IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 October 2023, accepted 22 November 2023, date of publication 12 December 2023, date of current version 5 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3341491

RESEARCH ARTICLE

Hydra-RAN Perceptual Networks Architecture: Dual-Functional Communications and Sensing Networks for 6G and Beyond

RAFID I. ABD[1], (Member, IEEE), DANIEL J. FINDLEY[2], (Senior Member, IEEE), AND KWANG SOON KIM[1], (Senior Member, IEEE)

[1]School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea
[2]Department of CCEE, North Carolina State University Centennial Campus, Raleigh, NC 27695, USA

Corresponding author: Kwang Soon Kim (ks.kim@yonsei.ac.kr)

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) under Grant 2021-0-02208, and in part by the National Research Foundation of Korea (NRF) funded by the Korean Government [Ministry of Science and ICT (MSIT)] under Grant 2022R1A5A1027646.

ABSTRACT After researchers devoted considerable efforts to developing 5G standards, their passion began to focus on establishing the basics for the standardization of 6G and beyond. The utilization of millimeter wave (MMW) and terahertz (THz) frequency bands, combined with sensors and artificial intelligence (AI), has gained significant attention in the research community for the development of the next generation of sensory and radio access networks (NG-SRANs). Leveraging the advantages of communication and sensor systems' common characteristics will open horizons for merging the two networks, thereby creating a unified perceptive and intelligence network. Overall, while using MMW and THz frequencies is certainly valuable, the ability to gather and transmit data in real-time makes sensors extremely effective in communication networks. In contrast, AI, machine learning (ML), and deep learning (DL) have become predominant methods for solving data analysis problems across a wide range of domains, such as analyzing large amounts of different sensor data, decision-making, channel estimation, self-organization, and self-healing. This paper proposes a novel design for a potential 6G network and beyond called the Hydra radio access network (H-RAN) perceptual networks architecture, which is designed based on NG-SRAN. From a design perspective, H-RAN aims to merge communication and sensing networks into a single network in which two functionalities are attempted to mutually complement each other, namely communication-aided sensing and sensing-aided communications networks. However, such a network provides an adequate platform for a wide range of AL/ML algorithms, such as real-time decision-making, self-organization, and self-healing. As a result, H-RAN perceptual networks architecture is expected to be more efficient, reliable, and secure than existing conventional networks, and is likely to play a critical role in a wide range of applications, including but not limited to mobile broadband, sensing systems, smart cities, autonomous vehicles, the internet of things (IoT) connectivity, vehicle-to-everything (V2X) communication, etc. This study gives a detailed overview of how H-RAN will revolutionize conventional future sensors and cellular networks through a comprehensive analysis of H-RAN architectural components and functionalities.

INDEX TERMS 6G networks and beyond, integration of sensor and communications networks, dual-functional networks, broad exploitation of AI/ML engines, perceptive networks, sensing/radio access networks (SRANs), self-organization/self-healing/IoT.

I. INTRODUCTION

Among the many assumptions regarding the future vision of 6G networks and beyond, there are a variety of theories, but a common theme among them is that higher frequency

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

bands, sensors, artificial intelligence (AI), machine learning (ML), and the Internet-of-Things (IoT) will have a more crucial role in the next generation of sensory and radio access networks (NG-SRANs) than any previous networks [1], [2], [3], [4]. A speculative study trend indicates that future radio access networks (RANs) will incorporate millimeter wave (MMW)/terahertz (THz) frequency bands, sensing functionality, and AI/ML engines in the upcoming network that can automatically adapt to changes in their environment with the vision of creating self-adaptive and self-healing networks [5], [6], [7]. Moreover, NG-SRANs hold significant potential to enable novel applications and use cases through perceptive networks. To this end, it seems that NG-SRANs will gain significant advantages through the integration of information from several sensor functions. Also, it will allow the efficient use of dense network infrastructure deployment to create intelligent and observable RANs. As a result, recent developments in the open radio access network (O-RAN) [8], [9], [10], [11], [12], [13], [14], [15] coupled with multi-sensor data fusion, and extensive use of AIL/ML workflows, are capable of generating perceptive networks, thereby enabling real-time decision-making, automated network management, and self-learning capabilities.

However, the current cellular protocol stack architecture in RANs has limitations in overall network state visibility and awareness of the underlying conditions [16], [17], [18], [19]. Without a comprehensive understanding of the overall network state, individual layers may struggle to adapt dynamically to changing conditions [18], [19], [20]. For example, if the physical layer detects a significant change in channel conditions, it may not be effectively communicated to higher-layer protocols, leading to inefficiencies in resource allocation and suboptimal performance [23], [24], [25]. Limited awareness of the actual network state can hinder the ability to dynamically adapt and optimize network responses in real-time [24], [25], [26], [27]. To address these limitations, there are ongoing research and development efforts in the field of network architecture and protocol design. These efforts aim to introduce more flexibility in cross-layer coordination, such as network slicing [28], software-defined networking (SDN) [29], and network function virtualization (NFV) [29]. Therefore, to supplement current solutions, we introduce NG-SRANs as cognitive networks to be more intelligent and adaptable, with the capability of automatically detecting and responding to changes in network conditions. Overall, future cognitive networks are expected to extensively employ AI/ML workflows throughout the network components to continuously learn from network data and user behavior, allowing real-time decisions and adapting to changing conditions.

AI has the potential to revolutionize future RANs by making them more intelligent, adaptive, and efficient. It enables networks to cope with the expected growth in complexity and increasing demand for future applications. For example, ML allows systems to learn and improve automatically from their experience without being explicitly programmed. It also performs complex functions such as signal classification [30], blockage prediction [31], waveform design [32], spectrum sensing [33], etc. Among the various ML techniques, DL is one of the most popular ML methods designed to automatically learn and extract hierarchical representations of data, enabling them to recognize patterns and make predictions with high accuracy, such as capturing high-dimensional spectrum representations [34], and determining the optimal beam orientation between nodes [35]. In addition, the challenges imposed by ultra-massive access applications and use cases in terms of energy consumption can be mitigated by utilizing AI-/ML-based strategies. Furthermore, sophisticated intelligent mobility management and resource allocation will ensure service dependability, ultra-reliability, and low-latency applications [5].

Indeed, emerging media technologies, such as holographic communications, involve transmitting three-dimensional (3D) holographic scenes in real-time, which demand extremely high data rates and significantly higher transmission speeds, up to terabits per second, compared to augmented reality (AR) and virtual reality (VR) applications. Thus, it seems 5G networks are far from meeting the demands of these applications, which require large-scale communication systems, super-massive access, ultra-broadband, and faster and more reliable communication [5]. However, due to the limitations of perceiving the surrounding environment, O-RAN based on MMW/THz frequency bands may not be able to provide real-time environment information in a continuous stream of the surrounding environment [23], [24], [25]. Thereupon, MMW/THz O-RAN networks, initial access, beam tracking, handover, blockage avoidance, predicting and avoiding faults, environmental perception, etc., may face various challenges that deserve further research [24], [25], [26], [27]. In addition to this setback is the fact that conventional beam training methods, such as exhaustive beam sweeps (EBS) [36], or feedback-based techniques [37], can be time-consuming and computationally intensive [25], [38]. As a result, it seems that the current conventional network design may have limitations when it comes to operating in the MMW/THZ frequency bands. These limitations primarily stem from an unawareness of the surrounding environment, deployment challenges, protocol stack considerations, beam management, and self-organization/self-healing capabilities, among others.

Given the extensive challenges, in addition to the future vision of what next-generation communications networks will look like, 6G networks and beyond require novel futuristic designs tailored to the characteristics of the MMW/THZ bands and future use cases and applications. To highlight the potential vision for the basic standardization of the 6G network and beyond, we propose a novel H-RAN employing the state-of-the-art concepts of integrated communications and sensing networks (communication-aided sensing and sensing-aided communications networks), with high-scale exploitation of AI/ML workflows. H-RAN architecture aims to provide edge computing by enabling data processing to occur closer to the source of data, reducing latency, and improving real-time processing capabilities, which could provide an effective starting point for being poised to revolutionize the telecom and sensor ecosystem. Additionally, H-RAN network slicing allows telecom and sensor ecosystem providers to create multi-function virtual networks within a single physical infrastructure, each customized for specific applications or user groups, more than any other existing network. This technology will support diverse services, such as autonomous vehicles, augmented reality, security enhancements, etc. Moreover, the H-RAN vision aims to drastically change the design, deployment, and operations of future cellular networks, ultimately enabling embedded intelligence decisions and real-time analytics. In addition to creating cognitive communication networks built by incorporating cognitive NG-SRAN technology, which enables networks and devices to dynamically sense and adapt to make intelligent decisions in response to changing network conditions. Last but not least, future heterogeneous H-RAN networks can be much more intelligent and perceptive by combining real-time/historical sensory data and extensive AI/ML capabilities. In such networks, the surrounding environment can be sensed everywhere, supporting a wide range of IoT services and applications in the future, as well as providing various solutions to address many of the challenges imposed by the MMW/THz frequency bands.

A. RELATED WORK

Thanks to a recent surge in AI/ML and sensor technologies, AI/ML workflows and high-dimensional sensor information are becoming increasingly accessible and are being applied across a wide range of essential applications [1]. Researchers have demonstrated that context information obtained from infrastructure sensors in conjunction with AI/ML algorithms can significantly enhance MMW/THZ 6G wireless in a variety of functions, and address several challenges related to MMW/THZ bands [2], [3]. 6G networks powered by AI/ML are considered a fundamental enabler to provide more powerful and intelligent capabilities to emerging services and applications [4], [5]. The authors of [4] claimed AI/ML techniques have gained significant attention for beam management frameworks in MMW/THz bands due to their capability of extracting and tracking nonlinear environmental characteristics. Therefore, there is a strong need for protocols and standardization activities for an enhanced AI/ML-based beam management platform. The researchers of [5] focused on some solutions for applying AI/ML models to 6G networking and resource management optimization, as well as channel estimation and spectrum management. Various AI/ML algorithms have been introduced to 6G resource allocation techniques (e.g., channels, bandwidth, computing resources, memory, processing power, etc.) to cope with the sophisticated optimization of decision-making by 6G's dynamics, multidimensionality, and random uncertainty. The study in [6] claimed that there could be no truly intelligent system without integrating AI/ML techniques. Hence, the study discussed the next-generation smart grid that leverages revolutionary technologies, such as AI/ML, sensors, and IoT to achieve robust reliability, resilience, and overall system performance. The study indicates that the next-generation smart grid includes unique features including microgrids, smart transmission lines, smart feeders and substations, programmable sensors (AI sensors), and AI-controlled grid management centers, among others. The study in [7] proposes that cognitive radio can be integrated into the current power grid to enable smart communication and decision-making. This study has shown that most power utility establishments around the globe would not only want to make their network smart for easy and convenient data logging and monitoring but would also like to reduce their operating costs. Zhang et al. [39] showed that by slightly modifying current cellular networks, joint communication, and radar/radio sensing capability networks may become more perceptive. Reference [40] asserted that integrating sensing into IoT devices and wireless networks could be accomplished quickly and inexpensively by reusing synchronization and reference signals for sensing patterns. Moving forward, a mechanism for integrating sensing and communication for individual IoT devices and categorizing extensions over existing devices into four aspects, namely time, frequency, space, and protocol has been proposed in [41]. The article in [42] introduces an integrated communications and sensing framework for IoT solutions that establish deep reinforcement learning for the decision process without requiring complete knowledge of the surrounding environment. On the other hand, sensors and DL algorithms have been proposed to address several challenges associated with MMW system actuation in many studies. For instance, studies in [43], [44], and [45] have designed efficient DL methods that leverage radar sensory data to direct MMW beam prediction while greatly reducing beam training overhead. For the same reason, the authors of [46] and [47] used prior information extracted from LIDAR sensory data to remarkably reduce beam training overhead by implementing deep neural networks (DNNs). These approaches maximize throughput with reasonable overhead and computational costs. Koda et al. [48] propose a proactive framework wherein handover timings, camera images, and deep reinforcement learning are used to determine the handover timing. To address the sensitivity of MMW/THZ systems to blockages, the researchers in [49], [50], and [51] utilize sensors to provide sensing information about the surrounding environment. To reduce the search overhead of iterative beam discovery procedures, the authors of [52] and [53] used contextual information from sensors and deep neural networks (DNNs) that improve classification accuracy for beam selection. Moreover, sensory data and DL architecture to address challenges related to MMW beam alignment, beamforming, and beam tracking are considered in [54], [55], and [56], respectively. On the concept of intelligence and learning in O-RAN, Bonati et al. [57] claimed that the O-RAN paradigm will drastically change the design, deployment, and operations of future cellular networks, thereby enabling embedded intelligence and real-time analytics. The study in [58] proposes an intelligent model for traffic congestion and radio resource management. The authors of [59] propose the open RAN closed-loop control on programmable experimental platforms (ColO-RAN), a pipeline for the design, training, and evaluation of DRL-based control loops in O-RAN.

B. CONTRIBUTION

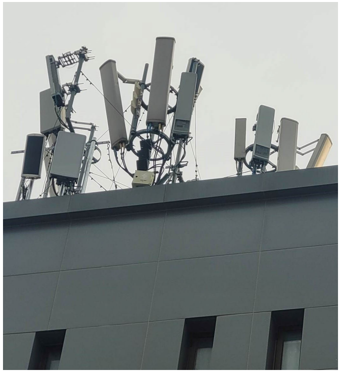
This paper introduces a concept of H-RAN architecture that integrates communication and sensing network functionality into a single network and uses the comprehensive and independent distribution of AI/ML engines for self-optimizing and self-healing strategies based on a perceptive dual-functional wireless network. The main contributions can be summarized as follows. We propose an H-RAN architecture integrating disaggregated sensing and communication systems into four H-RAN components: sensing and radio units (SRUs), Hydra distributed units (H-DU), Hydra central units (H-CU), and Hydra RAN intelligent controllers (H-RICs). Augmented with sensing and extensive AI/ML functionality, future heterogeneous H-RANs' "open eyes" are designed to be more intelligent and perceptive. Such networks, which can sense and perceive the surrounding environment ubiquitously, could serve as the backbone for revolutionizing the communications ecosystem.

