End-to-End Simulation of 5G mmWave Networks

Marco Mezzavilla, *Member, IEEE*, Menglei Zhang, Michele Polese , *Student Member, IEEE*,
Russell Ford, *Member, IEEE*, Sourjya Dutta, *Student Member, IEEE*,
Sundeep Rangan, *Fellow, IEEE*, and Michele Zorzi, *Fellow, IEEE*

Abstract—Due to its potential for multi-gigabit and low latency wireless links, millimeter wave (mmWave) technology is expected to play a central role in 5th generation (5G) cellular systems. While there has been considerable progress in understanding the mmWave physical layer, innovations will be required at all layers of the protocol stack, in both the access and the core network. Discrete-event network simulation is essential for end-to-end, cross-layer research and development. This paper provides a tutorial on a recently developed full-stack mmWave module integrated into the widely used open-source ns–3 simulator. The module includes a number of detailed statistical channel models as well as the ability to incorporate real measurements or ray-tracing data. The physical and medium access control layers are modular and highly customizable, making it easy to integrate algorithms or compare orthogonal frequency division multiplexing numerologies, for example. The module is interfaced with the core network of the ns–3 Long Term Evolution (LTE) module for full-stack simulations of end-to-end connectivity, and advanced architectural features, such as dual-connectivity, are also available. To facilitate the understanding of the module, and verify its correct functioning, we provide several examples that show the performance of the custom mmWave stack as well as custom congestion control algorithms designed specifically for efficient utilization of the mmWave channel.

Index Terms—mmWave, 5G, cellular, channel, propagation, PHY, MAC, multi-connectivity, handover, simulation, ns–3.

I. INTRODUCTION

MILLIMETER Wave (mmWave) communications are emerging as a key technology in 5th generation (5G) cellular wireless systems due to their potential to achieve the massive throughputs required by future networks [1]–[5]. In particular, mmWave has become a key focus of the

M. Mezzavilla, M. Zhang, S. Dutta, and S. Rangan are with NYU WIRELESS, New York University Tandon School of Engineering, Brooklyn, NY 11201 USA (e-mail: mezzavilla@nyu.edu; menglei@nyu.edu; sdutta@nyu.edu; srangan@nyu.edu).

M. Polese and M. Zorzi are with the Department of Information Engineering, University of Padova, 35131 Padua, Italy (e-mail: polesemi@dei.unipd.it; zorzi@dei.unipd.it).

R. Ford was with NYU WIRELESS, New York University Tandon School of Engineering, Brooklyn, NY 11201 USA. He is now with Samsung Research America, Mountain View, CA 94043 USA (e-mail: russell.ford@samsung.com).

3rd Generation Partnership Project (3GPP) NR[1] effort currently under development [7]. Due to the unique propagation characteristics of mmWave signals and the need to transmit in beams with much greater directionality than previously used in cellular systems, much of the recent work in mmWave communications has focused on channel modeling, beamforming and other physical layer procedures. However, the design of End-to-End (E2E) cellular systems that can fully exploit the high-throughput, low-latency capabilities of mmWave links will require innovations not only at the physical layer, but also across all layers of the communication protocol stack. For mmWave systems, E2E design and analysis are at a much earlier stage of research [8]–[10].

Discrete-event network simulators are fundamental and widely used tools for the development of new protocols and the analysis of complex networks. Importantly, most network simulators enable *full-stack simulation*, meaning that they model all layers of the protocol stack as well as applications running over the network. This full-stack capability will play a critical role in the development of 5G mmWave systems. The unique characteristics of the underlying mmWave channel have wide ranging effects throughout the protocol stack. For example, the use of highly directional beams increases the complexity of a number of basic Medium Access Control (MAC)-layer procedures such as synchronization, control signaling, cell search and initial access, which in turn affect delay and robustness [8]. MmWave signals are also highly susceptible to blockage [1], [11]–[13], which results in high variability of the channel quality. This erratic behavior complicates the design of rate adaptation algorithms and signaling procedures, requiring advanced solutions for multi-connectivity, fast handover and connection re-establishment [14]–[17]. New transport layer mechanisms may also be required in order to utilize the large capacity, when available, and to react promptly to rapid fading to avoid congestion [10], [18]–[20]. The need for ultra-low latency applications [1], [21], [22] may require solutions based on edge computing and distributed architectures that will determine a considerable departure from current cellular core network designs.

To better capture these design challenges, this work presents a comprehensive tutorial on the open-source mmWave simulation tool developed by New York University and the University of Padova for Long Term Evolution (LTE)-like 5G

---

[1]While NR was initially introduced as the acronym for "New Radio" [6], according to the latest 3GPP specifications [7] it has lost its original meaning and is now used to refer to the 5G Radio Access Network.

mmWave cellular networks, which can be used to evaluate cross-layer and end-to-end performance. This mmWave simulation tool is developed as a new module within the widely used ns–3 network simulator [23]. ns–3 is an open-source platform, that currently implements a wide range of protocols in C++, making it useful for cross-layer design and analysis. The new mmWave module presented here is based on the architecture and design patterns of the LTE LENA module [24], [25] and implements all the necessary Service Access Points (SAPs) needed to leverage the robust suite of LTE/Evolved Packet Core (EPC) protocols provided by LENA. The code (publicly available at GitHub [26], along with examples and test configurations) is highly modular and customizable to help researchers to design and test novel 5G protocols.

The ns–3 mmWave module was first presented in [27] and [28]. The 3GPP channel model implementation is introduced in [29], and the dual connectivity functionality is described in [15] and [30]. This paper extends these works by presenting the ns–3 mmWave module from a single and organic point of view, and is intended as a tutorial for any researcher that plans to use the simulator. In addition to its comprehensive description and discussion, we provide in Section X a brief guide on how to set up a simulation, followed by a number of representative examples.

The rest of the paper is organized as follows. In Section II, we provide some background on mmWave cellular communications and highlight some key problems at the higher protocol layers to motivate the need for a robust full-stack simulator. We also describe the main challenges related to the design of a mmWave cellular network simulator. Then, in Section III, we introduce ns-3, the network simulator on which our mmWave module is developed, and in Section IV we present the overall architecture of the mmWave module. We then take a closer look at each component, starting with the suite of Multiple Input, Multiple Output (MIMO) channel models in Section V. In addition to an implementation of the latest 3GPP "above 6 GHz" model [31], several custom channel models are also provided. Section VI discusses the features of the Orthogonal Frequency Division Multiplexing (OFDM)-based Physical layer, which has a customizable frame structure for evaluating different numerologies and parameters. In Section VII, we provide a MAC-layer discussion that includes our proposed flexible/variable Transmission Time Interval (TTI) Time Division Multiple Access (TDMA) MAC scheme, which is supported by several scheduler implementations. Section VIII presents the enhancements that we introduced to the LTE Radio Link Control (RLC) layer. The dual-connectivity architecture is reported in Section IX. In Section X, we show how the module can be used for cross-layer evaluation of multi-user cellular networks through a number of representative examples, and provide pointers to a large set of general results that have been obtained so far with this module. The integration of native Linux Transmission Control Protocol (TCP) implementations, performed through the ns–3 Direct Code Execution (DCE) framework, is discussed in Section XI. In Section XII, we provide details on our future plans for the simulator and

suggest possible research areas in which it could be used. Finally, we conclude this tutorial paper in Section XIII.

II. MILLIMETER WAVE CELLULAR BACKGROUND

Millimeter wave communication is an advanced PHY layer technology, which has recently come to the forefront of research interest and may be able to rise to the challenge of providing high-rate mobile broadband services, in addition to offering opportunities for reducing over-the-air latency for NR.

MmWave makes use of the radio frequency spectrum roughly between 30 and 300 GHz, even though the research challenges extend also to lower frequencies (i.e., above 6 GHz) which are considered for 3GPP NR. Systems that can operate in these bands are attractive because of the large quantities of available spectrum at these higher frequency ranges and the spatial degrees of freedom afforded by very high-dimensional antenna arrays, which are possible thanks to the smaller size of antenna elements at higher frequencies. Most current commercial wireless systems operate below 6 GHz, where lower frequencies allow for long-range propagation and low penetration loss (i.e., attenuation by walls and other obstacles), which makes them well-suited for radio communications. As a result, the sub-6 GHz spectrum has become heavily congested and individual bands are generally not available in contiguous chunks wider than 200 MHz. However, large swaths of spectrum are available at the higher mmWave frequencies, which offer the possibility of very wide bandwidths, in some cases even larger than 1 GHz.

Although the mmWave bands are already used by a variety of commercial applications, such as satellite and point-to-point backhaul communications, until recently they were considered impractical for mobile access networks due to the poor isotropic propagation and the vulnerability to shadowing at these higher frequencies. However, it has now been shown that the limitations of the mmWave channel can be overcome with the help of high-gain, directional antennas so that this vast region of spectrum can now be exploited to provide an order of magnitude or more increase in throughput for mobile devices [3], [32].

Directional smart antennas are the major technology enabler that will make it possible for mmWave devices to overcome the poor propagation effects and unlock this high-frequency spectrum. The theoretical free space path loss (as governed by Friis' Equation) is proportional to the square of the frequency, resulting in the magnitude of received power for a mmWave signal being over 30 dB (1000x) less than conventional cellular systems at equivalent distances between transmitter and receiver [33]. Multi-element antenna arrays and MIMO beamforming techniques offer a means of compensating for this high attenuation. With millimeter waves, the antenna size and spacing shrinks to be on the order of millimeters, making it possible to pack hundreds of elements onto a small cell base station and dozens onto a handheld device. Smaller antenna size also allows for multiple arrays to be integrated onto mobile devices to provide diversity and maintain connectivity even if the signal from one array is blocked (for instance, by the user's hand) [3].

It is clear that mmWave will be highly disruptive in the wireless space thanks to the prospect of massive bandwidth and high-gain antennas. Nevertheless, before mmWave technology can be effectively realized in 5G networks, there are numerous challenges to be addressed, not only at the physical layer, but also at higher layers of the radio stack, namely:

- *Adaptive beamforming and beam tracking:* The requirement of directionality introduces new challenges for supporting mobility in mmWave networks. The transmitter and receiver must continually track the channel as the mobile user moves in order to align their antenna arrays to achieve the maximum directional gain. MmWave signals are also known to be particularly susceptible to shadowing and can be completely blocked by many materials such as brick, tinted glass and even the human body [34], [35]. Fortunately, recent field measurements have demonstrated that reflected power can be sufficient for Non Line of Sight (NLOS) communications to be possible. A blocked link may therefore be able to recover by steering the beam from the primary Line of Sight (LOS) path to an alternate NLOS path. The User Equipment (UE) and base station must then jointly initiate a procedure to search for and select another path to reestablish the link.

- *Directional synchronization and broadcast channels:* Directionality also complicates the design of many control channels and procedures. The cell discovery and initial access procedures, where the UE must search for nearby base stations to which it can attach, will require an innovative approach to be handled efficiently. Traditional cells periodically broadcast synchronization signals (known as the Primary Synchronization Signal (PSS) in LTE systems) omnidirectionally, which are received by all devices within the cell's coverage range and used to initially connect to the cell. If a 5G mmWave evolved Node Base (eNB) were to broadcast the PSS with an omnidirectional antenna pattern, the signal would not benefit from the directional gain and might not have adequate range to be detected by many UEs. Therefore, the eNB and UE must perform an angular search in order for users to detect the PSS and hone in on the optimal Transmitter (TX)/Receiver (RX) beamforming angles [36]. A similar problem also arises for other control signals, such as the Downlink Control Information (DCI) assignments, which indicate the resources assigned to each user for Downlink (DL)/ Uplink (UL) transmission in the subframe or slot.

- *Issues for the MAC, Network and Transport Layers:* The rapid channel dynamics and vulnerability of mmWave links to shadowing will require frequent, near instantaneous handovers between neighboring 5G or 4G cells. *Dual-connectivity*, where mobiles are continuously connected to both the 5G and the legacy 4G network, may therefore be essential to recover from an abrupt failure of the primary 5G link [15], [30]. Additionally, at the transport layer, the congestion control and avoidance mechanisms provided by TCP must be able to quickly adapt to sudden fluctuations in capacity to maximally utilize the link bandwidth while avoiding overwhelming the network by sending too many packets, resulting in congestion and affecting other flows in the network. Current versions of TCP may not be optimized for mmWave channel dynamics [10], [19], so new algorithms may be called for to provide high rates for E2E sessions [18], [37], [38].

A. Potentials and Challenges of System-Level Simulations of mmWave Networks

An End-to-End network simulator for mmWave cellular networks is an invaluable tool that can help address these challenges by allowing the evaluation of the impact of the channel and of the PHY layer technology on the whole protocol stack. However, given the characteristics of mmWave communications described in the previous paragraphs, in order to have accurate results it is of paramount importance to model in detail the behavior of the different elements that interact in a cellular system. In the following paragraphs we will introduce and discuss some of the most important elements that need to be considered when designing a mmWave cellular system simulation, and show how they depend on one another:

- The channel model is the fundamental component of every wireless simulation. Given the harsh propagation conditions at mmWaves, the channel is one the main elements that affect the end-to-end network performance. Firstly, it has to account for the different LOS and NLOS states for the propagation loss and the fading [31]. Moreover, beamforming should be applied on top of the channel to accurately model directional transmissions, which have an impact on the link budget, the interference, and the control procedures. Finally, the Doppler effect is particularly relevant at mmWave frequencies, especially with high mobility [3]. An important consideration related to the channel model is the trade off between the accuracy and the computational complexity: very accurate models that require the computation of the complete channel matrix are usually also computationally intensive [29].

- The users' mobility and the network deployment have an important impact on the communication performance, intertwined with that of the channel model. Given the small range of the mmWave cells, the deployment will be dense and will require frequent access point updates, which should be simulated for a realistic performance assessment [39]. Moreover, mobility affects the performance of beam tracking algorithms [16]. Therefore, when simulating a mmWave network it is important to use realistic deployments and mobility models.

- The level of detail when modeling the protocol stack of the mmWave links and of the end devices is another important parameter for network simulations. A simplified model of the protocol stack can be accurate enough for studies that involve limited interplay between different layers, but cannot capture the behaviors that emerge from complex interactions among them, and therefore may not be sufficient to generate realistic results for end-to-end performance evaluations. For example, at mmWave frequencies, it has been shown that the channel behavior has

an impact on the TCP performance [10], [19], therefore a model of the TCP/IP stack is needed when analyzing the data rate that an application can reach in an end-to-end mmWave network.

To the best of our knowledge, there are no open source simulators capable of thoroughly modeling the mmWave channel along with the cellular network protocol stack as well as other protocols (e.g., the TCP/IP stack), realistic scenarios and mobility. There exists an ns–3-based simulator for IEEE 802.11ad in the 60 GHz band [40], [41], which however cannot be used to simulate cellular and 3GPP-like scenarios. Other papers [42]–[45] report results from system level simulations, with custom (often MATLAB-based and not publicly available) simulators which are not able to capture the complexity of the whole stack with a very high level of detail. This is what motivated us to develop an open source cellular mmWave module for the ns–3 simulator, which we will describe in the following sections.

III. NS–3

The ns–3 discrete-event network simulator [23], [46] is a very powerful tool available to communication and networking researchers for developing new protocols and analyzing complex systems. It is the successor to ns–2, a well-tested tool that has been in use by the networking community for over a decade in the design and validation of network protocols. ns–3 is open source, and can be downloaded from the website of the project.[2] An active community of researchers from both industry and academia has enriched the basic core of the simulator with several modules, and ns–3 can now be used to simulate a wide variety of wireless and wired networks, protocols and algorithms. There is a complete documentation[3] on the models in the ns–3 website, in terms of both the design of the models and what a user can do with the models. Moreover, a complete tutorial on how to install ns–3, set up ns–3 scenarios and topologies, handle the collection of statistics and log useful messages is provided in the documentation.[4] The tutorial is a good starting point for a researcher who approaches ns–3 for the first time.

The ns–3 simulator is organized into multiple folders. The src folder provides a collection of C++ classes, which implement a wide range of modular simulation models and network protocols. The different modules can be aggregated and instantiated to build diverse simulated network scenarios, making ns–3 especially useful for cross-layer design and analysis. The modularity and use of object-oriented design patterns also allow for new algorithms to be easily incorporated into the network stack and experimented with. Each module is itself organized into multiple subfolders, which contain the documentation and the source code of the model itself, the helpers, the examples and the tests. The helpers associated with each model have a very important role. They are classes which hide to the final user the complexity involved in setting up a complete scenario, for example by automatically assigning IP addresses, or connecting the different classes of a protocol

stack. The build folder contains the binaries of the simulator. Finally, the scratch folder is a special folder in which scripts with examples and scenarios can be built on-the-fly.

Besides the core module, which provides the basic structure of the simulator, there are modules for networking protocols (e.g., the TCP/IP stack protocols [47]), wireless protocols (LTE [24], Wi-Fi [48], WiMAX [49]), routing algorithms [50], mobility, embedding obstacles and buildings in the simulation scenarios, and data collection. All the modules are listed in the model library.[5]

In the following sections, we will describe in detail the mmWave module for ns–3, following the same approach which is used for the other ns–3 modules. We will first describe the model in terms of implementation of the different components of a mmWave cellular network and protocol stack, and then the examples and scenarios that can be simulated with it and how they can be set up.

IV. MMWAVE MODULE OVERVIEW

The ns–3 mmWave module is designed to perform end-to-end simulations of 3GPP-style cellular networks. As shown in Fig. 1, the architecture builds upon the ns–3 LTE module (LENA) [24], [25]. It leverages the detailed implementation of LTE/EPC protocols, and implements custom PHY and MAC layers. Additionally, it is possible to connect the module to a patched version of Direct Code Execution [51], a tool that allows the Linux stack TCP/IP implementation to run as the TCP/IP stack of ns–3 nodes, as well as to execute POSIX socket-based applications (i.e., wget, iPerf, etc). Figure 1 depicts the high-level composition of the MmWaveEnbNetDevice and MmWaveUeNetDevice classes, which represent the mmWave eNB[6] and UE radio stacks, respectively, along with a perspective on the end-to-end structure of the simulator. A more detailed UML class diagram is provided in Figure 2.

The ns–3 mmWave module also includes a McUeNetDevice, which is a NetDevice with a dual stack (LTE and mmWave), i.e., a device capable of connecting to both technologies. More details will be given in Section IX.

