

A Survey of Machine Learning Applications to Handover Management in 5G and Beyond

MICHAEL S. MOLLEL¹, (Student Member, IEEE), ATTAI IBRAHIM ABUBAKAR²,
METIN OZTURK³, SHUBI FELIX KAIJAGE¹, (Senior Member, IEEE), MICHAEL KISANGIRI¹,
SAJJAD HUSSAIN², (Senior Member, IEEE),
MUHAMMAD ALI IMRAN^{2,4}, (Senior Member, IEEE),
AND QAMMER H. ABBASI², (Senior Member, IEEE)

¹The Nelson Mandela African Institution of Science and Technology (NM-AIST), Arusha 23311, Tanzania

²Communication, Sensing and Imaging (CSI) Research Group, James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K.

³Electrical and Electronics Engineering, Ankara Yıldırım Beyazıt University, 06010 Ankara, Turkey

⁴Artificial Intelligence Research Center (AIRC), College of Engineering and Information Technology, Ajman University, Ajman, United Arab Emirates

Corresponding author: Michael S. Mollel (michaelm@nm-aist.ac.tz)

ABSTRACT Handover (HO) is one of the key aspects of next-generation (NG) cellular communication networks that need to be properly managed since it poses multiple threats to quality-of-service (QoS) such as the reduction in the average throughput as well as service interruptions. With the introduction of new enablers for fifth-generation (5G) networks, such as millimetre wave (mm-wave) communications, network densification, Internet of things (IoT), etc., HO management is provisioned to be more challenging as the number of base stations (BSs) per unit area, and the number of connections has been dramatically rising. Considering the stringent requirements that have been newly released in the standards of 5G networks, the level of the challenge is multiplied. To this end, intelligent HO management schemes have been proposed and tested in the literature, paving the way for tackling these challenges more efficiently and effectively. In this survey, we aim at revealing the current status of cellular networks and discussing mobility and HO management in 5G alongside the general characteristics of 5G networks. We provide an extensive tutorial on HO management in 5G networks accompanied by a discussion on machine learning (ML) applications to HO management. A novel taxonomy in terms of the source of data to be utilized in training ML algorithms is produced, where two broad categories are considered; namely, visual data and network data. The state-of-the-art on ML-aided HO management in cellular networks under each category is extensively reviewed with the most recent studies, and the challenges, as well as future research directions, are detailed.

INDEX TERMS Handover, machine learning, mobility management, fifth generation.

I. INTRODUCTION

Wireless communication networks have been witnessing an unprecedented demand in terms of bandwidth and number of connections in this so-called *information age*—in particular the age of big data¹ where data is regarded as *new oil* [1]. It is reported in the Ericsson Mobility Report that the mobile network traffic soared by 56% in the first quarter of 2020 [2], indicating the imminent issue that needs to be addressed. There are strong evidence for the correlation

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

¹Steve Lohr, *The Age of Big Data*, The New York Times, 11 Feb. 2012. Available online at <https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>. Accessed on 25 Oct. 2020.

between such growth in the global data traffic and the proliferation of emerging applications, including tactile-internet, virtual reality, high-definition video streaming. For example, we learnt from the same report [2] that video streaming alone constitutes more than half of the mobile data traffic, and there is a tendency towards higher resolutions—putting the issue at an alarming level in terms of data demand. This, in turn, poses serious challenges to legacy networks and paves the way for the fifth generation of cellular networks (5G), which offers a thousandfold increase in capacity [3]–[5]. As such, enhanced mobile broadband has been included in 5G New Radio (NR) as one of the scenarios—along with ultra-reliable low-latency communications (URLLC) and massive machine-type communications (mMTC)—in

order to support the aforementioned bandwidth-hungry applications [6].

On the other hand, Internet of things (IoT) devices have already pervaded our daily life, as they can be seen in numerous domains, including agriculture [7], healthcare [8], smart living [9]–[11], and smart city [12], to name a few. In the case of smart city, for example, city waste, building health monitoring, traffic, etc., are managed smartly using IoT technology by deploying the IoT devices to the required places accordingly [12]. A good example of this can be found in the publication by the Mayor of London on the road map for smart city agenda² with the slogan “*Smarter London Together*”, which dictates a heavy use of IoT technology in London to make the City more efficient and to boost the standard of living of its residents. IoT technology owes this popularity to the promises in terms of making our everyday life as well as industrial processes more manageable and efficient with continuous monitoring and quick response [13]–[15]. The alarming point here is that IoT devices are becoming more pervasive each year and are projected to gain more dramatic prevalence in the near future, albeit a slight deceleration due to COVID-19 pandemic [2].

The challenges of 5G and beyond (B5G) cellular communication networks, therefore, are primarily twofold: 1) the bandwidth demand due to more advanced smartphones with more computational capabilities, and the rise in data demanding applications, such as online gaming, augmented reality, etc. [16]; 2) the number of cellular connections that is exponentially growing mainly due to IoT technology. Various solutions have already been proposed in order to combat these issues: network densification and millimetre wave (mm-wave) communications are among the most important candidates for network capacity enhancement [17]. Network densification is a phenomenon, whereby the base station (BS) density in a given environment is increased in order to provide more radio access network (RAN) capacity. This concept mainly uses the idea of frequency reuse, which states that the frequency spectrum of one BS can be reused by other BSs as well only if they avoid interfering with each other. This avoidance is provided by lowering the transmit power in order to reduce the footprints of BSs, so that the overlapping regions are minimized—the less footprint of BSs results in more BSs deployment opportunity, which then leads to more RAN capacity. mm-wave, on the other hand, offers a great enhancement in the RAN capacity of cellular networks by exploiting the abundant bandwidth available in the mm-wave frequency spectrum. Moreover, as antenna sizes reduce with increasing carrier frequency, the use of mm-wave communication enables Multiple-Input Multiple-Output (MIMO) technology, which in turn enhances the reliability and capacity of the network [18]. In other words, the capacity enhancement supplied by mm-wave communications are mainly due

²The road map can be found at the following link. Accessed on 22/11/2020.
https://www.london.gov.uk/sites/default/files/smarter_london_together_v1.66_-_published.pdf.

to two factors: 1) increased bandwidth made available, and 2) MIMO technology [18], [19].

Even though these are sensible and effective methods of enhancing the capacity of cellular networks, a serious side effect immediately emerges: mobility management [16]. The common ground for network densification and mm-wave communication concepts is that both lead to more frequent handovers (HOs), which is defined as the user equipment's (UE's) change of channel, resource, or cell³ association while keeping an ongoing call or session. The underlying reasoning behind this consequence is mainly due the reduction of the footprint of BSs. First, in the case of network densification, the footprint is deliberately reduced with the use of small cells (SCs) in order to facilitate more BS deployments through frequency reuse. Second, concerning the mm-wave communications, the footprint of BSs reduces due to the higher propagation losses incurred at mm-wave frequencies (more dependency on line of sight (LOS)). Furthermore, the increased amount of bandwidth also shortens the range of mm-wave signals [20].

As such, the frequency of HOs grows due to the smaller footprints of BSs: mobile UEs would need to perform more HOs, given that there are now more BSs in a certain environment. Given that the average throughput of a user is a function of the number of HOs with an inverse proportionality [21], this issue has severe consequences in terms of communication quality—degrades the quality-of-service (QoS). Besides, as service interruptions are experienced during HOs, the user satisfaction rates are also affected negatively, undermining the great promises of 5G networks. These adverse effects are mainly caused by two reasons: 1) the number of HO experienced during a call or data transfer session; and 2) the HO cost incurred for each HO experienced. In this regard, the research activities on HO management have predominantly focused on these two aspects, such that minimizing the number of HOs and/or the cost incurred per HO.

Although the figures in terms of the growing number of IoT devices and BSs along with increasing demand for data-oriented applications have been discussed negatively so far, there are some positive impacts as well. The volume of data being generated by cellular networks is also growing considerably, making it a gold mine for network operators to exploit in such a way that more efficient management can be facilitated [16], [22]–[24]. In other words, although growing network sizes result in more complexity, the immense data volume generation becomes a key to alleviate such complexity: this so-called challenge brings its own opportunity and solution. In that regard, machine learning (ML) techniques have gained significant attention in the field of wireless communications, since such amount of data can be very well utilized for training ML models, which could help the networks gain experience and take proactive and more informed actions.

³Cell and BS are interchangeably used throughout this paper unless stated otherwise.

Therefore, in this survey, we focus on the application of ML algorithms to HO management in cellular networking with special attention aimed at 5G and B5G networks in order to keep the discussion timely as 5G has already become a reality, and visionary works about 6G has started to appear in the literature [25], [26], [26]. One of the common grounds of these studies, which try to plot the framework of 6G, is that they all agree that artificial intelligence (AI) and subsequently ML will play a key role in 6G networks as intelligence is expected to lie at the core of 6G networks. Moreover, terahertz (THz) frequencies have been projected to be used in 6G due to the abundance of the bandwidth available in these frequencies [26]. However, this makes the HO management concept even more crucial as THz band includes much higher frequencies than mm-wave band, and therefore the smaller footprint (and more frequent HOs) will be much more significant. To this end, we reviewed the state-of-the-art on ML-based HO management in cellular networks (with a special focus to 5G and B5G) by taking into account the data used during the implementation of such algorithms; a top-level taxonomy on the source of data generation is provided with two primary classes: visual data and wireless network data aided HO optimization. Visual data refers to the data that is captured from the environment of interest in a visual format, such as image and video. The data in visual format is then used to assist the HO process by, for example, detecting objects/blockages affecting signal propagation [27]. Wireless network data, on the other hand, is any kind of data that can be acquired by the wireless network, including received signal strength, channel state information, BS traffic load, neighbouring information, user locations, etc. As such, in addition to reviewing the most recent literature, to the best of the authors' knowledge, this is the unique attempt to survey the visual data assistance in HO management. Furthermore, discussions on HO management of legacy networks, including 3G and 4G, are omitted in this article, since a plethora of works surveying such networks are already available in the literature [5], [28]–[30].

A. OBJECTIVES AND CONTRIBUTIONS

As HO management is deemed as one of the most severe design challenges of 5G and B5G mobile communication networks, in this survey paper, we aimed at highlighting the current status of cellular communication networks as well as forthcoming issues related to HO management. Moreover, provided that ML assisted wireless communications has been projected to be at the heart of network management, this survey focuses primarily on the ML applications to HO management in the next generations of cellular networks. In this regards, the mobility management in 5G networks is thoroughly discussed with a special interest given to HO management in order to reveal the distinctive mobility management policy included in 5G standards, which makes it quite different from the legacy networks. Furthermore, we present the main characteristics of 5G networks, including mm-wave

communications, heterogeneous networking, IoT, vehicular communications, device-to-device communications, and high-speed train communications, that make the HO management even more challenging compared to the legacy networks. One of the most distinctive contributions of this paper is that we provide an outlook in terms of HO management in B5G, especially 6G where THz communications is projected to be a key component. HO management in THz communications is particularly covered in this work, since the transmission range at THz frequency is quite short due to the large-scale molecular absorption loss [31], increasing the likelihood of HOs. Furthermore, as AI is considered to be very instrumental in designing 6G networks [25], [26], [26], the validity of this work spans from 5G to B5G networks. In this regard, to the best of the authors' knowledge, this paper is one of the few attempts discussing the HO management in B5G networks, and with this we intend to produce a timely and novel survey paper that both reveals the current status and mentions futuristic applications/technologies.

After that, ML algorithms are categorized as supervised, unsupervised, and RL and briefly introduced, followed by discussions on ML-based HO management. Through these discussions, we aimed at providing a basic understanding of the generic principles of the most popular ML algorithms as well as how those algorithms can be applied to HO management process in cellular networks. Besides, the state-of-the-art about ML-aided HO management is extensively surveyed by reviewing the most recent studies in order to showcase the current status and opportunities. A top-level taxonomy is followed while reviewing the state-of-the-art, such that the ML-aided HO management methodologies are classified based on the source of the data they utilize. As such, two broad categories are encompassed: visual data based and wireless network data based HO management techniques. With this novel taxonomy, the major objective is to recognize the visual data aided HO management schemes—which has been long overlooked in the literature—by giving it a special place along with the traditional network data driven HO schemes. On the other hand, for the network-data based HO management, the most recent works are extensively reviewed under certain use cases: beam selection and BS selection. In addition, we also briefly discuss how intelligent HO schemes can help in emergency situations in the case of mobile clinics, ambulances, and remote hospitals, which could be also beneficial for pandemic scenarios, such as the current COVID-19 pandemic.

Another objective of this paper is to present the grand challenges for the application of ML algorithms to HO management, through which we aimed to address the current and future requirements of such implementations, and to identify possible research directions in order to make the ML integration to HO management in 5G and B5G more efficient, effective, and feasible. Therefore, with this section, we try to canalize the research focus to the identified topics to open a road for practical solutions.

B. RELATED WORKS

ML applications to self-organizing cellular networks were surveyed in [32], in which the authors provided in-depth coverage on the ML algorithms along with the characteristics of self-organizing networks. The authors presented the ML applications to cellular networks under the categories of the major functionalities in self-organizing networks: self-configuration, self-optimization, and self-healing. Various use-cases under each of the aforementioned functionality were provided, leading to a comprehensive picture of ML implementations to cellular networks. Thus, while this work focused on ML implementations and included a brief discussion on HO management, it did not primarily focus on HO management in 5G networks; instead, it drew a comprehensive framework for ML applications to multi-domain cellular networks, such as radio resource management, anomaly detection, backhauling, etc.

The work in [33] focused on the use-cases of mobility predictions, and provided an extensive review on the characteristics of mobility predictions (e.g., mobility predictability, user location, prediction output, and performance metrics) along with the methods of mobility prediction. Even though the review was not meant for ML alone, the methods covered are predominantly ML algorithms, hence it can be classified as focusing on ML. However, the scope of the work is not limited to HO management—albeit being included as one of the use-cases—and visual data driven HO management is mostly ignored. The authors in [28] presented a brief survey on HO management in 5G and B5G networks, where they provided a background on 5G networks along with some enabling technologies, such as mm-wave communications, heterogeneous networks (HetNets), software-defined networking, and ML. Although the main focus of their work is HO management in the next generation of cellular networks, the authors reviewed the literature without extensive discussions. ML implementations were included in general, and the authors failed to demonstrate an in-depth analysis of how ML can be incorporated in HO management in 5G and B5G networks. In addition, the visual data aided HO management and HO in emergency scenarios were completely overlooked in their work.

A very comprehensive and detailed survey was given in [5], in which HO management was elaborated for both long-term evaluation (LTE) and 5G networks with comparative discussions. HO procedures in both LTE and 5G were presented step-by-step, and HO types were covered in a detailed manner. The literature was also reviewed without any particular attention to ML algorithms; as such, even though some ML applications were mentioned while reviewing the state-of-the-art methods, the scope of the paper was solely HO management, not the ML applications to HO management. Similarly, an extensive review of mobility management in ultra-dense networks was given in [29]. In particular, the authors included a meticulous tutorial on the mobility management in cellular networks, followed by discussions

on proactive mobility management in the next generations of cellular networks, which comprises of a brief introduction to various ML techniques. Furthermore, the authors included an analysis of AI assistance in mobility management, where they mainly reviewed the literature by identifying the employed AI methods and use-cases. This work seems to be one of the most overlapping survey papers with our present work; however, the focus of the survey in [29] is broader as it tries to encompass every single issue in mobility management. In our present work, on the other hand, the scope is kept limited to HO management in order to make the review more comprehensive in terms of HO management. Moreover, ML application is not the main focus in [29], whereas, in our present work, we try to exclusively analyse the integration of ML to HO management in cellular networks by discussing ML-oriented opportunities as well as challenges. Besides, visual data based HO management parts covered in our present work constitutes one of the most important novelties and contributions in this present work, as it is not available in [29] or any other mobility or HO management based survey paper in the literature.

Another comprehensive survey was conducted in [34] for mobility management in 5G HetNets. In particular, the authors provided a detailed tutorial on radio resource control (RRC) states included in 5G NR along with the initial access and reachability. RRC protocol is essential for cellular communication networks, and it performs several key functionalities including connection establishment/release, configuration/establishment/release of radio bearer (RB), broadcasting of system information, etc. This topic is more elaborated in Section III-A. Connected mode mobility (i.e., HO) with various types of HOs were also elaborated, followed by beam level mobility management issues. The ML implementations were not the main focus of their work, and thus the scope was primarily kept on mobility management. As such, since ML was not the target, the source of data generation (visual data and wireless network data) were also not discussed. The authors in [35] analysed femtocell HOs in HetNets and provided a detailed background on LTE HO procedure with a particular interest to femtocell HOs. After identifying some challenges with the HO decision process in two-tier networks, an inclusive review was conducted on the existing HO decision techniques. Although the paper was meant for 5G networks, the main story was originated around LTE networks, as there was no particular discussion on the mobility management in 5G networks. In addition, the scope was quite limited, as only HO decision techniques were discussed, and even though some of the cited literature included ML implementations, ML was not the main consideration.

A succinct survey on HO-oriented mobility management was conducted in [30], where the authors provided very generic discussions on mobility and HO mechanisms in HetNets. In particular, mobility management was divided into location management and HO management, and each group

was elaborated subsequently. However, 5G or B5G cellular communication networks were not mentioned, and no special HO management scheme, such as ML-based HO management, was provided. Therefore, our present survey on intelligent HO management is more advanced compared to the one in [30] in terms of style, the methodology being followed, and comprehensiveness. Another brief survey on mobility management in 5G networks was given in [36]. An overview of the generations of cellular networks from 1G to 5G was first presented, followed by 5G structure and mobility management related discussions in 5G networks. HO management was also reviewed by introducing different types of HOs as well as HO parameters. However, the discussions were kept very short, and an in-depth coverage was not provided for 5G networks or mobility management. Furthermore, the author did not intend to make the survey around the ML applications to HO management.

The authors in [37] produced an extensive survey on mobility management by questioning the readiness of the state-of-the-art solutions for the next-generations of cellular networks, namely 5G and B5G. First, the requirements of the next generations of cellular networks in terms of mobility management were first identified, followed by an introduction of their own qualitative performance metrics for the existing mobility management solutions. Moreover, a discussion on the effectiveness and sufficiency of the standards for both legacy networks and 5G as well as the research activities for meeting these requirements was included in their work. Lastly, potential enabling technologies and existing challenges were reviewed in detail. Compared to our present survey; i) the authors did not focus only on HO management, ii) ML applications were not mainly iterated although a mild discussion on deep learning was included, and iii) visual data assistance in HO management in addition to HO management in emergency scenarios were not covered.

A tabular overview of the relevant survey papers on mobility and HO management is given in Table 1, where the included works are analyzed in terms of their focus on 5G and B5G networks, HO management, ML applications to HO management, the use of visual data for HO management, and HO management in emergency scenarios.

C. PAPER ORGANIZATION

The remainder of the paper is structured as follows: the basics of 5G networks along with mobility management-oriented characteristics of 5G networks, including HetNets, IoT, vehicular communications, device-to-device communications, and high-speed train communications, are presented in Section II, while Section III provides an inclusive discussion on the mobility management in 5G. Section IV provides a comprehensive tutorial on HO management in 5G networks by detailing the HO types, requirements and performance metrics, and radio resource management. ML applications to HO management is elaborated in Section V with a brief introduction to different branches of ML (namely, supervised, unsupervised, and RL), followed by an extension literature

review on the state-of-the-art in ML-based HO management techniques. Section VI highlights the challenges which ML-assisted HO management schemes would confront, and identifies future research directions. Lastly, Section VII concludes the paper with concluding remarks.

