Accepted for presentation in: IEEE Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM), Paris, France, October 2026.

Priority-Driven and Energy-Aware Cellular Communication for Mixed Critical Infrastructures

David Ronschka, Dennis Overbeck, Pascal Jörke and Christian Wietfeld
Chair of Communication Networks, TU Dortmund University, Germany
Email: {David.Ronschka, Dennis.Overbeck, Pascal.Joerke, Christian.Wietfeld}@tu-dortmund.de

***Abstract*—Future smart energy infrastructures will increasingly rely on public cellular networks to support communication for mixed critical infrastructure services. These applications impose heterogeneous communication requirements and must remain reliable even under high network load and constrained energy availability. This paper presents a framework for reliable low-latency cellular communication with energy awareness. It is evaluated in Network Simulator 3 (ns-3) in a mixed-criticality scenario that combines energy and mobility applications with consumer applications. The proposed approach combines Network Slicing for guaranteed resource isolation with predictive proactive grants for periodic mission-critical traffic, based on realistically emulated IEC 61850 communication patterns. Base stations are among the most power-intensive components of the cellular infrastructure, yet critical infrastructure communication depends fundamentally on their continued operation. Under energy-constrained conditions such as blackout scenarios, phased component deactivation offers a viable strategy to reduce their consumption without abandoning critical communication services. A novel base station energy model formalizes this deactivation process, with Network Slicing ensuring that dedicated resources remain guaranteed for the highest-priority grid devices throughout. Both implementations are released as open-source ns-3 modules. Simulation results show that end-to-end latencies below 5 ms can be achieved while consistently exceeding the packet delivery ratio target of 99.9999 % through guaranteed resource allocation. Progressive base station deactivation reduces power consumption by up to 77 % in the most critical scenario, while Network Slicing preserves full communication functionality for the highest-priority grid devices throughout.**

***Index Terms*—Network Slicing, gNB Energy Model, URLLC, ns-3, Simulation**

I. Introduction

THE ongoing digitalization of critical infrastructures is increasing their dependence on shared cellular networks for reliable communication, with Smart Grids representing one prominent example. This evolution introduces new communication demands, driven by the increasing integration of distributed energy resources, intelligent substations, electric vehicles, and large-scale sensor deployments. To support these developments, communication networks must provide not only wide-area connectivity and scalability, but also highly reliable and low-latency transmission capabilities for critical infrastructure operations.

Public cellular networks provide a promising foundation for addressing these requirements. The continuous evolution of mobile networks towards increasingly capable and globally available infrastructure makes them a promising communication platform for critical infrastructures, including Smart Grids, avoiding the need for dedicated, application-specific networks. At the same time, integrating critical infrastructure communication into such shared public infrastructure introduces substantial challenges, particularly due to the heterogeneous criticality of applications and their strict Quality of Service (QoS) requirements.

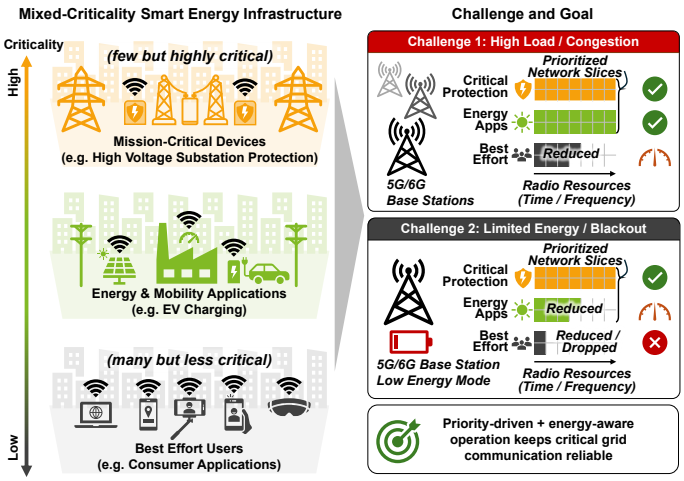

Fig. 1: Mixed-criticality smart infrastructure and priority-driven communication over cellular networks under congestion and energy-constrained conditions.

Critical infrastructure functions, such as protection and automation mechanisms in High Voltage (HV) infrastructure, require ultra-low latency and extremely high communication reliability. Highly critical components such as HV substations occur only infrequently within the overall network, whereas non-critical best-effort devices are widely distributed. This mixed-criticality structure requires prioritization and resource allocation strategies within the Radio Access Network (RAN) that can simultaneously serve devices with fundamentally different communication requirements, as illustrated in Fig. 1. For critical applications such as HV protection systems, latency targets of up to 5 ms [1] and reliability requirements of 99.9999 % [2] must be fulfilled to ensure stable and secure grid operation. Mechanisms such as network slicing, 5G Quality of Service Identifier (5QI) differentiation, and Ultra-Reliable Low-Latency Communication (URLLC), introduced with 5G NR and continuously refined in subsequent releases, provide an important basis for meeting these requirements, but

©2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, including reprinting/republishing this material for advertising or promotional purposes, collecting new collected works for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

their operation under energy-constrained conditions remains insufficiently explored.

The energy consumption of mobile networks constitutes a further factor that must be considered alongside communication performance. Modern cellular base stations are among the most power-intensive components of the cellular infrastructure, with typical power values for complex cells reaching around 8000 W per site [3]. As critical infrastructures such as Smart Grids are inherently linked to energy management and resilience, the cellular infrastructure that supports their communication becomes an energy-relevant component of the overall system. This dependency becomes critical during power shortages or partial blackouts, where grid-side energy availability may be limited to battery backup. Under such conditions, reducing base station power consumption through phased component deactivation is a viable strategy, but only if the communication reliability of the highest-priority grid devices can be guaranteed throughout the process. Network Slicing provides precisely this guarantee by ensuring dedicated, isolated resources for mission-critical traffic independently of the operational state of the base station.

This paper investigates priority-driven and energy-aware cellular communication for mixed critical infrastructures, using Smart Grid communication as a representative and demanding evaluation scenario. A novel base station energy model enables phased component deactivation under energy-constrained conditions, while Network Slicing and predictive proactive grants ensure that communication reliability and latency requirements for mission-critical IEC 61850 traffic are preserved under both nominal and degraded network conditions.

II. Related Work

This section reviews existing work across two research areas relevant to this paper, namely, uplink scheduling for URLLC services in 5G networks with prioritization utilizing network slices and base station power consumption modeling.