The MmWaveEnbMac and MmWaveUeMac MAC layer classes implement the LTE module Service Access Point (SAP) *provider* and *user* interfaces, which enable the inter-operation with the LTE RLC layer. Support for RLC Transparent Mode (TM), Saturation Mode (SM), Unacknowledged Mode (UM), Acknowledged Mode (AM) is built into the MAC and scheduler classes (i.e., MmWaveMacScheduler and derived classes). The MAC scheduler also implements a SAP for configuration at the LTE Radio Resource Control (RRC) layer (LteEnbRrc). Hence, every component required to establish Evolved Packet Core (EPC) connectivity is available.

The MmWavePhy classes handle directional transmission and reception of the DL and UL data and control channels

---

[2]http://www.nsnam.org

[3]https://www.nsnam.org/documentation/

[4]https://www.nsnam.org/docs/tutorial/html/

[5]https://www.nsnam.org/docs/release/3.27/models/html/index.html for ns–3 version 3.27

[6]Recently, 3GPP has proposed the term Next Generation Node Base (gNB) for the 5G NR base station.
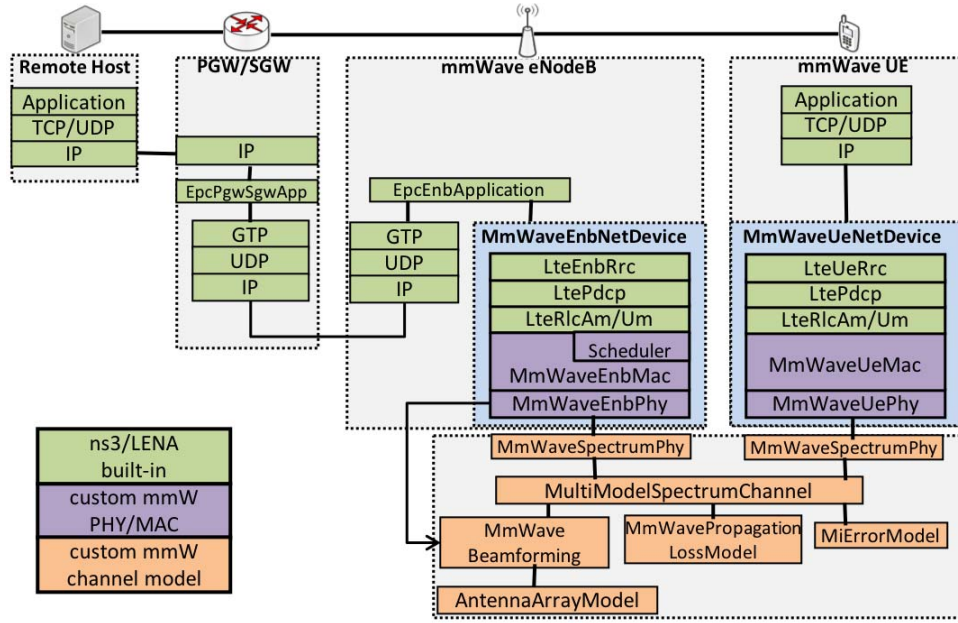
Fig. 1. Class diagram of the end-to-end mmWave module.

based on control messages from the MAC layer. Similar to the LTE module, each PHY instance communicates over the channel (i.e., `SpectrumChannel`) via an instance of the `MmWaveSpectrumPhy` class, which is shared for both the DL and the UL (since our design of the mmWave PHY layer is based on Time Division Duplexing (TDD), as detailed in Section VI-A). Instances of `MmWaveSpectrumPhy` encapsulate all PHY-layer models: interference calculation (`MmWaveInterference`), Signal to Interference plus Noise Ratio (SINR) calculation (`MmWaveSinrChunkProcessor`), the Mutual Information (MI)-based error model (`MmWaveMiErrorModel`), which computes the packet error probability, as well as the Hybrid Automatic Repeat reQuest (HARQ) PHY-layer entity (`MmWaveHarqPhy`) to perform soft combining.

Since the structure, high-level functions and naming scheme of each class closely follow the LTE LENA module, the reader is also referred to the LENA project documentation for more information [52].

V. CHANNEL AND MIMO MODELING

A. Channel Models

The ns–3 mmWave module allows the user to choose among different channel models, which provide a trade-off between computational complexity, flexibility and accuracy of the results. The most flexible and detailed channel model is the one described in detail in [29], which is based on the official 3GPP channel model for the 6-100 GHz frequency band [31]. It accounts also for spatial consistency of mobility-based simulations and provides a random blockage model, as well as the modeling of outdoor to indoor communications. The second model is based on traces from measurements or third-party ray-tracing software. This makes the channel model detailed and realistic, but constrains the simulation to limited measurements/ray-tracing routes. The third is the statistical channel model introduced in [27] and based on MATLAB traces, which makes the computation less demanding, but is available only for the 28 and 73 GHz frequencies. In the following paragraphs we will provide architectural details of all the available channel models.

1) 3GPP Statistical Channel Model: The 3GPP model for the 6-100 GHz band, described in [31], is applicable for bandwidths up to 10% of the carrier frequency and accounts for mobility. It provides several optional features that can be plugged into the basic model, in order to simulate, for example, spatial consistency (i.e., the radio environment conditions of close-by users are correlated) and random blockage. The model defines different scenarios, which describe different possible cellular network deployments: urban (with macrocells and microcells), rural and indoor.

Pathloss: The pathloss of the propagation channel is implemented in the `MmWave3gppPropagationLossModel` class. The model provides a statistical **LOS/NLOS** condition characterization, as well as pathloss computation considering outdoor to indoor penetration loss, as described in [31, Sec. 7.4]. The `MmWave-3gppBuildingPropagationLossModel` class, instead, determines the LOS condition according to the relative position of the UE and the eNB and to the presence of buildings or obstacles in the scenario. These classes also optionally apply an additional shadowing component to the pathloss. For a moving UE, the shadowing is correlated in space. Given the distance $\Delta d_{2D} > 0$ on the horizontal plane from the last position in which the shadowing was computed, the exponential correlation parameter is computed as $R(\Delta d_{2D}) = e^{-\Delta d_{2D}/d_{cor}}$, where d_{cor} is the correlation distance. In our implementation, pathloss and shadowing (if enabled) are updated at every transmission. Figure 3 shows the pathloss in dB for the 3D distance from the smallest value
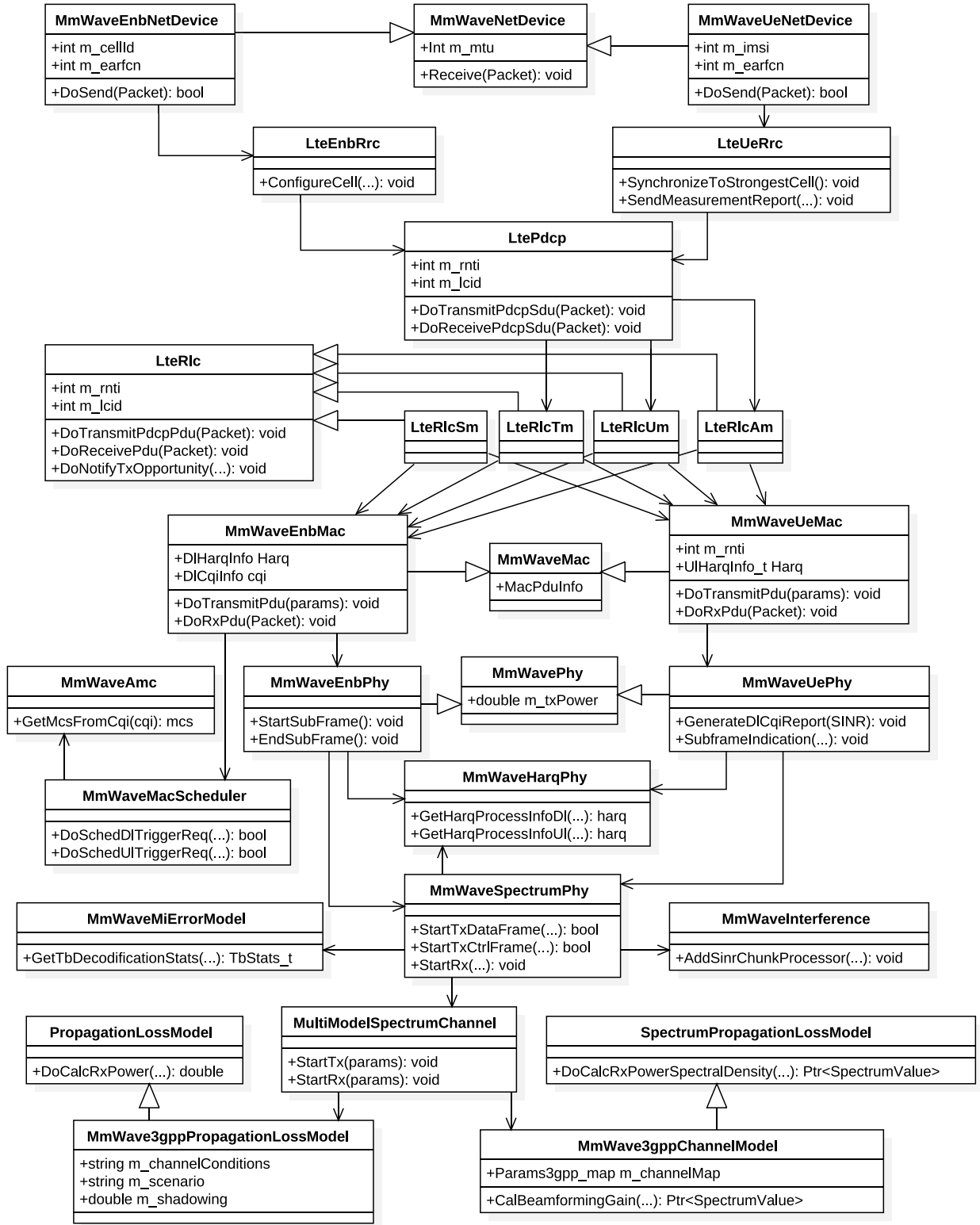
Fig. 2. Unified Modeling Language (UML) class diagram for the end-to-end mmWave module.

supported in each scenario to 1 km for outdoor and 100 m for indoor.

Small-scale fading: The small-scale fading model is implemented in the `MmWave3gppChannel` class, and follows the step by step approach of [31, Sec. 7.5]. Small-scale fading is the bottleneck of this channel model implementation, since it is very detailed and computationally demanding. The fading is

generated following the 3D statistical spatial approach originally proposed in [53]. The channel is described by a channel matrix $\mathbf{H}(t,f)$, where t is the time and f is the frequency, of size $U \times S$, where U and S are the number of antennas at the receiver and the transmitter. Each entry depends on $N \leq 20$ different multipath components, called *clusters*, which have different delays and received powers, according to an
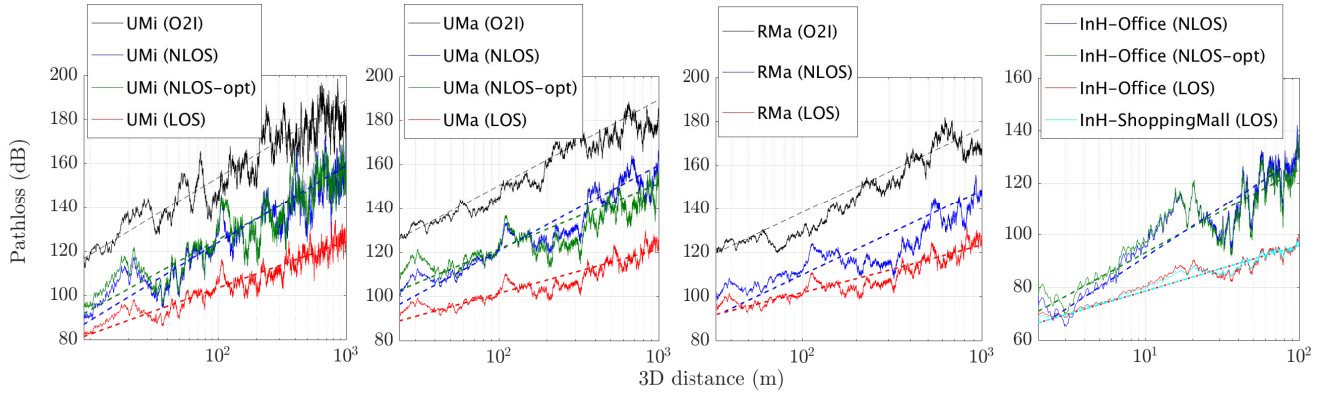
Fig. 3. Typical realizations of the 3GPP pathloss model described in [31]. We consider three outdoor scenarios (UMi stands for Urban Micro, UMa for Urban Macro, and RMa for Rural Macro) and two indoor scenarios (InH stands for Indoor Hotspot, either in the office or in a shopping mall). The realizations for each single 3GPP scenario differ because of the channel condition between the UE and the base station: it can be a LOS, NLOS or O2I channel. Moreover, 3GPP specifies an additional set of equations for the NLOS channel in the UMi, UMa and InH-Office scenarios, which are marked as optional (opt).

exponential power delay profile. A cluster is itself a combination of $M = 20$ *rays*, each with a slightly different arrival and departure angle in the vertical and horizontal planes.

The `MmWave3gppChannel` class has a method that generates the channel matrix, and stores the coefficient for each transmit element s, receive element u and cluster n in a data structure, that can be accessed by other methods in order to update the channel matrix or compute the beamforming gain. We introduced some assumptions with respect to the 3GPP model, in order to decrease the computational overhead introduced by the high level of detail of the channel. For example, we consider only antennas with vertical polarization, and the speed-dependent Doppler effect is not computed for each ray, but only for the central angle of each cluster. Further details on this implementation are given in [29].

Spatial consistency: The basic channel model described in the previous paragraphs can be used for drop-based simulations with limited mobility, i.e., for UEs that move in an area in which the channel is very correlated and the fading parameters do not change. However, for simulations in which mobility is an important factor, the spatial consistency of the channel throughout the path on which the UE moves can be simulated by enabling this option in the `MmWave3gppChannel` class. In the current implementation, we support spatial consistency with Procedure A of [31, Sec. 7.6.3.2] for both LOS and NLOS communications. It is possible to set the period of update t_{PER}, and every t_{PER} the cluster delays, powers and departure and arrival angles are updated with a transformation that accounts for the speed of the UE and for the distance traveled on the horizontal plane.

Blockage: This optional feature can be used to model the attenuation in certain clusters, according to their angle of arrival. The attenuation can be caused by the human body that holds the UE, or by external elements such as for example cars, other human bodies, trees. The blockage model is implemented in the `MmWave3gppChannel` class and can be optionally activated. In our implementation we consider blockage model A, which only distinguishes between self-blocking and non-self-blocking, and is generic and computationally efficient [31]. In particular, this model randomly generates

$K + 1$ blocking regions, one for self-blocking, with different parameters according to the orientation of the UE (i.e., portrait or landscape mode), and K for non-self-blocking. The attenuation is 30 dB for self-blocking, whereas it depends on the scenario and on the horizontal and vertical angles of arrival for non-self-blocking. Moreover, the blocking of a certain cluster is correlated in both space and time, according to the UE mobility, the blocker speed and the simulation scenario. Notice that, if both the blockage and the spatial consistency options are used, then the update of the channel with both features is synchronized, i.e., the cluster blockage is updated before the channel coefficients are recomputed with the spatial consistency procedure.

2) Ray-Tracing or Measurement Trace Model: `MmWave-ChannelRaytracing` uses software-generated or measurement traces to model the channel in ns–3, for pathloss and fading. The trace samples need to contain the number of paths and the propagation loss, delay, angle of arrival and angle of departure for each path. The following trace files have been tested in our implementation and are available in `mmwave/model/Raytracing/`.

Ray-tracing: Any ray-tracing software (e.g., WinProp [54]) can be used to generate the channel information for a specific route. This means that the simulation scenario must be chosen *a priori*, and cannot be random since it has to be given as input to the ray-tracing software. An example of ray-tracing route[7] is shown in Figure 4b.

QuaDRiGa: The Quasi Deterministic Radio Channel Generator model [56], supports consistent user mobility and massive MIMO at several frequencies (10, 28, 43, 60, 82 GHz). It also adds some time evolution characterization on top of the statistical channel to capture user mobility, which makes it suitable for system level simulations.

3) NYU Statistical Model: This channel model is based on the approach described in [32] and implemented in our previous work [27]. A MATLAB implementation of the same channel model is also available

---

[7]The ray tracing data was provided by the Communication Systems and Networks Group, University of Bristol, U.K. [10], [55].

(a) 3GPP statistical channel model (b) Ray-tracing Trace Model (c) NYU Statistical Model
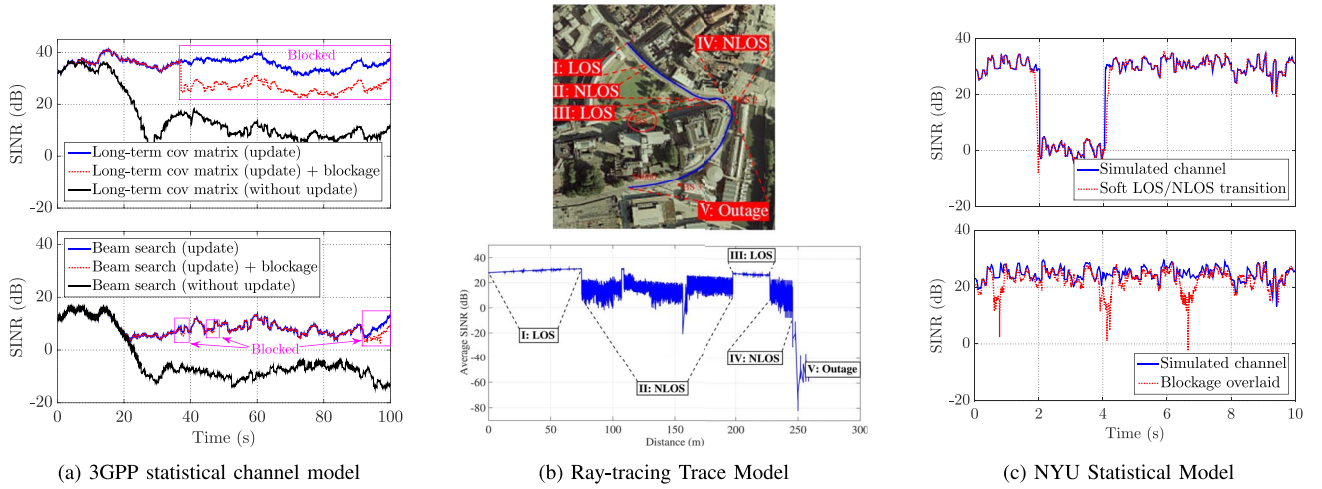
Fig. 4. Example of average SINR plots for the three channel models.

in [57]–[59]. It provides two pathloss models, which differ in how they capture the LOS/NLOS condition. The first, `MmWavePropagationLossModel`, is based on a statistical characterization of the LOS state, while the second, `BuildingsObstaclePropagationLossModel`, leverages the ns–3 buildings module in order to decide whether there is an obstacle between the UE and the eNB or not. In particular, it is possible to deploy – deterministically or randomly – objects of different sizes to mimic humans, cars, and buildings. A virtual line is drawn between the transmitter and the receiver: If this line intersects any object, the state is NLOS, otherwise it is LOS. In both classes, once the channel state is selected, the propagation loss is computed as in [32].