NOMENCLATURE

| | |
|---------|------------------------------------------------------------------------------------------|
| 3GPP | 3rd Generation Partnership Project |
| 5G | Fifth Generation |
| A2C | Advantage Actor-Critic |
| AI | Artificial Intelligence |
| AMF | Access Mobility Function |
| ANN | Artificial Neural Network |
| AS | Access Stratum |
| AUSF | Authentication Server Function |
| B5G | Beyond 5G |
| BBU | Baseband Unit |
| BS | Base Station |
| CCTV | Closed-Circuit Television |
| CMAB | Contextual Multi-Armed Bandit |
| CMAS | Commercial Mobile Alert System |
| CN | Core Network |
| CNN | Convolution Neural Networks |
| CP | Control Plane |
| CQI | Channel Quality Information |
| CSI | Channel State Information |
| D2D | Device-to-Device |
| DRX | Discontinuous Reception |
| DSRC | Dedicated Short-Range Communication |
| E-UTRA | Evolved Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access |
| EM | Expectation-Maximization |
| EPC | Evolved Packet Core |
| ETWS | Earthquake and Tsunami Warning Service |
| FDD | Frequency Division Duplex |
| HetNets | Heterogeneous Network |
| HO | Handover |
| HOO | Hierarchical Optimistic Optimization |
| HST | High Speed Train |
| ICA | Independent Component Analysis |
| IoT | Internet-of-Things |
| IP | Internet Protocol |
| KNN | K -Nearest Neighbour |
| KPI | Key performance indicators |
| LIDAR | Light Detection and Ranging |
| LOS | Line-of-Sight |
| LTE | Long Term Evaluation |
| M2M | Machine-to-Machine |
| MIMO | Multiple-Input Multiple-Output |
| mm-wave | Millimetre Wave |
| MME | Mobility Management Entity |
| mMTC | Massive Machine-type Communications |
| MR | Measurement report |
| NAS | Non-Access Stratum |
| NEF | Network Exposure Function |

| | |
|---------|-------------------------------------------|
| NextGen | Next Generation |
| NR | New Radio |
| NRF | NR Repository Function |
| NSI | Network Slice Instance |
| NSSF | Network Slice Selection Function |
| PCA | Principal Component Analysis |
| PCF | Policy Control Function |
| PCI | Physical Cell Identifier |
| PPO | Proximal Policy Optimization |
| QoS | Quality-of-Service |
| RAI | RAN Area Identifier |
| RAN | Radio Access Network |
| RAT | Radio Access Technology |
| RB | Radio Bearer |
| RNA | RAN-based Notification Area |
| RRC | Radio Resource Control |
| SCs | Small Cells |
| SGW | Serving Gateway |
| SMF | Session Management Function |
| SSB | Synchronization Signal Block |
| SVM | Support Vector Machine |
| TAI | Tracking Area Identifier |
| TDD | Time Division Duplex |
| UCB | Upper Confidence Bound |
| UDM | Unified Data Management |
| UDR | Unified Data Repository |
| UDSF | Unstructured Data Storage Function |
| UE | User Equipment |
| UP | User Plane |
| UPF | User Plane Function |
| URLLC | Ultra-Reliable Low-Latency Communications |
| V2X | Vehicle-to-everything |
| XGBOOST | Extreme Gradient Boosting |

II. CHARACTERISTICS OF 5G AND BEYOND: SOME GENERAL CONCEPTS

This section presents general concepts behind 5G and B5G system in a cellular network. A short review of the architecture, channel characteristics and various features and applications of 5G are presented.

A. 5G SYSTEM

Although this survey focuses on HO management in NR, it is useful to provide a brief overview of 5G’s architecture, interfaces and connections to serve as a background. The Next generation (NextGen) architecture is based on network function (NF) instead of a network entity (NE) that is obtained in LTE, according to 3GPP specification for LTE and new 5G systems [38]–[40]. In LTE’s core network (CN) also known as evolved packet core (EPC), the appropriate network protocols and interfaces are defined among the entities for each network entity (e.g. serving gateway (SGW) and the mobility management entity (MME)). In contrast, network protocols and interfaces in 5G CN (5GC) are specified for each NF. The NF is the processing functionality in 5G networks, and it can

be implemented in three ways [39]: 1) as a network element on dedicated hardware; 2) as a software instance running on dedicated hardware; or 3) as a virtualized function built on an appropriate platform, such as a cloud infrastructure. The advantage of NF over NE is that it dramatically decreases latency. This is achieved by carefully controlling the UE mobility (e.g. tracking and paging procedures) scheme and separating the user plane (UP) (also known as data plane is the dedicated channel that carries the network user traffic) from the control plane (CP) (which is responsible for routing data traffic through the network and for carrying out other control activities) to ensure that each plane’s resources are independently scaled and that more NF can be deployed in a distributed manner [41]. Fig. 1 shows the 5G system architecture along with NFs and reference points. A reference point shows the interaction between the services in two NFs (e.g. N4 is the reference point that connects UPF and SMF). The NF in the UP consists of user plane function (UPF) acting as a gateway for the UE traffic passing through RAN to external networks such as the Internet. It is responsible for packet routing and forwarding, packet inspection, QoS handling, packet filtering, and traffic measurement. Several components of NFs run in the CP. Some of the components are: access mobility function (AMF), session management function (SMF), network slice selection function (NSSF), unified data management (UDM), policy control function (PCF), authentication server function (AUSF). For further information on these functions, the reader is referred to [41]–[43].

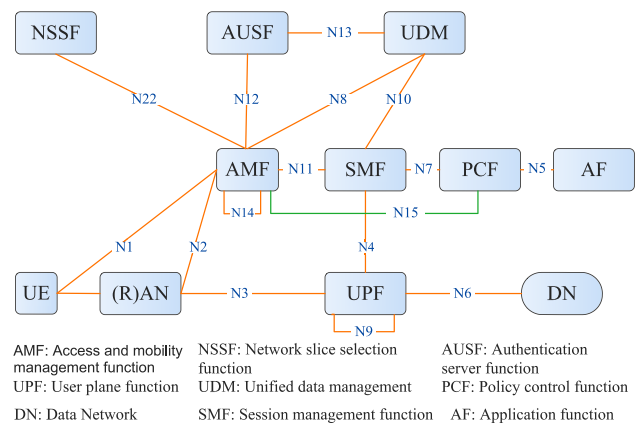


FIGURE 1. The 3GPP-5G architecture with reference points adapted from [41].

Overall, 5G architecture is divided into two parts, as shown in Fig. 2. The first part is the CN whose components have just been discussed while the second part is NextGen Radio Access Network (NG-RAN). The NextGen NodeB (gNB) serves as the access point for the 5G network, transmitting CP and UP traffic originating from N1, N2, N3 reference interfaces as shown in Fig 1. The purpose of the ng-eNB is to provide Evolved Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access (E-UTRA) UP and CP protocol terminations for UEs. In addition, 5G technology

TABLE 1. An overview of the related survey papers on mobility and HO management.

| Paper | Year | 5G+ Focus | HO Focus | ML Focus | Visual-data | HO in Emergency | Key Points |
|-------|------|-----------|----------|----------|-------------|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [32] | 2017 | ✓ | ✗ | ✓ | ✗ | ✗ | <ul style="list-style-type: none"> • ML applications to SON • Not focused on HO |
| [33] | 2019 | ✓ | ✗ | ✓ | ✗ | ✗ | <ul style="list-style-type: none"> • Detailed on mobility prediction • Various use-cases |
| [28] | 2020 | ✓ | ✓ | ✓ | ✗ | ✗ | <ul style="list-style-type: none"> • HO management in 5G+ • No in-depth coverage on ML |
| [5] | 2019 | ✓ | ✓ | ✗ | ✗ | ✗ | <ul style="list-style-type: none"> • Detailed HO for LTE and 5G • Not focused on ML |
| [29] | 2020 | ✓ | ✓ | ✓ | ✗ | ✗ | <ul style="list-style-type: none"> • Mobility in ultra-dense networks • Proactive mobility management • AI assistance |
| [34] | 2020 | ✓ | ✓ | ✗ | ✗ | ✗ | <ul style="list-style-type: none"> • Mobility in 5G HetNets • Tutorials on RRC states and HO • Not focused on ML |
| [35] | 2020 | ✗ | ✓ | ✗ | ✗ | ✗ | <ul style="list-style-type: none"> • Femtocell HOs in HetNets • HO decision techniques • Not focused on ML |
| [30] | 2019 | ✗ | ✓ | ✗ | ✗ | ✗ | <ul style="list-style-type: none"> • A generic tutorial on HO • Not focused on 5G and ML |
| [36] | 2020 | ✓ | ✓ | ✗ | ✗ | ✗ | <ul style="list-style-type: none"> • Mobility management in 5G • No in-depth coverage • ML was not scoped |
| [37] | 2020 | ✓ | ✓ | ✓ | ✗ | ✗ | <ul style="list-style-type: none"> • Extensive in mobility management • Questions the state-of-the-art • ML was not mainly focused |

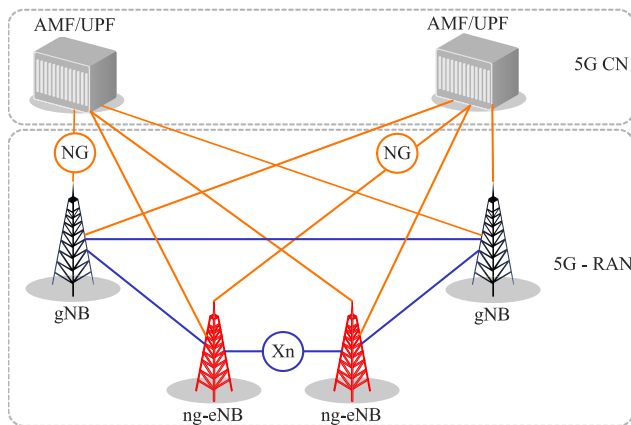


FIGURE 2. Overall architecture of 5G system showing network elements and interfaces based on the logical CN/RAN split, as in [38].

also supports LTE via ng-eNB. It allows existing 4G radio networks to coexist with the gNB. For example, if both LTE and 5G radio coverage are available, a 5G UE may use either LTE and 5G radio resources. Therefore, when there is no

5G coverage, LTE serves the 5G UE using the ng-eNB. The connection interface between gNB and ng-eNB is known as an Xn interface, and NG interface is the connection interface between gNB/ng-eNB and CN more specifically to the UPF the NG user-plane part (NG-U) and to the AMF the NG control-plane part (NG-C). The last interface that needs to be mentioned is the radio frequency interface, which is the circuit between the UE and the active gNB or ng-eNB which is also known as Uu interface. This interface supports a broad spectrum from low to high frequencies [44].

B. CHANNEL CHARACTERISTICS OF 5G WIRELESS SYSTEMS

As mentioned earlier, 5G systems use mm-wave frequencies, along with sub-1 and 6 GHz spectrum. It is envisaged that B5G networks will use THz frequencies [45], [46]. Compared to the sub-6GHz band, the advantages of the mm-wave band include more available bandwidth and use of small antennas in devices. Antenna size is inversely proportional to frequency; therefore, mm-wave antennas for UE and BS are small and can be placed in small devices. However,

the mm-wave band has some drawbacks that necessitate the use of sub-6 GHz frequencies in 5G. In this subsection, we present the rationale for the co-existence of multiband frequencies in 5G, as well as the characteristics and applications of different spectrum bands from sub-1 GHz to mm-wave.

1) SUB-1 GHz AND SUB-6 GHz IN 5G

In its early phases of implementation, 5G's main spectrum options were around 3.5 GHz and 4.5 GHz for sub-6 GHz with time division duplexing (TDD) technology. For the 3.5/4.5 GHz band, 5G aims to use existing BSs to help in the roll-out and implementation [47]. The 3.5 GHz band provides comparatively less coverage than the 2 GHz band used in legacy networks, and this is because radio propagation decreases as frequency increases. However, introducing MIMO beam-forming antennas at 3.5 GHz and higher spectrum reduces propagation losses, thereby significantly increasing coverage for 3.5/4.5 GHz.

The sub-1 GHz bands are also used through frequency division duplex (FDD) in 5G, especially for deep indoor penetration [48]. With its broader coverage, low data rate IoT connectivity and other critical communication like remote control or automotive applications can be introduced. Therefore, extensive coverage becomes imperative for these new use cases which can be served by the sub-1 GHz band [48]–[50].

2) mm-WAVE IN 5G

The propagation of waves at mm-wave is more prone to adverse effects of obstacles which can be caused by movement of people, presence of trees, foliage (outdoor scenario), furniture and walls (indoor scenario). Since the mm-wave spectrum is severely affected by rain and other atmospheric conditions, previous studies suggested that it was impractical to use this frequency band for mobile communications. However, this has been proven to be wrong, as recent studies have shown that atmospheric absorption does not create a significant loss when used in picocells—coverage below 200 m from transmitter— [51], [52]. These studies also show that even under very extreme rainfall, the rain attenuation would cause 1.4 dB and 2 dB loss at 28 GHz and 73 GHz, respectively. The impact of rain attenuation on mm-wave propagation, especially in urban picocell areas, will therefore become insignificant [52]. The short-range coverage of mm-wave has both advantages and disadvantages. Spatial reuse of frequency band, strong multi-path behaviour due to reflection are among the advantages of using mm-wave while one of the disadvantages of using this band is that many SCs are required to provide coverage due to the high propagation loss of mm-wave.

mm-wave is an inherently directional wave which means that there is a need for the transmitter and receiver to focus the beam towards each other, this is commonly known as beam steering. The main advantage of beam steering is to achieve high gain by focusing the transmitter and receiver towards each other. The beam steering is completed through a

beam training/tracking process. Beam training is a process of finding the desired beam to connect the UEs in order to reduce initial access delay. Another critical parameter to consider is sensitivity to blockage. mm-wave has a higher frequency, making the size of its wavelength small compared to many physical objects, and thus the low ability of mm-wave to diffract through large objects makes it sensitive to blockage. For example, at the 60 GHz band, it is observed that there is a 20-35 dB increase in the path-loss if an obstacle (e.g., humans or furniture) is introduced between the mm-wave link [52].

While it has been demonstrated that using mm-wave frequencies such as 28 GHz and 38 GHz is possible even in complex urban environments, many challenges such as low throughput and high signaling overheads associated with HO still needs to be addressed to realize the full potentials of the mm-wave band [52], [53].

3) CO-EXISTENCE OF SUB-1 GHz, SUB-6 GHz AND mm-WAVE

Given the rigid transmission efficiency standard for certain use cases such as vehicular networks, the use of mm-wave poses some significant difficulties in implementing reliable but high data rate communication. Critical IoT applications, including remote healthcare systems (for clinical remote monitoring and assisted living), traffic and industrial control (drone/robot/vehicle), and tactile Internet, etc., require higher availability, higher reliability, safety, and lower latency to ensure end-user experience as failure to satisfy these requirements would result in severe consequences, such as vehicle collision, and accident [15].

A CP and UP decoupled network is designed to circumvent these challenges by using the sub-6 GHz for the CP and mm-wave frequencies for the UP [54]. This guarantees that signaling from the CP reaches the UE with high reliability by using the sub-6GHz spectrum. On the other hand, the use of mm-wave frequencies for UP provides unprecedented data speeds, due to the vast bandwidth availability at the mm-wave spectrum. Therefore, the main purpose of sub-6 GHz and sub-1 GHz bands is to provide uninterrupted access to the CP or to provide coverage for areas where mm-wave cannot offer adequate coverage.

C. HETEROGENEOUS NETWORKS

HetNets comprises the deployment of BSs with different sizes. They are exciting low-cost approaches to meet the industry's growth requirements and offer a consistent connectivity experience. HetNet comprises SCs that support aggressive spectrum spatial reuse coexisting within macrocells, as shown in Fig. 3. A macrocell is a BS used in cellular networks with the function of providing radio coverage to a large area of mobile network access. The macrocell overlaps several SCs, and it has high output power, usually in the range of tens of watts and can provide coverage to a large area. However, the macrocell suffers from interference caused by the use of sub-6 GHz, which can travel far by nature. While the macrocell transmits radio waves over a long distance,

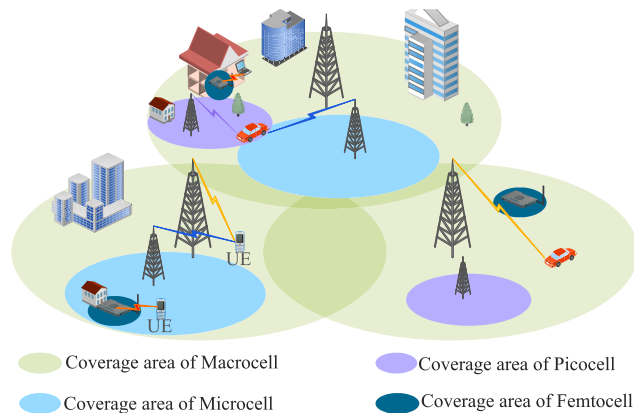


FIGURE 3. An illustration of heterogeneous networks –the macrocells are overlaid by three tiers of low-power BSs which are microcells, picocells and femtocells in the same geographic area.

if not managed properly, signal interference with other cells is very likely, which in turn could result in the degradation of network performance [55]. Nevertheless, macrocell has low spectral efficiency or area spectral efficiency, typically measured in (bit/s/Hz) per unit area, which results in less bandwidth and low data rate per UE. The data rate is the function of bandwidth, and SCs allows frequency reuse due to limited range hence more bandwidth and data rate per UE. Therefore, to increase the data rate, the idea of reducing BS footprint for macrocells was introduced [56].

In order to address the challenges facing the macrocell, the easiest and simplest way was to get the transmitter and receiver closer to each other. This solution creates a dual-benefit of high-quality links and more spatial reuse. Different cells with different sizes are considered based on the transmitted power, and the frequency of transmission used: macrocell, microcell, picocell, and femtocell are popular cells created by gradually reducing coverage range and transmission power. A summary of the types of cells in terms of coverage and capacity is presented in Table 2.

TABLE 2. Cell types in wireless networks in term of coverage and capacity.

| Cell type | Coverage range(meter) | Capacity |
|-----------|-----------------------|-------------|
| Femtocell | 10-20 | A few UEs |
| Picocell | 200 | 20 - 40 UEs |
| Microcell | 2000 | > 100 UEs |
| Macrocell | $(30-35) \times 10^3$ | Many UEs |

As the BS footprint becomes smaller with smaller BSs, the use of mm-wave become more feasible. The mm-wave frequency suffers from high penetration loss which brings the advantage of enabling the reuse of mm-wave frequency in indoor environment for femtocell. For further studies on HetNets please refer to [55], [57]–[59].

D. INTERNET OF THINGS

In this modern era, various applications used by billions of people are daily made available via the internet, thereby

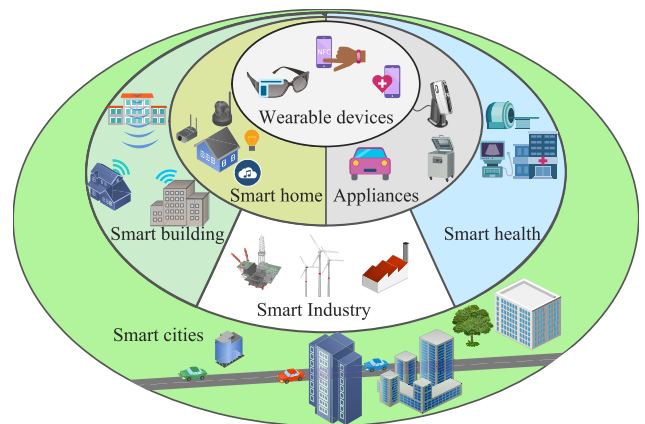


FIGURE 4. Sensors and IoT use case.

making the Internet an essential tool to interconnect these applications, among which services like video streaming, file sharing, electronic commerce, etc. are increasingly taking place online. The types of interconnected devices includes smart phones and IoT devices such as sensors, wearables, etc. These IoT devices are able to communicate with each other to share information with little or no human involvement. Fig 4 illustrates some common IoT uses cases. As the number of IoT devices keep increasing, the traffic generated by these devices also increases, hence, the underlying protocols that support IoT should be reconsidered to support the massive interconnection of both new and conventional devices [60], [61]. Conventional devices need to be made smarter by incorporating advanced technologies such as ubiquitous computing, artificial intelligence, embedded devices, different communication standards and technologies, various application services, and different Internet standards. However, the problem is that these devices used in IoT are memory-limited and energy-limited, so information should be routed efficiently, and the proper channel between source and sink should be carefully chosen [61]. IoT has different use cases such as smart cities, smart home, vehicular sensors, health monitoring, and sport & leisure scenarios. Several of these use cases are discussed in the following subsection, while focusing on the differences in application domains requirements.

1) SMART CITIES

This involves the use of smart technologies to provide relevant information and automated services that would improve the standard of living of the people in a particular area. These smart technologies include: deployment of sensors for traffic monitoring to prevent traffic jams and detect bad roads, for automated street lights, smart grid, waste management systems where environmental sensors deployed in various locations to detect pollution, water level, or fire. In these cases, the early detection of abnormal environmental situations can be used to alert the appropriate authorities in order

to enable them take the necessary actions when any incident occurs etc [15].

2) SMART HOME

This use case is sometimes classified as a part of smart cities. However, it is mostly limited to user-oriented applications, particularly for home networks [15]. Different services that can be classified under the smart home use-case include: (1) Connected home appliances where appliances such as smart fridges can be used to automatically order for the restocking of the fridge with food items or beverages when it detects that it is running out of supplies by checking a pre-defined threshold for the amount of each item. (2) Home video monitoring Homes can be equipped with small cameras that are mounted in different locations, and can be used to stream the video to the Internet for a remote monitoring. They can also be used to send alarms upon the detection of unusual movement or abnormal behavior, smoke, carbon monoxide, etc. in the monitored area.