1) Predictive Dynamic Network Slices:
Regarding reliability in grant-free access, [4] proposes a framework for URLLC User Equipments (UEs) leveraging Autoregressive Integrated Moving Average (ARIMA)-based prediction of future network load to enable adaptive resource allocation for low latency constraints. [5] uses a Deep Reinforcement Learning (DRL)-based uplink scheduler to eliminate latency-inducing scheduling requests, while prior work from our group investigated proactive, prediction-based resource management for 5G uplink slicing [6] using Long Short-Term Memory (LSTM) models. [7] employs Medium Access Control (MAC)-level predictions of buffer status reports to reduce head-of-line delay for vehicular UEs using bidirectional LSTM within the simulation framework ns-3. Other works successfully incorporate Random Forest models for data rate prediction outlining its applicability while maintaining low complexity in model training [8]. Additionally, a recent ns-3 based study [9] presents an end-to-end evaluation framework demonstrating that dedicated QoS flows can isolate 4K video traffic from best-effort background loads. However, the slicing here relies on a standard Proportional Fair scheduling, without 5QI-based traffic class differentiation, URLLC-specific latency guarantees, or any mechanism for coordinated resource shutdown. All those aspects are central to the blackout scenario addressed in this work.

2) Energy Modeling of Mobile Radio Networks:
A central challenge in energy-saving for mobile radio networks is the energy consumption of base stations, which account for the largest share of RAN power usage [10]. Thus, research in this area falls into the categories of modeling the base station energy consumption as well as developing strategies to reduce it. The authors of [11] propose a data-driven power consumption model derived from real-world measurements, where a technology- and frequency-specific linear regression model is shown to outperform neural network approaches. However, the model is derived from a limited set of hardware configurations and measurements aggregated at hourly granularity, whereas this work targets optimization at the millisecond timescale. At a larger scale, [12] combines ML-based and expert-driven models to predict network energy consumption and UE throughput under carrier shutdown configurations. The results were simulated using a custom agent-based model with focus on carrier shutdown dynamics and UE handovers, thus additional effects e.g. in scheduling operations were neglected in this perspective. Focusing on energy-saving of base stations, [13] implements a network-state-aware base station switching mechanism directly in ns-3, toggling power states based on UE connectivity and traffic activity, though the scope is limited to mmWave scenarios. [14] takes a learning-based approach, formulating carrier shutdown as a deep reinforcement learning problem where actions cover not only carrier deactivation but also the reallocation of traffic and users to remaining active carriers and thus optimizing Quality of Experience (QoE) for users. For the energy modeling of base stations within ns-3 itself, [15] provides a basic energy module for 5G mmWave gNBs, though with a focus on mmWave and without a detailed, measurement-based characterization of power consumption. In this work, we introduce an enhanced network state modeling in ns-3 to sub-6 GHz deployments in alignment with O-RAN working group specifications [16], and instantiate the energy efficiency modeling principles of [17] within a controlled simulation environment. Whereas, literature focuses on general improvements in energy consumption of mobile radio networks, we couple the energy saving strategies to prioritized, low latency network slices in the event of a grid blackout and subsequent re-establishment of normal operation.

Unlike prior work, we publicly release all ns-3 extensions developed in this study, providing the research community with a reproducible, openly available platform for energy-aware slicing evaluation under varying network conditions.

III. System Architecture

The proposed system architecture describes the ns-3-based simulation framework used to implement and evaluate the considered cellular communication mechanisms for mixed

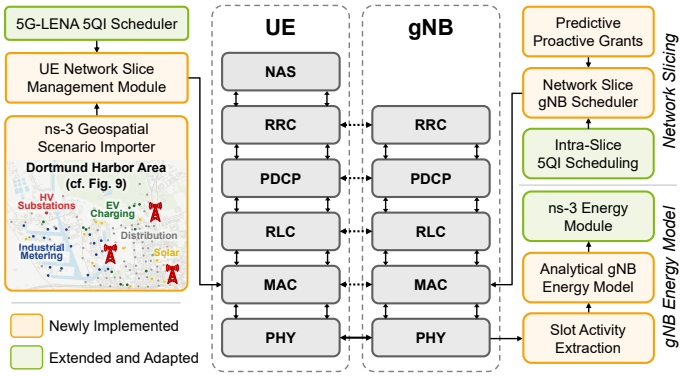

Fig. 2: Architecture of the proposed simulation framework. The figure shows the extended UE and gNB protocol stacks in ns-3 and the additional modules implemented for network slicing, predictive proactive grants, and RAN energy modeling.

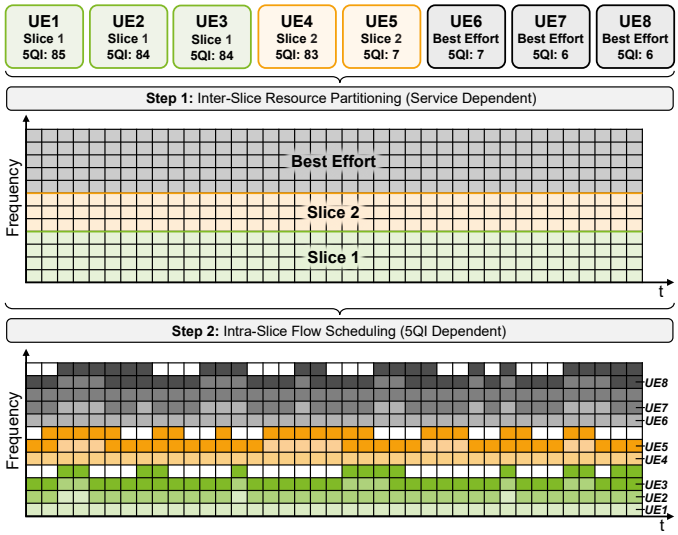

Fig. 3: Two-stage network slicing procedure. The inter-slice stage allocates resources across slices, while the intra-slice stage schedules UEs within each slice using 5QI-aware prioritization.

critical infrastructures in a representative Smart Grid scenario. Fig. 2 provides an overview of the implemented extensions and their integration into the simulated UE–gNB protocol stack.

The architecture consists of three main components. First, a Network Slicing mechanism is implemented at the radio access level to separate traffic flows with different service requirements. A custom OFDMA-based Network Slice Scheduler at the gNB MAC layer enables slice-specific Physical Resource Block (PRB) reservations combined with intra-slice prioritization based on 5QI-aware Proportional Fair (PF) scheduling. Second, the scheduling framework is extended by predictive proactive grants that use a trained prediction model to anticipate recurring uplink transmissions from mission-critical devices based on IEC 61850 traffic patterns. Third, an analytical gNB energy model based on [17] is integrated into ns-3 to estimate base station energy consumption from both static configuration parameters and dynamic operating conditions. The model considers factors such as transmit power, antenna configuration, system bandwidth, network load, and component shutdown states.

In addition, the simulation environment incorporates real-world geospatial data from OpenStreetMap (OSM) [18] and cellular base station location data from [19]. Infrastructure assets and base stations are placed according to actual deployment locations within the considered area, providing a realistic basis for evaluating the Network Slicing, predictive scheduling, and energy-aware mechanisms considered in this work.

A. Implementation of Network Slicing in ns-3

The network slicing functionality is implemented as an extension of the MAC-layer scheduler in the 5G-LENA framework and is publicly released as an open-source contribution[1]. The scheduler introduces slice-aware resource allocation while preserving the existing 5QI-based prioritization mechanisms of the simulator. For this purpose, the implementation employs a hierarchical two-stage design that provides guaranteed resource isolation across slices while retaining full prioritization flexibility within each slice. The first stage determines the resource budget of each slice, whereas the second stage selects the concrete UEs and flows within the assigned slice resources.

