X5G: An Open, Programmable, Multi-vendor, End-to-end, Private 5G O-RAN Testbed with NVIDIA ARC and OpenAirInterface

Davide Villa, Imran Khan, Florian Kaltenberger, Nicholas Hedberg, Rúben Soares da Silva, Stefano Maxenti, Leonardo Bonati, Anupa Kelkar, Chris Dick, Eduardo Baena, Josep M. Jornet, Tommaso Melodia, Michele Polese, and Dimitrios Koutsonikolas

Abstract—As Fifth generation (5G) cellular systems transition to softwarized, programmable, and intelligent networks, it becomes fundamental to enable public and private 5G deployments that are (i) primarily based on software components while (ii) maintaining or exceeding the performance of traditional monolithic systems and (iii) enabling programmability through bespoke configurations and optimized deployments. This requires hardware acceleration to scale the Physical (PHY) layer performance, programmable elements in the Radio Access Network (RAN) and intelligent controllers at the edge, careful planning of the Radio Frequency (RF) environment, as well as end-to-end integration and testing. In this paper, we describe how we developed the programmable X5G testbed, addressing these challenges through the deployment of the first 8-node network based on the integration of NVIDIA Aerial RAN CoLab Over-the-Air (ARC-OTA), OpenAirInterface (OAI), and a near-real-time RAN Intelligent Controller (RIC). The Aerial Software Development Kit (SDK) provides the PHY layer, accelerated on Graphics Processing Unit (GPU), with the higher layers from the OAI open-source project interfaced with the PHY through the Small Cell Forum (SCF) Functional Application Platform Interface (FAPI). An E2 agent provides connectivity to the O-RAN Software Community (OSC) near-real-time RIC. We discuss software integration, network infrastructure, and a digital twin framework for RF planning. We then profile the performance with up to 4 Commercial Off-the-Shelf (COTS) smartphones for each base station with iPerf and video streaming applications, as well as up to 25 emulated User Equipments (UEs), measuring a cell rate higher than 1.65 Gbps in downlink and 143 Mbps in uplink.

Index Terms—Private 5G; Multi-vendor; GPU acceleration; O-RAN.

This is a revised and substantially extended version of [1], which appeared in the Proceedings of the 2nd Workshop on Next-generation Open and Programmable Radio Access Networks (NG-OPERA) 2024.
This work was partially supported by the U.S. National Science Foundation under grant CNS-2117814, and by the U.S. National Telecommunications and Information Administration (NTIA)'s Public Wireless Supply Chain Innovation Fund (PWSCIF) under Award No. 25-60-IF054.
D. Villa, I. Khan, F. Kaltenberger, S. Maxenti, L. Bonati, E. Baena, J. M. Jornet, T. Melodia, M. Polese, and D. Koutsonikolas are with the Institute for the Wireless Internet of Things, Northeastern University, Boston, MA. Email: {villa.d, khan.i, f.kaltenberger, maxenti.s, l.bonati, e.baena, j.jornet, melodia, m.polese, d.koutsonikolas}@northeastern.edu
F. Kaltenberger is also with Eurecom, Sophia Antipolis, France. Email: florian.kaltenberger@eurecom.fr
N. Hedberg, A. Kelkar, and C. Dick are with NVIDIA, Inc., Santa Clara, CA. Email: {nhedberg, anupak, cdick}@nvidia.com
R. Soares da Silva is with Allbesmart, Castelo Branco, Portugal. Email: rsilva@allbesmart.pt
D. Villa and I. Khan are co-primary authors.

1 Introduction

The evolution of the Radio Access Network (RAN) in Fifth generation (5G) networks has led to key performance improvements in cell and user data rates, now at hundreds of Mbps on average, and in air interface latency [2], thanks to specifications developed within the 3rd Generation Partnership Project (3GPP). From an architectural point of view, 5G deployments are also becoming more open, intelligent, programmable, and based on software [3], through activities led by the O-RAN ALLIANCE, which is developing the network architecture for Open RAN. These elements have the potential to transform how we deploy and manage wireless mobile networks [4], leveraging intelligent control, with RAN optimization and automation exercised via closed-loop data-driven control; softwarization, with the components of the end-to-end protocol stack defined through software rather than with dedicated hardware; and disaggregation, with the 5G RAN layers distributed across different network functions, i.e., the Central Unit (CU), the Distributed Unit (DU), and the Radio Unit (RU).

Open and programmable networks are often associated with lower capital and operational expenditures, facilitated by the increasing robustness and diversity of the telecom supply chain [5], now also including open-source projects [6, 7] and vendors focused on specific components of the disaggregated RAN. This, and increased spectrum availability in dedicated or shared bands, has opened opportunities to deploy private 5G systems, complementing public 5G networks with more agile and dynamic deployments for site-specific use cases (e.g., events, warehouse automation, industrial control, etc).

While the transition to disaggregated, software-based, and programmable networks comes with significant benefits, there are also several challenges that need to be addressed before Open RAN systems can align their performance or improve over traditional cellular systems. First of all, the radio domain still exhibits a low degree of automation and zero-touch provisioning for the RAN configuration, complicating the successful deployment of end-to-end cellular systems. Second, the diverse vendor ecosystem comes with challenges related to interoperability and end-to-end integration across several products, potentially from

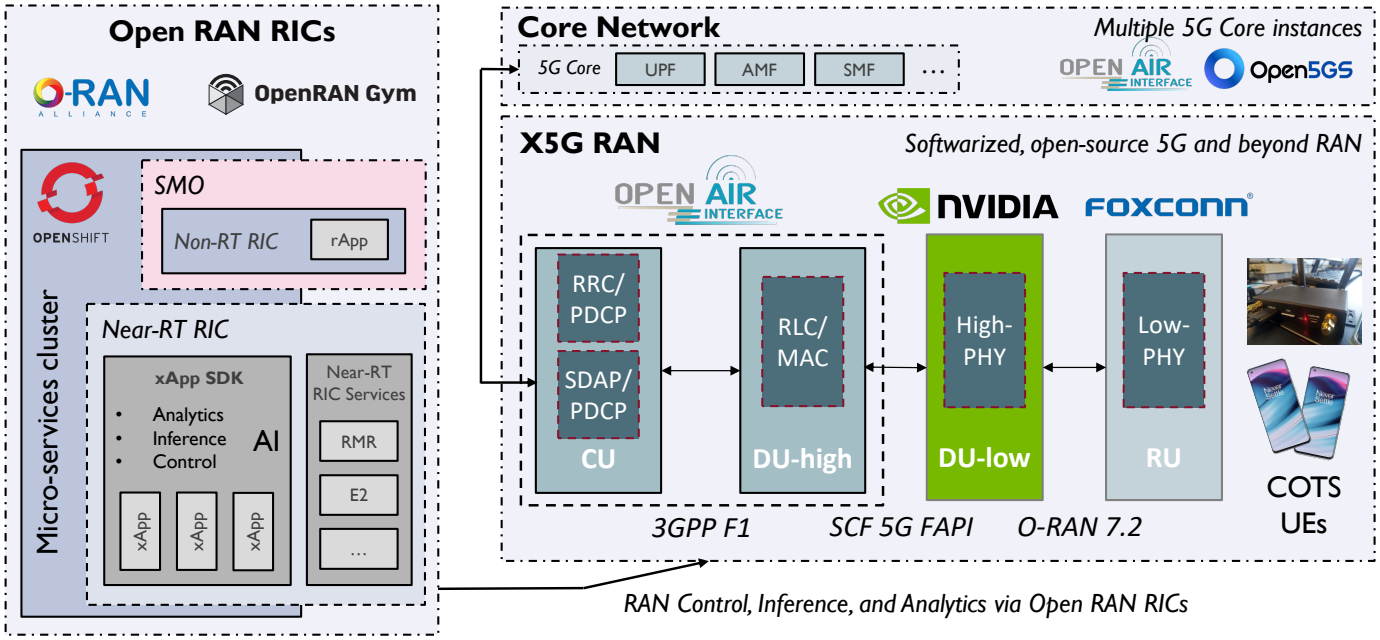


Fig. 1: X5G end-to-end programmable testbed overview.

different vendors [8–10]. Third, the Digital Signal Processing (DSP) at the Physical (PHY) layer of the stack is a computationally complex element, using about 90% of the available compute when run on general-purpose CPUs, and thus introducing a burden on the software-based and virtualized 5G stack components. Finally, there are still open questions in terms of how the intelligent and data-driven control loops can be implemented with Artificial Intelligence (AI) and Machine Learning (ML) solutions that generalize well across a multitude of cellular network scenarios [11]. These challenges call for a concerted effort across different communities (including hardware, DSP, software, DevOps, AI/ML) that aims to design and deploy open, programmable, multi-vendor cellular networks and testbeds that can support private 5G requirements and use cases with the stability and performance of production-level systems.

In this paper, we introduce X5G, a private 5G network testbed deployed at Northeastern University in Boston, MA, and based on multiple programmable and open-source components from the physical layer all the way up to the Core Network (CN), as shown in Figure 1. We discuss in detail the integration of a PHY layer implemented on Graphics Processing Unit (GPU) (i.e., NVIDIA Aerial) with OpenAirInterface (OAI) for the higher layers of the 5G stack [12]. This integration is based on the Small Cell Forum (SCF) Functional Application Platform Interface (FAPI), which regulates the interaction between the PHY and Medium Access Control (MAC) layers. This paper extends our recent work [1] by introducing new Open RAN elements and experimental results, including: (i) the integration of a near-real-time RAN Intelligent Controller (RIC) from the O-RAN Software Community (OSC) on an OpenShift cluster; (ii) the validation of additional CNs, such as Open5GS and a commercial core from A5G; (iii) enhancements to the hardware architecture, including a more robust networking infrastructure and additional RAN servers; (iv) the evaluation of X5G under diverse operational conditions, such as stress testing its performance ensuring reliability for a Private 5G (P5G) network; and (v) a more comprehensive related work section.

X5G leverages the inline acceleration of demanding PHY tasks on GPU, hardware that is well equipped with massive parallelization of DSP operations, enabling scalability and the embedding of AI/ML in the RAN. The X5G infrastructure is continuously expanding through the integration of an increasing number of components from various vendors, manufacturers, and open-source projects—such as NVIDIA, OAI, OpenShift, Keysight, OSC, Open5GS, and Foxconn—thereby creating a truly multi-vendor network architecture. It currently comprises more than 8 RAN servers for the NVIDIA/OAI CU and DU (known as NVIDIA Aerial RAN CoLab Over-the-Air (ARC-OTA) and referred to as ARC in this paper), several RUs from different vendors that can be installed in a lab space, as well as a Keysight RU emulator for further testing and profiling, O-RAN 7.2 fronthaul and timing hardware, along with multiple 5G CNs. The system delivers Key Performance Indicators (KPIs) representative of 5G sub-6 GHz systems, with cell throughput north of 1.65 Gbps with up to 25 connected User Equipments (UEs) and a 100 MHz carrier bandwidth.

The tools we developed, integrated, and deployed on X5G can be readily used for the development of intelligent use cases for 5G and beyond, thanks to the combination of NVIDIA ARC, OAI, and the OSC projects. As a result, this combination offers performance improvements over most open-source, non-accelerated solutions while maintaining the openness and code accessibility typical of Open RAN systems, further enhanced by the seamless integration of GPUs. X5G provides researchers with the necessary capabilities to develop, test, and evaluate a wide range of AI/ML and RAN solutions on a production-ready platform, including spectrum sharing techniques [13], secure cellular networks [14, 15], resource optimization [16], interference detection and mitigation, handover strategies, and the development of intelligent and autonomous networks. In addition, documentation and tutorials allow for the replication and bootstrapping of the testbed and its functionalities

across research institutions and beyond. In fact, the value propositions of a platform similar to X5G, with its openness, multi-vendor support, and GPU-accelerated capabilities, have been demonstrated for industrial stakeholders, as shown by SoftBank and Fujitsu in [17].

The rest of the paper is organized as follows. Section 2 introduces the software frameworks we developed and integrated to enable X5G. Section 3 concerns the deployment and configuration of the X5G network infrastructure. Section 4 describes an RF planning study to determine an optimal location for deploying the RUs. System performance is evaluated in Section 5 through various use case scenarios with multiple Commercial Off-the-Shelf (COTS) UEs and applications. Section 6 compares X5G with the state of the art. Section 7 draws conclusions and outlines our future work.

2 X5G Software

This section describes the software components of X5G, also shown in Figure 1. These components can be divided into three main groups: (i) a full-stack programmable Next Generation Node Base (gNB) (X5G RAN); (ii) the Open RAN RICs deployed on a micro-services cluster based on OpenShift; and (iii) various Core Networks (CNs) deployed in a micro-services-based architecture essential for the effective functioning of the 5G network.

2.1 Full-stack Programmable RAN with NVIDIA Aerial and OpenAirInterface

The right part of Figure 1 shows a detailed breakdown of the architecture of the X5G RAN, which follows the basic O-RAN architecture split into CU, DU, and RU. The DU is further split into a DU-low, implementing Layer 1 (PHY, or L1) functionalities, and into a DU-high, implementing Layer 2 (MAC and Radio Link Control (RLC), or L2) ones. As shown in Figure 2, DU-low and DU-high communicate over the 5G FAPI interface specified by the SCF [18]. The DU-low is implemented using the NVIDIA Aerial Software Development Kit (SDK) [19] on in-line GPU accelerator cards, whereas DU-high and CU are implemented by OAI on general-purpose Central Processing Units (CPUs). We deploy each function in separate Docker containers, sharing a dedicated memory space for the inter-process communication library that enables the FAPI interface. In our setup, we also combine the CU and the DU-high into a combined L2/L3 gNB Docker container, but the F1 split has also been deployed and tested.

