Exploring Active Inference for Efficient Resource Allocation in AAV-Enabled Cognitive NOMA Uplink

Felix Obite, *Member, IEEE*, Ali Krayani, *Member, IEEE*, Atm S. Alam, *Member, IEEE*, Lucio Marcenaro, *Senior Member, IEEE*, Arumugam Nallanathan, *Fellow, IEEE*, and Carlo Regazzoni, *Senior Member, IEEE*

***Abstract*—The integration of autonomous aerial vehicles (AAVs), cognitive radio (CR), and non-orthogonal multiple access (NOMA) presents a promising solution to significantly enhance the performance of future wireless networks. Achieving this integration requires cognitive self-awareness for intelligent resource allocation. In this paper, we address the problem of sum rate maximization in AAV-enabled cognitive NOMA uplink systems through the joint optimization of subchannel assignment and power allocation, while considering the AAV's mobility. The traditional approach to finding the optimal solution requires an iterative or exhaustive search across all possible combinations of subchannel assignment, power allocation, and AAV position at each time slot, leading to excessive computational complexity. Furthermore, machine learning models, often trained on datasets that do not fully capture the complexity of real-world scenarios, struggle to handle non-stationary events effectively. To solve this nonconvex optimization challenge, we draw inspiration from active inference in cognitive neuroscience and propose a novel data-driven approach called the Active Generalized Dynamic Bayesian Network (Active-GDBN). The main idea is to process the unknown nonlinear input of an exhaustive search optimization algorithm using an Active-GDBN framework. This framework leverages a probabilistic generative model to learn the complex relationships and dependencies among subchannel assignments, power distributions, and the AAV's mobility. The model is facilitated by continuous neuronal message passing in both discrete and continuous states to predict the optimal configuration. Numerical results show that the proposed approach achieves sum rate performance near the optimal exhaustive search and surpasses other baseline approaches.**

***Index Terms*—Active inference, generalized dynamic Bayesian network (GDBN), resource allocation, AAV, cognitive-NOMA.**

I. Introduction

With the increasing proliferation of mobile devices and wireless data traffic for the emerging Internet of Things (IoT) and 6G technologies, enhancing the channel capacity and connectivity of future wireless networks has become essential [1]. The non-orthogonal multiple access (NOMA) scheme has been considered as an innovative technology to enhance channel capacity and meet the demands for low latency, high throughput, and massive connectivity. This is because it allows multiple users to access the same orthogonal frequency concurrently [2], [3]. Unlike orthogonal multiple access (OMA) systems, NOMA exploits the power domain to serve different users with different power allocations, using superposition coding (SC) at the transmitting end and successive interference cancellation (SIC) at the receiving end [4]. Cognitive radio (CR) has been considered a progressive technique to improve radio spectrum efficiency [5]. By exploiting NOMA, multiple secondary users (SUs) can access the radio spectrum simultaneously and opportunistically without causing destructive interference to primary users (PUs). Similarly, adopting Autonomous aerial vehicles (AAVs) and high-altitude platforms as flying base stations has recently gained recognition as an emerging technology to guarantee cost-effective and reliable communications [6], [7], [8], [9]. As a result of their excellent line-of-sight (LOS) communications and high altitudes, AAVs can enhance the quality of service for ground users. To attain fully autonomous systems, recent research [10] focuses on equipping AAVs to perceive their environments and respond instantly, extracting real-time data to perform intelligent decisions. Cognitive self-awareness is a vast concept that defines the cognitive characteristics of typical biological agents [11]. An autonomous agent is considered self-aware if it can continuously adapt itself to new situations

Received 24 April 2024; revised 3 October 2024; accepted 24 November 2024. Date of publication 2 December 2024; date of current version 8 August 2025. This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC), U.K., under Grants EP/W004100/1, EP/W034786/1 and EP/Y037243/1, in part by the European UnionŠs Horizon Europe research and innovation programme under the Grant Agreement No. 101121134, and in part by the Italian Ministry of Universities and Research under the National Recovery and Resilience Plan (PNRR) Raise Innovation Ecosystem D.D. 1053 23/6/2022. The associate editor coordinating the review of this article and approving it for publication was X. Chen. *(Corresponding author: Felix Obite.)*

Felix Obite is with the Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture, University of Genoa, 16145 Genoa, Italy, and also with the School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS London, U.K. (e-mail: felix.obite@edu.unige.it).

Ali Krayani, Lucio Marcenaro, and Carlo Regazzoni are with the Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture, University of Genoa, 16145 Genoa, Italy, and also with the Italian National Inter-University Consortium for Telecommunications, 43124 Parma, Italy (e-mail: ali.krayani@ieee.org; lucio.marcenaro@unige.it; carlo.regazzoni@unige.it).

Atm S. Alam is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS London, U.K. (e-mail: a.alam@qmul.ac.uk).

Arumugam Nallanathan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS London, U.K., and also with the Department of Electronic Engineering, Kyung Hee University, Yongin 17104, Gyeonggi, South Korea (e-mail: a.nallanathan@qmul.ac.uk).

