop capabilities, where validated remediation strategies from the virtual twin are automatically applied back to the physical network, this thesis focuses on the simulation and analysis phase, utilizing the NDT as a controlled testbed for evaluating different strategies before any real-world deployment. Closed-loop automation remains out of scope for this work.

Real-World Applications and Examples:

Network Digital Twins are gaining traction in the telecommunications industry, for example Deutsche Telekom's 5G Campus Networks [39][17]: In industrial deployments such as the EUROGATE container terminals in Hamburg, Bremerhaven, and

Wilhelmshaven, Deutsche Telekom integrates digital twin concepts to connect industrial port handling equipment to the cloud, enabling data-driven optimization and predictive maintenance. These projects demonstrate the practical application of digital twin methodologies in real operational environments. ETSI and ITU standards: The European Telecommunications Standards Institute (ETSI) and International Telecommunication Union (ITU) have published frameworks and best practices for implementing digital twins in network management (e.g., ETSI GS NFV-IFA 031).

2.5. Simulation Tools

This study uses the "ns-3 Playground" a multi-layer simulation environment developed by Fraunhofer HHI and Deutsche Telekom's T-Labs. The ns-3 Playground integrates three specialized tools: ns-3 [44] for packet-level simulation, Sionna [21] for realistic radio propagation, and SUMO [29] for dynamic user mobility, within a unified digital twin framework. This integrated approach allows for realistic modeling of high-load mobile scenarios, enabling detailed analysis of network behavior under varied remediation strategies.

The following subsections introduce each major component individually, outlining their purpose, main features, and specific role in the overall simulation workflow.

2.5.1. ns-3

ns-3 is an open-source, discrete-event network simulator widely used in academic research and education for modeling Internet protocols and communication systems[44]. It provides detailed implementations of the Physical Layer (PHY), Medium Access Control (MAC), radio link control Radio Link Control (RLC), and higher-layer protocols for various wireless and wired technologies, including 4G LTE and 5G NR [44].

In this thesis, ns-3 version 3.33 serves as the core simulation engine, modeling the entire 5G protocol stack from the physical radio layer up to the Packet Data Convergence Protocol (PDCP). ns-3 is primarily written in C++, but this work integrates a Python-based wrapper (Python 3.10.18, GCC 9.5.0) to streamline scenario configuration, automate simulation runs, and facilitate post-processing of results [44]. This hybrid approach combines the computational efficiency of C++ with the flexibility and ease of scripting offered by Python.

2.5.2. Sionna

Sionna is an open-source library developed by NVIDIA for research in physical-layer wireless communications [21]. Its primary strength lies in its ability to perform GPU-accelerated ray tracing for radio wave propagation, enabling highly realistic modeling of signal strength, path loss, diffraction, reflection, and scattering in complex 3D environments.

Ray tracing is essential for accurate network simulations because it accounts for real-world propagation effects such as multipath fading, shadowing by buildings, and material-dependent attenuation—factors that simplified statistical models (e.g., free-space or log-distance models) cannot capture [36][21]. Sionna ingests 3D geometric and material data (typically from Blender files) along with antenna positions and orientations, and outputs detailed channel impulse responses that ns-3 uses to compute received signal strength and signal-to-interference-plus-noise ratio (SINR) for each user [21].

In this thesis the RT module of Sionna version 0.19.2 (TensorFlow 2.15.1 backend) is used to pre-compute ray-tracing data for the simulated campus network, ensuring that radio propagation remains realistic and consistent across all simulation runs [36].

2.5.3. SUMO

SUMO (Simulation of Urban MObility) is an open-source traffic simulation tool developed by the German Aerospace Center (DLR) [29]. While originally designed for urban outdoor traffic, SUMO's trajectory generation capabilities extend to indoor environments [29]. In this thesis, Eclipse SUMO version 1.23.1 is integrated to generate realistic mobility patterns for Autonomous Mobile Robots (AMRs) within a factory hall, utilizing predefined waypoint-based navigation rather than road networks or traffic lights [33]. This approach leverages SUMO's collision avoidance and movement modeling while adapting to the indoor, coordinate-based environment [29]. SUMO is primarily written in C++ and Python, enabling flexible integration into existing simulation workflows.

2.6. Integration of Tools - The ns-3 Playground Environment

The ns-3 Playground environment integrates **ns-3**, **Sionna**, and **SUMO** into a unified multi-layer simulation framework, enabling comprehensive modeling of network

protocol behavior, realistic radio propagation, and dynamic user mobility. This section describes how these tools interact and the data flow that connects them.

Role of Each Tool in the Integration: ns-3 serves as the core simulation engine, orchestrating the entire simulation workflow. It manages the discrete-event simulation timeline, schedules radio resources, handles user data traffic, and collects performance metrics such as throughput, SINR, and cell load [44].

Sionna provides realistic radio channel modeling through GPU-accelerated ray tracing. Given the 3D geometry of the simulation environment, antenna positions, and material properties, Sionna computes detailed channel impulse responses for each transmitter-receiver pair [21]. These responses capture real-world propagation effects including path loss, multipath fading, reflection, diffraction, and shadowing [21]. ns-3 uses Sionna's output to calculate received signal strength and SINR for each User Equipment (UE), which directly impacts link quality, modulation and coding scheme (MCS) selection, and achievable data rates.

In this thesis, SUMO generates mobility trajectories for User Equipments, which are Autonomous Mobile Robots (AMRs) within a factory hall scenario [29][33]. SUMO defines each UE's position, speed, and direction over time, accounting for road networks, obstacles, and movement constraints [29]. ns-3 reads these trajectories and updates UE positions during the simulation, ensuring that channel conditions, handover events, and traffic distribution across cells reflect realistic user mobility.

Input Data Sources: The simulation environment requires accurate 3D geometric and material data to enable realistic ray tracing. Blender, an open-source 3D modeling and rendering software, is used to create this environment. A Blender file serves as the foundational 3D model for the simulation environment and contains the following essential data: This file contains geometrical outlines, material properties and antenna coordinates. **Geometrical outlines** include buildings, walls, roads, terrain elevation, and other physical obstacles that affect radio wave propagation. **Material properties** are surface characteristics (e.g., concrete, glass, metal) that determine how radio signals reflect, absorb, or refract when interacting with structures. Different materials have distinct permittivity and conductivity values, which influence signal attenuation and scattering behavior. The **antenna coordinates** include the position (x, y, z coordinates), the height above ground, and the **orientation** (azimuth and tilt angles) of each base station antenna. Antenna orientation determines the directionality of the radiation pattern and affects coverage and interference characteristics.

This 3D model is then imported into Sionna for ray tracing computation. The accuracy of the geometric and material data directly impacts the realism of the resulting channel models and, ultimately, the validity of the simulation results.

Data Flow and Integration Workflow: Figure 2.3 illustrates the integration

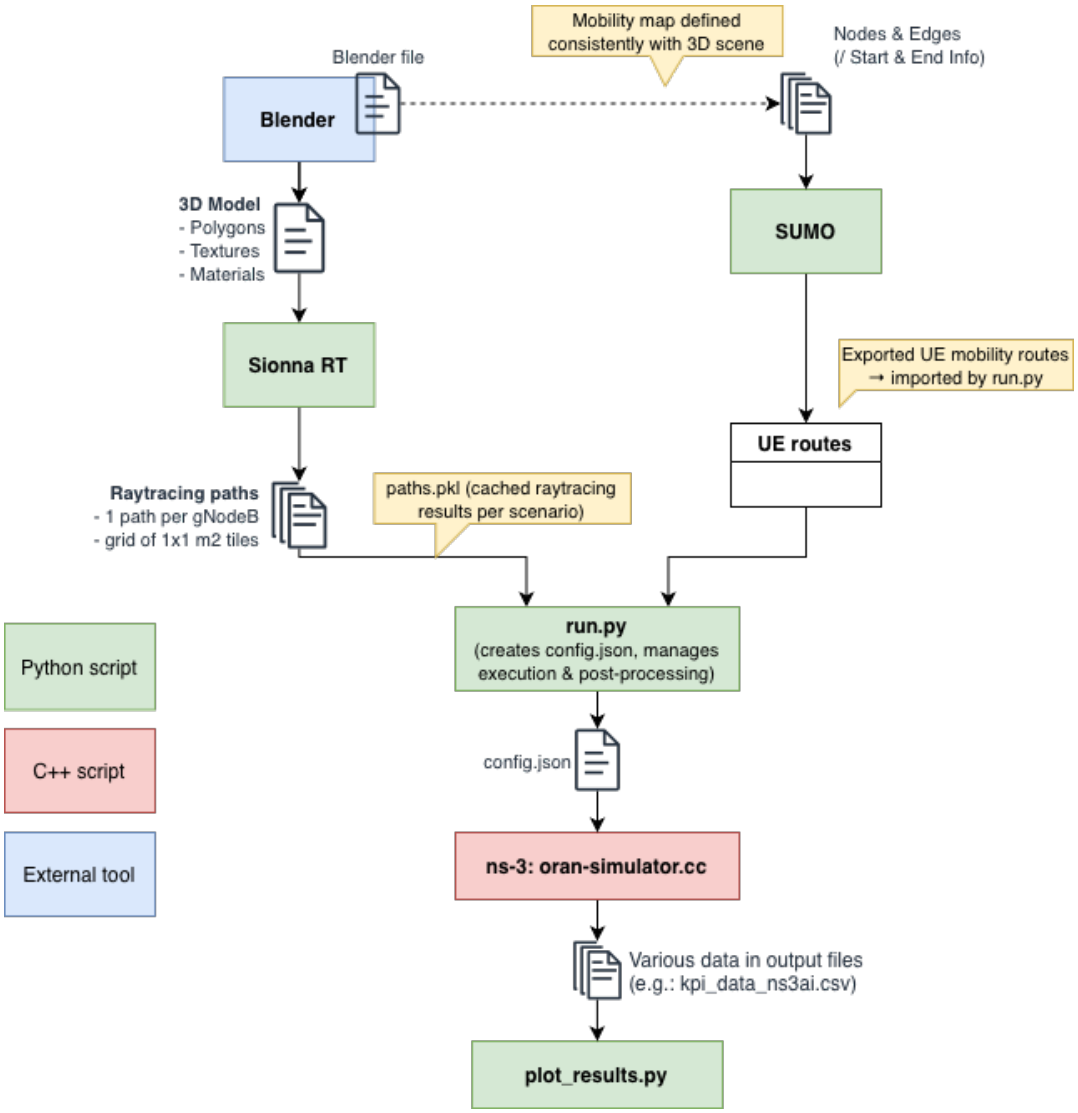

Figure 2.3.: Integration of Tools Overview

workflow of the ns-3 Playground. The simulation proceeds as follows:

1. **Mobility generation (SUMO):** SUMO generates time-stamped mobility trajectories for all UEs based on predefined routes, speeds, and movement behaviors [29]. These trajectories are exported as data files that specify each UE's position at discrete time steps.
2. **Ray tracing and channel modeling (Sionna):** Sionna reads the Blender file containing the 3D environment and antenna configurations. For each UE position along the SUMO trajectories, Sionna performs ray tracing to compute the channel impulse response between the base station antennas and the UE [21]. This includes calculating the signal strength contribution of each propagation path (line-of-sight, reflected, diffracted, scattered). The resulting **channel data** (e.g., path loss, delay spread, angular spread) is stored for use by ns-3.
3. **Network simulation (ns-3):** ns-3 loads both the SUMO mobility trajectories and the Sionna channel data. During the discrete-event simulation, UEs move according to SUMO trajectories. ns-3 queries the pre-computed Sionna channel data to determine signal quality (SINR) for each UE at each time step. The 5G NR scheduler allocates Physical Resource Blocks (PRBs) to UEs based on channel quality, traffic demand, and scheduling policy. Application-layer traffic (e.g., UDP streams) is generated, and ns-3 tracks throughput, latency, packet loss, and other performance metrics [21]. ns-3 produces log files containing time-series data for all performance indicators (e.g., per-UE throughput, cell load, SINR, BLER). These outputs are post-processed using Python scripts to generate plots, tables, and statistical summaries for analysis.

This integration ensures that radio propagation, user mobility, and network protocol layers are consistently modeled together in one unified simulation workflow, enabling realistic evaluation of network performance and remediation strategies.

Advantages of the Integrated Approach The ns-3 Playground's multi-tool integration offers several distinct advantages over using individual simulators in isolation. First, **realism** is substantially improved: combining ray-traced propagation with realistic mobility traces and full-stack protocol modeling captures the complexity of real-world networks far better than simplified analytical models or single-layer simulators, reducing the simulation-to-reality gap. Beyond realism, **flexibility** is a key benefit: each tool can be updated or replaced independently. For example, alternative mobility models or different ray-tracing engines could be integrated without redesigning the entire framework, allowing the simulator to evolve as technologies advance. **Scalability** is another significant advantage—GPU acceleration in Sionna enables efficient ray tracing even for large, complex 3D environments [21], while ns-3's modular architecture supports simulations with hundreds of UEs and multiple cells [44].

Finally, **reproducibility** is guaranteed: pre-computed Sionna channel data ensures that all simulation runs use identical propagation conditions, eliminating randomness from the channel model and improving comparability across different remediation methods [21].

The ns-3 Playground thus forms a powerful Network Digital Twin tool capable of evaluating network performance and testing remediation methods under realistic conditions. While it does not yet incorporate real-time data feeds from operational networks (as discussed in Section 2.4), it represents a significant step toward building a fully operational digital twin for network planning and optimization.

3. Simulation setup

Having introduced the fundamental concepts of mobile networks, congestion scenarios, and remediation methods in Chapter 2, this chapter describes the specific simulation setup used to evaluate these methods. The simulations were conducted using the ns-3 Playground environment in a controlled 5G campus network scenario designed to reproduce realistic resource-limited overload conditions.

3.1. Use Case Description

This thesis investigates congestion remediation in a 5G campus-network-like scenario using the ns-3 Playground environment, which combines ns-3 [44], SUMO [29], and Sionna [21]. The simulated area covers 80 meters times 70 meters, representing a section of an indoor factory hall campus site. A single macro gNodeB is placed at $[x, y, z] = [22, 63, 4.5]$ m, serving both mobile and static UEs. The simulation aims to reproduce realistic radio and mobility conditions in a controlled and reproducible digital-twin setup [22][47].

The chosen layout reflects a typical private 5G campus network as deployed in industrial or university environments, where localized traffic peaks can occur near production halls, laboratories, or event locations [10]. Such networks are smaller than public macro deployments and typically operate in frequency bands such as n78, which can be allocated for private campus network use in licensed spectrum[15](Chapters 2-3). Campus networks are particularly important for this study because they represent controlled environments where remediation strategies can be tested before deployment in larger public networks, and they exhibit similar resource-contention patterns as public networks during peak demand periods [15].

Simulation Purpose: The use case focuses on resource-limited overload, where radio resources (PRBs) become saturated while signal quality (SINR) remains good. Such high-load conditions are typical for temporary gatherings, campus hotspots, or production peaks in industrial private networks[32]. The resulting baseline scenario provides a reproducible overload condition used later to test different remediation methods.

KPI-Based High-Load Definition: Following operator practice and 3GPP TS 28.552 [1], a cell is considered under high load when:

1. PRB utilization $\geq 80\%$ (or equivalently, DL Cell Load ≥ 0.8), and
2. At least 30 UEs per 5 MHz bandwidth are active (i.e., ≥ 60 UEs for 10 MHz).

The thresholds are based on typical network operator KPIs [43][16](LinkedIn post on network optimization thresholds). Because the ns-3 Playground provides cell-load values but not explicit PRB statistics, this work treats the metrics as equivalent [1]:

$$\text{DL Cell Load} \approx \frac{\text{PRB Utilization}}{100}$$

This methodological substitution is valid since both metrics represent the scheduler's radio-layer occupancy.

For the 5G n78 band (3.7 GHz) with a 10 MHz channel bandwidth, the threshold of 30 UEs per 5 MHz corresponds to approximately 60 active UEs [13](for n78 band frequency). Using a wider bandwidth would require proportionally more simulated UEs and significantly longer process runtime; 10 MHz therefore represents a realistic yet computationally feasible configuration.

To exclude transient effects at simulation startup and focus on steady-state behavior, only the time window [3–10] s after simulation start is used for all KPI evaluations.

