IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 October 2024, accepted 25 October 2024, date of publication 7 November 2024, date of current version 19 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3493609

RESEARCH ARTICLE

Advancing Next Generation Wireless Networks With Digital Twin: Construction, Validation, and Real-World Applications on an Indoor Over-the-Air Testbed

BERK AKGUN, ADITYA JOLLY, BALWINDER SACHDEV, DIVYA RAVICHANDRAN, ROOHOLLAH AMIRI, (Member, IEEE), VIKAS JAIN, MURUGANANDAM JAYABALAN, YITAO CHEN, HETAL PATHAK, VINAY CHANDE, MOHAMMAD FAHIM, SRINIVAS YERRAMALLI, RUPESH ACHARYA, CHANDRESH TIWARI, CONNOR WOODAHL, ARUMUGAM KANNAN, XIAOXIA ZHANG, DEEPU ALEX, ABHISHEK KUMAR, HAI HONG, JOHN BOYD, RAJAT PRAKASH, SURESH BABU MUMMANA, SUMANTH GOVINDAPPA, JAMES Y. WILSON, JALAJ SWAMI, VIVIAN PHAM, ANDREI VADEANU, AND GILAD GOVRIN

Qualcomm Technologies Inc., San Diego, CA 92121, USA

Corresponding authors: Berk Akgun (bakgun@qti.qualcomm.com), Aditya Jolly (aditjoll@qti.qualcomm.com), and Vikas Jain (vjain@qti.qualcomm.com)

ABSTRACT Digital Twin (DT) technology has recently emerged as a powerful tool with the potential to revolutionize wireless systems as it enables accurate simulations, better decision-making, and tangible operational improvements. Prior studies on DT within the context of next generation wireless technologies have primarily focused on identifying potential use cases, application scenarios, standardization challenges, and conceptual implementation steps. However, the existing research is limited in translating theoretical ideas into real-world applications. Our research, in this paper, contributes to the practical realization of DT technology in the context of 6G wireless networks, demonstrating its potential impact on network planning, performance, and user experience. In particular, we explore the construction, validation, and applications of DT utilizing an indoor over-the-air (OTA) 5G NR testbed powered by an in-house developed Next Generation Radio Access Network (NG-RAN) that is fully compliant with 3rd Generation Partnership Project (3GPP) and Open RAN standards. First, we explain the integration and implementation steps followed to integrate Qualcomm Edgewise^{TM} Suite and Service Management and Orchestration (SMO) tools into the NG-RAN architecture, that will eventually enable the applicability of DT for wireless network operations. We then describe our procedure to construct and validate a high-fidelity DT of our OTA testbed modeling both Radio Frequency (RF) environment and system components. We demonstrate two pre-deployment use cases by describing our extensive coverage estimation and network capacity planning tests in OTA. Lastly, we explore how DT enables practical machine learning solutions for post-deployment use cases and share our comprehensive OTA performance results, highlighting that our proposed mobility and positioning techniques outperform the classical approaches in terms of throughput, number of undesired handovers, and positioning accuracy.

INDEX TERMS Digital twin, open RAN, 3GPP, RAN disaggregation, network planning, mobility, over-the-air, testbed.

I. INTRODUCTION

Digital Twin (DT) technology has recently gained significant attention for their potential to enhance efficiency, performance, and cost savings for wireless systems. It uses physical and operational data with historical patterns to

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaodong Xu.

model a physically accurate network simulation platform, which provides an up-to-date network status and predicts future network states [1], [2]. It also provides interfaces for interactions with the physical network and network applications/users. Unlike conventional network simulators, a Digital Twin Network (DTN) supports two-way communication between the physical network and the virtual twin network to enable closed-loop decisions and achieve real-time interactive mapping [3]. The authors in [4] offer an exhaustive examination of DT technology. Their treatment of DT encompasses multifaceted aspects, including precise definitions, relevant industry domains, enabling technological foundations, practical use cases, service applications, empirical case studies, and forward-looking trajectories. However, the discussion in [4] is a general survey on DTs, and it does not treat DT applications in telecommunications.

The authors in [5] present a vision for real-time DTs of physical wireless environments which are continuously updated using multi-model sensing data for distributed infrastructure and user devices. They present approaches for constructing and utilizing these real-time DTs and delineate on its applications and open problems highlighting the potential research platforms that can be used for such an investigation. Furthermore, the authors in [6] conduct a comprehensive survey on DT-enabled 6G services, while the authors in [7] delineate the potential use cases for DTs in 6G wireless networks. These use cases include network simulation and planning, network operation and management, data generation through simulation, Artificial Intelligence (AI) training and inference, and what-if analysis. In addition, the authors emphasize the critical prerequisites necessary to establish DTs for future 6G networks. On a similar vein, the authors in [8] identify potential research directions and standardization efforts related to the seamless integration of DTs within 6G communication networks. However, it is important to note that despite the comprehensive analysis, these studies do not present experimental evaluation and practical implementation of DTs.

The authors in [9], [10], and [11] present experimental evaluation of DT-enabled use cases for Multiple-Input Multiple-Output (MIMO) systems. The authors in [9] present a fingerprinting technique for localization with synthetic data from DT Radio Frequency (RF) maps and show its efficacy using practical simulations, realistic propagation models, and user measurements. The authors in [10] utilize a site-specific DT to generate synthetic channel state information (CSI) data and use it to train a Deep-Learning model. Afterwards, by using the DeepMIMO dataset [12], the performance of the model is evaluated in a real-world deployment. The authors in [11] propose a multi-agent Deep Reinforcement Learning framework to optimize the broadcast beam by constructing a data driven DT as a virtual environment for initial training. Their proposed technique enables safe exploration in the DT environment and faster convergence, as significant communication overhead involved with real networks is eliminated by using a DT.

Similarly, different DT frameworks for automatic reconstruction of a 3D environment and continuous interaction between the real-world and its twin are presented in [13] and [14], for beam management and directional beam selection, correspondingly. The authors in [13], propose a DT framework that automatically reconstructs the 3D environment and demonstrate the use of DT for beam management by implementing solutions for downstream sub-problems in beam acquisition, using the real-world dataset from the DeepSense 6G challenge.The authors in [14] present a framework with continuous interaction between the DT and the real-world at the edge and they show gains in beam selection accuracy using a publicly available RF dataset for DT creation.

The existing literature on DTs for 6G networks predominantly focuses on potential use cases, application scenarios, integration challenges, standardization efforts and potential applications. Real-world implementation aspects of DT models are discussed only in theory or the experimental evaluation presented in these studies are based only on publicly available dataset. To the best of our knowledge, the existing work lacks practical implementation of DTs and its integration with a live wireless network towards enabling closed-loop optimizations. To address this gap, our current research constructs a DT for an experimental OTA testbed and integrates it with our NG-RAN research platform. Through pre-deployment and post-deployment use cases, we demonstrate the practical viability of DT technology within a live network environment. With such DT technology implementation, our work endeavors to contribute towards the realization of DTs for future 6G networks.

A. CONTRIBUTIONS

Towards enabling DT in our testbed, we have enhanced our NG-RAN research platform [15]. In particular, we have integrated Qualcomm Edgewise Suite into the platform, which accelerates the deployment of cloud-native Artificial Intelligence (AI) and Machine Learning (ML) solutions and makes DT technology feasible for the next generation wireless infrastructure. In this paper, we first cover the specific enhancements introduced in the NG-RAN research platform, including the support for Open RAN (O-RAN) compliant O1 interfaces, the integration of Qualcomm Edgewise Suite, updates on RAN Deployment & Orchestration Framework, and the addition of Near Real-time RAN Intelligent Controller (RIC) capability. We then describe the construction of the DT model for our OTA testbed, including the radio environment and the system components. With extensive physical testing by utilizing the advanced NG-RAN architecture and OTA testbed, we validate the fidelity of the constructed DT in predicting pre-deployment network coverage and end-user performance, both of which set the foundation for high-fidelity DT. Furthermore, our efforts on sophisticated dynamic modeling of 5G RAN infrastructure and user devices lead to precise wireless

channel predictions towards enabling high-fidelity network capacity estimation and planning. We showcase that these precise channel predictions could help network operators estimate the number of users with the expected data rates that real-world deployments could sustain. We also discuss how this effort can help operators to identify potential performance bottlenecks and evaluate alternate scenarios to answer "what-if" questions.

We also demonstrate that the constructed high-fidelity DT is an enabler of practical ML solutions to improve closed-loop network operations. By being able to generate synthetic data at scale, the DT addresses the data collection challenge for ML applications in real-world deployments. In this work, we investigate two distinct ML use cases with DT assistance. First, we propose a novel mobility method, which utilizes a Long Short-Term Memory (LSTM) based neural network (NN) model trained using the data generated by the DT. Second, we show how DT enhances the Reference Signal Received Power (RSRP) based positioning accuracy of a mobile device utilizing an NN model compared to a weighted centroid algorithm (WCA).

The main contributions of this study can be summarized as follows:

- We provide our detailed implementation methods for the enhancements introduced to the NG-RAN research platform. This NG-RAN architecture is disaggregated, virtualized, AI-powered, and 3GPP & O-RAN compliant and it supports commercial 5GC and Service Management and Orchestration (SMO) solutions.
- We demonstrate high-fidelity DT construction and validation of an indoor OTA testbed, supported by the NG-RAN research platform.
- We showcase how to utilize the DT for pre-deployment use cases such as coverage estimation and capacity planning, which enables new capabilities for network operators that surpass the state-of-the-art techniques with significant cost reduction and enhanced operational efficiency.
- We also propose a novel DT-assisted mobility technique, or DTAM, that employs an LSTM based NN model. We demonstrate that the proposed technique reduces the number of ping-pong handovers, a problem observed in conventional mobility approaches, especially under challenging wireless conditions. Our OTA testing shows that this predictive mobility scheme delivers higher average throughput (up to 14% during handover events) and a more consistent user experience (by mitigating undesired handover ping-pongs) compared to conventional handover schemes.
- We also propose an RSRP-based DT-assisted positioning technique, or DTAP, that outperforms the conventional WCA positioning approach and does not require field data collection for NN model training. Our experimental results show that at 80-th percentile, DTAP significantly reduces the positioning error by up to 50% for some of our OTA routes.

The rest of the paper is organized as follows. Section II provides an overview of our OTA deployment while Section III details the advancements in our NG-RAN research platform. Section IV describes the DT construction and validation steps. Section V presents two pre-deployment use cases of the developed high-fidelity DT. DTAM and DTAP techniques are described and analyzed in Sections VI and VII, respectively. The OTA performance of each pre-deployment and post-deployment DT applications are evaluated in their corresponding sections. The paper is concluded in Section VIII.

II. OTA DEPLOYMENT

Our 5G network is deployed in a warehouse environment in one of the Qualcomm buildings in San Diego, as shown in Fig. 1. The warehouse environment is spread across an area of 150 ft×125 ft with three rows of metal shelves. The aisle region between the metal shelves acts as an impediment to the RF propagation and serve to produce conditions typical in real-world environment. The deployment is a 5G stand-alone network consisting of the NG-RAN research platform described in Section III connected to a commercial 5GC.

Our network deployment comprises of six Radio Units (RUs), which establish a six cell radio network. Each of these cells operates in the FR1 band (at 3.5 GHz) using Time Division Duplex (TDD) mode and a 100 MHz carrier bandwidth. To achieve PHY layer baseband functionality for these six cells, we utilize Qualcomm® FSM100xx-based PHY baseband PCIe cards, which are hosted in the RU servers [15]. These six RUs are strategically distributed across the OTA testbed. A subset of these RUs are directed at the center of the deployment to provide adequate RF coverage. Our testbed also consists of a commercially available Automated Guided Vehicle (AGV) with a Qualcomm® X60 5G modem based User Equipment (UE) onboard which is used for data collection and performance evaluation.

This AGV with the UE on board is programmed to follow a set of routes which are utilized for coverage and mobility studies. The coverage route programmed is shown in Fig. 1 using a dashed line. This is an extensive route covering the whole deployment with the goal to enable exhaustive data collection. The mobility route programmed is denoted by A-B-C-D-E-F-G, where the AGV with the UE on board enters the area between the metal blockers creating an adverse scenario for mobility and positioning studies. These routes are utilized and described further in Section IV-B, VI-D and VII-C.

III. NG-RAN RESEARCH PLATFORM

The NG-RAN research platform [15] is disaggregated, 3GPP and O-RAN interface compliant, virtualized and consists of CU (Centralized Unit), DU (Distributed Unit), and Radio Unit (RU) that interoperate with each other over Option 2 high-layer split (3GPP F1) mid-haul interface and

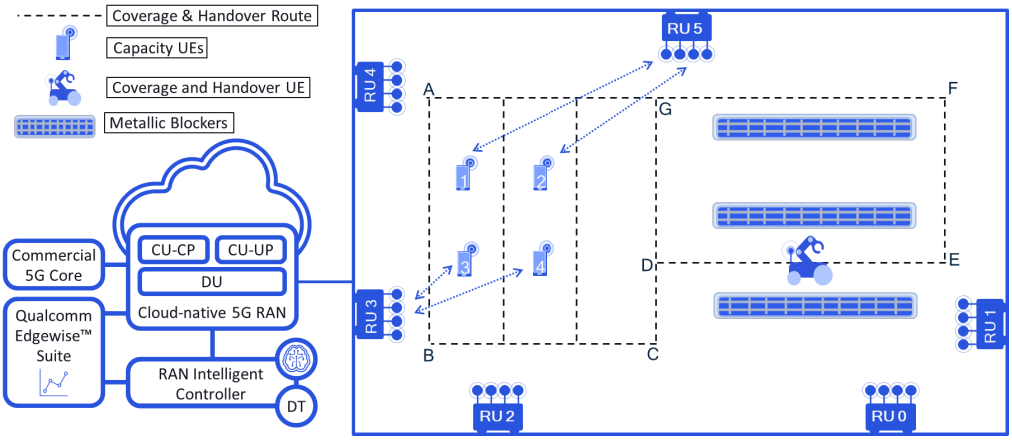


FIGURE 1. Indoor OTA deployment with 3GPP and O-RAN compliant NG-RAN research platform, a commercial 5GC, Qualcomm Edgewise Suite and six RUs.