The FAPI interface between the DU-high and DU-low defines two sets of procedures: configuration and slot procedures. Configuration procedures handle the management of the PHY layer and happen infrequently, e.g., when the gNB stack is bootstrapped or reconfigured. On the contrary, slot procedures happen in every slot (i.e., every 500 μs for a 30 kHz subcarrier spacing) and determine the structure of each Downlink (DL) and Uplink (UL) slot. In our case, L1 serves as the primary and L2 as the subordinate. Upon the reception of a slot indication message from L1, L2 sends either an UL or DL request to dictate the required actions for the PHY layer in each slot. Additionally, L1 might transmit other indicators to L2, signaling the receipt of data related

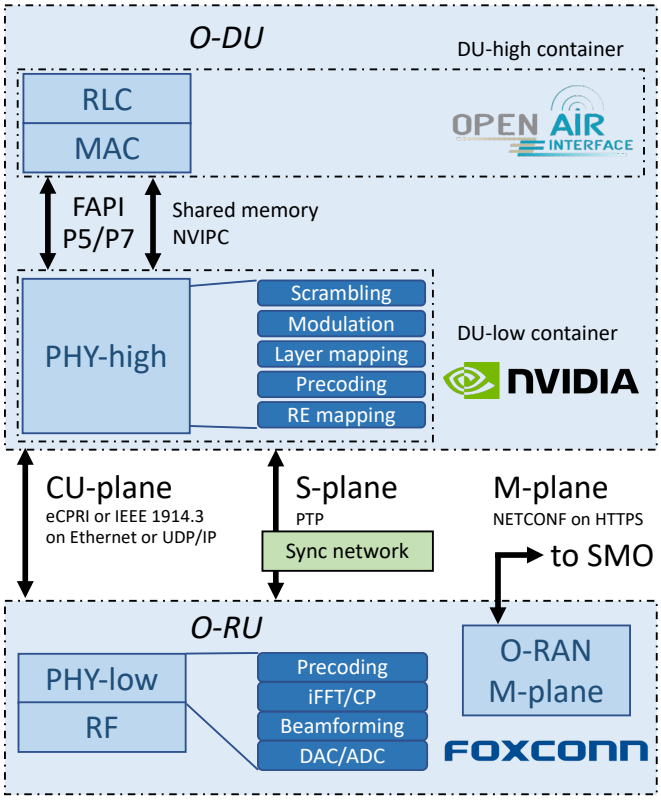


Fig. 2: Architecture of the lower layers of the X5G RAN following O-RAN specifications and consisting of: (i) a Foxconn O-RU; (ii) an O-DU-low based on NVIDIA Aerial SDK; (iii) an O-DU-high based on OpenAirInterface with their corresponding interfaces.

to Random Access Channel (RACH), Uplink Control Indication (UCI), Sounding Reference Signal (SRS), checksums, or user plane activities.

In our implementation, we use FAPI version 222.10.02 with a few exceptions as outlined in the NVIDIA Aerial release notes [20]. The transport mechanism for FAPI messages is specified in the networked FAPI (nFAPI) specification [21], which assumes that messages are transported over a network. However, in our implementation, the L1 and L2 Docker containers communicate through the NVIDIA Inter-Process Communication (NVIPC) library. This tool provides a robust shared memory communication framework specifically designed to meet the real-time performance demand of the data exchanges between MAC and PHY layers. In our implementation, we choose to transport the messages using little-endian with zero padding to 32 bits. The NVIPC library is also capable of tracing the FAPI messages and exporting them to a `pcap` file that can be analyzed offline with tools such as Wireshark.

The NVIDIA physical layer in the DU-low implements the O-RAN Open Fronthaul interface, also known as the O-RAN 7.2 interface [22], to communicate directly with the O-RU, in our case manufactured by Foxconn. This interface transports frequency domain In-phase and Quadrature (IQ) samples (with optional block floating point compression) over a switched network, allowing for flexible deployments. The interface includes synchronization, control, and user planes. The synchronization plane, or S-plane, is based on

PTPv2. We use synchronization architecture option 3 [23], where the fronthaul switch provides timing to both DU and RU. The interface also includes a management plane, although our system currently does not support it.

Table 1 summarizes the main features and operational parameters of the ARC deployment in the X5G testbed. The protocol stack is aligned with 3GPP Release 15 and uses the 5G n78 Time Division Duplexing (TDD) band and numerology 1. The DDDSU TDD pattern, which repeats every 2.5 ms, includes three downlink slots, one special slot (which is not used due to limitations in the Foxconn RUs), and an uplink slot. The uplink slot format implemented in OAI carries only two feedback bits for ACK/NACK per UE, thus allowing only the scheduling of two downlink slots per UE, eventually limiting the single UE throughput. Alternative TDD patterns, including DDDDDDSUUU and DDDDDDDSUU, repeating every 5 ms, are also already in use to provide additional ACK/NACK bits for reporting from the UEs and mitigate this limitation.

To compute the maximum theoretical cell throughput in downlink (T_{DL}) and uplink (T_{UL}), we first derive a few additional parameters from Table 1. The number of resource blocks (N_{RB}) is computed using

$$N_{RB} = \frac{\beta}{\chi \cdot \Delta f} = 273. \quad (1)$$

By default, the number of Orthogonal Frequency Division Multiplexing (OFDM) symbols per slot (N_{sym}) is 14. The number of slots per second (N_{slot}) is inversely proportional to the slot duration, which for numerology $\mu = 1$ is 0.5 ms. Hence, $N_{slot} = 1s/0.5\text{ms} = 2000$ slots/second. The maximum theoretical cell throughput for downlink and uplink is given by

$$T_{DL,UL} = N_{RB} \cdot \chi \cdot N_{sym} \cdot N_{slot} \cdot Q_m \cdot R \cdot L_{DL,UL} \cdot \eta, \quad (2)$$

where R is the effective code rate, which can approach 0.93 (as specified in the 3GPP standard [24]), and η is the fraction of time allocated for downlink or uplink operations based on the chosen TDD pattern. Considering the DDDDDDSUUU pattern to circumvent current OAI limitations on the ACK/NACK feedback bits, 60% of time is allocated for downlink and 30% for uplink since the special slot is unused due to Foxconn RU constraints. Consequently, the resulting theoretical peak cell throughput is 1.64 Gbps for downlink (T_{DL}) and 204 Mbps for uplink (T_{UL}). These values do not account for overheads typical of real networks—such as DeModulation Reference Signal (DMRS), Physical Uplink Control Channel (PUCCH), and Physical Downlink Control Channel (PDCCH)—which may further reduce net throughput. As shown by the experimental results in Section 5.5, X5G peak performance nearly reaches the theoretical downlink throughput, while the uplink is still under improvement.

TABLE 1: X5G ARC deployment main features.

Feature	Description
3GPP Release	15
Frequency Band	n78 (FR1, TDD)
Carrier Frequency	3.75 GHz
Bandwidth (β)	100 MHz
Subcarrier spacing (Δf)	30 kHz
Resource Block size (χ)	12 subcarriers
Modulation order (Q_m)	8 (256-QAM)
TDD config	DDDSU, DDDDDDSUUU*
Number of antennas used	4 TX, 4 RX
MIMO config (L_{DL}, L_{UL})	4 layers DL, 1 layer UL
Max theoretical cell throughput** (T_{DL}, T_{UL})	1.64 Gbps DL, 204 Mbps UL

*Currently the special slot is unused due to limitations in Foxconn radios.

**The single-user maximum theoretical DL throughput can currently only be reached in the DDDDDDSUUU TDD configuration. In the DDDSU TDD configuration, it is limited to 350 Mbps since we can schedule a maximum of 2 DL slots per user in one TDD period, as only 2 ACK/NACK feedback bits are available per user.

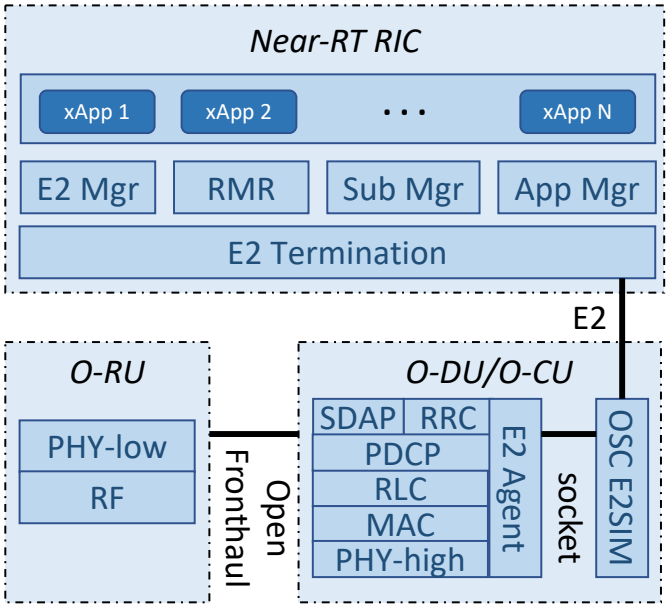


Fig. 3: Integration of the OSC Near-RT RIC in the OpenShift cluster with the X5G RAN.

2.2 Integration with the OSC Near-RT RIC

One of the key components of an O-RAN deployment is the Near-Real-Time (or Near-RT) RIC, and the intelligent applications hosted therein, namely xApps. These can implement closed-control loops at timescales between 10 ms and 1 s to provide optimization and monitoring of the RAN [25, 26]. In the current X5G setup, we deploy the "E" release of the OSC Near-RT RIC on a RedHat OpenShift cluster [4], which manages the lifecycle of edge-computing workloads instantiated as containerized applications. The Near-RT RIC and the ARC RAN are connected through the O-RAN E2 interface (see Figure 3), based on Stream Control Transmission Protocol (SCTP) and an O-RAN-defined application protocol (E2AP) with multiple service models implementing the semantic of the interface (e.g., control, reporting, etc) [3]. On the gNB side, we integrate an E2 agent based on the *e2sim* software library [27, 28], which is used to transmit the metrics collected by the OAI gNB to the RIC via the Key Performance Measurement (KPM) E2 service model. These metrics are then processed by xApps deployed on the RIC, and used to compute some control action (e.g., through AI/ML agents)

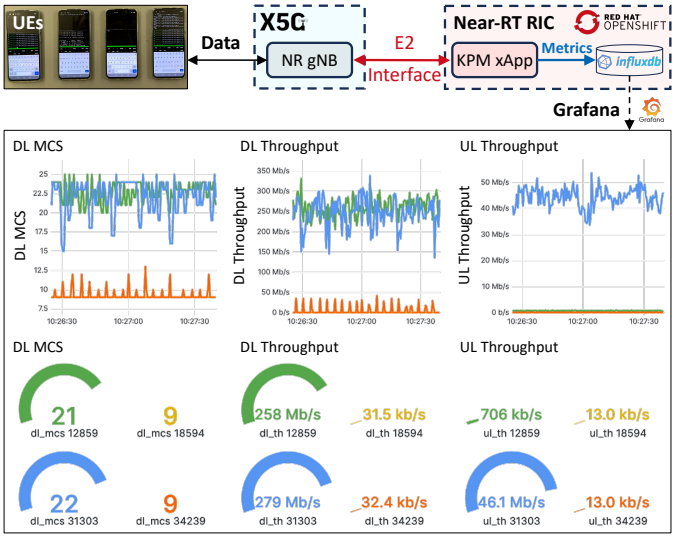


Fig. 4: KPM xApp example architecture including an X5G gNB with four connected UEs, each performing a different operation (ping, video streaming, DL test, and DL/UL tests), and a KPM xApp that pushes UE metrics into an Influx database, which are then visualized in a Grafana dashboard.

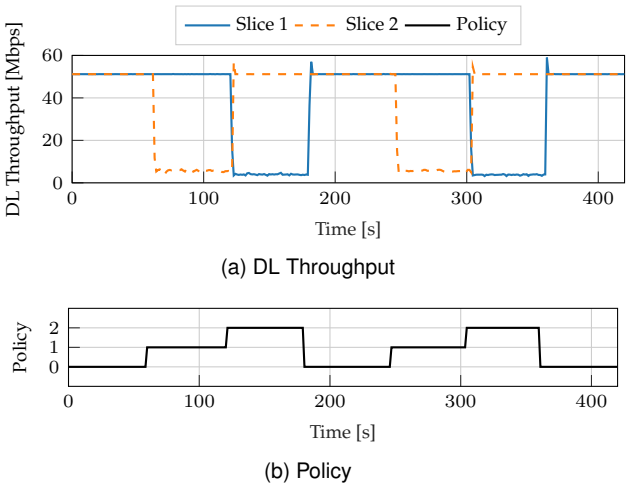


Fig. 5: Slicing xApp example showing: (a) DL throughput for two different slices, each with a single UE connected and pushing 50 Mbps of UDP traffic; (b) the network policy applied by the slicing xApp, switching between no-priority (0), prioritize slice 1 (1), and prioritize slice 2 (2).

that is sent to the RAN through the E2 interface and processed by the *e2sim* agent.

As an example, Figure 4 shows the architecture of a KPM xApp integrated with the X5G testbed. This xApp receives metrics from the E2 agent in the gNB, including throughput, number of UEs, and Reference Signal Received Power (RSRP), and stores them in an InfluxDB database [29]. The database is then queried to display the RAN performance on a Grafana dashboard [30] (see Figure 4). This setup creates a user-friendly observation point for monitoring network performance and demonstrates the effective integration of the near-RT RIC in our configuration. A tutorial on how to deploy and run this xApp in X5G or on a similar testbed can be found on the OpenRAN Gym website [31], which hosts an open-source project and framework for collaborative research in the O-RAN ecosystem [32].

The metrics collected by a KPM xApp can then be leveraged by a second xApp or an rApp to perform smart closed-loop RAN controls at runtime, based on an arbitrary optimization strategy or specific requirements. ORANSlice—an open-source, network-slicing-enabled Open RAN system that leverages open-source RAN frameworks such as OAI [16]—was successfully integrated and tested in X5G, enabling near-real-time slicing control of the resources allocated by a gNB to multiple slices of the network, according to different policies set by the network manager. Figure 5 presents the effects of various network policies applied by an ORANSlice slicing xApp in a X5G gNB. Figure 5a shows the DL throughput results for the two slices (slice 1 in blue and slice 2 in orange), with a single UE per slice connected and transmitting 50 Mbps of DL User Datagram Protocol (UDP) data, according to the policy shown in Figure 5b. The slicing xApp switches between three policies: (0) no-priority, where all slices share all resources, so both UEs achieve the target throughput of 50 Mbps; (1) prioritize slice 1: where 98% of resources are reserved for the first slice and 2% for the second one, causing the latter performance to drop to only 6 Mbps; (2) prioritize slice 2, where the opposite behavior of policy 1 is observed, with slice 1 now unable to achieve the target throughput. In this example, the policy is applied arbitrarily as a proof-of-concept for the network slicing control capabilities of X5G, while more intelligent strategies employing AI/ML components can be easily integrated into the decision process. Additional applications, including the emerging dApps [13], are currently being integrated into X5G to fully leverage its openness and programmability, further demonstrating the benefits of smart closed-loop control within the O-RAN ecosystem.