II. SYSTEM MODEL

A. IDEOLOGY AND PRINCIPLE

Although MMW/THZ technologies offer many desirable features such as high data rates and capacity [46], [47], [48], their unique characteristics come with their own set of limitations. These limitations include but are not limited to, low penetration efficiency, easy blockage, high path loss, limited communication range, frequent misalignment, etc. [24], [25], [38]. These challenges (among others) restrict MMW/THZ networks' ability to efficiently support mobile applications over 200m away from the source in harsh environments [60], [61]. Therefore, the dense deployment of MMW/THZ networks in cities, such as rooftops, towers, or streetlights is recommended to mitigate some of these challenges [60]. Accordingly, in the real world, the widespread deployment of sub-6 GHz networks in cities is indeed observed. It can be seen in various structures as illustrated in Fig. 1-(a), (b), (c), and (e). However, while current sub-6 GHz network deployment can address some challenges, it also comes with several disadvantages that need to be carefully considered. For instance, as shown in Fig. 1-(a), a large number of antennas are located at one location, thus this may reduce the network's ability to perceive the surrounding environment, increase interference, limit coverage range, reduce the probability of line-of-sight (LOS), power consumption, etc. Meanwhile, while strolling through cities, a clear sense of reality sinks in the presence of densely deployed sensing networks throughout various areas as seen in Fig. 1-(b), (c), (d), (e), and (f). This network consists of a multitude of sensors strategically placed to monitor and collect data across different parts of the city. Fig. 1-(b), (c), (d), (e), and (f) illustrate a real-world widespread sensor network installation in Seoul city. They appear omnipresent by monitoring parks, streets, alleys, subways, universities, shopping malls, etc. Sensors can collect real-time data on various parameters, which can be analyzed to gain insight into desired evolutionary patterns. Indeed, communication and sensor networks play a vital role in enabling connectivity, monitoring, and managing various aspects of the modern world. Still, the separate deployment of both networks in the same geographical area has several disadvantages associated with their coexistence at the same location. For instance, these disadvantages can include 1) the dense and chaotic deployment of cables and devices, 2) cable lines visually pollute the area where they are installed, 3) pollution caused by an increase in the number of power and data transmission cables, 4) high costs associated with establishing and maintaining both networks separately, 5) overlapping frequencies and interference, and 6) managing both networks independently can come with complexity, among others. Added to these drawbacks is the fact that communication and sensing networks are deployed side-by-side in the real world, as observed in Fig. 1-(b), (c), and (e). Thus, setting up and maintaining the two networks independently is not practical and economical. However, the is common in current deployments which result in infrastructure costs that are almost doubled due to separate deployment and maintenance of both networks.

Therefore, to address these disadvantages, and enhance the efficiency and performance of communication and sensing networks, as well as open horizons for developing future applications, we propose to establish a joint cooperative network by merging sensing and communication networks into a single network. This proposal derives its strength from the fact that the increasing trend of sensor and communications systems toward exploiting MMW/THZ bands has given them common characteristics [1]. This includes decoupling and matching a miniaturized antenna array, using multiple antenna arrays, leveraging similarities in the azimuth power spectrum, and monitoring with performance measurement metrics. As a result, channel characteristics, signal processing, and hardware components are highly similar between the two systems [42], [43], [44], [45], [46], [47], [48]. Moreover, the MMW/THZ system is only effective in short or medium transmission within a range of roughly 100-200m [60], [61]. This is similar to sensor capability in cities (e.g., radars, cameras, Lidars, lasers, etc.). These sensors are expected to cover short or medium distances of about 100-200m due to the abundance of alleys and street intersections. The fact that most sensing and communications devices are placed side-by-side in the real world, as indicated by Fig. 1-(b), (c), (d), (e), and (f), can be used as an



(a) Conventional roof-mounted base station deployment with many antennas at a single location.

(b) Side-by-side dense deployment of sensing and communications networks in Seoul, South Korea.

(c) Side-by-side dense deployment of sensing and communications networks in the Seoul subway system.

(d) Sensing networks in subway corridors in Seoul, South Korea.

(e) Side-by-side dense deployment of sensing and communications networks on urban roads in Seoul city.

(f) Sensing network in a public park in Seoul, South Korea.

FIGURE 1. **An illustrative example of the deployment of sensing and communications networks in Seoul, South Korea.**

advantage. Consequently, merging advantage merging the two networks will be more efficient, cost-effective, and less complex. Indeed, recent developments in sensors, AI/ML algorithms, computer vision, and fusion technology in conjunction with the crucial imperative of line-of-sight LOS links in MMW/THZ systems significantly motivate the trend towards sensor-aided wireless communication. Since a communication network involves exchanging information using specially tailored signals and retrieving it in noisy environments, a sensor network gathers and derives data from noisy and FOV observations. Thus, by pursuing direct trade-offs between both networks, the H-RAN vision might unify these two processes and maximize mutual performance gains.

B. ARCHITECTURE DESCRIPTION

The proposed H-RAN perceptual network architecture has been designed to supplement the existing O-RAN architecture [8]. This can be accomplished by adding additional hardware, layers, protocols, algorithms, interfaces, and widespread utilization of AI/ML engines, which attempt to mutually complement each other. The intended objective is achieved by utilizing the dense infrastructure of sensor and communication networks. This is done by building an interconnected and collaborative perception network that uses AI/ML techniques for learning, problem-solving, decision-making, and perception. We are also pursuing the integration of the two network functionalities through a common infrastructure and developing an intelligent network that enables future applications. As depicted in Fig. 2, we aim to broadly exploit AI/ML workflow capabilities in the proposed H-RAN architecture, which enables software applications to become more accurate at predicting outcomes without requiring explicit programming. By doing so, H-RAN will move beyond traditional communication networks and provide ubiquitous sensing and communication services through a combination of visual observations and AI/ML engines. H-RAN networks offer an exciting opportunity for 6G networks and beyond to implement communication by leveraging the MMW/THZ bands. Such networks observe the surrounding environment continuously, offering various service applications, including but not limited to real-time monitoring and optimization, predicting, and avoiding faults, intelligent network planning, self-optimizing/self-healing, meteorological observations, surveillance, security, detecting human activity, IoT applications, etc. The H-RAN vision is expected to unlock innovative applications and use cases in the future. By embracing O-RAN specifications into H-RAN

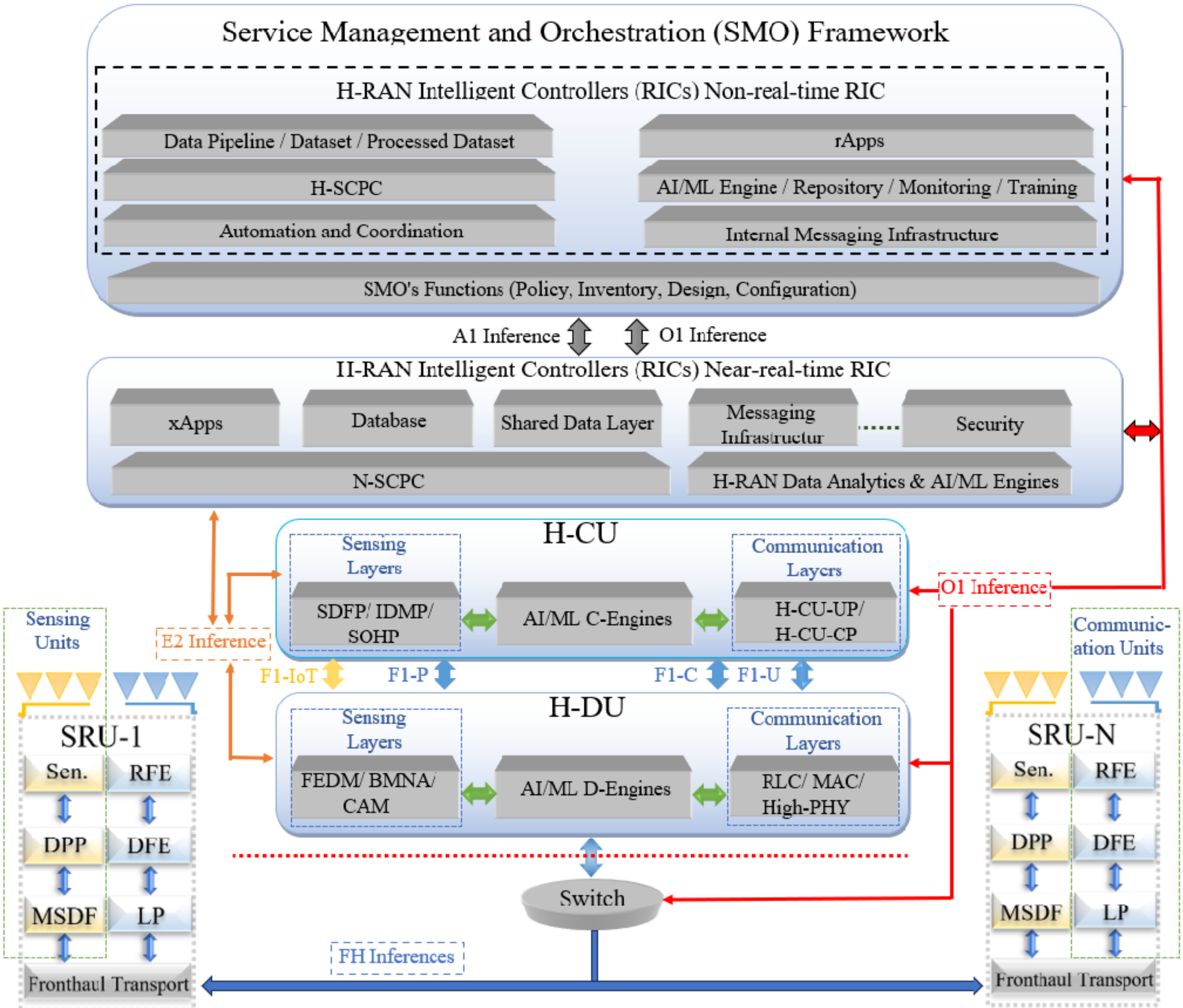


FIGURE 2. The disaggregated architecture of H-RAN perceptual networks distributes the deployment of functional units on cloud platforms in a virtualized network spanning from the edge to virtualization platforms. Augmented with sensing and extensive AI/ML functionality, future heterogeneous H-RANs' "open eyes" are designed to be more intelligent and perceptive. Such networks, which can sense and perceive the surrounding environment ubiquitously, could serve as the backbone for revolutionizing the communications ecosystem.

specifications, expanded horizons for future development can be explored. The goal of intelligent networks is to create a full self-learning network that will be adaptive, responsive, self-healing, fully autonomous, and cost-effective. Future directions and opportunities to develop advanced intelligent networks are predicated on revolutions in AI/ML-based techniques, programmable/intelligent sensors, and IoT processing capabilities [6].

Fig. 2 illustrates H-RAN disaggregation split into three components (cloud, edge, and cell sites) with extensive use of sensor data and AI/ML workflow for different functional units, which effectively embraces and extends the functional disaggregation paradigm for NG-SRAN. In terms of edge and cell sites, H-RAN disaggregation splits the conventional NR next-generation node base station (gNB) into sensors and radio unit (SRU), Hydra-distributed unit (H-DU), and Hydra-central unit (H-CU). The SRU is a logical node located in cell sites that hosts sensing and communication components integral to its function within the H-RAN architecture with limited signal processing capabilities, so deployment is simple and cost-effective. The Fronthaul (FH) interface [10] is the data communication link between the SRUs and H-DU. As shown in Fig. 1, H-DU is a logical node hosting sensing and communication layers in addition to AI/ML D-engines based on a functional split and is deployed at the edge of the network. The management and orchestration (SMO) [62], [63] is responsible for controlling and managing SRUs, H-DUs, and H-CUs. AI/ML D-engines are designed to make intelligent decisions by utilizing both communication parameters and sensor data. As a result, the H-DU can make intelligent decisions that adapt to changing network conditions and user requirements. The interface F1 [14] is used as a communication link between H-DUs and H-CU. The H-CU is typically located at the edge cloud or co-located with H-DUs and can oversee the control and management of multiple H-DUs. The H-CU gains a comprehensive understanding of network conditions and user behavior by combining communication parameters that encompass various metrics and sensor data. This enables H-CU to make intelligent decisions that enhance network performance. The E2 interface [15] is an open interface between the near-real-time RAN intelligent controllers (near-RT RICs) and the H-DUs, along with H-CUs. The other interface that connects RAN intelligent controllers with RAN nodes is the O1 interface [10].

As described in Fig. 2, the H-RAN architecture includes two RAN intelligent controllers (RICs) that perform management and control of the network in the near-RT, namely H-near-RT RIC and the non-real-time, namely H-non-RT RIC. Differing from the traditional RICs-based O-RAN specifications, the H-RIC controllers in H-RAN architecture are designed to make intelligent decisions using various inputs, including communication key performance measurements (KPMs) metrics, sensor data extracted from cell sites of the RAN, and user reports, among others. H-RICs can be defined as a collection of computing resources and virtualization infrastructure located in a single or multiple physical data centers [8]. In general, H-RICs typically employ AI/ML-based policy and control capabilities for handling real-time/historical communication parameters and sensor features.

C. SENSING AND RADIO UNIT (SRU) ARCHITECTURE

In H-RAN terminology, the SRU is a logical node hosting various components integral to its function within the H-RAN architecture to transmit and capture communication and sensor signals with limited processing capabilities. More specifically, the SRU is designed to merge communication and sensing units into a single unit.

The SRU architecture's modular and disaggregated approach allows for interoperability between equipment from different vendors, potentially leading to increased innovation and cost-effectiveness in the deployment and management of sensory and radio access networks. As shown in Fig. 3, the SRU hosts three main units, which are the communication unit, the sensing unit, and the FH transport unit. The communication unit is composed of data transmission and receiving components, while the sensing unit consists of sensor data collection and preprocessing components, and both the communication unit and sensing unit are interfaced through the FH transport unit.

Indeed, the functions of an SRU can vary depending on the specific implementation and deployment scenario. According to the H-RAN specifications, most of the baseband and sensing processing for the cluster of SRUs is performed and centralized in H-DU, which are connected through high-speed FH interfaces [10]. This permits more refined signal processing and load balancing while saving expenses. Fig. 3, illustrates the distribution of signal processing and sensor physical layers within a SRU. Typically, the SRU sits at the cell sites of a network, hosting sensors and radio antennas, from which it performs various sensing and radio-related functions.