Option 6 lower layer split. In this section, we highlight the architecture, services, and integration of our NG-RAN research platform with an SMO platform like Qualcomm Edgewise Suite, which will pave the path for augmenting advance technology like Digital Twin into this framework for wireless network operators. SMO, as defined in [16], is the central entity that is responsible for managing and orchestrating the platform, components, and services for the entire cellular RAN network. SMO provides support for the FCAPS (Fault, Configuration, Accounting, Performance & Security) management procedures for O-RAN Network Functions as well as associated services such as Trace and File management, Heartbeat management, Physical Network Function (PNF) startup and registration, PNF software and reset management, etc. over O-RAN defined O1 [17] interface. O1 interface specifies the management services (MnS) between the O1 compliant Managed Elements (MnS producers) and the SMO (MnS consumer) and defines the requirements, procedures, and operations. SMO also includes the Non-Real Time RAN Intelligent Controller (Non-RT RIC) which enables RAN optimization with intelligent radio resource management (operates at timescale of greater than 1 second) and provides policy-based guidance, ML model management, and enrichment information to Near-RT RIC function over O-RAN A1 interface [18].

A. QUALCOMM EDGEWISE SUITE

The NG-RAN is integrated with the commercial Qualcomm Edgewise Suite [19] which is multi-vendor RAN automation and centralized management platform to alleviate system complexity, simplify and automate cellular RAN infrastructure deployments. Qualcomm Edgewise Suite is also designed to support and manage both self-developed and third-party rApps [20] and provides interfaces to assure their interoperability towards enabling different use-cases as per customer requirements. The NG-RAN research platform has been enhanced to support O-RAN compliant Provisioning or Configuration Management (CM) services and Performance Assurance Management (PM) services over the O-RAN O1 interface. Fig. 2 highlights the NG-RAN research platform architecture with support for O-RAN CM and PM services over O-RAN O1 interface with Qualcomm Edgewise Suite. O-RAN Configuration Management services [17] allow SMO (MnS consumer) to configure attributes of managed objects for Network Functions (MnS producer) and enable the latter to report configuration changes to the SMO. The NG-RAN supports O-RAN O1 CM operations (e.g., NETCONF get, get-config, edit-config, etc.) for NETCONF [21] /YANG [22] solution set as specified in [17] and [23]. The NG-RAN implements the Generic Network Resource Model [24] and 5G Network Resource Model (Stage 2 and 3) [25], along with proprietary extensions to model all the manageable entities in the NG-RAN research platform. As shown in Fig. 2, all the O-RAN O1 NETCONF provisioning operations are serviced at the NETCONF server which are then propagated to the respective Network Function (NF)'s O1 CM Agent for processing using the NETCONF server's backend interface. Qualcomm Edgewise Suite incorporates NETCONF client capability and supports the O-RAN O1 CM operations towards providing the various afore-mentioned O-RAN CM services.

In Fig. 3 we show a snapshot of Qualcomm Edgewise CM dashboard that demonstrates a subset of OTA deployment in our test location described in Section VI-D. The right panel shows the network layout with 4 out of 6 cells radiating at the test location. The left side panel shows the configuration attributes for each cell that is selected by the network operator. The exhaustive list of attributes for CU and DU functions are available in the tabs at the top of the screen.

O-RAN PM services allow the RAN NF's (Performance Assurance MnS producer) to either stream (in real time) or report in bulk (through file transfer) performance data to SMO (Performance Assurance MnS consumer) that can, for example, enable data collection and analytics for AI/ML based use-cases. PM services also allow the SMO to conduct performance assurance operations on the RAN NF's such

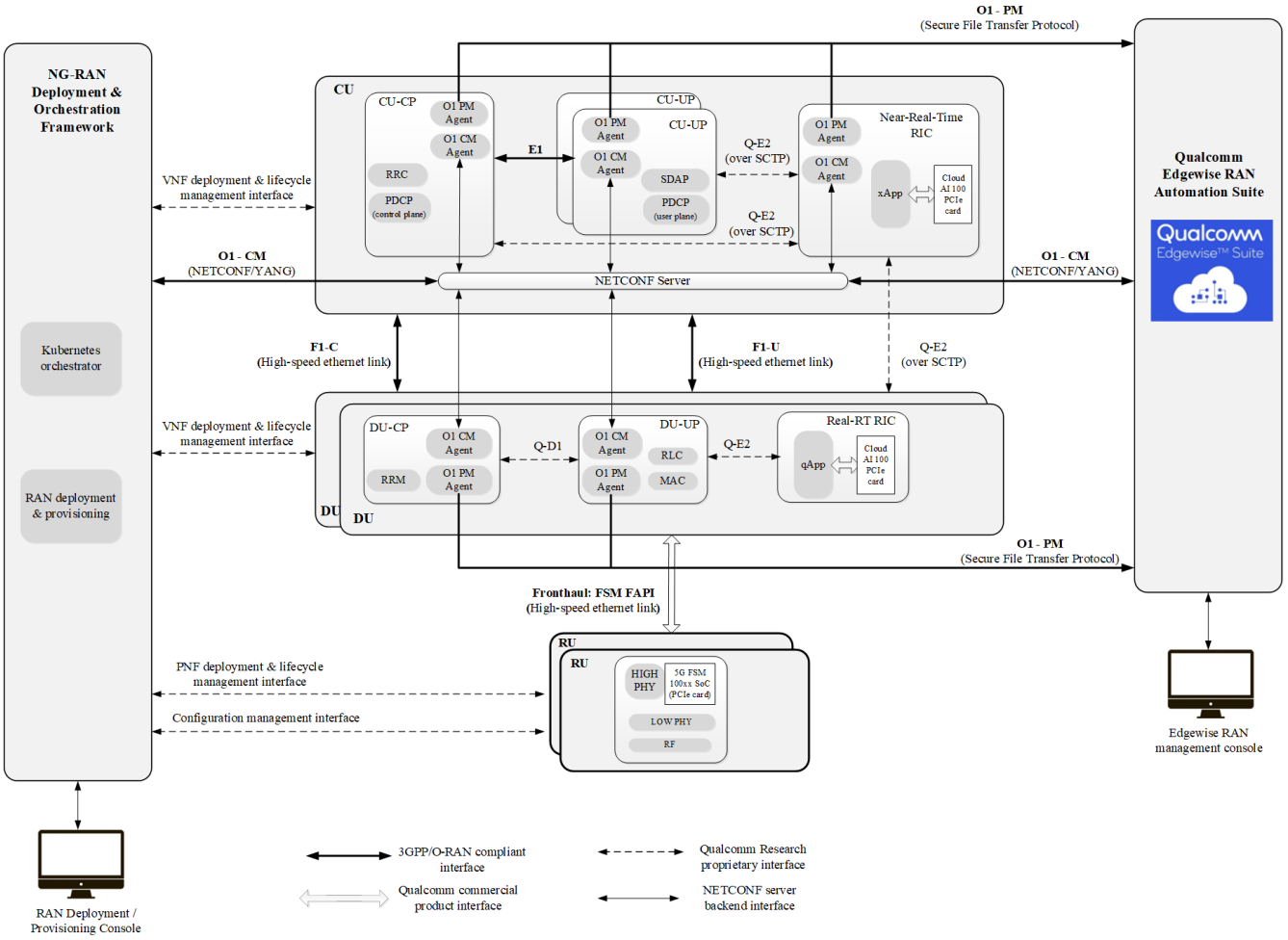


FIGURE 2. NG-RAN research platform with O-RAN O1 compliance & integrated with Qualcomm Edgewise Suite.

as selecting the measurements to be reported, setting the periodicity of reporting and other associated attributes. The NG-RAN implements the O1 Push-based Performance Data File reporting mechanism [17] using Secure File Transfer Protocol (SFTP) with PM file naming, content, and XML format compliant to O-RAN O1 [17], [23]. Fig. 2 shows this capability with the O1 PM Agent function within each NF, which provides the PM reporting services. The NG-RAN NF's implement support for PerfMetricJob IOC (as specified in 3GPP TS 28.622 [24]) over O1 CM interface for controlling the various aspects of performance metric production e.g., activating or deactivating the performance metrics, selection of metrics to be collected, periodicity of reporting, managing the scope of data collection across multiple managed object instances etc. In our OTA deployed NG-RAN network, the O1 PM Agent function incorporates the SFTP client functionality and communicates with the PM services in Qualcomm Edgewise Suite to periodically report key performance indicator (KPI) metrics in O1 PM XML file format. The data and reports generated using the PM services are used for monitoring UE Downlink (DL) throughput KPI for our deployment. The UE DL KPI throughput plot generated out of Qualcomm Edgewise PM is provided in Section VI-D.

B. NG-RAN DEPLOYMENT & ORCHESTRATION FRAMEWORK

The NG-RAN research platform performs the PNF and Virtual Network Function (VNF) workload deployment on the RAN infrastructure using the internally developed NG-RAN deployment and orchestration framework (refer Fig. 2). Virtualization for the RAN infrastructure is performed using standard Kubernetes (K8s) framework which manages the CU and DU hosts (x86 based servers) as part of K8s cluster and deploys each VNF as a containerized workload (specifically as a Kubernetes pod) on these hosts. The NG-RAN deployment & orchestration framework utilizes HELM charts [26] to package all the VNF workload artifacts for on-target installation, which also facilitates subsequent upgrades and management of software releases. The NG-RAN deployment framework also includes NETCONF client functionality for configuration management required during the initial phases of the NG-RAN platform setup and NF provisioning.

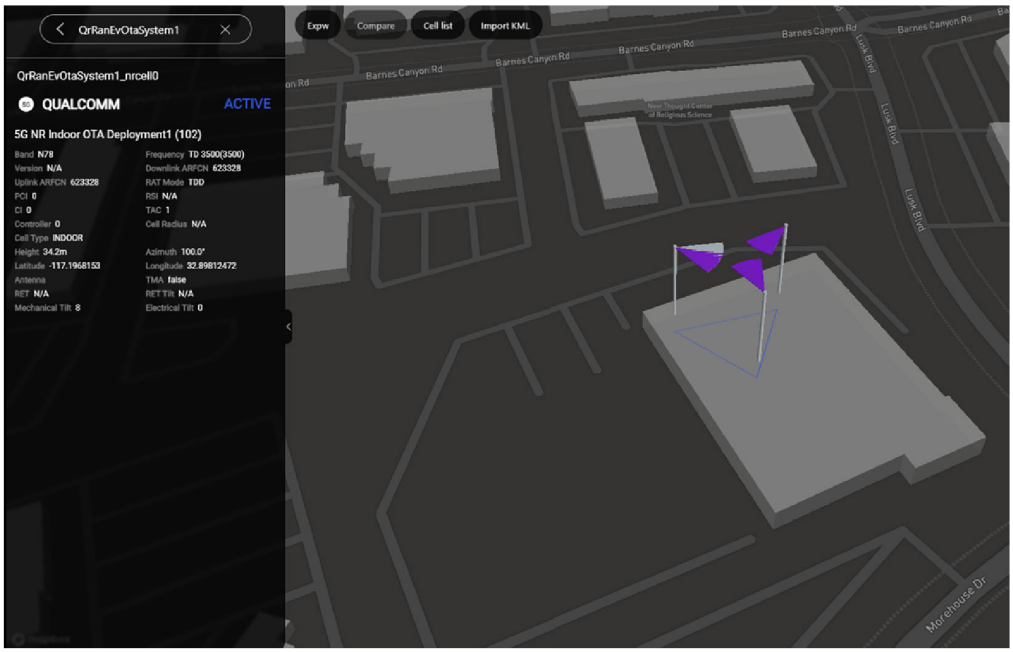


FIGURE 3. Qualcomm Edgewise CM Dashboard for OTA Deployment.

C. NG-RAN NEAR-RT RIC

The NG-RAN research platform has been enhanced to support our proprietary variant of Near-RT RIC function, described in O-RAN architecture [16]. Near-RT RIC enables near-real time control and optimization of RAN function via fine-grained data collection and actions on a time scale between 10ms and 1 sec. Near-RT RIC function hosts multiple applications, called xApps, which contain customized RAN optimization logic (e.g., using AI/ML algorithms), communicates through standardized interfaces and service models towards closed-loop control for optimizing the RAN functions. In the NG-RAN research platform (refer Fig. 2), Near-RT RIC is connected to the RAN or Q-E2 nodes (i.e., CU and DU) over the Q-E2 interface (modeled as an O-RAN E2-like interface) and allows the NG-RAN Near-RT RIC to control radio resource management and other functionalities of the RAN nodes. Q-E2 interface is logically composed of two protocols, namely Q-E2 Application Protocol (AP) and Q-E2 SM (Service Model), similar to E2-AP [27] and E2-SM [28]. Q-E2 AP incorporates protocol procedures for interface management (e.g., Q-E2 Setup procedure) and multiple services (Q-E2 report, Q-E2 control) that collectively implement the Q-E2 SM. For interface management procedures, after the SCTP [29] connection is established between Near-RT RIC and Q-E2 node (which is configured via O1 interface of the Q-E2 IP address and port of the Q-E2 termination of the Near-RT RIC), Q-E2 node transmits the Q-E2 Setup Request listing its supported RAN Functions along with node identifiers. After processing the Q-E2 Setup Request, Near-RT RIC replies with the Q-E2 Setup Response. Q-E2 report service involves Q-E2 RIC Indication message that could be triggered periodically or based on certain events, and contains relevant telemetry measurements from the Q-E2 RAN node. For the current NG-RAN use-cases described in Section VI, Q-E2 control service is initiated upon the reception of Q-E2 RIC Indication message and generates Q-E2 RIC Control message which contains the outcome of the RAN optimization logic executed at xApp, which is then applied at the Q-E2 RAN node. Similarly, for Q-E2 service models, the NG-RAN platform supports custom service for DT-assisted mobility use-case described in Section VI, wherein the Q-E2 RAN node (CU-CP) upon the receipt of L3 RRC measurement report from UE, constructs the set of serving and neighbor cell RSRP measurements along with UE location for the configured time window duration (to be reported to Near-RT RIC in Q-E2 RIC Indication message). The NG-RAN Near-RT RIC is currently deployed as containerized workload (specifically as a Kubernetes pod) on the CU host and utilizes the Qualcomm® Cloud AI 100 PCIe (Peripheral Component Interconnect Express) card [15] for performing AI/ML inference required for RAN optimization. The NG-RAN's virtualized architecture allows Near-RT RIC to be deployed on its own dedicated hosts, separate from CU and DU hosts, as needed to meet the use-case requirements.