Bandwidth and Load Thresholds: According to network operators, the threshold of 30 UEs per 5 MHz corresponds to a moderate overload. As this thesis employs the 5G n78 band (3.7 GHz), which supports channel bandwidths from 10 MHz to 100 MHz, a 10 MHz bandwidth was chosen. This implies a practical load threshold of ≈ 60 UEs per 10 MHz. Using a higher bandwidth would require proportionally more simulated UEs and significantly longer runtime; 10 MHz therefore represents a realistic yet computationally feasible configuration.

Evaluation Goal: The main goal is to reproduce a measurable high-load situation ($DLCellLoad \geq 0.8$) as a baseline reference to compare remediation methods in later chapters. All following methods (additional small cell, transmission power adjustment, and bandwidth scaling) are evaluated against this baseline.

3.2. Network Configuration

The simulated network represents a small 5G campus deployment operating in the n78 band (3.7 GHz, FR1) [13]. All configuration parameters were chosen to reproduce realistic channel and traffic conditions while keeping the computational effort feasible [6](for PRB and numerology 2),[15](for scheduler and general parameters),[44](for ns-3 default settings).

Radio and Propagation Settings

- Carrier frequency: 3.7 GHz (5G NR band n78)
- Bandwidth: 10 MHz
- Numerology: $\mu = 2$ (60 kHz subcarrier spacing)
- Transmission power (gNB): 30 dBm
- Scheduler: Proportional Fair (default ns-3 configuration)
- HARQ: enabled, RLC mode: Acknowledged Mode (AM)
- CQI reporting: 1 ms period
- Simulation time: 10 s

These configuration parameters include key reliability and feedback mechanisms: Hybrid Automatic Repeat reQuest (HARQ) provides automatic retransmission of corrupted blocks, RLC Acknowledged Mode (AM) ensures reliable in-order delivery, and Channel Quality Indicator (CQI) feedback at 1 ms intervals enables the scheduler to make informed resource allocation decisions based on current channel conditions.

This model is suited for dense urban outdoor environments where buildings form reflective "canyons" [5]. The urban propagation model—characterized by reflective surfaces and signal blockage from dense obstacles—is applicable to factory hall scenarios, where machine frames, storage racks, and metal structures create analogous multipath propagation conditions [38]. Ray tracing provides accurate channel modeling for both environments [36][20].

The confined indoor geometry of the factory hall thus exhibits comparable propagation characteristics to urban street canyons, making this model appropriate for the studied environment [5]. Antennas are modeled as isotropic (no tilt or sectorization) to focus the analysis on load and capacity effects rather than antenna patterns.

Topology and UE Placement: One macro gNodeB provides coverage from the fixed position $[x, y, z] = [22, 63, 4.5]$ m, where z denotes the antenna height above ground. Three static UEs are placed at

$$\ell_1 = [52, 63, 1.5], \quad \ell_2 = [22, 33, 1.5], \quad \ell_3 = [52, 33, 1.5].$$

The remaining 57 UEs are mobile, following SUMO-generated trajectories within an 80 m × 70 m area [29]. The SUMO parameter `CreateRandomRoutes = True` generates random yet logical mobility paths (no collisions, street-based) [29]. All UEs, mobile and static, use identical traffic models, ensuring that differences in KPIs originate from spatial and radio effects only.

Figure 3.1 shows the simulated UE mobility paths in the baseline scenario. All subsequent remediation methods use the same trajectories to ensure fair comparison.

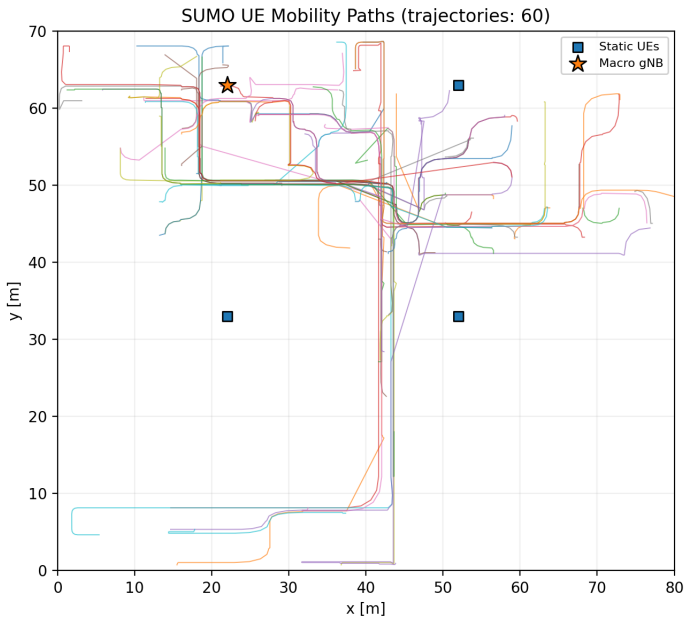

Figure 3.1.: SUMO UE Mobility Paths

A single macro gNodeB was chosen for this study. Including multiple gNodeBs would have required at least twice as many UEs to maintain comparable cell loads, which would not only increase the conceptual complexity but also substantially extend simulation times. Each baseline run with 60 UEs already required approximately three hours per replicate, and the iterative nature of this project—with several trial-and-error adjustments during development—made a multi-gNodeB setup computationally impractical within the available timeframe. Therefore, the focus was placed on a

single-cell scenario that still captures the essential congestion mechanisms and allows clear comparison between the remediation methods.

Parameter Precision Note: Simulation input parameters such as traffic rates are configured with precision up to two decimal places (reflecting practical implementation constraints). These configuration values differ from the KPI results, which may have higher precision due to statistical processing. For example, per-UE data rates are specified as 2.43 Mbits or 0.35 Mbits in the traffic model, while measured KPIs are reported with three decimal places (e.g., 0.836 for cell load) to reflect the statistical aggregation across replicates.

Cell Terminology: Throughout this thesis, the simulation employs two gNodeBs when testing Method 1 (small cell deployment scenarios):

- **Cell 0:** The macro cell (primary, always present in all simulations)
- **Cell 1:** The small cell (present only in Method 1 scenarios for positioning comparisons)

In Methods 2 and 3, only Cell 0 is active, and all traffic remains on the macro cell.

SUMO Routing Constraints: Due to SUMO routing constraints in the 80m × 70m area, only 57 mobile UE trajectories could be generated[29]. Therefore, 3 UEs were configured as static at fixed positions. All UEs use identical traffic models, ensuring that total offered load remains consistent. This mixed static-mobile configuration does not significantly affect the conclusions, as static and mobile UEs experience different mobility contexts but identical traffic demand. Details on simulation calibration and the SUMO mobile UE enumeration discrepancy, including correction methodology, are provided in the Appendix.

Traffic Model and Load Dimensioning: All UEs (static and mobile) employ identical UDP On/Off traffic with constant data rates and a packet size of 1400 bytes [32](for traffic models). The per-UE data rates were designed such that the overall downlink cell load reaches approximately 0.8 (80%), representing the high-load baseline condition defined in Section 3.1. To verify and calibrate these rates, the Shannon–Hartley capacity [20] was used.

The following equations are used throughout Chapters 3 and 4 for traffic model dimensioning and result verification:

$$\gamma_i = 10^{\mathrm{SINR}_{\mathrm{dB},i}/10} \tag{3.1}$$

$$\widehat{C} = B \cdot \frac{1}{N} \sum_{i=1}^{N} \log_2(1+\gamma_i) \tag{3.2}$$

$$C_{\mathrm{target}} = \alpha\, \widehat{C} \quad \text{(e.g., } \alpha = 0.8 \text{ for 80\% of Shannon)} \tag{3.3}$$

where γ_i is the linear SINR per packet or user (Eq. 3.1), $\widehat{C}$ (Eq. 3.2) is the empirical Shannon channel capacity, and C_{target} (Eq. 3.3) defines the desired throughput threshold as a fraction α of the theoretical maximum [20].

Attempted Calibration Using Real-Network Data: To provide context on the design choices made, the following subsection documents an initial attempt to derive traffic profiles from real-network data, and explains why an alternative approach was ultimately used.

To maximize realism, an initial attempt was made to derive a "standard user" traffic profile from real mobile network data (Aachen region). One sector's downlink time series (24 h, 1 h granularity) served as the basis. The goal was to map these values to the smaller simulated cell by accounting for radio coverage differences.

To achieve this goal the receiver sensitivity and the maximum path loss had to be calculated. Given minimum SNR $SNR_{\min} = 5\,\mathrm{dB}$, noise figure $NF = 5\,\mathrm{dB}$, and bandwidth $B = 10\,\mathrm{MHz}$, the noise power [20](for noise floor reference) is

$$P_n = N_0 + 10\log_{10}(B) + \mathrm{NF}, \quad N_0 \approx -174\,\mathrm{dBm/Hz}.$$

Numerically, $P_n = -174 + 10\log_{10}(10^7) + 5 = -99\,\mathrm{dBm}$. The minimum required received power becomes

$$P_{\mathrm{Rx,min}} = P_n + \mathrm{SNR}_{\min} = -99\,\mathrm{dBm} + 5\,\mathrm{dB} = -94\,\mathrm{dBm}.$$

With $P_{\mathrm{TX}} = 30\,\mathrm{dBm}$ and antenna gain $G_{\mathrm{TX}} = 10.9\,\mathrm{dBi}$, the maximum tolerable path loss is

$$PL_{\max} = P_{\mathrm{TX}} + G_{\mathrm{TX}} - P_{\mathrm{Rx,min}} = 30 + 10.9 - (-94) = 134.9\,\mathrm{dB}.$$

Using the free-space path-loss model [20] (carrier $f = 3700\,\mathrm{MHz}$),

$$PL_{\mathrm{FS}} = 20\log_{10}(d) + 20\log_{10}(f_{\mathrm{MHz}}) + 32.45,$$

the corresponding maximum distance is

$$d = 10^{\left(PL_{\max} - 20\log_{10}(f_{\mathrm{MHz}}) - 32.45\right)/20} = 10^{(134.9 - 20\log_{10}(3700) - 32.45)/20} \approx 35.8\,\mathrm{m}.$$

The next step is to scale the real profile to the simulated cell.

Let d_{sim} and d_{Aachen} denote the effective cell radii in simulation and reality, respectively. Assuming approximately uniform user distribution, the per-UE demand can be area-scaled as

$$\text{rate}_{\text{UE,sim}} \;=\; \text{rate}_{\text{UE,Aachen}} \cdot \left(\frac{d_{\text{sim}}}{d_{\text{Aachen}}} \right)^2 .$$

Equivalently, one can scale the sector's aggregate hourly demand by the same factor and divide by the (average) number of active UEs in that hour.

After applying the coverage-based scaling, the *per-UE* rates remained too low to create a realistic high-load condition in the small simulated cell (it would require an impractically large number of UEs to reach $\rho_{\text{DL}} \geq 0.8$). Therefore, this approach was discarded for the baseline. Instead, realistic fixed per-UE rates were chosen based on the Shannon lower bound and practical experience (DL ≈ 2.43 Mbps, UL ≈ 0.35 Mbps; Section 3.2), which produce a reproducible resource-limited overload suitable for evaluating remediation methods.

3.3. Simulator Environment

The simulations were executed in the ns-3 Playground framework described above. It integrates ns-3 (packet-level core), SUMO (mobility), and Sionna (channel modeling) in a unified Python-orchestrated workflow.

The simulations were executed on a Linux server (Ubuntu 22.04, kernel 6.8.0-83, x86_64 architecture) accessed remotely via VPN. The ns-3 Playground framework was pre-installed on the server with all dependencies configured. Simulation runs were launched via SSH terminal, and output files were transferred locally for post-processing and visualization using Python scripts (pandas, numpy, matplotlib).

Table 3.13.1 summarizes the software components and versions used in the ns-3 Playground environment.

The simulator used pre-processed Sionna ray-tracing files [21] and SUMO mobility files [29] to ensure consistent geometry and motion across replicates and enable faster iteration during the study. These inputs were loaded before each run, while the remaining parameters were defined through Python configuration classes (oransim.cfg.General, oransim.cfg.Cell, ...).

Determinism and Controlled Variability: To ensure statistical robustness while maintaining scenario consistency, each configuration was simulated with three replicates [30](for statistical methodology with small sample sizes). Per-UE traffic rates

Table 3.1.: Summary of the most important Components, including their Versions and Purposes

Component	Version/ Details	Purpose
ns-3	3.33	Core simulator (PHY-MAC-RLC stack)
Sionna	0.19.2 (TensorFlow 2.15.1 backend)	Ray-tracing/channel paths
SUMO	1.23.1 (Eclipse SUMO)	UE mobility traces
Python	3.10.18 (numpy 1.26.4, pandas 2.2.3, matplotlib 3.10.3)	Automation & post-processing
GCC	9.5.0 (Ubuntu)	C++ compilation
OS & Hardware	Ubuntu 22.04 LTS, AMD Ryzen 7, 32 GB RAM	Execution environment
IDE	Visual Studio Code 1.83	Code editing/debugging

were perturbed by ±3% (replicate 0: 0.97x, replicate 1: 1.00x, replicate 2: 1.03x nominal rate), and the global random seed was offset accordingly. This approach introduces realistic demand fluctuation without altering the fundamental scenario characteristics, enabling meaningful confidence intervals in the results.

3.4. KPI Extraction

This section describes how Key Performance Indicators (KPIs) are obtained from the simulator outputs and how they are aggregated for evaluation.

Data Sources and Analysis Window: Each simulation run produces a structured output folder with several text or CSV traces. For all KPIs, only the steady-state window [3 s, 10 s] of the 10 s simulation is evaluated to exclude start-up transients. Unless noted otherwise, results are reported as per-scenario averages across three replicates (rep 0/1/2 with ±3% traffic perturbation).

KPI Definitions:

- **DL Cell Load** (ρ_{DL}): Fraction of radio resources effectively occupied by downlink traffic (mean, p95). Used as proxy for PRB utilization [1].
- **DL Sum Throughput** [Mbps]: Sum of all UE downlink throughputs [1].
- **Spectral Efficiency** (η) [bit/s/Hz]: η = DL Sum Throughput/B with B the carrier bandwidth [7].

- **SINR** [dB]: Signal-to-Interference-plus-Noise Ratio per received packet; reported as mean and quantiles (p5, p50, p95) [20].
- **BLER**: Block error rate,

$$\mathrm{BLER} = \frac{N_{\mathrm{error}}}{N_{\mathrm{tx}}},$$

where N_{tx} is the number of transmitted transport blocks and N_{error} the count received in error [27].

Mapping of KPIs to Files and Post-Processing: If both `DlPdcpStats.txt`

Table 3.2.: Mapping of extracted KPIs to simulator output files and computation steps.

KPI	Source File(s)	Computation Method
DL Cell Load (mean, p95)	`kpi_data_ns3ai.csv`	Read per-TTI DL load $\rho_{\mathrm{DL}}(t)$; compute mean and 95th percentile over $t \in [3, 10]$ s.
DL Sum Throughput [Mbps]	`DlPdcpStats.txt` or `kpi_data_ns3ai.csv`	Sum per-UE PDCP throughput within the window and divide by window duration; convert to Mbps.
Spectral Efficiency η [bit/s/Hz]	Derived (uses throughput)	η = DL Sum Thrpt/B with $B \in \{10, 15, 20\}$ MHz.
SINR (mean, p5, p50, p95) [dB]	`RxPacketTrace.txt`	Parse per-packet SINR; select packets in $[3, 10]$ s; compute mean and quantiles.
BLER (mean)	`DlPhyTransmissionTrace.txt` or PHY HARQ trace	Count transmitted vs. error blocks in $[3, 10]$ s; compute $N_{\mathrm{error}}/N_{\mathrm{tx}}$.

and `kpi_data_ns3ai.csv` expose throughput, the PDCP file is preferred for packet-accurate results; the CSV serves for consistency checks [44]. When bandwidth differs (Method 3), η is always recomputed from the measured throughput and the actual B of that run. For SINR, quantiles are computed on the set of per-packet SINR values within the analysis window (not on per-UE means).

KPI Reporting Precision: Measured KPIs including cell load, throughput, and SINR are consistently reported with three decimal places to balance precision and readability. This precision reflects the statistical analysis across replicates and the output resolution of the simulation environment. For instance, cell load values are reported as 0.836 (not 0.84), throughput as 27.697 Mbps (not 27.7 Mbps), and SINR as 62.572 dB (not 62.6 dB). This standardization ensures consistency across all results presented in Chapter 4.