Digital Object Identifier 10.1109/TCCN.2024.3510577

in the environment, which becomes apparent when the agent's attention is on the exterior environment as well as its internal states [12]. Integrating AAVs, CR, NOMA, and cognitive self-awareness will significantly improve spectral resources. Nevertheless, to meet the increasing demand for intelligent devices, future wireless networks will need to improve their capacity for adaptive resource allocation and decision-making. The uncertainty and random nature of wireless networks make the sum rate maximization problem difficult to capture with a precise model. For years, numerical optimization has primarily focused on iterative and heuristic schemes. These approaches take a set of network parameters as input and perform a series of costly iterations to find the optimal solution as output [13]. Although these approaches have recorded tremendous success in specific situations, implementing them for real-time applications still poses significant computational complexity. They often require manual tuning of network parameters and lack the adaptability required for online operation in the envisaged future dynamic networks.

The inherent complexity of resource allocation issues makes learning models an attractive solution [14]. However, a significant challenge arises in generating the training dataset. This dataset may be created using heuristics [13] or exhaustive search [15] to generate approximate or optimal resource allocations. While machine learning (ML) approaches are highly effective in modeling dynamic systems, their models are not easy to interpret [16]. Additionally, reinforcement learning (RL) agents, often trained with predefined state spaces, struggle with generalization when encountering new experiences [17]. Therefore, developing a robust resource allocation scheme that is explainable, supports online incremental learning, and promotes cognitive self-awareness is paramount for maximizing the sum rate in emerging wireless networks.

A. *Motivations and Main Contributions*

The rapid explosion of mobile wireless networks has set the stage for integrating novel technologies such as CR, AAVs, NOMA, and cognitive self-organization to meet the ever-increasing demands for high throughput, reliability, and flexible applications. The integration of these technologies will create a robust network that leverages the strengths of each application to address the current limitations of wireless networks. NOMA and CR optimize network capacity and spectrum utilization, respectively, while AAVs provide the required coverage and mobility. In addition, cognitive self-organization is aimed at enhancing network efficiency and adaptability in dynamic and complex environments.

A significant portion of the existing literature is centered on solutions that involve stationary ground nodes [18]. Nonetheless, the joint problem of AAV mobility and resource allocation in NOMA becomes significantly more challenging to handle with traditional methods. This difficulty arises due to the unpredictable nature and dynamics of wireless communication networks, coupled with the high computational demands involved in finding optimal solutions.

In complex scenarios like deploying AAV systems in highly dynamic environments, ML methods can leverage past experiences to provide a solution that is near-optimal. However, in RL optimization, an agent maximizes a utility, value, or accumulated reward. This assertion, however, may signify a slight misconception in describing adaptive behaviors [19]. On the other hand, humans develop and learn new skills without the aid of an external reward [19]. Active inference [20], [21] an emergent framework from neurocognitive science, proposes that agents choose actions to optimize a generative model that is inclined to the agent's preferences. Under this postulation, optimal behavior occurs in agents, which follows the free energy principle and offers a fundamental basis for evaluating perception and action [22]. The active inference framework filters continuous observations using variational inference in a generalized coordinate of motion, making it well suited for representing continuous states [23]. Furthermore, preliminary results confirm that the active inference algorithm is more flexible and robust in diverse environments that are difficult for RL models [17]. Existing active inference systems typically involve training an agent with predefined generative state space models [19], [20]. However, these may not scale well in the real world when environmental dynamics are complicated, such as in a randomly changing radio environment.

In this paper, we introduce an intuitive approach to active inference using a unique hierarchical generalized dynamic Bayesian network to characterize the intricate temporal dynamics of the state space. In contrast to conventional ML models, which are not designed for causal reasoning and interpretability, the proposed active inference framework enables agents to proactively interact with their environment, dynamically gathering data and making real-time decisions.

The main contributions of this work are summarized as follows:

1) We propose a novel, data-driven algorithm, inspired by active inference in cognitive neuroscience, to address the complex sum rate maximization problem in AAV-based cognitive NOMA systems. Given the complex relationships among AAV mobility, subchannel assignment, and power allocation, the model dynamically adjusts by continuously updating its beliefs and predictions based on new observations. It selects appropriate actions (such as resource allocation based on AAV mobility), thus enabling accurate predictions while adapting to new or previously unseen experiences.
2) We formulate the non-convex and NP-hard optimization problem as a prediction error minimization task, using a Partially Observable Markov Decision Process (POMDP) and a generalized state-space model to characterize the time-varying environment. This problem is then solved by the Active-GDBN framework. Moreover, our algorithm is explainable by estimating and representing the dynamic causal structure of the radio environment at discrete and continuous levels, enabled by constant neuronal message passing.
3) Unlike most articles that rely on discrete power values, our proposed solution learns a continuous power variable. Converting power into discrete forms would impose computational burdens on the algorithm as the number of multiplexed users increases, due to the need to evaluate a growing number of potential power

combinations [2]. The proposed algorithm is computationally affordable since it does not require complex mathematical computations and exhaustive iterations.

4) Numerical results validate the effectiveness of the proposed algorithm, demonstrating near-optimal performance across multiple episodes and surpassing other baselines.

B. Related Work

Research on AAV-based NOMA systems is essential for enhancing spectral efficiency and system throughput. As a result, numerous studies have explored both traditional and machine learning models to address this nonconvex optimization challenge.