Channel configuration: Since the channel matrices and optimal beamforming vectors do not depend on the distance between the UE and the eNB, they are pre-generated in MATLAB to reduce the computational overhead in ns–3. At the beginning of each simulation we load 100 instances of the spatial signature matrices, along with the beamforming vectors. Moreover, in order to simulate realistic channels with large-scale fading, the channel matrices are updated periodically and independently (*block fading*). Currently, no results are available for modeling how the large-scale statistics of the mmWave channel change over time for a mobile user, thus it should be noted that the accuracy of this method is not verified at this time. The matrix update can take place at some fixed intervals, specified by the `LongTermUpdatePeriod` attribute of the `MmWaveBeamforming` class. The small-scale fading, instead, is calculated at every transmission, where we obtain the speed of the user directly from the mobility model. The remaining parameters that depend on the environment are assumed to be constant over the entire simulation time.

Semi-empirical feature: Finally, as shown in Fig. 4c, the soft transition between LOS/NLOS conditions can be modeled in a "semi-empirical" fashion, meaning that we overlay the statistical channel with blockage measurements performed in our lab [60]: Waving a hand in front of the receiver (hand blockage), walking between the transmitter and the receiver
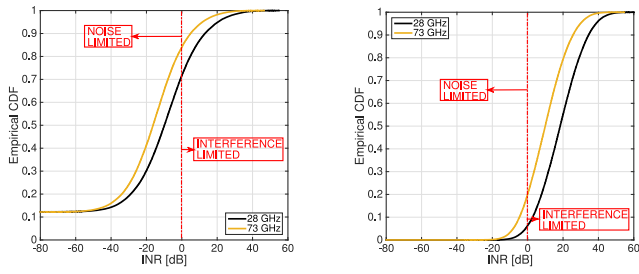
(human blockage), and placing a metal plate between the transmitter and the receiver to emulate an obstacle, like a car or a building.

B. Beamforming Gain

For the long-term statistical channel model, the beamforming vectors are directly loaded from MATLAB generated files. For the other channel models, two methods are implemented to compute beamforming vectors, i.e., the *long-term covariance matrix method* and the *beam search method*. Currently, the only available beamforming architecture for data transmission is *analog*, meaning that devices can transmit or receive in only one direction at a time. As part of our future work, we plan to integrate *hybrid* and *digital* transceiver designs.

In the long-term covariance matrix method, we assume that the transmitter estimates the spatial correlation matrix $\mathbf{Q}_{tx} = \mathbb{E}[\mathbf{H}^\dagger(t,f)\mathbf{H}(t,f)]$, where the expectation is taken over the frequencies and some interval of time. An analogous operation is done for the receiver. In practice, the TX and RX would estimate the spatial covariance matrix from reference or synchronization signals and beam scanning. Estimation of this covariance matrix is discussed in [61]. We do not, however, model the covariance estimation directly; instead we simply assume that the TX and RX know the correct long-term channel with some configurable delay. Beamforming vectors can then be computed from the maximal eigenvectors of the covariance matrices [62]. A computationally simple procedure is to use the power method [63]. The algorithm selects a random initial beamforming vector and iteratively multiplies it with the spatial correlation matrix $\mathbf{Q}_{tx}$, normalizing the results at each iteration. Finally, the output will converge to the correct eigenvector. The computation for the receiver is done in the same way, starting from $\mathbf{Q}_{rx} = \mathbb{E}[\mathbf{H}(t,f)\mathbf{H}^\dagger(t,f)]$.

In the beam search method, we assume that the TX and the RX scan a discrete number of beams from a pre-designed codebook [64]. Codebook design is discussed in detail in [65]. The beamforming vector is selected as the one with the highest power, possibly with some time-averaging.

(a) Empirical CDF of the INR for λ_{UE} = 300 UEs/km^2 and λ_{eNB} = 30 BSs/km^2.

(b) Empirical CDF of the INR for λ_{UE} = 1200 UEs/km^2 and λ_{eNB} = 120 BSs/km^2.

Fig. 5. INR trends at different user and base station density levels [67].

C. Interference

MmWave systems may be interference- or power-limited. Albeit potentially less significant for directional mmWave signals, which are generally assumed to be power-limited, there are still some cases where interference is non-negligible [66]. We report in Figure 5 some representative results obtained in [67], where, by plotting the Interference to Noise Ratio (INR), we show that the majority of the links are still interference-limited for some dense topologies. Additionally, although intra-cell interference (i.e., from devices of the same cell) can be neglected in TDMA or Frequency Division Multiple Access (FDMA) operation, it does need to be explicitly calculated in the case of Spatial Division Multiple Access (SDMA)/Multi-User MIMO, where users are multiplexed in the spatial dimension but operate in the same time-frequency resources. Therefore, we propose an interference computation scheme that takes into account the beamforming vectors associated with each link.

As an example, we compute the SINR between nodes eNB_1 and UE_1 in the presence of an interferer, eNB_2. To do so, we first need to obtain the beamforming gains associated with both the desired and interfering signals, i.e.,

$$G_{11} = \left| \mathbf{w}_{rx_{11}}^{\dagger} \mathbf{H}(t,f)_{11} \mathbf{w}_{tx_{11}} \right|^2,$$
$$G_{21} = \left| \mathbf{w}_{rx_{11}}^{\dagger} \mathbf{H}(t,f)_{21} \mathbf{w}_{tx_{22}} \right|^2, \qquad (1)$$

where $\mathbf{w}_{rx_{i,j}}$ is the beamforming vector of receiver i towards transmitter j, and $\mathbf{w}_{tx_{i,j}}$ is the beamforming vector of transmitter i towards receiver j. The SINR is then computed as:

$$SINR_{11} = \frac{\frac{P_{Tx,11}}{PL_{11}} G_{11}}{\frac{P_{Tx,22}}{PL_{21}} G_{21} + BW \times N_0}, \qquad (2)$$

where $P_{Tx,ii}$ is the transmit power of eNB_i, PL_{ij} is the pathloss between eNB_i and UE_j, and $BW \times N_0$ is the thermal noise.

D. Error Model

The mmWave module exploits the error model introduced in the ns-3 LTE LENA project, which follows a link abstraction technique for simulating Transport Block (TB) errors in the downlink of an LTE system. In a nutshell, the model described in [68] defines an accurate and lightweight procedure for the computation of the residual errors after PHY layer processing. This is achieved by exploiting:

- Mutual Information-based multi-carrier compression metrics to derive a unique SINR value of the channel, known as *effective SINR*, which is represented as γ_i in Eq. (3), and
- Link-Level performance curves obtained with a MATLAB bit-level LTE PHY simulator [69], which have been used to match a mathematical approximation of the Block Error Rate (BLER), as reported in Eq. (3).

The ultimate goal is to let the receiver derive the error probability of each TB to determine whether the packet can be decoded or not. Because each TB can be composed of multiple Code Blocks (CBs), whose size depends on the channel capacity, the BLER can be formulated as follows:

$$C_{BLER,i}(\gamma_i) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\gamma_i - b_{C_{SIZE},MCS}}{\sqrt{2}c_{C_{SIZE},MCS}} \right) \right], \qquad (3)$$

where γ_i corresponds to the mean mutual information per coded bit of code block i, as explained earlier, and $b_{C_{SIZE},MCS}$ and $c_{C_{SIZE},MCS}$ represent the mean and standard deviation of the Gaussian cumulative distribution, respectively, which have been obtained from the link level performance curves mentioned above. Finally, the TB block error rate is given by:

$$T_{BLER} = 1 - \prod_{i=1}^{C}(1 - C_{BLER,i}(\gamma_i)). \qquad (4)$$

E. Examples

An example of SINR plots for the three channel models was presented in Figure 4. An example related to the setup of the channel model can be found in the `examples` folder of the mmWave module, in the file `mmwave-3gpp-channel-example.cc`.

Figure 4a shows an example of a rural scenario with an eNB at coordinates $(0, 0)$ m and at the height of 35 m, and a UE in position $(100, 0)$ m, at the height of 1.5 m and moving at 1 m/s along the y axis, maintaining LOS connectivity. The channel is updated consistently every 100 ms. The top figure shows the SINR when the Beamforming (BF) vector is updated with the long-term covariance matrix method, while in the bottom one it is updated with the beam search method. Notice that the current implementation of the beam search method uses a fixed elevation angle of 90 degrees and sweeps only the horizontal plane. Therefore, the beam search method cannot align with the LOS cluster and the power is reduced by 20 dB. Moreover, after enabling the blockage model, the SINR achieved by the long-term covariance matrix method dropped by 20 dB when the LOS cluster was blocked. However, the beam search method experienced less blockage impact, as it did not align with the LOS cluster. In the other case, without update, the BF vector is computed at $t = 0$ s but never updated, and this causes the SINR to drop as the UE moves. Comparing the blue and black curves, it is possible to observe that for the first 20 s the performance with and without BF update is similar, because of the consistency of the channel and of the low mobility of the UE, but after $t = 20$ s the SINR without

update degrades by nearly 30 dB. The last observation is that the long-term covariance matrix method finds the optimal BF vector whenever the channel is changed, therefore the SINR is very stable. On the other hand, the beam search method shows an SINR drop after 20 s even with update, because when the UE moves both the UE and the eNB are unable to optimally adapt the BF vector and just select one of the available sectors.

Figure 4b plots the average SINR of a ray tracing channel indicating both LOS intervals and NLOS channel states. The ray tracing data contains 5000 samples along a 500 meter route. The SINR has a sudden change when the channel state switches. We note that the SINR curve within LOS is relatively stable, whereas more random variations are introduced for NLOS.

Finally, Figure 4c shows the average SINR trace generated with the NYU channel model [32] in two cases, a walking user blocked by a building (top) or by other humans (bottom). The main difference is that, with buildings, the link capacity drops rapidly and the blocking interval lasts seconds; on the other hand, with humans, the channel deteriorates slowly and the blockage lasts only for a short interval. From the top figure, we can observe that with soft LOS/NLOS transition enabled, the SINR curve changes less suddenly when the channel condition switches. In the bottom graph, three human blockage events, at 1, 4 and 7 seconds, are added on top of the statistical channel.

VI. PHYSICAL LAYER

In this section, we discuss the key features of the mmWave PHY layer. Specifically, we have implemented a TDD frame and subframe structure, which has similarities to TDD-LTE, but allows for more flexible allocation and placement of control and data channels within the subframe and is suitable for the *variable TTI* MAC scheme described in Section VII. Moreover, we implemented an error model and HARQ model based on those in LENA, but compatible with our custom mmWave PHY and numerology (for instance, they support larger TB and codeword sizes as well as multi-process stop-and-wait HARQ for both DL and UL).

A. Frame Structure

It is widely contended that 5G mmWave systems will target Time Division Duplex (TDD) operation because it offers improved utilization of wider bandwidths and the opportunity to take advantage of channel reciprocity for channel estimation [3], [14], [70]–[72]. In addition, shorter symbol periods and/or slot lengths have been proposed in order to reduce radio link latency [73]–[75]. The ns–3 mmWave module therefore implements a TDD frame structure which is designed to be configurable and supports short slots in the hope that it will be useful for evaluating different potential designs and numerologies. These parameters, shown in Table I, are accessible through the attributes of the common `MmwavePhyMacCommon` class, which stores all user-defined configuration parameters used by the PHY and MAC classes. Examples related to the setup of the PHY layer parameters can be found in the `mmwave-tdma.cc` and `mmwave-epc-tdma.cc` files.
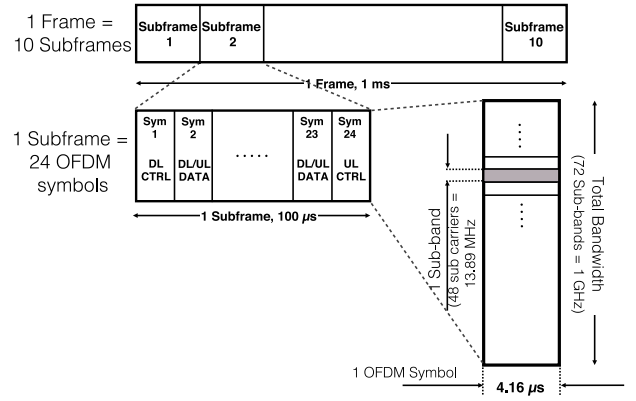

Fig. 6. Proposed mmWave frame structure.

The frame and subframe structures share some similarities with LTE in that each frame is subdivided into a number of subframes of fixed length [76]. However, in this case, the user is allowed to specify the subframe length in multiples of OFDM symbols.[8] Within each subframe, a variable number of symbols can be assigned by the MAC scheduler and designated for either control or data channel transmission. The MAC entity therefore has full control over multiplexing of physical channels within the subframe, as discussed in Section VII. Furthermore, each variable-length time-domain data slot can be allocated by the scheduler to different users for either the uplink or the downlink.

Figure 6 shows an example of frame structure with the numerology taken from our proposed design in [74]. Each frame of length 1 ms is split in time into 10 subframes, each of duration 100 μs, representing 24 symbols of approximately 4.16 μs. In this particular scheme, the downlink and uplink control channels are always fixed in the first and last symbol of the subframe, respectively. A switching guard period of one symbol period is introduced each time the direction changes from UL to DL. In the frequency domain, the entire bandwidth of 1 GHz is divided into 72 subbands of width 13.89 MHz, each composed of 48 subcarriers. It is possible to assign UE data to each of these subbands, as is done with Orthogonal Frequency-Division Multiple Access (OFDMA) in LTE, however only TDMA operation is currently supported for reasons we shall explain shortly.

B. PHY Transmission and Reception

The `MmWaveEnbPhy` and `MmWaveUePhy` classes model the physical layer for the mmWave eNB and the UE, respectively, and encapsulate similar functionalities as the `LtePhy` classes from the LTE module. Broadly, these objects (i) handle the transmission and reception of physical control and data channels (analogous to the Physical Downlink Control Channel (PDCCH)/Physical Uplink

---

[8]Though many waveforms are being considered for 5G systems, OFDM is still viewed as a possible candidate. In [71] and [77], Verizon and the consortium led by Korea Telecom propose a frame structure and OFDM numerology. However, this is still under debate in 3GPP [78]. We naturally chose to adopt OFDM, at least initially, for the mmWave module, which allows us to leverage the existing PHY models derived for OFDM from the LTE LENA module. As soon as the 3GPP NR will be standardized, the protocol stack in our module can be easily adapted to the updated parameters.

Parameter Name	Default Value	Description
SubframePerFrame	10	Number of subframes in one frame
SubframeLength	100	Length of one subframe in μs
SymbolsPerSubframe	24	Number of OFDM symbols per slot
SymbolLength	4.16	Length of one OFDM symbol in μs
NumSubbands	72	Number of subbands
SubbandWidth	13.89	Width of one subband in MHz
SubcarriersPerSubband	48	Number of subcarriers in each subband
CenterFreq	[6-100]	Possible carrier frequencies in GHz*
NumRefScPerSymbol	864 (25% total)	Reference subcarriers per symbol
NumDlCtrlSymbols	1	Downlink control symbols per subframe
NumUlCtrlSymbols	1	Uplink control symbols per subframe
GuardPeriod	4.16	Guard period for UL-to-DL mode switching in μs
MacPhyDataLatency	2	Subframes between MAC scheduling request and scheduled subframe
PhyMacDataLatency	2	Subframes between TB reception at PHY and delivery to MAC
NumHarqProcesses	20	Number of HARQ processes for both DL and UL

*The NYU channel model [32] supports only 28 and 73 GHz.

Control Channel (PUCCH) and Physical Downlink Shared Channel (PDSCH)/Physical Uplink Shared Channel (PUSCH) channels of LTE), (ii) simulate the start and the end of frames, subframes and slots, and (iii) deliver received and successfully decoded data and control packets to the MAC layer.

In the `MmWaveEnbPhy` and `MmWaveUePhy` classes, calls to `StartSubFrame()` and `EndSubFrame()` are scheduled at fixed periods, based on the user-specified subframe length, to mark the start and end of each subframe. The timing of variable-TTI slots, controlled by scheduling the `StartSlot()` and `EndSlot()` methods, is dynamically configured by the MAC via the MAC-PHY SAP method `SetSfAllocInfo()`, which enqueues an `SfAllocInfo` allocation element for some future subframe index specified by the MAC. A *subframe indication* to the MAC layer triggers the scheduler at the beginning of each subframe to allocate a future subframe. For the UE PHY, `SfAllocInfo` objects are populated after reception of DCI messages. At the beginning of each subframe, the current subframe allocation scheme is dequeued, which contains a variable number of `SlotAllocInfo` objects. These, in turn, specify contiguous ranges of OFDM symbol indices occupied by a given slot, along with the designation as either *DL* or *UL* and control (*CTRL*) or data (*DATA*).

The data packets and the control messages generated by the MAC are mapped to a specific subframe and slot index in the *packet burst map* and *control message map*, respectively. Presently, in our custom subframe design, certain control messages which must be decoded by all UEs, such as the DCIs, are always transmitted in fixed PDCCH/PUCCH symbols at the first and last symbol of the subframe, but this static mapping can be easily changed by the user.[9] Other UE-specific control and data packets are recalled at the beginning of each

allocated TDMA data slot and are transmitted to the intended device.