Sensors typically receive their primary inputs from physical parameters or phenomena of interest (e.g., RF signal, light signal, emitted, reflected, refracted, or scattered from the environment), and typical outputs with appropriate signal processing include angles, distance, velocity, location, motion, direction, identification, and many others. For instance, the input to the MMW radar sensor is electromagnetic waves, which generate and transmit signals in the form of continuous waves or pulsed signals. As the transmitted signals interact with objects in the sensor's field of view, they reflect back toward the sensor. Reflected signals provide information about objects, such as their distance, velocity, size, and angle. Similarly, other sensors (e.g., GPS, cameras, radars, lidars, etc.) capture data from the environment in a variety of formats and modalities, such as (visual, radio, thermal, depth data, etc.). As shown in Fig. 2, data pre-processing (DPP) is performed on various sensor data to ensure consistency and compatibility for preparing the data for further analysis. This can involve many processes (calibration, synchronization, resampling, filtering, dimensionality reduction, feature extraction, data splitting, etc.), to account for differences in sensor characteristics, data formats, or sampling rates. In the DPP, adjusting the synchronization between the communication signals and the sensing signals flowing into SRU is performed to ensure that data from sensors and communication devices are consistently timestamped. This accuracy is crucial for correlating events, making decisions, and analyzing data. In addition, time synchronization protocols can provide high-precision time synchronization for SRU in the network. Next, once the data is pre-processed, a possible next step is to fuse data from different sensors. A simple data fusion based on AI/ML algorithms can be applied in an SRU to combine features from multiple sensor data elements at an early stage to create a single integrated feature vector that represents a unified multiple modalities representation of the environment or objects of interest.

As for the interfaces, an SRU adheres to standardized interfaces to enable interoperability between different H-RAN components, from edge control to virtualization, such as RICs, H-CUs, and H-DUs as illustrated in Fig. 2. In addition to the C/U/M/S planes of the current O-RAN FU [15], H-RAN FH protocol also features additional planes that have been developed for sensing data delivery in a synchronized manner with IQ sample data in the user plane, while the control and management planes are extended to include the sensing portion, namely a perception plane (P-plane) and an internet of things-plane (IoT-plane). However, the P-plane is used for the transport of data, in addition to sensor layer control commands, to have a comprehensive understanding of the network's real-time status and evolving conditions. This enables operators to anticipate changes in network demand, plan for future expansions, and proactively

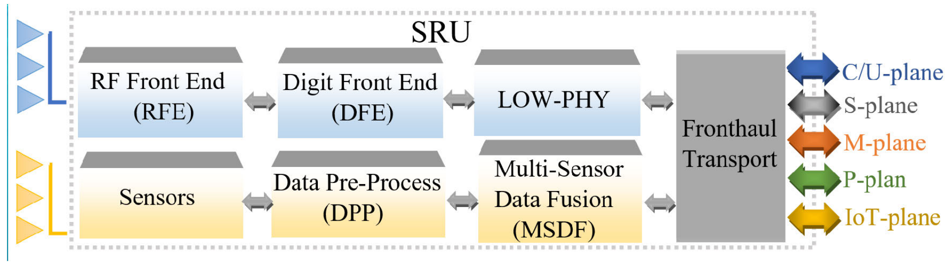


FIGURE 3. Sensing and Radio Unit (SRU) architecture.

address potential issues. Furthermore, the IoT-plane plays a crucial role in enabling users and systems to interact with and control IoT devices and networks. These interfaces facilitate data exchange, remote monitoring, and management of IoT-connected objects and growing ecosystems.

D. HYDRA DISTRIBUTED UNIT (H-DU) ARCHITECTURE

The H-DU is a logical node that hosts various network functions related to sensing and RAN functionalities, typically located at the edge of the network. It plays a crucial role in the overall network architecture by allowing a cluster of SRUs to be configured and controlled in accordance with network requirements and policies. As demonstrated in Fig. 4-(a), (b), in a similar manner to the modularity of the SRU architecture, the H-DU architecture is composed of communication layers and sensory layers, as well as H-DU interfaces and AI/ML D-engines. Unlike traditional monolithic O-DU architectures, H-DU is designed to control and manage sensor data and communication parameters simultaneously derived from a cluster of SRUs with broad AI/ML capabilities. The H-DU architecture is intended to achieve several goals, including perception, artificial intelligence, openness, virtualization, and disaggregation. As demonstrated in Fig. 4-(a), (b), in addition to conventional O-RAN layers of the wireless protocol stack, radio link control (RLC), medium access control (MAC), and high-PHY [11], H-DU architecture is designed to further host new innovative. The new innovative layers enabled by the H-DU architecture can be incorporated in conjunction with traditional layers to supplement and implement common and innovation functions, which include feature extraction and decision-making (FEDM), status estimation and decision-making (SEDM), as well as control and adaptive mechanisms (CAM). Here are some major H-DU protocols and functions:

1) FEATURE EXTRACTION AND DECISIONS-MAKING (FEDM)

The FEDM layer is responsible for collecting the data fusion set from a group of SRUs, completing feature selection and extraction, classification, decision level fusion, and decision-making, which represents the process of extracting sensory relevant features and information from a cluster of SRUs for further analysis. However, the specific FEDM architecture design and implementation may vary among different vendors, deployments, and technologies employed, but generally, FEDM is a sensing data processing method. Here's a general overview of the steps involved as depicted in Fig. 4(b): assuming the output fusion dataset of the cluster of SRUs is an input to the FEDM layer. Next, feature selection and extraction in FEDM is used to choose a subset of the most relevant features from the fused dataset while discarding irrelevant or redundant features. This subset of features should capture the essential information needed for analysis, classification, and modeling tasks. Let us assume that we have feature sets extracted from the fused dataset and measured simultaneously, where each feature set consists of an array of samples. The uni-modal feature selection and extraction of each sensor modality are used in FEDM as data representation, with each mode capturing a different aspect of the environment. When dealing with fused data from multiple sensors, the feature selection and extraction unit becomes a critical component of the data processing pipeline. High-dimensional data, especially when fused from multiple sensors, can result in an increase in computational complexity. Therefore, the feature selection and extraction unit reduce dimensionality by selecting a subset of features that are most relevant to the task at hand and extracting valuable information from the combined sensor data. Afterward, classifiers in multi-source data fusion are used to assign higher weights to classifiers with high accuracy, and then combine multiple classifiers with a weighted combination to obtain a strong classifier with high precision. Next, decision-level fusion is a process where decisions or outputs from multiple sensors are combined to make a final decision. This approach is often used when multiple sensors provide complementary information, and combining their outputs can lead to a more accurate and robust decision. Indeed, the fusion of different modes provides a means of compensating for incomplete or inaccurate information. As indicated in Fig. 4(b) AI/ML D-engines for FEDM can make decisions based on a set of algorithms that are trained to automatically analyze and interpret data, make predictions, estimate the status, detect patterns, and uncover trends discovered in the data. AI/ML D-engines can continuously learn and adapt to changing conditions and environments based on real-time feature extraction. This can include methods (e.g., object detection, tracking, classification, segmentation, etc.) to identify and extract meaningful features or objects

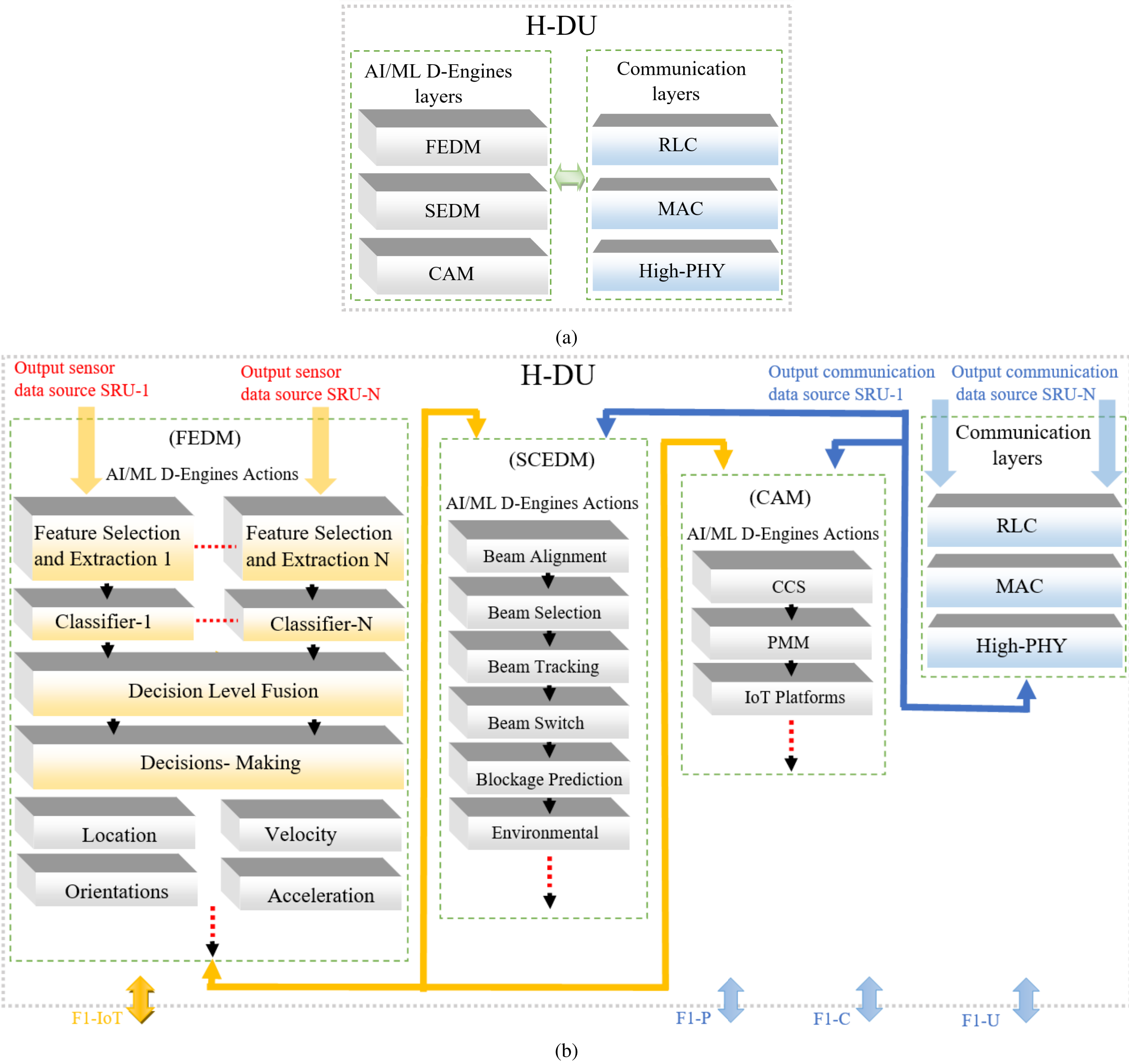


FIGURE 4. **Hydra Distributed Unit (H-DU) architecture.**

from the combined fused data. Once the features have been extracted, the next step of AI/ML D-engines is to fuse them to create a comprehensive representation of the scene or objects. The fused and extracted features can then be used in decision-making for various applications (e.g., beam management (BM), advanced driver assistance systems (ADAS), autonomous driving (AD), surveillance, environmental monitoring, situational awareness, pattern recognition tasks, IoT applications, etc.). Finally, a decision-making module based on AI/ML D- engines controls a set of decisions and takes responsibility for several functions, including:

- **Location:** AI/ML D-engines can process sensor data to extract meaningful information related to object location and predict the next position by combining data from multiple sensors.
- **Recognition:** ML algorithms categorize objects based on their features, such as size, shape, and texture. This can help distinguish between various types of users (e.g., cars, bicycles, pedestrians, etc.).
- **Velocity:** By tracking an object's movement over time, it is possible to estimate its velocity. This can be accomplished using various available sensors, for instance, by analyzing the Doppler shift in radar signals,

comparing successive images captured by a camera, or tracking how the distance to an object change between consecutive lidar measurements.

- **Direction:** Multi-sensor data fusion can help estimate the direction of an object by analyzing the relative positions of the sensors and the object. This can be further refined using machine learning algorithms that recognize patterns of movement associated with different directions.
- **Acceleration:** An object's acceleration can be estimated by measuring its speed changes over time. The process can be carried out using sensors (e.g., accelerometers, radar signals, Doppler shift changes, analyzing camera visual data, lidar, etc.).
- **Target tracking:** H-DU can use data fusion with overlapping fields of view of multi-sensors installed on a cluster of associated SRUs along with advanced AI/ML, to track targets' movement in real-time. Tracking targets using sensors involves continuously monitoring and estimating the position, velocity, and other relevant attributes of objects or entities of interest over time.

2) SENSING AND COMMUNICATION-BASED AI/ML D-ENGINS DECISION MAKING (SCEDM)

As seen in Fig. 4(b) SCEDM interfaces with communication layers and sensing layers jointly to perform several functions related to beam management and network adjustment. With H-DUs, network operators are able to programmatically control and configure various aspects of the RAN through an AI/ML D-engine model. AI/ML D-engines in SCEDM can dynamically adapt to communication parameters and sensor data and make real-time decisions to optimize, and respond to changing conditions, leading to improved efficiency and performance [6]. Listed below are some of the functions that SCEDM is able to provide.