Recent Conventional Approaches on AAV-Based NOMA: In [24], the pairing of users and power allocation for AAV-enabled Cognitive-NOMA systems was investigated to optimize the minimum sum data rate per pair. To maximize the minimum throughput of all terrestrial users, [25] optimized AAV positions using a successive convex approximation approach. Additionally, in [26], the authors implemented heuristic-based strategies for user association to enhance the spectral efficiency of a AAV-based NOMA system. In [27], the authors maximized throughput by validating NOMA schemes against OMA for multiple AAV setups. This is achieved through the joint optimization of power allocation and coverage radii, considering static ground users. Furthermore, the authors in [28] considered multiple AAV-based stations (AAV-BSs) and employed an optimal placement technique using exhaustive search to tackle the issue of multiple coverage circles. They also optimized for the maximum number of end-users covered with varying Quality of Service (QoS) requirements. In [29], the authors focused on enhancing the coverage efficiency of a wireless network featuring multiple AAV-based stations by reducing the average distance between the AAVs and the end-users. A combined optimization strategy for precoding was analyzed to ensure secure simultaneous wireless information and power transfer within AAV-supported NOMA systems [30]. In [31], the research focused on a CR-based AAV network with the AAV acting as a secondary aerial transmitter to a ground receiver, aiming to optimize the secondary receiver's worst-case secrecy rate and reduce transmit power while ensuring minimal interference to primary receivers. An SCA-based iterative algorithm was introduced to achieve a locally optimal solution. In [32], a study also addressed the use of AAV-NOMA for relaying signals from a base station to ground users, with the goal of reducing AAV power consumption. The authors recommended using penalty and successive convex approximation techniques to simultaneously optimize the AAV's location and transmission power. Conversely, [33] explored the issue of maximizing the weighted sum-rate for both AAV and ground users by concurrently optimizing the AAV's transmit power and uplink rate. The study introduced egoistic and altruistic transmission approaches, categorized as variants of heuristic search, which achieved higher rates compared to both OMA and non-cooperative NOMA approaches, though at an increased time complexity. Nevertheless, sum rate maximization problems are typically NP-hard [34]. Consequently, obtaining an optimal solution analytically is often challenging due to the intricate dynamics of wireless networks and computational complexity [2].

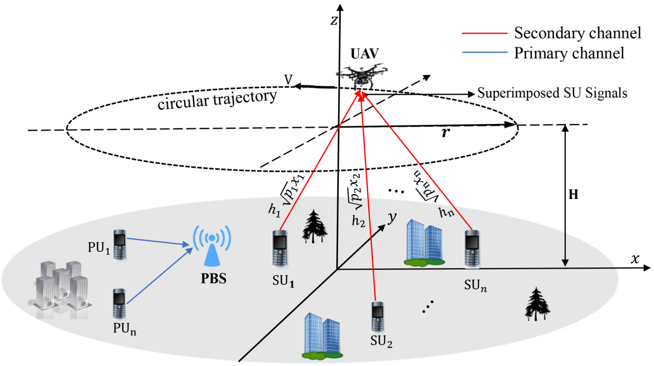


Fig. 1. Illustration of the system model with uplink NOMA signaling.

Recent Machine Learning Models on AAV-Based NOMA: Machine learning (ML) approaches have been applied to wireless system design to enhance system performance [35]. In [36], [37], the authors explored model-free and data-driven deep learning (DL) methods to minimize computational complexity using available input and output training datasets [38]. The preliminary results show improved performance compared to baseline traditional techniques [39]. Another promising subfield of ML is RL [40], which can be a practical choice for online decision-making applications. Specifically, deep reinforcement learning (Deep RL), integrating DL with RL, can offer higher convergence speeds and greater efficiency for NOMA systems with vast state and action spaces. Numerous studies have demonstrated the effectiveness of RL models in optimizing various features of AAV-based NOMA networks. These improvements include optimizing 2D and 3D AAV mobility planning [41], managing power allocation for NOMA [42], improving AAV-user association through clustering [43], and mitigating interference [44].

The rest of this paper is structured as follows: the system model and problem formulation are described in Section II. In Section III, the proposed active inference-based resource allocation approach (Active-GDBN) is presented. Section IV presents the simulation results and performance analysis. Finally, Section V concludes the paper and includes further research investigation.

II. System Model and Problem Formulation

A. System Model

In the uplink, we consider a multi-channel cognitive-NOMA system, as shown in Fig. 1, with a primary base station (PBS) serving P PUs and a cognitive AAV receiving signals from N randomly moving SUs in a cell. Similar to the approach in [45], it is assumed that the AAV maintains a steady speed V while navigating a 3D circular flight path, with its ground projection centered around the PBS.