3) HEALTHCARE/TELEMEDICINE/WEARABLE

This use case is becoming more popular as more devices such as watches and other wearable devices become increasingly available. Patients do not necessarily need to be monitored manually, but smart wearable devices track their health conditions for any abnormality. Such devices send an alarm message to a nearby hospital as soon as they detect anomalies with the patients being monitored.

All the use cases mentioned above face challenges that need to be addressed before IoT can become very efficient and able to integrate heterogeneous devices—device with different communication standards (protocols, technologies and hardware)—and applications envisaged for 5G [62]. These challenges include scalability, network management, security and privacy, interoperability and heterogeneity, network congestion and overload, and network mobility and coverage. To interconnect a massive number of devices and accommodate enormous traffic generated within 5G system, conventional sub-6 GHz is no longer sufficient, hence the need for the utilization of a new frequency band (mm-wave) [60]. This would lead enhanced QoS for IoT devices.

E. DEVICE-TO-DEVICE COMMUNICATION

Device-to-device (D2D) communication involves the direct communication between two devices without passing through a BS. These devices could be smartphones, vehicles, etc. This kind of communication usually occurs when both device are in close proximity to each other [63]. The introduction of D2D communication is necessary to cope with the rise in the number of devices as well as the increase in demand for high speed connections. It is one of the technologies that is being exploited in 5G and B5G networks as its use would lead to enhanced link reliability, spectral efficiency, system capacity, energy efficiency and reduced network delays [64]. The use of mm-wave in 5G would facilitate D2D communications as more direct links would be supported, thereby enhancing the

capacity of the network. In addition, due to the directional nature of 5G antennas, it would be possible to support more simultaneous connections in mm-wave systems. Despite the inherent advantages of D2D communications, due to UE mobility, and the fact that the UEs still need to connect to the BS in order to transmit control signals, the issue of HO needs to be carefully considered in order to prevent ping-pong effects which results in frequent HOs [65], [66].

F. VEHICULAR COMMUNICATIONS

Vehicular-to-everything (V2X) is a special case of D2D communication. It is a technology that provides communication between vehicles and surrounding devices, including hand-held devices, moving/stationary cars, and all other external IoT appliances. V2X is categorized into two main components: vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I). The former allows communication between two or more vehicles. On the other hand, the latter deals with communication between cars and other devices in its external environment, such as traffic/street lights [67]. The most common and popular communication protocol that supports vehicular networks is the dedicated short-range communication (DSRC), and it can support all V2X architecture. DSRC uses 75 MHz bandwidth at 5.9 GHz band, and was expected to provide the data rate up to 27 Mb/s and a transmission range up to 1000 m [67], [68]. As a result of the high mobility of vehicles, one of the major challenges that vehicular networks suffer from is HO. This occurs because in the course of the vehicles movement from one location to another, the often move out of the coverage area of one network also known as road side unit (RSU) to another thereby leading to frequent change of connection from one RSU to another. This issue would become more pronounced with the use of mm-wave in 5G as the coverage area of the RSUs would become smaller [69]. Hence, HO management must be carefully considered for fast moving vehicles in 5G mm-wave communication networks in order to ensure seamless HO.

G. HIGH SPEED TRAIN COMMUNICATION

High speed train (HST) communications is one of the verticals that would be supported by 5G networks. The availability of large spectrum in the mm-wave frequency band would make the provision of enhanced mobile broadband services possible for passengers in high speed trains [49]. HST communication networks basically encompasses two kinds of communications, namely: critical and non-critical communications. The former is the communication between the HST and its associated infrastructures and is necessary to control the speed and ensure the safety, reliability and smooth functioning of the HST. The latter is required to provide services to the passengers on-board such as high quality video, and other data services [70]. Even though mm-wave has great potentials for application in HST communications, due to the high mobility of trains, HST communication is often prone to frequent HOs and fast fading channels, that potentially undermines its availability. As a result, some new

technologies such as hybrid beamforming, beam management, network slicing, and distributed antenna system have been introduced in mm-wave communications to enhance its application in HST communications [50].

H. BEYOND 5G SYSTEM

5G is game-changer as it can provide data rates up to tens of gigabits per second, which is far beyond what is provided by legacy networks [71]. However, with the introduction of new use cases and applications such as virtual and augmented reality, remote surgery and holographic projection, 5G would not be able to meet the projected explosion in wireless data demands. As a result, research into higher frequency (beyond mm-wave) has risen, and THz frequency has become the B5G researcher’s focus as the new spectrum for B5G systems. Only frequency bands in the THz range can provide the large amount of bandwidth that is needed to support the terabit-per-second data rates in order to support huge traffic types such as uncompressed videos that is envisioned in B5G networks [46], [72].

The use of THz band in 6G is required to provide the reliable communication that is required to support various critical applications, accommodate high data rates per area, and support massive amounts of connected UEs. The THz frequency band has quite similar characteristics to that of mm-wave. However, because it has a higher frequency compared to mm-wave, this means that it would be prone to all the challenges facing mm-wave alongside additional challenges. Therefore, there is a need for more advanced error control mechanisms, mobility management techniques, as well as other new features to enable the utilization of this frequency band in mobile cellular networks.

III. MOBILITY MANAGEMENT IN 5G

Mobility management in 5G is quite different from that of legacy networks (2G-4G) and in this section, we present the concepts behind the radio access mobility in 5G cellular network. We also briefly explain the mobility state procedures in 5G system that makes it more efficient than legacy systems.

Definition 1 (Access Stratum): Access stratum (AS) is the set of protocols in 5G that contains the functionality associated with the UE’s access to the RAN and the control of active connections between a UE and the RAN.

Definition 2 (Non-Access Stratum): Non-access stratum is the set of protocols in 5G that handles functionality operating between UE and CN.

Definition 3 (RRC Context): The RRC context are the parameters necessary for establishing/maintaining communication between the UE and the CN.

Definition 4 (Cell Selection): Cell selection is the process of choosing a suitable cell⁴ for the UE to camp on. This process is performed as soon as the UE is switched on [74].

⁴A cell with the measured cell attributes satisfy the cell selection criteria [73]

Definition 5 (Cell Re-Selection): Cell re-selection is the process of choosing a suitable cell after the UE camps on a cell and stays in the idle or inactive state.

A. RADIO RESOURCE CONTROL STATE MACHINE

The RRC protocol is in the IP-level (Layer 3 /Network Layer) and is the protocol between UE and NG-RAN as specified by 3GPP TS 38.331 [75]. The RRC protocol’s essential functions include; 1) broadcast of system information; 2) Control of the RRC connection—this procedure includes paging, establishment, modification and release of the RB. It also involves establishing an RRC context; 3) measurement configuration and reporting, and other functions specified by 3GPP TS 38.331 that can be summarized in [43], [75]. The RRC’s operation is guided by a state machine that defines specific states where a UE may be present. The different states in this state machine have different amounts of radio resources that can be utilized by the UE once it enters into a particular state. Since different amounts of resources are available in different states, the state machine impacts the QoS that the user experiences and the energy consumption of the UE [76]. In addition, RRC states provide a clear distinction between HO and cell (re-) selection. The UE can be in one of the three RRC states, namely: RRC_Idle, RRC_Connected, and RRC Inactive state. Fig. 5 depicts the UE state machine and state transitions in 5G while Table 3 summarizes the RRC protocols and functions in each RRC state.

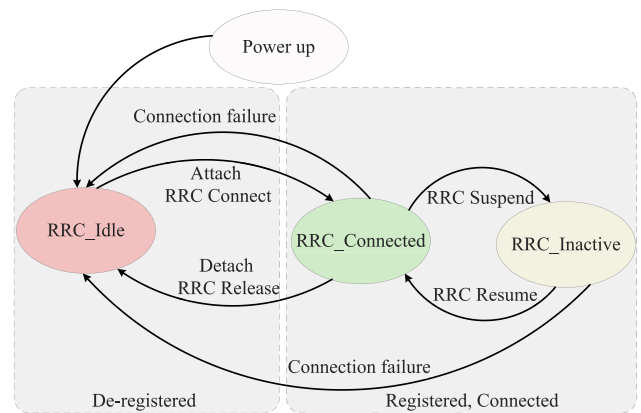


FIGURE 5. UE state machine and state transitions in 5G [75].

1) RRC_IDLE

In RRC_Idle state, the UE is not registered to a particular cell; hence, it does not have an AS context or receive any network information. This means that no specific link is established for communication between the UE and CN, and the UE does not belong to any specific cell. From the CN perspective, the UE is in the CN_Idle state,⁵ and the UE is in

⁵UE is said to be in CN_Idle state from CN perspective when no connection is established between UE and the CN [43], [77].

TABLE 3. Summary of the RRC state and mobility handling in 5G.

| RRC Protocol | RRC State | | |
|---------------------------------------|-----------|----------|-----------|
| | Idle | Inactive | Connected |
| Network selection/registration | ✓ | | |
| Cell re-selection | ✓ | ✓ | |
| 5GC based Paging | ✓ | | |
| NG-RAN based Paging | | ✓ | |
| 5G-RAN manages the UE RNA | | ✓ | |
| 5G-RAN knows UE serving cell | | | ✓ |
| Keep 5GC/5G-RAN connection for UE | | ✓ | ✓ |
| UE AS context stored in 5G-RAN and UE | | ✓ | ✓ |

(a kind of) sleep mode and wakes up periodically (according to a configured discontinuous reception (DRX cycle)) to listen for paging messages from the network through the downlink channel. During this period, no data transfer takes place and the UE enters into sleep mode regularly to reduce battery consumption. The network can reach the UEs in the RRC_Idle state by sending paging messages to notify them of changes in system information, warning messages such as earthquake and tsunami warning service (ETWS), and commercial mobile alert system (CMAS) which are sent as short messages. In this state, the UE manages mobility based on the network configurations via cell re-selections. It also performs the neighbouring cell measurements needed for cell re-selection in order to determine which cell it is to connect (explained in Section III-B). To reduce the network signaling overhead and the latency experienced by legacy networks (such as LTE) during the transition to RRC_Connected state, the RRC_Inactive state was introduced in 5G. In 5G, the network initiates the RRC release procedure to transit a UE from the connected to the idle state. In addition, as UE moves from the idle to the connected state, both the UE and the network establish the RRC context.

2) RRC_INACTIVE

5G-NR introduced RRC_Inactive state from lessons learned during the development of LTE. The findings revealed that the transition of wireless devices from idle state to connected state is the most frequent high-layer signaling event in existing LTE networks, occurring about 500 – 1,000 times a day.⁶ This transition involves a significant amount of signaling overhead between the UE and the network, as well as between network nodes, which can lead to increased latency and power consumption in the UE. The solution is to switch to RRC_Inactive state which will result in a significant reduction in both latency and UE battery consumption. When the UE is in inactive state, its behaviour is similar to that in idle mode in terms of power-saving. However, unlike the idle state,

⁶Meeting 5G latency requirements with inactive state, Published on 19 June, 2019. Available online at <https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/meeting-5g-latency-requirements-with-inactive-state>. Accessed on 25 Nov. 2020.

in the inactive state, RRC context is kept in both UE and gNB, and the UE is in CN_Connected state⁷ from the CN perspective, meaning that its connection to the CN is kept intact. Different from RRC_Idle state, the primary purpose of RRC_Inactive state is to reduce the network signaling load and latency involved during RRC_Idle to RRC_Connected state transition. In RRC_Inactive state, the network signaling becomes faster since the AS context is stored in both the UE and gNB. While 5G CN connection is still retained - (UE remains in CN_Connected state), the UE in RRC_Inactive state is in sleep mode and wakes up repeatedly- according to configured DRX cycle (which in this case is controlled by the 5G-RAN), and regularly monitors for paging messages from the network. The procedure for notifying the UEs about any change of system information or warning message is the same as that of the idle state [78].

3) RRC_CONNECTED

In the RRC_Connected state, the RRC context and all parameters needed to establish communication between the UE and the RAN are known to all entities. This means that in RRC_Connected state, the network configures all required parameters for communication between the network and the UE. In RRC_Connected state, the UE is in CN_Connected state from the CN point of view. The cell to which the UE belongs and the UE's identity is known. In addition, the cell radio-network temporary identifier (C-RNTI) used for signaling purposes between the UE and the CN is configured. The connected state is intended to transmit data to or from the UE, and to minimize excessive power consumption of the UE. DRX is optimized while maintaining user's quality-of-experience (QoE) [79]. With a configured DRX cycle, the UE only monitors downlink signaling when active, and then goes into sleep mode for the rest of the time with the receiver circuitry turned off. This process allows significant power consumption reduction, as the longer the DRX cycle, the lesser the power consumption. For exhaustive discussion on how DRX reduces excess power consumption, please refer to [78], [80]. Also, the RRC context is established in gNB for the connected state, therefore, data transmission/reception can commence relatively fast, as no connection setup, and signaling is needed. In this state, the network manages mobility by the process known as HO, explained in the Section IV-E.

As regards cell re-selection when leaving RRC_Connected state, the UE attempts to camp on a suitable cell according to *redirectedCarrierInfo* when transitioning from RRC_Connected state to RRC_Idle or RRC_Inactive state [73]. In the connected state, if the network initiates the RRC release message or the UE and CN are no longer attached, the UE moves into an idle state, on the other hand, if the network initiates the RRC suspend procedure, the UE would transit from connected to inactive state [73], [75] (see Fig. 5). One significant difference among the different

⁷CN_Connected state is when the UE establishes connection to the CN [43], [77].

states, as seen from the preceding discussions, is the mobility mechanisms involved. Efficient mobility management is an essential aspect of any mobile communication system. In the following subsections, we describe the different mobility mechanisms including idle- and inactive-state mobility.

B. IDLE AND INACTIVE STATE MOBILITY

Most importantly, RRC states ensure that the mobile UE is accessible via network mobility mechanisms, mainly when the UE is in the idle or inactive states, during which it has limited connection to the network. The network, through paging, communicates with the UE occasionally, and also sends short broadcast message which carries information about changes in the system [73]. The area over which a paging message is sent is an essential feature of the paging process. Also, in both states, the device can switch from one cell to another via cell re-selection. The UE scans for candidate cells for cell re-selection, and if the UE discovers a cell with received power sufficiently higher than its current one, it deems this the best cell and contacts the network through random access [73].

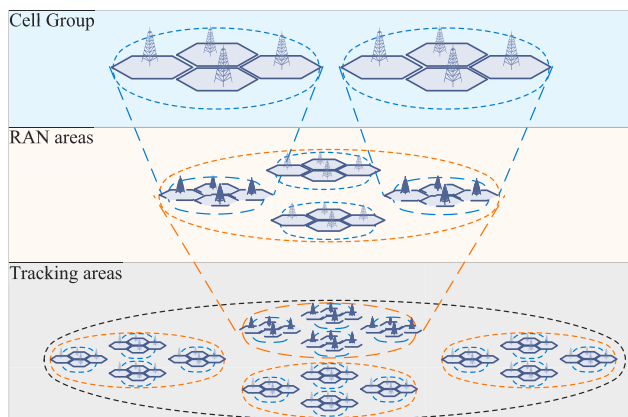


FIGURE 6. RAN Areas and Tracking Areas.

UE tracking needs to be intelligently carried out to avoid high overhead due to paging, and signaling at the network and cell level respectively [81]. Hence, the cell-group level tracking system was introduced in 5G-NR to tackle the challenge of high overhead due to signalling and paging. Figure. 6 illustrates how tracking of UE in the idle and inactive state is carried out in 5G-NR. In order to enable effective UE tracking, the cells are organized into cell groups, and the UEs are only monitored on the cell-group level, as shown in Figure. 6. The network only receives new UE location information when the UE moves into another cell group outside its previous cell-group. In case of paging the UE, the broadcasted paging message is sent to all cells within the specific cell group — this is done to reduce the paging overhead. This is the primary tracking procedure in the NR for both states. However, there is a difference in the way that cells are grouped in both states as well as how paging is initiated.

For the idle state, cell groups are grouped into *RAN areas*, where a *RAN area identifier* (RAI) identifies each RAN area. The *RAN areas*, in turn, are grouped into an even larger group known as tracking areas, where a tracking area identifier (TAI) is used to identify tracking area. Thus, each cell belongs to one cell group which also belongs to one RAN area as well as a tracking area, and their respective identities are provided as part of cell system information.

Tracking areas are the basis for CN-based UE tracking, and the CN is responsible for managing and initiating paging. The CN assigns each UE to a *UE registration area*, which consists of a list of TAIs. When a UE enters a cell belonging to a tracking area outside its assigned registration area, it accesses the CN and performs a *Non-Access Stratum (NAS) registration update*. The CN records the UE's location and updates the UE's registration area, then it provides the UE with a new TAI list that includes the TAIs that the UE has been assigned. The UE is assigned a set of TAIs to avoid repeated NAS registration updates in case the UE moves back and forth between two neighbouring tracking areas. If the UE moves back to the old TAI within the updated UE registration area, no new update is needed.

In the inactive state, RAN Area becomes the basis for UE tracking, which is carried out in the 5G-RAN level. 5G-RAN is responsible for initiating the paging and managing RAN-based notification area. UEs are assigned a RNA comprising the following: a list of cell identities, a list of RAN areas, or a list of tracking areas. The RNA is assigned to a UE by its serving NG-RAN based on the UE's registration area and can cover a single or multiple cells (a subset of the tracking areas). As a result, the UE can move freely within the allocated RNA without contacting the NG-RAN. However, if it moves to an area outside its current RNA, it initiates RAN-based Notification Area Update (RNAU). Once the serving cell (ng-eNB or gNB) receives the RNAU request from the UE, it may send the UE to one of the following RRC states: RRC_Inactive, RRC_Connected, or RRC_Idle. If UE remains in the inactive state, the serving NG-RAN may continue to send a periodic RNAU timer to the UE, which is used to notify the network that the UE is still active. The value of the RNAU time is assigned based on the RRC_Inactive assistant information (RIAI) [39]. In summary, two levels of paging can be applied for reaching the UE depending on its RRC state: CN-based paging for idle state and 5G-RAN-based paging for the inactive state (see Table 3).

C. CONNECTED STATE MOBILITY

The connection between UE and network is established in the connected state. Connected-state mobility aims to maintain connectivity without interruption or noticeable degradation as the UE moves within the network. To maintain the connection between UE and network in the connected state, the UE is continuously searching for new BSs to connect to. The BS search is based on current carrier frequency (intra-frequency measurements) and different

carrier frequencies (inter-frequency measurements) from the UE perspective.

Cell search in the connected state results in HO if suitable condition are met while for idle and inactive state, it results in cell re-selection. When it becomes necessary to perform HO in the connected state, the UE is not responsible for the decision. Instead, the UE performs signal measurement of the serving cell and neighbouring cells and generates the measurement report (MR)—containing cell level measurement results such as reference signal received power, signal-to-interference-plus-noise ratio, reference signal received quality, etc.— sent to the network. Based on this report, the network decides whether or not the UE is to HO to a new cell. The above procedure is not applied to the very small SCs (e.g 5G femtocell) that are tightly synchronized to each other [38], [76].

IV. HANDOVER MANAGEMENT IN 5G AND BEYOND

This section describes the step-by-step procedure for HO in 5G NR, introduces the various categories of HO, and also discusses HO requirements alongside its relationship with radio resources management.

A. TYPES OF HANDOVER

There are two broad categories of HO, namely; intra-/inter-frequency and intra-/inter-radio access technology (RAT) HO.

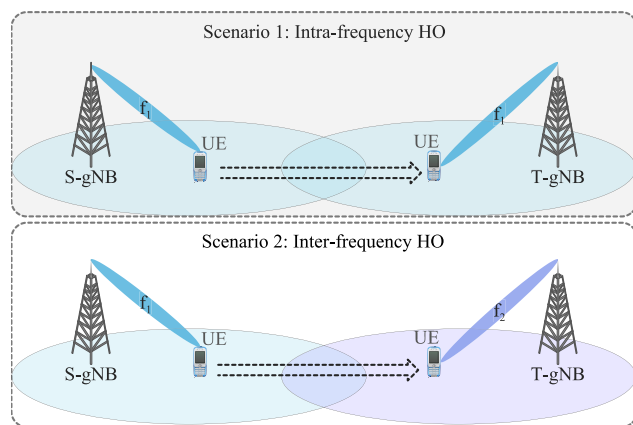


FIGURE 7. An illustration depicting intra-frequency HO in scenario 1 and inter-frequency HO in scenario 2.