- **Beam alignment:** Object detection in sensing systems is logically analogous to localization and location information estimation in communication systems [45], [46]. Sensor fusion data can provide information about channel conditions (e.g., device status, signal strength, phase, directionality, etc.) [1], [2], [3]. The radio unit collects the radio channel's characteristics matrixes between the SRU and the UE (e.g., channel state information (CSI), MIMO channel, signal-to-interference-plus-noise ratio (SINR), beamforming, path loss, delay spread, doppler shift, etc.) [11]. AI/ML D-engines can process this data in real-time to automatically estimate the optimal beam alignment parameters, (e.g., beam direction, beam width, beamforming weights, etc.). This can be achieved using techniques such as reinforcement learning [64], where the AI/ML D-engines learn from the sensor data and adjust the beam alignment parameters to reduce the beam search space and adapt to changing channel conditions and user mobility [52].
- **Beam selection:** Once the beam search space is minimized and adjusted, AI/ML D-engines can select the optimal beam pairs [46]. Sensor data and communication KPM metrics (e.g., channel conditions, user locations, environmental factors, etc.) can be trained by AI/ML D-engines to make intelligent beam selection decisions. Contextual information from sensors (e.g., user mobility patterns, network congestion levels, etc.) can also be used by AI/ML D-engines to make context-aware beam selection decisions.
- **Beam tracking:** Similar to object detection methods, object tracking in sensing systems is logically equivalent to beam tracking in communication systems [56]. Sensors provide information to communication systems by capturing visual information about the surrounding environment. The radio unit can provide several communication KPM metrics that need to be considered and monitored in real-time to dynamically adjust the beamforming direction. AI/ML D-engines are capable of processing and extracting relevant information from communication parameters and sensing observation (e.g., beamforming weighting coefficients, received signal strength (RSS), CSI, round-trip time (RTT), Doppler frequency shift, receiver's position, speed, acceleration, orientation, etc.). This can help predict the antenna direction and adjust the beam direction accordingly.
- **Beam switch:** Sensor data (e.g., UE location, velocity, orientation, etc.) can provide valuable information about the environment and wireless channel conditions. In addition, continuous monitoring of KPM metrics is available through the communication system. Therefore, the combination of this information can be trained by AI/ML D-engines to predict the optimal beam direction and switch between SRUs within the same cluster in real-time to maintain a strong and stable wireless connection. Specifically, this can be achieved by implementing interaction between AI/ML D-engines and the MAC layers based on radio conditions measurements.
- **Blockage prediction:** These techniques typically involve multiple sensors to detect obstacles and predict when communication beams might be blocked [50], [51]. Sensors can provide real-time data on obstacle location, distance, size, and shape. Based on this information, AI/ML D-engines can be used to analyze sensor data and make predictions about when and where obstacles might obstruct communication beams. This is done by taking proactive measures to avoid or mitigate beam blockages [50]. Based on these predictions, the communication system can proactively adjust beamforming parameters, routing, or transmit power.
- **Environmental adjustment:** Sensors can provide information about environmental conditions that may affect radio frequency (RF) beam performance (e.g., rain, fog, interference from other sources, etc.). These

performance issues can be identified by the communication system's continuous monitoring of KPIs, including link quality metrics. Cooperation between sensory and communication systems in H-DU, H-CU, and RICs can use this information to dynamically adjust beam parameters, such as beam direction or transmit power, to optimize the communication link and facilitate smooth handover decisions. For example, if a beam is experiencing high interference due to weather conditions, a handover can be triggered to transition to a less congested beam by H-CU. The synergy between continuous monitoring of environmental conditions and KPIs empowers communication systems with the intelligence to adapt and optimize RF beam performance in real time.

3) CONTROL AND ADAPTIVE MECHANISMS (CAM)

Indeed, H-RANs introduce a significant trend in the evolution of future networks by merging various technologies and bringing them together (e.g., communication networks, sensor networks, IoT networks, and possibly other emerging technologies). Furthermore, a single H-DU is designed to control and coordinate the functions of a larger group of SRUs than a traditional O-DU, thus covering a larger area. As a result, this design requires the H-DU to include additional components and functions to meet the requirements of an extensive heterogeneous network. As depicted in Fig. 4(b), the CAM layer in H-DU interfaces with the communication and sensing layers simultaneously to perform several functions in response to sensor inputs and communication parameters. In addition, the CAM serves as an intermediate point, facilitating communication between different H-RAN components, such as H-CUs and H-RICs. Specifically, it can perform the following functions in conjunction with other H-RAN components:

- **Coordinate and control signaling (CCS):** The CCS in CAM is responsible for overseeing the flow of control signaling messages between the SRUs, the H-DU, the H-CU, and the core network. Sensory and communication networks can collect real-time and historical data related to users' status (e.g., mobility, movement history, position, velocity, QoS monitoring, etc.). Different AI/ML algorithms can be utilized locally in various H-RAN components, such as RICs, H-CU, and H-DU for subsequent analysis of various aspects of the network. Therefore, the H-DU architecture can achieve a higher level of adaptability to RLC, MAC, and High-PHY functions. This contributes to the overall goal of creating a more intelligent and responsive wireless RAN, which makes it well-suited to address diverse requirements.
  It's crucial to note that H-DU functions, such as connection management, QoS management, traffic management, and resource allocation management are a collaborative effort involving multiple components in the H-RAN architecture. Therefore, the H-DU, in conjunction with the H-CU and other core network functions, contributes to optimizing the H-DU components' functions.
  Integrating real-time and historical sensory data with RLC layer decision-making processes might enable more intelligent and context-aware management to prioritize RLC layer processing to ensure low-latency communication. Moreover, at the MAC layer, the combination of real-time and historical sensory data, and communication parameters can provide insights into current network conditions, including signal strength, interference levels, user mobility, user density, etc. Users' quality of service (QoS) requirements may vary based on their mobility and location, among other factors. Sensory data can be used to differentiate QoS levels for different user groups or locations. For instance, users in fast-moving vehicles may require low-latency connections, while stationary users may have different QoS needs. QoS parameters can be dynamically adjusted to optimize traffic flow within each group. Additionally, the high-PHY layer is capable of optimizing signal transmission and reception by utilizing real-time and historical sensor data. For example, sensor data, such as location, speed, and direction, can be used to optimize beamforming strategies. This information enables the high-PHY layer to adjust transmission beam direction for enhanced signal strength and reliability.
- **Performance monitoring and management (PMM):** The PMM in CAM is responsible for monitoring and analyzing performance data from the cluster of SRUs to optimize the RAN for maximum performance within its designated area of responsibility by cooperating with H-CU, and the core network. It collects performance sensing data and key performance indicators (KPIs) from SRUs within its domain. Based on the collected data, the CAM may make local decisions in cooperation with higher layers of the protocol stack to optimize the SRAN within its coverage area.
- **IoT platforms:** IoT platforms provide edge computing to manage and analyze data from connected devices in collaboration and coordination with other H-RAN components.

E. HYDRA CENTER UNIT (H-CU)

The H-CU is a logical node in the H-RAN architecture that provides support for the higher layers of the protocol stack and typically has responsibility for functions (e.g., radio resource management, connection management, data plane processing, sensor decision management, to name a few). At a high level, the H-CU oversees the connection between the core network and the edge and cell sites of the H-RAN. It provides a centralized point of control for the RAN by being accountable for coordinating the functions of the H-DUs and SRUs to ensure intelligent and reliable operation within a specific geographical area. In addition to

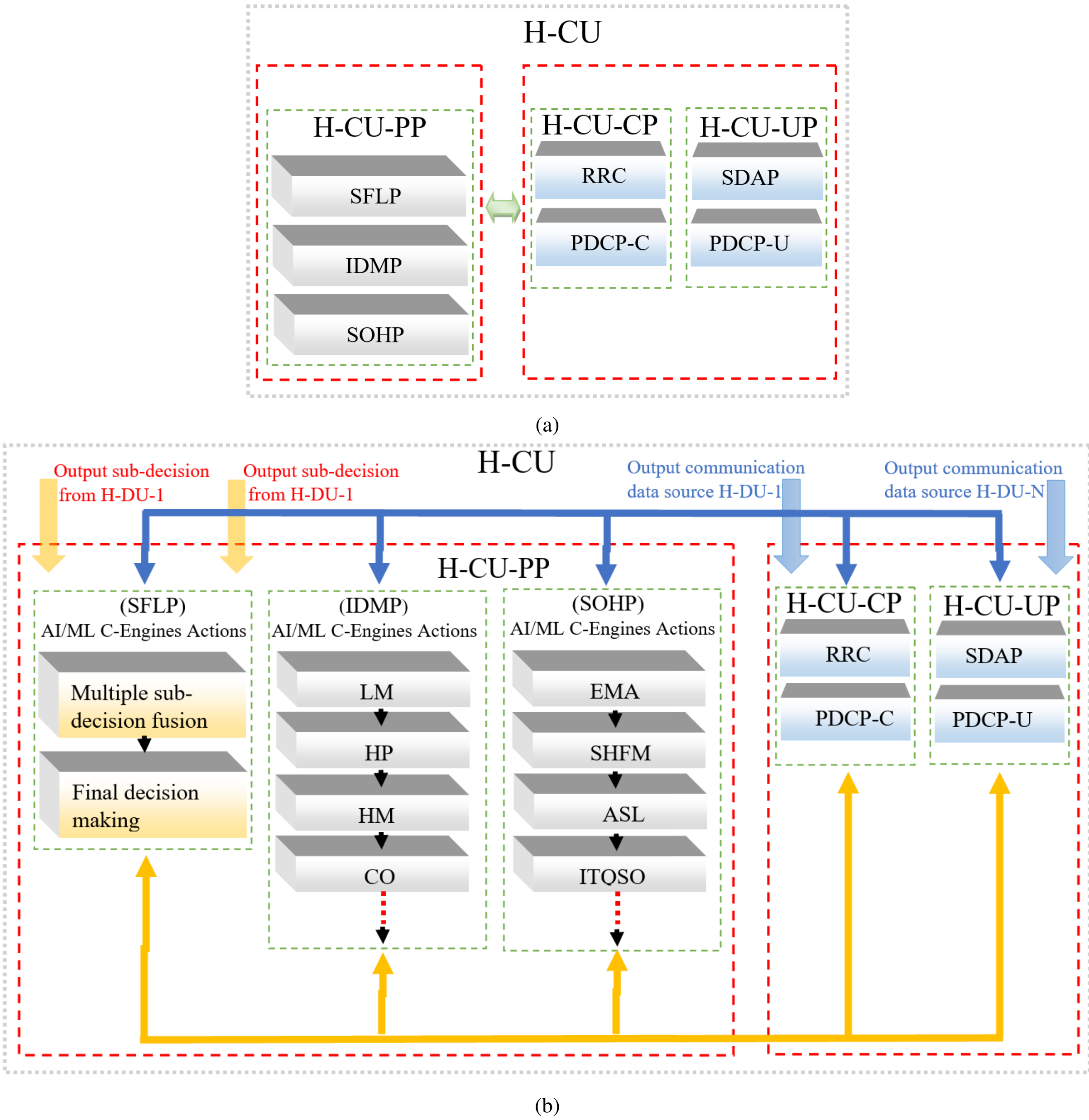


FIGURE 5. Hydra Center Unit (H-CU) architecture.

the connection management specified by the conventional open central unit (O-CU) specifications [8], [9], the H-CU also performs a variety of innovative functions in support of AI/ML C-engines, sensing layers, and higher layers of the protocol stack. As depicted in Fig. 5-(a), (b), in addition to hosting conventional O-RAN components such as radio resource control (RRC), packet data convergence protocol (PDCP), and service data adaptation protocol (SDAP) [12], the H-CU has developed new intelligent components, which enable more sophisticated and intelligent cognitive networks that are capable of supporting a wide range of use cases and applications, namely sub-decision fusion level protocol (SFLP), intelligent decision management protocol (IDMP), and self-optimizing and healing protocol (SOHP). H-CU interfaces with H-DUs, which perform distributed baseband processing functions. This interface facilitates communication and coordination between the centralized and distributed components of the H-RAN architecture. In contrast to the conventional O-CU architecture, the H-CU architecture is split into three logical components. In addition to the conventional control plane (C-plane)/user plane (U-plane) [8], H-CU invented an additional plane, called the P-plane. This logical split provides different functionalities to be deployed on various hardware platforms across the network. As demonstrated in Fig. 5-(a), (b), the H-CU-user plane (H-CU-UP) runs SDAP and PDCP-U, the H-CU-control plane (H-CU-CP) runs RRC and PDCP-C, while the H-CU-perception plane (H-CU-PP) runs SFLP, IDMP, and SOHP. H-CU is designed to support interfaces defined by the O-RAN interfaces Alliance for communication and coordination with other network components (e.g., E1, E2, O1, F1-C, F1-U, etc.) [14], as well as introducing additional interfaces, namely (F1-P and F1-IoT). Specifically, F1-P is an interface between the H-CU-PP and the H-DU. F1-P is primarily responsible for handling error reporting, recovery procedures, and control plane sensing messages between the H-CU and H-DU. These control messages are used for network management, configuration, control, and coordination. F1-P may also carry sensing and communication synchronization-related signaling, ensuring that the H-DU is synchronized with the network's timing and frequency references. F1-IoT is a potential networking interface for IoT applications that require real-time or near-real-time communication. Thanks to this connectivity, devices can transmit data and receive instructions. Here's a breakdown of key points of the main functions of the H-CU-PP components. Here's a breakdown of the main potential functions of the H-CU-PP components.

1) SUB-DECISION FUSION LEVEL PROTOCOL (SFLP)

SFLP performs multiple sub-decision fusions and final decision-making. Multiple sub-decision fusion refers to the integration process where different sources of sub-decision information are combined into one representational format [65]. SDLF is used to combine inputs from several sub-decisions coming from different H-DUs connected to a single H-CU at the H-CU final decision level to yield the final or higher decision that can be used to take several actions. In this scenario, decision-level fusion information relies on reasoning and inference while handling uncertainty, which increases confidence. Additionally, it is possible to enhance accuracy by integrating multiple sub-decisions into the final decision-making process [66]. According to the H-RAN design architecture, the cluster of SRUs is connected to one or more H-DUs, and the cluster of H-DUs is connected to one or more H-CUs. Considering the distribution manner of cooperative SRU deployment, we consider a scenario in which the UE exists simultaneously in the field of view (FOV) of multiple SRUs. Therefore, each SRU operates independently, detecting and classifying objects simultaneously. The UE can be detected and located within the area of interest based on multiple SRU cooperative detection technology, in which the UE's information is extracted from multi-SRUs overlapping FOVs in real-time. Due to the centralized control and management of the H-RAN network, all associated SRUs collaborate to transfer target information to the H-DU, and all H-DUs transfer target information to the corresponding H-CU. For instance, assume the scenario in which the UE is located between two SRUs connected to different H-DUs simultaneously. As a result, the UE's information extracted from each SRU is transferred to the associated H-DUs, which are then combined into the associated H-CU. Sharing this information is essential in many applications, including but not limited to beam tracking and intelligent handover.