To initiate transmission of a data slot, the eNB PHY first calls `AntennaArrayModel::ChangeBeamforming-Vector()` to update the transmit and receive beamforming vectors for both the eNB and the UE. In the case of control slots, no beamforming update is applied since we currently assume an "ideal" control channel. For both DL and UL transmissions, either the `MmWaveSpectrumPhy` method `StartTxDataFrame()` or `StartTxCtrlFrame()` is called to transmit a data or control slot, respectively. The functions of `MmWaveSpectrumPhy`, which is similar to the corresponding LENA class, are as follows. After the reception of data packets, the PHY layer calculates the SINR of the received signal in each subband, taking into account the path loss, MIMO beamforming gains and frequency-selective fading. This triggers the generation of Channel Quality Information (CQI) reports, which are fed back to the base station in either UL data or control slots. The error model instance is also called to probabilistically compute whether a packet should be dropped by the receiver based on the SINR and, in the case of an HARQ retransmission, any soft bits that have been accumulated in the PHY HARQ entity (see Section VII-B). Uncorrupted packets are then received by the `MmWavePhy` instance, which forwards them up to the MAC layer SAP.

VII. MAC LAYER

TDMA is widely assumed to be the de-facto scheme for mmWave access because of the dependence on analog beamforming, where the transmitter and receiver align their antenna arrays to maximize the gain in a specific direction (rather than with a wide angular spread or omni-directionally, as in conventional systems). Many early designs and prototypes have been TDMA-based [14], [70], [72], with others incorporating SDMA for the control channel only [73]. SDMA or FDMA schemes (as in LTE) are possible with *digital beamforming*, which would allow the base station to transmit or receive in multiple directions at the same time.

[9]As in [73] and [74], we assume either FDMA or SDMA-based multiple access in the control regions. However, we do not currently model these modulation schemes nor the specific control channel resource mapping explicitly. We intend for this capability to be available in later versions, which will enable more accurate simulation of the control overhead.

Furthermore, one of the foremost considerations driving innovation for the 5G MAC layer is latency. Specifically, the Key Performance Indicator of 1 ms over-the-air latency has been proposed as one of the core 5G requirements by such standards bodies as the International Telecommunication Union [79], as well as by recent pre-standardization studies such as those carried out under the METIS 2020 project [80]. However, a well-known drawback of TDMA is that fixed slot lengths or TTIs can result in poor resource utilization and latency, which can become particularly severe in scenarios where many intermittent, small packets must be transmitted to/received from many devices [74].

Based on these considerations, variable TTI-based TDMA frame structures and MAC schemes have been proposed in [21], [73]–[75], and [81]. This approach allows for slot sizes that can vary according to the length of the packet or TB to be transmitted and are well-suited for diverse traffic since they allow bursty or intermittent traffic with small packets as well as high-throughput data like streaming and file transfers to be scheduled efficiently.

The MAC layer implementation can be found in the `MmWaveEnbMac` and `MmWaveUeMac` classes, whose main role is the coordination of procedures such as scheduling and retransmission. Moreover, they interact with the RLC layer to receive periodic reports on the buffer occupancy, i.e., the Buffer Status Reports (BSRs), and with the physical layer classes for the transmission and reception of packets. To carry out their functionalities, the MAC classes interact with several other classes, that we will describe in the following paragraphs.

A. Adaptive Modulation and Coding

The role of the Adaptive Modulation and Coding (AMC) mechanism is to adapt the modulation scheme and the coding applied on top to the channel quality, measured using CQIs. In the simulator, this translates into (i) mapping the CQI into the Modulation and Coding Scheme (MCS), using the error model implemented in the `MmWaveMiErrorModel` and described in Section V-D, and (ii) computing the available TB size for a subframe given the MCS. This information is then used by the scheduler to perform radio resource management.

The AMC is implemented in the `MmWaveAmc` class, which uses most of the code of the corresponding LENA module class. Some minor modifications and additional methods were necessary to accommodate the dynamic TDMA MAC scheme and frame structure. For instance, the `GetTbSizeFromMcsSymbols()` and `GetNumSymbolsFromTbsMcs()` methods are used by the scheduler to compute the TB size from the number of symbols for a given MCS value, and vice versa. Also the `CreateCqiFeedbackWbTdma()` method is added to generate wideband CQI reports for variable-TTI slots.

Figure 7 shows the results of the test case provided in `mmwave-amc-test.cc`. This simulation serves to demonstrate the performance of the AMC and CQI feedback mechanisms for a single user in the uplink (although a
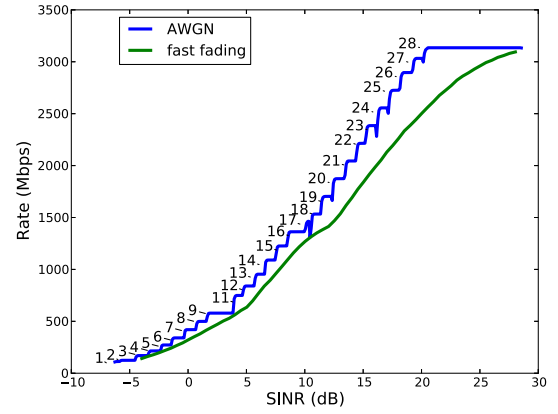

Fig. 7.   Rate and MCS vs. SINR for a single user under AGWN and fast-fading mmWave channels. ©[2016] ACM. Reprinted, with permission, from [28].

multi-user scenario could easily be configured as well). The default PHY/MAC parameters in Table I are used along with the default scheduler and default parameters for the statistical path loss, fading and beamforming models (i.e., `MmWavePropagationLossModel` and `MmWaveBeamforming`).

We compute the rate versus the average SINR over a period of 12 seconds (long enough for the small-scale fading to average out). The average PHY-layer rate is computed as the average sum of the sizes of successfully-decoded TBs per second. Every 12 seconds we artificially increase the path loss while keeping the UE position fixed. As the SINR decreases, the MAC will select a lower MCS level to encode the data. The test is performed for the Additive White Gaussian Noise (AGWN) case (i.e., no fading) as well as for small-scale fading. Although the UE position relative to the base station is constant, we can generate time-varying multi-path fading through the `MmWaveBeamforming` class by setting a fixed speed of 1.5 m/s to artificially generate Doppler, which is a standard technique for such an analysis. Also, we assume that the long-term channel parameters do not change for the duration of the simulation.

Figure 7 therefore shows the data rate that it is possible to achieve with a certain SINR and with a certain modulation and coding scheme. If this plot is compared to the one generated from a similar test in Figure 3.1 of the LENA documentation [52], we notice that the AGWN curve from the mmWave test is shifted by approximately 5 dB to the left, indicating that the LENA version is transitioning to a lower MCS at a much higher SINR. This is because the LENA test is using the more conservative average SINR-based CQI mapping, which targets a much smaller TB error probability. In our test, we use the Mutual Information-Based Effective SINR scheme described in Section V-D with a target maximum TB error of 10% in order to maximize the rate for a given SINR [68].

B. Hybrid ARQ Retransmission

Full support for HARQ with soft combining is now included in the mmWave module. HARQ is a technique introduced in [82] and extensively used in LTE networks [21], which

enables fast retransmissions with incremental redundancy in order to increase the probability of successful decoding and the efficiency of the transmissions. In LTE, the HARQ mechanism is based on multiple stop and wait retransmission processes, and a maximum of 8 simultaneous HARQ processes can be active at any given time [83]. The HARQ retransmissions have priority with respect to new transmissions, thus the available resources are given first to HARQ processes and then to the data queued in the RLC buffers. Despite being fundamental in protecting from the losses of packets due to rapid variations in the channel quality, the HARQ mechanism introduces additional latency [19], [21], therefore the optimization of its performance is necessary to enable the target of sub-1-ms latency for ultra-low-latency communications.

The `MmWaveHarqPhy` class along with the functionalities within the different scheduler classes are based heavily on the LENA module code. The scheduler at the eNB uses the information provided by HARQ feedback messages to assign new resources to the HARQ processes that require retransmissions. Each transport block is granted a maximum number of transmission attempts, which is set to 3, as in LTE. However, some novelties are introduced in `MmWaveHarqPhy` in order to account for the more challenging channel conditions of the mmWave scenario. First, multiple HARQ processes per user can be created not only for the downlink but also for the uplink. Second, the number of processes per user is not fixed to 8, but can be configured through the `NumHarqProcesses` attribute in `MmWavePhyMacCommon`. This makes it possible to control (and, if needed, increase) the number of the simultaneous stop and wait retransmission processes and optimize the bandwidth utilization. Third, additional modifications were needed to support larger codeword sizes in both the `MmWaveHarqPhy` and `MmWaveMiErrorModel` classes. Finally, the integration with the flexible TTI physical layer allows a reduction in the latency of the retransmissions, as discussed in [21].

C. Schedulers

We now present the implementations of four scheduler classes for the variable TTI scheme. These differ significantly from the OFDMA-based schedulers available in ns–3 LENA [52] as, instead of allocating Resource Blocks/Resource Block Groups of frequency-domain resources, these TDMA-based schedulers allocate time-domain symbols within a periodic subframe to different users in the DL or UL direction.

Before scheduling new data, Buffer Status Report and CQI messages are first processed. The MCS is computed by the AMC model for each user based on the CQIs for the DL or SINR measurements for the UL data channel. The MCS and the buffer length of each user are used to compute the minimum number of symbols required to schedule the data in the user's RLC buffers. This procedure for estimating the optimal MCS and determining the minimum number of symbols is common to each of the schedulers described in the following.

Round Robin (RR) Scheduler: The `MmWaveFlexTtiMacScheduler` class is the default scheduler for the mmWave module. It supports the variable TTI scheme previously described in Section VI and assigns OFDM symbols to user flows in *Round-Robin* order. Upon being triggered by a subframe indication, any HARQ retransmissions are automatically scheduled using the available OFDM symbols. While the slot allocated for a retransmission does not need to start at the same symbol index as the previous transmission of the same TB, it does need the same number of contiguous symbols and MCS, since an adaptive HARQ scheme (where the retransmission can be scheduled with a different MCS) has not yet been implemented.

To assign symbols to users, the total number of users with active flows is first calculated. Then the total available data symbols in the subframe are divided evenly among users. If a user requires fewer symbols to transmit its entire buffer, the remaining symbols (i.e., the difference between the available and required slot length) are distributed among the other active users.

One also has the option to set a fixed number of symbols per slot by enabling the *fixed TTI* mode. Although the same general subframe structure is maintained, slots will then be allocated in some multiple of `SymPerSlot` symbols. Setting the `SymPerSlot` attribute of the scheduler class to the number of slots per subframe, for instance, will result in only one UE being scheduled per subframe, which would be highly inefficient in a multi-user cell.

Proportional Fair (PF) Scheduler: Proportional Fair is another well-known discipline, and is provided by the `MmWaveFlexTtiPfMacScheduler` class. The PF scheduler attempts to prioritize traffic for high-SINR users while maintaining some measure of fairness by ensuring that low-SINR, cell-edge users are also scheduled [84].

Earliest Deadline First (EDF) Scheduler: The `MmWave-FlexTtiEdfMacScheduler` class implements an *Earliest Deadline First* policy, which is a priority queue-based policy that weighs flows by their relative deadlines for packet delivery. The deadlines are initially set according to the delay budget of the QoS Class Indicator (QCI) configured by the RRC layer [85], [86]. The deadline of the Head-of-Line (HOL) packets of each RLC buffer is then compared, and that with the earliest deadline is scheduled first. Any remaining symbols in the subframe are allocated to the packet with the next smallest relative deadline and so forth until all N_{sym} symbols are assigned. The EDF scheduler is the only deliberately delay-sensitive scheme included in the mmWave module and can be useful for evaluating the latency performance of mmWave links, as in the simulations presented in Section X.

Maximum Rate (MR) Scheduler: The *Maximum Rate* policy realized in the `MmWaveFlexTtiMrMacScheduler` class schedules only the users with the highest SINRs to maximize cell throughput. Initially, UEs are sorted based on their optimal MCS values. Symbols are distributed in round-robin fashion among UEs at the highest MCS level until the minimum number of symbols required to transmit the entire buffers of these users has been assigned. This is then repeated for UEs at the second highest level, and so forth until all symbols of the subframe are allocated.

The MR scheduler potentially suffers from extremely poor fairness when there are both high- and low-rate users, and

some users may not be scheduled at all, thus making it impractical for any real-world multi-user system. However, it may still be useful for testing system capacity and performance.

VIII. RLC LAYER

The RLC layer is inherited directly from the LTE module described in [24], and therefore all the LTE RLC entities are included. Moreover, the RLC AM retransmission entity is modified to be compatible with the mmWave PHY and MAC layers, and Active Queue Management (AQM) for the RLC buffers is introduced as a new optional feature.

A. Modified RLC AM Retransmission

Reordering and retransmission play an important role in RLC AM. Due to the shortened mmWave frame structure, the timers of the RLC entity should also be reduced accordingly, e.g., the `PollRetransmitTimer` is changed to 2 ms from 20 ms. Moreover, the original LTE module does not perform re-segmentation for retransmissions, and the RLC segment waits in the retransmission buffer until the transmission opportunity advertised by the lower layers is big enough. This becomes problematic when the transmission is operated over an intermittent channel, as a sudden channel capacity drop would halt the retransmission entirely. Therefore, we added to the RLC AM layer implementation the capability of performing segmentation also for the retransmission process, in order to support an intermittent mmWave channel. The re-segmentation process deployed in our RLC AM class works as follows: If the number of bytes that can be transmitted in the next opportunity is smaller than the bytes of the segment that should be retransmitted, then the segment will be split into smaller subsegments with a re-segment flag set to be true. The RLC layer at the receiver side will check the flags of the subsegments, and wait until the final one if the flag is set to be true. Finally, the subsegments are assembled to construct the original segment and forwarded to the upper Packet Data Convergence Protocol (PDCP) layer if all subsegments are received correctly. Otherwise, all subsegments are discarded and another retransmission is triggered.

B. Active Queue Management

Active Queue Management techniques are used in the buffers of routers, middleboxes and base stations in order to improve the performance of TCP and avoid the manual tuning of the buffer size. Different strategies have been defined in [87] and [88], and several of them are implemented in ns–3 [89]–[91]. AQM strategies allow the network to avoid congestion at the buffers, because they react early to the increase in the buffer occupancy by dropping some packets before the buffer is full. With respect to the default Drop-tail approach, in which no packet is dropped until the buffer is full, AQM techniques make TCP aware of possible congestion earlier, avoiding the latency increase which is typical of the bufferbloat phenomenon [92].

Some early AQM, such as Random Early Detection (RED) [93], [94], were widely studied in the literature, but failed to find market traction because of the intrinsic complexity of their tuning parameters. Recently, a simpler AQM technique, namely Controlled Delay (CoDel) [95], was proposed to replace RED queues. CoDel adapts to dynamic link rates without parameter configuration, and is able to discriminate "good" and "bad" queues: good queues can quickly empty the buffer, whereas "bad" queues persistently buffer packets. CoDel works by monitoring the minimum queue delay in every 100 ms interval, and only drops packets when the minimum queue delay is more than 5 ms.

In the RLC layer of the LTE module, the default queue management is Drop-tail. In the mmWave module, the RLC layer can use either the default Drop-tail approach or more sophisticated AQM techniques, that can be enabled by setting the `EnableAQM` attribute to true. The default AQM is the CoDel scheme, however it is possible to use any of the queues available in ns–3 by modifying the queue attribute in the `LteRlcAm` class. The evaluation of the AQM scheme is further discussed in Section X-D.

IX. DUAL CONNECTIVITY EXTENSION

The ns–3 mmWave module is also capable of performing simulations with dual-stack UEs connected both to an LTE eNB and to a mmWave eNB. This feature, partially described in [30], was introduced because mmWave 5G networks will likely use multi-connectivity and inter-networking with legacy Radio Access Technologies (RATs) in order to increase the robustness with respect to mobility and channel dynamics [15]–[17], [96]–[98]. The source code can be found in the `new-handover` branch of the ns–3 mmWave module repository.

The Dual Connectivity (DC) implementation of this simulation module assumes that the core networks of LTE and of mmWave will be integrated, as in one of the options described in [99]. Therefore the LTE and the mmWave eNBs share the same backhaul network, i.e., they are connected to each other with X2 links and to the Mobility Management Entity (MME) nodes with the S1 interface. As to the Radio Access Network (RAN), the DC solution of this module is an extension of 3GPP's LTE DC proposal [100]. In particular, a single bearer per DC flow is established, with a connection from the core network to the LTE eNB, where the flow is split and forwarded either to the local stack or to the remote mmWave stack. We chose the PDCP layer as the integration layer, since it allows a non-colocated deployment of the eNBs and a clean-slate approach in the design of the PHY, MAC and RLC layers [101].

A basic diagram for a DC UE device, an LTE eNB and a mmWave eNB is shown in Figure 8. The core of the DC implementation is the `McUeNetDevice` class, which is a subclass of the ns–3 `NetDevice` and provides an interface between the ns–3 TCP/IP stack and the custom lower layers. The `McUeNetDevice` holds pointers to the custom lower layer stack classes, and has a `Send` method that forwards packets to the TCP/IP stack. This method is linked to a callback on the `DoRecvData` of the `EpcUeNas` class, which as specified by the 3GPP standard acts as a connection between the LTE-like protocol stack and the TCP/IP stack.
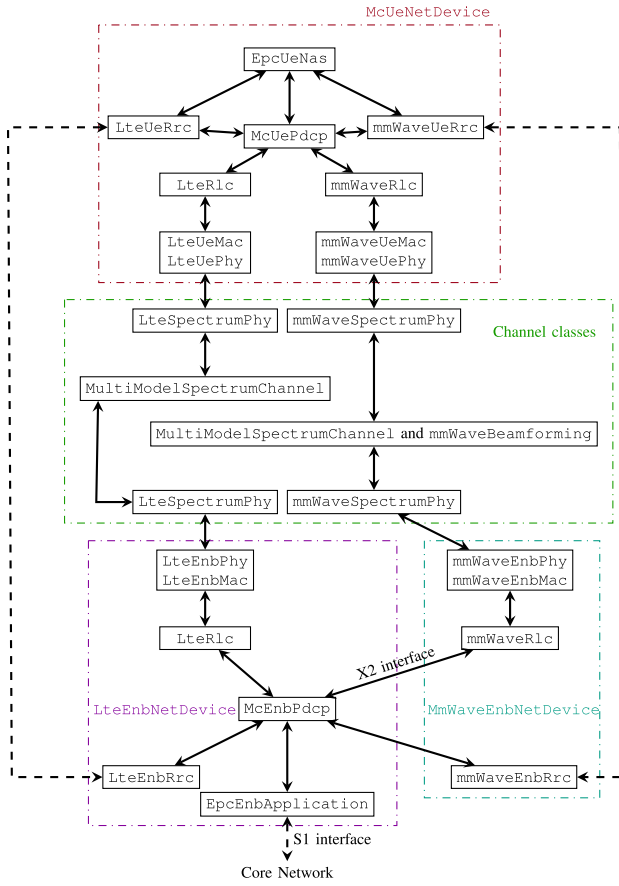
Fig. 8. Block diagram of a dual-connected device, an LTE eNB and a mmWave eNB [30].