2.3 Core Network

The X5G testbed facilitates the integration and testing of different CNs from various vendors and projects. We leverage virtualization to deploy all the necessary micro-services, e.g., Access and Mobility Management Function (AMF), Session Management Function (SMF), User Plane Function (UPF), in the OpenShift cluster that also supports the Near-RT RIC. We have successfully tested and integrated the X5G RAN with two open-source core network implementations, i.e., the 5G CNs from OAI [12], as also discussed in [1], and, in this paper, also with Open5GS [33] and the CoreSIM software from Keysight [34]. As part of our ongoing efforts, we plan to incorporate additional cores, including the commercial core from A5G [35].

2.4 X5G Software Licensing and Tutorials

X5G, including the Aerial PHY, the OAI higher layers, as well as the OSC RIC, is open and can be extended with custom features and functionalities. The NVIDIA ARC framework is documented on the NVIDIA portal [20], which is accessible through NVIDIA's 6G developer program. As mentioned in Section 2.2, the step-by-step integration between the OSC RIC and the ARC stack through the X5G E2 agent is discussed in a tutorial on the OpenRAN Gym website [31, 32].

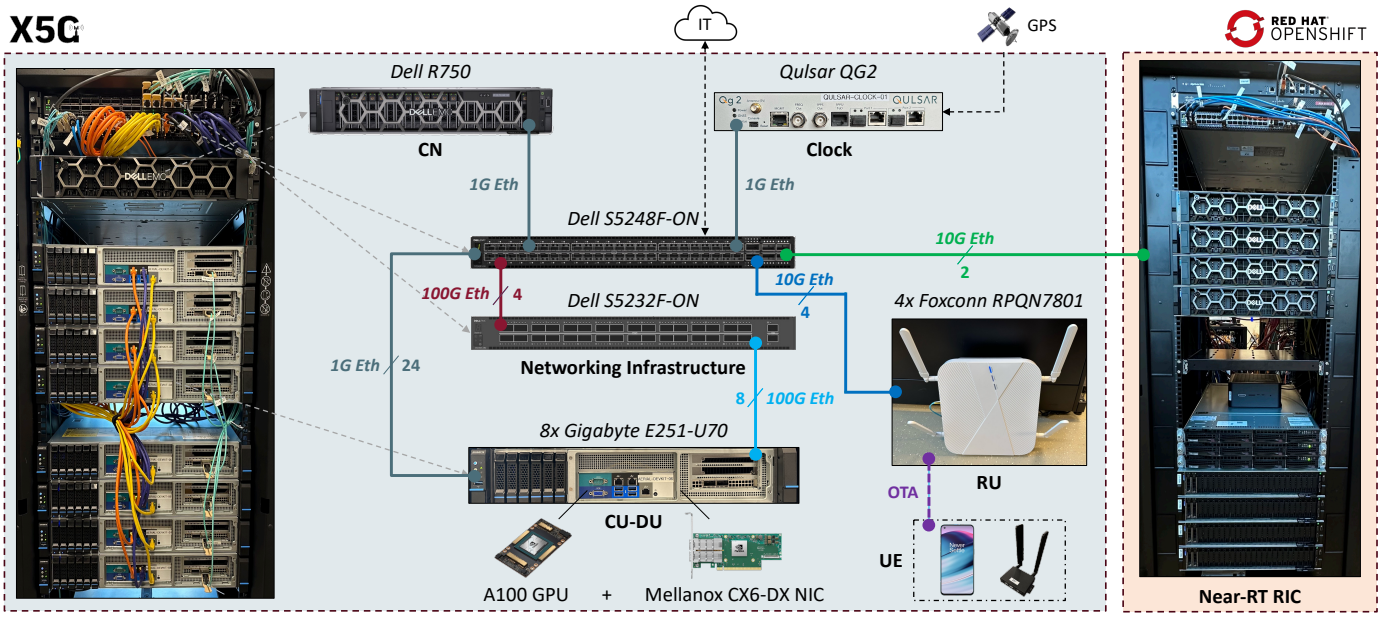


Fig. 6: Hardware and architecture infrastructure of the X5G deployment at Northeastern University.

The components implemented by OAI are published under the OAI public license v1.1 created by the OAI Software Alliance (OSA) in 2017 [36]. This license is a modified Apache v2.0 License, with an additional clause that allows contributors to make patent licenses available to third parties under Fair, Reasonable, And Non-Discriminatory (FRAND) terms, similar to 3GPP for commercial exploitation, to allow contributions from companies holding intellectual property in related areas. The usage of OAI code is free for non-commercial/academic research purposes. The Aerial SDK is available through an early adopter program [20]. The OSC software is published under the Apache v2.0 License.

3 X5G Infrastructure

This section describes the X5G physical deployment that is currently located on the Northeastern University campus in Boston, MA.[1] The deployment includes a server room with a dedicated rack for the private 5G system and an indoor laboratory open space area with benches and experimental equipment that provide a realistic Radio Frequency (RF) environment with rich scattering and obstacles. Figure 6 illustrates the hardware infrastructure that we deployed to support the X5G operations. This includes synchronization and networking infrastructures, radio nodes, eight ARC servers with integrated DU and CU, and additional compute infrastructure for the RIC and CN deployments. This infrastructure, which will be described next, has been leveraged to provide connectivity for up to eight concurrent COTS UEs, such as OnePlus smartphones (AC Nord 2003) and Sierra Wireless boards (EM9191) [37].

Synchronization Infrastructure. The synchronization infrastructure consists of a Qulsar (now VIAVI) QG-2 device acting as grandmaster clock. The QG-2 unit is connected to a GPS antenna for precise class-6 timing and generates both Precision Timing Protocol (PTP) and Synchronous Ethernet (SyncE) signals to provide frequency, phase, and time synchronization compliant with the ITU-T G.8265.1, G.8275.1, and G.8275.2 profiles. It sends the synchronization packets to the networking infrastructure through a 1 Gbps Ethernet connection. The networking infrastructure then offers full on-path support, which is necessary to distribute phase synchronization throughout the X5G platform.

Networking Infrastructure. The networking infrastructure provides connectivity between all the components of the X5G platform. It features fronthaul and backhaul capabilities through the use of two Dell switches (S5248F-ON and S5232F-ON) interconnected via four 100 Gbps cables in a port channel configuration. This configuration allows for the aggregation of multiple physical links into a single logical one to increase bandwidth and provide redundancy in case some of them fail. All switch ports are sliced into different Virtual Local Area Networks (VLANs) to allow the proper coexistence of the various types of traffic (i.e., fronthaul, backhaul, management). The Dell S5248F-ON switch primarily provides backhaul capabilities to the network and acts as a boundary clock in the synchronization plane, receiving PTP signals from the synchronization infrastructure. This switch includes 48 SFP+ ports: 12 ports are dedicated to the RUs and receive PTP synchronization packets, 10 are used to connect to the OpenShift cluster and service network, 10 are used for the out-of-band management network, and 16 connect to the CN and the Internet. Additionally, the switch includes 6 QSFP28 ports, 4 of which interconnect with the second switch. The Dell S5232F-ON switch mainly provides fronthaul connectivity to the gNBs. It includes 32 QSFP28 ports: 8 ports connect to the Mellanox cards of the ARC nodes via 100 Gbps fiber links, and 4 connect to the Dell S5248F-ON switch. The latter also acts as a boundary clock, receiving the synchronization messages from the S5232F-ON and delivering them to the gNBs.

1. X5G website: https://x5g.org.

RU. We deployed eight Foxconn RPQN 4T4R RUs, operating in the n78 band, with additional units being tested in the lab, and the Keysight RuSIM emulator. The Foxconn units have 4 externally mounted antennas, each antenna with a 5 dBi gain, and 24 dBm of transmit power. The Over-the-Air (OTA) transmissions are regulated as part of the Northeastern University Federal Communications Commission (FCC) Innovation Zone [38], with an additional transmit attenuation of 20 dB per port to comply with transmit power limits and guarantee the coexistence of multiple in-band RUs in the same environment. As we will discuss in Section 5, we leverage two of these RUs for the experimental analysis presented in this work. These RUs are deployed following RF planning procedures discussed in Section 4. Plans are in place to procure Citizen Broadband Radio Service (CBRS) RUs and deploy them in outdoor locations. RUs from additional vendors are being tested and integrated as part of our future works. Finally, we also tested and integrated the ARC stack with the Keysight RuSIM emulator, which supports the termination of the fronthaul interface on the RU side and exposes multiple RUs to the RAN stack, for troubleshooting, conformance testing, and performance testing [39].

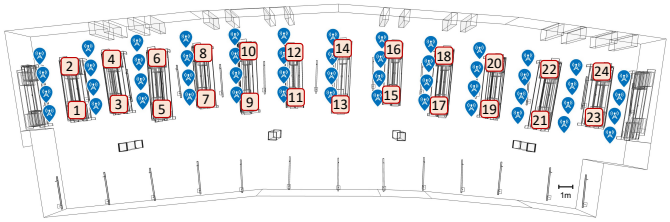
CU and DU. The 8 ARC nodes that execute the containerized CU/DU workloads are deployed on Gigabyte E251-U70 servers with 24-core Intel Xeon Gold 6240R CPU and 96 GB of RAM. The servers—which come in a half rack chassis for deployment in RAN and edge scenarios—are equipped with a Broadcom PEX 8747 Peripheral Component Interconnect (PCI) switch that enables direct connectivity between cards installed in two dedicated PCI slots without the need for interactions with the CPU. Specifically, the two PCI slots host an NVIDIA A100 GPU, which supports the computational operations of the NVIDIA Aerial PHY layer, as well as a Mellanox ConnectX-6 Dx Network Interface Card (NIC). The latter, which is used for the fronthaul interface, connects to the fronthaul part of the networking infrastructure via a QSFP28 port and 100 Gbps fiber-optic cable. In this way, the NIC can offload or receive packets directly from the GPU, thus enabling low-latency packet processing. Finally, each server is connected to the backhaul part of the networking infrastructure through three 1 Gbps Ethernet links, which provide connectivity with the OpenShift cluster (and thus the Near-RT RIC and the core networks), the management infrastructure, and the Internet. In addition to the Gigabyte servers, seven Grace Hopper (GH) machines—one of the latest NVIDIA high-computing ARM-based devices—are currently being integrated into X5G to run the ARC CU/DU worloads. Each GH combines a 72-core NVIDIA Grace CPU Superchip and an NVIDIA H100 Tensor Core GPU, linked through NVIDIA NVLink-C2C technology, which ensures seamless data sharing with up to 900 GB/s of bandwidth. It also features 480 GB of RAM, two BlueField-3 Data Processing Units (DPUs), and ConnectX-7 NICs. This configuration provides significantly higher computational capabilities compared to the Gigabyte servers, enabling the efficient support of concurrent RAN and AI/ML workloads.

Additional Compute. We leverage additional servers that are part of the OpenShift cluster and are used to instantiate the various CNs and the Near-RT RIC. The OpenShift cluster includes three Dell R740 servers acting as control-plane nodes and two Microway Navion Dual servers as worker nodes. The OpenShift rack is linked to the X5G rack through two 10 Gbps connections, one dedicated to OpenShift operations, and the other for the out-of-band management. Additionally, a Dell R750 server with 56 cores and 256 GB RAM is available for the deployment and testing of additional core network elements. This server connects to the networking infrastructure via a 1 Gbps Ethernet link and has access to the Internet through the Northeastern University network.

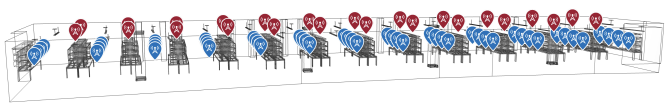
4 RF Planning with Ray-tracing

In this section, we present RF planning procedures to identify suitable locations for the RU deployment. This approach leverages an exhaustive search within a ray-tracing-based digital twin framework, with the objective of maximizing the RUs coverage while minimizing the overall interference. The study is conducted only once during the system deployment phase and remains valid as long as no significant changes occur in the environment. We perform ray-tracing in a detailed digitized representation of our indoor laboratory space in the Northeastern University ISEC building in Boston, MA, to achieve high fidelity between the real-world environment and the digital one. We carefully study how to deploy 2 RUs by considering the Signal to Interference plus Noise Ratio (SINR) between the RUs and the UEs as the objective function in the optimization problem. We limit the optimization space by using a grid of 24 possible RU locations and 52 UE test points, enabling an exhaustive search approach instead of a formal integer optimization problem, since these constraints keep the computation manageable.

First, we leverage the 3D representation of our laboratory space, created as part of the digital twin framework developed in [40] using the SketchUp modeling software. We then import the model in the MATLAB ray-tracing software and define the locations of RUs and UEs as shown in Figure 7a (from a top perspective) and in Figure 7b (from a side view). The 24 possible RUs locations (2 for each bench) are shown in red, while the 52 test points for the UEs (arranged in a 4×13 grid) are in blue. Tables 1 and 2 summarize the parameters used in our ray-tracing model. For the deployment planning purpose, we consider the RUs as transmitter nodes (TX) and the UEs as receiver ones (RX), i.e., we tailor our deployment to downlink transmissions.



(a) Site viewer top view.



(b) Site viewer side view.

Fig. 7: Site viewer with RU (red squares) and UE (blue icons) locations.

TABLE 2: Parameters of the MATLAB ray-tracing study to determine RU locations.