1) INTRA-/INTER-FREQUENCY HANDOVER

Intra-frequency and inter-frequency HO are the HO types for which the carrier frequency is the subject of interest. If the UE is to move to the target cell with the same frequency as that in the serving cell, it is generally known as intra-frequency HO as seen in Fig. 7 Scenario 1. In contrast, inter-frequency HO occurs if the UE is to use a different carrier frequency in the target cell as shown in Scenario 2 in Fig. 7. Event A3 and A6 initiate intra-frequency HO. Both Event A3 and A6 are triggered when the neighbouring BSs RF condition is higher than that of the serving BS. Moreover, Event A6 is

used for intra-frequency HO of the secondary frequency on which the UE camps. Event A4 and A5 are typically used for inter-frequency HO. Event A4 is triggered when the RF condition of one of the neighbouring BSs is higher than the threshold compared to that of the other BSs. On the other hand, Event A5 is triggered when the serving BS RF condition becomes lesser than the lower threshold and the RF condition of one of the neighbouring BS becomes higher than the upper threshold (where the threshold values are parameters that are optimized based on the network) [75], [82]. As mentioned in Section III-A3, HO occurs in the connected state and in that state, UE regularly sends the measurement report (MR)—containing cell level measurement results such as reference signal received power, signal-to-interference-plus-noise, reference signal received quality, etc.— of all neighbouring cells to the serving cell. More information regarding the HO trigger events can be found in [29], [38], [75].

The UE essentially carries out the measurements in the measurement gap at different frequencies for inter-frequency cases [82] and [5]. The measurement gap is necessary because without it, the UE would not be able to measure the target carrier frequency while transmitting/receiving to/from the serving cell simultaneously. The measurement gap specifies the time interval when no downlink (DL) or uplink (UL) signal is transmitted. Measurement gap only applies to some cases of intra frequency HO where enhanced UE coverage is not guaranteed to be aligned with the serving gNB's centre frequency [5], [82]. However, the measurement gap is required for all cases of the inter-frequency HO as specified in 3GPP [82]. Researches are concerned with fundamentals question on how the measuring gaps can be reduced, as large measuring gap results in lower throughput and higher UE energy consumption.

2) INTRA-/INTER-RAT HANDOVER

In the case of intra-RAT HO, UE hands over from serving BS (S-BS) to the target BS (T-BS) which both use the same RAT. Intra-RAT HO is commonly referred to as horizontal HO [5] as shown in Fig. 8. Intra-RAT HO can be either intra- or inter-frequency HO. Intra-RAT HO aims to preserve the connectivity of the UE with the existing network and the primary reason for this kind of HO can be attributed to load balancing or measurement trigger conditions [83]. Once UE HO occurs, it prefers to camp on the cell which provides the strongest received signal.

In contrast to intra-RAT HO, inter-RAT (or vertical) HO occurs when the UE hands over to a T-BS which uses a different RAT from the S-BS. Unlike in intra-RAT HO where the cell with the highest received signal is selected, in inter-RAT HO, other factors such as user mobility, service type, as well as the network property and state are considered when selecting the target cell. It also involves the switching of the logical interface between the two RATs [84]. The latency incurred during inter-RAT HO is still prohibitive for many application and services, thus, it poses a severe problem

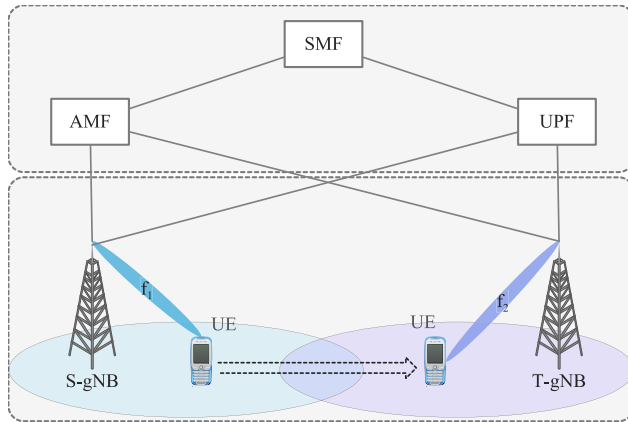


FIGURE 8. UE undergoes HO from one cell to another with both cells using the same RAT (intra-RAT).

in the NexGen mobile systems [84]. In order to improve the user experience, centralized architecture for inter-RAT HO, which integrates legacy and NR network protocol was proposed [84]. Fig. 9 demonstrates how the UE performs inter-RAT HO. From the figure, it can be seen that both distributed, and centralized CN architecture for multi-RATs are possible. The advantage of using centralized architecture is that it can lead to a significant reduction in HO signaling and interruption time.⁸ The centralized architecture comprises unified CN along with the baseband unit (BBU) and remote radio head (RRH) separated through a transport mechanism such as optical fiber. In a C-RAN architecture, the RRHs are connected to the BBU pool through high-bandwidth transport links known as fronthaul.

B. HANDOVER REQUIREMENTS AND KEY PERFORMANCE INDICATORS

Since HO has adverse effect on the overall performance of wireless networks, different features and requirement are necessary to reduce the impact of HO. Also, various key performance indicators (KPIs) are used to measure how the network performs during a HO. The various HO requirements and KPIs are presented as follows:

- **Seamless HO:** a seamless HO occurs when UE perceives continuation of connection during HO with little or no interruption during gNB switch. This guarantees the UE's active connection.
- **HO interruption time:** is a period where the UE is not permitted to send user plane packets to the BS. The UE experiences seamlessly HO if the interruption time is very small ($\leq 1\text{ms}$) [85], [86].
- **HO cost:** is defined as mobility interruption time per HO multiplied by the number of HOs for a particular UE's trajectory. This metric is imperative in the network as it

has a direct relationship to the system throughput [53]. HO cost decreases as the number of HOs and/or the mobility interruption time per HO decreases.

- **HO failures rate:** For any given UE trajectory or unit time, the HOs failure rate is the number of HO failures—unsuccessful HOs—divided by the number of times the UE experienced the HOs.
- **Signaling overhead:** HO signaling overhead are the various data generated during the process of HO to facilitate the operation. However, the HO process interrupts the data flow and results in the reduction of the UE throughput. [29].

There are other performance metrics that are essential to ensure optimal performance in wireless networks, particularly for HO optimization. Further details can be found in *Tayyab et al.* [5].

C. HANDOVER AND RADIO RESOURCE MANAGEMENT

In 5G, the term radio resource includes both traditional (from the legacy system) and extended resource concept [87]. These legacy resources include energy consumption (cell and UE transmitting power), frequency (channel bandwidth, frequency of the carrier) and antenna configurations. In addition, the extended resource definition in 5G covers the hard resource (number/type/configuration of antennas, the existence of nomadic/unplanned nodes, or mobile terminal relays) and soft resources (network node and UE software capabilities). It is also important to meet UE requirements such as QoS or QoE for all the UEs while properly managing resources. On the other hand, proper resource management can help networks fulfill HO KPIs, for example, by reducing the probability of HO failures while maintaining the QoE during and after HO [88], [89]. To increase wireless system efficiency, it is necessary to address and take into account the fundamental issues related to HO and resource management such as admission control, bandwidth and power control.

D. DUAL CONNECTIVITY

Dual connectivity means that the UEs can establish connection to two different cells at the same time [90]. Usually, in dual connectivity, UEs either connect to BSs of different sizes (macro cell and SC) or two different RATs simultaneously (e.g. 4G and 5G network), as illustrated in Fig. 10. Since the UE can be connected to two different RAT over different frequency bands simultaneously, the interruption time is reduced to zero. However, this would trigger an additional likelihood of HO where new HO cases are introduced relative to a single connection. These new HO scenarios (see Fig. 10) occur in two situations; when the UE switches the connection either from SC to macrocell or from SC to SC. With the introduction of mm-wave, the use of dual connectivity could lead to an increase in HO probability, thereby causing additional problems with mobility management, including an increase in signaling overhead, synchronization complexity between RATs for multi-RAT connectivity, simultaneous utilization of resources in multiple BSs, and reduction in battery

⁸Michael Wang, *5G, C-RAN, and the Required Technology Breakthrough*, Published on 21 Jun. 2018. Available online at <https://medium.com/@microwang/5g-c-ran-and-the-required-technology-breakthrough-a1b2babf774>. Accessed on 25 Oct. 2020.

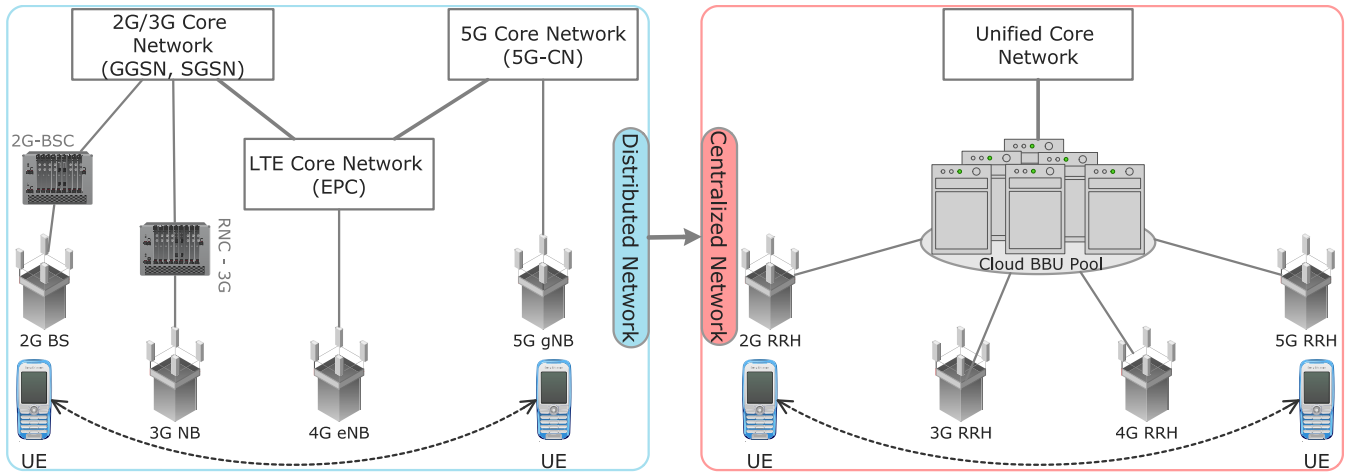


FIGURE 9. Inter-RAT HO scenarios in distributed and centralized RAN architectures¹⁰ [84].

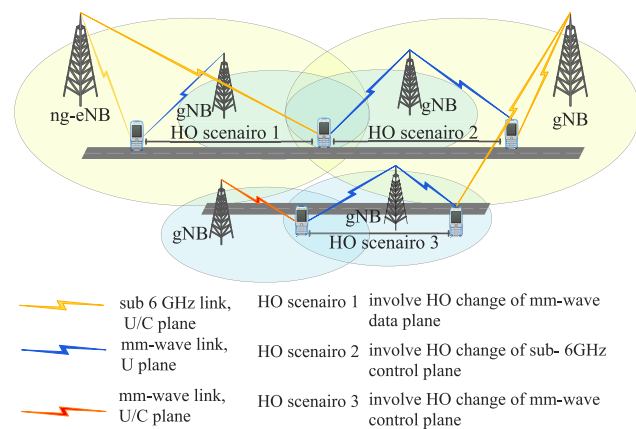


FIGURE 10. Dual connectivity with HO scenarios in future communication networks.

lifespan. The increase in signaling overhead is due to flow control between the RATs [90], [91], and these issues could be addressed using intelligent approaches.

E. HANDOVER MANAGEMENT IN NR

NR physical layer uniquely differs from the legacy RAT with the following features: dual connectivity, high-frequency spectrum, forward compatibility, ultra-lean design, use of mm-wave and relay for devices (device-to-device). NR supports both multi-connectivity and single-connectivity selection depending on the configuration set. For both configurations, hard HO is used during path switching [29]. In both licensed and unlicensed spectrum, NR operates between 600 MHz and 73 GHz. Forward compatibility means designing radio-interface architecture that enables new service requirements and accommodate new technologies while supporting legacy network UEs. While the ultra-lean design principle aims to decrease the always-on transmissions (for example, signals for BS detection, broadcast of system information) to achieve high data rates with low

energy consumption in the network. The main challenge for NR is the coverage due to the use of high frequency with high penetration loss that makes the cell footprint to become smaller. In this section, we describe the NR HO with a brief introduction of critical features and the entities involved in NR mobility. Also, a step by step HO procedure is provided for intra-AMF/UPF. The types of HO in NR are described as follows: **1) Intra-gNB HO:** This occurs when both the source and target cells¹¹ belong to the same gNB, as shown in Fig. 11.

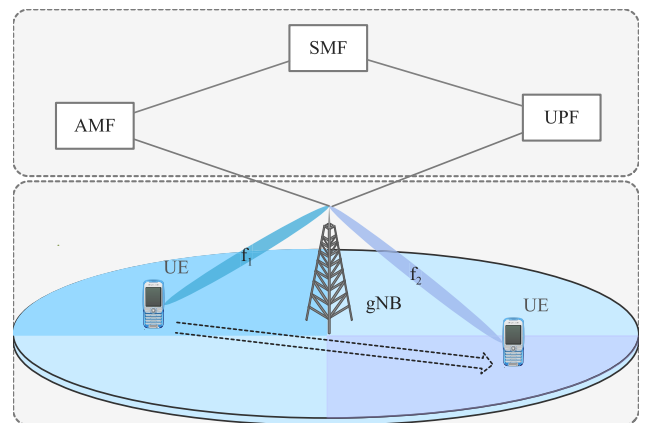


FIGURE 11. UE performs intra-gNB HO which involves the change of cells in the same gNB.

2) Inter-gNB HO without AMF Change: Inter-gNB HO generally occurs when serving and target cells are from different gNBs. There are two different types of HO within inter-gNB HO without AMF change, depending on whether the HO involves a change of UPF or not. However, the inter-gNB HO discussed here does not include a change of AMF in both cases, as shown in Fig. 12. Inter-gNB with

¹¹Cell here means the part of sector gNB that has specific beams and covers the specific environment.

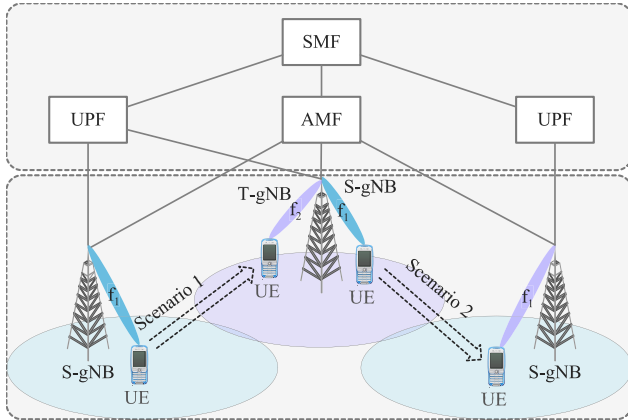


FIGURE 12. UE performs inter-gNB HO, which involves the change of gNBs with same UPF and AMF for scenario 1 and change of UPF for scenario 2.

intra-UPF HO is presented in Fig 12 scenario 1, where the HO involves a cell change with the same UPF, while Fig. 12 scenario 2 presents inter-gNB with inter-UPF HO where the cell switch involves a change of UPF.

3) Inter-gNB HO with AMF Change: In this case, the HO requires a change of AMF from the source to the target AMF. However, the HO involves no change of SMF, and only the NG interface is used as depicted in Fig. 13. There are two cases of inter-gNB HO with AMF change; in the first case (Fig. 13 scenario 1), the same UPF is maintained while the second case (Fig. 13 scenario 2) involves a change of UPF during HO.

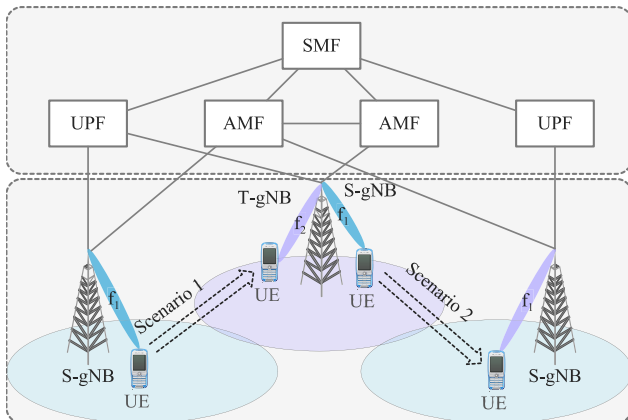


FIGURE 13. UE performs inter-gNB HO with AMF change, involving the change of gNBs while UPF is maintained in scenario 1 and change of UPF in scenario 2.

The basic HO procedure in NR is shown in Fig. 14 [5], [38]. It consists of three phases, namely: HO preparation (Steps 0-5), HO execution (Steps 6-8) and HO completion (Steps 9-12), which are described as follows:

- **Step 1:** the UE measuring procedure is configured according to access restriction and roaming information by the serving gNB (S-gNB), and the UE sends an MR to the target gNB (T-gNB).

- **Step 2:** the S-gNB determines to HO the UE, based on the MR and radio resource management information.
- **Step 3:** the S-gNB sends a HO request message to the T-gNB (which includes the necessary information to prepare for HO to the T-gNB).
- **Step 4:** the T-gNB executes the admission control procedure if the T-gNB can grant the resources.
- **Step 5:** the T-gNB sends a HO request acknowledgement to the S-gNB. As soon as the S-gNB receives the HO request acknowledgement message, data forwarding may be initiated.
- **Step 6:** S-gNB sends a HO command to the UE.
- **Step 7:** S-gNB sends the Sequence number status transfer message to the T-gNB.
- **Step 8:** UE detaches from the S-gNB and synchronizes with the T-gNB.
- **Step 9:** the T-gNB informs the AMF that the UE has changed the cell, through the Path switch request message.
- **Step 10:** 5GC switches the DL data path towards the T-gNB.
- **Step 11:** the AMF acknowledges the Path switch request.
- **Step 12:** the T-gNB informs the S-gNB that the HO was successful and triggers the release of resources by the S-gNB by sending a UE Context Release message. Finally, the S-gNB release the radio resources associated with the UE.

It is essential to point out that the above procedure is applied for HO between NR and NR technologies.

F. MOBILITY AND HANDOVER MANAGEMENT IN B5G

Researchers have anticipated some use cases and applications that make B5G to be different from 5G. Some of these use cases include integrated unmanned aerial vehicles (UAVs) communications, high mobility of devices (above 500 kmph), holographic projection, etc [92]. The high mobility of devices, UAVs, and other applications that use radio waves at the mm-wave and THz spectrum presents unprecedented wireless communication challenges in B5G. Among these challenges, mobility and HO management are anticipated to be the two most challenging issues in B5G networks since B5G networks would be highly dynamic, and multi-layered, which would lead to more frequent HO. High mobility of devices and UAVs results in uncertainties of their locations and keep in mind that high frequencies such as mm-wave and THz that would be used in B5G can be easily blocked by humans, buildings, etc.

Heuristic and traditional HO methods would not be able to react quickly. An alternative solution is to adopt artificial intelligence models for mobility prediction and optimal HO strategy in order to guarantee communication connectivity. Even though the introduction of multi-connectivity is a very promising solution, the procedure still needs intelligent HO management strategies to optimize the cell (re-)selection process in order to reduce signaling, guarantee high data

TABLE 4. Handover management in NR challenges associated, existing solution and future research directions.

| Challenge | Possible causes | Consequences/Effects | Solutions/Future research directions | Ref |
|-----------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| Frequent /unnecessary HO, HO failures and Ping Pong | The susceptibility of mm-wave channel to significant pathloss, atmospheric absorption and blockages. Deployment of large number of mm-wave small cells in HetNets to meet the traffic demand from UEs. | Varying signal strength, service interruptions during communication between the UEs and s-gNB which leads to a reduction in the overall QoS and increased service failures. | Development of advance techniques for array beam-forming and large MIMO to increase channel gain. Multi-connectivity that allows UE to communicate with more than one BS at the same time. Improved solutions for mobility and KPI's prediction for the proactive HO scheme. | [34], [90], [91], [93] |
| Load balancing | The uneven distribution of UEs in the network due to the random placement of cells or UEs mobility. | Cell congestion is likely to occur in some cells, due to load imbalance, resulting in the high HO failure rate. | Intelligent load balancing schemes that can provide fairness between congested and uncongested cells become vital to ensure efficient resources allocation and improvement in network performance. | [83], [94], [95] |
| Signaling overhead | Signalling overhead occurs during HO procedure and the problem escalates as the number of UEs increase. The paging procedure involved when the BSs searches for the UEs in the network also cause additional signalling overhead. | The higher the number of HOs, the more the signalling overhead that is generated, which causes more interruption in data transfer and hence high latency in the wireless system. | Optimal HO schemes that can reduce the number of HOs by considering various HO parameters. Development of models that requires the paging of fewer BSs in order to locate a UE in the network. | [34], [96], [97], [98] |
| Power consumption | Frequent HO procedures due to the deployment of massive small cells in the HetNets drains UE battery. The introduction of handover skipping scheme requires the UEs to increase transmission power to maintain connection with a distant BS results in more power consumption in the UEs. | The higher the number of small cells, the more the number of measurement and the faster UE battery drains. The HO skipping schemes increase the distance between UE and BS, thereby causing the UE to consume more power to reach the BS. | Measurements should be done within the DRX cycle. Solutions that use ML techniques should be developed to ensure a proper trade-off between HO and power consumption. | [99], [100], [81] |
| HO latency | The transmission of measurement reports between the UEs and the BSs during HO process causes signalling overhead which results in service delays. | During signaling, there is interruption in data transmission which leads to a reduction in the average throughput of the users as well as delays in data transmission. | Proactive HO solutions are essential to reduce the amount of signaling overhead associated with the HO process and to select the optimal BS before HO happens. | [101], [102], [103] |

rate, high reliability, and low latency in the B5G [92]. The HO procedure for the B5G might be similar to that of 5G, but there are no standards for B5G system yet. A summary of the challenges associated with HO management in NR alongside their causes and potential solutions are presented in the Table. 4.