The `McUeNetDevice` describes a dual connected UE with a single `EpcUeNas`, but with a dual stack for the lower layers, i.e., there are separate LTE and mmWave PHY and MAC layers. Moreover, there is an instance of the RRC layer for both links. This grants a larger flexibility, because the functionalities and the implementation of the two layers may differ. Besides, the LTE RRC manages both the LTE connection and the control plane features related to DC, while the mmWave RRC handles only the mmWave link. The usage of a secondary RRC, dedicated to the mmWave link, avoids latency in control commands (i.e., the mmWave eNB does not have to encode and transmit the control Packet Data Unit (PDU)s to the master LTE eNB). The `EpcUeNas` layer has an interface to both RRC entities to exchange information between them.

The LTE RRC manages also the data plane for the DC devices. In particular, for each bearer, a dual connected PDCP layer is initialized and stored in the LTE RRC. The classes describing the DC PDCP layer are `McEnbPdcp` and `McUePdcp`, respectively at the eNB side and at the UE side. They both extend the `LtePdcp` class with a second interface to the RLC. However, while `McUePdcp` simply has to communicate with a local RLC in the UE, the implementation of `McEnbPdcp` requires new interfaces to the class describing the X2 links between eNBs (i.e., `EpcX2`). In particular, in downlink the eNB PDCP has to send packets to the X2 link and the mmWave RLC layer has to receive them, and vice versa in uplink.

The DC module can be used to simulate different dual connected modes, i.e., it can support both Fast Switching (FS) and throughput-oriented dual connectivity, according to which RRC and X2 procedures and primitives are implemented. With FS, the UE is in the RRC_CONNECTED state with respect to both eNBs, but only transmits data to one of the two. With the other option, the UE can transmit data simultaneously on both RATs, and different flow control algorithms can be plugged in and tested.

As to the physical layer, the two stacks rely on the mmWave and LTE channel models. Notice that since the two systems operate at different frequencies, modeling the interference between the two RATs is not needed. Each of the two channel models can therefore be configured independently.

In order to use an `McUeNetDevice` as a mobile User Equipment in the simulation, the helper class of the mmWave module was extended with several features, such as (i) the initialization of the objects related to the LTE channel; (ii) the installation and configuration of the LTE eNBs, so that they can be connected to the LTE stack of the `McUeNetDevice`; and (iii) the methods to set up a `McUeNetDevice` and link its layers as shown in Fig 8. An example on how to set up a dual-connectivity based simulation is provided in the file `mc-twoenbs.cc`.

RRC Layer for Dual Connectivity and Mobility: The RRC layer implementation of the original LTE ns–3 module was extended in order to account for DC-related control procedures. In particular, the multi-connectivity uplink-based measurement framework described in [16] was added with changes to the `MmWaveEnbPhy`, `EpcX2` and `LteEnbRrc` classes. The `MmWaveEnbPhy` instance simulates the reception of uplink reference signals (which are accounted for as overhead in the data bearers resource allocation), computes the SINR for each UE in the scenario,[10] and sends this information to the LTE eNB on the X2 link. This also allows the simulation of a delay in the reporting, since the control packets with the SINR values must be transmitted on an ns–3 `PointToPointLink`, which adds a certain latency and has a certain bitrate.

Thanks to this framework, the LTE eNB is able to act as a coordinator for the surrounding mmWave eNBs, and learns which is the best association (in terms of SINR) between UEs and mmWave secondary eNBs. This enables automatic cell selection for mmWave eNBs at the beginning of a simulation, and the control of mobility-related operations. The DC module is indeed capable of simulating FS procedures between mmWave and LTE links and Secondary Cell Handover (SCH) (i.e., handovers between mmWave eNBs that do not involve the MME in the core network) initiated by the central controller in the LTE eNB. It is also possible to use the DC module to simulate X2-based RAT handovers between the LTE and mmWave eNBs, i.e., to use standalone UEs based on `McUeNetDevice` that can perform handovers from the LTE to the mmWave eNBs, and vice versa.

---

[10]The framework assumes that the optimal beam is always chosen, so the actual directional scan procedure described in [16] is not simulated.

TABLE II
DL PHY THROUGHPUT FOR RR, PF, MR AND EDF SCHEDULING POLICIES

		PHY-layer throughput [Mbps]			
	Policy	Cell	Mean UE	Mean 5% Worst UE	Max UE
70 UE – 10 Mbps	RR	1815.92	25.94	1.11	48.80
	PF	1494.61	21.35	3.26	43.94
	MR	2273.18	32.47	0.00	151.36
	EDF	925.80	13.23	7.31	31.02
7 UE – 100 Mbps	RR	715.26	102.18	49.18	134.28
	PF	758.32	108.33	52.32	158.16
	MR	766.26	109.47	47.26	158.22
	EDF	647.98	92.57	63.89	121.76

Different handover (either inter-RAT or SCH) algorithms can be tested, by implementing them in the `LteEnbRrc` class. In order to make the handover simulation more compliant with the 3GPP specifications, the lossless handover option implemented for ns–3 in [102] was adapted to the DC module in order to forward the RLC buffer content to the target RAT/eNB RLC layer for both the SCH and the FS. Moreover, in order to model the additional latency given by the interaction with the MME for inter-RAT handovers for standalone UEs, the link between the eNBs and the MME is modeled in this module as a `PointToPointLink`, while in the original ns–3 LTE module it is an ideal connection.

X. USE CASES

In this section, we illustrate various examples of scenarios[11] that can be simulated to show the utility of the module for the analysis of novel mmWave protocols and for testing higher-layer network protocols, such as TCP, over 5G mmWave networks. After each particular use case example, we also provide to the interested reader some references to recent papers that report additional results obtained using the ns–3 mmWave simulation module.

A. Simulation Setup Walk-Through

In order to proficiently use the mmWave ns–3 module, a basic knowledge of ns–3 is required. We therefore advise the interested users to first study the extensive documentation referenced in Section III. Moreover, we provide some basic ns–3 scripts in the `examples` folder of the mmWave module, that can be a basis for the design of any simulation script that uses the mmWave module. In the following paragraphs, we will describe the basic structure of a typical example in simple steps.

The first step is to configure all the attributes needed in a simulation. A complete list of attributes related to the mmWave module can be found in the `mmWaveAttributesList` file in the module repository. The second step involves the setup of the `MmWaveHelper` object, which provides methods to create the entities involved in the simulation (e.g., the channel-related objects and the `MmWavePhyMacCommon` object), to install

the mmWave stack over ns–3 nodes (for both UEs and eNBs), to perform the initial attachment of a UE to the closest eNB and to enable or disable the generation of simulation traces. Moreover, if the scenario of interest is an end-to-end scenario, the core network and the Internet must be set up as well. The first is created by the `MmWavePointToPointEpcHelper`, which also provides a pointer to the Packet Gateway (PGW) node. This is then usually connected to a remote host, and the Internet stack (i.e., the TCP/IP protocol suite) is added to the UEs and to the remote host.

In the third step, the positions and velocities of the eNBs and UEs are specified using one or more `MobilityHelper` objects and different mobility models. Moreover, buildings and obstacles can be added to the scenario using the ns–3 buildings module and the `Buildings` and `BuildingsHelper` objects. The fourth step requires the setup of applications in the UEs and in the remote host (if an end-to-end scenario is considered), in order to simulate downlink and uplink traffic. ns–3 provides a wide range of different applications, and helpers that take care of their setup. They can run on either UDP or TCP sockets, and several TCP congestion control versions are available. Finally, the simulation can be run using the `Simulator` object of ns–3, and traces are generated.

B. Multi-User Scheduling Simulation

In this experiment, the throughput and latency of users of a mmWave cell with 1 GHz of bandwidth are simulated for variable TTI and each of the scheduling policies described in Section VII.[12] We shall see how the choice of the scheduler has a significant impact on the subframe utilization and latency of the multi-user cell. In these scenarios, UEs have similar distances from the eNB but are assigned the constant speed of 25 m/s (typical of vehicular users), which results in a lower achievable rate, on average, as well as increased packet errors compared to walking users due to the more rapid variation in the channel.

The simulation is again run over 10 drops for each of two scenarios and using default parameters from Table I. In the first scenario, 70 UEs are simulated with each UE generating IP-layer traffic at an average arrival rate of 10 Mbps. In the

---

[11]The simulations in this section are all configured with the basic PHY and MAC parameters in Table I, with other notable parameters given in the sequel.

[12]The multi-user scheduling experiment can be reproduced by running the `mmwave-epc-tdma.cc` example simulation [26].

(a) Empirical CDF of rate

(b) Empirical CDF of latency

Fig. 9.   Distributions of PHY-layer throughput and IP-layer latency for 70 UEs, 10 Mbps/UE arrival rate.

(a) Empirical CDF of rate
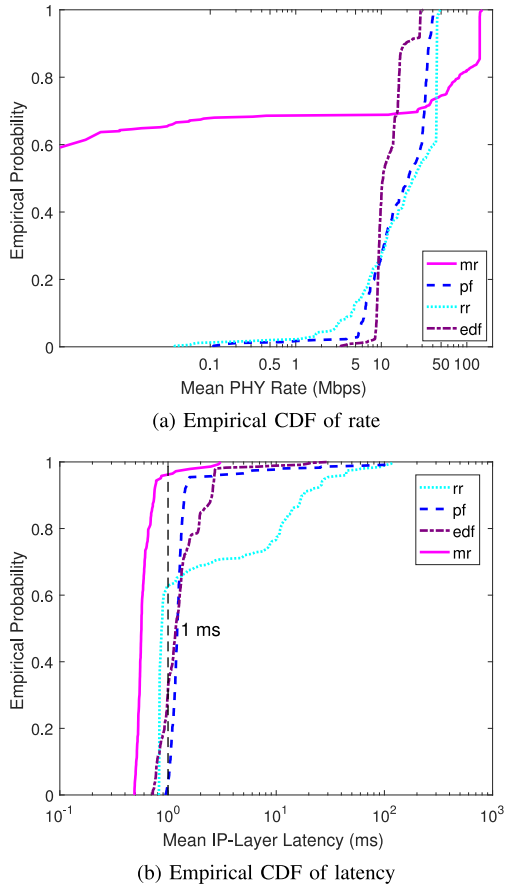

(b) Empirical CDF of latency

Fig. 10.   Distributions of PHY-layer throughput and IP-layer latency for 7 UEs, 100 Mbps/UE arrival rate.

TABLE III
IP-LAYER LATENCY FOR RR, PF, MR AND EDF SCHEDULING POLICIES

		IP-layer latency [ms]		
	Policy	Mean UE	Mean 5% Worst UE	Max UE
70 UE – 10 Mbps	RR	7.47	69.35	118.62
	PF	2.83	34.48	106.54
	MR	0.65	1.89	3.07
	EDF	1.63	7.91	30.65
7 UE – 100 Mbps	RR	0.67	2.01	2.37
	PF	0.55	0.68	0.77
	MR	0.56	0.78	1.09
	EDF	0.69	1.41	1.44

TABLE IV
FAIRNESS INDEX AND UTILIZATION (RECEIVED IP-LAYER RATE/ALLOCATED PHY RATE) FOR RR, PF, MR AND EDF SCHEDULING POLICIES

	Policy	Fairness	Utilization
70 UE – 10 Mbps	RR	0.71	0.53
	PF	0.76	0.73
	MR	0.28	0.39
	EDF	0.96	0.87
7 UE – 100 Mbps	RR	0.95	0.74
	PF	0.90	0.77
	MR	0.91	0.76
	EDF	0.99	0.84

second scenario, 7 UEs are simulated with a 100 Mbps arrival rate per UE.

These specific combinations of users and rates are deliberately chosen because they illustrate the cut-off point at which the system is no longer able to service most users at the requested rate, leading to backlogged queues and increased latency. That is, we wish to analyze the performance at the knee in the curve of the delay taken as a function of the system utilization. In the variable TTI system, this bottleneck effect has the following potential causes: (i) the number of users that must be serviced exceeds the number of available slots (ultimately limited by the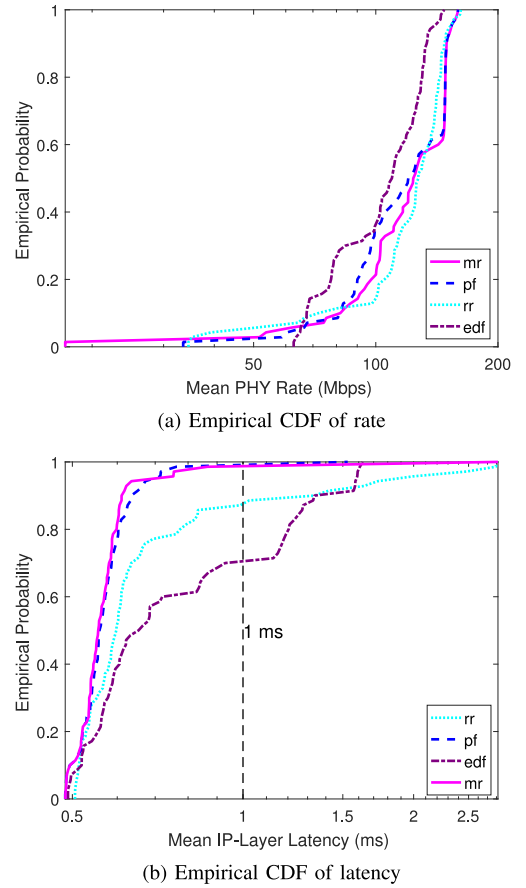 number of time-domain symbols), independently of the total throughput requested by the users; (ii) the number of users that are connected to eNB is smaller than the number of available slots, but the total throughput they request exceeds the available resources in the given time period; or (iii) a combination of the previous cases.

These effects are demonstrated in Figures 9 and 10 for the 70 UE/10 Mbps and 7 UE/100 Mbps arrival rate scenarios, respectively. The mean, maximum and cell-edge (i.e., 5% worst-case) user PHY rates and IP-to-IP layer latencies are also provided in Tables II and III along with the utilization and Jain's Fairness Index in Table IV.

For the 70 UE case, Figure 9a shows the distribution of the mean rate experienced by each UE over the simulation

TABLE V
ADDITIONAL PARAMETERS FOR VARIABLE AND FIXED TTI LATENCY EXPERIMENT

Description	Value
Subframe length in μs	100 or 66.67
OFDM symbols per slot	24 or 16
HARQ processes (DL and UL)	20 DL/20 UL
Number of UEs	Case 1: $\{10, 20, 30, 40, 50, 60, 70, 80, 100\}$
Number of UEs	Case 2: $\{1, 2, 3, 4, 5, 6, 7, 8, 10\}$
Traffic model	Case 1: Poisson, $\lambda = 12.5K$ pck/s, 100 B packets
Traffic model	Case 2: Poisson, $\lambda = 83K$ pck/s, 1200 B packets

duration. It can be seen that the MR and RR policies exhibit the greatest disparity between users scheduled with high and low rates.

It can also be observed that the PHY rate significantly exceeds the 10 Mbps arrival rate for some users, which leads to the poor utilization for these two policies, as shown in Table IV. The reason why the utilization (defined as the ratio of the received IP-layer rate to the allocated PHY-layer rate for each terminal) suffers in these cases is that the UEs with higher achievable rates are heavily favored by the MR and RR schedulers. As these users are typically scheduled at a higher MCS level, even a single 4.16 μs-long time-domain symbol has the capacity to transmit kilobytes of data, which cannot be fully taken advantage of given the low 10 Mbps arrival rate. Insufficient data is buffered at the MAC layer to utilize the full slot and useless padding bits must be added. This effect is felt less by users under the PF and EDF policies, which are inherently more fair and allow more resources to be scheduled for lower-MCS users.

The ensuing effect of these trends on latency is shown in Figure 9b. Here latency is measured as the time between the arrival time of packets at the PDCP layer of the eNB stack and the time they are delivered to the IP layer at the UE. Naturally, the MR scheduler offers the best delay performance because only 40% of the users, i.e., those with the highest rates, are ever scheduled (unscheduled users with zero rate are not included in the figure). The RR policy offers the highest worst-case delays but is able to achieve mean latencies of less than 1 ms for more than 60% of the users. Of all the policies besides MR, Earliest-Deadline First offers the best worst-case delays, as it attempts to balance the delay of all users by scheduling them exclusively based on their relative deadlines (not taking into account achievable rates). The EDF scheduler is able to achieve a mean UE latency of 1.6 ms, which, as we will see from the experiment in the next section, drops below 1 ms for 60 or fewer users (with the same arrival rate).

Finally, it can be observed from Figure 10b that, despite having the same total packet arrival rate of 700 Mbps as in the 70 UE case, latencies are much lower overall in the 7 UE/100 Mbps per UE case. This can be clearly explained by the higher utilization factor in Table IV. In this scenario, the number of available slots for scheduling different users is no longer the bottleneck. Though we still see that a significant number of users are scheduled at rates that exceed their 100 Mbps arrival rates, the utilization is notably better than in the 10 Mbps case for all scheduling policies. Thus, the channel capacity itself is better utilized, allowing most users to

be scheduled at the requested rate, thereby avoiding additional queue wait time and delay.

Additional results on the impact of the scheduling on throughput and latency can be found in [103].

C. Latency Evaluation for Variable and Fixed TTI Schemes

The results and the discussion introduced in this section are taken from our previous article [21], where the interested reader will find a more comprehensive treatment of techniques to achieve low latency in mmWave 5G cellular systems. The parameters for the scenario considered in this evaluation are provided in Table V. While the qualitative benefits of variable TTI over fixed TTI may seem self-evident, in this section we quantify the performance gains for a multi-user TDMA mmWave system with 1 GHz of bandwidth. We also demonstrate that, with the low-latency scheduling loop enabled by the proposed frame structure, LTE-style Hybrid ARQ can still be employed for enhanced link-layer reliability without excessively exceeding the delay constraints. We model the subframe formats shown in Figure 6 for two subframe periods: the default 100 μs subframe, equivalent to 24 OFDM symbols, and a 66.67 μs subframe, equivalent to 16 OFDM symbols. The symbol length of 4.16 μs is based on the numerology in [104]. Each subframe has one fixed DL-CTRL and one UL-CTRL symbol, with the remaining symbols used for DL or UL data slots. For fixed TTI mode, the entire subframe is allocated to a single user, whereas for variable TTI mode, the scheduler may allocate any number of data symbols within the subframe to match the throughput required by each user.