Parameter	Value
RU antenna spacing	0.25 m
RU antenna TX power (P_{RU})	24 dBm
RU antenna gain (G_{RU})	5 dBi
RU antenna pattern	Isotropic
RU TX attenuation (A_{RU})	$[0-50]$ dB
Set of RU locations ($\mathcal{R}$)	24 in a 2×12 grid
RU height	2.2 m
UE number of antennas	2
UE antenna spacing	0.07 m
UE antenna gain (G_{UE})	1.1 dBi
UE noise figure (F_{UE})	5 dB
Set of UEs locations ($\mathcal{U}$)	52 in a 4×13 grid
UE height	0.8 m
Environment material	Wood
Max number of reflections	3
Max diffraction order	1
Ray-tracing method	Shooting and bouncing rays

The ray-tracer generates a 24×52 matrix $\mathbf{C}$ where each entry $c_{i,j}$ corresponds to the channel information between RU_i with $i \in \mathcal{R}$, $\mathcal{R} = 1, ..., 24$, and UE_j with $j \in \mathcal{U}$, $\mathcal{U} = 1, ..., 52$. We use this to derive relevant parameters such as the thermal noise ($\mathcal{N}$) and the path loss ($\mathcal{L}$) to compute the Received Signal Strength Indicator (RSSI) $\mathcal{S}_{i,j}$ for UE_j connected to RU_i, as follows:

$$\mathcal{S}_{i,j} = P_{RU,i} + G_{RU,i} - A_{RU,i} - \mathcal{L}_{i,j} + G_{UE,j}, \quad (3)$$

where $P_{RU,i}$, $G_{RU,i}$, and $A_{RU,i}$ are the antenna TX power, gain, and attenuation of RU_i, respectively. Then, considering the linear representation of $\hat{\mathcal{S}}_{i,j}$, the SINR $\Gamma_{i,j}$ is

$$\Gamma_{i,j} = \frac{\hat{\mathcal{S}}_{i,j}}{\mathcal{N} F_{UE,i} + \sum_{u=1, u \neq i}^{M} \hat{\mathcal{S}}_{u,j}}, \quad (4)$$

where M is the number of RUs being deployed, $\mathcal{N}$ is the thermal noise, and $F_{UE,i}$ is the noise figure of UE_i. The SINR $\Gamma_{i,j}$ considers the interference to the signal from RU_i to UE_j due to downlink transmissions of all other $M-1$ RUs being deployed.

In our RF planning, we deploy two RUs (i.e., $M = 2$). In the following study, we consider scenarios where the first RU serves one UE from the test locations, while we assume that the second RU creates interference with the first, even without being assigned any UE from the list. We test all possible combinations of the 24 RU test locations, which, following the combinatorial equation of choosing 24 elements (n) in groups of 2 (r) as $C(n,r) = \frac{n!}{r!(n-r)!}$, results in a total of 276 pairs. The proposed approach for determining the optimal RU locations and the maximum average SINR ($\Phi_{\max}(\Gamma)$), called score, is presented in Algorithm 1. It takes as input the set of RU locations ($\mathcal{R}$), the set of UE test points ($\mathcal{U}$), and the SINR matrix $\mathbf{\Gamma}$. Then, it performs an exhaustive search, testing all pairs of RU against all UEs to determine the optimal RU pair (p^*, q^*) with the best maximum average SINR $\Phi_{\max}(\Gamma)$.

We test this algorithm with different values of the attenuation A_{RU}, from 0 to 50 dB in 10 dB increments. Figure 8 visualizes the normalized values of the score $\Phi(\Gamma)$ for all possible combinations of RU pairs and different attenuation values. Additionally, Table 3 provides the best RU locations including the minimum and maximum values of $\Phi(\Gamma)$ for the corresponding combinations. As expected, locations with further RUs exhibit higher average SINR values, as they are less affected by interference. However, it is important to note that the score also considers coverage, as it is computed based on the SINR. Consequently, the optimal combination of locations identifies RUs that are further apart but not necessarily the furthermost pair. Considering these results, for the experiments in Section 5, we select a TX attenuation of

Algorithm 1 Exhaustive Search Algorithm for RF Planning

Input: Set of RU locations ($\mathcal{R}$), set of UE test points ($\mathcal{U}$), precomputed SINR matrix $\mathbf{\Gamma}$
Output: Optimal RU pair (p^*, q^*) and maximum average SINR $\Phi_{\max}(\Gamma)$

Initialize $bestScore \leftarrow -\infty$ and $bestPair \leftarrow$ None
for all RU pairs $(p, q) \in \binom{\mathcal{R}}{2}$ **do**
  $sumSINR \leftarrow 0$
  for all UE $j \in \mathcal{U}$ **do**
    Compute $\Gamma_{p,j}$ (RU p serving, RU q interfering)
    Compute $\Gamma_{q,j}$ (RU q serving, RU p interfering)
    $sinrMax \leftarrow \max(\Gamma_{p,j}, \Gamma_{q,j})$
    $sumSINR \leftarrow sumSINR + sinrMax$
  end for
  $avgSINR \leftarrow sumSINR/|\mathcal{U}|$
  if $avgSINR > bestScore$ **then**
    $bestScore \leftarrow avgSINR$
    $bestPair \leftarrow (p, q)$
  end if
end for
return $bestPair, bestScore$

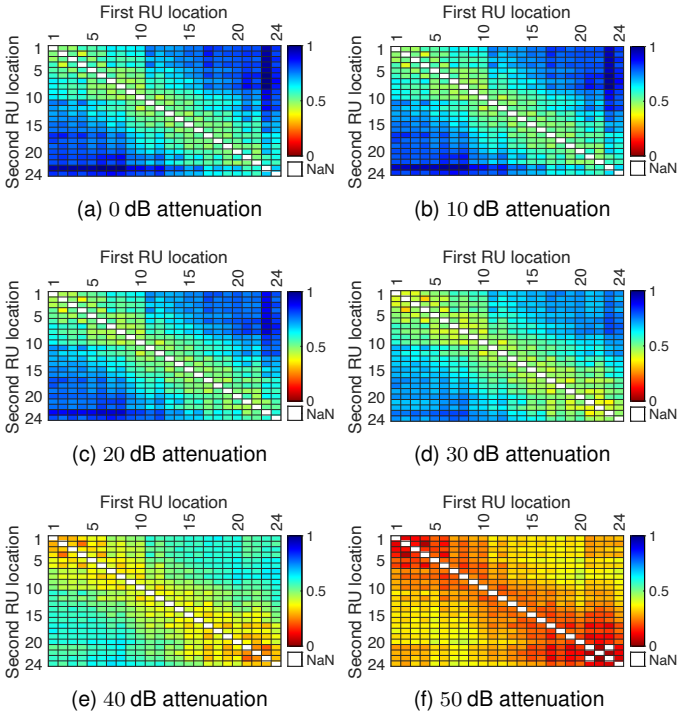


Fig. 8: Heatmap results of the normalized average SINR $\Phi(\Gamma)$ with 2 RUs.

TABLE 3: Best RUs and average SINR $\Phi(\Gamma)$ range values.

A_{RU} [dB]	RU locations with best SINR	[Min, Max] $\Phi(\Gamma)$ [dB]
0	[8, 23]	[6.08, 23.33]
10	[6, 23]	[5.71, 22.66]
20	[6, 23]	[5.00, 21.03]
30	[8, 23]	[3.58, 17.82]
40	[7, 24]	[0.19, 12.94]
50	[8, 20]	[-6.23, 6.63]

20 dB, which exhibits a good trade-off between coverage and average SINR values. Moreover, during our real-world experiments, we observed that a 20 dB attenuation leads to increased system stability and reduced degradation compared to lower attenuation values, resulting in improved overall performance, as it reduces the likelihood of saturation at the UE antenna. Therefore, we select locations [6,23] for our RUs deployment.

5 Experiment Results

In this section, we describe the design and execution of a comprehensive set of experiments that illustrate the capabilities of the X5G infrastructure in a variety of operational scenarios. We assess the adaptability of the testbed through rigorous testing, utilizing iPerf to measure network throughput and MPEG-DASH to gauge video streaming quality. The experiments are mainly conducted in the same indoor laboratory area modeled in Section 4. They include static configurations with a single UE as well as more complex setups with multiple UEs and RUs, leveraging the Keysight RuSIM emulator, and scenarios with UE mobility.

5.1 Setup Overview

We consider two different setups: (i) Gigabyte RAN servers with a 2x2 MIMO configuration (L_{DL}, L_{UL}), 2 layers DL, 1 layer UL, a DDDSU TDD pattern, and a modulation order (Q_m) up to 64-QAM (results for this setup are shown in Sections 5.2, 5.3, and 5.4); and (ii) GH RAN servers with a 4x4 MIMO configuration, 4 layers DL, 1 layer UL, a DDDDDDSUUU TDD pattern, and a Q_m up to 256-QAM (Sections 5.5 and 5.6). All experiments utilize a carrier frequency of 3.75 GHz with a bandwidth (β) of 100 MHz.

The experiments are mainly conducted in the laboratory area shown in Figure 9, which highlights the RU locations (outcome of the ray-tracing study discussed in Section 4), as well as the UE locations and the mobility pattern for the non-static experiments. All tests involving a single RU are conducted at location 6, as illustrated in Figure 9. An edge server, configured to support the iPerf and MPEG-DASH applications, is deployed within the campus network to ensure minimal latency, ranging from 1 to 2 ms. During static throughput tests, Transmission Control Protocol (TCP) backlogged traffic is transmitted first in the downlink and then in the uplink directions for 40 seconds each across different UE configurations. For video streaming, the server employs FFmpeg [41] to deliver five distinct profiles simultaneously at various resolutions—ranging from 1080P at 250 Mbps to 540P at 10 Mbps—to the UEs. On the device side, we leverage a pre-compiled iPerf3 binary for Android to generate

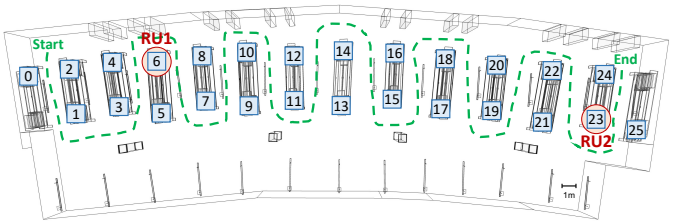


Fig. 9: Node locations considered in our experiments: RUs (red circles in 6 and 23); possible static UEs (blue squares); and mobile UEs (green dashed line).

TCP traffic, and Google's ExoPlayer for client-side video playback. Each set of experiments is replicated five times to ensure data reliability, with results including mean values and 95% confidence intervals of the metrics plotted. These metrics encompass application layer measurements such as throughput, bitrate, and rebuffer ratio, alongside MAC layer metrics like SINR, RSRP, and Modulation and Coding Scheme (MCS), collected at the OAI gNB level.

5.2 Static Experiments

1 UE, static, iPerf. In the initial series of tests, we analyze the performance of a single UE in ten static locations at varying distances from the RU, as shown in Figure 10, using the first configuration setup of 2x2 MIMO, 2 layers DL, 1 layer UL, a DDDSU TDD pattern, and a Q_m up to 64-QAM.

The iPerf throughput results in Figure 10a highlight the upper layer's responsiveness, showing a significant reduction from an average downlink throughput of 300 Mbps and an uplink one of 38 Mbps at locations near the RU, to significantly lower rates of 177 Mbps in downlink and 1.5 Mbps in uplink at the most remote point, i.e., location 18. This high-level data throughput behavior is supported by corresponding shifts in the lower layers.

For example, this trend is clearly noticeable from the results of Figure 10b, which shows the RSRP values reported by the UE to the gNB during DL (blue bars) and UL (orange bars) data transmissions. As the distance from the RU initially decreases starting from location 0 to location 6, the RSRP values peak at around -80 dBm. Subsequently, as the distance starts to increase again, moving from location 6 towards location 18, the RSRP values begin to decrease, reaching as low as -100 dBm.

In the same way, the MCS values for both downlink and uplink initially mirror each other, then begin to diverge from around location 10, as shown in Figure 10c. In uplink, the MCS values tend to remain stable or slightly increase, whereas downlink MCS values experience a slight decline possibly due to power control strategies that better favor the uplink at more distant locations.

Similarly, the CQI (Figure 10d) for both downlink and uplink starts closely matched but begins to show variation past the midpoint of the locations. The downlink CQI experiences a modest decline, while the uplink CQI sustains higher values. This suggests better channel conditions or more effective adaptation mechanisms in the uplink, due to adaptive power adjustments in uplink transmissions that maintain signal quality over distance.

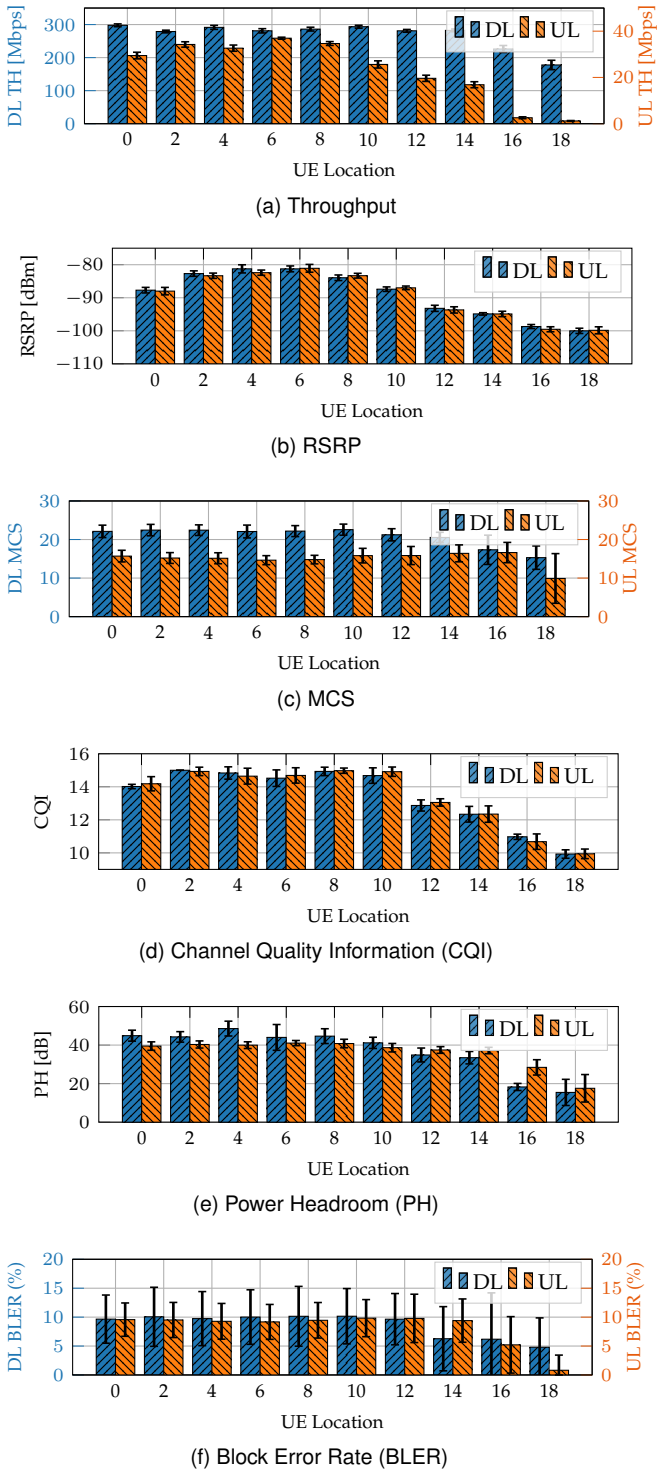


Fig. 10: Performance profiling with one UE and single RU for the static iPerf use case during DL (blue bars) and UL (orange bars) data transmissions.