The scheduling principle is illustrated in Fig. 3. In the first step, the available frequency-domain resource grid is partitioned between the configured slices and a best-effort resource pool. This partitioning is performed at slice level and defines which part of the available radio resources may be used by each slice during the current scheduling interval. In the second step, the resources assigned to each slice are filled with the corresponding UEs according to the intra-slice scheduling decision. The resulting allocation therefore reflects both the configured slice structure and the service requirements of the active flows.

Each slice is assigned an identifier and a reservation share defining its guaranteed fraction of available Resource Block Groups (RBGs), while UEs are associated with the corresponding slice according to their service class. The reservation shares are derived from the traffic characteristics of the served devices. Since IEC 61850 devices typically exhibit deterministic and largely periodic transmission patterns, the required resource share of each slice can be estimated reliably and configured accordingly. The shares therefore remain fixed during operation, preserving strict resource isolation and ensuring that the capacity reserved for mission-critical services remains available when required.

The intra-slice scheduler operates within the resource budget assigned by the inter-slice controller. Within this budget, UEs and logical channels are scheduled using the proportional-fair mechanism of the 5G-LENA 5QI scheduler, balancing channel

[1]https://github.com/tudo-cni/ns3-network-slicing

utilization and long-term fairness among users. In addition to the proportional-fair component, the scheduler retains the standard-compliant 5QI-based QoS handling of 5G-LENA. This hierarchical design separates resource isolation from service differentiation, allowing critical services to receive guaranteed resources while retaining flexible prioritization within each slice.

B. ML-based Predictive Proactive Scheduling

Protection and control devices in smart grid infrastructure depend on timely uplink transmissions to report fault events and state changes under strict latency constraints. To eliminate the overhead introduced by the conventional scheduling request procedure for such devices, the proposed architecture employs a machine learning-based proactive grant allocation mechanism that predicts upcoming transmissions and pre-allocates radio resources, enabling the UE to transmit immediately upon packet arrival.

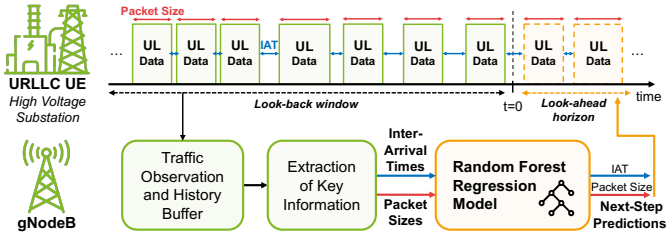

Fig. 4: Concept of the predictive proactive grant mechanism using previously observed uplink packet sizes and inter-arrival times to estimate future transmission activity within a configurable look-ahead horizon.

The basic concept is illustrated in Fig. 4. A history buffer at the gNB observes previous uplink transmissions of a selected URLLC UE. From this look-back window, scheduling-relevant features such as packet sizes and inter-arrival times are extracted and provided to a predictive model. The model estimates whether additional uplink packets are expected within a configurable look-ahead horizon. If future traffic is predicted, the scheduler can prepare uplink resources in advance and assign a proactive grant to the corresponding UE.

The traffic patterns used for this mechanism are derived from the IEC 61850-based substation scenario shown in Fig. 5, in which the Remote Terminal Unit (RTU) serves as the communication gateway to the remote Supervisory Control and Data Acquisition (SCADA) system via the cellular uplink. The architecture is emulated using the open-source *libIEC61850* library [20], with more than 1000 individual scenarios executed across systematic parameter sweeps of transmission intervals, message sizes, and timing jitter, accumulating over ten hours of network traffic. From the recorded traces, inter-arrival time and packet size are extracted as features, capturing the essential timing and load characteristics of the traffic without requiring application-level payload inspection. To account for event-driven behavior, selected runs include a critical occasion transition, in which reporting frequency increases under critical system conditions. This transition represents realistic emergency or fault situations, where selected devices, particularly protection-related infrastructure components, temporarily increase their communication activity. The dataset is partitioned into training (70 %), validation (15 %), and test (15 %) subsets and uses the inter-arrival time and packet size of the subsequent transmission as prediction targets for proactive grant allocation, with the test set reserved for final evaluation.

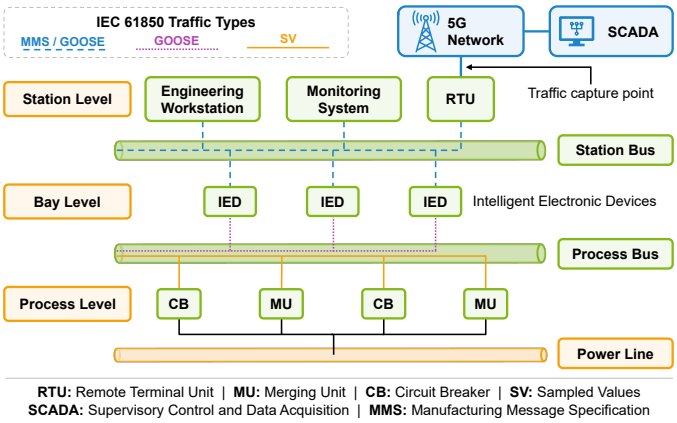

Fig. 5: IEC 61850 substation communication architecture, illustrating the hierarchical levels and their associated traffic types. The architecture serves as the basis for the traffic emulation used to generate representative communication patterns for training the predictive scheduling model.

IEC 61850 communication in the considered substation scenario is characterized by highly structured traffic patterns, consisting primarily of periodic monitoring messages and recurring event-driven communication. As a result, future transmissions can often be inferred from a limited history of recent packet arrivals and sizes, making the prediction task well suited for a regression-based approach.

A Random Forest regressor was selected as the prediction model due to its ability to capture nonlinear relationships between traffic features while remaining robust and computationally efficient. To determine a suitable model configuration, different lookback and lookahead settings were evaluated, followed by hyperparameter optimization using grid search with five-fold cross-validation. The final model employed a lookback window of 10 samples, a lookahead horizon of three future transmissions, 200 decision trees, and a maximum tree depth of 30.

Since proactive grant allocation depends on correctly predicting both the timing and size of future transmissions, inter-arrival time and packet size were jointly estimated. To account for timing variations and communication jitter, grant reservations were applied within a tolerance window of 2 ms around the predicted transmission instant.