V. MACHINE LEARNING FOR HO MANAGEMENT

The use of mm-wave and higher frequencies in 5G and B5G networks is going to introduce new challenges and complexity to the HO management that would be difficult to handle by conventional methods. Firstly, these frequency ranges suffer from severe attenuation (e.g., larger penetration losses), which means their transmission distance will be small. As a result, more BSs need to be deployed to cover the same area that would have been covered by those utilizing the microwave frequencies [104]. This implies that the size and the complexity of the network is going to greatly increase and the users will be prone to more frequent HOs which would greatly affect their QoS, particularly for high mobility users and applications.

Secondly, due to the use of directional beams for transmission in mm-wave networks, the presence of obstacles on the path of the transmitted beam can partially or completely hinder the user from gaining access to the network or negatively

impact the signal quality. As such, in mm-wave communication networks, the users are not only faced with the challenge of selecting the optimal BS but also the optimal beam to connect to per time in order to maximize their QoS. Hence, optimal beam selection has become another factor to consider in HO management process which would further add more complexity to the HO process because of the massive number of beams that the user has to select from during each HO instance [105], [106].

Finally, there is also the need to provide some high mobility based essential services for emergency scenarios such as medical services to patients in ambulances en-route hospital through real-time consultations with the doctors that are situated in a remote hospital. Especially, in the pandemic situation that we find ourselves in now, this kind of services may be needed to sustain the lives of the patients in critical conditions before they get to the hospital to receive proper medical attention [107], [108]. Intelligent HO optimization would help predict the route of the ambulance, determine the optimal BSs to connect and also pre-allocate the resources that will be needed at the BSs. This will help prevent intermittent service interruptions and guarantee the QoS need to support the communication between the paramedics in the ambulance and the doctors at the remote location [109], [110].

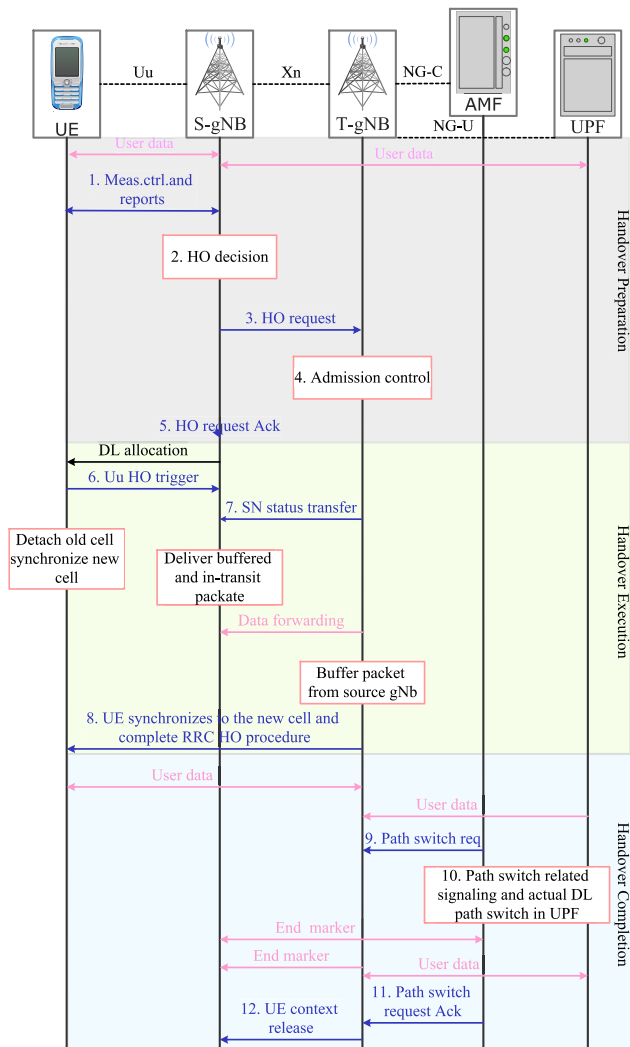


FIGURE 14. HO procedure in 5G-NR involving no change of AMF and UPF, based on [38].

Therefore, effective HO optimization would enable the selection of the optimal BS and beam for user connection that will maximize user connection, reduce excessive or unnecessary HOs, and enable the detection of obstacles and their avoidance. These are some of the issues that make HO optimization in mm-wave communications networks more challenging to handle compared to the previous generations of cellular networks. Moreover, since the HO process involves various network parameters that must be considered and optimized in real time in order to ensure seamless HO, this would be very challenging for most conventional methods to handle. The challenge with conventional methods of HO management is that they are computationally demanding to implement, particularly when the network dimension becomes very large. As such, before they can decide which target BS to associate the user with, the user must have moved from that location. This would result in sub-optimal HO decision and degradation in user QoS. In addition, they cannot accurately capture certain details of the network such as the presence of

different types and sizes of obstacles, as well as the dynamic traffic demand patterns that are typical of 5G and B5G networks, which are also important for making an optimal HO decision [109], [111], [112].

However, ML techniques can assist in bringing intelligence and helping the network to self-optimize. ML techniques are able to learn various network characteristics from data generated from the network, in order to optimize various aspects of the network. They are able to capture hidden details and patterns in the network from the network data that cannot be represented by analytical models [111]. They are self-adaptive and as such can react to changes in network environment and even predict future network or user demands before hand, thereby enabling the network to adequately prepare to handle such demands when they occur [109]. They can be designed in a computationally efficient manner such that the training phase of the algorithm, which is often computationally demanding, can be carried out offline, and then the trained model deployed online to carry out real-time optimization after which the model can be updated periodically, as it experiences new data [113].

In this section, we first present an overview of the major categories and types of ML algorithms used for HO optimization. Then, we delve into reviewing the state-of-the-art on ML-aided HO management. A top-level taxonomy is followed while reviewing the state-of-the-art, such that the ML-aided HO management methodologies are classified based on the source of the data they utilize. As such, two broad categories are considered: visual data based and wireless data based HO management techniques. The major objective of this novel taxonomy is to recognize the visual data aided HO management schemes—which has been long overlooked in the literature—by giving it a special place along with the traditional wireless data driven HO schemes. The visual aided wireless communications is an emerging research area in wireless communications where visual information (pictures/videos) captured from cameras, light detection and ranging (LIDAR), etc., are combined with wireless sensory data for wireless network optimization such as channel prediction, HO optimization, etc [114], [115]. This is necessary because mm-wave communication networks possess unique challenges that would be difficult to handle using only wireless sensory data but with the assistance of visual data, some of these challenges can be handled properly. On the other hand, for the wireless data based HO management, the most recent works are extensively reviewed under two use cases: beam selection and BS selection.

A. AN OVERVIEW OF MACHINE LEARNING ALGORITHMS

It has become very important to include AI/ML in the BS’s and beam selection process during HO, in order to achieve the primary objective of providing a seamless HO and to ensure that the UE achieves maximum throughput during the entire duration of its connection to the network. The HO optimization problem is a decision-making problem, and intelligence is imperative to ensure that the optimal decision

TABLE 5. Types of machine learning algorithm.

| Learning category | Inference task | Learning algorithms |
|------------------------|------------------------------|----------------------------------------------------------------------------------------------------------|
| Supervised learning | Classification/Regression | SVM, ANN, XGBOOST, k NN, decision tree, random forest |
| Unsupervised learning | Clustering/anomaly detection | k -means Principal Component analysis Expectation maximization Independent component analysis |
| Reinforcement learning | Decision making | Q -learning SARSA Policy gradient Proximal policy optimization Actor-critic |

is taken at each HO instance in a more efficient and effective manner.

We begin by defining ML and discussing the various categories. According to [116], ML is a set of computation procedures that evolved from formidable techniques in the field of AI that allow the computer to self-learn, discover patterns, and generate models from historical data without being explicitly programmed. The objective of ML is to identify features of a given data set that are likely to influence an outcome of interest given the input, and then use those learned features to predict the result in a new situation not previously encountered [32]. A substantial collection of ML techniques (model and algorithms) has been created to solve various challenges in different domains. These algorithms can be classified according to how learning is performed. They have been broadly categorized into three major classes [117]. Table 5 presents an overview of ML approaches based on their learning styles.

Definition 6 (Labelled Data Set): A labelled data set is a data set with clearly defined features (input) and target (output). The features are usually related to the target and enables the ML algorithm to identify the target or map the input to the output during the training phase.

Definition 7 (Unlabelled Data Set): An unlabelled data set is a data set that does not have labels. That is, there is no clear description of the features or targets in the data set.

Definition 8 (Model Training): Model training is the process of exposing an ML algorithm to the training data set (i.e., labelled or unlabelled data set) in order to enable it to learn the mapping between the features and the target. Thereafter, a model is obtained that can correctly predict the right target, even when it is feed with a new data set that it had not previously seen.

1) SUPERVISED LEARNING

As the name suggests, it is the learning technique which requires a labelled training set consisting of inputs features and output. The learning model tries to search for a function that maps the input to the desired output by minimizing

both the bias and variance error of the predicted results. After that, new data set is then applied to the trained model in order to predict the output. Supervised learning is basically classified into regression—where the predicted output is continuous—, and classification— where the predicted output is discrete or categorical. Examples of supervised learning algorithms include: artificial neural networks (ANN), support vector machine (SVM), extreme gradient boosting (XGBOOST), k -nearest neighbour (k NN), decision tree, random forest, etc, [118]. Supervised learning algorithms can help provide user mobility information through prediction of future location, trajectory, cell, etc., which is needed for proactive HO optimization and efficient resource allocation in 5G and B5G networks order to enhance the QoS of users [32].

2) UNSUPERVISED LEARNING

Different from supervised learning, in unsupervised learning, the training data set is unlabelled. The learning model in this case, tries to find hidden patterns, structures, and correlations within the training data set. They are mainly employed for anomaly detection, pattern recognition, and the reduction of the dimension of a data set. Common examples of unsupervised learning algorithms are k -means clustering, principal component analysis, expectation-maximization (EM), etc [119]. With the deployment of ultra-dense cellular networks and use of diverse kinds of devices (conventional UEs and IoT devices) in 5G and B5G, clustering algorithms can enable scalable and decentralized HO optimization particularly for cases where user mobility patterns are heterogeneous thereby reducing complexity. As an example, the authors in [120] proposed a two-layer approach to HO optimization in an ultra-dense network where k -means was first used to cluster the devices with similar mobility pattern then deep reinforcement learning was implemented to determine the optimal HO policy of the devices within each cluster.

3) REINFORCEMENT LEARNING

Unlike supervised and unsupervised learning that deal with continuous or discrete output prediction and identification

of hidden pattern or structures in data, RL is concerned with making decisions in order to obtain an optimal policy in a given environment. It is a trial and error kind of learning whereby an agent interacts with the environments, takes action and gets feedback in terms of reward or penalty, depending on whether the action taken is good or bad for a given objective. The outcome of RL is to learn the optimal policy that would enable the agent to make an optimal decision at any given state of the environment. RL algorithms can be value-based (e.g. Q -learning, SARSA), policy-based (e.g. policy gradient, proximal policy optimization (PPO) and actor-critic (A2C)) [121], [122]. Reinforcement learning algorithms are suitable for mobility management and HO optimization as they are able to adapt to varying user mobility and network condition in order to determine the optimal HO policy. They are particularly relevant in 5G and B5G networks because of increasing network dimension and complexity in order to reduce HO delays and minimize frequent HOs [32].

4) DISTRIBUTED LEARNING

Conventional approach to ML requires that the training data be stored in a central location either at the data centre or in the cloud. However, this approach has several challenges including data privacy, latency, increased signaling overhead, increase in the energy consumption of the UEs, etc. This has led to a growing interest in distributed learning where the data is processed at the location where it is generated, and only the trained ML models are transmitted to the central entity [123] and [124]. In this regard, federated machine learning [123] and [125] has been introduced to handle the aforementioned challenges and is gaining increasing application in the field of wireless communication [124] and [126]. Federated learning is a distributed machine learning approach which enables data generating entities to jointly learn a shared prediction model without having to send their data to the central entity. In this case, only the trained models are transmitted to the central entity. This ensures the preservation of data privacy and requires less communication resources for model transmission [127] and [128]. Federated learning has been applied in [129] and [130] for human mobility prediction in order to preserve the privacy of users and in [103] for proactive HO mm-wave vehicular networks in order ensure the privacy of user location information, minimize communication overhead while minimizing frequent HOs.

B. MACHINE LEARNING BASED HANDOVER OPTIMIZATION

HO optimization is necessary when selecting the BS/beam that a user should connect to, in order to minimize frequent HO due to the small footprint of mm-wave BSs in 5G and THz wave BSs that are envisioned to be used in B5G. This is because frequent HOs increase the HO cost, thereby reducing the network throughput. Throughout this paper, we will refer to HO as defined in [53] which establish the term HO cost.

With efficient HO optimization, the network is able to select the best T-BS that will provide a higher throughput for UE.

Before ML came into play, classical methods for BS selection were based on specific parameter measurement. These methods include selecting the T-BS based on distance, or the BS that provides a higher KPI such as reference signal received power, received signal strength indicator, and signal-to-noise ratio. In the measurement-based approaches, the channel state information (CSI) from the MR of all neighbouring BSs is measured, and the one with the best CSI is selected as the potential T-BS. These approaches are practical for sub-6 GHz frequencies; however, they are inefficient solutions in mm-wave and THz application band due to severe path loss and susceptibility to LOS blockage [102].

ML techniques can play a significant role in HO optimization and BS station selection by reducing delays, computational overhead, and frequent HOs. They help predict the T-BS and also ensure that adequate resources are available at the T-BS before HO occurs in order to ensure a seamless HO. In this section, we consider ML-based HO management in 5G networks from the perspective of visual, and wireless data aided HO optimization. In Table 6, we present a summary of the state-of-the-art ML-based HO optimization in 5G mm-wave communication systems.

1) VISUAL DATA AIDED HANDOVER OPTIMIZATION

Successive generations of cellular networks have mainly depended on wireless sensory information such as CSI, received power, etc., for network design and optimization. However, the use of mm-wave and THz frequencies in 5G and B5G networks would mean that BSs will have many antennas, communication will be through a large number of LOS beams, which would be subject to blockages of various types and would limit signal reception at the user end. In addition, much signaling overhead would be involved in the selection of the optimal beam for user connection in mm-wave networks if only wireless sensory data are exploited for optimal beam selection considering the massive number of beams that would be involved [114], [115].

The vision assisted HO optimization has become necessary because of the complexity of the mm-wave networks, and it might not be possible to capture all the conditions of the environments like obstacles, buildings, etc., using wireless sensory data. As a result, detecting or predicting the presence of obstacles that would block the received beam and reduce the throughput at the user end is very difficult to achieve using only wireless sensory data. However, with vision assisted HO optimization, visual data (image/video) is combined with wireless sensory data to enable proactive obstacle prediction and optimized beam/BS selection that would help enhance user QoS [131]. In addition, with the advancement in computer vision, the training overhead that is normally associated with training ML models for optimal beam selection can be greatly reduced by utilizing the images of networks in developing deep learning algorithms for efficient HO operation [132], [133]. In the following paragraphs,

TABLE 6. Summary of the state-of-the-art ML-based HO optimization in 5G mm-wave communication systems.

| Paper | Year | Visual Data | Wireless Data | | ML Algorithm |
|--------------|------------|-------------|----------------|--------------|---------------|
| | | | Beam Selection | BS Selection | |
| [27] | 2020 | ✓ | | | DRL |
| [134] | 2019 | ✓ | | | DRL |
| [135] | 2019 | ✓ | | | CNN |
| [136] | 2019 | ✓ | | | DNN |
| [137] | 2019 | ✓ | | | CNN |
| [131] | 2020 | ✓ | | | DNN |
| [132] | 2020 | ✓ | | | DNN |
| [139] | 2018 | | ✓ | | SVM |
| [140] | 2019 | | ✓ | | kNN,SVM, ANN |
| [141] | 2019 | | ✓ | | Random Forest |
| [142], [143] | 2019, 2020 | | ✓ | | SVM |
| [144] | 2019 | | ✓ | | DNN |
| [145] | 2018 | | ✓ | | Random Forest |
| [146] | 2019 | | ✓ | | ANN |
| [147] | 2020 | | ✓ | | DNN |
| [148] | 2020 | | ✓ | | DNN |
| [149] | 2020 | | ✓ | | DNN |
| [150] | 2020 | | ✓ | | DNN |
| [151] | 2019 | | ✓ | | MAB |
| [154] | 2020 | | ✓ | | MAB |
| [155] | 2020 | | ✓ | | CMAB |
| [156] | 2020 | | ✓ | | MARL |
| [157] | 2020 | | ✓ | | Q-learning |
| [158] | 2018 | | ✓ | | Q-learning |
| [159] | 2020 | | ✓ | | Q-learning |
| [161] | 2020 | | ✓ | | ANN |
| [162] | 2019 | | ✓ | | GNN |
| [163] | 2019 | | ✓ | | CNN |
| [164] | 2019 | | ✓ | | DNN |
| [165] | 2019 | | ✓ | | CMAB |
| [160] | 2019 | | ✓ | | Q-learning |
| [101] | 2018 | | | ✓ | DNN |
| [166], [176] | 2019, 2020 | | | ✓ | CMAB |
| [167] | 2018 | | | ✓ | MAB |
| [97], [177] | 2020 | | | ✓ | DDRL |
| [168], [169] | 2020 | | | ✓ | MARL |
| [170] | 2020 | | | ✓ | Q-learning |
| [171] | 2020 | | | ✓ | DNN |
| [172] | 2020 | | | ✓ | MARL - PPO |
| [173] | 2020 | | | ✓ | RL |
| [95] | 2020 | | | ✓ | DDPG RL |

we review the research works that have been proposed on the use of visual data for HO optimization in mm-wave networks.

One application of visual data for HO optimization is the prediction of obstacles that might affect the magnitude of the received power or data rate at the user end. In this regard, the authors in [27], proposed a cooperative sensing scheme for proactive HO in mm-wave networks using a combination of images captured from multiple cameras and received power. The idea is to map camera images with HO decision using DRL such that a proactive HO decision can be initiated before the received signal is blocked by an obstacle. The advantage of using multiple cameras is to cover areas that are inaccessible by other cameras so as to get a complete view of the network environment. The camera images also enable the prediction of obstacles that will affect the mm-wave links. The authors in [134] developed a DRL framework using camera images for optimizing the HO timing by predicting the future data rate of mm-wave links and ensuring that proactive HO is performed before data rate degradation occurs due to presence of obstacles.

Another application of visual data for HO optimization is the prediction or selection of the optimal beam that the user should connect based on user mobility and the presence of obstacles in the mm-wave network environment. Following this research direction, the authors in [135] demonstrated how data obtained from LIDAR sensors could be used to reduce the overhead associated with mm-wave beam selection and LOS detection, and proposed a decentralized architecture using deep CNN. Their work was extended in [136] by developing a deep learning-based centralized architecture for beam selection and the detection of the LOS in vehicular networks by combining location information and LIDAR data. In [137], the authors proposed a novel beams selection scheme that is capable of predicting the optimal beam to connect to at any position in the cell using image-based 3D reconstruction and CNN. They argue that the proposed method takes images from ordinary cameras and is cheaper to implement compared to LIDAR-based approaches in [135], [136].