It is noteworthy that with a circular mobility, the AAV can continuously travel around the PBS with flexible turning angles. This allows for a flexible balance between maximizing the data rate and minimizing energy consumption by adjusting the radius of the circular flight path [46]. In practice, a single AAV communication has opened a plethora of applications, such as meeting mission-specific demands and enhancing security surveillance [47]. This is particularly useful for shorter missions that require detailed monitoring of a specific target or observation. Consistent with the work in [48], we assume the AAV maintains a steady altitude, H, to avoid colliding with obstacles, as mandated by the regulatory body to ensure safety. This location is strategically selected to offer the optimum viewpoint for security surveillance. Particularly in the 3D Cartesian coordinate, each SU is located around the origin (0, 0, 0). The AAV is designed to travel from a predefined initial position $\mathbf{q}_I = (x_I, y_I, H)^T$ to a final position $\mathbf{q}_F = (x_F, y_F, H)^T$, with a time-varying circular path $\mathbf{q}(t) = (x(t), y(t), H)^T$, where $t \in \mathcal{T} = [0, T]$ and T represents the duration of the flight. To aid in designing the AAV's mobility, the total flight duration T is divided into L small time slots, each with a duration of d_t. These time intervals are chosen to be sufficiently small such that the AAV's state (position and velocity) can be considered essentially constant within each individual time slot.

Furthermore, the link's transceiver orientation is perfectly aligned with the other. In an opportunistic manner, the AAV exploits Active-GDBN using the NOMA scheme to allocate resources to SUs on K licensed subchannels. The entire bandwidth is consistently separated orthogonally, hence the interference among K sub-channels is insignificant. We represent the set of SUs as $\mathcal{N} = \{1, 2, \ldots, N\}$ and sub-channels by $\mathcal{K} = \{1, 2, \ldots, K\}$. During each time step, we assume that the position of each SU is constant, the AAV, embedded with Active-GDBN, transmits resource allocation policies to all SUs, and all the SUs move randomly to their next states in the subsequent time slot. The AAV builds its world model by sensing its surroundings to detect ground SUs, enabling them to connect to the uplink NOMA network and transmit signals to the AAV. Each SU n in the uplink transmits a signal to the AAV via sub-channel k using the allotted transmit power p_n^k with channel gain g_n^k. Let $\mathcal{U}_k \triangleq \{n \in \mathcal{N} : p_n^k > 0\}$ signify a group of SUs multiplexed through sub-channel k and $|\mathcal{U}_k|$ represents cardinality. For $k \in \mathcal{K}$, we denote the bandwidth for sub-channel k by b_k and the total bandwidth by $\mathrm{B_w}$. The AAV's objective is to assign SUs to unoccupied sub-channels and choose appropriate power levels to avoid interference and enable efficient decoding at the receiver. We adopted the 3GPP standardization release 15 [49], which defines the channel model between AAVs and users. This model accounts for path loss through both line-of-sight (LoS) and non-line-of-sight (NLoS) conditions. As a result, the 3D spatial distance between SU n and AAV u at time instance t is given by:

$$d_n^u(t) = \sqrt{[x_n^u(t) - x_u(t)]^2 + [y_n^u(t) - y_u(t)]^2 + h_u^2(t)}, \tag{1}$$

where $h_u(t)$ represents the AAV's height.

In the propagation model, LoS is indicated by P_{LoS}. The probability of experiencing NLoS conditions is defined as $P_{\mathrm{NLoS}} = 1 - P_{\mathrm{Los}}$. Therefore, the anticipated path loss $L_n^u(t)$, between SU n and AAV u at time instance t is described by:

$$L_n^u(t) = P_{\mathrm{NLoS}} \cdot L_{\mathrm{NLoS}} + P_{\mathrm{LoS}} \cdot L_{\mathrm{LoS}}. \tag{2}$$

Taking into account small-scale fading, the channel gain between SU n and AAV u at time t is expressed as:

$$g_n^u(t) = \frac{H_n^u(t)}{10^{0.1 \cdot L_n^u(t)}}, \tag{3}$$

where $H_n^u(t)$ represents the fading coefficient between SU n and AAV u. Similarly, in the case of uniform AAV motion at a constant velocity V, we denote $\|\mathbf{v}(t)\| = V, \forall t$. To guarantee the least minimum distance among the superposed SU signals, we assign distinct QPSK constellations to each SU. To this end, no two SUs have similar phase and amplitude. The minimum safe distance is maintained at a constant and sufficiently large value to minimize interference and enable effective SIC decoding by the AAV. If the difference between the received power levels Δ_y, is less than or equal to a predefined power difference threshold p_{th} [50], the AAV will be unable to correctly decode the superimposed signals. QPSK modulation is chosen for its simplicity and effectiveness in NOMA systems, offering a balanced trade-off between performance and interference compared to higher-order modulation schemes. We adopt a SIC detection order proposed in [51] utilizing a permutation function, $\sigma_k : \{1, \ldots, |\mathcal{U}_k|\} \rightarrow \mathcal{U}_k$, where $|\,\boldsymbol{.}\,|$ represents a finite number of SUs.