Fig. 6 shows representative one-step-ahead $(t+1)$ and three-step-ahead $(t+3)$ predictions on the test dataset. A grant prediction is considered successful when both the transmis-

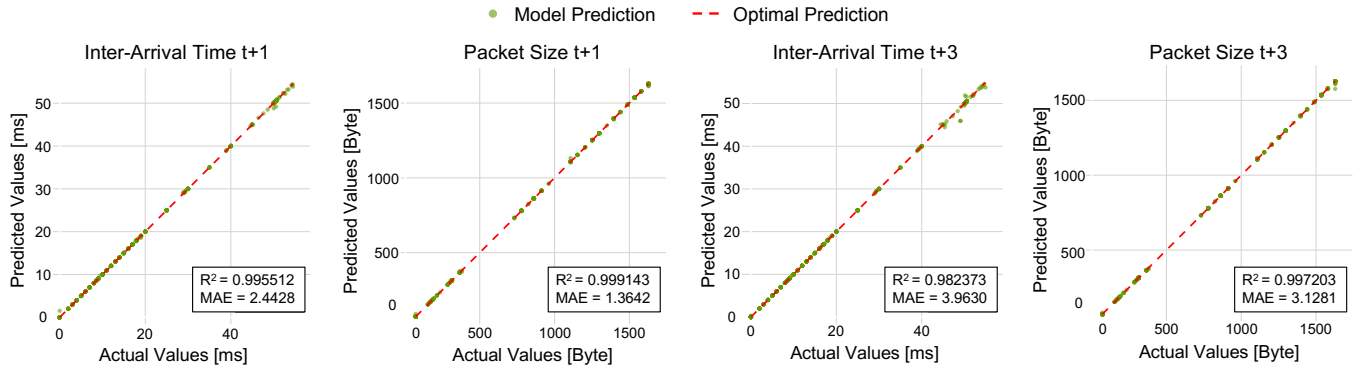

Fig. 6: Predicted versus actual inter-arrival time and packet size at one-step (t+1) and three-step (t+3) horizons, evaluating the Random Forest model used for Proactive Grant scheduling on IEC 61850 traffic.

sion time and the required resource reservation are predicted correctly within the configured tolerance window. The combined consideration of inter-arrival time (IAT) prediction and packet size estimation resulted in an overall grant prediction hit accuracy of 93.3 %, enabling reliable proactive resource reservation for future transmissions.

C. Base Station Energy Model in ns-3

To evaluate the energy consumption of 5G base stations under realistic network conditions, we implemented an analytical power model for the gNB Active Antenna Unit (AAU) within the ns-3 simulator. The implementation is based on the analytical model presented by Piovesan et al. [17] and is released as an open-source extension[2]. It integrates with the 5G-LENA NR module and the ns-3 energy framework.

The reference model decomposes the AAU power consumption into five components. The full expression used in our implementation is given in (1) and evaluated per scheduling slot; the corresponding symbols are summarized in Table I.

$$
\begin{aligned}
P_{AAU} = P_0 + P_{BB} &+ \underbrace{\sum_{t=1}^{T} M_{av,t}\, D_{Tran,t}}_{P_{Tran}} \\
&+ \underbrace{\sum_{t=1}^{T} M_{ac,t}\, D_{PA,t}\, f_{sym,t}}_{P_{PA}} \\
&+ \underbrace{\sum_{c=1}^{C} \frac{1}{\eta_c}\, P_{TX,c}^{max}\, \min\big(\rho_c + \rho_{ctrl,c},\, 1\big)}_{P_{out}}
\end{aligned}
\tag{1}
$$

While the reference model in [17] provides the analytical structure, it is formulated as a load-averaged expression and does not, in its original form, account for the slot- and symbol-level dynamics that arise in an event-driven simulator. To integrate it faithfully into the per-slot execution of ns-3 and to make it consistent with the symbol- and control-level behavior

[2]https://github.com/tudo-cni/ns3-gNB-energy-model

TABLE I: Parameters of the AAU power consumption model.

Symbol	Description	Symbol	Description
P_{AAU}	Total AAU power	η_c	MCPA efficiency, car. c
P_0	Baseline power	$P_{TX,c}^{max}$	Max. TX power, car. c
P_{BB}	Baseband power	C	No. of carriers
P_{Tran}	Transceiver power	$f_{sym,t}$	Symbol-activity factor
$M_{av,t}$	Avail. RF chains, t	$N_{data,t}$	Data symbols/slot
$D_{Tran,t}$	Power per RF chain	$N_{total,t}$	Total symbols/slot
T	No. of trx. groups	ρ_c	Data load fraction
P_{PA}	PA power	$\rho_{ctrl,c}$	Control load fraction
$M_{ac,t}$	Active RF chains, t	$N_{ctrl,c}^{RE}$	Control REs/slot
$D_{PA,t}$	Static MCPA power/chain	$N_{RB,c}$	Resource blocks, car. c
P_{out}	Transmit power		

observed at the PHY layer, we introduce two small refinements to the original formulation. Both leave the model's parameters and its limiting cases unchanged, but they improve its accuracy under realistic, time-resolved scheduling.

The first refinement concerns the power-amplifier term P_{PA}. In the reference model this term is a constant offset whenever the carrier is active. We replace it by an activity-dependent term that scales with a symbol-activity factor $f_{sym,t}$, defined as

$$
f_{sym,t} = \begin{cases} \dfrac{N_{data,t}}{N_{total,t}} & \text{if symbol shutdown enabled} \\ 1 & \text{otherwise} \end{cases}
\tag{2}
$$

Here $f_{sym,t}$ is the amplifier's duty cycle within a slot, i.e. the fraction of data-bearing symbols. It captures symbol-level discontinuous transmission (DTX), where the PA is switched off during empty symbols, and excludes control symbols so that P_{PA} vanishes to the idle point $P_0 + P_{BB} + P_{Tran}$ at $\rho_c = 0$ and recovers $M_{ac,t} D_{PA,t}$ at full load.

In the reference model, the transmit power P_{out} scales only with the data load ρ_c. However, the downlink control channels (e.g. PDCCH) also radiate power through the PA. We therefore add a separate control load fraction $\rho_{ctrl,c}$, which measures the share of the slot capacity occupied by control resource elements and is counted only while data is being transmitted

($\rho_c > 0$), since idle control overhead is already part of P_0. It is defined as

$$\rho_{ctrl,c} = \begin{cases} \dfrac{N^{RE}_{ctrl,c}}{N_{RB,c}\, N_{total,c}} & \text{if } \rho_c > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

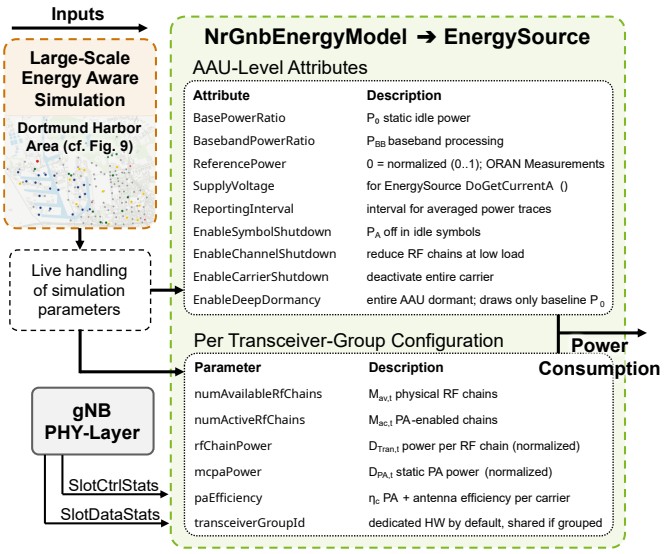

Fig. 7: Architecture of the proposed gNB energy model in ns-3. The model combines scenario-specific deployment information with live PHY-layer utilization statistics and supports runtime adaptation through configurable AAU- and transceiver-group parameters.