2) INTELLIGENT DECISION MANAGEMENT PROTOCOL (IDMP)

IDMP is interfaced with sensor layers and communication layers simultaneously, and it is responsible for several functions related to analyzing and training the integration of sensing data and communication parameters (e.g., location management (LM), handover preparation (HP), handover management (HM), centralized orchestration (CO), to name a few). It collects and analyzes sensing data from H-DUs to decide whether a handover is necessary. It also coordinates the transfer of ongoing communication sessions from the old cell to the updated cell. In general, H-RAN perception networks can provide additional context and real-time information about the network environment. To maintain seamless communication while moving about, AI/ML C-engines can use this data to predict user trajectories and proactively prompt handover decisions [48]. Particularly, if a user rapidly moves in a specific direction, a handover can be triggered to preemptively switch to a beam that aligns with the user's movement direction. A summary of IDMP's potential functions and operations is as follows:

First, after SFLP has reached a final decision, it provides information about the physical environment and conditions. The LM in IDMP is designed to detect the user's status, which is essential for predicting the user's trajectory and making handover decisions. Mobility pattern information can be extracted from SFLP, including user location, velocity, direction, and acceleration. Afterward, the HP is used for handover pre-processing, which involves gathering information from sensing and communication layers and training that information by AI/ML C-engines to identify patterns, trends, and potential handover scenarios. Sensor data along with ongoing communication characteristics can be used to predict when a handover is needed. This can be done by analyzing user movement patterns and trends and estimating their trajectory. AI/ML algorithms can then be applied to process this sensor data, along with other information (e.g., network conditions, traffic load, user preferences, etc.). The proactive handover decision-based H-RAN can be designed based on a tradeoff between probabilities of link blockage, and coverage. Following that, the HM is used to manage the handover decision through interfaces between H-CU and other components in the H-RAN architecture, such as the H-DU, RICs, and higher-layer controllers, collaborating to make handover decisions that are grounded in real-time data and intelligent algorithms, considering all aspects, such as handover triggers, resource allocation, fault tolerance, load balancing, etc. Without disrupting the user experience, HM is also responsible for feedback loops (FL) which are used to continuously monitor communication system performance and adjust parameters based on feedback. This allows for real-time optimization and adaptation to changing communication conditions. In summary, the combination of H-CU-UP, H-CU-CP, and H-CU-IP functions can allow for dynamic control and optimization of beam handover by proactively triggering beam handover decisions and sending these decisions to H-DUs. Thus, the adjacent H-DUs will be fully aware of the UE's movement, allowing them to make handover decisions proactively. Next, CO is used for a centralized orchestration function of AI/ML engines that can be used to manage and coordinate RICs, H-CUs, and H-DUs functions in the H-RAN [62]. This function can provide a unified view of the network and ensure that the RICs, H-CUs, and H-DUs are configured and operated in a coordinated and efficient manner by leveraging advanced analytics, automation, and intelligent decision-making.

3) SELF-CONFIGURATION AND SELF-HEALING PROTOCOL (SOHP)

This protocol enables H-CUs to adjust dynamically, make decisions, and self-heal by cooperating with other H-RAN components, such as RICs and H-DUs to maintain reliable and efficient operations at the network edge by simultaneously connecting to sensors and communication layers. For instance, when perceptual network failures or anomalies are detected through monitoring and analysis, the H-CU can use communication parameters along with real-time/historical sensory data and AI/ML C-engines to autonomously configure its parameters or settings to adapt to changing conditions. This can involve several functions (e.g., rerouting traffic, reconfiguring network elements, activating backup links or nodes to restore network connectivity and performance, switching to an alternate frequency band, dynamically adjusting transmission power, etc.). Below is an overview of SOHP's responsibilities and functions.

- **Environment monitoring and analytics (EMA):** The EMA in AI/ML C-engines and environmental adjustment in AI/ML D-engines cooperate with management and orchestration in RICs for monitoring and analyzing network performance in real-time in response to changes in the climate environment. For example, when H-DUs detect climate-related changes that could affect signal propagation, they report this information to the H-CU. The H-CU, with its centralized view and access to weather data by Management and Orchestration to RICs, can make informed decisions about adjusting transmit power levels. For example, in anticipation of heavy rain that might attenuate signals, the H-CU in cooperation with RICs can proactively increase transmit power in affected areas to maintain service. EMA might adjust transmit power levels, implement load balancing, take preventative measures in anticipation of adverse environmental conditions, etc. Indeed, this protocol is crucial due to the fact that the promising MMW/THz frequency bands are adversely affected by climate change. Using sensor data, user device reports, and network logs, the EMA can help operators identify patterns and anomalies in the network environment. For instance, sensors (e.g., lidar, temperature, humidity, pollution levels, RF sensors, visual monitoring sensors, etc.) are capable of continuously collecting data from their respective environments, along with communication parameters (e.g., network traffic load, cell utilization, device statistics, etc.). This integration provides a comprehensive view of the network's performance concerning its environment.
- **Self-healing and fault management (SHFM):** AI/ML engines can continuously monitor real-time/historical sensory data, communication parameters, and feedback to detect anomalies or network faults. AI/ML models in H-DUs, H-CU, and RICs can be collaboratively based on managing and orchestrating virtualized resources in the network to identify irregular patterns or unexpected changes in data. For example, a sudden drop in signal strength or a significant increase in latency might indicate a problem. Once an anomaly is detected, AI/ML algorithms can perform root cause analysis to determine the source of the issue. This could be due to equipment malfunction, interference, environmental factors, or other network conditions. Depending on the severity of the issue, AI/ML C-engines can trigger automated responses or re-report the issue to the core network.
- **Adaptive and self-learning (ASL):** As part of H-RAN cognitive and intelligent networks, ASL is crucial for the evolution of conventional wireless networks, especially in the context of 6G and beyond. It takes responsibility for adapting and learning based on real-time/historical data and feedback without human intervention, making the network more intelligent and adaptive over time. ASL, along with H-DUs and core networks jointly and cooperatively, operates to recognize patterns, predict future network behavior, and make informed decisions to optimize communication parameters. These adaptability and self-learning capabilities enable the network to continuously improve its ability to meet the diverse requirements of emerging applications and use cases.
- **Intelligent traffic steering and quality of service (QoS) optimization (ITQSO):** In the H-RAN architecture, QoS optimization in O-RAN networks is a collaborative effort involving multiple components and technologies (e.g., RICs, H-CU, H-DU, KPM metrics, sensors, AI/ML engines). These components work together to monitor network conditions, allocate resources, enforce policies, and make real-time adjustments to meet the QoS needs of different services and users. Indeed, the ITQSO in H-CU plays a central role in managing and controlling QoS, by continuously monitoring and analyzing real-time KPM metrics and sensor data in addition to overseeing the orchestration and control of traffic steering within the network. AI/ML engines in H-CU and H-DU can dynamically steer

traffic to optimal paths, prioritize critical services, and dynamically adjust QoS parameters.

F. HYDRA RAN INTELLIGENT CONTROLLERS (H-RICS)

According to the O-RAN specifications [15], The near-RT RIC serves as the core of control and optimization of the RAN, which collects communication data from the lower layers. In addition, the near-RT RIC enables management and control of the network in near-real-time (10 ms to 1 s) and comprises various applications supporting custom logic, known as xApps [59]. Meanwhile, the non-RT RIC is a part of the SMO [65] framework and operates on a time scale longer than 1 s, supplementing the near-RT RIC for intelligent optimization and operation. The non-RT RIC provides custom logic rApps applications to offer value-added services to support and facilitate RAN optimization and operations [19]. Unlike conventional RICs, Fig. 6 demonstrates the distributed architecture of H-RICs computing resources, with sensing and communication data exchange over open interfaces through internet-based cloud platforms, expanding the application of AI/ML engines on networks. The H-RAN perception incorporates two logical controllers with a centralized and abstract point of view on the overall network, namely (H-near-RT RIC and H-non-RT RIC) applications, and access to open interfaces for collecting, managing, classifying, and monitoring data. H-near-RT RICs interact with H-DUs and H-CUs using direct interface termination /citec9 to implement software control. H-RAN's unique characteristics enable the collection and analysis of data from various data sources, including KPM metrics, sensors, user devices, feedback, etc., to monitor network conditions continuously. Therefore, in this view, the proposed H-near-RT RIC platform introduces an additional component, called a near real time-sensor data collection and processing center (N-SCPC) as depicted in Fig. 6, which supplements the main components of the traditional near-RT RIC platform. The N-SCPC is developed to coordinate and cooperate with the rest of the H-near-RT RIC components to implement several applications and functions for continuous network monitoring. More specifically, the H-near-RT RIC is designed to work collaboratively with multiple H-RAN components within the network. These components provide sensing, radio access, and connectivity functions. Therefore, the H-Near-RT RIC operates in a closed-loop control fashion, which continuously monitors network conditions, gathers fine-grained data from the lower layers, and adjusts parameters based on predefined policies and objectives.

Meanwhile, the H-non-RT RIC manages and controls AI/ML workflow operations, model training, control inferences, and rApps updates [15]. In addition to collecting historical data on KPM metrics, user activity reports, and measurement reports defined by conventional non-RT RICs [19]. H-non-RT RICs also collect and analyze historical sensor data from the network. Therefore, the non-RT RIC incorporates an additional component, called the historical sensor data collection and processing center (H-SCPC) as depicted in Fig. 6, which supplements the main components of the traditional non-RT RIC platform. The H-SCPC can leverage historical data for trend analysis and pattern recognition. By understanding past users' and network behavior, the RIC can make more informed decisions regarding resource planning, capacity optimization, and network expansion. The policy and strategy engine are the core components of the H-non-RT RIC. It defines and manages network policies, strategies, and objectives. These policies guide network behavior and decision-making within the RAN. Therefore, H-non-RT RIC can use offline telemetry data to monitor network performance and identify areas where optimization is needed. Moreover, it utilizes historical sensor data and other sources of information to provide enrichment information for the H-near-RT RIC. The H-non-RT RIC can dynamically reconfigure the group of beams to optimize network performance and improve service quality for end-users (e.g., RAN sharing, service level agreement (SLA), antenna parameters, frequency planning, interference management, etc.).

III. H-RAN FRAMEWORK

A. MULTI-SENSOR DATA FUSION-BASED METHOD

Leveraging the prior information provided by a single sensor (e.g., GPS, GNSS, radar, lidar, cameras, etc.) is not always sufficient to cope with complex and challenging environments [65], [66], [67]. This is attributed to the fact that single-sensor systems cannot always efficiently cope with complex and challenging environmental conditions. For instance, radio sensing has relatively high distance accuracy, while its directional accuracy is lower in terms of azimuth and elevation [44], [45], [67]. In contrast, a camera provides high spatial resolution but is less accurate at estimating distances [54], [67]. The other trade-off to be observed is that radar detection accuracy is inferior to a camera and cannot accurately reflect the precise distribution of the surrounding scatters [67], while the camera is more sensitive to lighting and weather conditions. It can be observed from the tradeoff that radar and vision sensors complement each other [65]. The fusion of sensing data can take advantage of various sensors' information and characteristics, thus reducing missed detection rates under adverse environmental conditions [66]. The proposed H-RAN can be adapted to the fusion of various types of sensors (e.g., GNSS, radar, lidar, cameras, etc.). Each sensor offers a distinct perspective and captures specific aspects of the environment. By combining data from multiple sensors, a more comprehensive perception of the surroundings can be achieved. This allows for a richer understanding of the environment and enhances situational awareness.

1) H-RAN-BASED BEAMFORMING

The H-RAN architecture is capable of supporting five different beamforming solutions: 1) Predefined beamforming

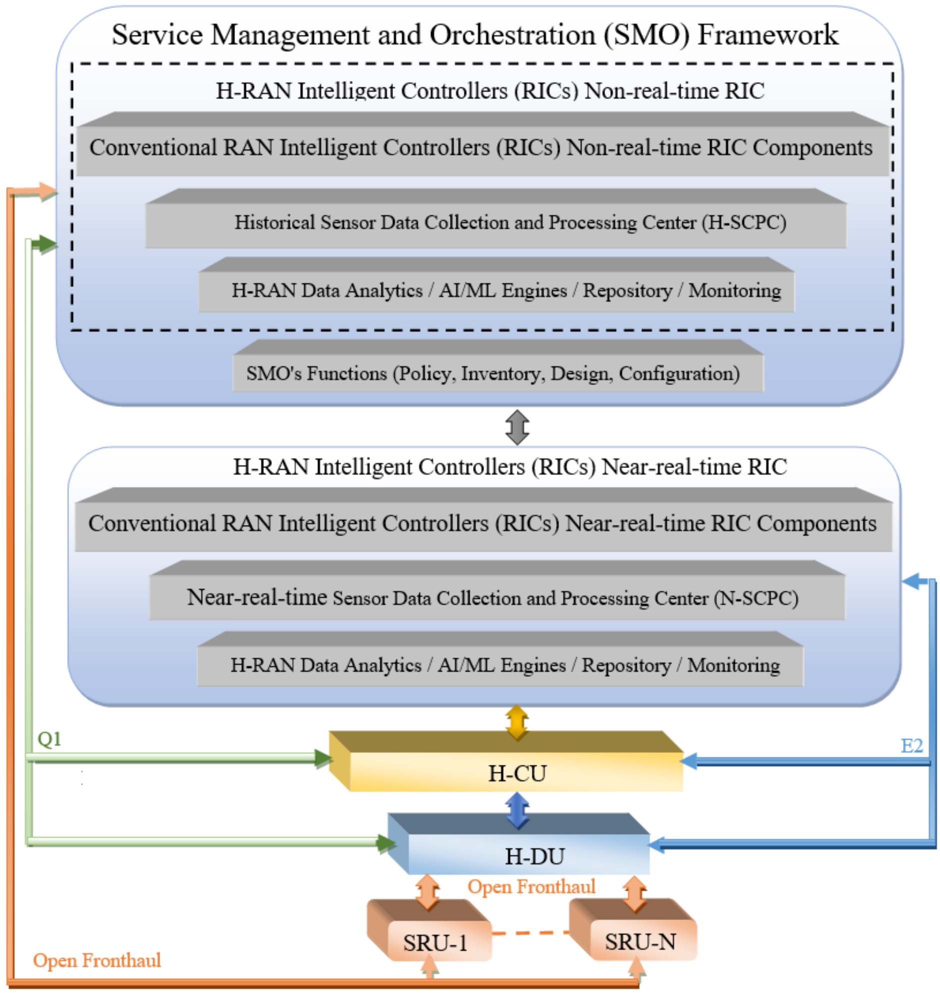


FIGURE 6. H-RAN Intelligent Controllers (RICs) architecture.