The work in [131] considered the problem of beam selection and blockage prediction using camera images, channel state, and deep learning for a single user communication in a mm-wave network. The beam selection problem was formulated as an image classification problem such that the UEs are mapped to a class of beams having a unique beam index, depending on their location in the image. However, to detect users that are blocked, the images are matched with channel information due to the difficulty of detecting obstacles in still images. The authors in [132] first developed a realistic image data set for ML-based mm-wave network optimization that considers many BSs, users, different obstacles, and rich environmental dynamics. Then leveraging the image data set and information regarding previously selected beams, a ML based vision-aided beam tracking framework was proposed to predict the future beams of mobile users in a mm-wave communication system.

2) WIRELESS DATA AIDED HANDOVER OPTIMIZATION

Non-vision assisted HO optimization, on the other hand, does not involve the uses visual sensory information such as images and videos for HO optimization. It uses the conventional wireless sensory information such as received signal level and channel state, and user location information to optimize the switching of user connection from one BS or beam to another. This is the general technique that is commonly used in wireless communications systems. In this session, we review the state-of-the-art in HO optimization in mm-wave communications networks from the perspective of the BS and beam selection by exploiting the CSI and user mobility information such as user location, trajectory, etc.

a: BEAM SELECTION

Due to the high path loss and sensitivity to blockage experienced by mm-waves, a large number of BSs comprising multiple directional antenna arrays have to be deployed. The use of multiple antenna arrays enables the formation of narrow signal beams with a high gain when the phase or amplitude of each antenna is adjusted. This approach, commonly known as beamforming [138], enables the formation of directional links between the BSs and UEs. However, because each BS comprises multiple beams, the challenge becomes selecting the optimal beam that will serve the UE in order to satisfy its QoS. In the following paragraphs, we review the most recent works on ML-based beam selection in mm-wave and THz communication systems.

The beam selection problem is sometimes modeled as a multi-classification problem, after which a supervised or deep learning algorithm is used to identify the beam class. In this regards, the authors in [139] proposed a data-driven approach for analog beam selection in hybrid MIMO systems. The beam selection problem was first formulated as a multi-classification problem and then solved using SVM in order to obtain the optimal analog beam for each user. The performance evaluation shows that the proposed method has similar data rate to that of traditional methods but with lesser complexity. In [140], the direction of arrival information was leveraged to developed a ML scheme for beams selection in mm-wave communications. The beam selection problem was expressed as a multi-class problem, and three supervised learning algorithms namely k NN, SVM, and ANN were used to solve the problem. The authors in [141], proposed a beam selection policy for THz systems based on ML approach with low complexity. The beam selection problem was first formulated as a multi-classification problem after which a random forest algorithm was used to determine the optimal beam class.

In [142] and [143], a ML framework for analog beam selection was proposed using SVM, which considers the transmit power of the SCs and channel information as inputs while the model training was performed using sequential minimal optimization in order to achieve high sum-rate at a lower computational complexity. A DNN model for

beam selection where channel knowledge is not required was developed in [144]. The beam selection problem was modeled as an image reconstruction problem, after which the DNN was used for interpolation. The proposed model was first trained offline—to reduce the training overhead—before online implementation of the trained model was performed. A beam selection framework for mm-wave vehicular networks using different ML-based classification models was proposed in [145]. The training data set comprised the vehicle location, type of receiver vehicle and its surrounding vehicles as well as previously selected beams. It was observed that the random forest algorithm outperformed other classification algorithms in terms of accuracy and efficiency. A neural network framework for beam selection in THz communication networks was proposed in [146]. The proposed model was trained using data samples obtained from the THz channel based on the multi-classification approach. The proposed model was able to determine the optimal beam for each user with low complexity compared to the conventional exhaustive search method.

Another category of mm-wave beam selection technique exploits the CSI of sub-6 GHz to minimize the search overhead involved in selecting the optimal beam as well as for initial beam establishment. In this regard, the authors in [147] proposed a DNN based framework for selecting the optimal mm-wave SCs and beam in a HetNet involving mm-wave SCs and sub-6 GHz macrocells. They utilized the CSI from sub-6 GHz macrocells for both SCs and beam selection in order to minimize the latency resulting from using conventional exhaustive search approach for beam selection. The authors in [148] introduced a deep learning approach to mm-wave beam selection in 5G and B5G using sub-6 GHz CSI. They argue that using the sub-6 GHz CSI for the mm-wave beam selection would help reduce the search space required for establishing the initial beam. In [149], a deep learning framework was proposed for predicting mm-wave beam and blockage while using sub-6 GHz channel. They proved that under certain conditions, a mapping function exists, that can be used to predict the optimal beam and blockages in any environment. Then they went further to show that this mapping function can be learnt using a large enough neural network after which a DNN model was designed to perform both predictions. The work in [150], suggested a deep learning approach for the prediction of the optimal mm-wave downlink beam. The developed DNN model takes as input a combination of features extracted from both the sub-6 GHz channel and mm-wave band in order to enhance prediction accuracy and achievable data rate.

RL techniques have also been applied to mm-wave beam selection in literature. In this regard, different (deep) RL algorithms such as multi-armed bandit (MAB), Q -learning, deep Q -learning approaches have been proposed. A novel ML-based beam tracking and alignment framework for a sparse and time-varying mm-wave channel was proposed in [151]. The channel tracking was performed using Bayesian learning and Kalman filtering after which the optimal beam

selection strategy was obtained using MAB. A fast ML algorithm for beam selection in 5G mm-wave vehicular networks using contextual MAB (CMAB) was proposed in [152] and [153]. The proposed model considers the traffic pattern and different types of blockages in order to select the optimal beam in real-time without prior training of the model. In [154], a beam tracking approach based on MAB is proposed to determine the optimal beams and data rates of the beams in a mm-wave communication system. The proposed model uses the beam quality information, and the feedback obtained from users during initial access to determine the optimal beam and transmission rate during the next transmission.

The authors in [155] proposed an online learning algorithm for optimal beam selection in mm-wave vehicular networks using CMAB. The developed algorithm is able to predict the beam direction of the target mm-wave BS from the serving mm-wave BS depending on the current traffic pattern while considering the user QoS requirements. In [156] multi-agent RL (MARL) approach for the joint optimization of user scheduling and beam selection in mm-wave networks was developed. The proposed method ensures that the delays associated with beam selection are minimized while ensuring that the users QoS are satisfied. The authors in [157] proposed a framework for mm-wave beam prediction in multi-UAV communication systems using Q -learning. The proposed model exploits the received coupling coefficients (a pair of analog beamforming vector from the transmitter and receiver side of the channel) to determine the optimal beam that will maximize the received signal-to-interference-plus-noise-ratio.

A learning-based approach for optimizing beam search in mm-wave BSs in an indoor network environment while considering user mobility has been proposed in [158]. The proposed approach uses multi-state Q -learning while exploiting user trajectory-based data from the radio. They argue that the proposed method is superior to traditional methods because it jointly considers both BS and beam selection, can be adapted to different indoor environments and user mobility and minimizes the delays due to beam search. A beam selection framework for high mobility vehicular networks which aims at enhancing data rate, minimizing the number of HOs and disconnection time was proposed in [159]. The proposed framework utilizes parallel Q -learning to determine the optimal beam for each vehicle. The algorithm leverages the possibility of simultaneously collecting information from multiple vehicles on the road to hasten its convergence to the optimal solution. In [160], an RL framework for beam selection in NLOS scenarios was introduced. The proposed framework employs Q -learning to determine the optimal NLOS beam for each user based on the user's QoS requirement.

The user location can also be exploited in order to identify the optimal beam selection for user association. The authors in [161], proposed a beam selection strategy based on ML that considers the user position and receiver orientation to select the optimal beam pair, thereby reducing the overhead associated with beam alignment. Moreover,

since their approach to beam selection is based on multi-classification, the neural network model is enriched with a large amount of CSI to enable it not only to select the strongest beam based on the magnitude of received signal but also an alternative beam. This makes the proposed approach resilient against blockages. A hierarchical learning-based beam selection scheme was proposed in [162] for multi-users in mm-wave vehicular networks. They developed a graph neural network (GNN) model for beam pair selection while considering CSI and user positions. A deep learning model based on CNN architecture was proposed in [163] for selecting the beam that gives the best communication performance to users in a massive MIMO system while considering user position. In [164], the authors developed a learning-based beam alignment scheme for mm-wave systems that can determine the optimal BS while only exploiting the user position. The proposed scheme can predict the optimal BS and beam even with incomplete user location information with reduced search time. A position-based online learning framework for optimal beam pair selection and refinement was proposed in [165] while considering only user position. The beam selection and refinement problem was first modeled as a continuum-armed bandit problem after which a risk-aware greedy upper confidence bound (UCB) algorithm was developed for beam selection while a hierarchical optimistic optimization (HOO) was used for beam refinement. The observed that where more information regarding the environment can be obtained from BS or user devices, the training overhead can be further reduced.

b: BASE STATION SELECTION

A proactive HO framework that enable users to switch connection to another BS before link disconnection was proposed in [101]. The proposed method uses deep learning to predict obstacles and trigger HO before link disconnection occurs, thereby ensuring the reliability of the link and preventing data transmission delays due to link disconnection. In [166] and [166], a HO mechanism for selecting the optimal BS in mm-wave network based on MAB approach was developed in order to ensure the user has a longer connection time with the BS after HO. The considered the user's post-HO trajectory, and the blockages along the LOS to predict future HO. A RL framework for minimizing frequent HO while satisfying users QoS was proposed in [167] using MAB. The proposed framework takes into account the channel conditions and user QoS requirements before triggering HO. Furthermore, two BS selection algorithms were also developed based on user density for both single-user and multiply user HO scenarios, respectively. An intelligent HO decision framework for BS selection was proposed in [97]. The proposed framework uses a double DRL (DDRL) algorithm to learn the optimal BS for user association in order to minimize the number of HOs and optimize the average throughput along the user trajectory. A distributed learning framework for HO optimization in dense mm-wave networks was proposed in [168] and [169] in

order to minimize frequent HO and optimize user throughput. The framework employs MARL where each user was modeled as an agent and takes an independent HO decision based on its local observation, thereby reducing signaling overhead.

The authors in [170] introduced a HO optimization algorithm based on RL for 5G systems. They modeled the HO problem as a CMAB, then developed a Q -learning solution. In [171], the authors proposed a deep learning model for user localization and proactive HO management, while considering user behaviour in the network. The proposed model uses the received signal measurements to reduce the number of unnecessary HOs and predict the user location while ensuring that the throughput of the network is maintained. A joint optimization framework for minimizing HO frequency and maximizing user throughput was proposed in [172]. The HO and power allocation problem was modeled as a cooperative multi-agent task, after which a MARL framework using proximal policy optimization (PPO) was developed. The model training was performed in a centralized manner after which decentralized policies were obtained for each user. The authors in [173], proposed a learning framework that jointly optimizes HO and beamforming for mm-wave networks. RL algorithm was employed to determine the optimal backup BSs along user trajectory that will help reduce the overhead signaling during channel estimation for user association and minimize the number of HOs. This would ensure an enhanced data rate along the user trajectory.

A learning-based load balancing HO mechanism was proposed in [95]. The user association problem was modeled as a non-convex optimization problem, after which a deep deterministic policy gradient (DDPG) RL algorithm was applied to solve the optimization problem. The algorithm's goal is to associate all the users in different trajectories in the network environment to the optimal BSs in such a way that maximizes their sum rate as well as reduces the number of HO occurrences. The authors in [174] exploit user-centric information to predict user future content request and mobility pattern, after which the optimal user association with UAVs, UAVs' position and content to cache at the UAVs were determined. The goal of their work was to enhance the QoE of the users while reducing the UAV's transmission power. A ML framework using echo state networks was proposed to predict the user future content requests and mobility pattern after which analytical derivations of the optimal UAV locations and contents to cache at the UAV were performed. In [175], a joint optimization framework was proposed for both resource and cache management over licensed and unlicensed for UAV networks. To solve the optimization problem, a liquid state machine learning algorithm was developed to predict the distribution of the user content as well as to enable the UAV select the optimal resource allocation strategy for serving user requests. In addition, a closed form expression was derived to determine the optimal user content to cache and optimal resource allocation.

VI. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Although many studies have been carried out to address the issue of HO management in 5G specifically for mm-wave applications, many significant challenges still needs to be addressed. In this section, we briefly highlight some of the challenges associated with the application of ML techniques for HO management in 5G and present future research directions.

A. DATA SET AVAILABILITY

ML-based implementations rely on the availability of sufficient¹² and quality data¹³ for model training. However, ML based mobility and HO optimization require data set containing user mobility history which is usually very difficult to obtain due to various data protection regulations [118]. Hence, synthetic data via network simulations are normally used for model training. Also, there is the issue of data uniformity where data set generated cannot be used across different platforms. Hence, there is a need to create quality data sets that can be used as benchmarks to assess the accuracy of different ML models that are being proposed for mobility predictions and HO optimization in order to verify their authenticity.

B. PRIVACY AND SECURITY

Mobile service providers are typically responsible for protecting the privacy of their customers. As a result, it is very difficult to release the complete and quality data sets from mobile networks without revealing the personal identity of the users and compromising their privacy. Moreover, ML model security is another issue that should be considered as deep learning models can be subject to adversarial attack. These attacks often inject fake data set to the training data set, thereby reducing model accuracy and resulting in sub-optimal network performance [178] and [179]. More researches need to be conducted on how to properly anonymise data sets from mobile operators in other to prevent vital user information from being revealed while retaining the relevant features of the data set. There is also a need for more research on how to secure deep learning model from adversarial attack that seeks to undermine their accuracy. In addition, more privacy-preserving ML algorithms such as federated learning needs to developed and employed for mobility management and HO optimization in order to guarantee the security and secrecy of user information.

C. GENERALIZATION OF THE ML MODEL

Generalization is the ability of a ML model to learn from seen data and be able to predict the unseen data. Nevertheless, it is not always clear if the trained model is truly generalized since it is difficult to determine if the data set that was used

¹²Sufficiency here means that the available data set should be large enough to enable proper training of the ML model.

¹³Quality here means that the data set must be free from missing entries, duplicated entries or any form of noise that may hinder the ML model from accurately learning from the data set.

for model training captured all the environment features and parameters such that the model would have been exposed to all these features during model training for effective generalization to happen. Therefore, it is essential to ensure that when generating synthetic data set or obtaining real data sets for mobility predictions and HO optimization that all the features in the environment are adequately represented in the data set in order to enhance generalization of the ML models.

D. OFFLINE VERSUS ONLINE LEARNING

Due to the large dimension of 5G and B5G networks as well as the large number of parameters that needs to be learnt during mobility prediction and HO optimization, network designers often have to resort to offline training to reduce both time and space complexity. The successful implementation of offline trained model depends on the adaptability and generalization ability of the model. However, HO optimization often require real-time training and decision making. Hence, there is a need to reduce the number of parameters that needs to be trained by employing clustering method [120] and the use of hardware acceleration [180] to facilitate the ML training process. There might also be a need for both offline and online learning where the model goes through a periodic update and refinement during real time implementation.

E. CENTRALIZED VS DISTRIBUTED DEPLOYMENT

ML models can either be implemented centrally or in a distributed manner, depending on the network configuration with each having its advantage or disadvantage. On the one hand, the advantage of decentralized implementation is low signaling overhead and lesser computation. At the same time, it faces the challenge of inaccurate network optimization due to localized or lack of global network information. On the other hand, the centralized learning case has a global information of the environment and is able to perform a coordinated and collaborative learning that leads to global network optimization. However, it results in massive signaling and computation overhead due to periodic data collection as well as end-to-end delays. Hence, there must be trade-off considerations between global accuracy and huge overhead [181]. With increased network dimensions, complexity and heterogeneous UEs in 5G and B5G networks, it would be more suitable to implement decentralized ML approaches for mobility management and HO optimization as they can help preserve user privacy, reduce latency and communication overhead and also minimize the energy consumption of UEs [124]. However, the issue of coordination for decentralized learning and the challenges involved in sending the locally trained models from the user devices to the central entity due to imperfect channel conditions [126], requires that more research and investigations needs to be carried out on how to effectively implement decentralized ML approaches for mobility management and HO optimization in 5G and B5G networks.

F. FREQUENT HANDOVER

The deployment of large number of mm-wave small cells (due to the short transmission distance of mm-wave signals) to serve the traffic demands of the growing number of UEs would result in increased and unnecessary HOs as well as HO failures. More frequent HO occurrences results in increased signalling overhead, reduced QoS of UEs and increased device power consumption [182]. Therefore, more improved models needs to be developed to reduce the number of HO occurrences as well as optimize the procedure of the HO decision-making process.

G. SIGNALING OVERHEAD

The increase in the number of UEs as well as the large deployment of mm-wave BSs would make the UEs more prone to frequent HOs. The process of HO usually involves the transmission of packets which results in increased signaling overhead in the network. The more the number of HO occurrences, the higher the signaling overhead. Signaling overhead during HO leads to interruptions in data transmission which reduces the user throughput and increases latency [183]. Newer approaches needs to be developed to reduce the number of HOs and shorten data interruption time due to transmission of HO signaling messages by optimizing HO parameters and eliminating the concept of area notification.

H. DEVICE POWER CONSUMPTION

The process of HO in mm-wave communications requires that the UEs makes intra-frequency or inter-frequency measurements depending on its carrier frequency and that of the BS. These measurements increases as the number of mm-wave BSs increases, thereby resulting in increased power consumption for the UEs [184]. This power consumption can be minimized by reducing the measurement gaps and ensuring that the UEs take measurements within the DRX cycle. The smaller, the DRX cycle, the lesser the power consumption [100]. Hence, mobility and HO management schemes that can reduce the DRX cycle needs to be developed in order to optimize the device power consumption.

I. LOAD BALANCING

The uneven distribution of UEs within the network due to random cell positioning and UEs mobility makes some cells to become more loaded than others as a result of more UEs associating with those cells than others. This load imbalance among the cells causes frequent HO and degradation in the QoS of the UEs [94]. Also, in an attempt to prevent frequent HO, most of the HO optimization methods proposed in literature suggest HO skipping or prolonged user connection to a BS [185], which could lead to load imbalance in the network. Therefore, further research needs to be conducted on the effects of the proposed HO optimization schemes on the load balance of the network in order to ensure the QoS of UEs and minimize network congestion.

VII. CONCLUSION

HO management has already been one of the main issues in cellular networks, and is envisioned to be more critical with the introduction of 5G networks due to the prospective capacity enhancement technologies. ML has been quite pervasive in numerous domains, including healthcare, agriculture, disaster prevention, etc., and it has become a reality in 5G networks with proven capabilities in terms of effectiveness and efficiency. Besides, almost all the visionary works that attempt to draw a framework for 6G network foresee that ML will lie at the heart of 6G. In this survey, we first tried to take a snapshot of the current status of cellular communication networks, and then gave a comprehensive tutorial in both mobility and HO management in 5G after elaborating some distinctive characteristics of 5G networks. After that, the major ML branches, namely supervised, unsupervised, and RL (RL), were introduced, and their applications to HO management process were presented. An extensive review on the recent studies on ML-aided HO management techniques was provided under a novel classification that is based on the source of the data for ML implementations. Lastly, the challenges that can be faced while incorporating ML into HO management procedure were identified and thoroughly discussed, followed by a discussion of future research directions.

Although there are multiple survey papers available in the literature reviewing mobility and/or HO management in 5G networks, this is a unique attempt to solely focus on ML applications to HO management. Further, the scope of the paper was kept limited to HO management in order to analyze the topic in a concise and clear manner. The state-of-the-art was reviewed which encompasses the most recent studies in order to demonstrate the current research focus within the community and to keep the readers up-to-date with the most relevant and timely data. In addition, we provided a novel taxonomy about the source of the data and visual data aided HO management techniques, which has been overlooked in the existing survey papers, was included in the discussions along with the traditional wireless data driven applications. With this, we aimed to divert the research focus from conventional approaches to the visual data sources due to the promising potentials of utilizing them in HO management. We also included a discussion on how intelligent HO management can be helpful in emergency scenarios, where there could be mobile clinics, ambulances, remote hospitals, etc. This is quite important on its own considering the current COVID-19 pandemic, which created a chaos in the world and put the need for intelligent and remote control under the spotlight.