Subsequently, the AAV receives $\mathrm{y}_{t,k}$ signals from different sub-channels, each representing one of four distinct hypotheses:

$$\begin{cases} \mathcal{H}_0 : \mathrm{y}_{t,k} = \eta_n^k, \\ \mathcal{H}_1 : \mathrm{y}_{t,k} = G\sqrt{p}\mathrm{s}_{t,k} + \eta_n^k, \\ \mathcal{H}_2 : \mathrm{y}_{t,k} = G\sqrt{p}\mathrm{s}_{t,k} + \mathbf{x}_{t,k} + \eta_n^k, \\ \mathcal{H}_3 : \mathrm{y}_{t,k} = \mathbf{x}_{t,k} + \eta_n^k, \end{cases} \tag{4}$$

where $\mathcal{H}_0$ denotes the random noise model (η_n^k), and $\mathrm{y}_{t,k}$ is the received combined signal. $\mathcal{H}_1$ represents the PU's signal model affected by random noise, where $\mathrm{s}_{t,k}$ is the PU's transmitted signal, G represent the channel gain from PU to PBS, and p indicates the transmitted power of the PU. The hypothesis $\mathcal{H}_2$ represents the combined signal model of the PU and the superimposed SUs, transmitted through sub-channel k, given by $\mathbf{x}_{t,k} = \sum_{n=1}^{|\mathcal{U}_k|} g_n^k \sqrt{p_n} x_n$, where g_n^k is the link gain between the SUs and the AAV, also affected by random noise η_n^k. Similarly, $\mathcal{H}_3$ relates only to the superimposed SUs' signal model and is influenced by random noise. The AAV will undergo offline training to learn four distinct models, each representing one of four hypotheses. This will enable it to individually predict the noise signal, PU signal, PU plus SU signals, and SU signals while operating online. Consequently, it will be able to effectively evaluate the actions taken (i.e., allocated power values and subchannel).

In line with the NOMA principle, the AAV first decodes the user with the strongest channel gain, treating the other contributing signals as interference. Subsequently, the user with the next best channel is decoded, continuing in descending order of channel gains. In this study, we have adopted a fixed SIC ordering approach, meaning that it considers the channel gains at the initial time instant, arranges the user based on their channel gains, and maintains this ordering constant throughout the mission. This approach helps to reduce complexity by avoiding the addition of another optimizing variable related to SIC ordering and by preventing the addition of a third online action to the proposed approach. Therefore, the power combination is optimized based on the AAV's mobility, while the decoding order remains constant over time. As the AAV is the receiver for all the uplink signals during uplink reception, it can successfully perform SIC in any desired order without significantly affecting performance [52]. This results in the consistent decoding of SUs in the same order, a typical assumption in NOMA systems [53], simplifying training and model complexity. By leveraging the learned generative model, the AAV iteratively performs SIC online to decode each SU's signal. Although we acknowledge that network conditions vary over time, the specific impact of SIC ordering is more intricate and is left for our future research. Similarly, the AAV can dynamically and continuously adjust its prediction model and actions based on real-time feedback of the channel state information (CSI). In the uplink case, the AAV acts as the only receiver. Therefore, all the transmitting SUs have the same noise on sub-channel k. Thus, the achievable data rate $\mathrm{R}_{k,n}$ of SU n and the sub-channel k is given by:

$$\mathrm{R}_{k,n} \triangleq b_k \log_2\left(1 + \frac{p_n^k g_n^k}{\sum_{j=\sigma_k^{-1}(n)+1}^{|\mathcal{U}_k|} p^k_{\sigma_k(j)} g^k_{\sigma_k(j)} + \eta_n^k}\right). \tag{5}$$

B. Problem Formulation

We aim to achieve a maximum cumulative sum rate of SUs by jointly optimizing subchannel assignment $b_{k,n}(t)$ and power allocation $p_{k,n}(t)$. This takes into account the AAV's position $\mathbf{q}(t)$ in each time slot, whether it follows a circular or random walk trajectory, subject to power and AAV mobility constraints, mathematically formulated as follows:

$$\max_{\{q(t), b_{k,n}(t), p_{k,n}(t)\}} \sum_{k=1}^{K} \sum_{n=1}^{N} R_{k,n} \tag{6}$$

$$\text{s.t.}\quad \text{C1:}\ \sum_{k=1}^{|\mathcal{U}_k|} b_{k,n}(t) p_{k,n}(t) \leq p_{max}^n,\ \ n \in \mathcal{N}, k \in \mathcal{K}, \tag{7}$$

$$\text{C2:}\ b_{k,n}(t) p_{k,n}(t) \leq p_{max}^{k,n},\ \ n \in \mathcal{N}, k \in \mathcal{K}, \tag{8}$$

$$\text{C3:}\ b_{k,n}(t) p_{k,n}(t) \geq 0,\ \ n \in \mathcal{N}, k \in \mathcal{K}, \tag{9}$$

$$\text{C4:}\ |\mathcal{U}_k| \leq M, k \in \mathcal{K}, \tag{10}$$

$$\text{C5:}\ \mathbf{q}(0) = \mathbf{q}_I, \tag{11}$$

$$\text{C6:}\ \mathbf{q}(T) = \mathbf{q}_F, \tag{12}$$

$$\text{C7:}\ \mathbf{v}(0) = \mathbf{v}_I, \tag{13}$$

$$\text{C8:}\ \mathbf{v}(T) = \mathbf{v}_F, \tag{14}$$

$$\text{C9:}\ V_{\min} \leq \|\mathbf{v}(t)\| \leq V_{\max}, \forall t, \tag{15}$$

$$\text{C10:}\ \sum_{t=1}^{T} \Delta(t) \leq T_{max}, \tag{16}$$

where C1 defines the total power budget for each SU n (i.e., must not exceed p_{max}^n). C2 is the power limit on each sub-channel k. C3 ensures that the assigned powers are non-negative. Due to the practical limitation of SIC and decoding complexity [53], the maximum number of SUs to be multiplexed per sub-channel is M, which is related to constraint C4. C5 and C6 denote the initial and final positions of the AAV, while C7 and C8 represent the desired AAV's initial and final velocities. Constraint C9 defines the minimum and maximum speed for the AAV, and C10 guarantees that the AAV finishes its mission within a specified maximum time frame, T_{max}.