PH metrics, shown in Figure 10e, reveal that downlink power headroom remains relatively stable across all locations, indicating a consistent application of power levels for downlink transmissions. In contrast, the uplink displays greater variability and generally higher values in distant locations to compensate for potential path loss and ensure that the transmit power remains adequate to maintain quality of service as the UE moves further from the RU.

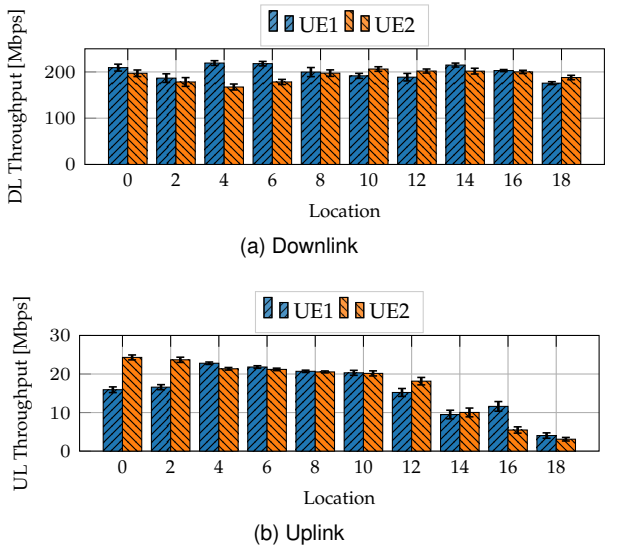


Fig. 11: Performance profiling for one RU and two static UEs for the static iPerf use case.

Finally, the BLER (Figure 10f) for downlink remains below 10% for most locations, pointing to good reliability and effective adaptation of the MCS. However, while the uplink BLER is generally low, it exhibits some peaks, especially around mid-range locations which may be related to specific lab obstacles or multipath effects and slow adaptation loops.

2 UEs, static, iPerf. We test the performance of 2 UEs for a single RU. We position the UEs at the same static locations as in the previous single UE static case. The results are plotted in Figure 11 for both downlink and uplink transmissions. We observe that, in most cases, the UEs are able to share bandwidth fairly. The best achievable average aggregate throughput from both UEs is around 400 Mbps in DL and 44 Mbps in UL. This shows that the total cell throughput can be higher than the single UE throughput. As discussed in Table 1, this is due to a limitation in the number of transport blocks that can be acknowledged in a single slot for a single UE in the case of the DDDSU TDD pattern used in this first set of experiments. Therefore, scheduling multiple UEs improves the resource utilization of the system.

1-4 UEs, static, iPerf. We further extend our evaluation to include additional tests with multiple UEs. At fixed location 4, we compare system performance with varying numbers of UEs (from 1 to 4) connected to our network. The average throughput and 95% confidence intervals are plotted in Figure 12. We observe that the UEs achieve steady throughput in all the cases, as indicated by the small confidence interval values. Additionally, the combined throughput increases with the number of UEs connected: with four UEs, the aggregate throughput reaches 512 Mbps in DL and 46 Mbps in UL. This scenario highlights the maximum throughput performance that X5G is able to achieve with the current 2x2 MIMO configuration, featuring 2 layers in DL and a DDDSU TDD pattern, ensuring a fair distribution of resources among all UEs. It is worth noting that the peak performance of X5G is detailed in Section 5.5.

2 RUs, 1 UE per RU, static, iPerf. Finally, we evaluate X5G performance with two RUs by connecting UE1 to RU1 and UE2 to RU2. RU1 is located at position 6, and RU2 is

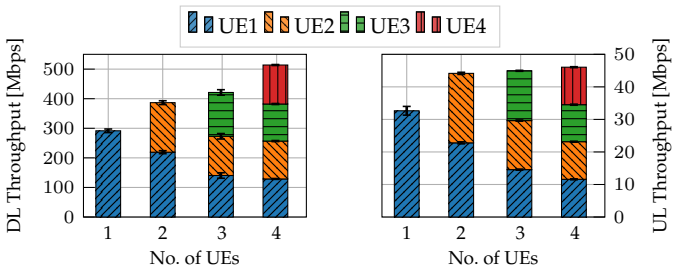


Fig. 12: Performance profiling with multiple UEs at fixed location 4 for the static iPerf use case using a DDDSU TDD pattern, a 2x2 MIMO configuration, 2 layers DL and 1 layer UL.

at position 23. We select six pairs of locations—(0,25), (2,23), (4,21), (6,19), (8,17), (10,15)—for the UEs to ensure different distances among them and the RU.

From Figure 13a, we observe that the DL throughput is significantly impacted by interference, particularly at cell edge locations. The throughput for UE1 shows a reduction of up to 90% as the UEs approach each other, while UE2 throughput decreases by up to 50%. These observations indicate that interference predominantly affects the DL direction. Conversely, as depicted in Figure 13b, UL throughput remains relatively stable across different location pairs, suggesting that UL is less susceptible to the types of interference affecting DL throughput.

Figure 14 further supports these observations by presenting additional KPIs from the MAC layer. Figure 14a shows that the MCS for both UEs decreases as the distance between the UEs diminishes, indicative of increasing interference levels. Figure 14b illustrates the RSRP, which varies in response to the UEs locations. Notably, despite adequate RSRP levels, the throughput remains low, highlighting the significant impact of interference, particularly in the DL direction.

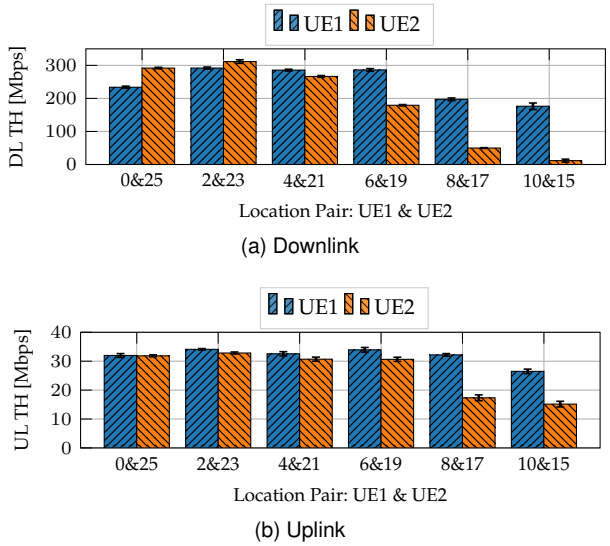


Fig. 13: Performance profiling for two RUs in the static iPerf use case, each with one assigned UE: UE1 to RU1 and UE2 to RU2.

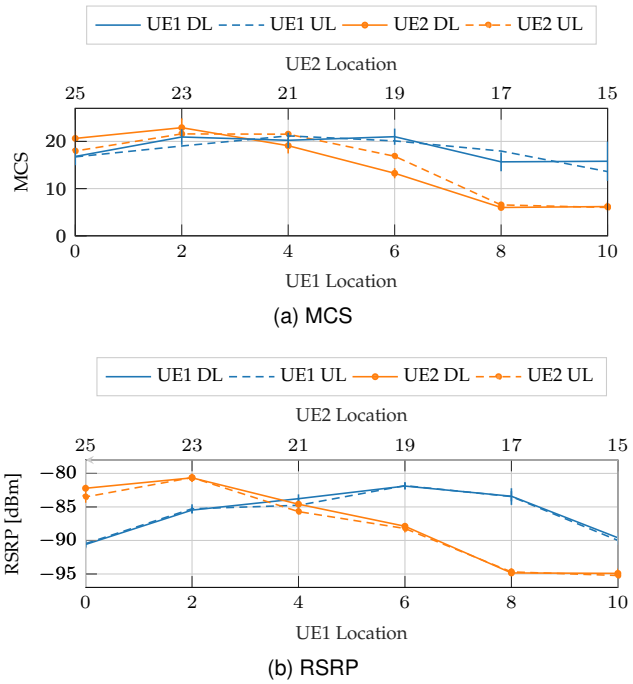
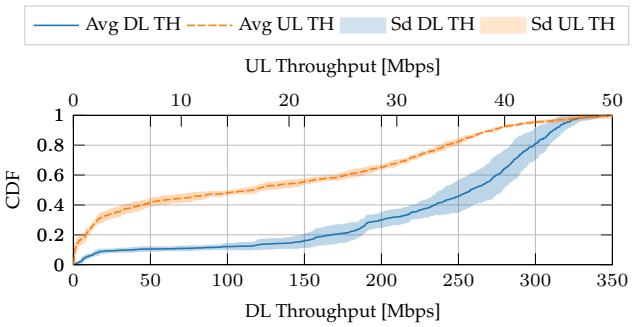


Fig. 14: MAC KPIs in the two RUs iPerf use case, each with one static UE (UE1 assigned to RU1 and UE2 to RU2): (a) averages and confidence intervals for DL MCS (solid lines) during DL data transmissions, and UL MCS (dashed lines) during UL transmissions, for UE1 (blue) and UE2 (orange); (b) averages and confidence intervals of RSRP reported by UE1 (blue) and UE2 (orange) during DL (solid lines) and UL (dashed lines) transmissions.

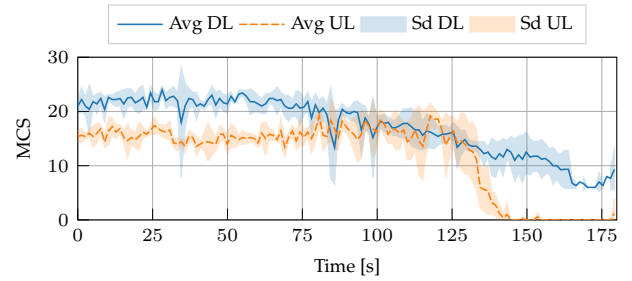
5.3 Mobile Experiments

We assess the network performance by measuring throughput as the UE follows the walking pattern around the laboratory space depicted by the dashed green line in Figure 9. The entire walk from the start to the end point spans approximately 3 minutes at regular walking speed. The mobile use case results are illustrated in Figure 15. The application layer throughput is depicted by Cumulative Distribution Function (CDF) plots in Figure 15a, where solid lines represent the averaged curve for all runs, while the shaded areas around these lines illustrate the variation across different runs, indicating the range of values within one Standard Deviation (SD) above and below the mean. The MCS and RSRP results at the MAC layer are shown in Figure 15b and Figure 15c, respectively. Also here, the average values are depicted using solid lines, while the shaded areas indicate the SD.

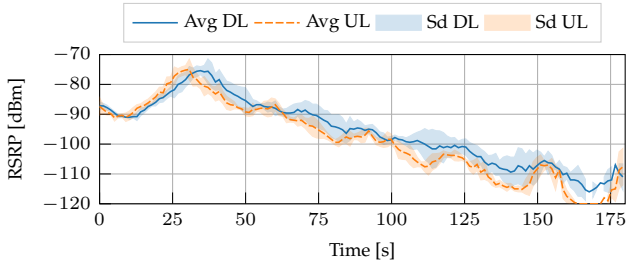
The throughput results of Figure 15a highlight notable variability in network quality influenced by mobility. Throughout the test, the UE achieves peaks of up to 350 Mbps in DL and 50 Mbps in UL. However, significant fluctuations in performance are observed, particularly as the UE moves further from the initial RU position. Figure 15b illustrates a significant drop in UL MCS values around the 100-second mark, where averages initially above 10 drop sharply, while DL MCS fluctuate more gradually until they fall below 10. This sudden decline in UL MCS at this specific time is likely due to deteriorating signal conditions, as corroborated by the corresponding RSRP trends in Figure 15c. This is most probably due to increased distance from the



(a) Throughput



(b) MCS



(c) RSRP

Fig. 15: Performance profiling with one RU and one mobile UE in the iPerf use case: (a) CDF of DL and UL throughputs with averages (solid lines) and SD (shaded areas); (b) averages (solid lines) and SD (shaded areas) of the DL MCS during DL transmissions (blue) and of the UL MCS during UL transmissions (orange); (c) averages (solid lines) and SD (shaded areas) of the RSRP reported by the UE during DL (blue) and UL (orange) data transmissions.

base station or physical obstructions, leading to a necessary reduction in MCS to maintain connectivity under compromised signal strength.

The MCS results of Figure 15b show that both DL and UL MCS values start relatively high but decrease as the UE moves further from the RU. The DL MCS exhibits more variability and sharper declines compared to the UL, which maintains a more stable profile until the final part of the walk. This suggests that the uplink benefits from more aggressive modulation and coding strategies due to 5G adaptive power control mechanisms that mitigate the impact of increasing distance and obstacles more effectively. Additionally, the RSRP data, shown in Figure 15c, indicates a gradual decline in signal strength as the UE moves along its trajectory. RSRP values for both DL (blue) and UL (orange) cases decrease over time, with the most significant drops observed after 100 seconds. This reduction in signal quality corresponds with declines in throughput and MCS, highlighting the strong dependency of these metrics on signal strength.

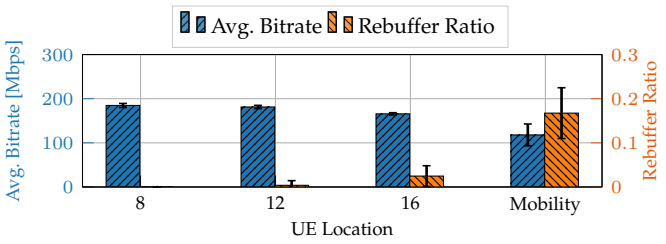


Fig. 16: Video streaming performance with one UE and single RU across both static (8—close, 12—mid, 16—far) and mobile use cases.