REFERENCES

- [1] J. Zhang and K. B. Letaief, "Mobile edge intelligence and computing for the Internet of vehicles," *Proc. IEEE*, vol. 108, no. 2, pp. 246–261, Feb. 2020.
- [2] *Ericsson Mobility Report*, Ericsson, Stockholm, Sweden, Jun. 2020.
- [3] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

- [4] S. A. Busari, S. Mumtaz, S. Al-Rubaye, and J. Rodriguez, "5G millimeter-wave mobile broadband: Performance and challenges," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 137–143, Jun. 2018.
- [5] M. Tayyab, X. Gelabert, and R. Jäntti, "A survey on handover management: From LTE to NR," *IEEE Access*, vol. 7, pp. 118907–118930, 2019.
- [6] *5G: Study on Scenarios and Requirements for Next Generation Access Technologies*, document TS 38.913, 3GPP, Sep. 2018.
- [7] M. H. Lashari, A. A. Memon, S. A. A. Shah, K. Neni, and F. Shafiqat, "IoT based poultry environment monitoring system," in *Proc. IEEE Int. Conf. Internet Things Intell. Syst. (IOTAIS)*, Nov. 2018, pp. 1–5.
- [8] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The Internet of Things for health care: A comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.
- [9] D. Pal, B. Papatratorn, W. Chutimaskul, and S. Funilkul, "Embracing the smart-home revolution in Asia by the elderly: An end-user negative perception modeling," *IEEE Access*, vol. 7, pp. 38535–38549, 2019.
- [10] Y. Meng, W. Zhang, H. Zhu, and X. S. Shen, "Securing consumer IoT in the smart home: Architecture, challenges, and countermeasures," *IEEE Wireless Commun.*, vol. 25, no. 6, pp. 53–59, Dec. 2018.
- [11] D. Mocrii, Y. Chen, and P. Musilek, "IoT-based smart homes: A review of system architecture, software, communications, privacy and security," *Internet Things*, vols. 1–2, pp. 81–98, Sep. 2018.
- [12] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.
- [13] L. D. Xu, W. He, and S. Li, "Internet of Things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.
- [14] S. Lee, M. Bae, and H. Kim, "Future of IoT networks: A survey," *Appl. Sci.*, vol. 7, no. 10, p. 1072, Oct. 2017.
- [15] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the Internet of Things: Communication technologies and challenges," *IEEE Access*, vol. 6, pp. 3619–3647, 2018.
- [16] M. Öztürk, "Cognitive networking for next generation of cellular communication systems," M.S. thesis, James Watt School Eng., Univ. Glasgow, Glasgow, Scotland, 2020.
- [17] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.
- [18] F. Jameel, F. Khan, M. A. A. Haider, and A. A. Butt, "Massive MIMO: A survey of recent advances, research issues and future directions," in *Proc. Int. Symp. Recent Adv. Electr. Eng. (RAEE)*, Oct. 2017, pp. 1–6.
- [19] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks—With a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.
- [20] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017.
- [21] R. Arshad, H. ElSawy, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, "Cooperative handover management in dense cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [22] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of machine learning in wireless networks: Key techniques and open issues," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3072–3108, 4th Quart., 2019.
- [23] M. E. M. Cayamcela and W. Lim, "Artificial intelligence in 5G technology: A survey," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2018, pp. 860–865.
- [24] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 3rd Quart., 2019.
- [25] Y. Lu and X. Zheng, "6G: A survey on technologies, scenarios, challenges, and the related issues," *J. Ind. Inf. Integr.*, vol. 19, Sep. 2020, Art. no. 100158.
- [26] A. Dogra, R. K. Jha, and S. Jain, "A survey on beyond 5G network with the advent of 6G: Architecture and emerging technologies," *IEEE Access*, early access, Oct. 15, 2020, doi: 10.1109/ACCESS.2020.3031234.
- [27] Y. Koda, K. Nakashima, K. Yamamoto, T. Nishio, and M. Morikura, "Cooperative sensing in deep RL-based image-to-decision proactive handover for mmWave networks," in *Proc. IEEE 17th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2020, pp. 1–6.
- [28] S. Sonmez, I. Shayea, S. A. Khan, and A. Alhammadi, "Handover management for next-generation wireless networks: A brief overview," in *Proc. IEEE Microw. Theory Techn. Wireless Commun. (MTTW)*, Oct. 2020, pp. 35–40.
- [29] S. M. A. Zaidi, M. Manalastas, H. Farooq, and A. Imran, "Mobility management in emerging ultra-dense cellular networks: A survey, outlook, and future research directions," *IEEE Access*, vol. 8, pp. 183505–183533, 2020.
- [30] P. P. Kumar and K. Sagar, "A relative survey on handover techniques in mobility management," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 594, Sep. 2019, Art. no. 012027.
- [31] K. Tekbük, A. R. Ekti, G. K. Kurt, and A. Görçin, "Terahertz band communication systems: Challenges, novelties and standardization efforts," *Phys. Commun.*, vol. 35, Aug. 2019, Art. no. 100700.
- [32] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2392–2431, 4th Quart., 2017.
- [33] H. Zhang and L. Dai, "Mobility prediction: A survey on state-of-the-art schemes and future applications," *IEEE Access*, vol. 7, pp. 802–822, 2019.
- [34] E. Gures, I. Shayea, A. Alhammadi, M. Ergen, and H. Mohamad, "A comprehensive survey on mobility management in 5G heterogeneous networks: Architectures, challenges and solutions," *IEEE Access*, vol. 8, pp. 195883–195913, 2020.
- [35] R. Ahmad, E. A. Sundararajan, and A. Khalifeh, "A survey on femtocell handover management in dense heterogeneous 5G networks," *Telecommun. Syst.*, vol. 75, no. 4, pp. 481–507, Oct. 2020.
- [36] A. N. Kasim, "A survey mobility management in 5G networks," Jun. 2020, *arXiv:2006.15598*. [Online]. Available: <http://arxiv.org/abs/2006.15598>
- [37] A. Jain, E. Lopez-Aguilera, and I. Demirkol, "Are mobility management solutions ready for 5G and beyond?" *Comput. Commun.*, vol. 161, pp. 50–75, Sep. 2020.
- [38] *5G/NR; Overall Description: Stage-2*, document TS 38.300, 3GPP, Version 15.9.0, Apr. 2020.
- [39] A. A. R. Alsaedy and E. K. P. Chong, "Tracking area update and paging in 5G networks: A survey of problems and solutions," *Mobile Netw. Appl.*, vol. 24, no. 2, pp. 578–595, Apr. 2019.
- [40] *LTE; General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access*, document TS 23.401, 3GPP, Version 14.4.0, Jul. 2017.
- [41] *5G; System Architecture for the 5G System (5GS)*, document TS 23.501, 3GPP (3GPP), Version 15.5.0, Apr. 2019.
- [42] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.
- [43] E. Dahlman, S. Parkvall, and J. Sköld, "Radio-interface architecture," in *5G NR: The Next Generation Wireless Access Technology*, E. Dahlman, S. Parkvall, and J. Sköld, Eds. New York, NY, USA: Academic, 2018, ch. 6, pp. 73–102.
- [44] Z. Ren, S. Wu, and A. Zhao, "Coexist design of sub-6 GHz and millimeter-wave antennas for 5G mobile terminals," in *Proc. Int. Symp. Antennas Propag. (ISAP)*, Oct. 2018, pp. 1–2.
- [45] R. Chataut and R. Akl, "Massive MIMO systems for 5G and beyond networks—Overview, recent trends, challenges, and future research direction," *Sensors*, vol. 20, no. 10, p. 2753, May 2020.
- [46] K. M. S. Huq, S. A. Busari, J. Rodriguez, V. Frascolla, W. Bazzi, and D. C. Sicker, "Terahertz-enabled wireless system for beyond-5G ultra-fast networks: A brief survey," *IEEE Netw.*, vol. 33, no. 4, pp. 89–95, Jul. 2019.
- [47] A. Schumacher, R. Merz, and A. Burg, "3.5 GHz coverage assessment with a 5G testbed," in *Proc. IEEE 89th Veh. Technol. Conf. (VTC-Spring)*, Apr. 2019, pp. 1–6.
- [48] *5G Deployment Below 6 GHz Ubiquitous Coverage for Critical Communication and Massive IoT*, NOKIA, Espoo, Finland, 2017.
- [49] R. Chen, W.-X. Long, G. Mao, and C. Li, "Development trends of mobile communication systems for railways," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3131–3141, 4th Quart., 2018.
- [50] G. Yue, D. Yu, L. Cheng, Q. Lv, Z. Luo, Q. Li, J. Luo, and X. He, "Millimeter-wave system for high-speed train communications between train and trackside: System design and channel measurements," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 11746–11761, Dec. 2019.

- [51] M. L. Attiah, A. A. M. Isa, Z. Zakaria, M. K. Abdulhameed, M. K. Mohsen, and I. Ali, "A survey of mmWave user association mechanisms and spectrum sharing approaches: An overview, open issues and challenges, future research trends," *Wireless Netw.*, vol. 26, no. 4, pp. 2487–2514, May 2020.
- [52] 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.
- [53] R. Arshad, H. Elsawy, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, "Handover management in 5G and beyond: A topology aware skipping approach," *IEEE Access*, vol. 4, pp. 9073–9081, 2016.
- [54] L. Yan, H. Ding, L. Zhang, J. Liu, X. Fang, Y. Fang, M. Xiao, and X. Huang, "Machine learning-based handovers for sub-6 GHz and mmWave integrated vehicular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4873–4885, Oct. 2019.
- [55] M. Ali, S. Mumtaz, S. Qaisar, and M. Naeem, "Smart heterogeneous networks: A 5G paradigm," *Telecommun. Syst.*, vol. 66, no. 2, pp. 311–330, Oct. 2017.
- [56] A. M. Hamed and R. K. Rao, "Spatial spectral and energy efficiencies of cellular networks limited by co-channel interference and path loss in Nakagami-m fading environment," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, Dec. 2018, Art. no. 118.
- [57] L. F. Ibrahim, H. A. Salman, Z. F. Taha, N. Akkari, G. Aldabbagh, and O. Bamasak, "A survey on heterogeneous mobile networks planning in indoor dense areas," *Pers. Ubiquitous Comput.*, vol. 24, no. 4, pp. 487–498, Aug. 2020.
- [58] W. Yu, H. Xu, H. Zhang, D. Griffith, and N. Golmie, "Ultra-dense networks: Survey of state of the art and future directions," in *Proc. 25th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Aug. 2016, pp. 1–10.
- [59] M. A. Adedoyin and O. E. Falowo, "Combination of ultra-dense networks and other 5G enabling technologies: A survey," *IEEE Access*, vol. 8, pp. 22893–22932, 2020.
- [60] T. Yilmaz, G. Gokkoca, and B. O. Akan, "Millimetre wave communication for 5G IoT applications," in *Internet of Things (IoT) in 5G Mobile Technologies*. Cham, Switzerland: Springer, 2016, pp. 37–53.
- [61] N. N. Srinidhi, S. M. D. Kumar, and K. R. Venugopal, "Network optimizations in the Internet of Things: A review," *Eng. Sci. Technol., Int. J.*, vol. 22, no. 1, pp. 1–21, Feb. 2019.
- [62] A. N. Uwaechia and N. M. Mahyuddin, "A comprehensive survey on millimeter wave communications for fifth-generation wireless networks: Feasibility and challenges," *IEEE Access*, vol. 8, pp. 62367–62414, 2020.
- [63] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs, "LTE release 12 and beyond [accepted for open call]," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 154–160, Jul. 2013.
- [64] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, 4th Quart., 2014.
- [65] W. K. Lai, C.-S. Shieh, F.-S. Chou, C.-Y. Hsu, and M.-H. Shen, "Handover management for D2D communication in 5G networks," *Appl. Sci.*, vol. 10, no. 12, p. 4409, Jun. 2020.
- [66] P. Gandotra and R. K. Jha, "Device-to-device communication in cellular networks: A survey," *J. Netw. Comput. Appl.*, vol. 71, pp. 99–117, Aug. 2016.
- [67] I. Abbasi and A. S. Khan, "A review of vehicle to vehicle communication protocols for VANETs in the urban environment," *Future Internet*, vol. 10, no. 2, p. 14, Jan. 2018.
- [68] K. Z. Ghafoor, M. Guizani, L. Kong, H. S. Maghdid, and K. F. Jasim, "Enabling efficient coexistence of DSRC and C-V2X in vehicular networks," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 134–140, Apr. 2020.
- [69] A. A. Ahmed and A. A. Alzahrani, "A comprehensive survey on handover management for vehicular ad hoc network based on 5G mobile networks technology," *Trans. Emerg. Telecommun. Technol.*, vol. 30, no. 3, Mar. 2019, Art. no. e3546.
- [70] C. Briso-Rodríguez, K. Guan, Y. Xuefeng, and T. Kürner, "Wireless communications in smart rail transportation systems," *Wireless Commun. Mobile Comput.*, vol. 2017, pp. 1–10, Jan. 2017.
- [71] G. Liu and D. Jiang, "5G: Vision and requirements for mobile communication system towards year 2020," *Chin. J. Eng.*, vol. 2016, pp. 1–8, Apr. 2016.
- [72] S. Mumtaz, J. M. Jornet, J. Aulin, W. H. Gerstacker, X. Dong, and B. Ai, "Terahertz communication for vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5617–5625, Jul. 2017.
- [73] NR; User Equipment (UE) Procedures in Idle Mode and in RRC Inactive State, document TS 38.304, 3GPP, Version 15.1.0, Oct. 2018.
- [74] 5G/NR; Requirements for Support of Radio Resource Management, document TS 38.133, 3GPP, Version 15.2.0, Jul. 2018.
- [75] 5G/NR; Radio Resource Control (RRC); Protocol Specification, document TS 38.331, 3GPP, Version 15.3.0, Oct. 2018.
- [76] S. Ahmadi, "New radio access layer 2/3 aspects and system operation," in *5G NR*, S. Ahmadi, Ed. New York, NY, USA: Academic, 2019, ch. 2, pp. 195–284.
- [77] J. Kim, D. Kim, and S. Choi, "3GPP SA2 architecture and functions for 5G mobile communication system," *ICT Exp.*, vol. 3, no. 1, pp. 1–8, Mar. 2017.
- [78] I. L. Da Silva, G. Mildh, M. Säily, and S. Hailu, "A novel state model for 5G radio access networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, May 2016, pp. 632–637.
- [79] J.-M. Liang, P.-Y. Chang, J.-J. Chen, C.-F. Huang, and Y.-C. Tseng, "Energy-efficient DRX scheduling for D2D communication in 5G networks," *J. Netw. Comput. Appl.*, vol. 116, pp. 53–64, Aug. 2018.
- [80] E. Dahlman, S. Parkvall, and J. Sköld, "Scheduling," in *5G NR: The Next Generation Wireless Access Technology*, E. Dahlman, S. Parkvall, and J. Sköld, Eds. New York, NY, USA: Academic, 2018, ch. 14, pp. 275–299.
- [81] S. Ahmadi, "Radio resource control functions," in *LTE-Advanced*, S. Ahmadi, Ed. New York, NY, USA: Academic, 2014, ch. 5, pp. 227–287.
- [82] LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description; Stage 2, document TS 36.300, 3GPP, Version 15.3.0, Oct. 2018.
- [83] M. Hamdan, E. Hassan, A. Abdelaziz, A. Elhigazi, B. Mohammed, S. Khan, A. V. Vasilakos, and M. N. Marsono, "A comprehensive survey of load balancing techniques in software-defined network," *J. Netw. Comput. Appl.*, vol. 174, Jan. 2021, Art. no. 102856.
- [84] S. Ahmadi, "5G network architecture," in *5G NR*, S. Ahmadi, Ed. New York, NY, USA: Academic, 2019, ch. 1, pp. 1–194.
- [85] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 119–125, Dec. 2018.
- [86] M. Lauridsen, L. C. Gimenez, I. Rodriguez, T. B. Sorensen, and P. Mogensen, "From LTE to 5G for connected mobility," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 156–162, Mar. 2017.
- [87] Y. Li, E. Pateromichelakis, N. Vucic, J. Luo, W. Xu, and G. Caire, "Radio resource management considerations for 5G millimeter wave backhaul and access networks," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 86–92, Jun. 2017.
- [88] B. V. Quang, R. V. Prasad, and I. Niemegeers, "A survey on handoffs—Lessons for 60 GHz based wireless systems," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 1, pp. 64–86, 1st Quart., 2012.
- [89] L. D. Nguyen, "Resource allocation for energy efficiency in 5G wireless networks," *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, vol. 5, no. 14, Jun. 2018, Art. no. 154832.
- [90] I. Shayeia, M. Ergen, M. H. Azmi, S. A. Çolak, R. Nordin, and Y. I. Daradkeh, "Key challenges, drivers and solutions for mobility management in 5G networks: A survey," *IEEE Access*, vol. 8, pp. 172534–172552, 2020.
- [91] R. Antonioli, G. Parente, C. Silva, D. Sousa, E. Rodrigues, T. Maciel, and F. Cavalcanti, "Dual connectivity for LTE-NR cellular networks: Challenges and open issues," *J. Commun. Inf. Syst.*, vol. 33, no. 1, pp. 282–294, 2018.
- [92] H. Yang, A. Alphones, Z. Xiong, D. Niyato, J. Zhao, and K. Wu, "Artificial-intelligence-enabled intelligent 6G networks," *IEEE Netw.*, vol. 34, no. 6, pp. 272–280, Nov. 2020.
- [93] Y. He, W. Huang, H. Wei, and H. Zhang, "Effect of channel fading and time-to-trigger duration on handover performance in UAV networks," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 308–312, Jan. 2021.
- [94] K. M. Addali, S. Y. B. Melhem, Y. Khamayseh, Z. Zhang, and M. Kadoch, "Dynamic mobility load balancing for 5G small-cell networks based on utility functions," *IEEE Access*, vol. 7, pp. 126998–127011, 2019.
- [95] S. Khosravi, H. S. Ghadikolaei, and M. Petrova, "Learning-based load balancing handover in mobile millimeter wave networks," Nov. 2020, *arXiv:2011.01420*. [Online]. Available: <http://arxiv.org/abs/2011.01420>
- [96] P. Zhang, J. Lu, Y. Wang, and Q. Wang, "Cooperative localization in 5G networks: A survey," *ICT Exp.*, vol. 3, no. 1, pp. 27–32, Mar. 2017.