C. Random Linear AAV Trajectory (3GPP Model)

In the preceding subsection, despite achieving a flexible balance between maximizing data rate and energy efficiency, adopting a circular AAV trajectory imposes some limits on the AAV's flexibility and degree of freedom. In this subsection, we propose a generalized random linear AAV trajectory that follows the 3GPP standardization simulation for AAV-related studies [54], where drones start at specified random positions in the network and move linearly in random uniform directions at a constant height and velocity throughout the entire simulation time.

In our previous work [55], we demonstrated the feasibility of leveraging active inference for suboptimal subchannel and power allocation optimization based on random linear AAV mobility. Although the proposed approach outperformed existing suboptimal baselines in simulations, its performance was limited due to the use of a suboptimal random scheme for dataset generation. In this paper, we adopt an exhaustive search scheme, a global optimization approach suitable for sum rate maximization problems [56], [57]. While our primary goal in this study is not explicitly trajectory planning, we selected exhaustive search optimization for its straightforward approach [58]. This is in contrast to alternating optimization, where the original problem is divided into sub-problems and solved using Lagrange multipliers, successive convex approximations, or subgradient descent methods [59], which often result in complex mathematical formulations and excessive computations. Although exhaustive search and other global approaches for finding the optimal solution are typically slow, even for smaller parameters, this is a one-time cost for data-driven methods, making it feasible to find suboptimal data-driven solutions that are near-optimal.

D. Exhaustive Search Optimization

Addressing the optimization problem in equation (6) poses a considerable challenge, mainly because it is nonconvex and falls into the category of NP-hard problems. Similar to the work in [2], obtaining a global optimum can be achieved through simultaneous exhaustive search operations (as described in **Algorithm 1**) across all possible combinations of subchannel assignment, power allocation, and AAV positions. For each combination, the sum rate is calculated for all episodes, taking into account building occlusion. This involves updating the best subchannel assignment and power allocation to determine the optimal solution. The optimized

Algorithm 1 Exhaustive Search (Resource Allocation)

```
Input: Set of subchannels $k$; Set of SUs $n$; Set of power values $p$; $P_{max}$; qI; qF; vI; vF; $V_{min}$; $V_{max}$; $T$; $H$; AAV radius $r$.
Initialization:
K = all possible subchannel assignments $b_{k,n}$
P = all possible power allocations $p_{k,n}$
best sum rate = 0
best subchannel assignment = []
best power allocation = []
for $i$ = 1: K do
    $b_{k,n}$ = K(i);
    for $j$ = 1: P do
        $p_{k,n}$ = P(j);
        sum Rate = 0
        for $t$ = 1: $T$ do
            Calculate AAV position qI and velocity vI at time slot $t$
            calculate sum rate ($b_{k,n}$, $p_{k,n}$, $q$, building occlusion)
            signal power = $p_{k,n}$ ($t$) * channel gain($t$)
            $sumrate+ = \log_2\left(1+\frac{\text{signalpower}}{\text{interference}+\text{noisepower}}\right)$
            if sum rate > best sum rate:
            best sum rate = sum rate
            best subchannel assignment = $b_{k,n}$
            best power allocation = $p_{k,n}$
        end for
    end for
end for
Update the AAV's position based on circular or random linear trajectory constraints.
if maximum iterations reached: terminate
Output: optimized solution (best sum rate, best subchannel assignment, best power allocation).
```

sum rate signifies the highest achievable sum rate for the joint objective.

Nevertheless, in this paper, our goal is to tackle the joint objective using exhaustive search optimization to generate the training dataset for both circular and random linear mobility paths, which also serves as a robust upper bound for evaluating the online performance of the proposed data-driven approach. Subsequently, the AAV utilizes the solutions derived from the exhaustive search method to acquire knowledge and construct a dynamic generative model that characterizes both the wireless environment and the decision-making processes of the global optimizer. In the online phase, the proposed method incrementally generates new generative models from training data based on incoming sensory inputs. This adaptive approach enables the AAV to handle new wireless experiences not encountered during training.

III. The Proposed Approach

In this section, we propose a new data-driven approach to resolve the optimization problem in (6). The framework applies a POMDP and continuous generalized states to characterize the variables that represent the optimal solution, leading to a new suboptimal solution (Active-GDBN) guided by the principle of minimizing free energy.