IV. DT CONSTRUCTION, FIDELITY, AND COVERAGE

A. CONSTRUCTION

In this section, we describe the steps involved in construction of a high-fidelity DT of our physical OTA testbed along with the benefits it can bring to coverage estimation. DTs in mobile networks can exist at various levels, given that 5G NR networks comprise of multiple hardware and software layers. It is therefore crucial to define which segments of the network are modeled in the DT. In our implementation, we create a DT that includes the physical deployment site (indoor

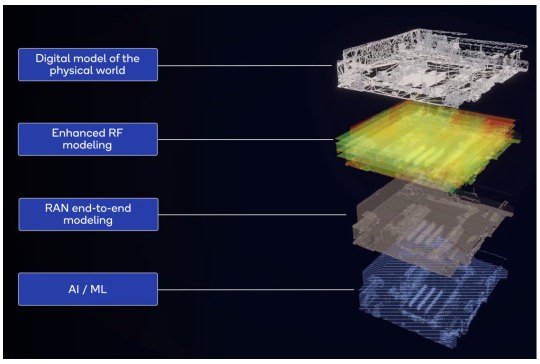


FIGURE 4. High-level depiction of steps involved in DT construction.

building), radio environment, RAN, and its associated system components including six RUs. Each step in constructing the DT is depicted in Fig. 4 and outlined below.

1) ACCURATE 3D ENVIRONMENT MODELING

The three-dimensional testbed environment is synthesized utilizing a collection of high accuracy LiDAR (Light Detection and Ranging) and drone derived image scans. The images are processed using open drone map (ODM) [30], a native application for processing drone images, to generate a predicted point cloud. The point cloud is imported into Cloudcompare (3D point cloud processing software) [31], where it is downsampled and aligned with previous scans. The point cloud is then used to create a mesh of all the environment using poisson surface reconstruction algorithm [32] and subsequently partitioned into various objects such as walls, shelves, poles, and ceiling. The complexity of the ray tracing based channel modeling [33] was significantly reduced by a factor of ten (from 20 minutes down to 2 minutes) through the decimation of the three-dimensional objects from three million triangles to sixty thousand. To accurately simulate the electromagnetic (EM) properties of an object in the EM wave propagation model, it is imperative to define the material of each object within the DT. The material allocated to each object segment is chosen from the International Telecommunication Union (ITU) standard library, specifically at the OTA testbed frequency of 3.5 GHz [34].

2) SPATIALLY CONSISTENT RADIO ENVIRONMENT MODELING

The process of modeling a radio environment with spatial consistency employs ray tracing, a scalable methodology for characterizing wireless channel models specific to an environment. The foundational steps of ray tracing encompass 3D modeling of the environment, which serves as an input to the ray tracing software [35], path discovery, and field computations over the identified paths. Path discovery is predicated on the principles of Shooting and Bouncing Rays (SBR) and image theory. During this phase, we integrate antenna characteristics for a specified transmitter and receiver location, orientation, array configuration, quantity of antenna elements, antenna pattern, and the polarization of each individual element. All potential geometric paths with pre-determined propagation characteristics are identified. These propagation characteristics encapsulate multiple orders of reflections and diffractions. Ultimately, the geometric paths and the type of interactions at each interaction point are utilized to compute the behavior of EM wave propagation throughout the path. Upon completion of this process, multiple propagation paths are identified, and corresponding features for each path are computed. These features include power, phase, time of arrival, angles of departure, and angles of arrival. In conclusion, at the end of this step in the DT creation process, wireless channel emulation is developed for a pair of transmitting and receiving antennas.

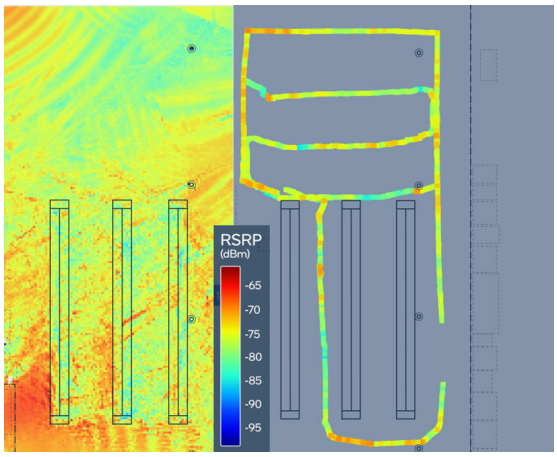


FIGURE 5. Digital Twin generated RSRP heatmap of the whole testbed area (left) vs. actual RSRP measurements from the physical testbed (right) where the UE onboard an AGV is programmed to traverse an extensive route as seen in the figure.

3) CALIBRATION OF DT

The calibration of the DT is a multi-level process. In this study, we calibrate the DT on an elemental basis, wherein each element of the network is calibrated against its counterpart, taking into account the effect of the element on the RSRP. For example, all the gains and losses of the digital and analog components of the RF paths, such as Intermediate Frequency (IF) components, Digital to Analog Convertor (DAC), Analog to Digital Convertor (ADC), and RF Front End (RFFE), are meticulously measured and incorporated into the DT. This calibration enables the RSRP at a specific user location within the DT to be numerically calibrated, thereby closely predicting the RSRP measured by the user device in the real world environment.

B. FIDELITY AND COVERAGE

To evaluate the fidelity of the constructed DT model, we executed a drive test within the OTA testbed as depicted in Fig. 1, utilizing the UE onboard an AGV that was programmed to traverse an extensive route. During the test, we collected data pertaining to the UE reported RSRP and

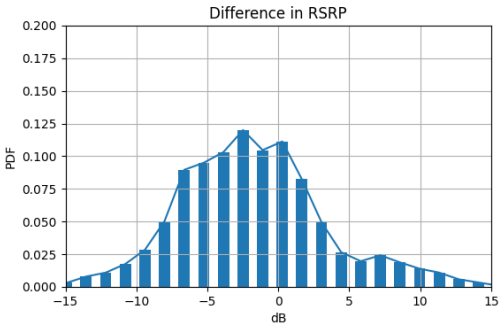


FIGURE 6. PDF of difference in RSRP.

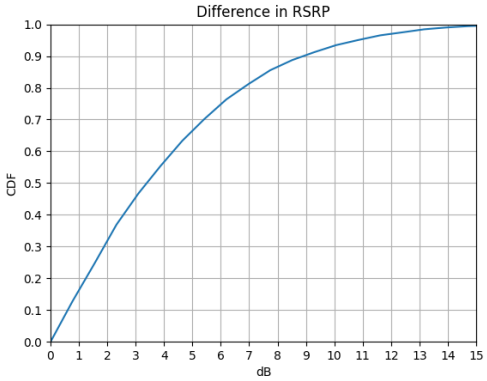


FIGURE 7. Absolute CDF of difference in RSRP.

the precise ground truth location from the AGV, which was carrying the UE. The UE reports the serving cell and neighboring cell Synchronization Signal Block (SSB) based RSRP measurements, which are subsequently utilized to validate the results derived from the DT. We then computed the RSRP in the DT by calculating the conducted transmission power, the path loss from the ray tracer, and normalizing the received power to represent SSB RSRP reports.

Fig. 5 illustrates a two-dimensional RSRP heat map generated from both DT as well as physical testbed, thereby indicating high fidelity of the constructed DT model. In Fig. 6 and Fig. 7, we plot the probability density function (PDF) and absolute cumulative distribution function (CDF), respectively, of difference between the DT computed RSRP and the UE reported RSRP over the course of the drive test. From the plot, it is observed that the DT predicted RSRP closely aligns with the UE measured RSRP and is within an average of 4.17 dB. The difference in RSRP can be further minimized by augmenting the DT predicted RSRP with RSRP measurements from the OTA testbed. This observation further emphasizes the precision in the DT's coverage prediction.

V. DT-ASSISTED CAPACITY

Accurate RF modeling of the physical environment obtained from the DT along with sophisticated dynamic modeling of 5G RAN infrastructure and user devices can enable high-fidelity network capacity estimation and planning. Even though commercial network planning tools provide coverage predictions by using 3D maps and ingest field measurements to provide better coverage estimates, they are not designed to predict capacity and optimize wireless network operations such as mobility and positioning. In this section, we describe how DT can be effectively used towards achieving these objectives. This entails DT-assisted capacity validation and leveraging it to predict what-if deployment scenarios.

A. DT-ASSISTED CAPACITY VALIDATION

Our first objective is to build a reliable DT-assisted capacity analysis platform that can faithfully represent the capacity Key Performance Indicators (KPIs) of a deployed OTA network, with verifiable fidelity between observed OTA KPIs and DT predicted KPIs.

Towards this objective, a subset of the OTA testbed that is discussed in Section II is considered, which is a deployment with 2 Cells and 4 UEs (with 2 UEs connected per cell). The RF model for this scenario is obtained from the DT framework developed as described in Section III. This channel model is fed into the 5G RAN research system emulator that closely emulates the OTA RAN configuration with the traffic of interest, which is UDP (User Datagram Protocol) based full-buffer downlink traffic. The emulator output is processed to compute the capacity KPIs such as throughput, scheduling, channel quality, and resource utilization. The DT-predicted KPIs are then compared with OTA KPIs as perceived by the users in real-time, when positioned at equivalent locations for the same deployment scenario and traffic of interest.

Multiple components in the end-to-end DT-assisted capacity analysis model have to be aligned with the OTA deployment. Apart from accurate RF modeling from DT, the next requirement is to emulate the OTA RAN behavior in the 5G RAN research system emulator in terms of parameters and configurations such as frame-structure, overheads, scheduler user selection, and link adaptation algorithm. Any mismatch in modeling the RAN configuration could create a significant departure between OTA performance and DT predictions. Further, it is of importance to accurately replicate the UE behavior in the RAN emulator.

Fig. 8 shows the comparison between downlink user throughput as predicted by the DT-assisted framework and the actual user perceived throughput in the OTA testbed at different times for the 2 Cell/4 UE deployment scenario with full-buffer traffic. It is seen that observed throughput trend for all the 4 users is similar to the DT-predicted values though the absolute numbers are slightly different. With RF conditions changing gradually over time it is also expected to see slight variations in the user perceived OTA throughput over time. In order to evaluate the deviations from DT-assisted predictions, the user perceived throughput in the OTA test bed are captured at two instances well-separated in time. From Fig. 8, it is observed that the maximum deviation of DT predicted user throughput from OTA dataset-1 is at 17% for

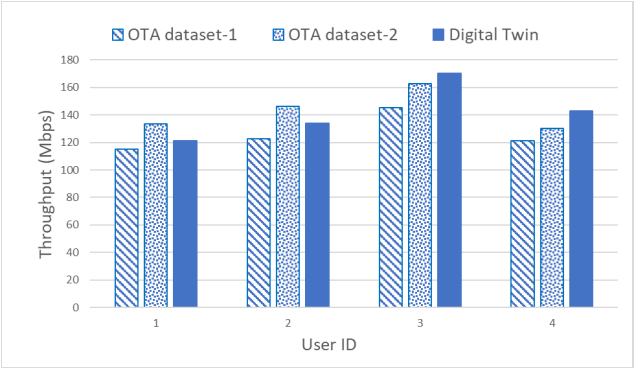


FIGURE 8. OTA vs. DT predicted downlink throughput for 2 Cell/4 UE scenario.

user 4 with DT predictions being more optimistic for all users, while the maximum deviation from OTA dataset-2 is close to 10% for users 1 and 4 with DT over-estimating for user 4 and under-estimating for user 1. The deviations from DT predictions are caused due to slight inaccuracies in modeling the interference and receiver behavior. Receiver modeling involves simulating the behavior of commercial UEs in terms of their ability to perform interference cancellation, channel estimation, demodulation, and decoding. When the simulator fails to accurately model the spatial nulling of interference that is seen in OTA environment, it could result in inaccurate SINR (Signal to Interference plus Noise Ratio) estimation causing mismatch between the DT predicted throughput and real-time observed throughput. The discrepancies observed in the deviation from the DT predicted throughput, upon the collection of capacity KPIs at disparate intervals, is ascribed to the environmental changes in the OTA testbed that happens over time. However, by refining our treatment of these modeling errors and updating the DT model periodically to account for the changing physical environment, it is possible to achieve even more accurate predictions from DT. Our study was primarily confined to the indoor OTA deployment, with the examination of outdoor DTs designated as a subject for future research.