(P-BF) [19]: P-BF refers to the use of fixed beamforming vectors that are pre-configured and advertised by the SRUs to the H-DU during startup. 2) Attribute-based beamforming (A-BF) [19]: A-BF is a technique in which the H-DU selects beamforming vectors according to specific attributes, such as azimuth and elevation angles. 3) Weight-based beamforming (W-BF) [19]: W-BF is a technique in which the H-DU specifies the weights for generic time and/or frequency domain beamforming vectors. 4) Channel-information-based beamforming (CSI-BF) [19]: CSI-BF is a technique in which the H-DU needs to estimate CSI for each user based on the received signal and feedback information. Finally, Sensor fusion and AI/ML engines-based beamforming, namely (H-BF): H-BF is an innovative emerging technology that can help to overcome the limitations of traditional beamforming techniques, which may be unable to adapt to complex and dynamic channel conditions in real-time. Moreover, H-RAN has the potential to open up horizons for innovation as well as the development of beamforming technologies tailored to H-RAN's unique characteristics.

B. H-AI/ML WORKFLOWS

AI/ML has emerged as a fundamental paradigm to orchestrate communication and information networks from cloud, edge, and cell sites. For instance, supervised learning (SL) is a type of ML where the algorithm is trained on a labeled dataset, meaning it learns from input-output pairs by mapping input data to the correct output. Fig. 7, illustrates a model of the proposed AI/ML D-engines, AI/ML C-engines, and H-RICs engines in H-DU, H-CU, and H-RICs of the H-RAN platform, respectively. In contrast to conventional AI/ML engines in O-RAN, which gather communication data, H-AI/ML engines in H-RAN architecture are in the position of gathering sensor data, communication information, user reports, and other sources of information from cluster SRUs. The H-RAN specifications consider preliminary data pre-processing and, in this step, data for both sensors and communications layers are merged, shaped, and formatted according to the input size of the specific H-AI/ML engine models. The SMO framework can be used to manage all orchestration, management, and automation procedures to

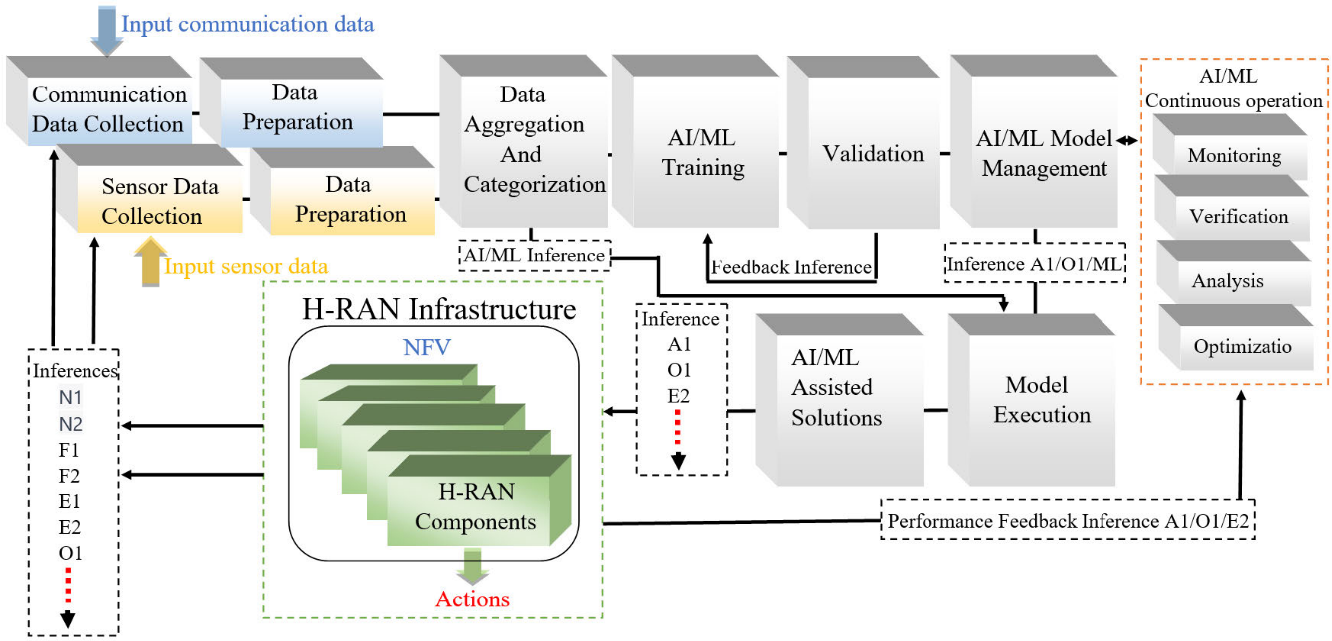


FIGURE 7. H-AI/ML Workflows Architecture.

monitor, and control RAN components from cloud to cell sites for different pieces of equipment [62]. Thus, SMO can enable H-DU, H-CU, and H-RICs to collect all data being produced, including relevant AI/ML pre-processing operations. Additionally, the H-DU and H-CU can influence the SMO framework, which gives the H-DU and H-CU the ability to indirectly control and manage all the SRUs connected to the SMO. The NFV in the H-RAN architecture can be deployed in several locations depending on the specific network design and requirements including H-DU, H-CU, cloud data centers, edge computing nodes, etc. Indeed, the widespread usage of AI/ML engines within H-RAN components contributes to adapting the network to rapidly changing network conditions and user behavior by continuously assessing both real-time and historical data. AI/ML-based resource allocation is capable of accommodating the complexity of modern networks, including large-scale deployments, and heterogeneous devices [5]. In contrast, conventional networks often rely on standardized configurations (e.g., static, or predefined configurations) that may not fully align with rapidly changing network conditions. For the sake of illustration and clarity, this section gives a brief overview of the operational procedures that regulate the AI/ML workflow based on online and offline training for the H-RAN platform, as follows:

- **Data collection and pre-processing:** The proposed H-RAN cognitive architecture aims to exploit AI/ML solutions broadly, and as a result, different types of interfaces are used for sensors and communication data collection (e.g., O1, A1, E2, FH, ML, etc.) [15]. Data collections are stored in various datasets of H-RAN components (e.g., H-DU datasets, H-CU datasets, H-RICs datasets, collection data lake centralized repositories, etc.) where it can be extracted upon request. The H-RAN specifications accommodate a preliminary phase of the data pre-processing stage. At this stage, data is shaped and structured to match the input size of the particular AI/ML model under evaluation for both training data and online inference data. Data preparation might require autoencoders for dimensionality reduction, in addition to standard AI/ML data processing techniques including scaling, reshaping, and standardization [9].
- **Online training:** In this model, AI/ML models can utilize online training techniques to fine-tune and update their parameters based on real-time sensor data and ongoing communication parameters. This approach involves continuously updating the model's parameters using new data as it becomes available, as opposed to traditional offline training (batch training) where the model is trained on a fixed dataset [9]. The pre-trained model provides a starting point for online training, as new sensor data and communication parameters become available. The model then updates its parameters using online training algorithms. The model learns from the updated data and adapts its predictions or decisions based on the latest information. The updated model's performance is evaluated using evaluation metrics or system feedback. This evaluation helps assess the effectiveness of online training. The online training process is iterative and ongoing, continuously updating the model's parameters as new data arrives. This allows the model to adapt to changing conditions, evolving

user behavior, and varying communication parameters. As observed in Fig. 7. Once models are trained, they pass through a validation phase to ensure their reliability and effectiveness. After models are deployed, they are fed with data to perform diverse online inference executions (e.g., classification, prediction, management, and control, to name a few).

- **Offline training:** In this model, AI/ML models can be fine-tuned through offline training, also known as batch training, based on sensor data and communication parameters [9]. The offline training process involves training the model on a fixed pre-collection dataset, followed by applying the trained model for inference or prediction on new data [10]. In dataset preparation, the collected data is split into two parts: a training dataset and a validation dataset [11]. The training dataset is used for model training, while the validation dataset is used to assess model performance during training. During the training process of the AI/ML model, the model learns to generalize patterns and interactions between the input data e.g., key performance measurements, sensor data, user reports, and the desired output. As a similar process to online training, the fine-tuned model is evaluated on the validation dataset to assess its performance and generalization. This evaluation helps determine the model's accuracy, precision, recall, or other relevant metrics. Finally, after successful training and evaluation, the fine-tuned model can be deployed for inference or prediction of new unseen data. It can make predictions or decisions based on the latest sensor data and communication parameters.
- **Continuous online/offline training operations:** As indicated in Fig. 7, the AI/ML online/offline model has the capability to monitor and analyze the model deployed across networks to verify that the network's performance is not adversely affected by AI/ML models' inference outputs. Continuous operation is an essential component of the AI/ML workflow for analyzing and monitoring intelligence deployment throughout the network and verifying that AI/ML models are precise and successful [9].

C. DEPLOYMENT SCENARIOS FOR H-RAN NETWORKS

In H-RAN network deployment scenarios, choosing an architecture that accommodates both communication and sensor elements is crucial for minimizing complexity and expense. As an example, dense deployments of sensor and communication networks in urban areas already exist. Therefore, realizing the concept of blending sensors and communication networks into a single cohesive network interconnected through open interfaces to construct SRUs reduces both capital and operational expenses. Through this feature, the H-RAN paradigm derives its strength from revolutionizing not only future RANs, but also sensing networks by drastically changing conventional network design, functions, operations, and deployment. Therefore, to reduce latency and bandwidth usage, SRU is designed to use edge computing to process sensor data locally before transmitting it to the central network. In addition, it must ensure that sensors and communication equipment adhere to standard protocols and data formats for seamless integration.

D. H-RAN PARADIGM FUTURE RESEARCH DIRECTIONS

O-RAN Alliance has identified several use cases and applications for AI/ML workflows to control RAN behavior [14]. However, each of these use cases and applications has its own unique requirements and challenges. A key objective of the H-RAN architecture is the provision of additional components for managing and optimizing network infrastructure and operations, spanning from edge systems to virtualization platforms incorporating real-time/historical sensory data, broadening the surface of AI/ML automated decision-making, and computing components that seek to enhance use cases and applications and improve overall network performance. In this section, we discuss the advantages and characteristics of H-RAN architecture and boost the surface dimension of AI/ML engines to improve network efficiency by demonstrating potential improvements, functionalities, and novel applications. In addition, we show some of the key distinctions between the proposed H-RAN and the existing O-RAN as listed in Table 1. Finally, we briefly summarize the future research directions and solutions that the H-RAN vision intends to achieve and document the details for future research and studies. It is anticipated that the H-RAN vision will serve as a major source of inspiration for researchers in searching for feasible solutions to various challenges associated with O-RAN implementation [22], [23]. In addition, H-RAN will open the horizons to a wide range of innovative use cases and applications. Theoretically, as indicated in Fig. 8, several areas can be identified as areas of interest, which are outlined below.

- **Beamforming and beam selection/tracking:** Multi-sensor data fusion can provide a wealth of useful features, and when combined with communication parameters, AI/ML engines can train models to adjust beamforming, beam selection, and beam tracking based on changing network conditions.
- **Handover management:** Diverse sensors can continuously monitor users' position and movement while providing real-time feedback to AI/ML engines. The combination of sensor feedback and communication parameters can train AI/ML models to accurately track the user's location and predict its future movement. Predicting future user movements allows the engine to initiate handovers proactively before signal quality deteriorates.
- **Blockage prediction and avoidance:** Real-time sensor data fusion and ongoing communication sessions can train AI/ML algorithms to build perceptive and predictive networks that anticipate blocking incidents. AI/ML

TABLE 1. Below are some of the key distinctions between the proposed H-RAN and the existing O-RAN.

Country List	
H-RAN	O-RAN
Sensor and radio access network (SRAN)	Radio access network (RAN)
The Input/Output are sensing and communication data	The Input/Output are communication data
Dual-functional network	Singel-functional network
Sensor and radio access network (SRAN)	Radio access network (RAN)
Cognitive network	Non-cognitive network
Extensive use of AI/ML engines in network architecture components	limited use of AI/ML engines in network architecture components
Communication-aided sensing and sensing-aided communications networks	Communications-aided networks
SRU, H-DU, H-CU, and H-RIC are equipped with sensing and communication protocols, layers, and units	O-RU, O-DU, O-CU, and RICs are equipped with communication protocols, layers, and units

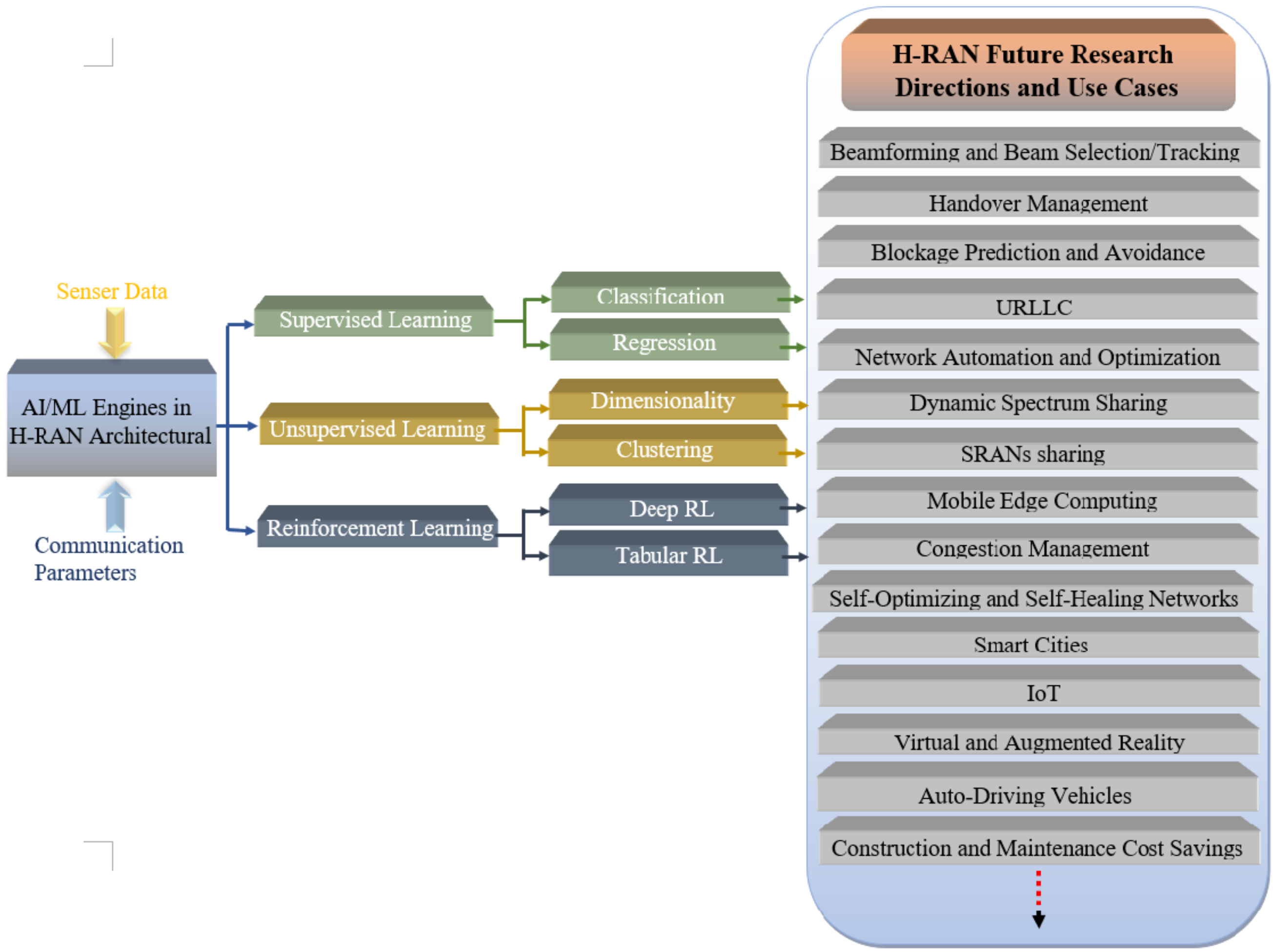


FIGURE 8. H-RAN future research directions and use cases.

algorithms can be used to analyze a mixture of data to identify patterns and anomalies that may indicate blockages, obstructions, or deteriorating link conditions on the communication link.