As shown in Fig. 7, the model is integrated with the ns-3 energy framework and attaches to the 5G-LENA NR base station at the physical layer. For each slot in which the base station accesses the channel, dedicated monitoring interfaces report the scheduler's resource block utilization, the number of data and control symbols, and the current transmit power. The model is configured through a set of base station level attributes and a hardware parameter set defined per carrier. Each carrier belongs to a transceiver group that represents one physical radio front end. The static transceiver and amplifier power is accounted once per transceiver group, whereas the radiated power is accounted per carrier. Carriers sharing a front end belong to the same transceiver group, while distinct bands or panels form separate groups within one base station. Internally, the power is always computed in normalized form, where full load corresponds to a value of one. This allows a direct comparison with the analytical reference curves. The normalized values can then be mapped to absolute power figures using per-component measurements reported by the O-RAN Alliance Working Group in [16], which provide power consumption characteristics for different base station configurations and load conditions. For each slot, the model derives the demand load fraction from the scheduler's allocation, updates a smoothed load estimate using

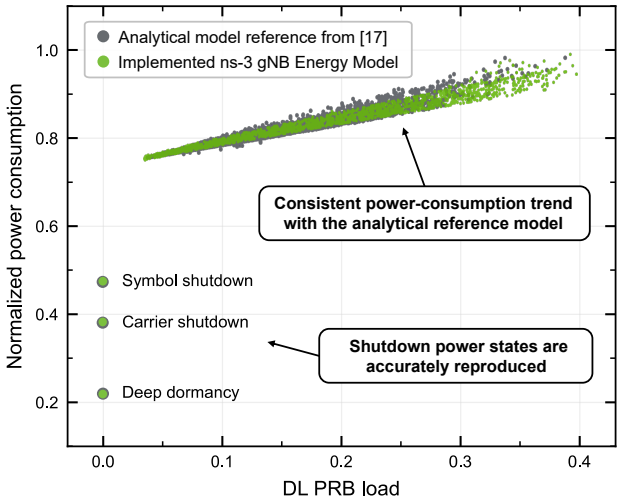

Fig. 8: Validation of the implemented ns-3 gNB energy model against the analytical reference model from [17], showing close agreement in load-dependent and shutdown-state power consumption.

an exponentially weighted moving average that drives the energy saving state, and evaluates the individual power components, aggregating the static hardware terms per transceiver group and the radiated power per carrier. Four energy saving mechanisms are available. Symbol shutdown scales the static amplifier power by the ratio of active data symbols to total symbols per slot. Channel shutdown reduces the number of active RF chains under low load. Carrier shutdown deactivates an entire carrier, and deep dormancy leaves only the baseline power once all carriers of a base station are switched off.

We validate the implementation against the total normalized power consumption of the reference model [17]. For validation, the transmit power is varied over different downlink PRB load levels while the resulting power consumption is monitored. The symbol shutdown, carrier shutdown, and deep dormancy mechanisms are evaluated individually and compared against their respective reference behavior. Fig. 8 compares the analytical reference values, shown in grey, with the output of our model, shown in green. The simulated behavior closely follows the reference, reproducing both the load-dependent trend and the observed scatter. The close agreement across all evaluated load levels and shutdown mechanisms confirms that the implementation accurately reproduces the behavior of the analytical reference model.

D. Realistic Simulation Environment and Scenario Modeling

To evaluate the proposed scheduling and energy management mechanisms under realistic conditions, the simulation scenario is based on the Dortmund harbor district, as illustrated in Fig. 9. A dedicated preprocessing pipeline combines publicly available infrastructure data from OSM [18] with cellular deployment information obtained from [19]. The extracted geographic coordinates are translated directly into the

ns-3 environment, allowing communication and power system entities to retain their actual spatial relationships.

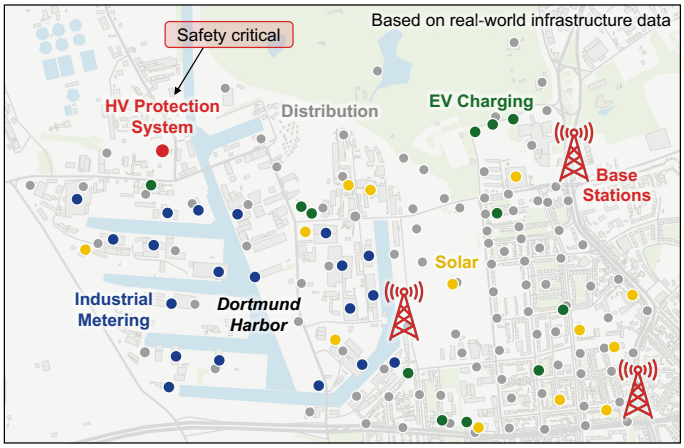

Fig. 9: Geospatial critical infrastructure communication scenario in the Dortmund Harbor district, combining cellular infrastructure with distributed energy assets and safety-critical protection applications.

The resulting topology incorporates both cellular and mixed-criticality infrastructure. Real gNB locations are used and configured according to common deployment practices, employing the n28 band for wide area coverage and the n78 band for high capacity services. The Smart Grid model includes high voltage substations, distribution stations, photovoltaic installations, industrial metering points, and electric vehicle charging stations. Particular attention is given to protection traffic associated with high voltage substations, which constitutes one of the most demanding IEC 61850 communication services in terms of latency requirements.

IV. Evaluation and Results

This section evaluates the implemented communication mechanisms within a realistic simulation environment, assessing their capability to fulfill the strict latency and reliability requirements of critical infrastructure applications, exemplified by Smart Grids. The analysis focuses on the implementation and evaluation of two resource management mechanisms: slice-based resource reservation for separating traffic classes and predictive proactive grants for reducing uplink latency in periodic mission-critical traffic. The evaluation is conducted using the Dortmund harbor district scenario introduced in the previous chapter. Beyond communication performance, energy awareness at the base station level is a central aspect of this evaluation, as constrained energy availability requires phased component deactivation while simultaneously preserving reliable communication for the highest-priority services. Network Slicing is the key enabler of this interdependence, with predictive proactive grants further tightening the latency budget for time-critical flows under these constrained conditions.