We also note that UEs are again modeled as moving at 25 m/s, typical of vehicular speeds, which causes fast channel variations and frequent packet errors from small-scale fading (it is observed that between 0.5% and 3% of the transport blocks are lost and require retransmission).

We consider a simple traffic model with Poisson arrivals, where each UE sends an average of 12.5K packets per second (100-byte packets resulting in an average rate of 10 Mbps), as well as a separate, higher-throughput case where each UE sends an average of 83K packets per second (1200-byte packets resulting in an average rate of 100 Mbps). Scheduling is performed based on the EDF policy where the scheduler attempts to deliver each IP packet within 1 ms from its arrival at the PDCP layer and packets are assigned a priority based on how close they are to the deadline. Priority is therefore always given to HARQ retransmissions. We simulate the performance for between 10 and 100 UEs for a 10 Mbps (per UE) arrival rate and between 1 and 10 UEs for the 100 Mbps case,

(a) 10 Mbps per UE arrival rate (100-bytes packets)

(b) 100 Mbps per UE arrival rate (1200-bytes packets)
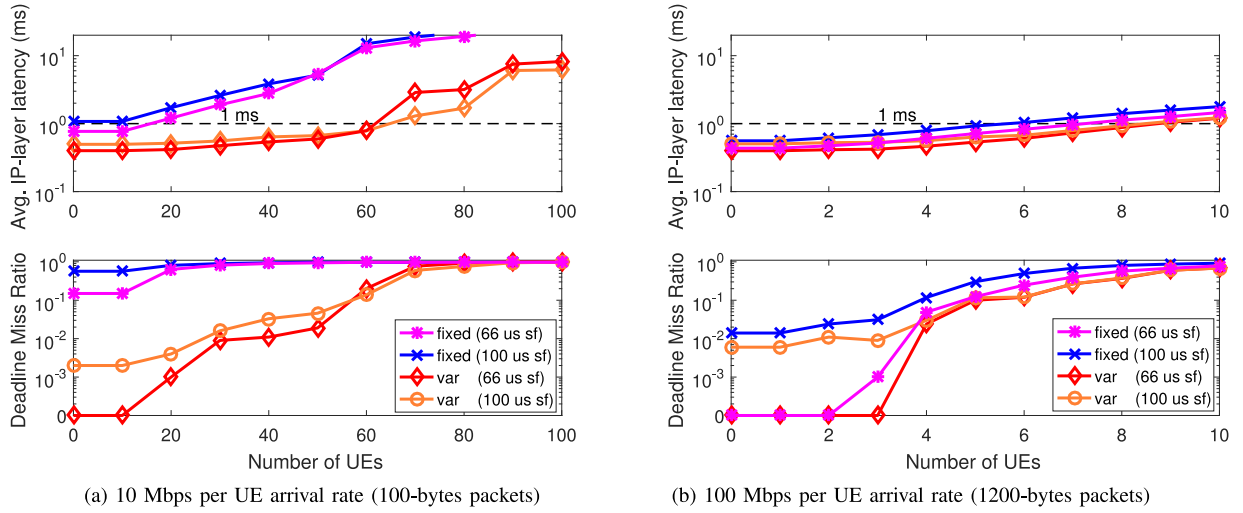
Fig. 11.   Latency and Deadline Miss Ratio as a function of the downlink IP-layer arrival rate for fixed and variable TTI radio frame structures.

equivalent to a total IP-layer arrival rate of between 100 and 1000 Mbps in both cases.

In Fig. 11, the downlink radio link latency is averaged among the best 95% of the users (i.e., the 5% of UEs experiencing the highest latency are not considered). The Deadline Miss Ratio (DMR), which represents the fraction of packets delivered after the 1 ms deadline, is also given for the top 95th percentile UEs. We see that, for a 10 Mbps arrival rate (Figure 11a), variable TTI is able to achieve sub-ms average latency and a DMR of about 10% with over 60 users (corresponding to a 600 Mbps total packet arrival rate) and consistently outperforms fixed TTI. Fixed TTI, despite the relatively short subframe compared to LTE, exceeds 1 ms average latency and has a DMR of over 60% even for the 20 UE case and of more than 90% for 40 or more users. This result shows that variable TTI will be essential for reliable, low-latency service, particularly when considering use cases with many lower-rate devices, such as Machine-Type Devices (MTDs) [4].

For the higher-load (100 Mbps arrival rate per UE) case in Figure 11b, we expect the deviation between the variable and fixed TTI schemes to be less pronounced, as the bottleneck is now the multi-user channel capacity and not the minimum slot size. However, we do find a reduction in radio link latency of around 500 μs, or 30%, for the variable TTI scheme in some cases.

We also find that, for a smaller number of users, the shorter 66.67 μs subframe offers some improvement over the longer 100 μs subframe thanks to the decreased turnaround time. In particular, the DMR is consistently less for the 100 Mbps/UE case for both variable and fixed TTI. However, this trend reverses with more users due to the lower ratio of control to data symbols in the 100 μs subframe case. We note that the control overhead could be somewhat mitigated by multiplexing data in the DL-CTRL region.

We also note that, in real-world implementations, there may be some additional delay related to beam tracking (i.e., for computing and applying the optimal TX/RX beamforming vectors), although the performance limitations of adaptive

beamforming transceivers and channel tracking techniques in future implementations are still unknown. We assume that this delay can be neglected in our analysis because data is constantly being transmitted to each UE and channel state feedback is being transmitted by the UEs to the eNB in each subframe period (which is well within the coherence time observed in many studies), thus ensuring that the channel state information is always up-to-date at the eNB.

The performance of a mmWave cellular network with respect to the end-to-end or the RAN latency has been studied in several papers with the ns-3 mmWave module. Ford *et al.* [21], [103] discuss architectural and protocol solutions for low latency networks based on mmWave, while [10], [19], [38], and [105] propose and evaluate latency-reduction techniques for transport protocols and video streaming.

D. TCP Performance Over mmWave

Another typical use case involves the simulation of end-to-end networks, in order to assess the performance of higher-layer protocols on top of mmWave links. An example of these kinds of simulations is shown in Figure 12, where we evaluated the TCP performance with an eNB that uses Drop-tail or CoDel queue management at the RLC layer, and a mobile UE is experiencing blockages from other humans. The sender opens a File Transfer Protocol (FTP) connection and sends a large file to the UE. The congestion control is TCP Cubic, with delayed acknowledgment disabled. The maximum queue length is 50K packets. The core network latency is 40 ms. The UE is walking at 1 m/s, 300 meters away from the base station, while maintaining LOS connectivity, and experiencing 3 human blocking events.

Drop-tail: Since the RLC queue size is large enough and all packets lost in the wireless link are recovered by means of lower layer retransmissions (RLC AM and MAC HARQ), the sender is unaware of the packet loss, thus keeping a large congestion window that results in high throughput, but also high buffer occupancy and consequent high delay.
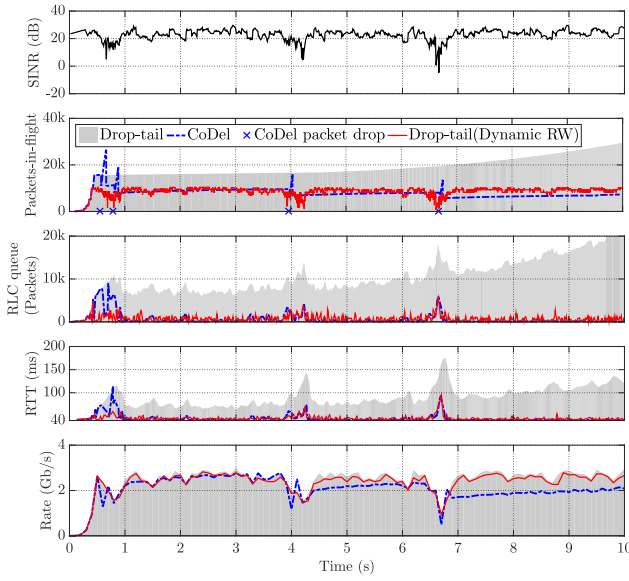
Fig. 12. TCP performance of a single UE with human blockages. ©[2017] IEEE. Reprinted, with permission, from [18].

CoDel: CoDel has the ability to actively drop packets when it detects high buffering delay. The CoDel packet drop events are also labeled in Figure 12. At 0.5 s, as the RLC queue is building up, the first packet is dropped, which informs the sender to reduce the congestion window. At 0.8 s, a human blockage deteriorates the wireless link capacity and causes the RLC queue to grow, thus triggering one more packet drop. Similarly, at 4 s and 6.7 s, two more packets are dropped. As a consequence, the congestion window will decrease and all the packets buffered in the RLC will be delivered. Nonetheless, after the wireless link has recovered from a human blockage, the congestion window ramps up to link capacity very slowly.

Dynamic Receive Window (RW): Our previous results in [18] showed that a user may mitigate the delay by sending TCP acknowledgments containing Dynamic RW. The optimal RW is fed back to the sender, and the sender takes the minimum of the receiver window and the congestion window as the sending window. As a result, the sending rate is precisely regulated so that the delay is reduced without rate degradation.

Additional results on the performance of TCP on mmWave cellular networks obtained with ns-3 mmWave can be found in the literature. In [10], [18], [19], [106], and [107] the authors investigate the performance of different congestion control algorithms, considering in particular the throughput-latency trade off. Proxy architectures that improve the throughput [108] and both throughput and latency [38] have been proposed, as well as uplink-based cross layer congestion control algorithms [37]. The performance of TCP with different mobility management schemes is studied in [39]. Finally, [19], [20] use ns-3 mmWave to analyze the performance of Multipath TCP [109] in mmWave and LTE cellular networks.

E. LTE-Aided Multi Connectivity

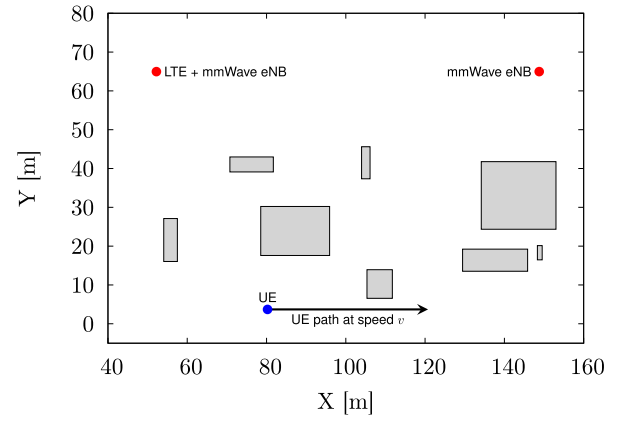Finally, a fourth use case regards the study of mobility management schemes for mmWave radio access networks. The

Fig. 13. Random realization of the simulation scenario. The grey rectangles are randomly deployed non-overlapping obstacles.

(a) PDCP throughput with *Fast Switching*.

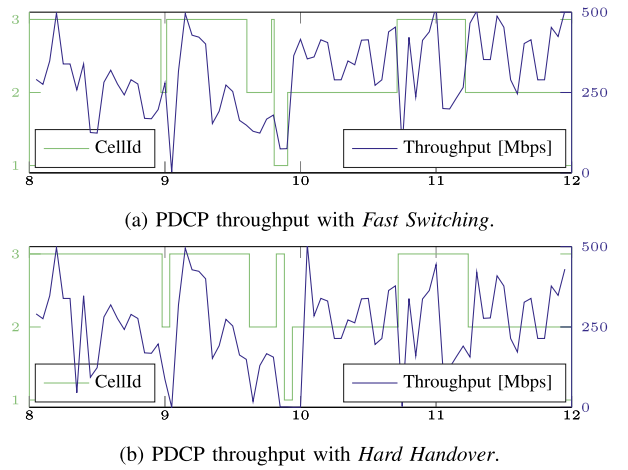(b) PDCP throughput with *Hard Handover*.

Fig. 14. PDCP throughput with multiple RATs and eNBs. eNBs with CellId 2 and 3 are mmWave eNBs, while CellId 1 stands for the LTE eNB co-located with mmWave eNB 2.

example described in the following paragraphs uses the Dual Connectivity extension introduced in Section IX. Thanks to the `McUeNetDevice`, it is possible to simulate scenarios similar to that in Figure 13, with an LTE eNB and multiple mmWave eNBs under its coverage, either co-located, or connected via the X2 interface. The UE can move with different patterns, according to the mobility models available in ns–3, and while moving experiences different channel conditions according to its LOS/NLOS position and distance from the eNBs. These scenarios can be used to evaluate the performance of different mobility management and multi connectivity schemes.

An example can be found in the `mc-twoenbs.cc` file. The purpose of this example is to compare a system with DC and fast switching among Radio Access Technologies (RATs) and a stand-alone architecture, with the UE connected to a single RAT (either LTE or mmWave) at any given time. The benefit of the first architecture is evident when the UE connected to the mmWave RAT experiences an outage: thanks to dual connectivity, it can immediately recover the communication using the LTE link, while the stand-alone UE must sense the link failure and then perform a handover to the LTE eNB. This takes more time and leads to a time interval in

which the throughput that can be achieved at the PDCP layer is zero. This is shown in Figure 14. The green line represents the current cell to which the UE is attached, and this information can be retrieved from the `CellIdStats.txt` and `CellIdStatsHandover.txt` files. The purple line, instead, is the instantaneous PDCP throughput, sampled from the `DlPdcpStats.txt` trace.

Besides this specific example, more general results on dual connectivity for LTE and mmWave can be found in [15] and [30], where smart handover strategies are proposed and extensively evaluated with UDP as the transport protocol, and in [39] and [106], which instead use TCP.

XI. INTEGRATION WITH DCE AND EXAMPLES

DCE was introduced in [51] as a powerful tool that combines the flexibility of a network simulator such as ns–3 with the robustness of the TCP/IP stack of the Linux kernel and the authenticity of real applications. There are several benefits in using this tool. First, the Linux kernel implements protocols which are not yet available for ns–3, or which are in an early development phase and present some limitations. An example is Multipath TCP (MPTCP), the multipath extension of TCP which makes it possible to transmit data on multiple subflows (i.e., a mobile user could simultaneously transmit on a Wi-Fi subflow and a cellular subflow) [109]. At the time of writing, it was implemented for ns–3 by different projects [110], [111], but none of them is completely compliant with the MPTCP specification, and they are not integrated in the main ns–3 release and validated. With DCE, instead, it is possible to use the MPTCP code developed and tested by the same MPTCP protocol designers [112]. Second, the Linux kernel TCP/IP stack is the most widely used in real production environments and datacenters, besides being the basis for the Android mobile operating system. Therefore, it is a very well tested codebase, with very few bugs. Moreover, its usage in network simulations provides a higher level of realism. Finally, with DCE it is possible to use real POSIX socket-based applications. For example, the well known iPerf tool [113] can be used to measure the maximum achievable datarate in the network. It is also possible to simulate a website, with an http daemon in the server and wget as a client. Besides, standard ns–3 applications (`OnOffApplication`, `BulkSendApplication`) can be used with the Linux TCP/IP stack thanks to DCE Cradle.

In order to integrate DCE with the ns–3 mmWave module, it is necessary to patch the `KernelFdSocketFactory` class so that it recognizes the `MmWaveUeNetDevice`. The patch can be found in the `utils` folder of the ns–3 mmWave module repository. Then, replace the standard ns–3 folder with our mmWave module. Notice that, if MPTCP is used as the transport protocol, the DC extension must be used with the patch provided in the `utils` folder.

MPTCP on mmWave links: The latest Linux kernel implementation of MPTCP compatible with DCE can be found in the `net-next-nuse` library of the LibOS project [114]. The standard DCE distribution already provides MPTCP examples, which can be promptly extended in order to account for mmWave and LTE subflows, as long as they operate on links
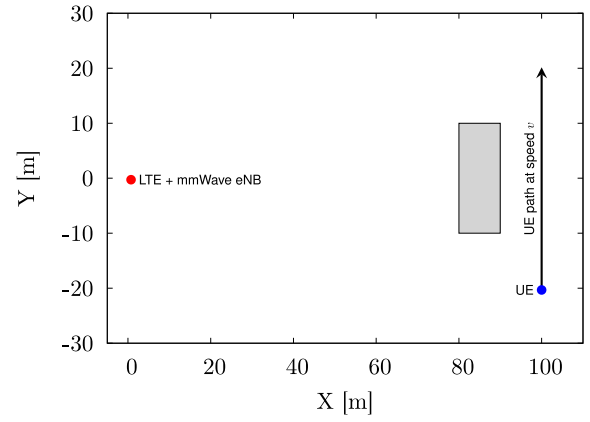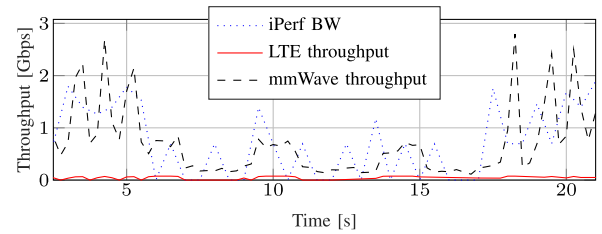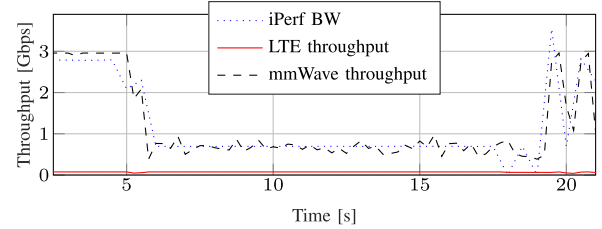

Fig. 15.   Random realization of the simulation scenario. The grey rectangle represents a building.

(a) Balanced Link Adaptation (BALIA)

(b) Uncoupled Cubic

Fig. 16.   Throughput with MPTCP.

with different carrier frequencies (i.e., it is possible to simulate an MPTCP connection on a 2.1 GHz LTE link and a 28 GHz mmWave link). It is possible to simulate different state of the art congestion control algorithms for MPTCP, either coupled or uncoupled, as shown in [19] and [20].