Replicate Handling and Confidence Intervals Each scenario uses three replicates with small traffic disturbances as described in Section 3.3. For each KPI, the following procedure is applied:

1. Compute per-replicate statistic within the analysis window (3–10 s).
2. Calculate the mean across replicates as the reported value.
3. Estimate the 95% confidence interval using the t-distribution, given the limited number of replicates (n = 3) [30](for t-distribution methodology).

Confidence Interval Calculation With only three replicates per scenario, the t-distribution is more appropriate than the normal distribution [30](Chapter 9, Section 9.4 – small sample statistics). The 95% confidence interval is computed as:

$$\text{CI} = \bar{x} \pm t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} \tag{3.4}$$

where $\bar{x}$ is the sample mean of the three KPI measurements, s is the sample standard deviation, $n = 3$ is the number of replicates, and $t_{\alpha/2,n-1}$ is the critical t-value [30]. For a 95% confidence level with 2 degrees of freedom (df $= n - 1 = 2$), the critical value is approximately $t_{0.025,2} \approx 4.30$ [30]. This critical value is substantially larger than the corresponding z-value of 1.96 used for large samples, reflecting the increased uncertainty from the small sample size and resulting in wider confidence intervals [30].

The width of confidence intervals decreases with increasing sample size n, following the relationship CI width $\propto 1/\sqrt{n}$ [30, Chapter 9]. The three replications per scenario employed here provide initial estimates of the performance range. However, following the principles of statistical experimental design, a substantially larger sample size (e.g., 20–50 replications) would provide narrower confidence intervals and more robust parameter estimation. This represents a natural extension of the methodology for future work seeking to explore the parameter space more comprehensively.

4. Implementation

4.1. Baseline Scenario (No Remediation)

To establish a reference point for evaluating remediation methods, this section first simulates a high-load network with a single macro gNodeB serving 60 UEs without any congestion relief measures. The baseline results quantify the performance degradation under resource saturation and provide the metrics against which all remediation strategies are compared.

4.1.1. Parameter Settings

To establish high-load conditions, per-UE data rates are calibrated such that the network reaches the target cell load of 0.8 (80% resource utilization), and later that it surpasses said target cell load. This calibration uses the Shannon capacity calculation to determine realistic rates. The following paragraphs describe the iterative procedure. The initial step is to convert the SINR from decibels (dB) to linear units using equation 3.1 [19, Appendix C]. The SINR is set at 10 dB:

$$\gamma = 10^{\mathrm{SNR_{dB}}/10} = 10^{10/10} = 10$$

Now, γ=10, BW = 10 MHz and N = 1 (number of Antennas (Single Input Single Output (SISO) leads to 1)) were used to calculate the Shannon capacity with equation 3.2.

$$\widehat{C} = B \cdot \frac{1}{N} \sum_{i=1}^{N} \log_2(1+\gamma_i) = 10\,\mathrm{MHz}\ \log_2\big(1+10\big) \approx 34.59\,\mathrm{Mbps}.$$

The Shannon Capacity of 34.59 Mbps can be viewed as an upper bound [20]. A moderate load assumption of 0.25 Mbps per UE was used to calculate a first potential Cell load. This value was chosen to represent typical user requirements for light online activities such as email, web browsing, and basic data services, providing a realistic moderate-load reference point for the theoretical calculations [18][32]. Each UE having a mean data rate of 0.25 Mbps sums up to $0.25Mbps \cdot 60 = 15Mbps1$ for 60 UEs.

```
# First Simulation Traffic Model
dlTrafficModelMobileUes  = dict(model="onOff", packetSize=1400, dataRate=580000)
ulTrafficModelMobileUes  = dict(model="onOff", packetSize=1400, dataRate=83333)
dlTrafficModelStaticUes  = dict(model="onOff", packetSize=1400, dataRate=580000)
ulTrafficModelStaticUes  = dict(model="onOff", packetSize=1400, dataRate=83333)
```

Figure 4.1.: Traffic Model for first Simulation (Code lines)

Taking the Ratio $DataRate/ShannonCapacity$ the approximate Cell load was calculated: $15Mbps \div 34.59Mbps \approx 0.43$ Concluding, the moderate load assumption is approximately 43% of the theoretical Shannon capacity. A simulation for these configurations should show a cell load of about 0.4. The Cell load should be approximately 0.4. The results of a simulation run for this show very good SINR values between 45 and 69 db.

For the initial simulation, 60 UEs (3 static and 57 mobile) were configured. This light-load baseline served to verify correct ns-3 Playground operation, confirm realistic SINR values, and ensure absence of congestion for clean baseline behavior. Code listing 4.1 specifies a packet size of 1400 Bytes, UDP protocol, DL data rate 0.58 Mbits, and UL data rate 0.083 Mbits for both mobile and static UEs.

Based on the measured mean SINR from the initial simulation (54.88 dB), the required traffic rates for the 0.8 load target were derived using the Shannon lower bound [20]. The Shannon lower bound as in equation 3.3 established as follows: First, the Conversion from db to linear from the mean SINR value of 54.88 db, could be calculated:

$$\gamma = 10^{54.88_{\mathrm{dB}}/10} = 10^{54.88/10} = 18.2307.$$

Now the resulting γ of 18.2307 was inserted in the equation 3.2 as well as the bandwidth of 10 MHz (and N=1):

$$\widehat{C} = 10\,\mathrm{MHz}\,\log_2(1 + 18.2307) \approx 182.3075\,\mathrm{Mbps}.$$

And in the last calculation step, this Shannon capacity was utilized to determine the lower-bound check according to equation 3.3 :

$$C_{\mathrm{target}} = 0.8 \cdot 182.3075\mathrm{Mbps} \approx 145.846\,\mathrm{Mbps},$$

According to this, a target cell load of 0.80 is reached when the total data rate is approximately 145.846 Mbps. As there are 60 UEs the data rate per UE is approximately 2.43 Mbps. For the next simulation, this subresult was used to attempt to realize a cell load of 0.80 or 80%. Thus, the traffic model was adjusted as seen in figure 4.2 for the second simulation. Beside the data rate of now 2.43 Mbps for DL and 0.35

```
# Second Simulation Traffic Model (80% load)
dlTrafficModelMobileUes  = dict(model="onOff", packetSize=1400, dataRate=2430000)
ulTrafficModelMobileUes  = dict(model="onOff", packetSize=1400, dataRate=350000)
dlTrafficModelStaticUes  = dict(model="onOff", packetSize=1400, dataRate=2430000)
ulTrafficModelStaticUes  = dict(model="onOff", packetSize=1400, dataRate=350000)
```

Figure 4.2.: Traffic Model for Target Cell Load of 0.8

```
# Baseline Simulation Traffic Model (80%+15% load)
dlTrafficModelMobileUes  = dict(model="onOff", packetSize=1400, dataRate=2800000)
ulTrafficModelMobileUes  = dict(model="onOff", packetSize=1400, dataRate=400000)
dlTrafficModelStaticUes  = dict(model="onOff", packetSize=1400, dataRate=2800000)
ulTrafficModelStaticUes  = dict(model="onOff", packetSize=1400, dataRate=400000)
```

Figure 4.3.: Traffic Model for Baseline Scenario with target Cell Load of 80%+15%

Mbps for UL, no other configuration was changed for simulation two. The output of the second simulation is as follows: DL Cell load= 0.894; UL Cell load= 0.106.

The results indicate a high downlink cell load (0.894). Although this exceeds the target 0.8, it represents an acceptable high-load baseline because the congestion remains resource-limited: despite the elevated cell load of 0.894, SINR values remain high around 62.6 dB, indicating that radio signal quality is not the limiting factor. This asymmetry between high load and good SINR is characteristic of PRB-constrained scenarios, where the scheduler has insufficient physical resources to serve all traffic demand rather than insufficient signal quality.

Despite high cell load (0.894), SINR remains excellent (62.6 dB), confirming resource-limited overload. Congestion stems from a scarcity of PRBs, not from noise and interference[1].

The subsequent simulation was configured for a target cell load of 80%+15%. For DL $0.15 \cdot 2.43Mbps = 0.36Mbps$ and for UL $0.15 \cdot 0.35bit/s = 0.05Mbps$ were calculated as the additional data rate values. Consequently, the data rates for the next simulation resulted in $2.43Mbps + 0.36Mbps = 2.79Mbps \approx 2.8Mbps$ for the Downlink and $0.35Mbps + 0.05Mbps = 0.40Mbps$ for the uplink. This was then adjusted in the traffic model as seen in figure 4.3, while the other configurations remained unchanged.

As fourth simulation, the traffic model remained with the traffic model as seen in figure 4.3, but the simulation time was increased to 20 seconds, to observe the influence of the simulation time on the high-load situation.

Finally, as a fifth simulation, the simulation time was changed to 30 seconds and the setting *idealRLC=true* was adjusted to *idealRLC=false*. The intention was to try and

```
replicate_ids = [0, 1, 2]                              # <-- Anti-determinism: 3 replicates usually [0,1,2]
RATE_FACTOR_BY_REP = {0: 0.97, 1: 1.00, 2: 1.03}   # <-- Anti-determinism: tiny ±3% rate nudge
```

Figure 4.4.: The Python Code lines setting up the replicate_ids and the rate factor of these replicates.

```
# Iterate replicates (minimal addition)
for r in replicate_ids:                                  # <-- Anti-determinism: replicate loop
```

Figure 4.5.: The Python Code line which starts a simulation loop for r (three) replicates

make the effects of network congestion better visible by using an even more congested scenario [44]. The increase in simulation time to 30 seconds caused a different number of effective mobile UEs to be created. Thus, the number of mobile UEs was changed to 59 in order to keep an effective number of mobile UEs of 57.

Instead of a nominal "+20%" over the 80% profile, a robust overload condition was defined by the following KPI thresholds: DL cell load $p_{95} \geq 0.95$ in the [3–10] s window (with mean ≈ 0.85–0.90), while keeping the DL/UL ratio constant ($\sim 7{:}1$). Starting from the 80%+15% profile (2.80/0.40 Mbps per UE), the offered rates were increased in small steps and the resulting KPIs were evaluated. The smallest pair that consistently met the overload target across runs was 3.19 Mbps (DL) and 0.459 Mbps (UL) per UE, which preserves the DL/UL ratio ($3.19/0.459 \approx 6.95 \approx 7$). Internally, this scenario was labeled "+20%", but in this thesis it is referred to as "Overload (empirically tuned)" to emphasize that the rates were chosen by KPI-driven tuning rather than by a strict 20% arithmetic increment.

Final data rate: DL = 3.19 Mbps, UL = 0.459 Mbps.

To assess sensitivity to different UE distributions, an additional simulation was conducted with 20 static and 40 mobile UEs. The detailed positions of the 20 static UEs are listed in Appendix A.2.

During initial development, it was discovered that identical input files (SUMO trajectories, Sionna channels) combined with fixed random seeds produced completely deterministic outputs (verified via MD5 checksums) [44][29].

```
# General
general = cfg.General(
    simTime=10,                # The simulation time in seconds. default value is 60 seconds.
    randomSeed=42 + r          # <-- Anti-determinism: seed offset per replicate
)
```

Figure 4.6.: The Python Code changing the random seed for each replicate run.

In order to ensure statistically robust results, the simulation is executed in three replicates, each with slightly different traffic scaling and random seeds as depicted in the figures 4.4, 4.5 and 4.6.

To introduce realistic variability for statistical analysis while maintaining scenario reproducibility, a lightweight anti-determinism approach was implemented using traffic perturbations (±3%) and seed offsets, as described in Section 3.3 [30].

4.1.2. Results

Three baseline configurations were tested to calibrate the high-load threshold: Table 4.1

Table 4.1.: Summary of KPIs for the Main Baseline Configurations (S1-S3)

KPI	S1: Under 80%	S2: 80%	S3: 80%+15% (10 s)
`DL_Load_mean`	0.138	0.663	0.836
`DL_Load_p95`	0.139	0.684	0.838
`DL_Thrpt (Mbps)`	2.958	22.750	27.697
η (bit/s/Hz)	0.296	2.275	2.770
`SINR_m (dB)`	61.898	64.600	62.572
`SINR_p5 (dB)`	49.970	39.604	38.951
`SINR_p50 (dB)`	66.715	68.051	64.868
`SINR_p95 (dB)`	72.259	81.149	79.290
`BLER_m`	0.917	0.849	0.839

Looking at the results in Table 4.1, the key observations are: S1 (Normal Load) represents a normally-loaded cell with 0.14 cell load and minimal congestion. S2 (Target 80%) reaches exactly the defined congestion threshold at 0.66 load, representing the boundary between acceptable and overloaded operation [1]. S3 (High-Load) exceeds the threshold significantly at 0.84 load, representing true congestion with resource-limited overload [1].

For the extended baseline configurations Table 4.2 presents three additional high-load scenarios to test robustness across different traffic durations and UE distributions:

These extended scenarios 4.2 demonstrate that the high-load behavior (between approximately 0.83–0.87 cell load) is consistent across different traffic intensities and UE distributions, validating S3 as a representative high-load baseline for remediation testing.

Table 4.2.: Summary of KPIs for Extended Baseline Configurations (S4-S6)

KPI	S4: 80%+15% (20 s)	S5: 80%+20% (30 s)	S6: 40+20 UEs
DL_Load_mean	0.836	0.874	0.836
DL_Load_p95	0.838	0.875	0.838
DL_Thrpt (Mbps)	27.694	32.899	27.697
η (bit/s/Hz)	2.769	3.290	2.770
SINR_m (dB)	62.606	60.921	56.767
SINR_p5 (dB)	38.791	37.337	34.973
SINR_p50 (dB)	65.269	62.117	60.341
SINR_p95 (dB)	79.499	79.643	72.703
BLER_m	0.839	0.832	0.839

Simulation time remained approximately 3 hours across all configurations. This consistency occurs because the main computational cost arises from per-UE operations (scheduling, packet processing, logging), which scales similarly whether UEs are static or mobile. Pre-computed trajectories (SUMO) and radio propagation (Sionna ray-tracing) do not significantly increase runtime.

Figure 4.7 and Figure 4.8 visualize the cell load and SINR across all six baseline scenarios. For the remediation method evaluations (Sections 4.2 – 4.4), **scenario S3** (80%+15%, 10 s) was selected as the reference baseline because:

1. It reliably exceeds the 80% congestion threshold (0.836 load).
2. It represents realistic high-capacity operation without extreme overload.
3. Its SINR distribution (62.572 dB mean, 39–80 dB range) reflects real-world diversity.

This baseline choice enables fair, reproducible comparison of remediation strategies.

4.2. Methods 1: Additional Small Cell

In order to remediate a congestion scenario, this thesis first examines the method of adding an additional small cell. In real networks this can be compared to having a second gNodeB kept in standby and activating it in case of a high-load event [11][17].