A. Simulation Configuration

The evaluation is conducted within the ns-3 simulation environment (v3.47), extended by the 5G-LENA module (v4.2), using the Dortmund harbor district scenario introduced in the previous chapter. To ensure a realistic representation of current cellular infrastructure, the network configuration follows practical deployment parameters for both coverage and capacity layers. The setup employs two frequency bands, with n28 operating in Frequency Division Duplex (FDD) mode to provide wide-area coverage and n78 operating in Time Division Duplex (TDD) mode to deliver additional capacity. Furthermore, devices participating in the mixed-criticality scenario are assigned application-specific 5QI values according to their communication requirements, enabling differentiated treatment of critical and non-critical traffic. Network Slicing guarantees resource availability for critical traffic, with predictive proactive grants additionally applied to high-voltage substation flows for further latency reduction. The complete set of simulation and traffic parameters is summarized in Table II.

TABLE II: Simulation setup and scenario configuration.

Parameter	**Band**	
	n28	**n78**
Frequency	700 MHz	3.5 GHz
Duplex mode	FDD	TDD
Bandwidth	20 MHz	60 MHz
MIMO	4×4	64×64
TDD pattern	–	7D2U
SCS / MCS table	30 kHz / 3GPP Table 1	
BS TX power	49 dBm	
Channel model	Urban / Suburban	
Sectors per gNB	3 (120° each)	
BS / UE height	40 m / 1.5 m	
Assigned 5QI values	6, 7, 8, 9, 83, 84, 85	
Traffic (Smart Grid)	IEC 61850-modeled traffic	
Traffic (Best-effort)	$\lambda = 100$–200 packets per second, 100–1000 Byte	
Slice: Critical	6 High-Voltage protection systems (Proactive Grants)	
Slice: Standard	EV chargers, PV, energy distribution, smart meters	
Slice: Low	Low-Voltage distribution boxes	
Slice: Best-effort	Non-critical traffic	
Simulation	60 s (+500 ms init/cool-off), 10 seeds	

B. Performance Analysis of Prioritization Mechanisms

The performance of the proposed prioritization mechanisms is evaluated under increasing network load in terms of latency and reliability. For this purpose, the number of active UEs is gradually increased to analyze the scalability and robustness of the different configurations. The considered setups include the default configuration without prioritization, 5QI-based traffic differentiation, Network Slicing, the combined 5QI and slicing approach, and the proposed predictive proactive grant mechanism for the most critical HV substation devices.

Fig. 10 shows the resulting mean packet delivery ratio and end-to-end latency for prioritized critical infrastructure traffic, represented by HV protection systems in Smart Grids, together with the corresponding Best Effort background traffic. In the

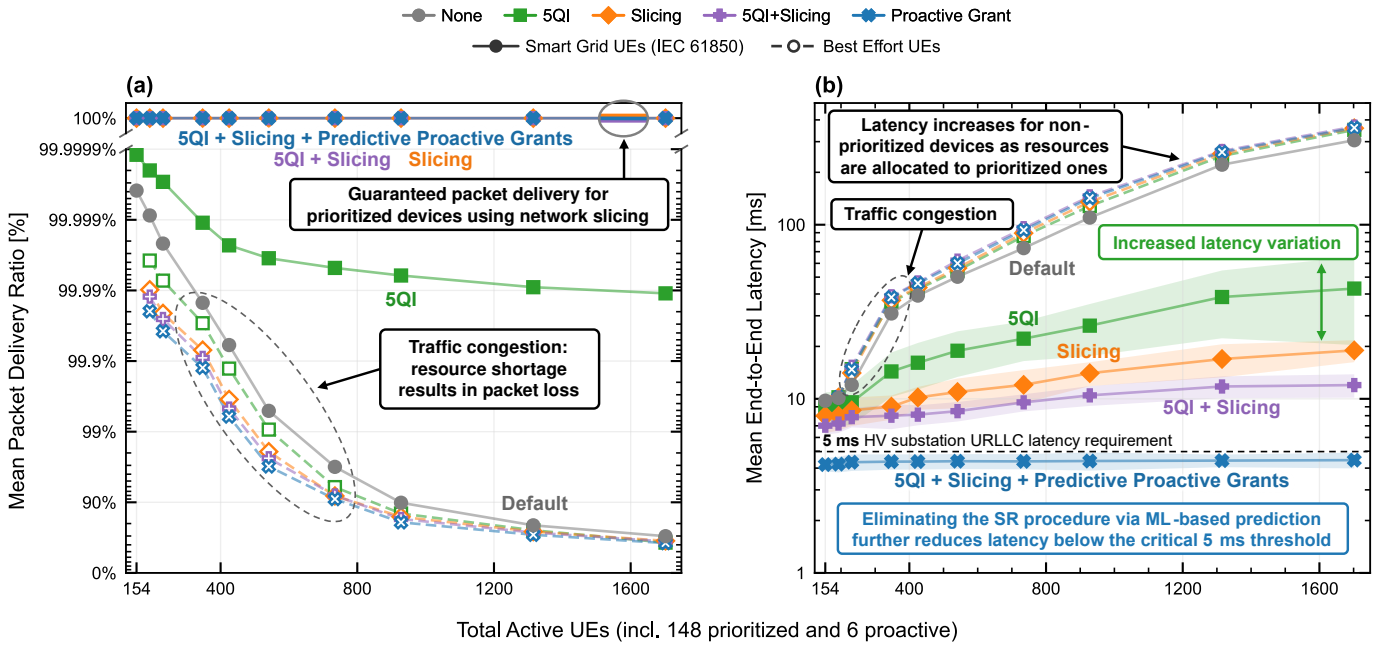

Fig. 10: Impact of different prioritization mechanisms on communication performance under increasing network load. (a) Mean packet delivery ratio and (b) mean end-to-end latency (bands: mean jitter between consecutive packets) for prioritized devices, comparing default scheduling, 5QI-based prioritization, Network Slicing, and Predictive Proactive Grants.

default configuration, both latency and packet losses increase rapidly as the number of UEs grows, indicating early network congestion and insufficient resource availability for critical traffic. Since all devices compete equally for radio resources, mission-critical communication cannot be reliably protected under high load.

Introducing 5QI-based prioritization improves latency compared to the default case, as traffic classes with higher priority receive preferential scheduling. However, because the applied scheduler is still based on weighted proportional resource allocation, it cannot guarantee packet delivery under severe congestion. This behavior is reflected in Fig. 10 (a), where packet losses remain observable at high network load despite the applied prioritization.

Network Slicing provides a stronger isolation mechanism by reserving dedicated resources for specific service classes. As a result, the slicing-based configurations eliminate packet loss for prioritized devices across the considered load range, consistently exceeding the reliability target of 99.9999 %. This behavior follows directly from the static resource reservation within the corresponding network slices, which guarantees the availability of resources for critical traffic independently of Best Effort load conditions. Predictive proactive grants have no direct impact on reliability, as packet delivery is already ensured by the resource isolation provided through Network Slicing.