- [97] M. S. Mollel, A. I. Abubakar, M. Ozturk, S. Kaijage, M. Kisangiri, A. Zoha, M. A. Imran, and Q. H. Abbasi, "Intelligent handover decision scheme using double deep reinforcement learning," *Phys. Commun.*, vol. 42, Oct. 2020, Art. no. 101133.
- [98] M. S. Mollel, S. Kaijage, and M. Kisangiri, "Deep reinforcement learning based handover management for millimeter wave communication," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, p. 786, 2021.
- [99] Z. Pan, M. Saito, J. Liu, and S. Shimamoto, "P-persistent energy-aware handover decisions employing RF fingerprint for adaptive-sized heterogeneous cellular networks," *IEEE Access*, vol. 7, pp. 52929–52944, 2019.
- [100] P. Skrimponis, S. Dutta, M. Mezzavilla, S. Rangan, S. H. Mirfarshbafan, C. Studer, J. Buckwalter, and M. Rodwell, "Power consumption analysis for mobile mmWave and sub-THz receivers," in *Proc. 2nd 6G Wireless Summit (6G SUMMIT)*, Mar. 2020, pp. 1–5.
- [101] A. Alkhateeb, I. Beltafy, and S. Alex, "Machine learning for reliable mmWave systems: Blockage prediction and proactive handoff," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 1055–1059.
- [102] Y. Oguma, T. Nishio, K. Yamamoto, and M. Morikura, "Proactive handover based on human blockage prediction using RGB-D cameras for mmWave communications," *IEICE Trans. Commun.*, vol. E99.B, no. 8, pp. 1734–1744, 2016.
- [103] K. Qi, T. Liu, and C. Yang, "Federated learning based proactive handover in millimeter-wave vehicular networks," in *Proc. 15th IEEE Int. Conf. Signal Process. (ICSP)*, vol. 1, Dec. 2020, pp. 401–406.
- [104] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, I. Chih-Lin, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.
- [105] Y. Kim, H.-Y. Lee, P. Hwang, R. K. Patro, J. Lee, W. Roh, and K. Cheun, "Feasibility of mobile cellular communications at millimeter wave frequency," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 589–599, Apr. 2016.
- [106] 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.
- [107] M. A. Usman, N. Y. Philip, and C. Politis, "5G enabled mobile healthcare for ambulances," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2019, pp. 1–6.
- [108] A. I. Abubakar, K. G. Omeke, M. Ozturk, S. Hussain, and M. A. Imran, "The role of artificial intelligence driven 5G networks in COVID-19 outbreak: Opportunities, challenges, and future outlook," *Frontiers Commun. Netw.*, vol. 1, p. 4, Nov. 2020.
- [109] H. Tabassum, M. Salehi, and E. Hossain, "Fundamentals of mobility-aware performance characterization of cellular networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2288–2308, 3rd Quart., 2019.
- [110] M. Ozturk, M. Gogate, O. Onireti, A. Adeel, A. Hussain, and M. A. Imran, "A novel deep learning driven, low-cost mobility prediction approach for 5G cellular networks: The case of the control/data separation architecture (CDSA)," *Neurocomputing*, vol. 358, pp. 479–489, Sep. 2019.
- [111] T. Wang, S. Wang, and Z.-H. Zhou, "Machine learning for 5G and beyond: From model-based to data-driven mobile wireless networks," *China Commun.*, vol. 16, no. 1, pp. 165–175, Jan. 2019.
- [112] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1790–1821, 3rd Quart., 2017.
- [113] F. Hussain, S. A. Hassan, R. Hussain, and E. Hossain, "Machine learning for resource management in cellular and IoT networks: Potentials, current solutions, and open challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1251–1275, 2nd Quart., 2020.
- [114] T. Nishio, Y. Koda, J. Park, M. Bennis, and K. Doppler, "When wireless communications meet computer vision in beyond 5G," Oct. 2020, *arXiv:2010.06188*. [Online]. Available: <http://arxiv.org/abs/2010.06188>
- [115] Y. Tian, G. Pan, and M.-S. Alouini, "Applying deep-learning-based computer vision to wireless communications: Methodologies, opportunities, and challenges," Jun. 2020, *arXiv:2006.05782*. [Online]. Available: <http://arxiv.org/abs/2006.05782>
- [116] A. Dey, "Machine learning algorithms: A review," *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 3, pp. 1174–1179, 2016.
- [117] K. Sultan, H. Ali, and Z. Zhang, "Big data perspective and challenges in next generation networks," *Future Internet*, vol. 10, no. 7, p. 56, Jun. 2018.
- [118] H. Fourati, R. Maaloul, and L. Chaari, "A survey of 5G network systems: Challenges and machine learning approaches," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 2, pp. 385–431, Feb. 2021.
- [119] J. Wang, C. Jiang, H. Zhang, Y. Ren, K.-C. Chen, and L. Hanzo, "Thirty years of machine learning: The road to Pareto-optimal wireless networks," Jan. 2019, *arXiv:1902.01946*. [Online]. Available: <http://arxiv.org/abs/1902.01946>
- [120] Z. Wang, L. Li, Y. Xu, H. Tian, and S. Cui, "Handover control in wireless systems via asynchronous multiuser deep reinforcement learning," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4296–4307, Dec. 2018.
- [121] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, 4th Quart., 2019.
- [122] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.
- [123] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-efficient learning of deep networks from decentralized data," *Proc. Mach. Learn. Res.*, vol. 54, pp. 1273–1282, Apr. 2017.
- [124] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, Jun. 2020.
- [125] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.
- [126] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269–283, Jan. 2021.
- [127] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, "Federated multi-task learning," in *Proc. NIPS*, 2017, pp. 1–11.
- [128] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.
- [129] J. Feng, C. Rong, F. Sun, D. Guo, and Y. Li, "PMF: A privacy-preserving human mobility prediction framework via federated learning," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–21, Mar. 2020.
- [130] A. Li, S. Wang, W. Li, S. Liu, and S. Zhang, "Predicting human mobility with federated learning," in *Proc. 28th Int. Conf. Adv. Geograph. Inf. Syst.*, Nov. 2020, pp. 441–444.
- [131] M. Alrabeiah, A. Hredzak, and A. Alkhateeb, "Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5.
- [132] M. Alrabeiah, J. Booth, A. Hredzak, and A. Alkhateeb, "ViWi vision-aided mmWave beam tracking: Dataset, task, and baseline solutions," Feb. 2020, *arXiv:2002.02445*. [Online]. Available: <http://arxiv.org/abs/2002.02445>
- [133] M. Alrabeiah, A. Hredzak, Z. Liu, and A. Alkhateeb, "ViWi: A deep learning dataset framework for vision-aided wireless communications," Nov. 2019, *arXiv:1911.06257*. [Online]. Available: <http://arxiv.org/abs/1911.06257>
- [134] Y. Koda, K. Nakashima, K. Yamamoto, T. Nishio, and M. Morikura, "Handover management for mmWave networks with proactive performance prediction using camera images and deep reinforcement learning," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 2, pp. 802–816, Jun. 2020.
- [135] A. Klautau, N. González-Prelcic, and R. W. Heath, Jr., "LIDAR data for deep learning-based mmWave beam-selection," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 909–912, Jun. 2019.
- [136] M. Dias, A. Klautau, N. González-Prelcic, and R. W. Heath, Jr., "Position and LIDAR-aided mmWave beam selection using deep learning," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5.
- [137] W. Xu, F. Gao, S. Jin, and A. Alkhateeb, "3D scene-based beam selection for mmWave communications," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1850–1854, Nov. 2020.
- [138] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
- [139] Y. Long, Z. Chen, J. Fang, and C. Tellambura, "Data-driven-based analog beam selection for hybrid beamforming under mm-wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 340–352, May 2018.

- [140] C. Antón-Haro and X. Mestre, "Learning and data-driven beam selection for mmWave communications: An angle of arrival-based approach," *IEEE Access*, vol. 7, pp. 20404–20415, 2019.
- [141] X. Ma, Z. Chen, Z. Li, W. Chen, and K. Liu, "Low complexity beam selection scheme for terahertz systems: A machine learning approach," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2019, pp. 1–6.
- [142] Y. Yang, Y. He, D. He, Z. Gao, and Y. Luo, "Machine learning based analog beam selection for 5G mmWave small cell networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2019, pp. 1–5.
- [143] Y. Yang, Z. Gao, Y. Ma, B. Cao, and D. He, "Machine learning enabling analog beam selection for concurrent transmissions in millimeter-wave V2V communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9185–9189, Aug. 2020.
- [144] C.-H. Lin, W.-C. Kao, S.-Q. Zhan, and T.-S. Lee, "BsNet: A deep learning-based beam selection method for mmWave communications," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–6.
- [145] Y. Wang, A. Klautau, M. Ribero, M. Narasimha, and R. W. Heath, Jr., "MmWave vehicular beam training with situational awareness by machine learning," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–6.
- [146] Z. Li, X. Ma, W. Chen, N. Kuang, and B. Zhang, "Neural network enhanced analog beam selection scheme for terahertz systems," in *Proc. IEEE/CIC Int. Conf. Commun. Workshops China (ICCC Workshops)*, Aug. 2019, pp. 136–141.
- [147] D. Jagyasi and M. Coupechoux, "DNN based beam selection in mmW heterogeneous networks," in *Proc. Int. Conf. Netw. Games, Control Optim. (Netgcoop)*, Cargèse, France, 2020, pp. 1–13.
- [148] M. S. Sim, Y.-G. Lim, S. H. Park, L. Dai, and C.-B. Chae, "Deep learning-based mmWave beam selection for 5G NR/6G with sub-6 GHz channel information: Algorithms and prototype validation," *IEEE Access*, vol. 8, pp. 51634–51646, 2020.
- [149] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5504–5518, Sep. 2020.
- [150] C. Bian, Y. Yang, F. Gao, and G. Ye Li, "FusionNet: Enhanced beam prediction for mmWave communications using sub-6 GHz channel and a few pilots," Sep. 2020, *arXiv:2009.02655*. [Online]. Available: <http://arxiv.org/abs/2009.02655>
- [151] M. B. Booth, V. Suresh, N. Michelusi, and D. J. Love, "Multi-armed bandit beam alignment and tracking for mobile millimeter wave communications," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1244–1248, Jul. 2019.
- [152] A. Asadi, S. Müller, G. H. Sim, A. Klein, and M. Hollick, "FML: Fast machine learning for 5G mmWave vehicular communications," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2018, pp. 1961–1969.
- [153] G. H. Sim, S. Klos, A. Asadi, A. Klein, and M. Hollick, "An online context-aware machine learning algorithm for 5G mmWave vehicular communications," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2487–2500, Dec. 2018.
- [154] I. Aykin, B. Akgun, M. Feng, and M. Krunz, "MAMBA: A multi-armed bandit framework for beam tracking in millimeter-wave systems," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Jul. 2020, pp. 1469–1478.
- [155] X. Li, R. Zhou, Y.-J. A. Zhang, L. Jiao, and Z. Li, "Smart vehicular communication via 5G mmWaves," *Comput. Netw.*, vol. 172, May 2020, Art. no. 107173.
- [156] C. Xu, S. Liu, C. Zhang, Y. Huang, and L. Yang, "Joint user scheduling and beam selection in mmWave networks based on multi-agent reinforcement learning," in *Proc. IEEE 11th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2020, pp. 1–5.
- [157] H.-L. Chiang, K.-C. Chen, W. Rave, M. K. Marandi, and G. Fettweis, "Multi-UAV mmWave beam tracking using Q-learning and interference mitigation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–7.
- [158] Y.-J. Chen, W.-Y. Cheng, and L.-C. Wang, "Learning-assisted beam search for indoor mmWave networks," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2018, pp. 320–325.
- [159] N. Van Huynh, D. N. Nguyen, D. T. Hoang, and E. Dutkiewicz, "Optimal beam association for high mobility mmWave vehicular networks: Lightweight parallel reinforcement learning approach," May 2020, *arXiv:2005.00694*. [Online]. Available: <http://arxiv.org/abs/2005.00694>
- [160] R. Wang, O. Onireti, L. Zhang, M. A. Imran, G. Ren, J. Qiu, and T. Tian, "Reinforcement learning method for beam management in millimeter-wave networks," in *Proc. UK/China Emerg. Technol. (UCET)*, Aug. 2019, pp. 1–4.
- [161] S. Rezaei, C. N. Manchón, and E. de Carvalho, "Location- and orientation-aided millimeter wave beam selection using deep learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.
- [162] Y. Wu and B. Ai, "A learning-based beam training method for multi-user millimeter wave systems," in *Proc. IET 8th Int. Conf. Wireless, Mobile Multimedia Netw.*, 2019, pp. 38–42.
- [163] H. P. Tauqir and A. Habib, "Deep learning based beam allocation in switched-beam multiuser massive MIMO systems," in *Proc. 2nd Int. Conf. Latest Trends Electr. Eng. Comput. Technol. (INTELLECT)*, Nov. 2019, pp. 1–5.
- [164] Y. Heng and J. G. Andrews, "Machine learning-assisted beam alignment for mmWave systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.
- [165] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, Jr., "Online learning for position-aided millimeter wave beam training," *IEEE Access*, vol. 7, pp. 30507–30526, 2019.
- [166] L. Sun, J. Hou, and T. Shu, "Optimal handover policy for mmWave cellular networks: A multi-armed bandit approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.
- [167] Y. Sun, G. Feng, S. Qin, Y.-C. Liang, and T.-S.-P. Yum, "The SMART handoff policy for millimeter wave heterogeneous cellular networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 6, pp. 1456–1468, Jun. 2018.
- [168] M. Sana, A. De Domenico, E. C. Strinati, and A. Clemente, "Multi-agent deep reinforcement learning for distributed handover management in dense mmWave networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 8976–8980.
- [169] M. Sana, A. De Domenico, W. Yu, Y. Lohan, and E. C. Strinati, "Multi-agent reinforcement learning for adaptive user association in dynamic mmWave networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6520–6534, Oct. 2020.
- [170] V. Yajnanarayana, H. Rydén, and L. Hevizi, "5G handover using reinforcement learning," in *Proc. IEEE 3rd 5G World Forum (5GWF)*, Sep. 2020, pp. 349–354.
- [171] R. Klus, L. Klus, D. Solomitskii, M. Valkama, and J. Talvitie, "Deep learning based localization and HO optimization in 5G NR networks," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, Jun. 2020, pp. 1–6.
- [172] D. Guo, L. Tang, X. Zhang, and Y.-C. Liang, "Joint optimization of handover control and power allocation based on multi-agent deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13124–13138, Nov. 2020.
- [173] S. Khosravi, H. S. Ghadikolaei, and M. Petrova, "Learning-based handover in mobile millimeter-wave networks," Mar. 2020, *arXiv:2003.11009*. [Online]. Available: <http://arxiv.org/abs/2003.11009>
- [174] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1046–1061, May 2017.
- [175] M. Chen, W. Saad, and C. Yin, "Liquid state machine learning for resource and cache management in LTE-U unmanned aerial vehicle (UAV) networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1504–1517, Mar. 2019.
- [176] L. Sun, J. Hou, and T. Shu, "Spatial and temporal contextual multi-armed bandit handovers in ultra-dense mmWave cellular networks," *IEEE Trans. Mobile Comput.*, early access, Jun. 5, 2020, doi: [10.1109/TMC.2020.3000189](https://doi.org/10.1109/TMC.2020.3000189).
- [177] M. S. Mollel, S. Kaijage, M. Kisangiri, M. A. Imran, and Q. H. Abbasi, "Multi-user position based on trajectories-aware handover strategy for base station selection with multi-agent learning," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.
- [178] M. Barreno, B. Nelson, A. D. Joseph, and J. D. Tygar, "The security of machine learning," *Mach. Learn.*, vol. 81, no. 2, pp. 121–148, 2010.
- [179] J. Suomalainen, A. Juhola, S. Shahabuddin, A. Mämmelä, and I. Ahmad, "Machine learning threatens 5G security," *IEEE Access*, vol. 8, pp. 190822–190842, 2020.
- [180] A. Reuther, P. Michaleas, M. Jones, V. Gadepally, S. Samsi, and J. Kepner, "Survey of machine learning accelerators," Aug. 2020, *arXiv:2009.00993*. [Online]. Available: <http://arxiv.org/abs/2009.00993>
- [181] N. Kato, B. Mao, F. Tang, Y. Kawamoto, and J. Liu, "Ten challenges in advancing machine learning technologies toward 6G," *IEEE Wireless Commun.*, vol. 27, no. 3, pp. 96–103, Jun. 2020.

- [182] T. Bilen, B. Canberk, and K. R. Chowdhury, "Handover management in software-defined ultra-dense 5G networks," *IEEE Netw.*, vol. 31, no. 4, pp. 49–55, Jul. 2017.
- [183] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Predictive and core-network efficient RRC signalling for active state handover in RANs with control/data separation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1423–1436, Mar. 2017.
- [184] S. Hailu, M. Saily, and O. Tirkkonen, "RRC state handling for 5G," *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 106–113, Jan. 2019.
- [185] R. Arshad, H. ElSawy, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, "Handover management in dense cellular networks: A stochastic geometry approach," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–7.



MICHAEL S. MOLLEL (Student Member, IEEE) received the B.Sc. degree in telecommunication engineering from the University of Dar es Salaam, Dar es Salaam, Tanzania, in 2011, the M.Sc. degree in information and communication engineering from The Nelson Mandela African Institution of Science and Technology (NM-AIST), Arusha, Tanzania, in 2014, where he is currently pursuing the Ph.D. degree. His research interests include 5G and beyond 5G mobile networks,

the Internet of Things, artificial intelligence, intelligent computer, and wireless networks. He received the 2019 Best Paper Award in IEEE WCNC 2019 Workshop.



ATTAI IBRAHIM ABUBAKAR received the B.Eng. degree in electrical and electronics engineering from the University of Agriculture, Makurdi, Nigeria, in 2011, and the M.Sc. degree in wireless communication systems from The University of Sheffield, U.K., in 2015. He is currently pursuing the Ph.D. degree with the James Watt School of Engineering, University of Glasgow, U.K. His research interests include energy efficiency in 5G heterogeneous cellular networks,

self-organising networks (SON), and application of machine learning algorithms to wireless communications.



METIN OZTURK received the B.Sc. degree in electrical and electronics engineering from Eskisehir Osmangazi University, Turkey, in 2013, the M.Sc. degree in electronics and communication engineering from Ankara Yildirim Beyazit University, Turkey, in 2016, and the Ph.D. degree from the Communications, Sensing, and Imaging Research Group, James Watt School of Engineering, University of Glasgow, U.K., in 2020. From 2013 to 2016, he worked as a Research Assistant with Ankara Yildirim Beyazit University, where he is currently a Lecturer. His research interests include intelligent networking for wireless communication networks, with a focus on energy efficiency, mobility management, and radio resource management in cellular networks.



SHUBI FELIX KAIJAGE (Senior Member, IEEE) received the Doctor of Engineering degree in electronics and information engineering from the University of the Ryukyus, Okinawa, Japan, in March 2011. From September 2011 to December 2013, he worked as a Postdoctoral Research Fellow with the Research Center of Terahertz Technology, Shenzhen University, China. He currently works as a Senior Lecturer and the Dean of the School of Computational and

Communications Science and Engineering (CoCSE), The Nelson Mandela African Institution of Science and Technology (NM-AIST), Arusha, Tanzania. He has published over 40 scientific articles in international peer-reviewed journals and more than 50 research papers presented in various international conferences. His research interests include optics and photonics, optical fiber and photonic crystal fibers (PCFs), fiber optics communication, terahertz wave technology, radio frequency identification (RFID), the Internet of Things (IoT), and wireless sensor networks. He was a recipient of numerous international awards and grants.



MICHAEL KISANGIRI received the Master of Science and Ph.D. degrees in telecommunication engineering from the Wroclaw University of Technology - Poland, in 2002 and 2008, respectively. Since 2008, he has been working as an Academician with several universities. He currently works with The Nelson Mandela African Institution of Science and Technology, Arusha, Tanzania. He is a Senior Lecturer in electronics and telecommunication. His research interests include evolutionary

computation in telecommunication networks, antenna design and triangular mesh modeling.



SAJJAD HUSSAIN (Senior Member, IEEE) received the master's degree in wireless communications from Supélec, Gif-sur-Yvette, in 2006, and the Ph.D. degree in signal processing and communications from the University of Rennes 1, Rennes, France, in 2009. He has worked as an Associate Professor with the Electrical Engineering Department, Capital University of Science and Technology (CUST), Islamabad, Pakistan. He is currently a Senior Lecturer (Associate Professor)

in electronics and electrical engineering with the University of Glasgow, U.K. He has over ten years of academic experience with roles as a Principal/Co-Investigator in several funded projects. He has authored/coauthored over 60 journal and conference publications, one authored book, and one edited book. His research interests include 5G self-organizing networks, industrial wireless sensor networks, and machine learning for wireless communications and technology-enhanced student learning techniques. He is a Fellow of the Higher Education Academy, and a Senior Fellow Recognising Excellence in Teaching. Since 2019, he has been an Associate Editor of IEEE ACCESS.



MUHAMMAD ALI IMRAN (Senior Member, IEEE) is currently the Dean of the University of Glasgow and UESTC. He is a Professor of wireless communication systems. His research interests include self-organised networks, wireless networked control systems, the Internet of Things (IoT), and wireless sensor systems. He heads the Communications, Sensing and Imaging (CSI) Research Group, University of Glasgow. He is the Director of Glasgow-UESTC Centre for Educa-

tional Development and Innovation. He is an Affiliate Professor with the University of Oklahoma, USA, a Visiting Professor with the 5G Innovation Centre, University of Surrey, U.K., and an Adjunct Research Professor with the Artificial Intelligence Research Center (AIRC), College of Engineering and Information Technology, Ajman University, United Arab Emirates. He has over 20 years of combined academic and industry experience with several leading roles in multi-million-pound-funded projects. He has filed 15 patents. He has authored/coauthored over 400 journal and conference publications. He has edited seven books. He has authored more than 30 book chapters. He has successfully supervised over 40 postgraduate students at the doctoral level. He has been a consultant to international projects and local companies in the area of self-organised networks. He is a Fellow of IET and a Senior Fellow HEA.



QAMMER H. ABBASI (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electronics and telecommunication engineering from the University of Engineering and Technology (UET), Lahore, Pakistan, and the Ph.D. degree in electronic and electrical engineering from the Queen Mary University of London (QMUL), in 2012.

He is currently a Senior Lecturer (Associate Professor) with the James Watt School of Engineering, University of Glasgow, U.K., the Deputy

Head of the Communication Sensing and Imaging Group, and a Research Investigator with the Scotland 5G Center. He has contributed to over 350 leading international technical journal and peer-reviewed conference papers, and ten books. He received several recognitions for his research, which include an appearance on BBC, STV, dawn news, local and international newspapers, the cover of MDPI journal, most downloaded articles, U.K. exceptional talent endorsement by Royal Academy of Engineering, National Talent Pool Award by Pakistan, International Young Scientist Award by NSFC China, URSI Young Scientist Award, National Interest Waiver by the USA, four best paper awards, and best representative image of an outcome by QNRF. He is an Associate Editor of the IEEE JOURNAL OF ELECTROMAGNETICS, *RF And Microwaves in Medicine and Biology*, IEEE SENSORS JOURNAL, IEEE OPEN ACCESS ANTENNA AND PROPAGATION, IEEE INTERNET OF THINGS JOURNAL, and acted as a guest editor for numerous special issues in top-notch journals.

• • •