In the baseline scenario, a single gNodeB provides coverage for the entire area. This node is hereafter referred to as the macro cell, as it plays the role of a wide-area coverage site. In contrast, the additional node introduced in the remediation strategies

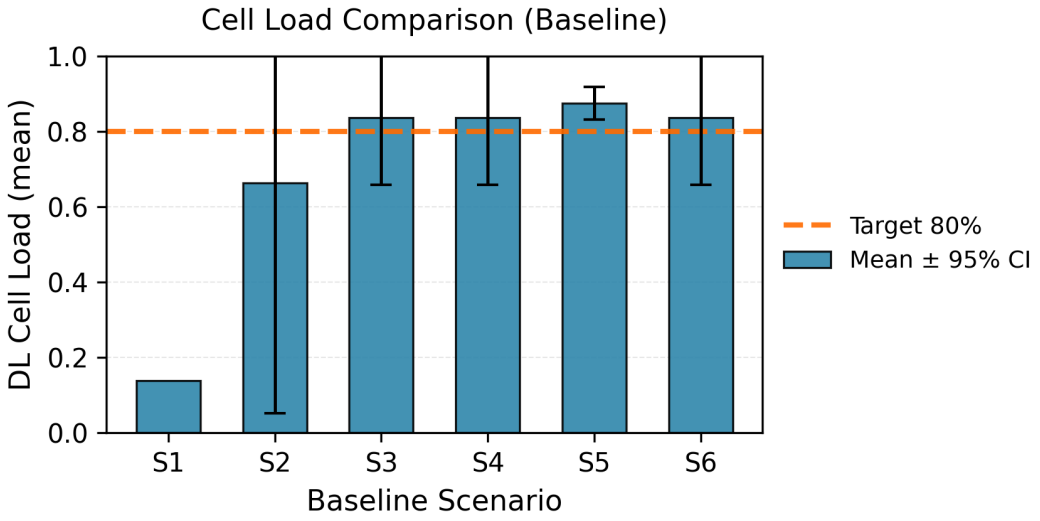

Figure 4.7.: Mean DL Cell Load for the various baseline simulations

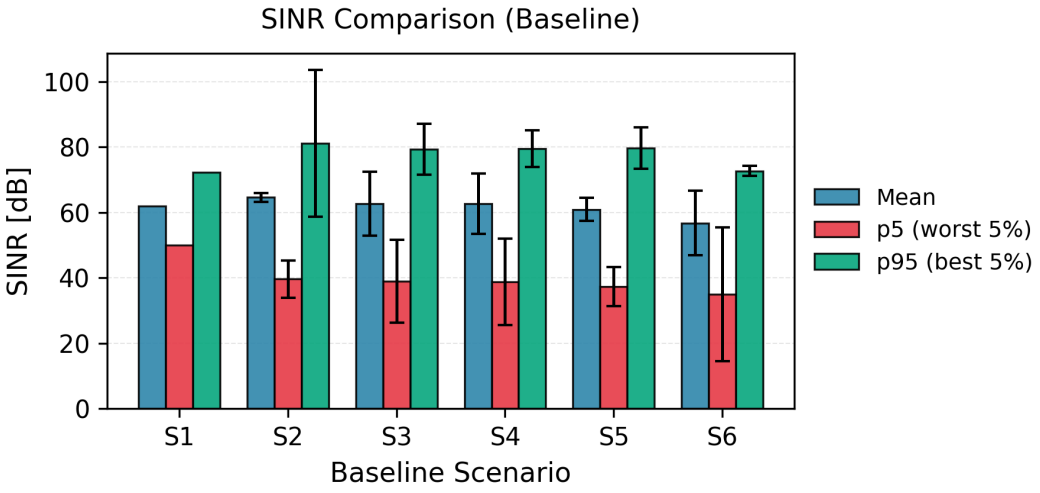

Figure 4.8.: Mean SINR values for the various baseline simulations

is referred to as a small cell. In the simulation, both nodes are modeled using the same gNodeB configuration for consistency. The distinction between the 'macro' and the 'small cell' is therefore not due to different hardware parameters, but results from their deployment role: the first node provides full-area coverage, while the second node is added at an alternative location to locally relieve congestion.

It is therefore important to investigate different strategies for determining an optimal positioning of the small cell within the given network.

Based on the literature review, three representative placement strategies were identified and are implemented and compared in this section of the thesis:

- **K1 – Geometric split:** The first placement strategy follows a simple geometric principle: the area between the first gNodeB and the additional small cell is divided into two roughly equal regions [45][28]. This is equivalent to a Voronoi-based partitioning, where UEs are associated with the geographically nearest site [45][28]. Such a geometric split provides a fair, distance-based distribution of users between the first and the additional small cell. It does not require knowledge of traffic hotspots, making it a generic baseline strategy that is often used in simulation studies to approximate balanced coverage.
- **K2 – Hotspot overlay:** The second strategy places the small cell directly at the location of a traffic hotspot, i.e. where user density and demand are particularly high [31][26]. This reflects a widely adopted operator practice in real networks, where small cells are deployed at stadiums, train stations, or other crowded venues in order to offload the macro and to ensure quality of service for the majority of users present in that hotspot. In this approach, the goal is not to balance load geometrically across the network, but to maximize local capacity exactly where it is most needed.
- **K3 – Intermediate split:** The third strategy represents a compromise between geometric balancing and hotspot targeting. Here the small cell is positioned between the first gNb site, the macro site, and the hotspot area. The idea is to offload a larger fraction of the macro's users than in the pure hotspot overlay, while still steering capacity towards the hotspot direction. This can be interpreted as a load-balancing approach: the macro is relieved more strongly than with the hotspot overlay, but the total throughput gain is not as high as in the pure geometric split. Such intermediate placements are studied in the literature as a means to trade off macro relief and overall efficiency.

4.2.1. Additional Small Cell Activated

To evaluate the impact of placement strategy, three candidate locations were selected based on geometric, traffic-driven, and compromised design principles. For each candidate, the baseline overload scenario was replicated three times using traffic perturbations (±3%) to assess robustness. Initial simulations with fixed seeds (41, 42, 43) produced identical results, confirming deterministic behavior. To introduce realistic variability without altering the scenario, subsequent simulations employed traffic perturbations (±3% per-UE rates) with offset random seeds, generating three distinct replicates per placement strategy. Implementation Remarks: When transitioning between gNB layout scenarios (K1, K2, K3), the ray-tracing channel cache must be cleared to prevent geometric mismatches.

4.2.1.1. Positioning Method 1 - Geometric split

To calculate the position of the small cell using a geometric split strategy, which is a method based on Voronoi tessellation, the coverage area is partitioned into regions closest to each base station location [38]. This Voronoi-based partitioning is a common approach in cellular network planning [28], as it naturally distributes UEs to the geographically nearest site [45]. First, the known parameters have to be specified. The macro gNodeB is located at $M = [22, 63, 4.5]$, and the target is to determine the small cell position $S = [x_2, y_2, z_2]$. The optimization focuses on the $[x, y]$-plane, while z adjusts the installation height. Since the EIRP is kept constant (about 30 dBm), the boundary Reference Signal Received Power (RSRP) line coincides with the geometric divider between M and S. RSRP is the power level at the UE receiver, which determines cell selection and handover decisions. For an equal y-coordinate ($y_2 = 63$), this boundary line is given by the midpoint formula

$$x_b = \frac{x_M + x_S}{2}. \tag{4.1}$$

The design goal is to achieve an approximately equal offload, assuming UEs are uniformly distributed along the x-axis within the target corridor $x \in [22, 52]$ of width $W = 30\,\mathrm{m}$. For an offload share $\alpha = 0.50$ towards $g\mathrm{NB}_2$, the boundary is placed at

$$x_b = x_{\max} - \alpha W = 52 - 0.5 \cdot 30 = 37. \tag{4.2}$$

Using the midpoint relation $x_b = (x_M + x_S)/2$, the small cell coordinate follows as

$$x_S = 2x_b - x_M = 2 \cdot 37 - 22 = 52. \tag{4.3}$$

The remaining coordinates are chosen as $y_2 = 63$ (same street axis) and $z_2 \approx 5\,\mathrm{m}$ (street-level mast/facade), yielding candidate K1 at

$$S_1 = (52, 63, 5).$$

The final boundary location $x_b = 37$ thus effectively splits the corridor into two equal halves, assigning the eastern part ($x > 37$) to the small cell and achieving the target offload of approximately 50%. This baseline assumption of uniform UE distribution makes the method natural for further comparative evaluations.

Using this first strategy, the program added coordinate S_1 to `gnbLocations` for the baseline scenario. The simulation was subsequently run with the different replicates.

4.2.1.2. Positioning Method 2 - Hotspot overlay

For the previous method we already established some known parameters. The objective of this strategy is to let the small cell dominate in the local hotspot around $\mathbf{H} = (52, 33)$, which corresponds to the static UE location. The hotspot location at 52, 33) was chosen based on spatial UE clustering analysis. In the simulated scenario, multiple user equipments are concentrated near this location—both static UEs positioned there as anchor points and mobile UEs following SUMO trajectories that converge in this region. This represents a realistic scenario where network operators deploy small cells at identified hotspots with persistent high user density. Using a simple RSRP dominance rule (bigger than or equal to 3 dB stronger than the macro), it can be shown that a small cell placed within about 30 m of this point will dominate the hotspot area [20, Chapter 6]. This 3 dB threshold is a standard industrial handover hysteresis margin ensuring reliable cell selection. As a result, candidate K2 was positioned directly at (52, 33, 5) to provide maximum coverage in the hotspot.

To achieve the objective of this strategy, an RSRP dominance criterion is applied, requiring the small cell to provide at least $\Delta = 3\,\mathrm{dB}$ stronger RSRP than the macro. The two-dimensional distance between the macro and the hotspot is

$$d_M = \sqrt{(52-22)^2 + (33-63)^2} = \sqrt{30^2 + 30^2} \approx 42.43\,\mathrm{m}. \tag{4.4}$$

Height differences are neglected, since including the z-dimension has little effect here. Assuming a log-distance path loss model under near-LOS conditions with path loss exponent $n \approx 2$ and equal transmit powers $P_M \approx P_S$ [20], the dominance condition can be written as

$$P_S - 10n\log_{10}(d_S) \;\geq\; P_M - 10n\log_{10}(d_M) + \Delta. \tag{4.5}$$

This simplifies to

$$d_S \leq d_M \cdot 10^{-\Delta/(10n)}. \tag{4.6}$$

Substituting the numerical values gives

$$d_S \leq 42.43 \cdot 10^{-3/20} \approx 42.43 \cdot 0.708 \approx 30.0\,\text{m}. \tag{4.7}$$

Hence, placing the small cell within about 30 m of the hotspot centroid ensures that it dominates the hotspot with at least a 3 dB margin.

Following this design rule, a practical coordinate choice for K2 is $\mathbf{S}_2 = (52, 33, 5)$, i.e. an installation inside a fabric hall near the hotspot.

The coordinate for K2 was added to gnbLocations in the program code for the baseline scenario. The simulation was run three times for the different replicates.

4.2.1.3. Positioning Method 3 - Intermediate split

The third positioning strategy aims at a somewhat more aggressive offload than in K1. Instead of splitting the UE corridor evenly, an offload share of approximately $\alpha = 0.63$ was chosen, so that about 63% of the corridor is served by the small cell. Here, the corridor refers to the main traffic route (x) where UEs are concentrated. The motivation behind this choice was to shift the dominance region clearly toward the small cell, but without placing the cell completely at the corridor boundary.

The corridor is again defined as $x \in [22, 52]$, giving a width of $W = 30\,\text{m}$. For the desired offload share, the boundary line (4.2) is

$$x_b = x_{\max} - \alpha W = 52 - 0.633 \cdot 30 = 33.0. \tag{4.8}$$

With the relation $x_b = (x_M + x_S)/2$, the small-cell x-coordinate follows 4.3 as

$$x_S = 2x_b - x_M = 2 \cdot 33 - 22 = 44. \tag{4.9}$$

Together with $y_2 = 63$ (street axis) and $z_2 \approx 5\,\text{m}$ (street-level mounting), the resulting candidate position is $\mathbf{S}_3 = (44, 63, 5)$. The corresponding boundary line at $x_b = (22 + 44)/2 = 33$ assigns the region $x > 33$ to the small cell. Under the assumption of a homogeneous UE distribution, the offload share is

$$\alpha = \frac{52 - x_b}{W} = \frac{52 - 33}{30} \approx 0.633, \tag{4.10}$$

which confirms that about 63% of the corridor is offloaded. This choice of $\alpha \approx 0.63$ reflects a compromise: it provides stronger offloading than the 50/50 baseline (K1), while avoiding an extreme placement at the corridor edge. It represents a design where the small cell is intentionally favored, to relieve the macro more aggressively.

Finally, note that unequal EIRP values between macro and small cell shift the boundary line. In a one-dimensional model, the RSRP equality condition reads

$$\frac{d_M}{d_S} = 10^{\Delta P/(10n)},$$

where $\Delta P = P_M - P_S$ is the transmit power difference and n is the path loss exponent. For a total macro small cell separation $D = |x_S - x_M|$, this implies

$$d_S = \frac{D}{1 + 10^{\Delta P/(10n)}}, \qquad d_M = D - d_S.$$

Thus, a higher macro EIRP pushes the boundary closer to the small cell, while a stronger small cell has the opposite effect. Following this method, a practical coordinate choice for K3 is $\mathbf{S}_3 = (44, 63, 5)$.

The coordinate for K2 was added to gnbLocations in the program code for the baseline scenario. The simulation was run three times for the different replicates.

4.2.2. Comparison and results

To evaluate the effectiveness of each positioning strategy, nine simulation runs were conducted (three positioning methods × three random seeds). Table 4.3 summarizes the key performance indicators (KPIs) extracted from each scenario.

Figure 4.9 illustrates the three small cell placement strategies overlaid on the baseline UE mobility trajectories. The UE paths remain identical to the baseline scenario (Figure 3.1) since the trajectory generation is decoupled from radio network design. However, the figure visually demonstrates the spatial relationship between each cell position and the UE paths, revealing why hotspot targeting (K2) should be more effective than geometric partitioning (K1): the hotspot position (K2) is deliberately placed near high-density UE routes, whereas the geometric split (K1) divides coverage equally regardless of traffic distribution.

Reproducibility and stochasticity: In initial testing with fixed random seeds, all three seeds produced identical KPIs, confirming the simulation's deterministic foundation (pre-computed SUMO trajectories, ray-traced propagation, and constant-rate traffic). However, for final results reported in Table 4.3, the simulations were re-executed with three separate replicates using lightweight anti-determinism (seed

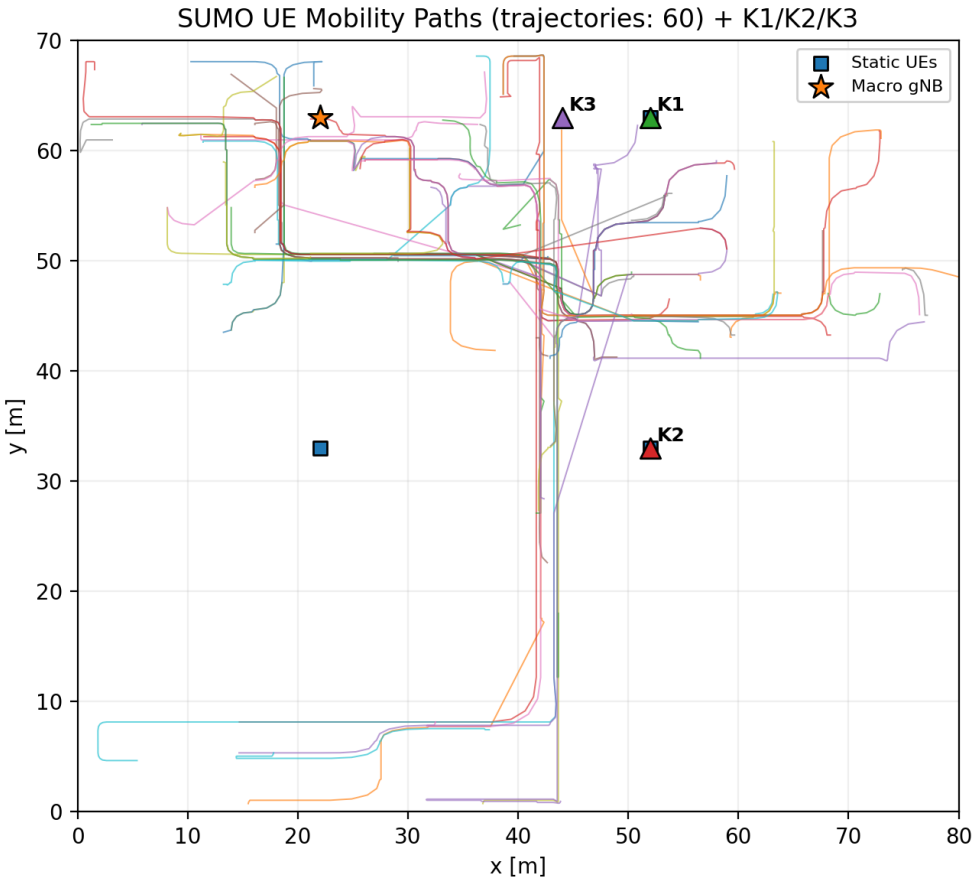

Figure 4.9.: SUMO UE Mobility Paths including the three possible additional Small Cell positions - K1 (Geometric Split), K2 (Hotspot Overlay) and K3 (Intermediate Split)

Table 4.3.: Summary of resulting KPIs for the three positioning methods (Method 1)

KPI	K1 Geometric Split	K2 Hotspot Overlay	K3 Intermediate Split
DL_Cell0_Load_mean	0.677	0.604	0.697
DL_Cell0_Load_p95	0.845	0.749	0.816
DL_Cell1_Load_mean	0.690	0.721	0.765
DL_Cell1_Load_p95	0.856	0.813	0.869
DL_Load_mean	0.683	0.663	0.731
DL_Load_p95	0.872	0.813	0.872
DL_Thrpt (Mbps)	19.348	24.272	25.651
η (bit/s/Hz)	1.935	2.427	2.565
SINR_m (dB)	32.232	33.080	32.989
SINR_p5 (dB)	3.835	4.017	4.347
SINR_p50 (dB)	21.957	23.794	24.114
SINR_p95 (dB)	73.289	69.388	74.848
BLER_m	0.888	0.860	0.855

variation in traffic generation), introducing realistic stochasticity. The error bars in Figures 4.10–4.12 reflect this variability, demonstrating statistically robust results across replicates.