5.4 Video Streaming Experiments

We place the UE at three static locations at different distances from the RU: location 8 (close); 12 (mid); and 16 (far). We run each video session for three minutes, streaming five distinct profiles simultaneously at various resolutions as described in Section 5.1. We then plot the mean bitrate over five runs, as well as the rebuffer ratio, in Figure 16. As expected, the average bitrate decreases and the rebuffer ratio increases as further distances between UE and RU are considered, transitioning from close to far static locations. We observe that the UE can achieve a steady mean bitrate of around 180 Mbps in all static cases. Note that, unlike test results achieved through iPerf backlogged traffic, the mean bitrate for video streaming is lower. The video client fetches segments in an intermittent fashion (causing flows to be short), which depends on parameters, e.g., video buffer and segment size. Because of this, throughput sometimes does not increase to the fullest during that short period of time, and the client algorithm Adaptive Bitrate Streaming (ABR) downgrades the bitrate based on the estimate it gets. This is due to a slow MCS selection loop in the OAI L2, which will be improved as part of our future work. However, this shows that our setup is capable of supporting up to 8K High Dynamic Range (HDR) videos that require $150 - 300$ Mbps bitrates according to YouTube guidelines [42]. During mobility, the average bitrate is 120 Mbps, and the rebuffer ratio increases to 15%. This is once again because the UE moves away from the RU, gradually entering low-coverage regions and eventually disconnecting.

5.5 Peak Performance Experiments

In this second set of experiments, we expand our evaluation to stress-test the system and attain peak performance results. To achieve these compared to previous tests, we leverage a GH RAN server with a DDDDDDSUUU TDD pattern, a 4x4 MIMO configuration, 4 layers DL, 1 layer UL and a Q_m) up to 256-QAM. We compare system output with a single and double commercial OTA UEs connected to our network at a fixed location using Open5GS as CN and iPerf to generate traffic, as well as with the Keysight RuSIM emulator device, emulating both RU and up to 25 UEs, using Keysight CoreSIM to emulate the CN. The average throughput and 95% confidence intervals are plotted in Figure 17. We observe that in OTA at a fixed location, the UEs achieve steady throughput in all cases, as indicated by the small confidence interval values, with a peak of up to 1.05 Gbps in DL and 100 Mbps in UL for a single UE (*1-ota*).

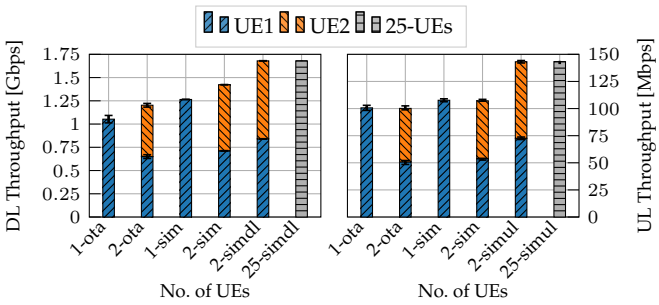


Fig. 17: Performance profiling to achieve peak network throughput, leveraging: one (*1-ota*) and two (*2-ota*) OTA UEs at a fixed location using iPerf and a DDDDDDSUUU TDD pattern; one (*1-sim*) and two (*2-sim*) emulated UEs using Keysight RuSIM and CoreSIM with a DDDDDDSUUU TDD pattern; and two (*2-simdl, 2-simul*) and twenty-five (*25-simdl, 25-simul*) emulated UEs with Keysight RuSIM and CoreSIM, a reduced number of guard symbols, a DDDDDDDSUU TDD pattern for DL cases, and a DDDSU TDD pattern for UL cases.

Furthermore, the combined throughput increases with the number of connected UEs (*2-ota*), reaching a maximum of 1.2 Gbps in DL, while remaining close to 100 Mbps in UL.

By using the Keysight RuSIM emulator with the same configuration as OTA, performance improves to over 1.26 Gbps with a single UE (*1-sim*) and 1.42 Gbps with two UEs in DL (*2-sim*), and close to 110 Mbps in UL for both one and two UEs. This performance increase can be attributed to the more controlled environment provided by RuSIM, which eliminates external interference and impairments. In this case, an *ExcellentRadioConditions* channel model—also used for Base Station (BS) conformance testing as specified in the 3GPP specifications [43]—is enabled to simulate ideal radio conditions. To achieve the current peak cell throughput, we leverage a DDDDDDDSUU TDD pattern in DL and a DDDSU pattern in UL, utilizing a reduced number of guard symbols (only one) enabled by RuSIM during two separate experiment runs with two emulated UEs. This approach results in an aggregate throughput of 1.68 Gbps in DL (*2-simdl*) and 143 Mbps in UL (*2-simul*). Moreover, we stress-test the system by simultaneously connecting up to 25 emulated UEs while exchanging traffic, achieving similar performance (*25-simdl, 25-simul*). This demonstrates that the network can reliably sustain multiple UEs and reaches its peak with two UEs, while fairly distributing resources when more devices are connected. These results highlight the maximum performance currently achievable by X5G, showcasing values comparable to those of production-level systems.

5.6 Long-running Experiments

To validate stability and reliability, we evaluate X5G through long-running experiments with a single UE performing continuous operations. The cell configuration remains the same as in Section 5.5, utilizing the DDDDDDSUUU TDD pattern, a 4x4 MIMO setup with 4 DL layers and 1 UL layer. The UE is a Samsung S23 phone, which cycles randomly every 10 minutes between three different operations:

- DL test, a 1-minute UDP downlink iPerf data test targeting 50 Mbps.
- UL test, a 1-minute UDP uplink iPerf data test targeting 10 Mbps.

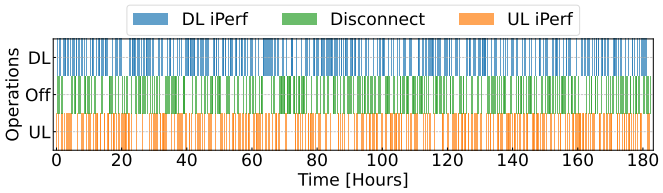


Fig. 18: Long-running stability experiment involving one UE randomly cycling for over 180 hours among three operations, repeated every 10 minutes: (blue) DL iPerf for 1 minute; (orange) UL iPerf for 1 minute; and (green) disconnection from the network for the remainder of the 10-minute cycle window.

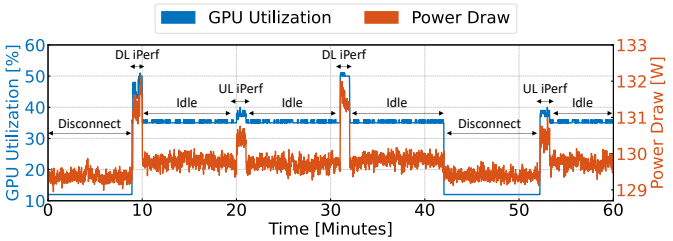


Fig. 19: GPU utilization (blue) and power draw (orange) of the NVIDIA Grace Hopper server node during a one-hour window of the long-running stability experiment. The results show the behavior of the system when the UE cycles through three operations: disconnecting for 10 minutes, performing a DL iPerf test for 1 minute, followed by 10 minutes of idling, and performing a UL iPerf test for 1 minute, followed by 10 minutes of idling.

- Disconnection, the UE disconnects from the network, remains disconnected for the remaining 10 minutes, and then reconnects.

Figure 18 shows the results of the long-running experiment, where the operations performed by the UE every 10 minutes are represented with colored bars. The system can sustain indefinite uptime, as highlighted in the figure with over 180 hours of operation before the cell was manually shut down to vacate the spectrum for other planned experiments in the area. Additionally, Figure 19 presents some of the metrics available on the RAN server side, showing the resource utilization required to run the NVIDIA ARC gNB with a single cell on a GH200 GPU. Specifically, GPU utilization and power draw are depicted for a GH server during a one-hour window of the previous long-running stability experiment. We can see how the utilization (in blue) drops to nearly 10% when no UE is connected, and rises to approximately 50% during DL data traffic, reflecting the system's computation demand. On the other hand, during idle periods and UL communication, GPU utilization remains stable between 35% and 40%, respectively. The power draw (in orange) follows a similar trend, ranging from 129 to 132 W. It is important to note that these results apply to a single cell, but the resource requirements for multiple cells do not scale linearly. Each GH server can support up to 20 cells [17] while maintaining a high-level of energy efficiency for RAN communications [44]. Overall, these results highlight the high reliability of X5G in terms of both performance and stability, positioning it as a suitable candidate for P5G deployments, as well as a valuable playground to develop, test, and evaluate novel AI/ML algorithms and solutions for the RAN.

6 Related Work

This section compares the features and capabilities of the X5G testbed within the context of similar programmable open RAN and 5G, highlighting its unique features and contributions beyond 5G research and experimentation. Surveys of testbeds for open and programmable wireless networks can also be found in [3, 45].

The Platforms for Advanced Wireless Research (PAWR) [46] offers a set of geographically and technically diverse testbeds designed to enhance specific wireless communication areas. These include POWDER, AERPAW, COSMOS, ARA, and Colosseum, each equipped with specialized technologies to address varied research needs.

The POWDER facility, located at the University of Utah in Salt Lake City, UT, supports a wide spectrum of research areas, including next-generation wireless networks and dynamic spectrum access [47]. Its 5G stack is based primarily on a combination of open-source stacks, combined with Software-defined Radios (SDRs) or RUs but not accelerated at the physical layer, and on a commercial Mavenir system, which does not support access to the source code from the PHY to the core network, differently from the X5G stack.

Similarly, AERPAW, deployed on the campus of North Carolina State University in Raleigh, NC, focuses on aerial and drone communications, diverging from our emphasis on private 5G network configurations [48]. The AERPAW facility hosts an Ericsson 5G deployment with similar limitations with respect to stack programmability for research use cases.

The COSMOS project [49] leverages an array of programmable and software-defined radios, including USRP and Xilinx RFSoC boards, to facilitate mmWave communication experiments across a city-scale environment. The outdoor facilities of COSMOS are deployed in the Harlem area, in New York City, while its indoor wireless facilities are on the Rutgers campus in North Brunswick, NJ. Unlike X5G, COSMOS is designed for broad academic and industry use and is more focused on mmWave deployments enabling diverse external contributions to its development without specific emphasis on any single network architecture.

The ARA testbed [50], deployed across Iowa State University (ISU), in the city of Ames, and surrounding rural areas in central Iowa, serves as a large-scale platform for advanced wireless research tailored to rural settings. ARA includes diverse wireless platforms ranging from low-UHF massive MIMO to mmWave access, long-distance backhaul, free-space optical, and Low Earth Orbit (LEO) satellite communications, utilizing both SDR and programmable COTS platforms and leveraging open-source software like OAI, srsRAN, and SD-RAN [51]. However, unlike the X5G testbed, ARA focuses primarily on rural connectivity without focusing on specialized hardware for PHY layer optimization or digital twin frameworks for RF planning.

Colosseum is the world's largest Open RAN digital twin [40, 52]. This testbed allows users to quickly instantiate softwarized cellular protocol stacks, e.g., the OAI one, on its 128 compute nodes. These nodes control 128 SDRs that are used as RF front-ends and are connected to a massive channel emulator, which enables experimentation in a variety of emulated RF environments. However, the Colosseum servers are not equipped to offload lower-layer cellular operations on GPUs, and the available SDRs are USRP X310 from NI, instead of commercial RUs.

The OSC is also involved in the creation of laboratory facilities [53] that comply with O-RAN standards and support the testing and integration of O-RAN-compliant components. These testing facilities, distributed across multiple laboratories, foster a diverse ecosystem through their commitment to open standards and collaborative development. However, unlike X5G, they do not explicitly focus on the deployment complexities of private networks, nor do they provide any PHY layer acceleration technology or utilize a digital twin for RF planning. Instead, they aim to promote multi-vendor interoperability within an open collaborative framework.

6G-SANDBOX [54] is a versatile facility that includes four geographically displaced platforms in Europe, each equipped to support a variety of advanced wireless technologies and experimental setups. It uses a mix of commercial solutions (for example, Nokia microcells, Ericsson Base Band Unit (BBU), and the Amarisoft stack) and open source solutions (for example, OAI and srsRAN) in diverse environments ranging from urban to rural settings. Unlike X5G, 6G-SANDBOX primarily facilitates wide-ranging 6G research through its extensive, multi-location infrastructure. Its predecessor, 5GENESIS [55], featured a modular and flexible experimentation methodology, supporting both per-component and end-to-end (E2E) validation of 5G technologies and KPI across five European locations. This testbed emphasizes a comprehensive approach to 5G performance assessment, integrating diverse technologies such as Software-defined Networking (SDN), Network Function Virtualization (NFV), and network slicing to enable rigorous testing of vertical applications but not including O-RAN architectures.

The Open AI Cellular (OAIC) testbed [56], developed at Virginia Tech, is an open-source 5G O-RAN-based platform designed to facilitate AI-based RAN management algorithms. It includes the OAIC-Control framework for designing AI-based RAN controllers and the OAIC-Testing framework for automated testing of these controllers. The OAIC testbed introduces a new real-time RIC, zApps, and a Z1 interface to support use cases requiring latency under 10 ms, integrated with the CORNET infrastructure for remote accessibility.

The CCI xG Testbed provides a comprehensive platform for advanced wireless research, particularly in the realm of 5G and beyond. It features a disaggregated architecture with multiple servers distributed across geographically disparate cloud sites, leveraging a combination of central and edge cloud infrastructures to optimize resource allocation and latency. The testbed includes several SDR-based CBRS Base Station Device (CBSD) integrated with an open-source Spectrum Access System (SAS) for dynamic spectrum sharing in the CBRS band [57]. Additionally, the testbed supports a full O-RAN stack using srsRAN and Open5GS and features both non-RT RIC and near-RT RIC for real-time and non-real-time radio resource management [58, 59].

The testbed in [60] provides a prototypical environment designed to experiment with vRAN deployments and evaluate resource allocation and orchestration algorithms. It focuses on the decoupling of radio software components from hardware to facilitate efficient and cost-effective RAN de-

ployments. This testbed includes datasets that characterize computing usage, energy consumption, and application performance, which are made publicly available to foster further research. Unlike the X5G testbed, the O-RAN platform primarily addresses the flexibility and cost efficiency of virtualized RANs without incorporating specialized hardware for PHY layer tasks.

The disaggregated 5G testbed for live audio production [61] emphasizes ultra-reliable low-latency communication for media applications. Its scope is narrower than X5G, which supports a broader range of experimental scenarios and computationally intensive network configurations.