B. DT-ASSISTED CAPACITY ANALYSIS FOR WHAT-IF SCENARIOS

Our second objective is to demonstrate the use of DT-assisted capacity analysis platform to estimate the capacity KPIs in what-if deployment scenarios – such as with different user densities, traffic profiles, and intensity. An example of such a scenario is to estimate KPIs when a larger number of UEs are deployed in an extended network with more cells, under a traffic demand profile. To illustrate this, a deployment consisting of 4 cells and 12 UEs with bursty traffic under different loading conditions is considered as shown in Table 1. The bursty traffic is generated from Poisson based file arrival rate with file size of 0.5 MB such that low loading corresponds to a mean rate of 4 files per sec or 16 Mbps and high loading, to 12 files per sec or 48 Mbps.

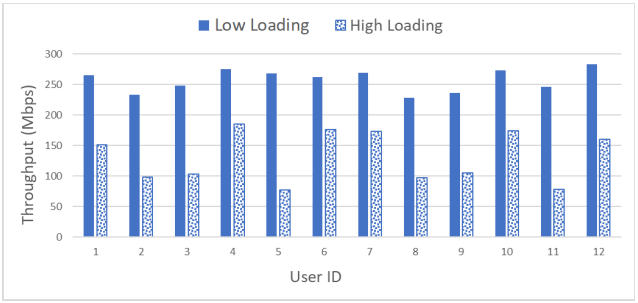


FIGURE 9. DT predicted mean user throughput in the what-if scenario under low/high loading with Poisson traffic.

The simulation for this scenario is run on the DT-assisted capacity framework as described in Section V-A and the predicted capacity KPIs are analyzed.

TABLE 1. DT-predicted resource utilization % in the what-if scenario under low/high loading with Poisson traffic.

RU	Low Load Resource Utilization(%)	High Load Resource Utilization(%)	No. of UEs	UE Ids
0	4	24	1	12
1	4	19	1	4
2	17	61	4	1,6,7,10
3	26	85	6	2,3,5,8,9,11

It can be inferred from Table 1 and Fig. 9 that the capacity KPIs as perceived by the 12 users for this extended scenario are aligned with the expected outcomes in terms of resource utilization, impact of interference, and throughput. At low loading conditions, the overall resource utilization is low across all cells resulting in less interference and higher user perceived throughput. At high loading conditions, overall resource utilization goes up because of more traffic and interference. Thus, the user perceived throughput is significantly reduced due to resource sharing and increased interference. The reduction in throughput for users 4 and 12 at high loading is caused primarily due to interference from traffic in the neighboring cells and not due to resource sharing, as they are the only users in their respective cells. The throughput reduction for the rest of the users at high loading conditions is caused by increased interference as well as resource sharing within the users of the same cell. Thus, the DT-assisted framework can effectively capture the effect of changes in traffic loading as observed by the corresponding increase or decrease in the user throughput.

VI. DT-ASSISTED MOBILITY

In this section, we explore how DT technology can enhance closed-loop network operations through practical ML solutions. Specifically, we address the challenge of data collection from real-world deployments, which is often time-consuming and costly. In many scenarios, collecting field data at scale is not even feasible. By integrating high-fidelity DT into ML model development procedure, we can conveniently

overcome these limitations and achieve more efficient network management. We can simply generate high-quality synthetic data from a DT model, which accurately represents spatial signal patterns in the real world. Then, this synthetic data from a high-fidelity DT can be used to train ML models, eliminating the need to collect field data and leading to enhanced scalability. In our proposed technique, we exploit this synthetic data from the DT to train a predictive mobility ML model, which is an LSTM-based NN model. This model aims at selecting the optimal cell for a device in mobility in terms of throughput and user experience. Our OTA test results show that our predictive mobility technique, i.e., DT-assisted mobility or DTAM, delivers higher average throughput and a more consistent user experience along the test route compared to the baseline Release-15 (Rel15) mobility approach.

A. BASELINE REL15 MOBILITY

Before delving into the details of DTAM, we first briefly explain the baseline Rel15 mobility approach, which is established by 3GPP in Release 15 RRC specifications (see Section 9.2 in [36]) and widely adopted for 5G cellular networks. Specifically, 5G 3GPP specifications define various intra Radio Access Technology (RAT) measurement events, namely A1 to A6 for supporting mobility use-cases. (Note that inter-RAT mobility is beyond the scope of this work.)

At a high level, intra-RAT handover procedure is used to transfer the UE connection from a source cell to a target cell to enable consistent Quality of Service (QoS) and maintain service continuity. A handover event is triggered by the source cell, based on the received measurement reports, which indicate that the corresponding radio conditions are satisfied with respect to neighbor cells. Among intra-RAT handover events, Event A3 is typically used for intra-frequency handover procedures, since it provides a handover triggering mechanism based upon relative measurement results, e.g., it can be triggered when the RSRP of a neighbor cell is stronger than the RSRP of the serving cell. Some key configuration parameters and our optimization methods are discussed in Section VI-D. A call flow describing a typical NR handover can be summarized with the following steps [36] (please also refer to Fig. 10): 1) The source cell provides the UE with measurement and reporting configuration. 2) The UE reports according to the specified configuration. 3) On receipt of the measurement report, the source cell decides to handover the UE based on the measurement reports. 4) The source cell then initiates a handover request to the target cell which performs admission control and then returns an acknowledgment message. 5) After the approval, the source cell sends a handover command (i.e., RRC Reconfiguration message) to the UE, and the UE disconnects from the old cell. 6) The UE then completes the RRC handover procedure by sending an RRC Reconfiguration Complete message to the target cell. 7) The network completes additional handover related procedures with switching the data traffic to the target cell and releasing the UE context

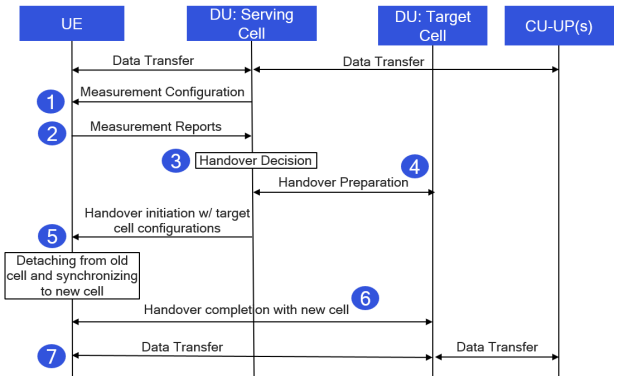


FIGURE 10. 5G NR intra-RAT handover procedure [36].

at the source cell. Note that during the handover procedure (i.e., starting from Step 5 and onward), UE cannot be scheduled, which leads to an interruption in data transfer, until connection to the target DU cell is established. Therefore, it is crucial that the handover decision mechanism and procedures be optimal to enable a more consistent user experience.

B. PROPOSED MOBILITY METHOD

In this section, we describe our proposed mobility method, i.e., DTAM, that addresses performance degradation issues which may occur in a challenging environment with dense deployment of cells with directional antennas and RF propagation impediments, e.g., metallic blockers located in the OTA testbed. Specifically, due to high spatial reuse and interference, the close proximity of cells and focused coverage areas lead to frequent handovers as users move through the network. In addition, these conditions often lead to handover ping-pong events, that occur when a UE switches from a source cell to a target cell and then switches back to its earlier source cell within a short amount of time. To mitigate these undesired handovers, which cause network signaling overhead and traffic disruption, we explore the idea of performing RSRP predictions for serving and neighbor cells based on the past measurement reports sent by UEs. Hence, the handover decisions are proactively determined by considering future wireless channel conditions rather than relying only on the current measurements.

Due to its superior performance in sequence prediction and learning long-term dependencies in datasets, we study LSTM-based neural networks to predict the future RSRP measurements, and utilize this information for handover decisions [15]. To enhance the accuracy of RSRP predictions, we configure periodic measurement reports as described by 3GPP specifications [37], so that UE performs RSRP measurements and conveys this information to the CU-CP network function residing in the NG-RAN as shown in Fig. 1. The CU-CP network function forwards these RSRP measurements to the RIC which hosts the LSTM-based NN model for predicting future RSRP measurements.

We set the periodicity of the measurement reports as 120 ms to maximize the prediction accuracy, as this is the minimum allowable value by the 5G NR 3GPP specifications to achieve the best baseline Rel15 handover performance.

Fig. 11 highlights the key functional components of our DTAM NN model design with the required inputs and the outputs. In the first step, UE provides periodic measurement reports belonging to the serving cell and all the available neighbor cells. In addition to these RSRP reports, the AGV, which carries the UE, is capable of recording and sending its two-dimensional location information every 80 ms to the CU-CP Network Function over a separate communication network (e.g. Wi-Fi). We exploit this additional data to further enhance the RSRP prediction accuracy in our design. After gathering both RSRP measurements (over 5G) and location information, they are transported to the RIC where the NN module, or RSRP predictor, is located. This NN module, pre-trained using DT-generated synthetic data, consumes this input sequence and provides predicted RSRP values for the prediction horizon h. In other words, the RSRP predictor takes an $d \times l$ input matrix $X_t = [x_{t-l}, x_{t-l+1}, \ldots, x_{t-1}]$, where l is the sequence length and d is the feature size. Each $x_k \in X_t$ is a vector with d entries and includes UE-reported RSRP measurements of $n = d-2$ cells and two-dimensional (x-y coordinates) UE location at time k. After the inference is performed by the RIC, it returns a $n \times h$ output matrix with predicted RSRP values $Y_t = [y_t, y_{t+1}, \ldots, y_{t+h-1}]$ in the prediction horizon h, where each $y_k \in Y_t$ is a vector with n entries, i.e., RSRP predictions for n cells for horizon k.

Upon receipt of RSRP predictions, the CU-CP Network Function processes the predicted RSRP values and applies majority voting logic to determine whether to trigger a handover decision. In particular, the CU-CP gets predicted RSRP values Y_t for inference at time t as input and outputs the predicted cell index z_t with the following rule. Let H and K denote the handover margin and minimum number of instances in the majority voting scheme that satisfy the requirements to trigger a handover, respectively. Also, let $\mathbb{1}[.]$ denote the indicator function. Given that s represents the serving cell index, if there is any cell $i \in \{0, 1, \ldots, d-2\}$, where $i \neq s$ such that

$$\sum_{j=0}^{h-1} \mathbb{1}\left[Y_t[j,i] > H + Y_t[j,s]\right] \geq K \tag{1}$$

then, trigger a handover to cell i, i.e., $z_t = i$.

Fig. 12 shows the high-level model structure of the RSRP predictor and NN model parameters are summarized in Table 2. In this structure, a single LSTM layer is used, while the LSTM hidden size is set to 16. Two fully connected layers follows this LSTM layer, by utilizing Rectified Linear Unit (ReLU) as activation function. The size of the fully-connected layers is set to 256. The other key design elements are depicted in Fig. 12.

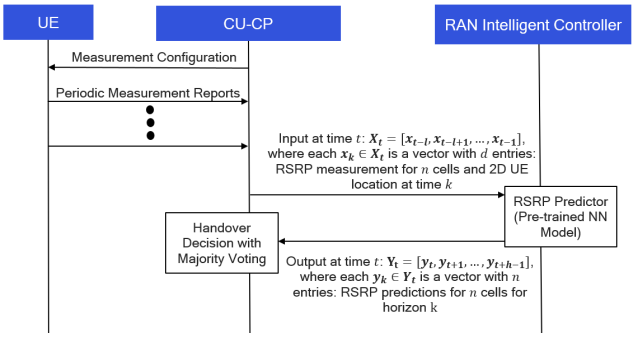


FIGURE 11. DT-assisted Mobility: Proposed handover technique that utilizes an LSTM-based neural network model.

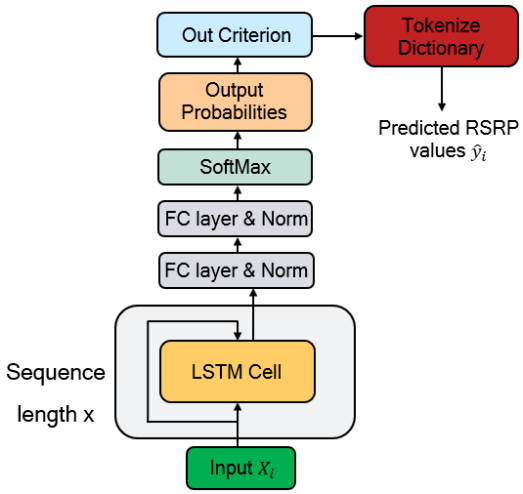


FIGURE 12. The NN model structure for the RSRP predictor.

TABLE 2. Summary of NN model parameters.