Key findings from tabular data: K2 (Hotspot Overlay) achieves the best macro cell relief with Cell 0 load reduced to 0.604, representing approximately 27.9% reduction relative to baseline (0.838). This compares favorably to K1 (0.677) and K3 (0.697). Throughput performance shows a different ranking: K3 delivers the highest total throughput (25.651 Mbps), followed closely by K2 (24.272 Mbps), while K1 lags significantly (19.348 Mbps). Overall cell load (mean across both cells) reveals K2's strategic advantage at 0.663, suggesting the best balance between relieving macro load and avoiding small cell overload. SINR values remain comparable across all methods (approximately between 32–33 dB mean), confirming signal quality is not the limiting factor; K2 shows a slight advantage at 33.080 dB.

Figure 4.10 visualizes the load distribution per cell, making the K2 advantage explicit. K2's macro cell load (0.604) is clearly lowest among the three, while its small cell load (0.721) remains manageable—not overloaded like K3. This figure demonstrates that K2 does not simply shift congestion from macro to small cell; instead, it balances load effectively across both cells. The confidence intervals (error bars) show that results are consistent across replicates, validating the stability of the placement strategy.

Figure 4.11 illustrates the fundamental trade-off between capacity relief and throughput delivery. Each point in this graph represents one positioning strategy, showing the

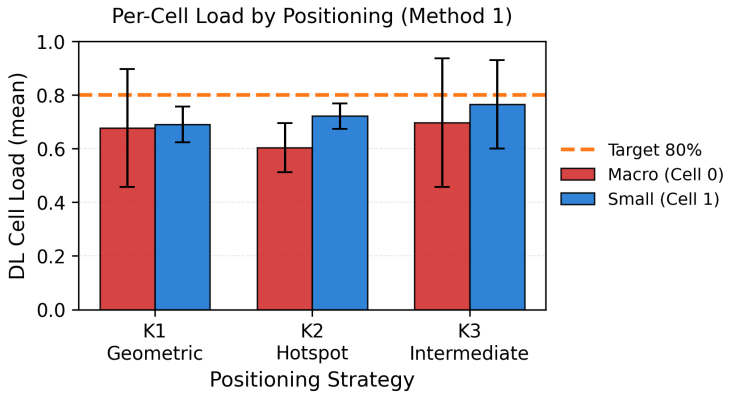

Figure 4.10.: Per-Cell Load Comparison for varying Positioning Strategies

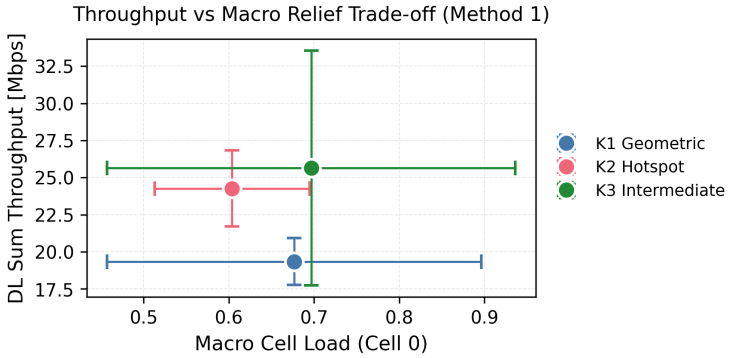

Figure 4.11.: Throughput vs. Macro Relief Trade-off

relationship between macro relief (x-axis) and total throughput delivered (y-axis). K2 occupies the optimal position: achieving strong macro relief (load ≈ 0.604) while maintaining good throughput. K3 is positioned further up (more throughput, 25.651 Mbps) but fails to relieve the macro cell as much (load 0.679), indicating sub-optimal spatial targeting. K1 performs poorly in both dimensions. This scatter plot is essential for showing that K2 is not just locally best; it achieves the best trade-off overall.

Figure 4.12 compares signal quality across placements. SINR values are remarkably similar (approximately between 32–33 dB), confirming that signal quality is not the distinguishing factor between strategies. K2's slight SINR advantage (33.080 dB) indicates that its hotspot location captures slightly better-positioned UEs, but this is secondary to its load-balancing superiority. This figure justifies focusing on load and throughput metrics rather than coverage—the bottleneck is clearly capacity, not radio quality.

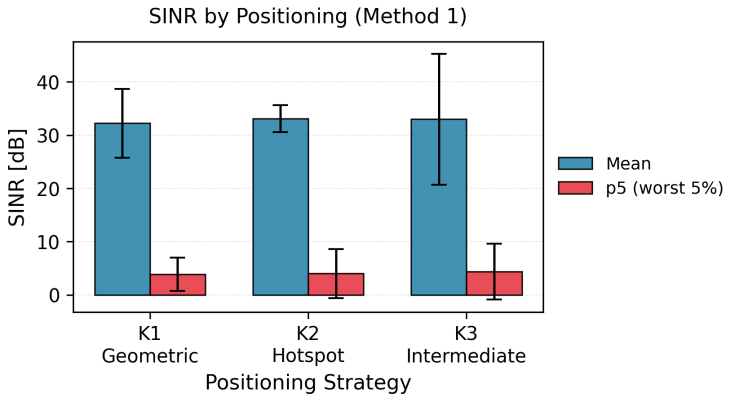

Figure 4.12.: SINR Comparison by Positioning Strategy

Among the three placement strategies, K2 (Hotspot Overlay) [31] emerges as the clear winner, achieving the best macro cell relief (0.604 load) while maintaining competitive throughput (24.272 Mbps). K3 attains higher throughput but compromises on macro relief (0.697 load), while K1 underperforms on both metrics. These results underscore an important principle: spatial targeting is more effective than geometric balancing for non-uniform traffic distributions [31][34]. Placing the small cell directly in the traffic hotspot concentrates capacity where it is most needed, matching real-world operator practice at high-demand venues, such as stadiums or concert halls.

4.3. Method 2: Transmission Power Adjustment

While Method 1 (small cell deployment) requires capital investment and new infrastructure, transmission power adjustment represents a parameter-based remediation approach using existing hardware. This section evaluates whether increased transmit power can mitigate resource-limited overload by improving signal strength and SINR.

4.3.1. Concept and Parameters

Transmission power adjustment represents a parameter-based remediation method that does not require additional hardware deployment. The hypothesis is that increasing base station transmit power improves received signal strength and SINR, thereby enabling higher modulation orders and per-user throughput[15] (Chapter 5). However,

this approach only benefits resource-limited overload if the constraint is signal quality rather than radio resource availability.

For this method, transmission power was increased incrementally in 3 dB steps (from 30 dBm baseline to 33, 36 and 39 dBm) while keeping all other baseline parameters [13]. The 3 dB step size was chosen to represent meaningful and practical increments (each 3 dB = 2x linear power increase) while respecting regulatory limits for n78 band (40–43 dBm max).

4.3.2. Results

Table 4.4.: Summary of KPIs for transmit-power sweep (Method 2)

KPI	Tx 30 dBm	Tx 33 dBm	Tx 36 dBm	Tx 39 dBm	Tx 60 dBm
DL_Load_mean	0.836	0.836	0.836	0.836	0.794
DL_Load_p95	0.838	0.838	0.838	0.838	0.824
DL_Thrpt (Mbps)	27.697	27.697	27.697	27.697	28.336
η (bit/s/Hz)	2.770	2.770	2.770	2.770	2.834
SINR_m (dB)	62.572	65.496	68.496	71.496	91.195
SINR_p5 (dB)	38.951	43.573	46.573	49.573	67.703
SINR_p50 (dB)	64.868	66.711	69.711	72.711	94.891
SINR_p95 (dB)	79.290	82.400	85.400	88.400	111.338
BLER_m	0.839	0.839	0.839	0.839	0.833

Surprising result: Power does not relieve capacity-limited congestion. Contrary to expectations, increasing transmission power from 30 to 39 dBm did not reduce cell load or improve throughput. Table 4.4 reveals a significant asymmetry in the results: cell load and throughput remained constant (load 0.836, throughput 27.697 Mbps) while SINR improved dramatically (62.572 → 71.496 dB mean). The scheduler was demand-saturated rather than signal-constrained—increasing transmit power improved signal conditions but could not create additional radio resources.

Figure 4.13 visualizes this critical finding: the cell load curve is completely flat across 30–39 dBm (operational range), with only a marginal dip at 60 dBm (theoretical boundary) [15][20]. This horizontal line directly contrasts with what one would expect in an interference-limited scenario, where improved signal quality should translate to lower load. The flatness of this curve is the key evidence that capacity, not coverage, is the bottleneck.

Figure 4.14 demonstrates the disconnect between signal quality and congestion relief. SINR increases monotonically from 62.572 dB (30 dBm) to 91.195 dB (60 dBm), show-

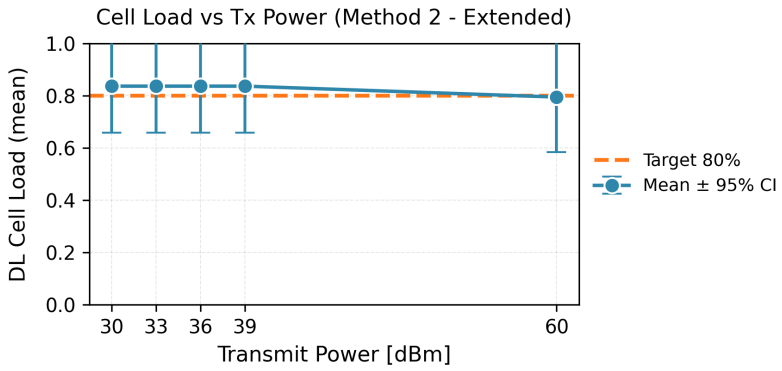

Figure 4.13.: Mean DL Cell Load vs. Transmit Power

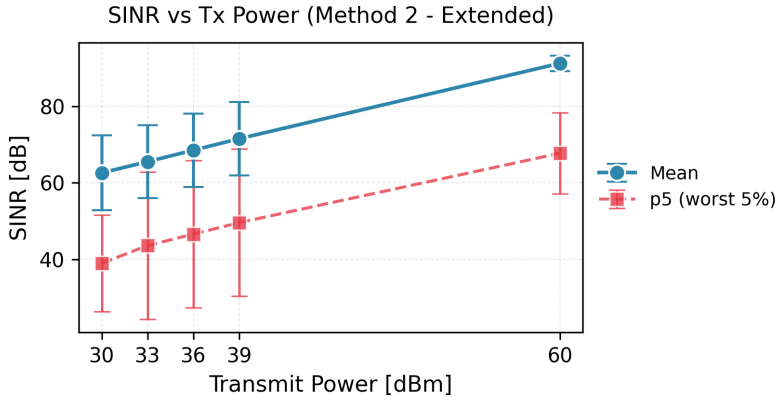

Figure 4.14.: Mean SINR vs. Transmit Power

ing excellent signal quality improvements. However, this signal-quality gain translates to zero cell-load improvement [15][20]. This figure pair (flat load + rising SINR) is essential for proving that the bottleneck is radio resources, not interference. In an interference-limited scenario, the SINR improvement would have immediately reduced cell load; here it does not, validating the diagnosis of resource-limited overload [20, Chapter 15].

Implications for remediation strategy selection: This finding demonstrates a critical principle for network operators: diagnosing congestion type is essential before selecting remediation methods. Transmission power adjustment is most effective for interference-limited or service-quality-limited congestion, where poor SINR prevents UEs from achieving adequate data rates [15, Chapter x],[37][11].

The 60 dBm result serves as a theoretical boundary condition, showing that even unrealistic power levels cannot overcome resource scarcity in this scenario. This validates that Method 2 (transmit power adjustment) is ineffective for resource-limited high-load scenarios and reinforces the need for capacity-adding solutions (small cells, bandwidth scaling) instead.

4.4. Method 3: Bandwidth Scaling

Bandwidth scaling represents a spectrum-based remediation strategy that directly increases cell capacity by allocating wider frequency channels. Unlike power adjustment (signal quality) or small cell deployment (spatial distribution), bandwidth scaling directly expands the PRB pool.

From equation 2.2, the Shannon-Hartley theorem establishes that channel capacity scales linearly with bandwidth:

$$C = B \cdot \log_2(1 + \text{SNR}).$$

In practical terms, doubling bandwidth (e.g., from 10 MHz to 20 MHz) approximately doubles the theoretical capacity. In 5G NR, this translates directly to proportionally more Physical Resource Blocks (PRBs): each 5 MHz of spectrum at 60 kHz subcarrier spacing provides approximately 26 PRBs. Thus, scaling from 10 MHz ($\approx$ 51 PRBs) to 15 MHz ($\approx$ 79 PRBs) and 20 MHz ($\approx$ 106 PRBs) effectively multiplies the scheduler's resource budget.

The fundamental question this method addresses is: In a resource-limited scenario, does expanding PRB availability reduce cell load and maintain throughput? This evaluation is performed under an optimistic assumption, that sufficient spectrum is available to quantify the technical potential of bandwidth scaling as a congestion remediation.

The per-slot throughput [43] at any base station is governed by the following relationship between physical layer parameters and protocol efficiency:

$$\text{Throughput} = (1 - \text{BLER}) \cdot \frac{\text{BitsPerBlock}}{\text{TransmitTimePerBlock}}$$

where:

- **BLER** is the probability that a transmitted Physical Resource Block is corrupted due to poor channel conditions

```
if nMobileUes > 0 and nStaticUes == 0:
    totalBandwidth = 10.e6 # 70.e6 default
elif nMobileUes == 0 and nStaticUes > 0:
    totalBandwidth = 30.e6
elif nMobileUes > 0 and nStaticUes > 0:
    totalBandwidth = 100.e6
```

Figure 4.15.: Python Code for setting the bandwidth in the simulation environment. Part 1

- **BitsPerBlock** depends on the Modulation and Coding Scheme (MCS) selected; higher MCS (e.g., 256-QAM) pack more bits per symbol, but require higher SINR
- **TransmitTimePerBlock** is fixed by the slot structure (typically 0.5 ms or 1 ms in 5G NR)

4.4.1. Parameters and Simulation Configuration

In 5G NR, at 60 kHz subcarrier spacing (numerology $\mu = 2$), each 5 MHz channel accommodates approximately 25–26 PRBs [43]. As established in Section 2.3.3, a 10 MHz channel provides ≈ 51 PRBs, 15 MHz provides ≈ 79 PRBs, and 20 MHz provides ≈ 106 PRBs.

The baseline scenario was repeated with bandwidth scaled to 15 MHz and 20 MHz, while traffic demand per UE remained constant (3.19 Mbps DL, 0.459 Mbps UL per UE, 60 UEs total). This provides an ideal-case evaluation: if sufficient spectrum is available, does cell load decrease proportionally, and is throughput maintained or improved?

The three configurations tested are:

1. **Baseline:** 10 MHz channel, ≈ 51 PRBs
2. **15 MHz scaled:** 15 MHz channel, ≈ 79 PRBs (55% increase)
3. **20 MHz scaled:** 20 MHz channel, ≈ 106 PRBs (108% increase)

The increase in available PRBs is not perfectly proportional to the bandwidth scaling, as guard bands and control channel overhead reduce the effective resource utilization, particularly at smaller bandwidths [4]. Thus, the 20 MHz configuration provides $106/51 \approx 2.08$ times the baseline PRBs, yielding an increase of $(106 - 51)/51 \approx 108\%$ rather than the linear expectation of 100%. All other parameters (UE positions, traffic profiles, antenna configuration, propagation model) remain identical to the baseline, isolating the effect of bandwidth on resource availability and cell load.

```
cc = cfg.ComponentCarrier(
    carrierFrequency=3.7e9,                    # Carrier frequency in Hz. The channel mod
    totalBandwidth=totalBandwidth,                    # The carrier bandwidth in Hz.
    numerology=2                               # The PHY layer numerology. Only numerolo
)
```

Figure 4.16.: Python Code for setting the bandwidth in the simulation environment. Part 2

Bandwidth scaling is performed by varying the carrier bandwidth parameter, as shown in Figure 4.15 for the baseline and in Figure 4.16 for scaling runs.