A. Problem Transformation Based on Active-GDBN

A fundamental premise of active inference is that agents possess a probabilistic generative model capable of generating predictions about sensory inputs. This model works by leveraging hidden variables that characterize the underlying causes behind observed outcomes. The model's predictions are continuously compared against real-world observations to estimate how perfectly the hidden causes align with the current situation. We explore active inference using the framework of a POMDP [60]. In each time instance t, the true state of the radio environment $\tilde{\mathrm{S}}_t \in \mathbb{R}^{d_s}$ evolves with a stochastic transition given by $\tilde{\mathrm{S}}_t \sim \Pr(\tilde{\mathrm{S}}_t|\tilde{\mathrm{S}}_{t-1}, \boldsymbol{\mathcal{A}})$, where $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{d_a}$ is the AAV's actions. The true state is generally hidden from the AAV, so the AAV can only receive observations $\tilde{\mathrm{Z}}_t \in \mathbb{R}^{d_z}$, given by $\tilde{\mathrm{Z}}_t \sim \Pr(\tilde{\mathrm{Z}}_t|\tilde{\mathrm{S}}_t)$. From Bayesian principle, for a given prior $\Pr(\tilde{\mathrm{X}}_t)$, with a likelihood $\Pr(\tilde{\mathrm{Z}}_t|\tilde{\mathrm{X}}_t)$, the posterior $\Pr(\tilde{\mathrm{X}}_t|\tilde{\mathrm{Z}}_t)$ is given by [61]:

$$\Pr\left(\tilde{\mathrm{X}}_t|\tilde{\mathrm{Z}}_t\right) = \frac{\mathrm{P}\left(\tilde{\mathrm{Z}}_t|\tilde{\mathrm{X}}_t\right)\Pr\left(\tilde{\mathrm{X}}_t|\tilde{\mathrm{Z}}_{t-1}\right)}{\Pr\left(\tilde{\mathrm{Z}}_t|\tilde{\mathrm{Z}}_{t-1}\right)}. \tag{17}$$

As revealed in (17), the posterior $\Pr(\tilde{\mathrm{X}}_t|\tilde{\mathrm{Z}}_t)$, is defined by three main components:

- *The prior:* $\Pr(\tilde{\mathrm{X}}_t|\tilde{\mathrm{Z}}_{t-1})$ which describes the prior knowledge or belief of the AAV, defined as:

$$\Pr(\tilde{\mathrm{X}}_t|\tilde{\mathrm{Z}}_{t-1}) = \int \Pr(\tilde{\mathrm{X}}_t|\tilde{\mathrm{X}}_{t-1})\Pr(\tilde{\mathrm{X}}_{t-1}|\tilde{\mathrm{Z}}_{t-1})\, d\tilde{\mathrm{X}}_{t-1}, \tag{18}$$

  where $\Pr(\tilde{\mathrm{X}}_t|\tilde{\mathrm{X}}_{t-1})$ represent the state temporal transition.
- *The likelihood:* $\Pr(\tilde{\mathrm{Z}}_t|\tilde{\mathrm{X}}_t)$, defined as the probability of the observation sequence $\tilde{\mathrm{Z}}_t$ given the hidden states $\tilde{\mathrm{X}}_t$.
- *The observation:* $\Pr(\tilde{\mathrm{Z}}_t|\tilde{\mathrm{Z}}_{t-1})$, which represent the normalizing factor defined as:

$$\Pr\left(\tilde{\mathrm{Z}}_t|\tilde{\mathrm{Z}}_{t-1}\right) = \int \Pr\left(\tilde{\mathrm{Z}}_t|\tilde{\mathrm{X}}_t\right)\Pr\left(\tilde{\mathrm{X}}_t|\tilde{\mathrm{Z}}_{t-1}\right) d\tilde{\mathrm{X}}_t. \tag{19}$$

In a practical scenario, the optimized solution is challenging to implement due to the requirement of multiple summations and integration. Hence, the alternative solution proposed in the literature uses Particle and Kalman filtering [62]. In our proposed model, we used a modified Markov Jump Particle Filter (M-MJPF) [63]. The M-MJPF uses a switching model and applies particle filtering for discrete state prediction and updating as well as Kalman filtering for continuous state prediction and updating.

We define the optimization objective of the active inference agent (AAV) as the minimization of state prediction errors. With the available set of observations, free energy can be employed as an objective function for minimizing the Kullback-Leibler (KL) divergence between the approximate posterior distribution $q(x)$ and the true posterior distribution $p(x)$ (i.e., the target distribution), which is expressed as:

$$\begin{aligned}\mathrm{D}_{KL}[q(x)\|p(x)] &= \arg\min_{q(x)}\left[\log\frac{q(\mathbf{x})}{p(\mathbf{x})}\right] \approx \frac{1}{N_p}\sum_{i=1}^{N_p}\log\frac{q\left(\mathbf{x}^{(i)}\right)}{p\left(\mathbf{x}^{(i)}\right)}\\ &= -\frac{1}{N_p}\sum_{i=1}^{N_p}\log\left(W\left(\mathbf{x}^{(i)}\right)\right),\end{aligned} \tag{20}$$

where $W(\mathbf{x})$ is a ratio function comparing two probability density functions. If $q(\cdot) = p(\cdot)$ and $W(\mathbf{x}^{(i)}) = 1$ for all values of i, then $\mathbf{D}_{KL}[q(x)\|p(x)] = 0$.

In general, KL $\neq 0$. Intuitively, if KL is small, the number of particles can be decreased; otherwise, it can be increased.