Model Parameters	Size/Range
Size of synthetic data from DT	700K to 1M
Train/validation/test split	0.8/0.1/0.1
Feature size d	n (RSRP for each cell) + 2D location
Sequence length l	50
Prediction horizon size h	16
LSTM hidden size	32
Fully-connected layer size	256
# of LSTM layers	1
# of fully-connected layers	2
Model input dimension: $[d, l]$	$[n+2, 50]$
Model output dimension: $[n, h]$	$[n, 16]$

C. NN MODEL TRAINING

We use DT-generated synthetic dataset for training, validation, and test purposes of the LSTM model with 0.8/0.1/0.1 split, respectively. In our synthetic data generation, we consider UEs with a height of 1.46 meters assuming a Single Input Single Output (SISO) transmission scheme. These UEs are distributed in an area of 25.92 m $\times$ 13.56 m with 12 cm steps in both dimensions. These dimensions lead to a UE grid with a size of 217 $\times$ 113. Let x_u and y_u denote the UE indices in this grid system where

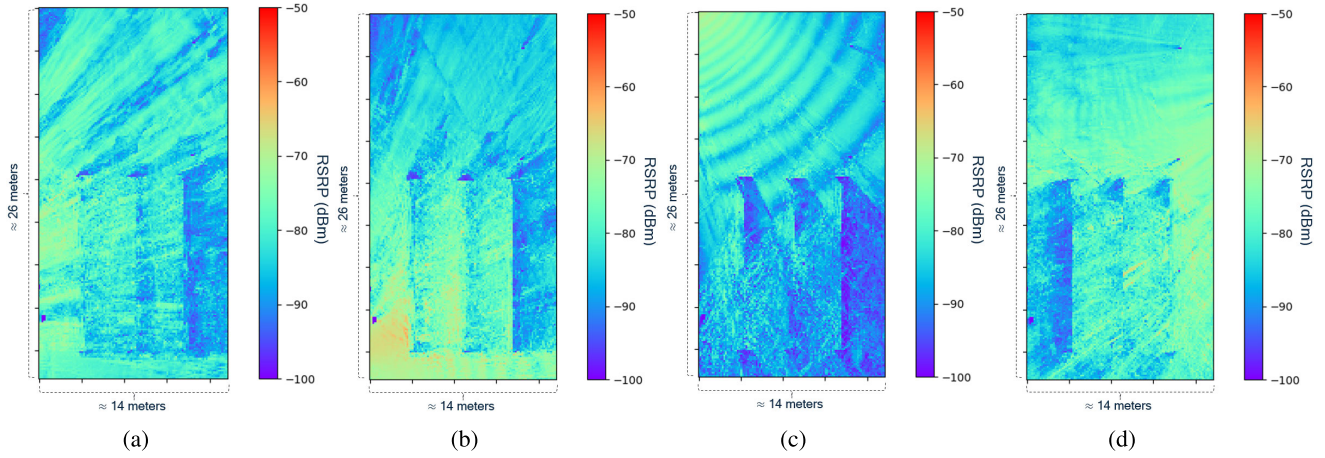


FIGURE 13. RSRP heatmaps generated by the DT with linear interpolation for RU 0, 1, 3, and 5, respectively.

$x_u \in [0, \cdots, 216]$ and $y_u \in [0, \cdots, 112]$. For each one of these UEs, we then compute RSRP values for all n cells, utilizing the DT. Hence, for a UE with the index of (x_u, y_u), we obtain a vector $s_{(x_u,y_u)} = \left[s^0_{(x_u,y_u)}, \cdots, s^{n-1}_{(x_u,y_u)}\right]$ where each $s^k_{(x_u,y_u)} \in s_{(x_u,y_u)}$ represents RSRP value of cell k. Note that due to computational resource constraints, 12 cm is the smallest spacing that can be handled in the DT. However, the configured RSRP reporting interval is 120 ms in the NG-RAN and the target AGV speed is 30 cm/s. This fact necessitates the interpolation of the RSRP values generated by the DT, as the maximum spacing between two consecutive RSRP measurements should be $120 \times 30 \times 10^{-3} = 3.6$ cm for a more accurate representation of the actual OTA scenario. To increase the statistical robustness, we further reduce this spacing to 2.4 cm in our dataset, which expands the original UE grid to an aggregated grid represented by $\mathcal{G}^{1081\times561}$. In this expansion, we utilize linear interpolation on the generated RSRP values in dB domain per cell. We simply apply the interpolation in one dimension (e.g., x axis) and then in the other dimension, while creating the DT-generated RSRP vectors of $p_{(x_r,y_r)} = \left[p^0_{(x_r,y_r)}, \cdots, p^{n-1}_{(x_r,y_r)}\right]$ where each $p^k_{(x_r,y_r)} \in p_{(x_r,y_r)}$ represents the corresponding RSRP for cell k $\forall\,(x_r, y_r) \in \mathcal{G}$. Fig. 13 shows the RSRP heatmap generated by the DT with linear interpolation for RU 0, 1, 3, and 5. When these figures are compared with Fig. 1, the area occupied by the metallic blockers are clearly seen with very low RSRP values (note that Figs. 13 are 90° rotated version of Fig. 1 in clockwise). Furthermore, depending on where the RU's are located and their orientation, the region with higher RSRP coverage changes. For example, RU 0 and RU 1 cover the left and bottom parts of the area, respectively, whereas RU 3 is stronger at the top left section and RU 5 is stronger at the right side of the area. This deployment leads to handover events when the AGV moves around various routes, especially inside the alleys created by the metallic blockers.

The next step is to sample random paths from the aggregated grid $\mathcal{G}$ to generate the datasets to be used in training, validation, and testing. Algorithm 1 summarizes the random path generation process with the required inputs and the output. Let α, N, and L denote the stride, number of paths to be generated, and the total length of each path, respectively. Then, the following steps are repeated N times until all the paths are generated. First, a random starting point, i.e., $g_0 \in \mathcal{G}$, is chosen. We then pick a random direction from the following options: Up, down, left, or right. However, to prevent unreasonable scenarios and to keep the analysis less complicated, the UE is not allowed to go immediately back. That is, if the previous direction is down, the new selected direction cannot be up. After the new direction is determined, the UE moves α steps towards this direction and then adds the grid point g_1 to the sample set. The UE then randomly selects a new direction and repeats the same process L times to complete a path. Therefore, for each path i, a sample set of $g_i = [g_0, \cdots, g_L]$ is generated where $i \in \{0, N-1\}$. During this process, a time series of RSRP values $\mathcal{O} = \{(X_k, Y_k)\}_{k=0}^{N\times L}$ of length $N \times L$ is

Algorithm 1 Random Path and Dataset Generation Process

Input: Aggregated grid:$\mathcal{G}$, DT-generated RSRP vectors: $p_{(x_r,y_r)}$ $\forall\,(x_r, y_r) \in \mathcal{G}$, stride: α, number of paths:N, length of each path: L, lookback size: l, horizon size: h
Output: A time series of RSRP values $\mathcal{O} = (X, Y)$
for each path i in $\{0, N-1\}$ **do**
  Pick a random starting point $g_0 \in \mathcal{G}$
  for for each j in $\{0, L-1\}$ **do**
    Pick a random direction from {up, down, left, right}, but never go back
    Go in the selected direction by α steps and append $\mathcal{O}$ with the corresponding (X_j, Y_j) pair
  end for
end for
return $\mathcal{O}$

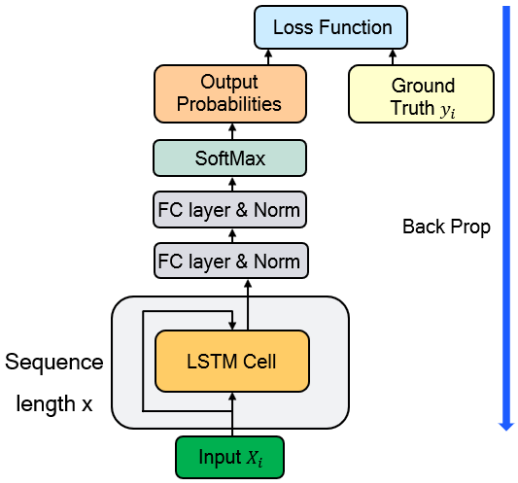


FIGURE 14. High-level block diagram for the NN model training.

generated by appending the corresponding RSRP samples at each step to the dataset $\mathcal{O}$. Here, each X_k denotes a $(n+2) \times l$ matrix where each column represent a feature set at the corresponding sampling time. For example, let us consider a generic sampling point $g_k \in g_i$ at time k where $i \in \{0, N-1\}$. Then, $X_k = [x_{k-l+1}, \cdots, x_k]$ consists of vectors corresponding from the sampling point g_{k-l+1} to the current sampling point g_k where there is a one-to-one correspondence between the sampling time k, the sampling location g_k, and the sample x_k. As discussed in the previous section, the vector x_k consists of n RSRP values corresponding to each one of n cells and two-dimensional coordinates at time k. Furthermore, Y_k denotes an $n \times h$ matrix where each column represent RSRP measurements from n cells for the horizon length of h at the sampling time k. That is, $Y_k = [x_{k+1}, \cdots, x_{k+h}]$.

After the completion of this dataset generation procedure, we randomly split $\mathcal{O} = \{(X_k, Y_k)\}_{k=0}^{N \times L}$ into training dataset, validation dataset, and test dataset with 0.8/0.1/0.1 ratio, respectively. Fig. 14 show the high-level block diagram for the NN model training. We use cross entropy loss as loss function. Therefore, the target truth $Y_k \; \forall k \in \{0, N \times L\}$ are tokenized to the corresponding indices in the token set during training. Furthermore, uniform Xavier weight initialization method are used in both LSTM and fully connected layers where all biases are set to 0. We also employ batch normalization and dropout techniques (where dropout ratio is 0.1) to train the fully connected layers.

D. PERFORMANCE EVALUATION

This section describes the performance comparison conducted between the baseline and DTAM schemes, focusing on the following KPIs:

1) NUMBER OF HANDOVERS/PING-PONGS

This KPI quantifies the aggregate count of handovers throughout the entire test duration, encompassing both regular handovers and ping-pong handovers. For our evaluation, a ping-pong handover is defined as any sequence of handovers where the UE handovers to a target cell and returns to the serving cell in less than 3 s.

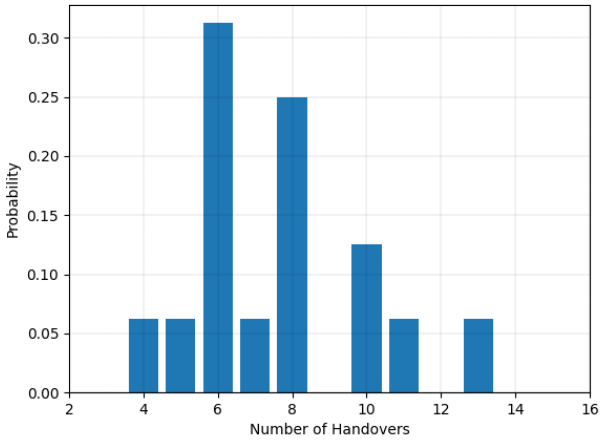
2) DIFFERENTIAL THROUGHPUT GAIN

This KPI quantifies the differential in DL throughput between the DTAM and the baseline schemes. Specifically, it examines regions of the test run where the NR Physical Cell ID (PCI) differs between the two schemes, highlighting the potential throughput advantage of DTAM over the baseline scheme. Our primary focus here is solely on DL, while the investigation of uplink (UL) throughput remains a topic for future research.

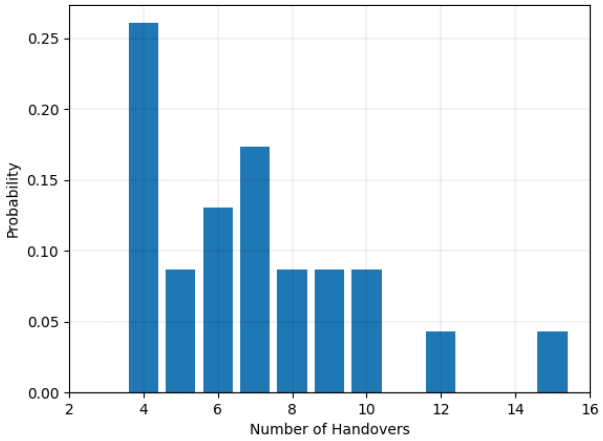
In our performance study, we conducted tests using a subset of four cells from our larger six cell OTA deployment. These selected cells were configured to establish a cell-edge scenario within a warehouse environment, where an AGV (with a UE onboard) followed a predefined route denoted by A-B-C-D-E-F-G-A covering the entire deployment area, as shown in Fig. 1. The RU IDs considered for these tests were 0, 1, 3, and 5. The selection of these RUs provided sufficient coverage while also necessitating handovers due to the directional nature of our RU transmit antennas. To ensure realistic conditions, we enabled DL Link Adaptation (LA) with MCS Table 2 (see Table 5.1.3.1-2 in [38]). Full-band Physical Downlink Shared Channel (PDSCH) spanning 12 symbols were configured on all D and S slots of the DDSU TDD slot pattern. The UE was configured to report SSB-based RSRP measurements for serving and neighbor cells defined by baseline configuration parameters, including A3 offset, hysteresis, and time-to-trigger. The UE reported RSRP measurements were used by the CU to trigger A3 event-based handovers for the baseline scheme [37]. The key configuration parameters for the DTAM scheme were H and K defined in Section VI. The ranges for these parameters were $H \in [0\text{ dB}, 15\text{ dB}]$ and $K \in [1, 16]$.

In our investigation, we systematically explored a range of configuration parameters to identify the most optimal settings for both the baseline and DTAM schemes. Notably, we observed substantial run-to-run variations during our experiments. To ensure robustness in our reported findings, we selected a single representative run for each scheme. This selection process involved removing outliers based on the observed number of handovers and ping-pongs across all runs and subsequently choosing the median run. By adopting this approach, we accurately reflect the average performance gain achieved by the DTAM scheme over the baseline scheme while mitigating the influence of anomalous results. The PMF of all the runs based on total number of handovers for both baseline and DTAM schemes are shown in Fig. 15a and Fig. 15b, respectively.

The selected baseline Rel15 mobility run shows six handovers with one of the handovers characterized as a ping-pong and the selected DTAM run shows four total handovers with no ping-pongs. The throughput and serving cell PCI over time curves obtained from Edgewise PM for



(a)



(b)

FIGURE 15. PMF of number of handovers observed during (a) baseline Rel15 mobility and (b) DTAM runs.

both baseline and DTAM schemes over the full test route are shown in Fig. 16. The handover characterized as ping-pong for the Rel15 baseline run can be seen in Fig. 16 where the UE onboard the AGV handovers from PCI 5 to PCI 3 and back to PCI 5 in less than 3 sec, whereas no such instance is observed in the DTAM run.