While Network Slicing primarily improves reliability, the latency results shown in Fig. 10 (b) reveal additional benefits of the proposed hierarchical prioritization approach, where the shaded bands indicate the latency variation and therefore the jitter seen by the application. Compared to standalone 5QI scheduling, the combination of 5QI and Network Slicing further reduces end-to-end latency and visibly narrows this band, since the reserved resources are available independently of the Best Effort load, keeping the waiting time until a transmission opportunity nearly constant. The lowest latency is achieved when predictive proactive grants are enabled for the most critical HV substation devices. By allocating uplink resources in advance based on predictable traffic patterns, the Scheduling Request (SR) and grant procedure can be partially bypassed, reducing signaling overhead and delays. Consequently, the average latency remains close to 4 ms even under high network load, complementing the reliability benefits of Network Slicing.

C. Energy-Constrained Network Performance

Beyond reliability and latency, the energy demand of the gNB is a relevant factor for the integration of mobile network infrastructure into Smart Grid communication, particularly under emergency or resource-constrained operating conditions. This subsection evaluates gNB power consumption under an incremental power reduction strategy, in which the radio configuration is progressively scaled down across several defined phases, and examines whether the prioritization guarantees provided by Network Slicing remain effective as the network is degraded.

The base station parametrization corresponds to the configuration given in Table II. The reduction strategy investigated here follows a defined deactivation plan comprising seven phases of increasing severity, summarized in Table III. Starting

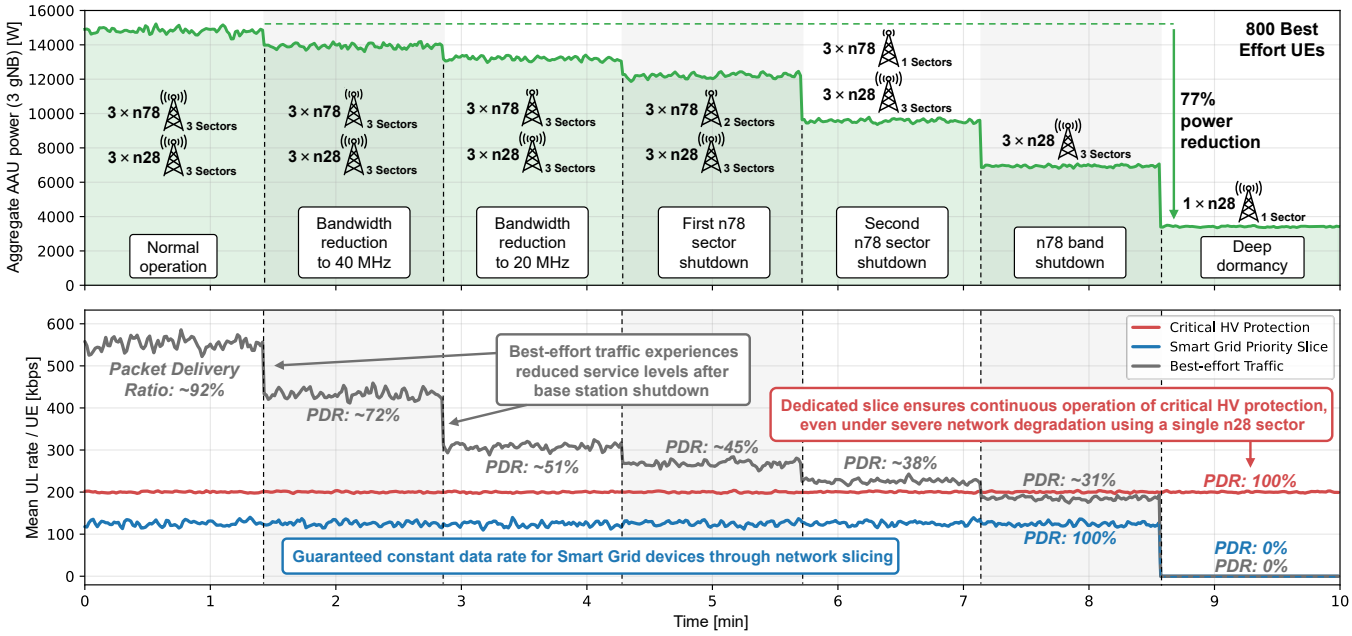

Fig. 11: Stepwise gNB power-reduction and per-slice uplink rates in the considered smart-grid scenario with 800 UEs. As the available base-station capacity is reduced, slicing maintains the critical HV-protection and smart-grid priority slices until a critical network state is reached, where deep dormancy is applied and only the most critical traffic remains served.

from the baseline configuration, the n78 bandwidth is gradually reduced in two steps, first from 60 MHz to 40 MHz, then from 40 MHz to 20 MHz, together with corresponding transmit power adjustments. Subsequently, the least-loaded n78 sector per site is deactivated, followed by deactivation of the two least-loaded sectors, and finally complete shutdown of the n78 layer through band-level deep dormancy. A final, critical phase reduces the network to a single active n28 sector across all sites, with the remaining sectors set to deep dormancy, representing a worst-case emergency configuration.

TABLE III: Stepwise gNB power reduction plan for energy conservation in critical infrastructure scenarios.

Phase	**Configuration**
Normal operation	All sectors active, Band n28 + n78
BW reduction I	n78 bandwidth reduced to 40 MHz
BW reduction II	n78 bandwidth reduced to 20 MHz
Partial n78 shutdown I	Least-loaded n78 sector per site deactivated
Partial n78 shutdown II	Two least-loaded n78 sectors per site deactivated
Full n78 shutdown	All n78 sectors in deep dormancy
Emergency mode	One n28 sector active; rest deep dormant

Each phase corresponds to a concrete modification of the radio configuration, including bandwidth reduction, transmit-power adaptation, or carrier deactivation, thereby directly affecting both power consumption and the available radio capacity. As n78 sectors are deactivated, their associated UEs are rerouted onto the n28 coverage layer, increasing the load on the remaining active sectors. Consequently, the available radio resources must be shared among a larger number of UEs, making efficient traffic prioritization increasingly important in the later deactivation phases. Since Network Slicing has already been shown to reliably prioritize critical IEC 61850 Smart Grid traffic over best-effort traffic, this subsection further investigates whether this prioritization remains effective as the network is progressively degraded. To this end, the achieved data rates of the critical, sliced devices are analyzed across the deactivation phases, in order to verify that prioritized critical communication remains stable even under significant power-driven network reduction.

The resulting power consumption across the seven phases is shown in the upper panel of Fig. 11, where the individual reduction measures are applied successively to demonstrate their cumulative effect. Total Radio Unit (RU) power consumption decreases with each phase, although the contribution of the individual measures differs considerably. Bandwidth reduction yields only limited savings, reducing consumption by approximately 4 % and 7 % relative to the baseline while all sectors remain active. Significantly larger gains are achieved through n78 deactivation, reaching reductions of approximately 18 %, 36 %, and 53 % for the successive shutdown phases. This behavior is expected, as the later phases progressively place larger portions of the radio infrastructure into dormancy rather than merely reducing the utilization of active resources. The final emergency configuration, in which only a single n28 sector remains operational, lowers the total RU power consumption by approximately 77 % compared to normal operation. In this state, only the minimum infrastructure required to maintain critical communication remains active.