The data usage control framework [62] addresses privacy challenges in hybrid private-public 5G networks. While it highlights the importance of secure orchestration and policy management, its focus on analytics differs from X5G capabilities in physical layer acceleration and network performance experimentation.

Finally, the Microsoft enterprise-scale Open RAN testbed [63] highlights the potential of virtualized RAN functions on commodity servers, employing disaggregated architectures to demonstrate scalability and flexibility. By integrating Kubernetes for dynamic orchestration and using Intel FlexRAN with ACC100 accelerators for Low-Density Parity-Check (LDPC) look-aside offloading, this testbed achieves functional disaggregation of RAN workloads.

While state-of-the-art software stacks such as srsRAN already offer similar performance in terms of core 5G functionalities, including handovers, X5G distinguishes itself through its integration of GPU acceleration, enabling enhanced flexibility and computational power for future innovations. Unlike traditional platforms that rely on CPU-based architectures, which achieve performance parity for standard RAN tasks, it leverages GPUs not only for optimized PHY processing but also as a unified platform for AI/ML workloads. Indeed, the GPU architecture of X5G supports the development and deployment of dApps [13] that utilize AI/ML models for real-time network optimization. This capability aligns directly with the vision outlined by the AI-RAN Alliance [64], which emphasizes the integration of AI-driven decision-making processes across three key development areas: (i) AI-for-RAN, (ii) AI-and-RAN, and (iii) AI-on-RAN, making our platform an ideal candidate for advancing these areas. Moreover, the modular design of X5G guarantees compatibility with both open-source and commercial cores, facilitating future experiments with advanced technologies like massive MIMO, mmWave, and beamforming that are currently under development.

7 Conclusions and Future Work

We introduced X5G, an open, programmable, and multi-vendor private 5G O-RAN testbed deployed at Northeastern University in Boston, MA. We demonstrated the integration of NVIDIA Aerial, a PHY layer implementation on GPUs, with higher layers based on OAI, resulting in the creation of the NVIDIA ARC platform. We provided an overview of the ARC software and hardware implementations, designed for a multiple-node deployment, including a Red Hat OpenShift cluster for the OSC RIC deployment, as well as examples of a KPM xApp and a slicing xApp. Additionally, we conducted a ray-tracing study using our digital twin framework to determine the optimal placement of X5G RUs. Finally, we discussed platform performance with varying numbers of COTS and emulated UEs and applications, such as iPerf and video streaming, as well as through long-running and stress-test experiments to evaluate its stability.

Next, we plan to continue the deployment of X5G gNBs comprising a mix of indoor and outdoor locations for more realistic experiments and comprehensive development of UE handover procedures. We are targeting the integration of RUs from different vendors and supporting bands for 5G New Radio (NR) Frequency Range 2 (FR2). We will also develop pipelines for the automatic deployment, testing, and management of workloads leveraging the Red Hat OpenShift cluster already in use for the OSC RIC integration. Our aims include (i) deploying a fully functional and reliable private 5G network that remains continuously up and running, providing an infrastructure for users to operate on and for researchers to collect realistic datasets, and (ii) enabling full RAN control to facilitate dynamic changes in network behavior by enhancing the capabilities of X5G, thereby offering the research community an end-to-end open and programmable platform for the development and testing of next-generation wireless networks and algorithms.

References

[1] D. Villa, I. Khan, F. Kaltenberger, N. Hedberg, R. S. da Silva, A. Kelkar, C. Dick, S. Basagni, J. M. Jornet, T. Melodia, M. Polese, and D. Koutsonikolas, "An Open, Programmable, Multi-vendor 5G O-RAN Testbed with NVIDIA ARC and OpenAirInterface," in *Proceedings of the 2nd Workshop on Next-generation Open and Programmable Radio Access Networks (NG-OPERA)*, May 2024.

[2] A. Narayanan, M. I. Rochman, A. Hassan, B. S. Firmansyah, V. Sathya, M. Ghosh, F. Qian, and Z.-L. Zhang, "A Comparative Measurement Study of Commercial 5G mmWave Deployments," in *IEEE Conference on Computer Communications*, May 2022.

[3] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, Second quarter 2023.

[4] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "NeutRAN: An Open RAN Neutral Host Architecture for Zero-Touch RAN and Spectrum Sharing," *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, pp. 5786–5798, 2024.

[5] S. Pongratz, "Open RAN Market Opportunity and Risks," Dell'Oro Group report, November 2021.

[6] F. Kaltenberger, A. P. Silva, A. Gosain, L. Wang, and T.-T. Nguyen, "OpenAirInterface: Democratizing innovation in the 5G Era," *Computer Networks*, vol. 176, p. 107284, July 2020.

[7] I. Gomez-Miguelez, A. Garcia-Saavedra, P. D. Sutton, P. Serrano, C. Cano, and D. J. Leith, "SrsLTE: An Open-Source Platform for LTE Evolution and Experimentation," in *Proceedings of ACM WiNTECH*, NYC, New York, October 2016.

[8] Meticolous Research. 5G Testing Market. [Online]. Available: https://www.meticulousresearch.com/product/5g-testing-market-5482

[9] P. Bahl, M. Balkwill, X. Foukas, A. Kalia, D. Kim, M. Kotaru, Z. Lai, S. Mehrotra, B. Radunovic, S. Saroiu, C. Settle, A. Verma, A. Wolman, F. Y. Yan, and Y. Zhang, "Accelerating Open RAN Research Through an Enterprise-Scale 5G Testbed," in *Proceedings of ACM MobiCom '23*, Madrid, Spain, October 2023.

[10] B. Tang, V. K. Shah, V. Marojevic, and J. H. Reed, "AI Testing Framework for Next-G O-RAN Networks: Requirements, Design, and Research Opportunities," *IEEE Wireless Communications*, vol. 30, no. 1, pp. 70–77, February 2023.

[11] C. Fiandrino, L. Bonati, S. D'Oro, M. Polese, T. Melodia, and J. Widmer, "EXPLORA: AI/ML EXPLainability for the Open RAN," vol. 1, pp. 1–26, December 2023.

[12] F. Kaltenberger, T. Melodia, I. Ghauri, M. Polese, R. Knopp, T. T. Nguyen, S. Velumani, D. Villa, L. Bonati, R. Schmidt, S. Arora,

M. Irazabal, and N. Nikaein, "Driving Innovation in 6G Wireless Technologies: The OpenAirInterface Approach," *arXiv:2412.13295 [cs.NI]*, pp. 1–30, December 2024.
[13] A. Lacava, L. Bonati, N. Mohamadi, R. Gangula, F. Kaltenberger, P. Johari, S. D'Oro, F. Cuomo, M. Polese, and T. Melodia, "dApps: Enabling Real-Time AI-Based Open RAN Control," *arXiv:2501.16502 [cs.NI]*, pp. 1–31, January 2025.
[14] J. Groen, S. D'Oro, U. Demir, L. Bonati, D. Villa, M. Polese, T. Melodia, and K. Chowdhury, "Securing O-RAN Open Interfaces," *IEEE Transactions on Mobile Computing*, vol. 23, pp. 1–13, December 2024.
[15] J. Groen, S. Di Valerio, I. Karim, D. Villa, Y. Zhang, L. Bonati, M. Polese, S. D'Oro, T. Melodia, E. Bertino, F. Cuomo, and K. Chowdhury, "TIMESAFE: Timing Interruption Monitoring and Security Assessment for Fronthaul Environments," *arXiv:2412.13049 [cs.NI]*, pp. 1–13, December 2024.
[16] H. Cheng, S. D'Oro, R. Gangula, S. Velumani, D. Villa, L. Bonati, M. Polese, T. Melodia, G. Arrobo, and C. Maciocco, "ORANSlice: An Open Source 5G Network Slicing Platform for O-RAN," in *Proceedings of ACM Workshop on Open and AI RAN*, Washington, DC, USA, November 2024.
[17] Fujitsu. https://www.fujitsu.com/global/about/resources/news/press-releases/2024/1113-01.html. Accessed December 2024.
[18] Small Cell Forum, "5G FAPI: PHY API Specification," techreport 222.10.02, March 2020.
[19] A. Kelkar and C. Dick, "NVIDIA Aerial GPU Hosted AI-on-5G," in *IEEE 4th 5G World Forum (5GWF)*, October 2021, pp. 64–69.
[20] NVIDIA Corporation. Aerial CUDA-Accelerated RAN. Accessed June 2024. [Online]. Available: https://docs.nvidia.com/aerial/cuda-accelerated-ran/index.html
[21] Small Cell Forum, "5G nFAPI specifications," techreport 225.2.1, November 2021.
[22] O-RAN Alliance, "O-RAN Working Group 4 (Open Fronthaul Interfaces WG) Control, User and Synchronization Plane Specification," techreport O-RAN.WG4.CUS.0-R003-v12.00, June 2023.
[23] O-RAN Working Group 4. (2021, July) O-RAN Fronthaul Control, User and Synchronization Plane Specification 7.0. ORAN-WG4.CUS.0-v07.00 Technical Specification.
[24] ETSI, "5G; NR; User Equipment (UE) radio access capabilities," 3GPP TR 38.306 version 17.0.0 Release 17, 2022.
[25] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward Next Generation Open Radio Access Networks: What O-RAN Can and Cannot Do!" *IEEE Network*, vol. 36, no. 6, pp. 206–213, November 2022.
[26] F. Mungari, "An RL Approach for Radio Resource Management in the O-RAN Architecture," in *18th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, July 2021.
[27] E. Moro, M. Polese, A. Capone, and T. Melodia, "An Open RAN Framework for the Dynamic Control of 5G Service Level Agreements," in *IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, Dresden, Germany, November 2023.
[28] O-RAN Software Community. (2022) sim-e2-interface repository. https://github.com/o-ran-sc/sim-e2-interface.
[29] InfluxData. InfluxDB. https://www.influxdata.com. Accessed May 2024.
[30] Grafana Labs. Grafana Dashboard. https://grafana.com. Accessed May 2024.
[31] OpenRAN Gym. OpenRAN Gym Website. https://openrangym.com/. Accessed May 2024.
[32] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "OpenRAN Gym: AI/ML Development, Data Collection, and Testing for O-RAN on PAWR Platforms," *Computer Networks*, vol. 220, p. 109502, January 2023.
[33] Open5GS. https://open5gs.org. Accessed May 2024.
[34] Keysight. P8850S CoreSIM — Core Simulation RAN Solutions. Accessed June 2024. [Online]. Available: https://www.keysight.com/it/en/product/P8850S/coresim-core-simulation-ran-solutions.html
[35] A5G Networks. https://a5gnet.com. Accessed May 2024.
[36] OpenAirInterface Software Alliance. Accessed June 2024. [Online]. Available: https://openairinterface.org
[37] Sierra Wireless. EM9191 5G NR Sub-6 GHz Module. https://www.sierrawireless.com/iot-modules/5g-modules/em9191/.
[38] FCC: Engineering & Technology Bureau. (2021, August) FCC Designates New Innovation Zones For Advanced Wireless Technology Research And Innovation. Public Notice: FCC-21-92.
[39] Keysight. P8822S RuSIM – UE / O-RU Emulation Over the O-RAN Fronthaul. Accessed June 2024. [Online]. Available: https://www.keysight.com/it/en/product/P8822S/rusim-over-o-ran-fronthaul.html
[40] D. Villa, M. Tehrani-Moayyed, C. P. Robinson, L. Bonati, P. Johari, M. Polese, and T. Melodia, "Colosseum as a Digital Twin: Bridging Real-World Experimentation and Wireless Network Emulation," *IEEE Transactions on Mobile Computing*, pp. 1–17, January 2024.
[41] *FFMPEG*. [Online]. Available: https://ffmpeg.org/ffmpeg.html
[42] *Guidelines for Video Bitrates*. [Online]. Available: https://support.google.com/youtube/answer/1722171?hl=en#zippy=%2Cbitrate
[43] ETSI, "NR, Base Station (BS) conformance testing, Part 2: Radiated conformance testing," 3GPP TS 38.141-2 version 16.10.0 Release 16, December 2021.
[44] L. Kundu, X. Lin, and R. Gadiyar, "Towards Energy Efficient RAN: From Industry Standards to Trending Practice," 2024. [Online]. Available: https://arxiv.org/abs/2402.11993
[45] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "Open, Programmable, and Virtualized 5G Networks: State-of-the-Art and the Road Ahead," *Computer Networks*, vol. 182, pp. 1–28, December 2020.
[46] Platforms for Advanced Wireless Research (PAWR). https://www.advancedwireless.org. Accessed June 2024.
[47] J. Breen, A. Buffmire, J. Duerig, K. Dutt, E. Eide, M. Hibler, D. Johnson, S. K. Kasera, E. Lewis, D. Maas, A. Orange, N. Patwari, D. Reading, R. Ricci, D. Schurig, L. B. Stoller, J. Van der Merwe, K. Webb, and G. Wong, "POWDER: Platform for Open Wireless Data-driven Experimental Research," in *Proceedings of ACM WiNTECH*, New York, NY, USA, 2020.
[48] A. Panicker, O. Ozdemir, M. L. Sichitiu, I. Guvenc, R. Dutta, V. Marojevic, and B. Floyd, "AERPAW emulation overview and preliminary performance evaluation," *Computer Networks*, vol. 194, p. 108083, July 2021.
[49] T. Chen, P. Maddala, P. Skrimponis, J. Kolodziejski, A. Adhikari, H. Hu, Z. Gao, A. Paidimarri, A. Valdes-Garcia, M. Lee, S. Rangan, G. Zussman, and I. Seskar, "Open-access millimeter-wave software-defined radios in the PAWR COSMOS testbed: Design, deployment, and experimentation," *Computer Networks*, vol. 234, p. 109922, October 2023.
[50] H. Zhang, Y. Guan, A. Kamal, D. Qiao, M. Zheng, A. Arora, O. Boyraz, B. Cox, T. Daniels, M. Darr, D. Jacobson, A. Khokhar, S. Kim, J. Koltes, J. Liu, M. Luby, L. Nadolny, J. Peschel, P. Schnable, A. Sharma, A. Somani, and L. Tang, "ARA: A Wireless Living Lab Vision for Smart and Connected Rural Communities," in *Proceedings of ACM WiNTECH*, New York, NY, USA, April 2021.
[51] J. O. Boateng, T. Zhang, G. Zu, T. U. Islam, S. Babu, H. Zhang, and D. Qiao, "AraSDR: End-to-End, Fully-Programmable Living Lab for 5G and Beyond," in *IEEE International Conference on Communications (ICC)*, Denver, CO, USA, June 2024.
[52] M. Polese, L. Bonati, S. D'Oro, P. Johari, D. Villa, S. Velumani, R. Gangula, M. Tsampazi, C. P. Robinson, G. Gemmi, A. Lacava, S. Maxenti, H. Cheng, and T. Melodia, "Colosseum: The Open RAN Digital Twin," *arXiv:2404.17317 [cs.NI]*, pp. 1–13, April 2024.
[53] F. Bimo, F. Feliana, S. Liao, C. Lin, D. F. Kinsey, J. Li, R. Jana, R. Wright, and R. Cheng, "OSC Community Lab: The Integration Test Bed for O-RAN Software Community," in *IEEE Future Networks World Forum (FNWF)*, Los Alamitos, CA, USA, October 2022.
[54] 6G-SANDBOX. 6G-SANDBOX: The Experimental Facility for Advanced Wireless Research. https://6g-sandbox.eu/. Accessed May 2024.
[55] A. Diaz Zayas, G. Caso, Ö. Alay, P. Merino, A. Brunstrom, D. Tsolkas, and H. Koumaras, "A Modular Experimentation Methodology for 5G Deployments: The 5GENESIS Approach," *Sensors*, vol. 20, no. 22, 2020.
[56] P. S. Upadhyaya, N. Tripathi, J. Gaeddert, and J. H. Reed, "Open AI Cellular (OAIC): An Open Source 5G O-RAN Testbed for Design and Testing of AI-based RAN Management Algorithms," *IEEE Network*, vol. 37, no. 5, pp. 7–15, September 2023.
[57] A. Tripathi, J. S. Mallu, M. H. Rahman, A. Sultana, A. Sathish, A. Huff, M. R. Chowdhury, and A. P. Da Silva, "End-to-End O-RAN Control-Loop For Radio Resource Allocation in SDR-Based 5G Network," in *IEEE Military Communications Conference (MILCOM)*, December 2023.
[58] O. R. Collaco, M. R. Chowdhury, A. P. da Silva, and L. DaSilva, "Enabling CBRS Experimentation through an OpenSAS and SDR-based CBSD," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, August 2023.