To understand where this ping-pong happened in our OTA deployment we plot the UE location vs. serving cell PCI for the whole run in Fig. 17a and Fig. 17b for the baseline and DTAM runs, respectively. It can be seen that the ping-pong handover for the baseline Rel15 run happened in the aisle region between the metal shelves of our OTA deployment evident from the consecutively highlighted handover instances where UE onboard the AGV had PCI 0 as its serving cell. The adverse RF conditions created by the aisle region between the metal shelves of our OTA deployment, made baseline Rel15 mobility to perform an unnecessary handover because of relatively shorter timescale RSRP change, whereas in the DTAM run no such handover is observed considering the ML model's ability to predict RSRP changes in advance.

FIGURE 16. Throughput and serving cell PCI over time for baseline Rel15 mobility and DTAM runs obtained from Qualcomm Edgewise PM dashboard.

There are two key findings from these observations. Firstly, DTAM demonstrates a significant reduction in the total number of observed handovers compared to the baseline scheme. This reduction is accompanied by the elimination of ping-pong handovers. The performance improvement of DTAM can be attributed to the predictive capabilities of our LSTM-based NN model. Our model predicts RSRP based on prediction horizon size, ensuring a selection of target cell that the UE can sustain its link. Notably, the absence of ping-pongs in DTAM stems from intelligent decision-making based on predicted RSRP values. Rather than switching between cells based on current RSRP conditions only, DTAM prioritizes staying with the current cell when instantaneous RSRP changes occur. Consequently, DTAM scheme proves more robust to RSRP fluctuations in our OTA deployment, mitigating potential performance degradation observed in baseline handovers. Secondly, DTAM demonstrates an average differential throughput gain that is 14% higher than that of the baseline scheme throughout the entire run. In particular, it enhances this KPI from 232.3 Mbps to 264.9 Mbps. This figure is derived by averaging the throughput across all occurrences where a differential PCI was observed for baseline and DTAM schemes, respectively. This gain is a direct consequence of the reduced number of handovers in DTAM. By minimizing handover occurrences, we effectively mitigate the instantaneous throughput degradation associated with handover events. Fig. 18 delineates the individual instances where DTAM and the baseline scheme exhibit divergent PCI values due to handover events and Fig. 17 portrays the regions in the OTA deployment where these divergent PCIs are observed. For the first and fifth instance where the PCI differs for baseline Rel15 mobility and the DTAM run, DTAM scheme triggered a handover before the baseline scheme acquiring the potential throughput gain as evident from the first and fifth bar in Fig 18. Similarly, in the second instance, the DTAM scheme handed over

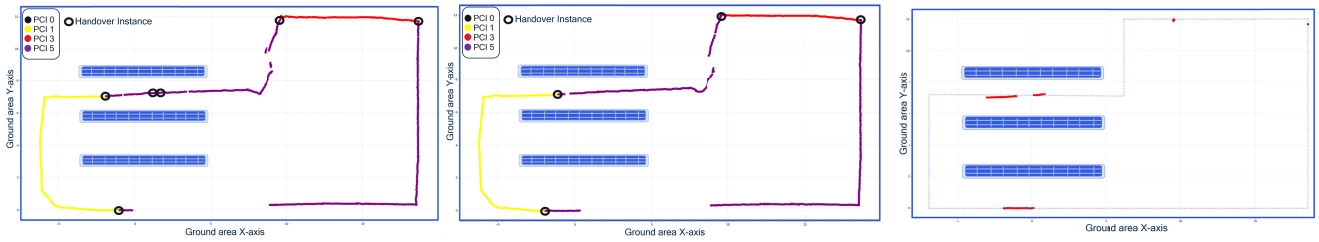


FIGURE 17. (a/b) UE location vs. serving cell PCI for (a) baseline Rel15 mobility and (b) DTAM runs. (c) Areas where the PCI differs for baseline Rel15 mobility and DTAM runs.

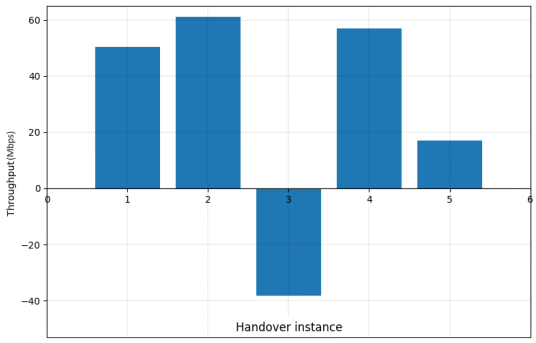


FIGURE 18. Differential throughput gain of DTAM over baseline scheme at each handover instance.

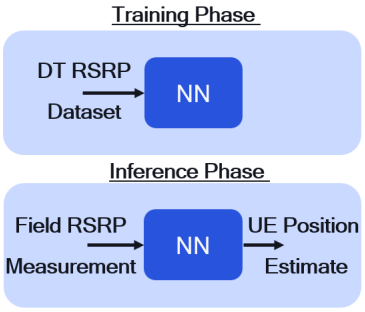


FIGURE 19. High-level block diagram of DT-assisted positioning procedure.

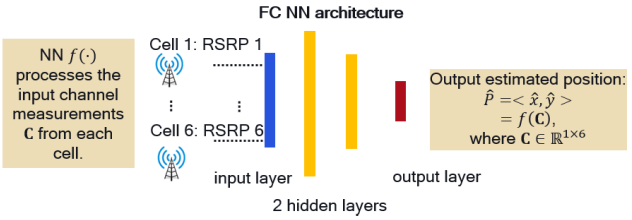


FIGURE 20. NN model architecture used in DTAP method.

to PCI 5 later than the baseline scheme. Looking at the corresponding bar on Fig. 18, we can see that this was a better decision, as DTAM run showed a better throughput in this region. In the third instance, depicted by the negative bar in Fig. 18, the DTAM scheme chooses to stick with PCI 1, while the baseline Rel15 mobility scheme performs a ping-pong handover to PCI 0. As a result, the baseline Rel15 mobility scheme experiences an immediate increase in throughput, leading to the negative bar. However, the DTAM scheme trades off this instantaneous gain to avoid ping-pong handovers, resulting in a more consistent user experience with fewer interruptions compared to the baseline Rel 15 mobility scheme. Specifically, these handovers led the baseline scheme to forfeit potential throughput gains that would have been achievable by intelligently choosing when to handover, as demonstrated by the discussion above. This underscores the efficacy of our proposed DTAM technique.

In summary, our DTAM scheme optimally utilizes available network capacity by judiciously avoiding unnecessary handovers, thereby demonstrating more consistent user experience that can provide better QoS for the users in a wider range of scenarios. As the NN Model was trained only using synthetic data generated from the developed high-fidelity DT, its effectiveness in predicting RSRPs in a real world scenario highlights the high-fidelity of our DT. We also observed that the model was tolerant to small changes in the environment because of the training strategy used, eliminating the need to recreate changes in the DT and retraining the model. This shows the ability of the developed DT to be used for closed-loop application in next generation of cellular networks.

VII. DT-ASSISTED POSITIONING

In this section, we show how a DT can be used to improve the RSRP based positioning accuracy of a UE in an indoor environment. While conventional RSRP based positioning algorithms, e.g., weighted centroid algorithm (WCA), do not provide a good positioning accuracy, ML approaches are more promising and show relatively enhanced performance. However, ML approaches heavily rely on field data collection for NN training, which are very expensive and time-consuming as discussed before. Here, we propose an RSRP based DT-assisted positioning technique (DTAP) that outperforms the conventional WCA approach and does not require field data collection. Before explaining the details of DTAP, we provide a brief description of the conventional WCA in the next subsection. Then, we demonstrate our DTAP approach in Section VII-B. Finally, in Section VII-C,

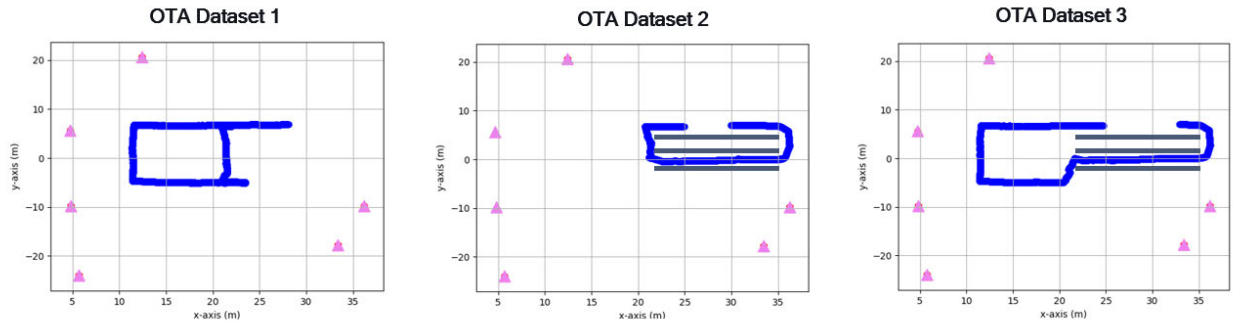


FIGURE 21. OTA routes for performance comparison between DTAP and WCA.

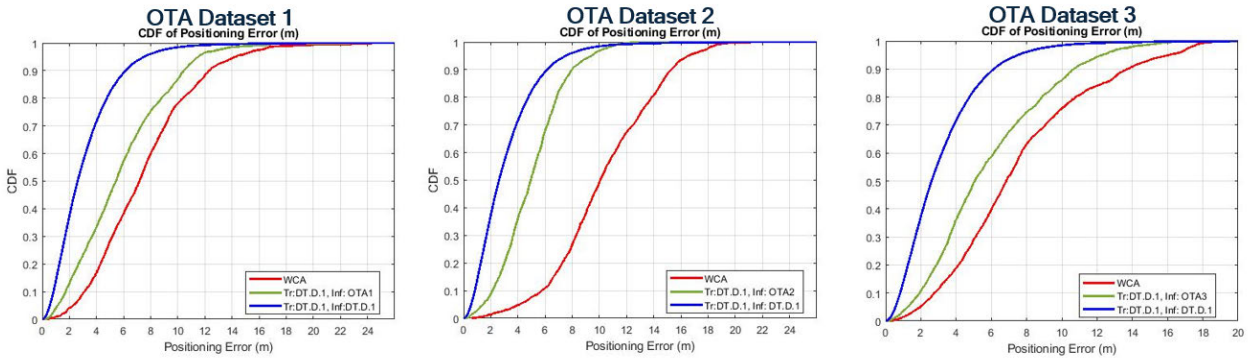


FIGURE 22. Positioning error CDF comparison between DTAP and WCA for three different OTA routes.

we provide experimental results where we show performance comparisons between WCA and DTAP approaches.

A. WEIGHTED CENTROID ALGORITHM

In RSRP based positioning, the goal is to position a UE given a set of SSB RSRP measurements at the UE from multiple cells with known locations. The idea can be summarized as follows: For a given target UE, let $r_1 \geq r_2 \geq \cdots \geq r_M$ denote the measured RSRP values (in dBm) from M different cells, and let $P_1, P_2, \cdots, P_M$ be the different cells' locations. The UE position estimate is then given by the weighted sum of the cells' positions, where the weights are a function of the RSRP values. Specifically, the UE position estimate $\hat{P}$ is calculated as:

$$\hat{P} = \frac{\sum_{k=1}^{N} w_k P_k}{\sum_{k=1}^{N} w_k}, \tag{2}$$

where N is the maximum number of cells used in the equation, $N \leq M$, and $w_1, \cdots, w_N$ are the cells' weights where $w_k = 2^{\frac{r_k - r_1}{\lambda}}$ for a scaling factor $\lambda > 0$.

B. PROPOSED POSITIONING METHOD

Our proposed procedure is depicted in Fig. 19 and can be summarized as follows. First, in the training phase, an NN model is trained on a DT-generated dataset. The dataset generation follows the same steps as in Section VI-C. Particularly, the dataset is composed of RSRP vectors, where each vector represents RSRP values received at a different receive point from the six different cells. Similar to the procedure in Section VI-C, the receivers are modeled as a dense grid of receive points with 12 cm spacing in both the horizontal x and y axes, corresponding to a grid with a size of 217×113. Also, each receiver models a UE at a height of 1.46 m with an almost omni radiation pattern to match its real counterpart and the RUs in the DT are located at the same locations as their real counterparts with heights ranging from 4.56 m to 4.83 m, with radiation patterns matching the real deployment.

After the NN model is trained only on DT-generated data, the inference can be performed using field RSRP measurements as an input to the trained NN model to obtain UE position estimates. The NN model used in DTAP consists of two fully connected hidden layers with an input layer of size 6 and an output layer of size 2. The NN takes six RSRP values from the six different RUs as input and outputs the estimated UE positions x and y coordinates. The NN architecture is depicted in Fig. 20. In training phase, we minimize an MSE loss function using Adam optimizer with initial learning rate of 0.01 and batch size 50 for 150 epochs.

C. EXPERIMENTAL RESULTS

Our field RSRP measurements are comprised of three separate OTA datasets, each of which constitutes a particular route within the test area, as shown in Fig. 21. Specifically, the first OTA dataset, denoted by OTA1, is collected along a route

TABLE 3. Positioning error percentiles for the three different OTA routes with DTAP and WCA.

OTA1	%50-percentile (m)	%80-percentile (m)	%90-percentile (m)
DT-assisted	5.45	8.85	10.45
WCA	7.16	10.38	12.38
OTA2	%50-percentile (m)	%80-percentile (m)	%90-percentile (m)
DT-assisted	4.94	6.84	7.97
WCA	10.04	13.85	15.29
OTA3	%50-percentile (m)	%80-percentile (m)	%90-percentile (m)
DT-assisted	5.02	8.76	10.67
WCA	6.88	10.80	13.76

outside of metallic blockers, i.e., A-B-C-D-G-A in Fig. 1. The second OTA dataset, denoted by OTA2, is collected along a route around the metallic blockers, i.e., G-D-E-F-G in Fig. 1. The last OTA dataset, denoted by OTA3, is collected along a full route around the testing area, i.e., A-B-C-D-E-F-G-A in Fig. 1.