4.4.2. Results

To evaluate the effectiveness of bandwidth scaling, simulations were conducted at 10 MHz (baseline), 15 MHz, and 20 MHz. Table 4.5 summarizes the key performance indicators.

Table 4.5.: Summary of KPIs for bandwidth scaling (Method 3). Bandwidth values represent allocated spectrum for the gNodeB; 10 MHz is the baseline operational configuration.

KPI	10 MHz	15 MHz	20 MHz
DL_Load_mean	0.836	0.500	0.188
DL_Load_p95	0.838	0.500	0.189
DL_Thrpt (Mbps)	27.697	28.728	10.672
η **(bit/s/Hz)**	2.770	1.915	0.534
SINR_m (dB)	62.572	56.435	56.117
SINR_p5 (dB)	38.951	28.411	33.931
SINR_p50 (dB)	64.868	57.804	57.844
SINR_p95 (dB)	79.290	77.816	71.704
BLER_m	0.839	0.832	0.938

Highly effective capacity relief at 15 MHz: Scaling bandwidth to 15 MHz was highly effective for relieving resource-limited overload. Compared to the baseline 10 MHz scenario, cell load dropped dramatically from 0.836 to 0.500—a **40% reduction**. Critically, throughput remained stable (27.697 $\rightarrow$ 28.728 Mbps), indicating that the network successfully utilized the additional spectrum to serve user demand without quality degradation. Spectral efficiency decreased slightly (2.770 $\rightarrow$ 1.915 bit/s/Hz), as expected when spreading the same load across wider bandwidth. SINR showed a

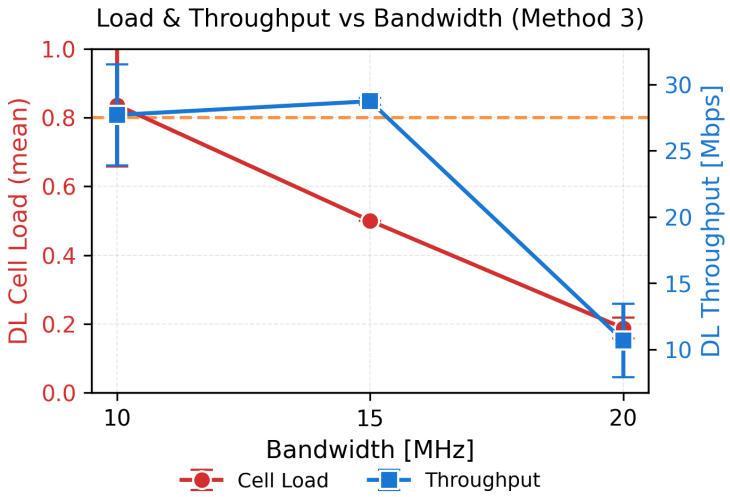

Figure 4.17.: Cell Load and Throughput vs. Bandwidth.

modest decrease (62.572 → 56.435 dB), expected due to wider bandwidth reducing power density, but remained in acceptable ranges [20].

Figure 4.17 visualizes the capacity-load trade-off across the bandwidth sweep. The dual-axis plot reveals distinct trends: cell load decreases monotonically from 0.838 (10 MHz) to 0.500 (15 MHz) to 0.189 (20 MHz), while throughput increases from 27.697 Mbps at 10 MHz to a peak of 28.728 Mbps at 15 MHz. The 15 MHz configuration represents an ideal "sweet spot," where additional PRBs directly reduce resource contention without degrading signal quality, allowing the scheduler to serve waiting UEs more efficiently. However, throughput anomalously collapses to 10.672 Mbps at 20 MHz, a counter-intuitive behavior that suggests factors beyond simple bandwidth scaling are at play (discussed in detail below).

Figure 4.18 shows SINR values across the bandwidth configurations. Mean SINR decreases from 62.572 dB at 10 MHz to 56.435 dB at 15 MHz, with a slight further reduction to 56.117 dB at 20 MHz. This decline is a predictable consequence of spreading fixed transmit power over a wider frequency spectrum, resulting in lower power density per subcarrier [20]. Critically, the SINR reduction does not prevent the load-mitigation benefit at 15 MHz, indicating that the primary constraint on cell performance is radio resource availability (PRB count), not signal quality or coverage. This finding validates the thesis that targeted spectrum allocation can effectively address high-load congestion in capacity-constrained scenarios.

Unexpected degradation at 20 MHz: The observed throughput for the 20 MHz

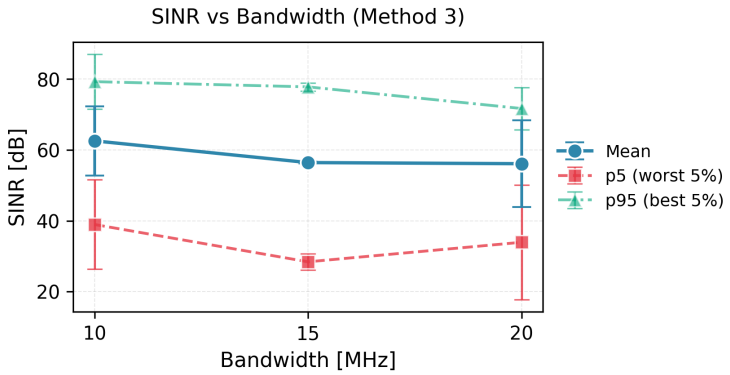

Figure 4.18.: Mean SINR vs. Bandwidth.

configuration is considerably lower than theoretically expected. According to Chapter 2 of Dahlman et al. [15, p. 35], the peak data rate of a wireless system is controlled by the available system bandwidth and the spectral efficiency:

$$\text{Peak data rate} = \text{System bandwidth} \times \text{Peak spectral efficiency}$$

This relationship implies that, under the condition that the spectral efficiency remains constant, increasing the bandwidth from 10 MHz to 20 MHz should roughly double the achievable data rate for users, as more physical resource blocks become available for user scheduling. However, in practice, the simulation's combination of a fixed per-user traffic demand as in figure 4.3 and limited scheduler adaptation led to underutilization of the provided bandwidth in the 20 MHz scenario. As a result, the additional resources were not effectively filled and overall network throughput did not scale as expected in a real-world deployment with sufficiently high traffic demand.

Therefore, the throughput values obtained for the 20 MHz case in this simulation are not representative of actual network potential, but rather reflect an artificial limit set by simulation parameters. In operational 5G NR systems, increasing the channel bandwidth produces proportional capacity increases - as long as user and traffic demand, along with network configurations, are sufficient to utilize those additional resources [15, p. 35].

Conclusion for Method 3: Bandwidth scaling to 15 MHz effectively mitigates resource-limited overload with 40% cell load reduction while maintaining user throughput. This method directly adds capacity by allocating additional PRBs, making it the **most technically effective** of the three remediation strategies. However, significant practical barriers limit deployment: acquiring additional licensed spectrum requires regulatory approval and negotiation with other operators, and associated costs scale

with bandwidth and geographic coverage. Despite its technical superiority, real-world constraints mean that this method is reserved for scenarios where temporary spectrum (e.g., event-specific licenses) can be obtained, or where permanent spectrum expansion is justified by market demand.

5. Evaluation

5.1. Evaluation and Comparison of Remediation Methods

Following the detailed analysis of the baseline scenario and three remediation methods in Chapter 4, this chapter provides a comprehensive comparative evaluation. The primary evaluation metric is cell load reduction relative to the high-load baseline scenario (80%+15%, 10s), with secondary consideration for throughput preservation, SINR impact, and practical feasibility.

Table 5.1 summarizes the improvement (delta) values for each remediation strategy compared to the baseline scenario. Delta values represent the difference in KPIs: negative values indicate improvement (e.g., lower load), positive values indicate degradation (e.g., higher BLER).

Table 5.1.: Improvement vs. baseline (80%+15%, 10s): Δ-values are averages across repetitions.

Remediation Strategy	**Δ Load (mean)**	**Δ Thr [Mbps]**	**Δ SINR [dB]**	**Δ BLER**
Method 1: Small Cell Placement				
K1 (Geometric)	−0.153	−8.400	−30.300	−0.049
K2 (Hotspot)	−0.174	−3.400	−29.500	+0.021
K3 (Intermediate)	−0.105	−2.000	−29.600	+0.016
Method 2: Transmit Power				
30 dBm (baseline)	0.000	0.000	0.000	0.000
33 dBm	0.000	0.000	+2.900	0.000
36 dBm	0.000	0.000	+5.900	0.000
39 dBm	0.000	0.000	+8.900	0.000
60 dBm (theoretical)	−0.042	+0.600	+28.600	−0.006
Method 3: Bandwidth Scaling				
10 MHz (baseline)	0.000	0.000	0.000	0.000
15 MHz	−0.336	+1.000	−6.100	−0.008
20 MHz (artifact)	−0.648	−17.000	−6.500	+0.098

Key observations from delta values:

Method 1 (Small Cells) shows modest but consistent load relief. K2 (Hotspot) achieves the largest reduction (27.9% relative improvement), slightly better than K1 (-0.153) and K3 (-0.105). However, all three strategies incur throughput penalties (-2 to -8.4 Mbps), reflecting the load-balancing trade-off: capacity is shared between macro and small cell [11]. . SINR decreases significantly (-29 to -30 dB) due to reduced transmission power concentration; nonetheless, resulting SINR values (32-33 dB) remain acceptable [20].

Method 2 (Transmit Power) shows zero improvement for cell load and throughput across all realistic power levels (30–39 dBm), confirming the resource-limited nature of overload. Only SINR improves (+2.9 to +8.9 dB), which does not translate to capacity relief. The theoretical 60 dBm case shows minimal load reduction (-0.042), validating that even unrealistic power levels cannot overcome resource scarcity.

Method 3 (Bandwidth Scaling) achieves the strongest load relief. The 15 MHz configuration reduces load by -0.336 (40% improvement) while maintaining throughput (+1 Mbps gain). The 20 MHz result should be disregarded as a simulator artifact. Critically, Method 3's benefits come without the throughput penalties seen in Method 1.

5.1.1. Technical Performance Comparison

Cell Load Improvement: Figure 5.1 visualizes cell load improvement across all strategies. The most striking observation is the flat response for Method 2 (transmit power): all bars at 0 across the 30–39 dBm range, directly proving that power adjustment does not address capacity-limited overload. In contrast, Method 1 bars show modest negative deltas (-0.10 to -0.17), with K2 performing best. K2 shows smaller error bars compared to other methods, though the limited number of replicates constrains the strength of this statistical inference. Method 3 bars dominate, with 15 MHz achieving -0.336 load reduction – more than double compared to any other strategy. The horizontal target line shows that Method 3 alone achieves relief to the 80% threshold without additional measures.

Signal Quality and Throughput Trade-offs:

Figure 5.2 reveals the divergent SINR behavior across methods. Method 2 shows a clear upward trend (+2.9 to +8.9 dB as power increases), demonstrating excellent signal quality improvement – yet this improvement provides no congestion relief (as seen in the flat cell load response). Methods 1 and 3 show SINR degradation (-6 to -30 dB), reflecting the trade-off of spreading capacity or power. Critically, even with

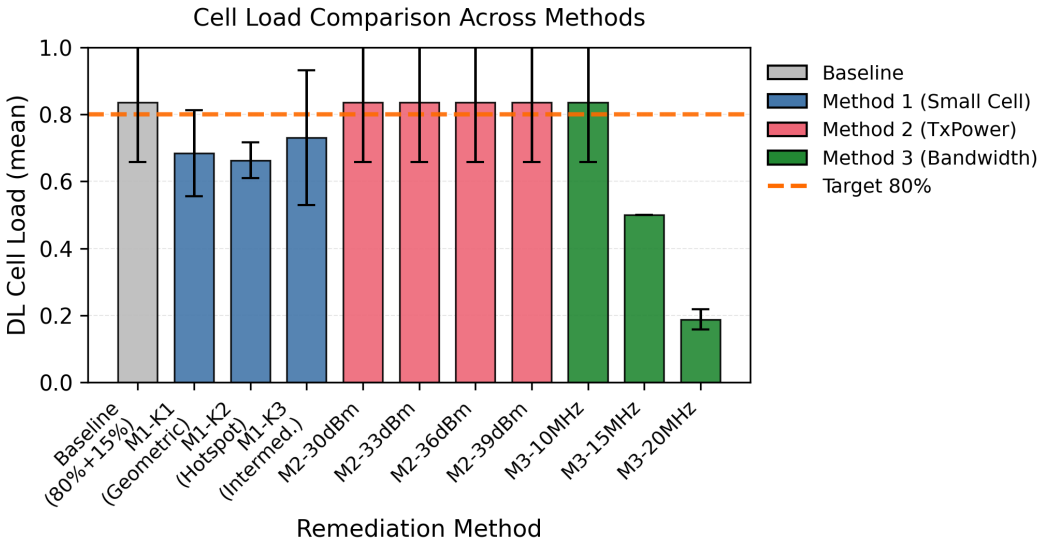

Figure 5.1.: Delta Cell load compared Methods Compared to Baseline

-30 dB SINR reduction, absolute SINR remains in the excellent range (32-33 dB for Method 1), so the degradation is not operationally problematic.

Figure 5.3 demonstrates the critical trade-off between capacity and throughput. Method 1 incurs throughput penalties (-2 to -8.4 Mbps), a consequence of splitting demand between macro and small cell [11]. Method 2 shows no throughput benefit. Method 3 at 15 MHz remarkably shows slight throughput gain (+1 Mbps), the only strategy to improve both load and throughput simultaneously.

Figure 5.4 synthesizes the load-throughput relationship, clearly showing Method 3 at 15 MHz as the only strategy achieving both load reduction AND throughput preservation. Method 1 strategies lie along a trade-off frontier (lower load, but reduced throughput). Method 2 sits at the baseline position (no improvement). This scatter plot makes visually clear why Method 3 is technically superior, despite practical barriers.

5.1.2. Deployment Feasibility and Practical Constraints

Method 1 (Small Cell Deployment):

Method 1 offers the most flexible deployment timeline. For long-term congestion, operators can install permanent small cells at known hotspots [11]. For temporary high-load events (concerts, sports matches), mobile base stations , such as COWs [40] can be deployed within hours. However, Method 1 requires significant capital expenditure

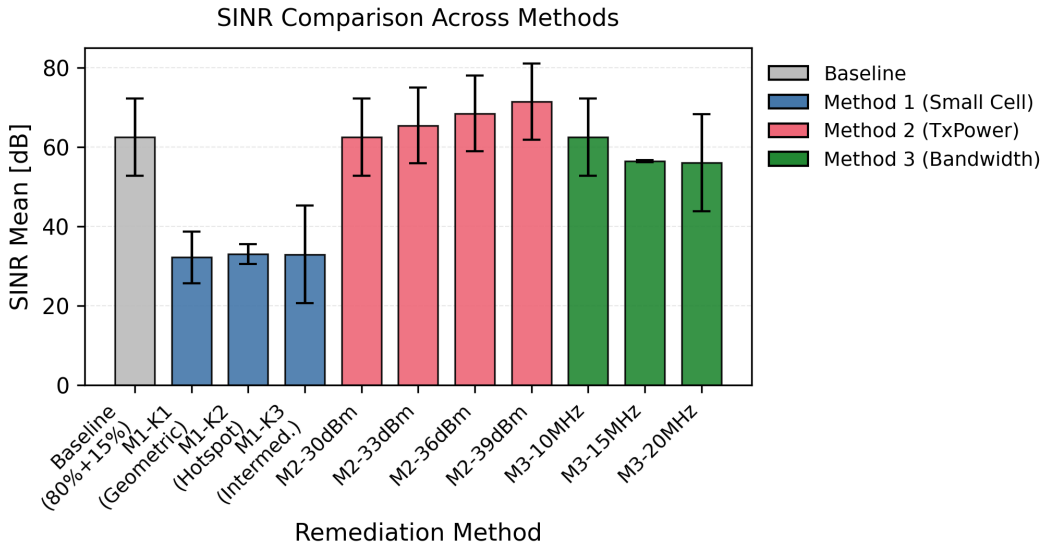

Figure 5.2.: Delta SINR Comparison. Shows SINR improvement (positive Δ) for Method 2 (transmit power) and SINR degradation (negative Δ) for Methods 1 and 3 due to load distribution or power density reduction.

for hardware acquisition and site preparation. Additionally, energy consumption of a mobile small cell (typical 1-2 kW) exceeds the cost of temporarily boosting existing cell parameters [9]. Method 1's load relief (-0.17 at best) is substantial but incomplete, leaving the system near the congestion threshold. Operators with existing mobile infrastructure inventory may find this method most economical for recurring events.