An example is in the file `dce-example-mptcp-mmwave`, which creates the scenario shown in Figure 15. The application used is iPerf, and the mobile device creates two uplink subflows to a remote server, the first on mmWave, and the second on LTE. The UE moves along the y-axis and switches from a LOS to a NLOS condition, and then returns to LOS. Figure 16 shows the output of a simulation, with the TCP throughput for two different congestion control algorithms for MPTCP, together with the per-subflow Radio Access Network throughput. In particular, Figure 16a shows the performance of Balanced Link Adaptation (BALIA) [115], which is a coupled congestion control algorithm, i.e., it tries to adapt the congestion window of each MPTCP subflow according to the congestion experienced on all links. In Figure 16b, instead, the congestion controls on the LTE and

on the mmWave subflows are uncoupled, i.e., each subflow is independent, and TCP Cubic is used. The first observation is that the LTE subflow has, as expected, a much smaller throughput compared to the mmWave subflow, and thus the total throughput measured with iPerf is similar to that of the mmWave connection. The second is that the uncoupled solution manages to reach a more stable throughput in NLOS conditions, compared to the coupled solution, as was observed in [19] and [20], showing that the current coupled congestion control algorithms are not well suited for a deployment over these kinds of links.

XII. POTENTIAL USES AND FUTURE EXTENSIONS

Given the full-stack nature of the simulation framework introduced in this paper, the 5G mmWave research community can leverage this tool to bring and test innovation at every layer. Each module can be easily extended while maintaining full backward compatibility. The fundamental components are in the form of functions, classes and modules, which can be implemented to design novel algorithms, procedures, and, more in general, architectures. For example, the *scheduling* and *allocation* strategies proposed in [116]–[118] can be readily integrated and tested in our framework with some simple tweaks. Similarly, due to the importance of coping with *mobility* and frequent *handovers*, innovative approaches like the one proposed in [119], which exploits *caching*, can take advantage of the modular structure of the ns–3 framework to test flexible and reprogrammable logics. Additionally, as previously mentioned, several papers already fully exploit the capabilities of this simulator to capture the performance of *TCP* in mmWave networks, and propose some novel approaches to mitigate the limiting effects of congestion control procedures with intermittent multi-Gbps mmWave links [10], [18], [38], [39], [108].

As part of our future work, we aim at expanding the code to include additional components such as:

- 3GPP-inspired *signaling/beamtracking* procedures to better accommodate novel techniques like those proposed in [118] and [120];
- multi connectivity based on Carrier Aggregation [121], as detailed in [122];
- novel applications such as *virtual & augmented reality*, to ultimately test key 5G metrics as done in [123], where the authors leverage our mmWave module to run a performance analysis of traditional video delivery over mmWaves, and in [105], where ns-3 mmWave is used to assess the performance of network coding and multi connectivity for reliable video streaming over mmWave;
- *multi-hop* architectures for both the access and the backhaul links, as presented in [124];
- *vehicular* channel and traffic models to test and capture the end-to-end performance of mmWave communications for high-mobility scenarios [125]–[127];
- *public safety* scenarios, including aerial communications and robotics, where the propagation environments and the performance requirements differ from those of traditional cellular networks, as detailed in [128].

Moreover, we plan to address the challenge of *scalability*. 5G networks will likely comprise a large number of nodes,

with high mobility, and thus channel states must be updated frequently. In addition, the use of low-latency applications requires that packet timelines must be scheduled at very fast interaction times. To do so, we will explore the following two design options:

- *Low-rank channel modeling:* We will develop computationally simple low-rank models that approximate the end-to-end phased array system well.
- *Migration to cluster computing:* Further scaling will be achieved by investigating the deployment of the simulator onto open-source large cluster platforms such as Amazon Web Services (AWS) [129].

Finally, we plan to officially merge our mmWave module into the main ns–3 release, in order to regularly maintain its compatibility with every new update. The interested reader can refer to the ns-3 mmWave repository [26] and its wiki page for the latest updates.

XIII. CONCLUSION

In this tutorial paper, we have presented the current status of the ns–3 framework for simulation of mmWave cellular systems. The code, which is publicly available at GitHub [26], is highly modular and customizable to allow researchers to test novel 5G protocols. We have shown some performance trends based on the mmWave channel models available. A detailed explanation of our configurable physical and MAC layers is provided, along with a corroborating set of simulation results for varying configurations. Implementations of advanced 5G architectural features, such as dual connectivity, are also available, and we have reported different representative results. We have also shown that the module can be interfaced with the higher-layer protocols and core network models from the ns–3 LTE module to enable full-stack simulations of end-to-end connectivity, along with the simulation of real applications through the implementation of direct code execution. The module is demonstrated through several example simulations showing the performance of our custom mmWave stack as well as custom congestion control algorithms, specifically designed for efficient utilization of the mmWave channel.

ACRONYMS

3GPP	3rd Generation Partnership Project
5G	5th generation
AGWN	Additive White Gaussian Noise
AM	Acknowledged Mode
AMC	Adaptive Modulation and Coding
AQM	Active Queue Management
BALIA	Balanced Link Adaptation
BF	Beamforming
BLER	Block Error Rate
BSR	Buffer Status Report
CB	Code Block
CQI	Channel Quality Information
DC	Dual Connectivity
DCE	Direct Code Execution
DCI	Downlink Control Information
DL	Downlink

DMR	Deadline Miss Ratio
E2E	End-to-End
EDF	Earliest Deadline First
eNB	evolved Node Base
EPC	Evolved Packet Core
FDMA	Frequency Division Multiple Access
FS	Fast Switching
FTP	File Transfer Protocol
gNB	Next Generation Node Base
HARQ	Hybrid Automatic Repeat reQuest
HOL	Head-of-Line
LOS	Line of Sight
LTE	Long Term Evolution
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MI	Mutual Information
MIMO	Multiple Input, Multiple Output
MME	Mobility Management Entity
MPTCP	Multipath TCP
MR	Maximum Rate
MTD	Machine-Type Device
NLOS	Non Line of Sight
O2I	Outdoor to Indoor
OFDM	Orthogonal Frequency Division Multiplexing
PDCCH	Physical Downlink Control Channel
PDCP	Packet Data Convergence Protocol
PDSCH	Physical Downlink Shared Channel
PDU	Packet Data Unit
PF	Proportional Fair
PGW	Packet Gateway
PHY	Physical
PSS	Primary Synchronization Signal
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
RAN	Radio Access Network
RAT	Radio Access Technology
RED	Random Early Detection
RLC	Radio Link Control
RR	Round Robin
RRC	Radio Resource Control
RW	Receive Window
RX	Receiver
SAP	Service Access Point
SCH	Secondary Cell Handover
SDMA	Spatial Division Multiple Access
SINR	Signal to Interference plus Noise Ratio
SM	Saturation Mode
TB	Transport Block
TCP	Transmission Control Protocol
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
TM	Transparent Mode
TTI	Transmission Time Interval
TX	Transmitter
UE	User Equipment
UL	Uplink
UM	Unacknowledged Mode
UML	Unified Modeling Language.

REFERENCES

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.
[2] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
[3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
[4] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
[5] T. S. Rappaport, R. W. Heath, Jr., R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Upper Saddle River, NJ, USA: Pearson Educ., 2014.
[6] "Study on scenarios and requirements for next generation access technologies, V14.1.0," 3GPP, Sophia Antipolis, France, Rep. TR 38.913, 2017.
[7] "NR and NG-RAN overall description—Rel. 15," 3GPP, Sophia Antipolis, France, Rep. TS 38.300, 2018.
[8] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: A MAC layer perspective," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3437–3458, Oct. 2015.
[9] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, Nov. 2015.
[10] M. Zhang *et al.*, "Transport layer performance in 5G mmWave cellular," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, San Francisco, CA, USA, Apr. 2016, pp. 730–735.
[11] K. Allen *et al.*, "Building penetration loss measurements at 900 MHz, 11.4 GHz, and 28 GHz," U.S. Dept. Commerce, Nat. Telecommun. Inf. Admin., Boulder, CO, USA, NTIA Rep. 94-306, 1994. [Online]. Available: https://www.its.bldrdoc.gov/publications/details.aspx?pub=2333
[12] S. Singh, F. Ziliotto, U. Madhow, E. M. Belding, and M. J. W. Rodwell, "Millimeter wave WPAN: Cross-layer modeling and multi-hop architecture," in *Proc. 26th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Barcelona, Spain, May 2007, pp. 2336–2340.
[13] J. S. Lu, D. Steinbach, P. Cabrol, and P. Pietraski, "Modeling human blockers in millimeter wave radio links," *ZTE Commun.*, vol. 10, no. 4, pp. 23–28, Dec. 2012.
[14] A. Ghosh *et al.*, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, Jun. 2014.
[15] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5G mmWave mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2069–2084, Sep. 2017.
[16] M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "An efficient uplink multi-connectivity scheme for 5G mmWave control plane applications," *IEEE Trans. Wireless Commun.*, submitted for publication, 2017. [Online]. Available: https://arxiv.org/abs/1610.04836
[17] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Mobility modeling and performance evaluation of multi-connectivity in 5G intra-frequency networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.
[18] M. Zhang, M. Mezzavilla, J. Zhu, S. Rangan, and S. Panwar, "TCP dynamics over mmwave links," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1–6.
[19] M. Polese, R. Jana, and M. Zorzi, "TCP in 5G mmWave networks: Link level retransmissions and MP-TCP," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Atlanta, GA, USA, 2017, pp. 343–348.
[20] M. Polese, R. Jana, and M. Zorzi, "TCP and MP-TCP in 5G mmWave Networks," *IEEE Internet Comput.*, vol. 21, no. 5, pp. 12–19, Sep. 2017.
[21] R. Ford *et al.*, "Achieving ultra-low latency in 5G millimeter wave cellular networks," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 196–203, Mar. 2017.
[22] R. Ford, A. Sridharan, R. Margolies, R. Jana, and S. Rangan, "Provisioning low latency, resilient mobile edge clouds for 5G," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Atlanta, GA, USA, 2017, pp. 169–174.
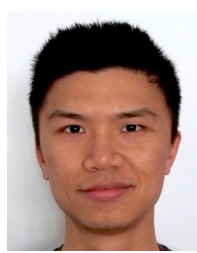
[23] T. R. Henderson, M. Lacage, G. F. Riley, C. Dowell, and J. Kopena, "Network simulations with the ns-3 simulator," *SIGCOMM Demonstration*, vol. 14, no. 14, p. 527, 2008.

[24] N. Baldo, M. Miozzo, M. Requena-Esteso, and J. Nin-Guerrero, "An open source product-oriented LTE network simulator based on ns-3," in *Proc. 14th ACM Int. Conf. Model. Anal. Simulat. Wireless Mobile Syst.*, 2011, pp. 293–298. [Online]. Available: http://doi.acm.org/10.1145/2068897.2068948

[25] (Feb. 2012). *LTE-EPC Network Simulator*. [Online]. Available: http://iptechwiki.cttc.es/LTE-EPC_Network_Simulator_(LENA)

[26] NYU WIRELESS, University of Padova, *ns-3 Module for Simulating mmwave-Based Cellular Systems*. [Online]. Available: https://github.com/nyuwireless-unipd/ns3-mmwave

[27] M. Mezzavilla, S. Dutta, M. Zhang, M. R. Akdeniz, and S. Rangan, "5G mmWave module for the ns-3 network simulator," in *Proc. 18th ACM Int. Conf. Model. Anal. Simulat. Wireless Mobile Syst.*, Cancún, Mexico, 2015, pp. 283–290. [Online]. Available: http://doi.acm.org/10.1145/2811587.2811619

[28] R. Ford *et al.*, "A framework for end-to-end evaluation of 5G mmWave cellular networks in ns-3," in *Proc. Workshop ns-3*, Seattle, WA, USA, 2016, pp. 85–92. [Online]. Available: http://doi.acm.org/10.1145/2915371.2915380

[29] M. Zhang, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "ns-3 implementation of the 3GPP MIMO channel model for frequency spectrum above 6 GHz," in *Proc. Workshop ns-3*, Porto, Portugal, 2017, pp. 71–78. [Online]. Available: http://doi.acm.org/10.1145/3067665.3067678

[30] M. Polese, M. Mezzavilla, and M. Zorzi, "Performance comparison of dual connectivity and hard handover for LTE-5G tight integration," in *Proc. 9th EAI Int. Conf. Simulat. Tools Techn. (SIMUTOOLS)*, Prague, Czech Republic, 2016, pp. 118–123. [Online]. Available: http://dl.acm.org/citation.cfm?id=3021426.3021445

[31] "Study on channel model for frequency spectrum above 6 GHz, V14.2.0," 3GPP, Sophia Antipolis, France, Rep. TR 38.900, 2017.

[32] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[33] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[34] H. Zhao *et al.*, "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city," in *Proc. IEEE ICC*, Budapest, Hungary, 2013, pp. 5163–5167.

[35] G. R. MacCartney, S. Deng, S. Sun, and T. S. Rappaport, "Millimeter-wave human blockage at 73 GHz with a simple double knife-edge diffraction model and extension for directional antennas," in *Proc. IEEE 81st Veh. Technol. Conf.*, Montreal, QC, Canada, Sep. 2016, pp. 1–6.

[36] C. N. Barati *et al.*, "Directional cell discovery in millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6664–6678, Dec. 2015.

[37] T. Azzino, M. Drago, M. Polese, A. Zanella, and M. Zorzi, "X-TCP: A cross layer approach for TCP uplink flows in mmwave networks," in *Proc. 16th Annu. Mediterranean Ad Hoc Netw. Workshop (Med-Hoc-Net)*, Budva, Montenegro, Jun. 2017, pp. 1–6.

[38] M. Polese *et al.*, "milliProxy: A TCP proxy architecture for 5G mmWave cellular systems," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Oct. 2017, pp. 951–957.

[39] M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Mobility management for TCP in mmWave networks," in *Proc. 1st ACM Workshop Millimeter-Wave Netw. Sens. Syst. (mmNets)*, 2017, pp. 11–16.

[40] H. Assasa and J. Widmer, "Implementation and evaluation of a WLAN IEEE 802.11ad model in ns-3," in *Proc. Workshop ns-3*, Seattle, WA, USA, 2016, pp. 57–64.

[41] H. Assasa and J. Widmer, "Extending the IEEE 802.11ad model: Scheduled access, spatial reuse, clustering, and relaying," in *Proc. Workshop ns-3*, Porto, Portugal, 2017, pp. 39–46.

[42] T. Kim *et al.*, "Tens of Gbps support with mmWave beamforming systems for next generation communications," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013, pp. 3685–3690.

[43] K. Zheng *et al.*, "10 Gb/s HetSNets with millimeter-wave communications: Access and networking—Challenges and protocols," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 222–231, Jan. 2015.

[44] C. Dehos, J. L. González, A. D. Domenico, D. Kténas, and L. Dussopt, "Millimeter-wave access and backhauling: The solution to the exponential data traffic increase in 5G mobile communications systems?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 88–95, Sep. 2014.

[45] R. Taori and A. Sridharan, "Point-to-multipoint in-band mmWave backhaul for 5G networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 195–201, Jan. 2015.

[46] T. R. Henderson, S. Roy, S. Floyd, and G. F. Riley, "ns-3 project goals," in *Proc. Workshop ns-2 IP Netw. Simulator*, Pisa, Italy, 2006, p. 13.

[47] M. Casoni, C. A. Grazia, M. Klapez, and N. Patriciello, "Implementation and validation of TCP options and congestion control algorithms for ns-3," in *Proc. Workshop ns-3*, Barcelona, Spain, 2015, pp. 112–119.

[48] G. Pei and T. Henderson, "Validation of ns-3 802.11b PHY model," Boeing Res. Technol., Seattle, WA, USA, Rep., 2009. [Online]. Available: https://www.nsnam.org/ pei/80211b.pdf

[49] J. Farooq and T. Turletti, "An IEEE 802.16 WiMAX module for the ns-3 simulator," in *Proc. 2nd Int. Conf. Simulat. Tools Techn.*, Rome, Italy, 2009, pp. 1–11, doi: 10.4108/ICST.SIMUTOOLS2009.5644.

[50] H. Narra, Y. Cheng, E. K. Çetinkaya, J. P. Rohrer, and J. P. G. Sterbenz, "Destination-sequenced distance vector (DSDV) routing protocol implementation in ns-3," in *Proc. 4th Int. ICST Conf. Simul. Tools Techn.*, Barcelona, Spain, 2011, pp. 439–446. [Online]. Available: http://dl.acm.org/citation.cfm?id=2151054.2151132

[51] H. Tazaki *et al.*, "Direct code execution: Revisiting library OS architecture for reproducible network experiments," in *Proc. 9th ACM Conf. Emerg. Netw. Exp. Technol. (CoNEXT)*, Santa Barbara, CA, USA, 2013, pp. 217–228. [Online]. Available: http://doi.acm.org/10.1145/2535372.2535374

[52] Centre Tecnologic de Telecomunicacions de Catalunya. (Jan. 2014). *The LENA ns-3 LTE Module Documentation*. [Online]. Available: http://iptechwiki.cttc.es/ LTE-EPC_Network_Simulator_(LENA)

[53] "Spatial channel model for multiple input multiple output (MIMO) simulations, V6.1.0," 3GPP, Sophia Antipolis, France, Rep. TR 25.996, 2003.

[54] *Winprop Software*. Accessed: Apr. 28, 2018. [Online]. Available: https://altairhyperworks.com/product/FEKO/WinProp-Propagation-Modeling

[55] R. N. Almesaeed, A. S. Ameen, E. Mellios, A. Doufexi, and A. Nix, "3D channel models: Principles, characteristics, and system implications," *IEEE Commun. Mag.*, vol. 55, no. 4, pp. 152–159, Apr. 2017.

[56] S. Jaeckel, L. Raschkowski, K. Borner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.

[57] S. Sun, G. R. MacCartney, Jr., and T. S. Rappaport, "A novel millimeter-wave channel simulator and applications for 5G wireless communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–7.

[58] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2207–2225, Jul. 2016.

[59] NYU Wireless. (2016). *NYUSIM: The Open Source 5G Channel Model Simulator Software*. [Online]. Available: http://wireless.engineering.nyu.edu/5g-millimeter-wave-channel-modeling-software/

[60] M. Giordani, M. Mezzavilla, A. Dhananjay, S. Rangan, and M. Zorzi, "Channel dynamics and SNR tracking in millimeter wave cellular systems," in *Proc. 22th Eur. Wireless Conf.*, Oulu, Finland, May 2016, pp. 1–8.