Fig. 22 shows a comparison between the cumulative distribution functions (CDFs) of the positioning errors of the conventional WCA and our proposed DTAP approaches. For all OTA routes, DTAP shows improvement in positioning accuracy over WCA. For example, at 80-th percentile, DTAP has gains of 1.53 m, 7.01 m, and 2.04 m over WCA for OTA1, OTA2, and OTA3, respectively. In addition, the high gain of 7.01 m at 80-th percentile for OTA2 shows the DTAP capability to mitigate the positioning errors due to NLOS/blocking conditions. Table 3 summarizes the CDF error comparison between our proposed approach and conventional WCA solution for three OTA routes.

VIII. CONCLUSION

With demonstrations in our OTA testbed, we bridged the gap between theory and implementation, towards achieving the promised benefits of DT technologies. Furthermore, as we move into the next phase of 5G Advanced and 6G, our work will empower the wireless operators and researchers to access the capabilities of DT technology for network planning, management, and enhancement, refining foundational aspects even further. In this study, we particularly showed how to construct and validate a high-fidelity DT model of our OTA testbed. We then explained how to exploit this model for pre-deployment use cases such as coverage estimation and network capacity planning by sharing our OTA results. We further highlighted that the DT can serve as an effective methodology to generate large-scale synthetic data to train NN models. To showcase a potential post-deployment application, we proposed a DT-assisted mobility technique which utilizes an LSTM-based NN model. Our OTA results showed that DTAM outperforms the baseline mobility scheme in terms of throughput and ping-pongs during handovers. We also showcased another post-deployment application where RSRP-based positioning approach is enhanced by an NN model trained with DT-generated synthetic data.

As discussed throughout the paper, we have identified several key areas for future exploration to enhance the application of DT technology in wireless networks. First of all, the discrepancies observed in the deviation from the DT predicted throughput, due to environmental changes in the OTA testbed over time, highlight the need for continuous refinement. Future work will focus on improving the treatment of these modeling errors and incorporate a closed loop periodic updating strategy for the DT model to account for the evolving physical environment. This approach aims at increasing the prediction accuracy of the DT. Additionally, while our current study was confined to an indoor OTA deployment, future research will extend to outdoor DTs. This will involve examining the unique challenges presented by outdoor network deployments (such as quality of outdoor maps, building shapes and heights, roof structures, vegetation, poles and other structures), dynamic motion of objects, seasonal and weather conditions, user distribution, and traffic demand behavior modeling. Outdoor networks also involve a wide variety of UEs (i.e., enhanced Mobile Broadband or eMBB, reduced capacity or RedCap, extended reality or XR etc.) and a plethora of UE operating modes resulting in more complex interference patterns. Finally, we acknowledge that the implementation of DT technology in wireless networks can raise concerns about energy consumption and costs. However, it is important to note that we are in the early stages of utilizing DT concepts in wireless networking. The benefits of DTs, such as improved network optimization and reliable capacity planning, have significant potential to enhance user experience, ultimately outweighing these initial costs and energy expenditures in the long term. By focusing on these practical aspects of integrating DT solutions into real-world networks, we plan to explore additional use cases for DT technology. Our ongoing studies in this research area include investigating how DTs can optimize efficient resource allocation for different service types in consideration of network energy savings and utilizing DTs to develop sophisticated scheduling algorithms.

ACKNOWLEDGMENT

The authors would like to thank Swarup Dwaral, Vinay Garg, Sukruth Babu, Prabhu Ramani Iyer, Sandeep Jangid, Rachit Srivastava, Rahul Ankushrao Kawadgave, Aaron Bailey, Shubham Agrawal, Hemanshu Tyagi, Yun Lin, Tao Liu, Varun Rajpal, and Vince Baglin, for their technical expertise and contributions to the work presented in this article. They would also like to thank Shlomo Horev and Daniele Harel (both from Edgewise DevOps Teams) and Karthikeyan Manoharan, Sri Harsha Bodduluru, Sai Vikram Katkam, Neelesh Singh, Miguel Abraham, Ankur Gupta, and Chinmaya Parida (all from Qualcomm IT Teams), for their help in provisioning Edgewise AWS cloud instance and establishing end-to-end connectivity with the NG-RAN research platform in the laboratories and OTA test-bed.

They would also like to thank Vijay Shirsathe, VP of Engineering for Qualcomm Technologies Inc., for his technical guidance and constant support toward the development of the Next Generation RAN research platform.

They would also like to thank John Smee, SVP of Engineering for Qualcomm Technologies Inc., for his unwavering support and overall direction without which this work would not have been possible.

Snapdragon and Qualcomm branded products are products of Qualcomm Technologies Inc., and/or its subsidiaries. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the author(s) and do not necessarily reflect the views of Qualcomm Technologies Inc.

REFERENCES

[1] L. U. Khan, W. Saad, D. Niyato, Z. Han, and C. S. Hong, "Digital-twin-enabled 6G: Vision, architectural trends, and future directions," *IEEE Commun. Mag.*, vol. 60, no. 1, pp. 74–80, Jan. 2022.

[2] H. X. Nguyen, R. Trestian, D. To, and M. Tatipamula, "Digital twin for 5G and beyond," *IEEE Commun. Mag.*, vol. 59, no. 2, pp. 10–15, Feb. 2021.

[3] Y. Wu, K. Zhang, and Y. Zhang, "Digital twin networks: A survey," *IEEE Internet Things J.*, vol. 8, no. 18, pp. 13789–13804, Sep. 2021.

[4] S. Mihai, M. Yaqoob, D. V. Hung, W. Davis, P. Towakel, M. Raza, M. Karamanoglu, B. Barn, D. Shetve, R. V. Prasad, H. Venkataraman, R. Trestian, and H. X. Nguyen, "Digital twins: A survey on enabling technologies, challenges, trends and future prospects," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2255–2291, 4th Quart., 2022.

[5] A. Alkhateeb, S. Jiang, and G. Charan, "Real-time digital twins: Vision and research directions for 6G and beyond," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 128–134, Nov. 2023.

[6] A. Masaracchia, V. Sharma, B. Canberk, O. A. Dobre, and T. Q. Duong, "Digital twin for 6G: Taxonomy, research challenges, and the road ahead," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 2137–2150, 2022.

[7] X. Lin, L. Kundu, C. Dick, E. Obiodu, T. Mostak, and M. Flaxman, "6G digital twin networks: From theory to practice," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 72–78, Jun. 2023.

[8] H. Ahmadi, A. Nag, Z. Khar, K. Sayrafian, and S. Rahardja, "Networked twins and twins of networks: An overview on the relationship between digital twins and 6G," *IEEE Commun. Standards Mag.*, vol. 5, no. 4, pp. 154–160, Dec. 2021.

[9] J. Morais and A. Alkhateeb, "Localization in digital twin MIMO networks: A case for massive fingerprinting," 2024, *arXiv:2403.09614*.

[10] S. Jiang and A. Alkhateeb, "Digital twin aided massive MIMO: CSI compression and feedback," 2024, *arXiv:2402.19434*.

[11] Y. Li, C. Zhang, Y. Huang, and Q. Zheng, "Optimization of broadcast beams in massive MIMO: Learning from a digital twin," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Mar. 2023, pp. 2043–2048.

[12] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," Feb. 2019, *arXiv:1902.06435*, doi: 10.48550/arXiv.1902.06435.

[13] M. Arnold, B. Major, F. V. Massoli, J. B. Soriaga, and A. Behboodi, "Vision-assisted digital twin creation for mmWave beam management," 2024, *arXiv:2401.17781*.

[14] B. Salehi, U. Demir, D. Roy, S. Pradhan, J. Dy, S. Ioannidis, and K. Chowdhury, "Multiverse at the edge: Interacting real world and digital twins for wireless beamforming," *IEEE/ACM Trans. Netw.*, vol. 32, no. 4, pp. 3092–3110, Aug. 2024.

[15] B. Akgun, D. S. M. Singh, S. Kotla, V. Jain, S. Namdeo, R. Acharya, M. Jayabalan, A. Kumar, V. Chande, A. Kannan, J. Swami, Y. Chen, J. Boyd, and X. Zhang, "Interference-aware intelligent scheduling for virtualized private 5G networks," *IEEE Access*, vol. 12, pp. 7987–8003, 2024.

[16] O-RAN Work Group 1, "O-RAN architecture description 11.0," O-RAN Alliance, Germany, Tech. O-RAN.WG1.OAD-R003-v11.00, 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[17] O-RAN Work Group 10, "O-RAN operations and maintenance interface specification 12.0," O-RAN Alliance, Germany, Tech. O-RAN.WG10.O1-Interface.0-R003-v12.00, 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[18] O-RAN Work Group 2, "O-RAN A1 interface: Application protocol 4.01," O-RAN Alliance, Germany, Tech. O-RAN.WG2.A1AP-v04.01, 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[19] Qualcomm Edgewise Suite. (2024). *Leading RAN Automation and Management Platform*. [Online]. Available: https://www.qualcomm.com/products/internet-of-things/networking/cellular-networks/qualcomm-edgewise-suite

[20] O-RAN Work Group 2, "O-RAN non-RT RIC architecture 5.0," O-RAN Alliance, Germany, Tech. O-RAN.WG2.Non-RT-RIC-ARCH-R003-v05.00, 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[21] *Network Configuration Protocol (NETCONF)*, document RFC 6241, IETF, Jun. 2011.

[22] *The YANG 1.1 Data Modeling Language*, document RFC 7950, IETF, Aug. 2016.

[23] 3GPP, "Management and orchestration; Generic management services," 3rd Gener. Partnership Project (3GPP), v18.2.0, Tech. TS 28.532, 2024.

[24] 3GPP, "Telecommunication management; generic network resource model (NRM) integration reference point (IRP); information service (IS)," 3rd Gener. Partnership Project (3GPP), v18.7.0, Tech. TS 28.622, 2024.

[25] 3GPP, "Management and orchestration; 5G network resource model (NRM); stage 1," 3rd Gener. Partnership Project (3GPP), Tech. Rep. TS 28.541, v18.6.0, 2024.

[26] Helm. (2024). *The Package Manager for Kubernetes*. [Online]. Available: https://helm.sh/

[27] O-RAN Work Group 3, "O-RAN e2 application protocol (E2AP) 5.0," O-RAN Alliance, Germany, Tech. Rep. O-RAN.WG3.E2AP-R003-v05.00, 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[28] O-RAN Alliance, "O-RAN e2 service model (E2SM) 5.0," Germany, Tech. O-RAN.WG3.E2SM-R003-v05.00, 2024. [Online]. Available: https://orandownloadsweb.azurewebsites.net/specifications

[29] *Stream Control Transmission Protocol*, document RFC 9260, IETF, 2022.

[30] Open Drone Map. (2018). *Open Source Toolkit for Processing Aerial Imagery*. [Online]. Available: https://www.opendronemap.org/odm/

[31] (Jul. 2018). *CloudCompare*. [Online]. Available: https://www.cloudcompare.org/main.html

[32] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proc. 4th Eurographics Symp. Geometry Process.*, 2006, vol. 7, no. 4, pp. 1–10.

[33] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: Principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015.

[34] International Telecommunication Union. (2023). *Effects of Building Materials and Structures on Radiowave Propagation Above About 100 Mhz*. [Online]. Available: https://www.itu.int/dmspubrec/itu-r/rec/p/R-REC-P.2040-3-202308-I

[35] Remcom. (2024). *Wireless InSite 3D Wireless Propagation Software*. [Online]. Available: https://www.remcom.com/wireless-insite-propagation-software

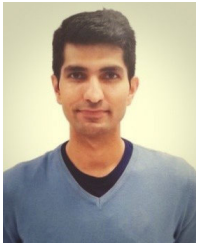
[36] 3GPP, "NR and NG-RAN overall description; stage-2," 3rd Gener. Partnership Project (3GPP), Tech. Rep. TS 38.300, v18.1.0, Mar. 2024.

[37] 3GPP, "Radio resource control (RRC) protocol specification (Release 15)," Tech. Rep. TS 38.331, v15.3.0, 2018.

[38] 3GPP, "Physical layer procedures for data," Tech. Rep. TS 38.214, v16.2.0, Jul. 2020.

BERK AKGUN received the B.S. and M.S. degrees in electrical and electronics engineering from Middle East Technical University, Ankara, Türkiye, in 2012 and 2014, respectively, and the Ph.D. degree in ECE from The University of Arizona, Tucson, AZ, USA, in 2019. From 2012 to 2014, he was a Software Design Engineer with the Communication and Information Technologies Division of Aselsan, Ankara. He is currently a Senior Engineer with Qualcomm Technologies Inc. His research interests include mmWave channel characterization, robust mmWave system design, secure multiuser MIMO systems, and wireless communications and networking, with an emphasis on designing and integrating next-generation private networks.

ADITYA JOLLY received the bachelor's degree in electronics and communications engineering from Indian Institute of Information Technology Guwahati (IIITG), India, in 2021, with a focus on wireless communication and signal processing, and the master's degree in electrical engineering from Columbia University, New York, NY, USA, in 2022, with a specialization in wireless and mobile communication. He joined Qualcomm Technologies as a System Integration and a Test Engineer after graduation. His research interests include full-duplex communication, intelligent reflective surfaces, advanced multiple access, and wireless communication, with an emphasis on the integration of these technologies into next-generation cellular networks. He was awarded the Chairman's Gold Medal at IIITG and the Tesla Scholarship at Columbia University.