[59] G. Z. Bruno, G. M. Almeida, A. Sathish, A. P. da Silva, L. A. D. A. Huff, K. V. Cardoso, and C. B. Both, "Evaluating the Deployment of a Disaggregated Open RAN Controller On a Distributed Cloud Infrastructure," *IEEE Transactions on Network and Service Management*, pp. 1–13, April 2024.
[60] J. X. Salvat, J. A. Ayala-Romero, L. Zanzi, A. Garcia-Saavedra, and X. Costa-Perez, "Open Radio Access Networks (O-RAN) Experimentation Platform: Design and Datasets," *IEEE Communications Magazine*, vol. 61, no. 9, pp. 138–144, June 2023.
[61] J. Dürre, N. Werner, P. Matzakos, R. Knopp, A. Garcia, and C. Avellan, "A Disaggregated 5G Testbed for Professional Live Audio Production," in *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2022, pp. 1–6.
[62] H. Zafar, U. Fattore, F. Cirillo, and C. J. Bernardos, "Data Usage Control for Privacy-Enhanced Network Analytics in Private 5G Networks," *IEEE Open Journal of the Communications Society*, pp. 1–1, 2024.
[63] P. Bahl, M. Balkwill, X. Foukas, A. Kalia, D. Kim, M. Kotaru, Z. Lai, S. Mehrotra, B. Radunovic, S. Saroiu, C. Settle, A. Verma, A. Wolman, F. Y. Yan, and Y. Zhang, "Accelerating Open RAN Research Through an Enterprise-scale 5G Testbed," ser. ACM MobiCom '23. New York, NY, USA: Association for Computing Machinery, 2023.
[64] AI-RAN Alliance. https://ai-ran.org/wp-content/uploads/2024/12/AI-RAN_Alliance_Whitepaper.pdf. Accessed December 2024.

Davide Villa received his B.S. in Computer Engineering from the University of Pisa, Italy, in 2015, and his M.S. in Embedded Computing Systems from Sant'Anna School of Advanced Studies and the University of Pisa, Italy, in 2018. He worked as a Research Scientist in the Embedded Systems and Network Group at United Technologies Research Center in Cork, Ireland, from 2018 to 2020. He is currently pursuing a Ph.D. in Computer Engineering at the Institute for the Wireless Internet of Things at Northeastern University in Boston, USA. His research interests include 5G and beyond cellular networks, channel characterization for wireless systems, O-RAN, and software-defined networking for experimental wireless testbeds.

Imran Khan received his B.S. in Electrical Engineering from Bangladesh University of Engineering and Technology, in 2014, and his M.S. in Computer Engineering from Southern Illinois University of Carbondale, USA, in 2020. He is currently pursuing a Ph.D. in Computer Engineering at Northeastern University in Boston, USA. His research interest revolves around ensuring comprehensive performance, seamless mobility, and dependable reliability in 5G/6G networks.

Florian Kaltenberger is an Associate Professor in the Communication Systems department at EURECOM (France). He received his Diploma degree (Dipl.-Ing.) and his Ph.D. both in Technical Mathematics from the Vienna University of Technology in 2002 and 2007 respectively. He is part of the management team for the real-time open-source 5G platform OpenAirInterface.org where he is coordinating the developments of the OAI radio access network project group, which delivered support for 5G non-standalone access in 2020 and for 5G standalone access in 2021. Florian is currently on sabbatical at Northeastern University, where he is working on bringing different open-source communities around 5G and open RAN together to build an end-to-end reference architecture for 6G research.

Nicholas Hedberg is a Senior Engineer in the Public Sector at NVIDIA in Zurich, Switzerland, having joined in October 2021. He brings extensive experience from his previous tenure at Viasat Inc., where he was a System Engineering Team Lead in Lausanne, focusing on cutting-edge phased array antennas for satellite communications. Prior roles at Viasat in Carlsbad involved significant contributions to mobile terminal Verilog modules and ASIC development. Nicholas holds a BS in Physics and a BA in Economics from UC San Diego (2003-2007).

Rúben Soares da Silva received his degree in Computer Engineering from Castelo Branco Polytechnic Institute (IPCB) in 2018. Having joined Allbesmart in 2021, began work for OpenAirInterface Software Alliance in integrating the OAI L2 with the NVIDIA Aerial L1 following Small Cell Forums' FAPI standard, and since continued development in support of the FAPI split, as well as continued support for the deployments using this L2/L1 integration.

Stefano Maxenti is a Ph.D. Candidate in Computer Engineering at the Institute for the Wireless Internet of Things (WIoT) at Northeastern University, under Prof. Tommaso Melodia. He received a B.Sc. in Engineering of Computing Systems in 2020 and a M.Sc. in Telecommunication Engineering in 2023 from Politecnico di Milano, Italy. His research is linked with AI applications for wireless communications and orchestration, integration, and automation of O-RAN networks

Leonardo Bonati is an Associate Research Scientist at the Institute for the Wireless Internet of Things, Northeastern University, Boston, MA. He received a Ph.D. degree in Computer Engineering from Northeastern University in 2022. His main research focuses on softwarized approaches for the Open Radio Access Network (RAN) of the next generation of cellular networks, on O-RAN-managed networks, and on network automation and orchestration. He served as guest editor of the special issue of Elsevier's Computer Networks Journal on Advances in Experimental Wireless Platforms and Systems.

Anupa Kelkar is the product manager for NVIDIA 5G Aerial converged radio access and edge compute platform. She has over 20 years of telecommunications and networking industry experience in software engineering and product management leadership spanning wireline, wireless, and satellite networks. Before NVIDIA, Anupa worked at Apple in incubation for AR/VR low-latency on-device, cloud, and edge-use cases, at Qualcomm in the connected cars (CV2X) ecosystem. And if you ever fly JetBlue, United, Continental, or Quantas, she was instrumental in the inflight broadband connectivity over satellite networks. Anupa graduated from University of California, Berkeley in electrical engineering and computer science.

Chris Dick joined NVIDIA in 2020 where he is a system architect working on the application of Artificial Intelligence and Machine Learning to 5G and 6G wireless. In his 30 years working in signal processing and communications he has delivered silicon and software products for 3G, 4G, and 5G baseband DSP and Docsis 3.1 cable access and vector processor architectures. He has performed research and delivered products for digital front-end (DFE) technology for cellular systems with a particular emphasis on digital pre-distortion for power amplifier linearization. Chris has also worked extensively on silicon architecture and compilers for machine learning and parallel computing architectures. Prior to moving to Silicon Valley in 1998, he was a tenured academic in Melbourne Australia for 13 years. He has over 250 publications and 100 patents. From 1998 to 2020 he was a Fellow and the DSP Chief Architect at Xilinx. In 2018 he was awarded the IEEE Communications Society Award for Advances in Communication for research in the area of full-duplex wireless communication.

Eduardo Baena is a postdoctoral research fellow at Northeastern University. Holding a Ph.D. in Telecommunication Engineering from the University of Malaga, his experience spans various roles within the international private sector from 2010 to 2017. Later he joined UMA as a lecturer and researcher contributing to several H2020 projects and as a Co-IP of national and regional funded projects.

Josep M. Jornet (M'13–SM'20–F'24) is a Professor in the Department of Electrical and Computer Engineering, the director of the Ultrabroadband Nanonetworking (UN) Laboratory, and the Associate Director of the Institute for the Wireless Internet of Things at Northeastern University (NU). He received his Ph.D. degree in Electrical and Computer Engineering from the Georgia Institute of Technology, Atlanta, GA, in August 2013. He is a leading expert in terahertz communications, in addition to wireless nano-bio-communication networks and the Internet of Nano-Things. In these areas, he has co-authored over 250 peer-reviewed scientific publications, including one book, and has been granted five US patents. His work has received over 17,000 citations (h-index of 61 as of June 2024). He is serving as the lead PI on multiple grants from U.S. federal agencies including the National Science Foundation, the Air Force Office of Scientific Research, and the Air Force Research Laboratory as well as industry. He is the recipient of multiple awards, including the NSF CAREER Award in 2019, the 2022 IEEE ComSoc RCC Early Achievement Award, and the 2022 IEEE Wireless Communications Technical Committee Outstanding Young Researcher Award, among others, as well as four best paper awards. He is a Fellow of the IEEE and an IEEE ComSoc Distinguished Lecturer (2022-2024). He is also the Editor-in-Chief of the Elsevier Nano Communication Networks journal and Editor for IEEE Transactions on Communications and Nature Scientific Reports.

Tommaso Melodia is the William Lincoln Smith Chair Professor with the Department of Electrical and Computer Engineering at Northeastern University in Boston. He is also the Founding Director of the Institute for the Wireless Internet of Things and the Director of Research for the PAWR Project Office. He received his Ph.D. in Electrical and Computer Engineering from the Georgia Institute of Technology in 2007. He is a recipient of the National Science Foundation CAREER award. Prof. Melodia has served as Associate Editor of IEEE Transactions on Wireless Communications, IEEE Transactions on Mobile Computing, Elsevier Computer Networks, among others. He has served as Technical Program Committee Chair for IEEE INFOCOM 2018, General Chair for IEEE SECON 2019, ACM Nanocom 2019, and ACM WUWnet 2014. Prof. Melodia is the Director of Research for the Platforms for Advanced Wireless Research (PAWR) Project Office, a $100M public-private partnership to establish 4 city-scale platforms for wireless research to advance the US wireless ecosystem in years to come. Prof. Melodia's research on modeling, optimization, and experimental evaluation of Internet-of-Things and wireless networked systems has been funded by the National Science Foundation, the Air Force Research Laboratory the Office of Naval Research, DARPA, and the Army Research Laboratory. Prof. Melodia is a Fellow of the IEEE and a Distinguished Member of the ACM.

Michele Polese is a Research Assistant Professor at the Institute for the Wireless Internet of Things, Northeastern University, Boston, since October 2023. He received his Ph.D. at the Department of Information Engineering of the University of Padova in 2020. He then joined Northeastern University as a research scientist and part-time lecturer in 2020. During his Ph.D., he visited New York University (NYU), AT&T Labs in Bedminster, NJ, and Northeastern University. His research interests are in the analysis and development of protocols and architectures for future generations of cellular networks (5G and beyond), in particular for millimeter-wave and terahertz networks, spectrum sharing and passive/active user coexistence, open RAN development, and the performance evaluation of end-to-end, complex networks. He has contributed to O-RAN technical specifications and submitted responses to multiple FCC and NTIA notice of inquiry and requests for comments, and is a member of the Committee on Radio Frequency Allocations of the American Meteorological Society (2022-2024). He is PI and co-PI in research projects on 6G funded by the NTIA, the O-RAN ALLIANCE, U.S. NSF, OUSD, and MassTech Collaborative, and was awarded with several best paper awards and the 2022 Mario Gerla Award for Research in Computer Science. Michele is serving as TPC co-chair for WNS3 2021-2022, as an Associate Technical Editor for the IEEE Communications Magazine, as a Guest Editor in an IEEE JSAC Special Issue on Open RAN, and has organized the Open 5G Forum in Fall 2021 and the NextGenRAN workshop at Globecom 2022.

Dimitrios Koutsonikolas is an Associate Professor in the Department of Electrical and Computer Engineering and a member of the Institute for the Wireless Internet of Things at Northeastern University. Between January 2011 and December 2020, he was in the Computer Science and Engineering Department at the University at Buffalo, first as an Assistant Professor (2011-2016) and then as an Associate Professor (2016-2020) and Director of Graduate Studies (2018-2020). He received his PhD in Electrical and Computer Engineering from Purdue University in 2010.

His research interests are broadly in experimental wireless networking and mobile computing, with a current focus on 5G networks and latency-critical applications (AR, VR, CAVs) over 5G, millimeter-wave networking, and energy-aware protocol design for smartphones. He has served as the General Co-Chair for IEEE LANMAN 2024, IEEE WoWMoM 2023, and ACM EWSN 2018, and TPC Co-Chair for IEEE LANMAN 2023, IEEE HPSR 2023, IEEE DCOSS 2022, IEEE WoWMoM 2021, and IFIP Networking 2021. He received the IEEE Region 1 Technological Innovation (Academic) Award in 2019 and the NSF CAREER Award in 2016. He is a senior member of the IEEE and the ACM and a member of USENIX.