- **Ultra-reliable and low-latency communications (URLLC):** The integration of multi-sensor data fusion with AI/ML engines represents a powerful approach to optimize networks, predict potential issues, and ensure ultra-reliable performance across various applications. This adaptive and proactive strategy aligns with the demands of modern communication systems that require resiliency, low latency, and efficient use of resources.
- **Network automation and optimization (NAO):** The merging of real-time/historical sensory data and communication parameters can play a significant role in

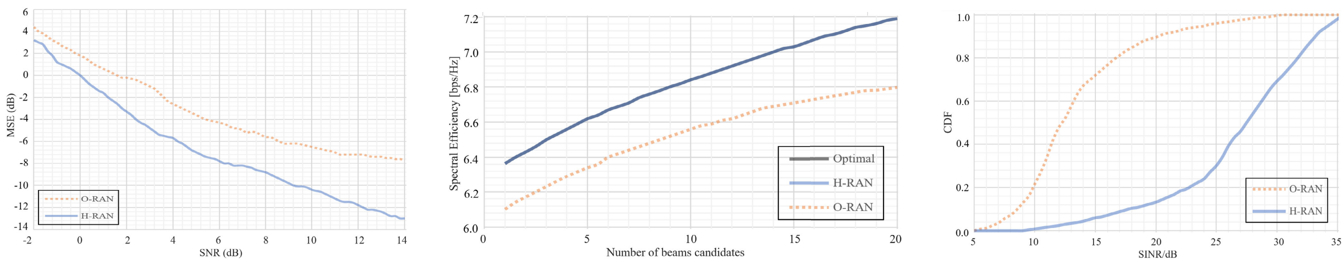
aiding NAO by providing real-time/ near real-time intelligent decision-making along with using SDN and NFV. The network can analyze performance and usage patterns to dynamically optimize resources for improved efficiency.
- **Dynamic spectrum sharing:** Network operators can create accurate spectrum maps and identify areas where spectrum resources are underutilized based on sensor data from a variety of sources, such as spectrum analyzers and other network sensors. For instance, if sensors detect a sudden increase in interference in a specific frequency range, the network can dynamically shift communication to less congested bands.
- **Sensor and radio access networks (SRANs) sharing:** Dual-functional communication and sensor networks have the potential to enable different operators or applications to share common infrastructure and standards, including telecommunication operators, sensor network providers, IoT applications providers, security networks, automated driving services, etc. Such sharing between multiple network operators includes devices, towers, data, spectrum, protocols, core networks, backhaul, power, management, control, maintenance, etc. Indeed, such collaboration among different network operators and service providers improves network coverage and increases network capacity.
- **Mobile edge computing (MEC):** H-RAN-based MEC can bring computing resources closer to the end user and support a variety of use cases, including content caching, real-time analytics, augmented and virtual reality, mission-critical applications, etc. AI/ML algorithms can personalize content recommendations and reduce content delivery times.
- **Congestion management:** With the incorporation of sensor data, and communication sessions, operators can create more comprehensive network diagrams, and identify areas in which network resources are most needed. For instance, network operators can take a data-driven approach along with historical data to plan infrastructure upgrades and expansions in areas prone to congestion.
- **Self-optimizing and self-healing networks:** In the H-RAN vision, self-optimization involves dynamically adjusting network parameters and configurations through AI/ML engines by analyzing real-time and historical data, predicting future network behavior, and making intelligent decisions. Once a fault is detected, self-healing networks can autonomously apply corrective actions. This reduces the reliance on manual interventions and speeds up the recovery process.
- **Smart cities:** The power of H-RAN architecture lies in its ability to collect vast amounts of data from the physical world. AI/ML engines can then turn this data into actionable insights. H-RAN architecture can optimize smart city applications by collecting data from a variety of sensors deployed throughout the city.
- **Internet of Things (IoT):** IoT devices process the data and transmit it to a centralized platform for further analysis by H-RAN. Data processing involves real-time analytics, machine learning, and other algorithms to derive insights. H-RAN-based edge computing, on the other hand, involves processing data near IoT devices. This reduces latency and enhances real-time processing capabilities.
- **Virtual and augmented reality:** By leveraging sensor data from various sources, network operators can create more immersive and interactive VR/AR experiences.
- **Auto-driving vehicles:** The design of H-RAN architecture based on the concept of a perceptive and dual-functional network includes several use cases and functions that could benefit the auto-driving domain. For instance, sensors could be used to define relevant parameters (e.g., location, speed, direction, acceleration, classification, etc.). AI/ML algorithms are used to fuse and interpret this data to create a comprehensive understanding of the vehicle's environment. For instance, AI/ML algorithms consider multiple factors, such as sensor data, traffic rules, navigation maps, and real-time traffic conditions, to plan the vehicle's path and make driving decisions.
- **Construction and maintenance cost savings:** Dual-functional communication and sensor networks can potentially reduce construction and maintenance costs compared to traditional separate networks (e.g., shared infrastructure, reduced installation effort, streamlined maintenance, shared power and connectivity, intelligent monitoring/predictive maintenance, network planning, etc.).

IV. NUMERICAL EVALUATIONS AND SIMULATIONS

The H-RAN paradigm represents a comprehensive revolution in both communications and sensing networks simultaneously. Therefore, we anticipate significant improvements in overall network performance in addition to new features to be available as a result of the H-RAN vision. However, this section shows only some examples of the performance enhancements that the H-RAN architecture can provide while leaving further performance improvement analysis and additional feature extraction evaluation for future research efforts.

A. DATASET

This section describes the datasets that were used to assess the H-RAN network. The proposed H-RAN architecture requires adequate and relevant datasets for machine learning and computer vision. Simulating scenarios is intended to collect data from the same scene as that captured by sensors, and communication channels. The ViWi dataset [68] was utilized to achieve this goal, which is the foundation of several cutting-edge approaches. Simulations for H-RAN's vision



(a) Normalized mean square error (MSE) versus signal-to-noise-ratio (SNR).

(b) Number of beams vs spectral efficiency.

(c) SINR vs CDF.

FIGURE 9. Simulation results.

begin with determining the physical study setting in which the problem occurs. Such a learning environment, however, must incorporate real-world components such as buildings, curbs, streets, automobiles, people, trees, etc. As soon as the targeted research environment is created, sensors and wireless data channel propagation should be gathered and evaluated based on the same simulated scenarios. Blender [69], was used to collect the sensing scenario samples, while the wireless InSite ray-tracing program [70] was employed to build a wireless raw dataset and integrate the channel quality of distinct beam pairings. Visual and wireless raw datasets were run separately to achieve distinct dataset settings.

B. NUMERICAL EVALUATIONS

Among the many analytical comparisons that can be obtained through simulating scenarios, we have selected only a few numerical comparisons for this study, which are normalized mean square error (MSE), spectral efficiency, and signal-to-interference-plus-noise ratio (SINR) as demonstrated in Fig. 9-(a), (b), and (c). The numerical results for the proposed H-RAN architecture are provided in comparison with the conventional O-RAN architecture. Fig. 9(a), shows that the MSE of the accuracy channel estimate is significantly reduced compared with conventional training designs. This is due to the fact that sensing-aided communications can contribute to reducing the MSE of the channel estimate through the following: (1) Incorporating sensing information into the channel estimation process, to mitigate interference effects and improve the accuracy of the estimated channel. (2) Sensing information can provide insights into the dynamic nature of the communication channel, such as variations in multipath propagation, fading, or shadowing. By adapting channel estimation algorithms based on these dynamic characteristics, H-RAN architecture can achieve more accurate estimates, reducing the MSE. (3) Sensing techniques can provide information about the environment surrounding the communication system, such as signal propagation characteristics, noise levels, or the presence of obstacles. This environmental awareness can be used to optimize the channel estimation process, considering specific conditions, and reducing estimation errors. Fig. 9(b) illustrates the spectral efficiency of the schemes involved in the comparisons against the number of beams. Specifically, when comparing different schemes against the number of beams, it is generally observed that the spectral efficiency of the system increases gradually as the number of beams increases. This is mainly attributed to the increase in beamforming gain. Beamforming allows for focused transmission and reception in specific directions, thereby improving signal quality and reducing interference. However, the proposed sensing-aided communications system offers a significant advantage by achieving higher spectral efficiency even with a small number of beams. This means that the system can achieve comparable or even better spectral efficiency compared to traditional schemes that require a larger number of beams. The proposed H-RAN matches optimal performance and achieves a value of 6.61 with only 5 beams, which significantly outperforms conventional methods. Finally, the cumulative distribution function (CDF) of the downlink SINR metric is used to evaluate the performance of a network. It provides insights into the probability distribution of SINR values and indicates how different algorithms or systems affect overall SINR performance. The evaluation of SINR using the CDF plot in Fig. 9(c) highlights the superiority of the envisioned H-RAN architecture compared to conventional architecture. The higher SINR values achieved by the proposed system indicate improved signal quality, reduced interference, and better overall network performance. From Fig. 9(c) we can note that the proposed method obtains a 93.5 % probability that the SINR is larger than 20dB, which indicates a small estimation error of predicted angles.

C. DISCUSSION AND FURTHER WORKS

The H-RAN architecture has been developed with a forward-looking approach to accommodate future technologies and advancements. It considers the evolving needs of 6G and beyond, including emerging communication technologies, massive IoT deployments, URLLC, immersive user experiences, etc. H-RAN vision provides a foundation for integrating these technologies and facilitates seamless integration by incorporating dynamic network optimization mechanisms to adapt to changing user demands and network requirements. These mechanisms continuously monitor network performance, identify bottlenecks, and optimize network parameters to ensure efficient resource allocation and meet user requirements. In general, H-RAN is designed to harness the power of sensing networks to enhance communication performance and leverage communication capabilities to enhance sensing functionalities. Therefore, H-RAN serves as a dual-functional perceptive network capable of interpreting and understanding the surrounding environment. This perceptive nature allows the network to adapt and respond to varying conditions and optimize performance. In this paper, the foundational principles and main specifications of H-RAN have been briefly outlined. However, there are still many issues open to debate about development, protocols, design, standardization, etc. The first phase of future H-RAN research aims to investigate the functions listed in Fig. 8. H-RAN supports a dual-functional network, this paper briefly discusses some of the main functions performed and only related to the communications network, while leaving the discussion open to future studies to innovate new functions and specifications for sensing networks, and IoT applications. The H-RAN paradigm represents a comprehensive revolution in both communications and sensing networks simultaneously. Therefore, we anticipate significant improvements in overall network efficiency, as well as creating novel applications commensurate with the H-RAN network's capabilities.

ACKNOWLEDGMENT

(Rafid I. Abd and Kwang Soon Kim are co-first authors.)

REFERENCES

[1] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.

[2] 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.

[3] B. Salehi, G. Reus-Muns, D. Roy, Z. Wang, T. Jian, J. Dy, S. Ioannidis, and K. Chowdhury, "Deep learning on multimodal sensor data at the wireless edge for vehicular network," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7639–7655, Jul. 2022.

[4] M. Qurratulain Khan, A. Gaber, P. Schulz, and G. Fettweis, "Machine learning for millimeter wave and terahertz beam management: A survey and open challenges," *IEEE Access*, vol. 11, pp. 11880–11902, 2023.

[5] J. Du, C. Jiang, J. Wang, Y. Ren, and M. Debbah, "Machine learning for 6G wireless networks: Carrying forward enhanced bandwidth, massive access, and ultrareliable/low-latency service," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 122–134, Dec. 2020.

[6] E. Esenogho, K. Djouani, and A. M. Kurien, "Integrating artificial intelligence Internet of Things and 5G for next-generation smartgrid: A survey of trends challenges and prospect," *IEEE Access*, vol. 10, pp. 4794–4831, 2022.

[7] E. Ebenezer, T. Swart, and T. Shongwe, "Leveraging on the cognitive radio channel aggregation strategy for next generation utility networks," *Energies*, vol. 12, no. 14, p. 2753, Jul. 2019.

[8] *O-RAN Working Group 1, O-RAN Architecture Description 5.00, O-RAN.WG1.O-RAN-Architecture-Description-v05.00 Technical Specification*, Jul. 2021.

[9] *O-RAN Working Group 2, O-RAN AI/ML Workflow Description and Requirements 1.03, O-RAN.WG2.AIML-v01.03 Technical Specification*, Jul. 2021.

[10] *O-RAN O2dms Interface Specification: Profile based on ETSI NFV Protocol and Data Models 2.0, O-RAN.WG6.O2DMS-INTERFACE-ETSINFV-PROFILE-v02.00*, Oct. 2022.

[11] *O-RAN Working Group 4, O-RAN Management Plane Specification 7.0, O-RAN*, document O-RAN.WG4.MP.0-v07.00, Alfter, Germany Technical Specification, Jul. 2021.

[12] *O-RAN-Architecture-Description-v05*, document TS O-RAN.WG1, O-RAN Alliance, O-RAN Architecture Description, Jul. 2021.