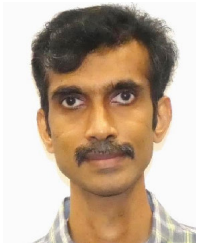


BALWINDER SACHDEV received the B.E. degree from Mumbai University, India, in 2005, and the M.S. degree in electronics and telecommunication engineering from Rutgers University, NJ, USA, in 2008, with a focus on digital signal processing and wireless communications. In 2008, he joined Qualcomm Technologies Inc., San Diego, CA, USA, where he is currently a Senior Staff Engineer, contributing to 5G NR research and prototyping efforts. Prior to joining Qualcomm, he gained valuable industry experience through internships with Broadcom and Conexant, where he worked on WCDMA and ADSL technology, respectively. His research interests include the application of digital twin technology to wireless communication, the Industrial IoT, RF sensing, ML-based precise positioning, and open-RAN-based private 5G networks.

DIVYA RAVICHANDRAN received the B.E. degree in electronics and communications from Madras University, India, in 2004, and the M.S. degree in electrical engineering from San Diego State University, in 2007, with a specialization in wireless communications and coding theory. She has been employed with Qualcomm Technologies Inc., USA, since 2008, where she holds the role of a Staff Engineer. Her significant work includes the testing and characterization of 3G/4G/5G prototype systems, with a specific focus on the PHY/MAC layers.

ROOHOLLAH AMIRI (Member, IEEE) received the B.Sc. and M.Sc. degrees (Hons.) in communication systems from Iran University of Science and Technology (IUST), in 2011 and 2013, respectively, and the Ph.D. degree from Boise State University, in 2020. Since then, he has been with Qualcomm Wireless Research and Development (WRD), San Diego, CA, USA. His research interests include communication systems, precise indoor positioning, joint communication and sensing, digital twins, and the application of artificial intelligence to the design and deployment of wireless networks.

VIKAS JAIN received the B.E. degree in electronics and communication engineering from the Thapar Institute of Engineering and Technology, India, in 1995, and the M.S. degree in computer science from the University of California at San Diego, San Diego, CA, USA, in 2003. In 1998, he joined Qualcomm Technologies Inc., to contribute the design and prototyping of various advanced features as part of RAN software development for 3G CDMA EV-DO, 4G LTE, and 5G NR wireless technologies. He is currently the Senior Director of Technology of the Wireless Research and Development Group, Qualcomm Technologies Inc., where he works on the software design and development for the 5G next-generation RAN research platform focusing on RAN disaggregation, virtualization, ML-enabled optimizations, and network automation. Prior to Qualcomm Technologies Inc., he contributed to the software development for digital telephony switching systems with C-DOT, India. He holds 24 granted U.S. patents and more than 120 granted international patents.



MURUGANANDAM JAYABALAN received the B.E. degree in electronics and communication engineering from the Sathyabama Engineering College, India, in 2000, and the M.S. degree in electrical engineering from Wichita State University, Wichita, KS, USA, in 2004. In 2006, he joined Qualcomm Technologies Inc., where he is currently the Director of Engineering and leading the System Integration and Test Team, with a focus on RAN disaggregation on 5G. He is also involved in ultra-mobile broadband (UMB), 4G LTE, and 5G NR: coordinated multipoint, precise positioning, the Industrial IoT, and private networks.

YITAO CHEN received the B.S. degree in ECE from Shanghai Jiao Tong University, in 2014, and the Ph.D. degree in ECE from The University of Texas at Austin, in 2020. He is currently a Senior Engineer with Qualcomm Technologies Inc., where he has contributed to the design, prototyping, and standardization of 5G cellular wireless technologies of MIMO, the Industrial IoT, and wireless ML research. He holds 13 U.S. patents. His research interests include wireless communication, information theory, coding theory, signal processing, and machine learning.

HETAL PATHAK received the B.E. degree from Mumbai University, India, and the M.S. degree in electrical and computer engineering from the University of Wisconsin–Madison, USA. She joined Qualcomm Technologies Inc., USA, in 2005, where she is currently a Senior Staff Engineer and the Manager. She has worked on integrating, testing, and characterizing the performance of multiple advanced wireless technologies, networks, and systems, such as 5G NR, LTEU, LTE, NGRAN, private networks, location-based services, and multi-media broadcast systems. She has also led technology trials for private LTE networks and the Industrial IoT.

VINAY CHANDE received the degree in engineering from Indian Institute of Technology Bombay, Mumbai, and the Ph.D. degree in electrical engineering from the University of Maryland. As a Systems Engineer with the Wireless Research and Development Group, Qualcomm Technologies Inc. His current work allows him to participate in and witness the advances in millimeter-wave radio bands, unlicensed spectrum access, and machine learning for the Industrial IoT.

MOHAMMAD FAHIM received the Ph.D. degree in electrical engineering from Pennsylvania State University, USA. He is currently a Systems Engineer with the Wireless Research and Development Department, Qualcomm Technologies Inc. He is working on the development of new technologies for 5G and beyond cellular systems, especially in the areas of the Industrial IoT, precise positioning, and digital twins.

SRINIVAS YERRAMALLI received the B.Tech. degree from the International Institute of Information Technology, Hyderabad, India (IIIT-H), in 2007, and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, CA, USA, in 2013. He is currently with Qualcomm Wireless Research and Development (WRD), San Diego, CA, USA. He has also attended 3GPP RAN1 as a delegate for multiple topics (unlicensed LTE and NR, precise positioning and AI/ML positioning, and framework) and holds more than 450 granted U.S. patents and many international patents in various areas of wireless system design. His current 6G research interests include radio and RAN network digital twin construction and using the digital twin for efficiently designing, deploying, and improving wireless networks, AI/ML algorithms for wireless systems, including standardization in 3GPP, precise positioning, and joint communication and sensing related topics.

RUPESH ACHARYA received the B.E. degree in electronics and communication from the Institute of Engineering, Pulchowk Campus, Nepal, and the M.S. degree in electrical engineering from The University of Texas at Dallas. In 2009, he joined Qualcomm Technologies Inc., where he is currently a Senior Staff Engineer with the System Integration and Test Team. He has contributed to 3G, 4G, 5G, and Wi-Fi-related end-to-end system integration and testing.

CHANDRESH TIWARI received the M.S. degree in electrical communication engineering from Indian Institute of Science (IISc), Bengaluru, India, in 1997. In 2009, he joined Qualcomm Technologies Inc., to contribute the design and prototyping of various advanced features as part of RAN software development for 4G LTE, 4G SON, and 5G NR wireless technologies. He is currently a Principle Engineer with the Wireless Research and Development Group, Qualcomm Technologies Inc., where he works on the software design and development for the 5G next-generation RAN research platform focusing on RAN disaggregation and RAN-control plane design and development. Prior to Qualcomm Technologies Inc., he worked in multiple wireless protocol stacks (CDMA, WCDMA, UMTS, LTE, and WiFi) in UE and RAN side in different companies (including Motorola, Comverse Network Systems, Lucent Technologies, Ericsson, Nokia, and Freescale).

CONNOR WOODAHL received the B.S. degree in computer science from Baylor University, Waco, TX, USA, in 2021. He has been a Software Engineer with the Wireless Research and Development Department, Qualcomm Technologies Inc., since February 2022. He has contributed to the design and prototyping of various projects related to precise positioning, RF sensing, and the creation of digital twins.

ARUMUGAM KANNAN received the B.E. degree in electronics and communication engineering from Anna University, India, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from Georgia Institute of Technology, in 2004 and 2007, respectively. He is currently a Principal Engineer with Qualcomm Technologies Inc., where he has contributed to the design, prototyping, and standardization of multiple LTE and 5G cellular wireless technologies ranging from advanced receivers, shared unlicensed spectrum operation of cellular systems, millimeter wave communications, the Industrial IoT, and wireless AI/ML research. He holds more than 200 granted U.S. patents and more than 600 granted international patents.

XIAOXIA ZHANG received the B.S. and M.S. degrees in ECE from the University of Science and Technology of China and the Ph.D. degree in ECE from The Ohio State University, in 2002. After graduation, she joined Qualcomm Technologies Inc., where she is currently the Senior Director. She is involved in 3G, 4G, and 5G, including PHY/MAC system design, evaluation, prototyping, and standardization. She is leading the project focusing on the Industrial IoT, shared/unlicensed spectrum, and RAN disaggregation on 5G/6G technology. She is a prolific inventor with more than 300 granted U.S. patents.

DEEPU ALEX received the B.Tech. degree in electronics and communications engineering from the National Institute of Technology, Calicut, India, and the M.S. degree in electrical and computer engineering from Texas A&M University, College Station, TX, USA. He is currently a Staff Engineer with Qualcomm Technologies Inc., with expertise in developing and verifying 3G/4G/5G technology prototypes.

ABHISHEK KUMAR received the master's degree in computer applications from Delhi University, India, in 2006. After completing the master's degree, he joined Hughes Systique Corporation, where he worked on various projects in developing 3G/4G-based UE and RAN side protocol stacks for the LEO and GEO satellite communications. In 2019, he joined Qualcomm Research and he has been contributing in the design and development of 5G and 6G cellular wireless technologies of the Industrial IoT, private networks, and RAN disaggregation.

HAI HONG received the B.S. degree in computer science and engineering from the University of California at San Diego. He is currently a Software Engineer with Qualcomm Technologies Inc. He has been contributing to the test tools and automation software development effort for the 5G and 6G cellular wireless technologies.

JOHN BOYD received the B.S. degree in computer science and engineering from the University of California at San Diego, San Diego, CA, USA, in 1994. He is currently the Director of Engineering of Qualcomm Technologies Inc., where he is also a Software Engineer and the Manager of the Wireless Research and Development Group. Having the opportunity to contribute to a diverse range of projects throughout the company during his 27 years of service. His current focus is a machine learning framework that facilitates the development of RAN optimizations.

RAJAT PRAKASH received the B.Tech. degree in electrical engineering from Indian Institute of Technology Kanpur, the M.S. degree in electrical engineering from Cornell University, and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana–Champaign. He is currently with the Wireless Research and Development Group, Qualcomm. His current work focuses on digital twin, virtualized RAN, and the Industrial IoT technologies, and on the evolution of 5G towards 6G. He has previously worked on small cells, self-organizing networks, neutral hosts, VoLTE, and VoWiFi technologies. He has participated in several industry bodies, including 3GPP, O-RAN alliance, CBRS alliance, multifile alliance, small cell forum, 5G ACIA, IEEE, and 3GPP2.

SURESH BABU MUMMANA received the B.Tech. degree in electronics and communications engineering from Jawaharlal Nehru Technology University, India, in 2005, and the M.S. degree in electrical engineering from University at Buffalo, The State University of New York, in 2008, with a major in wireless communication and MIMO systems. He is currently pursuing the M.B.A. degree with Kelly School of Business, Indiana University Bloomington, with a major in strategy and corporate entrepreneurship. He is a Senior Staff Software Engineer and the Manager of the Wireless Research and Development Group, Qualcomm Technologies Inc., where he works on the software design and development of 5G systems, wireless digital twin, and non-data related wireless features, such as precise positioning and RF sensing.

SUMANTH GOVINDAPPA received the B.E. degree in computer science from Bangalore Institute of Technology, India, in 2002, and the M.S. degree in computer science from The University of Texas at Dallas, USA, in 2006. He was an Intern with Ericsson, in 2007, and then joined Qualcomm Technologies Inc., in 2008. With Qualcomm, he has continuously contributed to the design, prototyping, and standardization of multiple features in WCDMA, LTE, and 5G cellular wireless technologies. He holds six granted U.S. patents and 16 granted international patents. His current research focus is RAN disaggregation, where he leads the team in prototyping the CU-UP network function. His research interests include cellular wireless technologies and AI in wireless.

JAMES Y. WILSON received the B.S. degree in business computer methods from California State University at Long Beach, Long Beach, CA, USA, in 1994. He is currently a Senior Staff Software Engineer with Qualcomm Technologies Inc., where he designs and implements the compute cluster used by the RAN Network Function Engineering Team. During his 15 years of service at Qualcomm, he has also contributed to the development of the system-level software used in a variety of products throughout Qualcomm.

JALAJ SWAMI received the B.Tech. degree in computer science and engineering from IIT-Kanpur, India, and the M.S. degree in information and computer science from the University of Hawaii at Manoa. He initially worked on Telecom Software for C-DOT, India. He is currently a Software Engineer with Qualcomm, where he has worked on device software development and infrastructure software for wireless, broadcast, and satellite communications. He is contributing to 5G and 6G development for the Industrial IoT and RAN disaggregation.

VIVIAN PHAM received the B.S. degree in computer science from the University of California at San Diego, San Diego, CA, USA, in 2019. She is currently a Senior Engineer with Qualcomm Technologies Inc., where she has contributed to various tools and frameworks used in the testing, data collection, and integration of multiple advanced wireless 5G and 6G systems. Her current work focuses on designing and implementing the NG-RAN deployment and orchestration framework.

ANDREI VADEANU received the B.S. and M.S. degrees in electronics and telecommunications engineering from Politehnica University of Bucharest, Romania, in 2012 and 2014, respectively. From 2012 to 2019, he was a RAN Engineer for different European operators, being in charge of testing, optimization, and configuration of the 2G, 3G, 4G, and 5G technologies. He is currently a Principal Engineer with the Wireless Technology Group, with a focus on research, design, and development for different ML optimizations, network automation techniques, and use cases of commercial interest to Qualcomm Edgewise products.

GILAD GOVRIN received the B.Sc. degree in information systems from Ben Gurion University, Israel. He is currently a Principal Engineer and the Manager of Qualcomm Technologies Inc., where is he leading the Architecture Team for the Edgewise platform, a RAN service automation, management, and orchestration (SMO) platform. He leads the evolution of cutting-edge Gen-AI technologies and software architecture. Prior to joining Qualcomm through acquisition, in 2022, he served as the Chief Architect of the Edgewise platform with Cellwize, for several years, where he played a key role in building and deploying Edgewise to several global mobile network operators.