[61] P. A. Eliasi, S. Rangan, and T. S. Rappaport, "Low-rank spatial channel estimation for millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2748–2759, May 2017, doi: 10.1109/TWC.2017.2662687.

[62] D. J. Love and R. W. Heath, "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102–1110, Jul. 2003.

[63] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, vol. 87. Oxford, U.K.: Clarendon Press, 1965.

[64] J. Wang, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.

[65] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[66] M. Rebato, J. Park, P. Popovski, E. D. Carvalho, and M. Zorzi, "Stochastic geometric coverage analysis in mmWave cellular networks with a realistic channel model," in *Proc. IEEE Glob. Commun. Conf. GLOBECOM*, Singapore, Dec. 2017, pp. 1–6.

[67] M. Rebato, M. Mezzavilla, S. Rangan, F. Boccardi, and M. Zorzi, "Understanding noise and interference regimes in 5G millimeter-wave cellular networks," in *Proc. 22nd Eur. Wireless Conf.*, Oulu, Finland, 2016, pp. 1–5.

[68] M. Mezzavilla, M. Miozzo, M. Rossi, N. Baldo, and M. Zorzi, "A lightweight and accurate link abstraction model for the simulation of LTE networks in ns-3," in *Proc. 15th ACM Int. Conf. Model. Anal. Simulat. Wireless Mobile Syst. (MSWiM)*, 2012, pp. 55–60.

[69] *The Vienna LTE Simulators*. Accessed: Apr. 28, 2018. [Online]. Available: https://www.nt.tuwien.ac.at/research/mobile-communications/vccs/vienna-lte-a-simulators/

[70] Z. Pi and F. Khan, "System design and network architecture for a millimeter-wave mobile broadband (MMB) system," in *Proc. IEEE Sarnoff Symp.*, Princeton, NJ, USA, May 2011, pp. 1–6.

[71] *KT 5th Generation Radio Access; Physical Layer; General Description—5G Pre-Specification*, 5G SIG document 5G.201, Korea Telecom, Seongnam, South Korea, 2016.

[72] M. Cudak *et al.*, "Experimental mmWave 5G cellular system," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, USA, 2014, pp. 377–381.

[73] T. Levanen, J. Pirskanen, and M. Valkama, "Radio interface design for ultra-low latency millimeter-wave communications in 5G era," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, USA, Dec. 2014, pp. 1420–1426.

[74] S. Dutta *et al.*, "Frame structure design and analysis for millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1508–1522, Mar. 2017.

[75] S. Dutta *et al.*, "MAC layer frame design for millimeter wave cellular system," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Athens, Greece, Jun. 2016, pp. 117–121.

[76] D. Astély *et al.*, "LTE: The evolution of mobile broadband," *IEEE Commun. Mag.*, vol. 47, no. 4, pp. 44–51, Apr. 2009.

[77] *5G TF; Air Interface Working Group; Verizon 5th Generation Radio Access; Physical Channels and Modulation (Release 1)*, Verizon, Basking Ridge, NJ, USA, 2016. [Online]. Available: http://www.5gtf.net/V5G_211_v1p7.pdf

[78] "Study on new radio access technology—Physical layer aspects, V14.0.0," 3GPP, Sophia Antipolis, France, Rep. TR 38.802, 2017.

[79] ITU-R WP5D, Working Document Toward Preliminary Draft New Recommendation ITU-R M.[IMT.VISION]. (2014). *Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-PDF-E.pdf

[80] P. Popovski *et al.*, "EU FP7 INFSO-ICT-317669 METIS, D1. 1 scenarios, requirements and KPIs for 5G mobile and wireless system," document ICT-317669-METIS/D1.1, METIS, 2013. [Online]. Available: http://publications.lib.chalmers.se/records/fulltext/213055/local_213055.pdf

[81] P. Kela *et al.*, "A novel radio frame structure for 5G dense outdoor radio access networks," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC-Spring)*, Glasgow, U.K., May 2015, pp. 1–6.

[82] S. Choi and K. G. Shin, "A class of adaptive hybrid ARQ schemes for wireless links," *IEEE Trans. Veh. Technol.*, vol. 50, no. 3, pp. 777–790, May 2001.

[83] "Medium access control (MAC) protocol specification, V14.0.0," 3GPP, Sophia Antipolis, France, Rep. TS 36.321, 2016.

[84] S. Sesia, M. Baker, and I. Toufik, *LTE—The UMTS Long Term Evolution: From Theory to Practice*. Chichester, U.K.: Wiley, 2011.

[85] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *4G LTE/LTE-Advanced for Mobile Broadband*. Oxford, U.K.: Academic Press, 201.

[86] "Technical specification group services and system aspects; policy and charging control architecture, V14.2.0," 3GPP, Sophia Antipolis, France, Rep. TR 23.203, 2017.

[87] G. Abbas, Z. Halim, and Z. H. Abbas, "Fairness-driven queue management: A survey and taxonomy," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 324–367, 1st Quart., 2016.

[88] F. Baker and G. Fairhurst, "IETF recommendations regarding active queue management," Internet Eng. Task Force, Fremont, CA, USA, RFC 7567, 2015.

[89] P. Imputato and S. Avallone, "Design and implementation of the traffic control module in ns-3," in *Proc. Workshop ns-3*, Seattle, WA, USA, 2016, pp. 1–8.

[90] P. Imputato and S. Avallone, "Traffic differentiation and multiqueue networking in ns-3," in *Proc. Workshop ns-3*, Porto, Portugal, 2017, pp. 79–86.

[91] A. Deepak, K. S. Shravya, and M. P. Tahiliani, "Design and implementation of AQM evaluation suite for ns-3," in *Proc. Workshop ns-3*, 2017, pp. 87–94.

[92] J. Gettys and K. Nichols, "Bufferbloat: Dark buffers in the Internet," *ACM Queue*, vol. 9, no. 11, pp. 40–54, Nov. 2011.

[93] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Trans. Netw.*, vol. 1, no. 4, pp. 397–413, Aug. 1993.

[94] T. J. Ott, T. V. Lakshman, and L. H. Wong, "SRED: Stabilized RED," in *Proc. 18th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 3. New York, NY, USA, 1999, pp. 1346–1355.

[95] K. Nichols and V. Jacobson, "Controlling queue delay," *Commun. ACM*, vol. 55, no. 7, pp. 42–50, Jul. 2012.

[96] S. Chandrashekar *et al.*, "5G multi-RAT multi-connectivity architecture," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, May 2016, pp. 180–186.

[97] J. G. Rois, B. Lorenzo, F. J. González-Castaño, and J. C. Burguillo, "Heterogeneous millimeter-wave/micro-wave architecture for 5G wireless access and backhauling," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Athens, Greece, Jun. 2016, pp. 179–184.

[98] A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

[99] "Migration and Interworking Aspects—SA WG2 Temporary," document S2-163348, AT&T, Dallas, TX, USA, 2016. [Online]. Available: http://www.3gpp.org/ftp/tsg_sa/WG2_Arch/TSGS2_116_Vienna/Docs/S2-163348.zip

[100] "Study on small cell enhancements for E-UTRA and E-UTRAN, v12.0.0," 3GPP, Sophia Antipolis, France, Rep. TR36.842, 2013.

[101] I. D. Silva *et al.*, "Tight integration of new 5G air interface and LTE to fulfill 5G requirements," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC-Spring)*, Glasgow, U.K., May 2015, pp. 1–5.

[102] B. Nguyen *et al.*, "Towards understanding TCP performance on LTE/EPC mobile networks," in *Proc. 4th Workshop All Things Cell. Oper. Appl. Challenges (AllThingsCellular)*, Chicago, IL, USA, 2014, pp. 41–46. [Online]. Available: http://doi.acm.org/10.1145/2627585.2627594

[103] R. Ford, "Low latency fifth-generation cellular networks," Ph.D. dissertation, Dept. Elect. Comput. Eng., Polytech. Inst. New York Univ., New York, NY, USA, 2017.

[104] Z. Pi and F. Khan, "A millimeter-wave massive MIMO system for next generation mobile broadband," in *Proc. 46th Asilomar Conf. Signals Syst. Comput. (ASILOMAR)*, Pacific Grove, CA, USA, Nov. 2012, pp. 693–698.

[105] M. Drago, T. Azzino, M. Polese, C. Stefanovic, and M. Zorzi, "Reliable video streaming over mmWave with multi connectivity and network coding," in *Proc. Int. Conf. Comput. Netw. Commun. (ICNC)*, Mar. 2018. [Online]. Available: https://arxiv.org/abs/1711.06154

[106] P. Jimenez Mateo, "Analysis of TCP performance in 5G mm-wave mobile networks," M.S. thesis, Grad. School Eng. Basic Sci., Universidad Carlos III de Madrid, Getafe, Spain, 2017.

[107] M. Pieskä and A. Kassler, "TCP performance over 5G mmWave links—Tradeoff between capacity and latency," in *Proc. 13th IEEE Int. Conf. Wireless Mobile Comput. Netw. Commun.*, Rome, Italy, Oct. 2017, pp. 385–394.

[108] M. Kim, S.-W. Ko, and S.-L. Kim, "Enhancing TCP end-to-end performance in millimeter-wave communications," in *Proc. IEEE 28th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Montreal, QC, Canada, Oct. 2017, pp. 1–5.

[109] A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar, "Architectural guidelines for multipath TCP development," Internet Eng. Task Force, Fremont, CA, USA, RFC 6182, 2011.

[110] B. Chihani and C. Denis, "A multipath TCP model for ns-3 simulator," in *Proc. Workshop ns-3 Held Conjunction SIMUTools*, 2011. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00641539/document

[111] M. Coudron and S. Secci, "An implementation of multipath TCP in ns3," *Comput. Netw.*, vol. 116, pp. 1–11, Apr. 2017.

[112] C. Paasch and S. Barre. *Multipath TCP in the Linux Kernel*. Accessed: Apr. 28, 2018. [Online]. Available: http://www.multipath-tcp.org

[113] *Iperf 2.0*. Accessed: Apr. 28, 2018. [Online]. Available: https://iperf.fr

[114] H. Tazaki, R. Nakamura, and Y. Sekiya, "Library operating system with mainline Linux network stack," in *Proc. Netdev 0.1*, 2015.

[115] Q. Peng, A. Walid, J. Hwang, and S. H. Low, "Multipath TCP: Analysis, design, and implementation," *IEEE/ACM Trans. Netw.*, vol. 24, no. 1, pp. 596–609, Feb. 2016.

[116] O. Semiari, W. Saad, and M. Bennis, "Joint millimeter wave and microwave resources allocation in cellular networks with dual-mode base stations," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4802–4816, Jul. 2017.

[117] O. Semiari, W. Saad, and M. Bennis, "Downlink cell association and load balancing for joint millimeter wave-microwave cellular networks," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[118] M. E. Rasekh, Z. Marzi, Y. Zhu, U. Madhow, and H. Zheng, "Noncoherent mmWave path tracking," in *Proc. 18th Int. Workshop Mobile Comput. Syst. Appl.*, Sonoma, CA, USA, 2017, pp. 13–18. [Online]. Available: http://doi.acm.org/10.1145/3032970.3032974

[119] O. Semiari, W. Saad, M. Bennis, and B. Maham, "Caching meets millimeter wave communications for enhanced mobility management in 5G networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 779–793, Feb. 2018.

[120] J. Palacios, D. D. Donno, and J. Widmer, "Tracking mm-Wave channel dynamics: Fast beam training strategies under mobility," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Atlanta, GA, USA, May 2017, pp. 1–9.

[121] Z. Khan *et al.*, "Carrier aggregation/channel bonding in next generation cellular networks: Methods and challenges," *IEEE Netw.*, vol. 28, no. 6, pp. 34–40, Nov./Dec. 2014.

[122] T. Zugno, M. Polese, and M. Zorzi, "Integration of carrier aggregation and dual connectivity for the ns-3 mmWave module," in *Proc. Workshop ns-3 (WNS3)*, 2018. [Online]. Available: https://arxiv.org/abs/1802.06706

[123] Y. Hou, W. Zhou, L. Song, and M. Gao, "A QoE estimation model for video streaming over 5G millimeter wave network," in *Proc. Int. Conf. Broadband Wireless Comput. Commun. Appl.*, 2016, pp. 93–104.

[124] A. Drozdy, J. Kapanen, and J. Manner, "User level performance analysis of multi-hop in-band backhaul for 5G," *Wireless Netw.*, pp. 1–15, Apr. 2017. [Online]. Available: https://link.springer.com/article/10.1007%2Fs11276-017-1513-2#citeas

[125] A. Tassi, M. Egan, R. J. Piechocki, and A. Nix, "Modeling and design of millimeter-wave networks for highway vehicular communication," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 10676–10691, Dec. 2017.

[126] I. Mavromatis, A. Tassi, R. J. Piechocki, and A. Nix, "mmWave system for future ITS: A MAC-layer approach for V2X beam steering," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Toronto, ON, Canada, Sep. 2017, pp. 1–6.

[127] I. Mavromatis, A. Tassi, R. J. Piechocki, and A. Nix, "Beam alignment for millimetre wave links with motion prediction of autonomous vehicles," in *Proc. IET Conf.*, 2017, p. 6 (8.), doi: 10.1049/ic.2017.0023.

[128] M. Mezzavilla *et al.*, "Public safety communications above 6 GHz: Challenges and opportunities," *IEEE Access*, vol. 6, pp. 316–329, 2018.

[129] *AWS Storage Services Overview*. Accessed: Apr. 28, 2018. [Online]. Available: https://goo.gl/h2yJje

Menglei Zhang received the B.S. degree in electrical engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2013 and the M.S. degree in electrical engineering from the New York University Tandon School of Engineering, Brooklyn, NY, USA, in 2015, where he is currently pursuing the Ph.D. degree in electrical engineering with Prof. Rangan. His research interests include wireless communications, channel modeling, transport congestion control, and system level simulation.
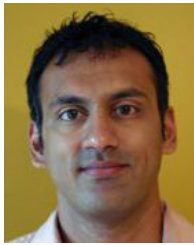
Michele Polese (S'17) received the B.Sc. degree in information engineering and the M.Sc. degree in telecommunication engineering from the University of Padova, Italy, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree with the Department of Information Engineering, under the supervision of Prof. M. Zorzi. He visited New York University (NYU), in 2017, and AT&T Labs, Bedminster, NJ, USA, in 2018, for one month. He is collaborating with several academic and industrial research partners, including Intel, InterDigital, NYU, AT&T Labs, University of Aalborg, King's College, and NIST. His research interests are in the analysis and development of protocols and architectures for the next generation of cellular networks (5G), in particular for millimeter-wave communication, and in the performance evaluation of complex networks.

Russell Ford (S'09–M'18) received the B.S. degree in electrical and computer engineering from Florida State University in 2010, the M.S. degree in electrical engineering from the Polytechnic Institute of New York University (NYU) in 2012, and the Ph.D. degree in computer science from NYU in 2017, under the supervision of Prof. S. Rangan. He is currently a Research Engineer with Samsung Research America, Mountain View, CA, USA. His research interests include 5G mobile cloud/edge computing, MAC and network-layer resource allocation, and network performance modeling and simulation.

Marco Mezzavilla (S'10–M'14) received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of Padua, Italy, in 2007 and 2010, respectively, and the Ph.D. degree in information engineering from the University of Padua in 2013, under the supervision of Prof. M. Zorzi. He held visiting research positions with the NEC Network Laboratories, Heidelberg, Germany, in 2009, the Centre Tecnolgic Telecomunicacions Catalunya, Barcelona, Spain, in 2010, and Qualcomm Research, San Diego, CA, USA, in 2012. He holds a research scientist position with the NYU Tandon School of Engineering, where he leads various mmWave related research projects, mainly focusing on 5G PHY/MAC design. He has been serving as a reviewer for many IEEE conferences, journals, and magazines. He has authored and co-authored multiple publications in conferences, journals, and some patent applications. His research interests include design and validation of communication protocols and applications to fourth generation broadband wireless technologies, millimeter wave communications for 5G networks, multimedia traffic optimization, radio resource management, spectrum sharing, convex optimization, cognitive networks, and experimental analysis.

Sourjya Dutta (S'16) received the B.Tech. degree in electronics and communications engineering from the National Institute of Technology, Durgapur, India, in 2012. He is currently pursuing the Ph.D. degree in electrical and computer engineering with New York University Tandon School of Engineering, Brooklyn, NY, USA, under the guidance of Prof. S. Rangan. His research interests include wireless communications, MAC layer design, network simulation and prototyping.

Sundeep Rangan (S'94–M'98–SM'13–F'16) received the B.A.Sc. degree in electrical engineering from the University of Waterloo, Canada, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of California, Berkeley. He has held Post-Doctoral appointments with the University of Michigan, Ann Arbor, and Bell Labs. In 2000, he co-founded (with four others) Flarion Technologies, a spin off of Bell Labs, that developed Flash OFDM, a precursor to OFDM-based 4G cellular technologies, including LTE and WiMAX. In 2006, Flarion was acquired by Qualcomm Technologies and he became a Director of engineering with Qualcomm involved in OFDM infrastructure products. He joined the ECE Department, Polytechnic School of Engineering (currently, NYU Tandon School of Engineering) in 2010. He is currently an Associate Professor and an Acting Director of NYU WIRELESS, a leading research center in 5G wireless systems. His research interests are in wireless communications, signal processing, information theory, and control theory.

Michele Zorzi (F'07) received the Laurea and Ph.D. degrees in electrical engineering from the University of Padova in 1990 and 1994, respectively. From 1992 to 1993, he was on leave with the University of California San Diego (UCSD). He was with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy, the Center for Wireless Communications, UCSD, and the University of Ferrara. In 2003, he joined the faculty of the Information Engineering Department, University of Padova, where he is currently a Professor. His present research interests include performance evaluation in mobile communications systems, random access in mobile radio networks, ad hoc and sensor networks and IoT, energy constrained communications protocols, 5G millimeter-wave cellular systems, and underwater communications and networking. He was the Editor-in-Chief of the IEEE WIRELESS COMMUNICATIONS from 2003 to 2005, and the IEEE TRANSACTIONS ON COMMUNICATIONS from 2008 to 2011, and is currently the founding Editor-in-Chief of the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. He was a Guest Editor for several Special Issues in IEEE Personal Communications, IEEE Wireless Communications, IEEE Network, and IEEE JSAC. He served as a member-at-large in the Board of Governors of the IEEE Communications Society from 2009 to 2011, and as its Director of Education from 2014 to 2015.