Method 2 (Transmit Power Adjustment):

This method proved ineffective for resource-limited overload, as demonstrated by zero cell load improvement despite significant SINR gains. While power adjustment can be implemented via software update (trivial cost and deployment time), it provides no benefit for the target scenario [15]. Method 2 is useful only for interference-limited or coverage-limited congestion, which is a different problem than what was examined here [15, Chapter 5][20]. This finding highlights a critical diagnostic lesson operators must distinguish congestion type before selecting remediation. Applying Method 2 to capacity-limited scenarios wastes operational effort.

Method 3 (Bandwidth Scaling):

Bandwidth scaling to 15 MHz achieves the strongest technical performance (-0.336 load reduction, 40% improvement) with stable throughput. However, obtaining additional licensed spectrum is expensive, requires regulatory approval, and involves negotiating

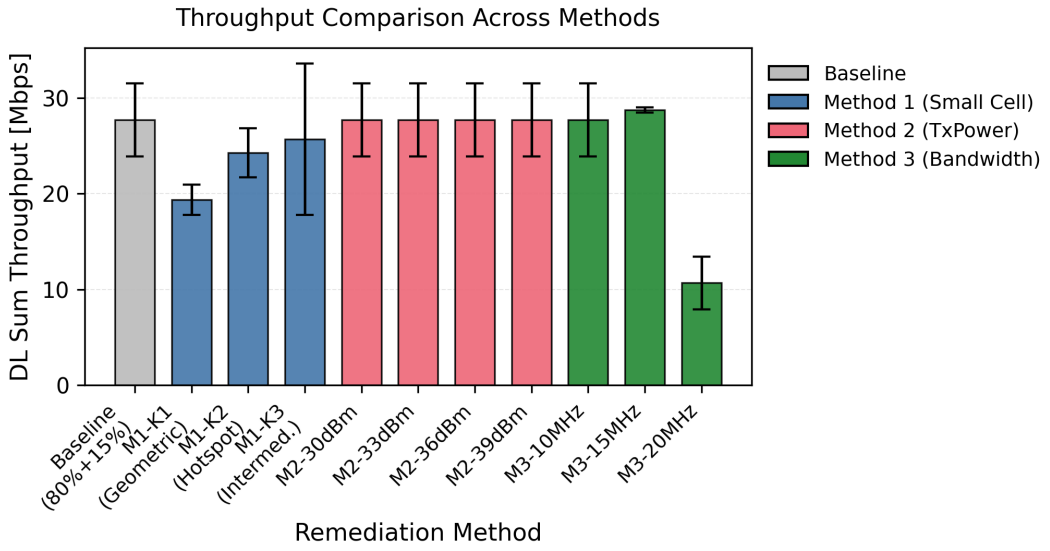

Figure 5.3.: Delta Throughput Comparison. Shows throughput change for each remediation. Method 1 incurs penalties (negative Δ) due to load sharing; Method 2 shows zero change; Method 3 shows stability at 15 MHz, collapse at 20 MHz.

with other operators for adjacent frequency bands [41]. The Implementation timescale is measured in weeks to months. Method 3 is suited for permanent capacity expansion rather than temporary event relief. In scenarios where temporary spectrum can be licensed (increasingly common in developed nations for major events), Method 3 becomes very useful.

Winner: Method 1-K2 (Practical) vs. Method 3 (Technical)

The evaluation reveals a clear distinction between technical optimality and practical feasibility.

Technically, Method 3 is superior: 40% load reduction with no throughput penalty represents the most efficient use of resources. If spectrum were freely available, bandwidth scaling would be the obvious choice for any operator.

Practically, Method 1-K2 (Hotspot Small Cell) is the winner for this high-load event scenario. It achieves 27.9% load relief (sufficient for most temporary situations), can be deployed in hours via mobile infrastructure, and leverages existing operator assets. While the throughput penalty (-3.4 Mbps) is non-trivial, the resulting 24.3 Mbps still meets typical mobile broadband requirements at events.

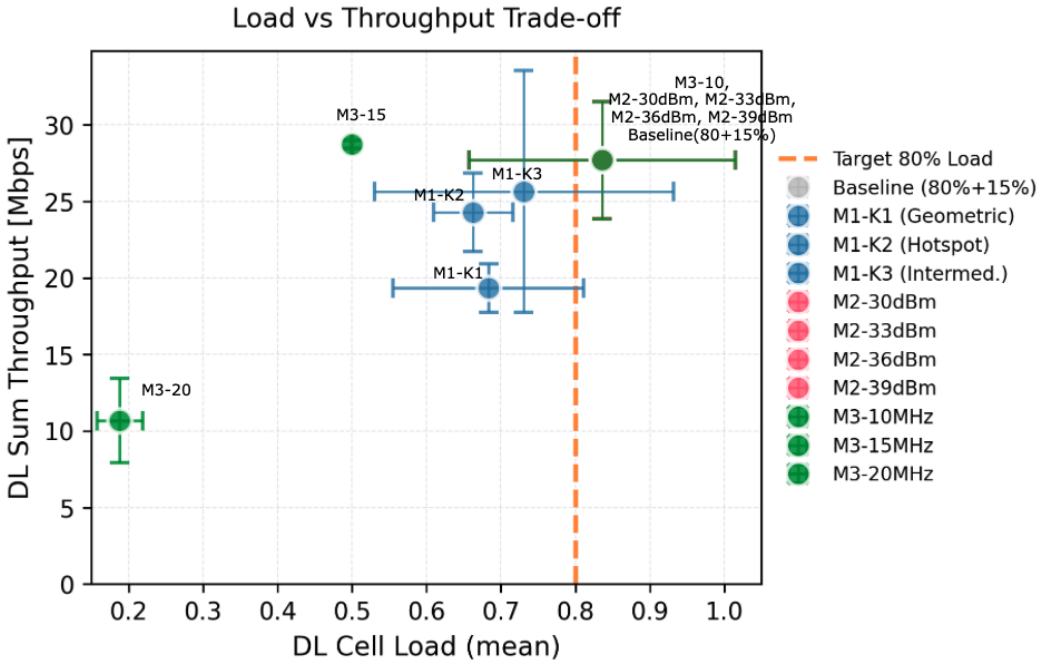

Figure 5.4.: Load-Throughput Trade-off Space. Scatter plot showing each remediation strategy positioned by cell load (x-axis, lower is better) vs. total throughput (y-axis, higher is better). Baseline at top-left (high load, good throughput); ideal remediation at bottom-right (low load, good throughput).

For long-term recurrent congestion at fixed locations, permanent small cell deployment becomes economically justified [11]. For temporary events, mobile small cells represent the pragmatic choice [40]. For permanent capacity expansion, Method 3 remains the long-term solution despite regulatory barriers.

5.2. Evaluation of the Simulation Approach

This study employed an ns-3-based network digital twin integrating realistic mobility (SUMO), ray-tracing propagation, and protocol-accurate 5G NR simulation. This section evaluates the strengths and limitations of this approach compared to traditional network simulators and analytical models.

5.2.1. Advantages of the Simulation Approach

The ns-3 Playground multi-tool environment employed in this thesis offers several key advantages over traditional network simulation approaches, particularly for evaluating spatially dependent remediation strategies in realistic scenarios.

Using Real Geographic Data: Traditional network simulators such as ns-2 [25], ns-3 [44] or OMNeT++ [46] provide protocol-level accuracy, but typically utilize abstract topologies and synthetic mobility patterns. In contrast, the approach presented here integrates real-world geographic data from OpenStreetMap, calibrated ray-tracing propagation models (Sionna), and realistic user mobility traces (SUMO). This combination enables spatially accurate analysis of localized congestion, which is critical for evaluating positioning-dependent remediation strategies such as small cell placement. The superiority of the K2 (Hotspot Overlay) positioning strategy over geometric/intermediate splits (K1, K3) would not have been discoverable without this spatial accuracy.

Accurate 5G Protocol Modeling: The ns-3 NR module implements the full 3GPP 5G NR protocol stack, including MAC scheduling, HARQ retransmissions, and RLC segmentation [44]. This level of detail captures interactions between layers that simplified analytical models cannot represent. For example, the observation that increasing transmit power improved SINR but did not reduce cell load (Method 2) required accurate modeling of PRB allocation dynamics which is a behavior that emerges only from full protocol simulation.

Reproducibility and Controlled Comparison: The simulation was configured to be deterministic in key aspects: ray-tracing propagation was pre-computed (not re-calculated per run), traffic patterns followed constant bitrate (not random arrivals), and SUMO vehicle movements used fixed seeds. This design choice, which is not a simulator limitation, enables reproducible baseline comparisons. Importantly, the study ran multiple replicates with different random seeds for scheduling and channel estimation to assess whether results were sensitive to these stochastic elements. Statistical analysis confirmed narrow 95% confidence intervals [30] across all replicates, showing that remediation method rankings were robust and not artifacts of particular random seeds. This combination of controlled propagation with stochastic scheduling variability provides a middle ground: reproducible scenarios without exaggerating the precision of results. In physical testbeds, this level of control is impossible as environmental variability, uncontrolled interference, and real-time traffic changes introduce confounding factors that make isolated method comparison extremely difficult.

Cost-Effectiveness and Rapid Iteration: Deploying physical testbeds for three remediation strategies across nine positioning variants would require substantial

infrastructure (multiple base stations, spectrum licenses, measurement equipment) and coordination with operators. The simulation approach enabled exploration of 27+ simulation variants (baseline + 9 Method 1 + 6 Method 2 + 3 Method 3 + replicates) at the cost of computational time alone. This cost advantage is particularly valuable during the exploratory phase of network planning, where multiple strategies must be evaluated before committing to field trials.

Extensibility to Additional Scenarios: The modular architecture of ns-3 Playground, as described in chapter 2 allows straightforward adaptation to other scenarios: different event types (indoor vs. outdoor), alternative network topologies (multi-cell, heterogeneous networks), or emerging 5G features (network slicing, beamforming, carrier aggregation). For example, extending this study to evaluate dual-connectivity or inter-cell coordination schemes would require primarily software configuration changes, not extensive software re-coding or hardware deployment.

5.2.2. Limitations and Scope of Validity

Despite its advantages, the simulation approach employed in this thesis has several important limitations that affect the applicability of results.

Fixed Traffic Model: The simulation used constant-bitrate UDP traffic (2.8 Mbps DL per UE) to represent mobile broadband demand. Real mobile traffic exhibits bursty patterns, for example, from web browsing, video streaming with adaptive bitrate, or background app updates, with high temporal variance. The fixed traffic model may overestimate the severity of congestion in some cases (real bursty traffic would have idle periods) or underestimate it in others (traffic bursts can overwhelm buffers). The 20 MHz bandwidth scenario showed this limitation: the traffic model could not saturate the expanded channel, leading to throughput collapse. This is a behavior which is not expected in real networks with adaptive traffic demand [15].

Scheduler Simplification: ns-3's proportional-fair scheduler approximates but does not fully replicate vendor-specific implementations used in commercial gNodeBs (e.g., Ericsson, Nokia, Huawei). Real schedulers incorporate proprietary optimizations, QoS prioritization for different traffic classes, and dynamic adaptation to interference conditions. While the proportional-fair scheduler is widely accepted for academic research, absolute KPI values (e.g., throughput, BLER) should be interpreted as comparative rather than predictive of operational network performance.

Single-Cell Focus: This study modeled a single macro cell with one optional small cell, deliberately isolating the high-load scenario for clarity. Real mobile networks are densely deployed multi-cell systems, where inter-cell interference, handover dynamics, and coordinated scheduling have a significant impact on overall performance. The

remediation strategies evaluated here might interact with neighboring cells in ways that are not captured by the simulation. For example, adding a small cell in a dense urban deployment can create interference with adjacent macro cells, potentially reducing the observed load relief benefit.

The decision to isolate the high-load scenario was justified by two main factors: it allowed the underlying mechanisms of each remediation strategy to be understood in a controlled environment and kept processing demands manageable. Simulating a multi-cell baseline would have required far more UEs and significantly greater computational resources, making the study less controllable. Future work could extend this approach to more complex, realistic multi-cell environments to study how remediation techniques perform under practical network conditions.

Applicability Statement: Despite these limitations, the simulation results provide valid insights for the target use case: temporary, localized high-load events with primarily mobile broadband traffic (such as concerts, sports matches, festivals). The observed KPI trends (cell load reduction from small cells and bandwidth scaling; ineffectiveness of power adjustment for resource-limited congestion) align with theoretical expectations from 5G NR capacity analysis [15] and operational reports from network operators.

The simulation approach is therefore well-suited for exploratory strategy evaluation and comparative analysis of remediation methods. However, findings should be validated through field trials before operational deployment, particularly for absolute performance predictions (e.g., "Method 1 will reduce load to exactly 0.66"). The relative rankings and trade-off insights (e.g., "Method 1 achieves moderate relief with immediate deployment; Method 3 achieves strong relief but requires regulatory approval") are robust and directly applicable to network planning decisions.

Future Extensions. The modular ns-3 Playground framework enables several promising extensions to address current limitations:

- Implementing realistic traffic models (HTTP/2, DASH video streaming) to capture bursty behavior.
- Extending to multi-cell scenarios with inter-cell interference coordination (ICIC) and handover modeling.
- Integrating machine learning-based schedulers to evaluate advanced 5G optimization techniques.
- Validating simulation results against field measurement campaigns from operator partners.

Such extensions would enhance the digital twin's predictive accuracy while maintaining its core advantages of reproducibility, spatial realism, and cost-effectiveness.

6. Conclusion and Outlook

6.1. Conclusion

This thesis evaluated three remediation strategies to address resource-limited high-load scenarios in mobile networks using an ns-3-based network digital twin. The results clearly identified Method 1 with positioning strategy K2 (Hotspot Overlay Small Cell Placement) as the most practical and effective option for the use case studied. Deploying a small cell directly at the traffic hotspot achieved a 27.9% reduction in macro cell load, maintained acceptable throughput, and preserved strong SINR values. This method outperformed transmission power adjustment (which did not improve capacity at all) and also had advantages over bandwidth scaling at 15 MHz, which—while technically effective—requires available spectrum and regulatory approval, making it less suitable for rapid response scenarios.

Mobile solutions such as Cells on Wheels (COWs) make Method 1 particularly well suited for short-notice, temporary events. Bandwidth scaling to 15 MHz emerged as a close second, and would likely be preferable in situations where additional spectrum can be obtained quickly. As with all technical solutions, there are trade-offs: deploying additional small cells increases both operational costs and maintenance needs. Moreover, if the baseline scenario involved multiple gNodeBs or represented a more complex, interference-prone environment, the results for each method might differ, and performance would likely be affected by inter-cell interactions.

The ns-3 Playground demonstrated several strengths as a simulation tool in this study. It enabled geographically realistic, reproducible, and protocol-accurate modeling and supported systematic comparison of remediation strategies in a controlled environment. Compared to traditional network simulators, the Playground's integration of realistic mobility and radio propagation was a key advantage, enabling spatially targeted analysis vital for real-world network planning. These findings show the value of digital twin simulations in the design and evaluation of practical network management solutions.

Ultimately, this work shows that careful digital twin-based modeling helps network operators make better decisions, especially when fast, data-driven responses to high-load events are needed.

6.2. Outlook

Several promising research directions remain open. Future studies should explore more complex remediation techniques, such as carrier aggregation, advanced beamforming, or AI-driven traffic management that adapts to dynamic user behavior and network load. Another valuable extension would be to expand the baseline scenario to multi-cell environments reflecting more realistic networks, with different cell types (macro, micro, pico), more diverse user mixes, and dynamic handover and interference patterns. Exploring scenarios with QoS-driven overload or mixed uplink and downlink congestion would also help evaluate the broad applicability of remediation strategies.

There is also substantial value in applying the ns-3 Playground to alternative use cases: event types with different mobility patterns, networks with automated resource allocation, or studies involving real operator data for model calibration. These directions, along with more realistic and dynamic channel models (such as including fast fading or evolving propagation conditions), would make simulation studies even more robust.