The lower panel of Fig. 11 reports the corresponding mean per-UE uplink data rate and packet delivery rate, separated by

traffic class. During the gradual reduction of radio resources, the critical HV-protection slice remains stable throughout all phases, while the smart-grid priority slice is maintained only as long as sufficient residual capacity is available. In the final emergency configuration, the network serves only the most critical traffic.

Best-effort traffic degrades as resources are removed, with the most pronounced reduction occurring already during the n78 bandwidth scaling phases. Here, available capacity is reduced while the number of served UEs remains unchanged, directly affecting both data rate and packet delivery rate. Subsequent sector deactivations have a comparatively smaller impact, as affected UEs can still be served by the remaining sectors and the n28 coverage layer. As a result, the loss of capacity is partially compensated despite the reduced number of active sectors. Only when the network reaches the emergency configuration are the remaining resources reserved exclusively for critical communication, causing all non-critical traffic to be dropped. Overall, the results demonstrate that the slicing scheduler first absorbs resource shortages within the lower-priority traffic classes while preserving the service guarantees of the highest-priority applications for as long as network resources permit.

V. Conclusion

Our proposed framework demonstrates the feasibility of jointly addressing traffic prioritization and energy-aware operation for critical infrastructure communication over public cellular networks. By combining Network Slicing with ML-based predictive proactive grants, the framework provides guaranteed resources for critical traffic while reducing uplink access delays, achieving end-to-end latencies below 5 ms and exceeding the 99.9999 % reliability target under high network load. Furthermore, the results show that critical communication services can be maintained under severely energy-constrained operation, where progressive component deactivation lowers base station power consumption by up to 77 %. This requires coordinated traffic prioritization and guaranteed resource allocation alongside energy-aware network operation. As radio resources are progressively reduced and infrastructure components enter dormancy, Network Slicing sheds non-critical services while preserving communication for the most critical protection and control functions. These mechanisms are not specific to this scenario: comparable mixed-criticality deployments with similarly heterogeneous requirements, such as industrial automation or transport infrastructure, can build on the same approach. The slicing scheduler and the gNB power model are released as open-source ns-3 modules to support reuse and reproducibility. Future work will extend the evaluation to larger and more dynamic deployments with UE mobility, aperiodic traffic, and more adaptive slice management.

Acknowledgment

This work has been partially funded by the Federal Ministry for Economic Affairs and Energy (BMWE) via the project ReCoDE under funding reference 03EI6093B.

References

[1] 3rd Generation Partnership Project (3GPP), "Service requirements for cyber-physical control applications in vertical domains; Stage 1 (Release 19)," 3GPP, Technical Specification (TS) 22.104, Jun. 2024.

[2] ——, "Service requirements for the 5G system; Stage 1 (Release 20)," 3GPP, Technical Specification (TS) 22.261, Mar. 2026.

[3] European Telecommunications Standards Institute (ETSI), "Environmental Engineering (EE); Sustainable Power Feeding Solutions for 5G Network," ETSI Standard ETSI ES 203 700 V1.1.1, Dec. 2020.

[4] Z. Zhao, Q. Du, and G. K. Karagiannidis, "Improved Grant-Free Access for URLLC via Multi-Tier-Driven Computing: Network-Load Learning, Prediction, and Resource Allocation," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 3, pp. 607–622, 2023.

[5] K. Boutiba, M. Bagaa, and A. Ksentini, "On using Deep Reinforcement Learning to reduce Uplink Latency for uRLLC services," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022.

[6] D. Overbeck, N. A. Wagner, F. Kurtz, and C. Wietfeld, "Proactive Resource Management for Predictive 5G Uplink Slicing," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022.

[7] V. Kumar Gautam, V. R. Chintapalli, B. R. Tamma, and C. S. Ram Murthy, "Enhancing Uplink Scheduling in 5G Enabled Vehicular Networks: A Cross-Layer Approach with Predictive Buffer Status Reporting," in *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, 2024, pp. 01–07.

[8] B. Sliwa, H. Schippers, and C. Wietfeld, "Machine Learning-Enabled Data Rate Prediction for 5G NSA Vehicle-to-Cloud Communications," in *2021 IEEE 4th 5G World Forum (5GWF)*, 2021, pp. 299–304.

[9] R. Ramalakshmi, V. A. Vinex, M. Bratheep, and K. Hareshkumar, "System-Level Modeling and End-to-End Evaluation of 5G Networks in NS-3: Access Technologies, Scheduling, Network Slicing, and Core VNF Scaling," in *2026 International Conference on Future and Advanced Computing Technologies (ICFACT)*, 2026, pp. 1–6.

[10] NGMN Alliance, "Network Energy Efficiency Phase 2," Next Generation Mobile Networks (NGMN) Alliance, Final Deliverable, October 2023, Green Future Networks Programme, Version 1.0. [Online]. Available: https://www.ngmn.org/publications/network-energy-efficiency-phase-2.html

[11] G. Vallero, G. Perin, M. Meo, M. S. Garino, S. D'Elia, and D. Vaccarono, "A New Explainable Power Demand Model for 4G LTE and 5G NR Base Stations," in *ICC 2025 - IEEE International Conference on Communications*, 2025, pp. 4173–4178.

[12] D. López-Pérez, A. De Domenico, N. Piovesan, and M. Debbah, "Data-Driven Energy Efficiency Modeling in Large-Scale Networks: An Expert Knowledge and ML-Based Approach," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, pp. 780–804, 2024.

[13] A. Sen, B. Pal, and S. Achari, "SmartMME: Implementation of Base Station Switching Off Strategy in ns-3," in *2024 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, 2024, pp. 1–6.

[14] A.-K. Dang, H. Khalifé, M. Sintorn, S. Rovedakis, and S. Secci, "Data-driven Energy Optimization in Mobile Networks with User Experience Guarantees," in *IEEE INFOCOM 2025 - IEEE Conference on Computer Communications*, 2025, pp. 1–10.

[15] A. Sen, S. Bonda, J. Jayatheerthan, and S. Chakraborty, "An ns3-based Energy Module for 5G mmWave Base Stations," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2022, pp. 1–2.

[16] O-RAN Alliance, Work Group 1 (Use Cases and Overall Architecture), "Network Energy Saving Use Cases Technical Report," 2023, Version 2.00.

[17] N. Piovesan, D. López-Pérez, A. De Domenico, X. Geng, H. Bao, and M. Debbah, "Machine Learning and Analytical Power Consumption Models for 5G Base Stations," *IEEE Communications Magazine*, vol. 60, no. 10, pp. 56–62, 2022.

[18] OpenStreetMap contributors, "OpenStreetMap," 2026. [Online]. Available: https://www.openstreetmap.org

[19] Surveying and Cadastral Office. (2021) Mobilfunkkataster Dortmund [German]. City of Dortmund. [Online]. Available: https://geoweb1.digistadtdo.de/doris_gdi/mapapps4/resources/apps/stadtinformation/index.html

[20] MZ Automation GmbH, "libIEC61850 – An open-source library for IEC 61850," 2025, Version 1.6.0.