[13] *O-RAN Operations and Maintenance Architecture Version 3.0*, O-RAN Alliance, Alfter, Germany, Apr. 2020.

[14] *O-RAN Working Group 1, O-RAN Operations and Maintenance Interface 4.0*, document O-RAN.WG1.O1-Interface.0-v04.00 Technical Specification, Nov. 2020.

[15] *O-RAN Non-RT RIC and A1 Interface: Use Cases and Requirements 4.00, O-RAN.WG2. Use-Case-Requirements-v04 Technical Specification*, Jul. 2021.

[16] Z. Wang, Z. Liu, Y. Shen, A. Conti, and M. Z. Win, "Location awareness in beyond 5G networks via reconfigurable intelligent surfaces," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2011–2025, Jul. 2022.

[17] C. Fiandrino, G. Attanasio, M. Fiore, and J. Widmer, "Toward native explainable and robust AI in 6G networks: Current state, challenges and road ahead," *Comput. Commun.*, vol. 193, pp. 47–52, Sep. 2022.

[18] W. Azariah, F. A. Bimo, C. Lin, R. Cheng, R. Jana, and N. Nikaein, "A survey on open radio access networks: Challenges, research directions, and open-source approaches," Tech. Rep., Aug. 2022.

[19] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," 2022, *arXiv:2202.01032*.

[20] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 214–222, Feb. 2020, doi: 10.1109/MWC.2019.1900027.

[21] U. Challita, H. Ryden, and H. Tullberg, "When machine learning meets wireless cellular networks: Deployment, challenges, and applications," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 12–18, Jun. 2020.

[22] B. Brik, K. Boutiba, and A. Ksentini, "Deep learning for B5G open radio access network: Evolution, survey, case studies, and challenges," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 228–250, 2022.

[23] S. K. Singh, R. Singh, and B. Kumbhani, "The evolution of radio access network towards open-RAN: Challenges and opportunities," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2020, pp. 1–6.

[24] Y. Heng, J. G. Andrews, J. Mo, V. Va, A. Ali, B. L. Ng, and J. C. Zhang, "Six key challenges for beam management in 5.5G and 6G systems," *IEEE Commun. Mag.*, vol. 59, no. 7, pp. 74–79, Jul. 2021.

[25] 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.

[26] A. Jahid, M. H. Alsharif, and T. J. Hall, "A contemporary survey on free space optical communication: Potentials, technical challenges, recent advances and research direction," *J. Netw. Comput. Appl.*, vol. 200, Apr. 2022, Art. no. 103311.

[27] J. M. Batalla, M. Moshin, C. X. Mavromoustakis, K. Wesołowski, G. Mastorakis, and K. Krzykowska-Piotrowska, "On deploying the internet of energy with 5G open RAN technology including beamforming mechanism," *Energies*, vol. 15, no. 7, p. 2429, Mar. 2022.

[28] S. D'Oro, L. Bonati, F. Restuccia, and T. Melodia, "Coordinated 5G network slicing: How constructive interference can boost network throughput," *IEEE/ACM Trans. Netw.*, vol. 29, no. 4, pp. 1881–1894, Aug. 2021.

[29] B.-S.-P. LinI, "Toward an AI-enabled O-RAN-based and SDN/NFV-driven 5& IoT network era," *Netw. Commun. Technol.*, vol. 6, no. 1, p. 6, Jun. 2021, doi: 10.5539/nct.v6n1p6.

[30] Y. Shi, K. Davaslioglu, Y. E. Sagduyu, W. C. Headley, M. Fowler, and G. Green, "Deep learning for RF signal classification in unknown and dynamic spectrum environments," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Nov. 2019, pp. 1–10.

[31] 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.

[32] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Physical layer deep learning of encodings for the MIMO fading channel," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2017, pp. 76–80.
[33] K. Davaslioglu and Y. E. Sagduyu, "Generative adversarial learning for spectrum sensing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.
[34] T. Erpek, T. J. OShea, Y. E. Sagduyu, Y. Shi, and T. C. Clancy, "Deep learning for wireless communications," in *Development and Analysis of Deep Learning Architectures*. Cham, Switzerland: Springer, 2020, pp. 223–266.
[35] T. S. Cousik, V. K. Shah, T. Erpek, Y. E. Sagduyu, and J. H. Reed, "Deep learning for fast and reliable initial access in AI-driven 6G mm wave networks," *IEEE Trans. Netw. Sci. Eng.*, to be published.
[36] L. Wei, Q. Li, and G. Wu, "Exhaustive, iterative and hybrid initial access techniques in mmWave communications," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.
[37] X. Cao, X. Hu, and M. Peng, "Feedback-based beam training for intelligent reflecting surface aided mmWave integrated sensing and communication," *IEEE Trans. Veh. Technol.*, vol. 27, no. 6, pp. 3367–3382, Jun. 2023.
[38] R. I. Abd and K. S. Kim, "Protocol solutions for IEEE 802.11bd by enhancing IEEE 802.11ad to address common technical challenges associated with mmWave-based V2X," *IEEE Access*, vol. 10, pp. 100646–100664, 2022.
[39] A. Zhang, M. L. Rahman, X. Huang, Y. J. Guo, S. Chen, and R. W. Heath, "Perceptive mobile networks: Cellular networks with radio vision via joint communication and radar sensing," *IEEE Veh. Technol. Mag.*, vol. 16, no. 2, pp. 20–30, Jun. 2021.
[40] Y. Cui, F. Liu, X. Jing, and J. Mu, "Integrating sensing and communications for ubiquitous IoT: Applications, trends, and challenges," *IEEE Netw.*, vol. 35, no. 5, pp. 158–167, Sep. 2021.
[41] Z. Chen, T. Zheng, C. Hu, H. Cao, Y. Yang, H. Jiang, and J. Luo, "ISACoT: Integrating sensing with data traffic for ubiquitous IoT devices," *IEEE Commun. Mag.*, vol. 61, no. 5, pp. 98–104, Oct. 2023.
[42] N. H. Chu, D. N. Nguyen, D. T. Hoang, Q.-V. Pham, K. T. Phan, W.-J. Hwang, and E. Dutkiewicz, "AI-enabled mm-waveform configuration for autonomous vehicles with integrated communication and sensing," *IEEE Internet Things J.*, vol. 10, no. 19, pp. 16727–16743, Oct. 2023.
[43] A. Graff, Y. Chen, N. González-Prelcic, and T. Shimizu, "Deep learning-based link configuration for radar-aided multiuser mmWave vehicle-to-infrastructure communication," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 7454–7468, Jun. 2023.
[44] U. Demirhan and A. Alkhateeba, "Radar aided 6G beam prediction: Deep learning algorithms and real-world demonstration, open-world class discovery with kernel networks," in *Proc. IEEE Wireless Conf. IEEE Trans. Wireless Commun. (WCNC)*, Apr. 2022, pp. 10–13.
[45] Y. Chen, A. Graff, N. González–Prelcic, and T. Shimizu, "Radar aided mmWave vehicle-to-infrastructure link configuration using deep learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 01–06.
[46] A. Klautau, N. González-Prelcic, and R. W. Heath, "LiDAR data for deep learning-based mmWave beam-selection," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 909–912, Jun. 2019.
[47] M. Dias, A. Klautau, N. González-Prelcic, and R. W. Heath, "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.
[48] 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. Cognit. Commun. Netw.*, vol. 6, no. 2, pp. 802–816, Jun. 2020.
[49] A. Alkhateeba and U. Demirhan, "Radar aided proactive blockage prediction in real-world millimeter wave systems, open-world class discovery with kernel networks," in *Proc. IEEE Int. Conf. IEEE Int. Commun. (ICC)*, May 2022, pp. 16–20.
[50] 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.
[51] 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.
[52] M. Zecchin, M. B. Mashhadi, M. Jankowski, D. Gündüz, M. Kountouris, and D. Gesbert, "LiDAR and position-aided mmWave beam selection with non-local CNNs and curriculum training," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2979–2990, Mar. 2022.
[53] M. B. Mashhadi, M. Jankowski, T.-Y. Tung, S. Kobus, and D. Gündüz, "Federated mmWave beam selection utilizing LiDAR data," *IEEE Wireless Commun. Lett.*, vol. 10, no. 10, pp. 2269–2273, Oct. 2021.
[54] W. Xu, F. Gao, X. Tao, J. Zhang, and A. Alkhateeb, "Computer vision aided mmWave beam alignment in V2X communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2699–2714, Apr. 2023.
[55] B. Salehi, M. Belgiovine, S. G. Sanchez, J. Dy, S. Ioannidis, and K. Chowdhury, "Machine learning on camera images for fast mmWave beamforming," in *Proc. IEEE 17th Int. Conf. Mobile Ad Hoc Sensor Syst. (MASS)*, Dec. 2020, pp. 338–346.
[56] T. Zhang, J. Liu, and F. Gao, "Vision aided beam tracking and frequency handoff for mmWave communications," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, May 2022, pp. 1–2.
[57] L. Bonati, S. D'Oro, M. Polese, S. Basagni, and T. Melodia, "Intelligence and learning in O-RAN for data-driven NextG cellular networks," *IEEE Commun. Mag.*, vol. 59, no. 10, pp. 21–27, Oct. 2021.
[58] S. Niknam, A. Roy, H. S. Dhillon, S. Singh, R. Banerji, J. H. Reed, N. Saxena, and S. Yoon, "Intelligent O-RAN for beyond 5G and 6G wireless networks," 2020, *arXiv:2005.08374*.
[59] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "ColO-RAN: Developing machine learning-based xApps for open RAN closed-loop control on programmable experimental platforms," *IEEE Trans. Mobile Comput.*, vol. 22, no. 10, pp. 5787–5800, Oct. 2023.
[60] X. Ge, S. Tu, G. Mao, C.-X. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Wireless Commun.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.
[61] M. G. S. Sriyananda, X. Wang, and R. K. Rao, "Crowdsensing-assisted path loss estimation and management of dynamic coverage in 3D wireless networks with dense small cells," *IEEE Access*, vol. 9, pp. 112670–112685, 2021.
[62] *5G; Management and Orchestration; 5G Performance Measurements*, Standard ETSI TS 128 552, 3GPP TS 28.552 Version 16.6.0 Release 16, 2020.
[63] M. Mohsin, J. M. Batalla, E. Pallis, G. Mastorakis, E. K. Markakis, and C. X. Mavromoustakis, "On analyzing beamforming implementation in O-RAN 5G," *Electronics*, vol. 10, no. 17, p. 2162, Sep. 2021.
[64] C. Dong, Y. Tang, L. Jing, and L. Zhang, "Adaptive transmission for underwater acoustic communication based on deep reinforcement learning," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Oct. 2022, pp. 1–5.
[65] T. Liu, S. Du, C. Liang, B. Zhang, and R. Feng, "A novel multi-sensor fusion based object detection and recognition algorithm for intelligent assisted driving," *IEEE Access*, vol. 9, pp. 81564–81574, 2021.
[66] B. Chandrasekaran, S. Gangadhar, and J. M. Conrad, "A survey of multisensor fusion techniques, architectures and methodologies," in *Proc. SoutheastCon*, Mar. 2017, pp. 1–8.
[67] Y. Song, Z. Xie, X. Wang, and Y. Zou, "MS-YOLO: Object detection based on YOLOv5 optimized fusion millimeter-wave radar and machine vision," *IEEE Sensors J.*, vol. 22, no. 15, pp. 15435–15447, Aug. 2022.
[68] M. Alrabeiah, A. Hredzak, Z. Liu, and A. Alkhateeb, "ViWi: A deep learning dataset framework for vision-aided wireless communications," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5.
[69] *Blender*. [Online]. Available: https://www.blender.org
[70] *Remcom, Wireless Insite*. [Online]. Available: http://www.remcom.com/wireless-insite

RAFID I. ABD (Member, IEEE) received the B.S. degree in electronics and communications engineering from the University of Technology, Baghdad, Iraq, in 2004, and the M.S. degree in communications and networks from the University of Portsmouth, U.K., in 2009. He is currently pursuing the Ph.D. degree with the School of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea. Since 2011, he has three patents registered in the field of communications and networks. His research interests include wireless communication and signal processing, including mmWave and vehicular communications.

DANIEL J. FINDLEY (Senior Member, IEEE) received the Ph.D. degree in civil engineering from North Carolina State University, Raleigh, NC, USA. He was an Adjunct Assistant Professor with the Department of Civil, Construction, and Environmental Engineering, NC State University. He is currently a licensed Professional Engineer (P.E.) in North Carolina. He is also the Associate Director of the Institute for Transportation Research and Education, NC State University. He specializes in economic impact analysis, multi-modal transportation, human behavior research, transportation engineering studies, and asset management and inventory. He served as a Principal Investigator (PI) or a co-PI on over 80 funded projects and has conducted research for sponsors, including NCDOT, FHWA, ACRP, and ITE. He has studied many of the modes of travel, including aviation, bicycle, ferry, highway, pedestrian, port, and railroad.

KWANG SOON KIM (Senior Member, IEEE) received the B.S. (summa cum laude), M.S.E., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in February 1994, February 1996, and February 1999, respectively. From March 1999 to March 2000, he was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA, USA. From April 2000 to February 2004, he was a Senior Member of Research Staff with the Mobile Telecommunication Research Laboratory, Electronics and Telecommunication Research Institute (ETRI), Daejeon. Since March 2004, he has been with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea, where he is currently a Professor. His research interests include signal processing, communication theory, information theory, and stochastic geometry applied to wireless heterogeneous cellular networks, wireless local area networks, wireless D2D networks, wireless ad hoc networks, and new radio access technologies for 5G. He was a recipient of the Postdoctoral Fellowship from the Korea Science and Engineering Foundation (KOSEF), in 1999. He received the Outstanding Researcher Award from ETRI, in 2002, the Jack Neubauer Memorial Award (Best System Paper Award, IEEE Tansactions on Vehicular Technology) from the IEEE Vehicular Technology Society, in 2008, and the LG R&D Award: Industry-Academic Cooperation Prize, LG Electronics, in 2013. From 2006 to 2012, he served as an Editor for the *Journal of the Korean Institute of Communications and Information Sciences* (KICS). From 2013 to 2016, he served as the Editor-in-Chief for the *Journal of KICS*. Since 2008, he has been serving as an Editor for the *Journal of Communications and Networks* (JCN). From 2009 to 2014, he served as an Editor for IEEE Transactions on Wireless Communications.