Additionally, future research could investigate using realistic mobility traces (such as SUMO trajectories) as direct input for base station placement optimization, enabling RAN planning to adapt proactively to actual movement patterns rather than static assumptions.

The continued development and refinement of digital twin models, combined with field validation, will be the key to supporting future mobile networks that are reliable, resource-efficient and able to meet the challenges of tomorrow's connected society. To conclude, the findings here contribute another piece to the broader field of digital twin simulations for network planning. As research continues and tools evolve, simulation-based insights will play an increasing role in developing realistic, effective solutions to complex network challenges.

7. References

[1] 3GPP. *5G Performance Measurements; KPI for PRB Usage (Total DL PRB Usage)*. Tech. rep. Version 16.10.0. Section 5.1.1.2.1 defines DL PRB Usage KPI as a measure of radio physical layer load; Accessed: 2025-10-21. 3GPP, 2021. URL: https://www.etsi.org/deliver/etsi_ts/128500_128599/128552/16.10.00_60/ts_128552v161000p.pdf.

[2] 3GPP. *5G System Overview*. Accessed: 2025-10-22. 2022. URL: https://www.3gpp.org/technologies/5g-system-overview.

[3] 3GPP. *NG-RAN; Architecture description*. Technical Specification TS 38.401, v18.1.0, Release 18. Accessed: 2025-10-22. 3rd Generation Partnership Project, 2025. URL: https://www.3gpp.org/ftp/Specs/archive/38_series/38.401/38401-18.1.0.zip.

[4] 3GPP. *NR; NR and NG-RAN Overall description; Stage-2*. Technical Specification TS 38.300, v16.2.0, Release 16. Accessed: 2025-10-22. 3rd Generation Partnership Project, 2022. URL: https://www.3gpp.org/ftp/Specs/archive/38_series/38.300/38300-16.2.0.zip.

[5] 3GPP. *NR; Study on channel model for frequencies from 0.5 to 100 GHz*. Technical Specification TS 38.901, v16.1.0, Release 16. Accessed: 2025-10-22. 3rd Generation Partnership Project, 2020. URL: https://www.3gpp.org/ftp/Specs/archive/38_series/38.901/.

[6] 3GPP and ETSI. *3GPP TS 38.211 V16.3.0 Release 16: NR; Physical channels and modulation*. Tech. rep. TS 138 211 v16.3.0. Section 5.1: Modulation schemes; Section 6.3.1: Physical uplink shared channel; Accessed: 2025-10-21. ETSI, Nov. 2020.

[7] 5G Networks. *Spectral Efficiency : 5G-NR and 4G-LTE compared*. Accessed: 2025-10-21. 2021. URL: https://www.5g-networks.net/spectral-efficiency-5g-nr-and-4g-lte-compared/.

[8] Alyssa Lamberti. *How to Detect, Fix & Reduce Network Overload*. Accessed: 2025-10-19. Obkio. 2023. URL: https://obkio.com/blog/reducing-network-overload/.

[9] Imran Ashraf, Federico Boccardi, and Lester Ho. "SLEEP mode techniques for small cell deployments." In: *IEEE Communications Magazine* 49.8 (2011), pp. 72–79. DOI: 10.1109/MCOM.2011.5978418.

[10] C. Bektas et al. “Rapid Network Planning of Temporary Private 5G Campus Networks.” In: *2021 IEEE 94th Vehicular Technology Conference (VTC-Fall)*. 2021. DOI: 10.1109/VTC2021-Fall52928.2021.9625234.

[11] Zubin Bharucha et al. “Small Cell Deployments: Recent Advances and Research Challenges.” In: *IEEE Communications Magazine* 51.12 (Nov. 2013), pp. 98–106.

[12] Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit. *Welche Frequenzen nutzen die 5G-Netze?* FAQ; Accessed: 2025-10-30. 2019. URL: https://www.bundesumweltministerium.de/faq/welche-frequenzen-nutzen-die-5g-netze.

[13] Bundesnetzagentur. *Administrative Rules for Spectrum Assignments for Local Spectrum Usages in the 3700-3800 MHz Band*. Section 1.2.1 (p. 12): No in-block EIRP limits; Sections 1.2.2-1.2.3 (p. 13–14): Out-of-block emission specifications; Accessed: 2025-10-27. May 2023. URL: https://www.bundesnetzagentur.de/SharedDocs/Downloads/EN/Areas/Telecommunications/Companies/TelecomRegulation/FrequencyManagement/FrequencyAssignment/LocalBroadband3,7GHz.pdf.

[14] Javier Campos. *Understanding the 5G NR Physical Layer*. Accessed: 2025-10-21. Nov. 2017. URL: https://www.3g4g.co.uk/5G/5Gtech_video0022_01.pdf.

[15] Erik Dahlman, Stefan Parkvall, and Johan Skold. *5G NR: The Next Generation Wireless Access Technology*. 1st ed. Chapters 2–6 and 8 (p. 13–105, 129–143): Chapter 2 (IMT overview); Chapter 3 (5G spectrum and frequency bands); Chapter 4 (LTE overview); Chapter 5 (NR overview, physical layer design, modulation schemes, MCS selection); Chapter 6 (NR architecture); Chapter 8 (NR physical layer in detail, modulation, LDPC coding, scheduling, transmission procedures). Academic Press, 2018. ISBN: 978-0-12-814323-0.

[16] Pathum Dilshan. *Telecom network optimization and user experience discussion [LinkedIn post]*. Accessed: 2025-10-20. LinkedIn. Mar. 2025. URL: https://www.linkedin.com/posts/pathum-dilshan_telecom-networkoptimization-userexperience-activity-7208317449006374913-NN-j.

[17] EUROGATE Group. *Connected Container Terminals: Telekom Builds 5G Campus Networks for EUROGATE*. Accessed: 2025-10-24. EUROGATE GmbH & Co. KGaA, KG. 2024. URL: https://www1.eurogate.de/en/connected-container-terminals-telekom-builds-5g-campus-networks-for-eurogate/.

[18] Federal Communications Commission. *Broadband Speed Guide*. Tech. rep. Last reviewed August 19, 2019; Accessed: 2025-10-21. Washington, DC: Federal Communications Commission, Consumer and Governmental Affairs Bureau, 2019. URL: https://archive.legmt.gov/content/Committees/Interim/2019-2020/Economic-Affairs/Meetings/Jan-2020/FCCbroadband-speed.pdf.

[19] Roger L. Freeman. *Fundamentals of Telecommunications*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2005. ISBN: 9780471710459.

[20] Andrea Goldsmith. *Wireless Communications*. Chapter 2 (p. 25–48): Path loss and shadowing, dB representations, free-space path loss; Chapter 4 (p. 93–96): SINR definition, channel capacity, modulation dependent throughput. Cambridge, UK: Cambridge University Press, 2005. ISBN: 978-0-521-83716-3.

[21] Jakob Hoydis et al. *Sionna: An Open-Source Library for Next-Generation Physical Layer Research*. Accessed: 2025-10-19. 2023. arXiv: 2203.11854 [cs.IT]. URL: https://arxiv.org/abs/2203.11854.

[22] Linbo Hui et al. "Digital Twin for Networking: A Data-Driven Performance Modeling Perspective." In: *IEEE Network* 37.3 (2023), pp. 202–209. DOI: 10.1109/MNET.119.2200080.

[23] IDC-Online. *Cellphone Signals & Cell Splitting*. Accessed: 2025-10-19. IDC-Online. URL: https://www.idc-online.com/technical_references/pdfs/electrical_engineering/Cellphone_signals_and_Cell_splitting.pdf.

[24] International Telecommunication Union, Telecommunication Standardization Sector (ITU-T). *Recommendation ITU-T E.417: Network Management for IP-based Networks*. ITU-T Recommendation E.417. Section 6.2.4: Definition of Overload Conditions. Geneva, Switzerland: International Telecommunication Union, 2005.

[25] Teerawat Issariyakul and Ekram Hossain. "Introduction to Network Simulator 2 (NS2)." In: *Introduction to Network Simulator NS2*. Boston, MA: Springer US, 2009, pp. 1–18. ISBN: 978-0-387-71760-9. DOI: 10.1007/978-0-387-71760-9_2. URL: https://doi.org/10.1007/978-0-387-71760-9_2.

[26] Aymen Jaziri, Ridha Nasri, and Tijani Chahed. "Offloading traffic hotspots using moving small cells." In: *Proceedings of the 2016 IEEE International Conference on Communications (ICC)*. May 2016, pp. 1–6. DOI: 10.1109/ICC.2016.7511566.

[27] Khushboo Kalyani. *BLER: A Critical Parameter in Cellular Receiver Performance*. Explains BLER definition, relationship to SNR/SINR, MCS, and throughput; Accessed: 2025-10-21. 2021. URL: https://www.5gtechnologyworld.com/bler-a-critical-parameter-in-cellular-receiver-performance/.

[28] Anders Landström, Håkan Jonsson, and Anders Simonsson. "Voronoi-Based ISD and Site Density Characteristics for Mobile Networks." In: *2012 IEEE Vehicular Technology Conference (VTC Fall)*. 2012, pp. 1–5. DOI: 10.1109/VTCFall.2012.6399149.

[29] Pablo Alvarez Lopez et al. "Microscopic Traffic Simulation using SUMO." In: *The 21st IEEE International Conference on Intelligent Transportation Systems*. Accessed: 2025-10-19. IEEE, 2018. URL: https://elib.dlr.de/124092/.

[30] Douglas C. Montgomery and Ronald E. Walpole. *Probability and Statistics for Engineers and Scientists*. 9th. Pearson Education, 2012. ISBN: 978-0-321-62911-1.

[31] E. Nan, Xiaoli Chu, and Jie Zhang. "Mobile Small-Cell Deployment Strategy for Hot Spot in Existing Heterogeneous Networks." In: *2015 IEEE Globecom Workshops (GC Wkshps)*. 2015, pp. 1–6. DOI: 10.1109/GLOCOMW.2015.7414060.

[32] Jorge Navarro-Ortiz et al. "A Survey on 5G Usage Scenarios and Traffic Models." In: *IEEE Communications Surveys & Tutorials* 22.2 (2020), pp. 905–929. DOI: 10.1109/COMST.2020.2971781.

[33] Ding Nian et al. *A Comprehensive Review on Automatic Mobile Robots: Applications, Perception, Communication and Control*. Tech. rep. Accessed: 2025-10-19. TechRxiv, Apr. 2022. DOI: 10.36227/techrxiv.19619880.v1. URL: https://www.researchgate.net/publication/360197839.

[34] V. B. Nikam et al. "Optimal Positioning of Small Cells for Coverage and Cost Efficient 5G Network Deployment: A Smart Simulated Annealing Approach." In: *2020 IEEE 3rd 5G World Forum (5GWF)*. 2020, pp. 454–459. DOI: 10.1109/5GWF49715.2020.9221257.

[35] NRExplained. *Bandwidth – NRExplained*. Accessed: 2025-10-06. URL: https://www.nrexplained.com/bandwidth.

[36] NVIDIA Labs. *Sionna Ray Tracing (RT) Module*. Documentation for Sionna RT module used for radio propagation modeling; Accessed: 2025-10-19. 2024. URL: https://nvlabs.github.io/sionna/rt/index.html.

[37] Nasr Obaid and Andreas Czylwik. "An Efficient Macrocell Power Control Algorithm for Enhancing Dense Picocell Deployments." In: *European Wireless 2014; 20th European Wireless Conference*. 2014, pp. 1–6.

[38] José Portela and Marcelo Alencar. "Cellular Coverage Map as a Voronoi Diagram." In: *Journal of Communication and Information Systems* 23 (Apr. 2008). DOI: 10.14209/jcis.2008.3.

[39] Private Networks Technology Blog. *Deutsche Telekom's Campus Network L for Industry: A Modular Solution for Private 5G Connectivity*. Accessed: 2025-10-24. Private Networks Technology. Apr. 2024. URL: https://blog.privatenetworks.technology/2024/04/deutsche-telekoms-campus-network-l-for.html.

[40] Ladan Rabieekenari, Kamran Sayrafian, and John S. Baras. "Autonomous relocation strategies for cells on wheels in environments with prohibited areas." In: *2017 IEEE International Conference on Communications (ICC)*. 2017, pp. 1–6. DOI: 10.1109/ICC.2017.7997091.

[41] Male Srinivasa Rao et al. "5G Deployment Challenges and Solutions: A Comprehensive Review." In: *2024 9th International Conference on Communication and Electronics Systems (ICCES)*. 2024, pp. 1–6. DOI: 10.1109/ICCES63552.2024.10859706.

[42] RF Wireless World. *4G LTE PRB Utilization: Formula, Calculator, and Definition*. Standard PRB utilization formula; 80–85% threshold; Accessed: 2025-10-21. 2024. URL: https://www.rfwireless-world.com/terminology/4g-lte-prb-utilization.

[43] RF Wireless World. *5G NR PRB Utilization and Resource Allocation*. Confirms PRB utilization formula applies to 5G NR; Accessed: 2025-10-21. 2024. URL: https://www.rfwireless-world.com/terminology/5g-nr-prb-utilization-and-resource-allocation.

[44] George Riley and Thomas Henderson. "The ns-3 Network Simulator." In: *Modeling and Tools for Network Simulation*. Vol. 5933. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 15–34. DOI: 10.1007/978-3-642-12331-3_2.

[45] Jun Su et al. "5G multi-tier radio access network planning based on voronoi diagram." In: *Measurement* 192 (2022), p. 110814. ISSN: 0263-2241. DOI: 10.1016/j.measurement.2022.110814. URL: https://www.sciencedirect.com/science/article/pii/S0263224122001117.

[46] Andras Varga. "OMNeT++." In: *Modeling and Tools for Network Simulation*. Ed. by Klaus Wehrle, Mesut Güneş, and James Gross. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 35–59. ISBN: 978-3-642-12331-3. DOI: 10.1007/978-3-642-12331-3_3. URL: https://doi.org/10.1007/978-3-642-12331-3_3.

[47] Wen Wu, Xinyu Huang, and Tom H. Luan. *AI-Native Network Digital Twin for Intelligent Network Management in 6G*. Accessed: 2025-10-20. 2024. arXiv: 2410.01584 [cs.NI]. URL: https://arxiv.org/abs/2410.01584.

A. Appendix

A.1. Note on Mobile UE Calibration in SUMO Mobility

During simulation setup using the SUMO mobility model, it was observed that the number of mobile UEs requested for trajectory generation did not always match the number of trajectories successfully created by SUMO. For example, requesting 60 mobile UEs typically resulted in only 57 valid SUMO trajectories due to constraints in route availability and network configuration. To ensure accurate load calibration and reproducibility, the simulation configuration was adjusted to use the actual number of mobile UEs generated rather than the requested number. This correction was consistently applied across all relevant simulation runs.

A.2. Static UE Positions in Mixed Mobility Scenario

To assess how remediation methods perform with different UE distributions, a sensitivity analysis was run using 20 static and 40 mobile UEs. The static UEs were placed at

Table A.1.: Static UE positions for mixed mobility sensitivity analysis.

Deployment Region	**UE Count**	**Position(s)** $[x, y, z]$ (m)
Macro vicinity	4	$[20, 61, 1.5]$, $[24, 62, 1.5]$ $[21, 65, 1.5]$, $[25, 59, 1.5]$
Hotspot vicinity	5	$[50, 32, 1.5]$, $[52, 31, 1.5]$, $[54, 34, 1.5]$ $[49, 36, 1.5]$, $[56, 33, 1.5]$
Facility edges	4	$[5, 5, 1.5]$, $[75, 65, 1.5]$ $[75, 5, 1.5]$, $[5, 65, 1.5]$
Mid-cell distribution	6	$[40, 60, 1.5]$, $[35, 40, 1.5]$, $[60, 50, 1.5]$ $[30, 25, 1.5]$, $[45, 15, 1.5]$, $[65, 25, 1.5]$
Additional placement	1	$[15, 30, 1.5]$

strategic locations, as described in A.1 to test coverage gaps and load concentrations: four near the macro cell base station, five clustered around the hotspot region, four

at the facility edges, six distributed around the middle cell area, and one additional position. This arrangement intentionally creates areas of varying signal strength and competition for resources, allowing us to observe how the remediation strategies adapt to non-uniform load patterns.

The 40 mobile UEs follow the standard SUMO-based mobility patterns described in Section 3.2, ensuring that the sensitivity analysis combines realistic user movement with controlled static anchor points to test robustness across diverse deployment scenarios.