To ensure that equation (20) is non-negative, given that KL ≥ 0, the normalized (N) weights are calculated:

$$\mathrm{D}_{KL}[q(x)\|p(x)] \approx -\frac{1}{N_p}\sum_{i=1}^{N_p}\log\left(\tilde{W}\left(\mathbf{x}^{(i)}\right)\right) \equiv N_{\mathrm{KL}}. \quad (21)$$

This allows the AAV to select appropriate actions (joint subchannel assignment and power allocation based on the AAV's position) to attain the minimum prediction error $N_{\mathrm{KL}}^{\min}$. In the context of uplink NOMA, the AAV compares the divergence between the transmitted and received radio signals.

B. Radio Environment Representation

We describe the radio environment using a generalized state-space model consisting of the following:

$$\tilde{\mathrm{S}}_{t,k}^{(e)} = \mathrm{f}\left(\tilde{\mathrm{S}}_{t-1,k}^{(e)}\right) + \mathrm{w}_{t,k}, \quad (22)$$

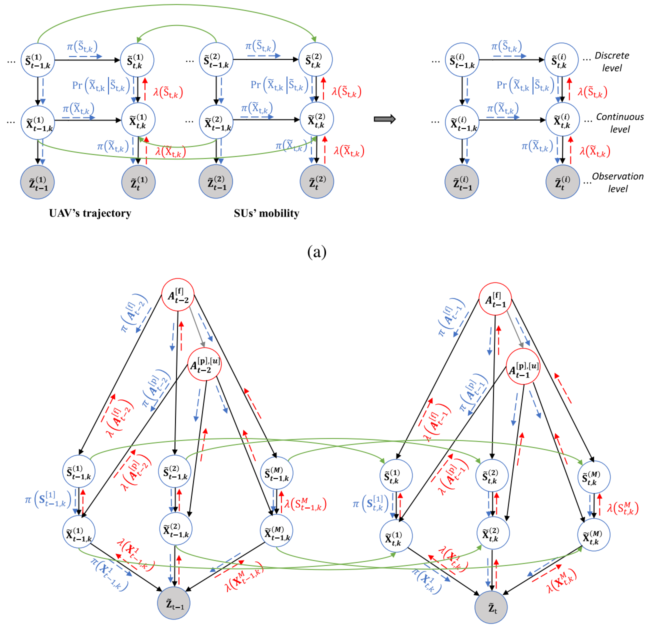
$$\tilde{\mathrm{X}}_{t,k}^{(e)} = \mathrm{C}\tilde{\mathrm{X}}_{t-1,k}^{(e)} + \mathrm{D}\mathrm{U}_{\tilde{\mathrm{S}}_{t,k}^{(e)}} + \mathrm{w}_{t,k}, \quad (23)$$

$$\tilde{\mathrm{Z}}_{t,k} = \mathrm{H}\left(\tilde{\mathrm{X}}_{t,k}^{(1)} + \cdots + \tilde{\mathrm{X}}_{t,k}^{(M)} + \tilde{\mathrm{X}}_{t,k}^{(pu)}\right) + \mathrm{v}_{t,k}, \quad (24)$$

where (22) denotes discrete random variables representing the clusters of the discrete states, while the sub-channel conveying the signal with power level is $\tilde{\mathrm{S}}_{t,k}^{(e)}$. The dynamic transition model of $\tilde{\mathrm{S}}_{t,k}^{(e)}$ change along with (22), where f(.) represent a non-linear function that states how $\tilde{\mathrm{S}}_{t,k}^{(e)}$ change with time depending on $\tilde{\mathrm{S}}_{t-1,k}^{(e)}$ and $\mathrm{w}_{\mathrm{t,k}}$ denotes the system process noise given by $\mathrm{w}_{t,k}\sim\mathcal{N}(0,\Sigma_{\mathrm{w}_{t,k}})$. The dynamic model equation stated in (23) defines how the continuous Generalized States (GS) $\tilde{\mathrm{X}}_{t,k}^{(e)}$ change with time depending on $\tilde{\mathrm{X}}_{t-1,k}^{(e)}$ and $\tilde{\mathrm{S}}_{t,k}^{(e)}$, stated as $e \in \{no, pu, c, 1, \ldots, M\}$, where no, pu, and c denote the noise model, PU model and superimposed SU signals. The matrices C and D encode the dynamic behavior and control principles, respectively, while the vector $\mathrm{U}_{\tilde{\mathrm{S}}_{t,k}^{(e)}}$ represents the control inputs. Equation (24) describes the observation model and how the sensory signals depend on the GS.

C. The Proposed Active-GDBN

The proposed graphical representations of the Active-GDBN model at different hierarchical levels (discrete and continuous levels) are shown in Fig. 2. As presented in Fig. 2(a), the process involves an offline stage (i.e., perceptual learning of preferred observation), where the AAV is equipped with an interactive coupled-state switching GDBN structure, which it uses to learn a generative model that encodes joint subchannel assignment, power allocation, and the AAV's mobility. Concurrently, the AAV also learns the PU and noise models. The coupled state switching model [64] allows the AAV to efficiently model the temporal relationships between its trajectory and the mobility patterns of the SUs. These models are designed to handle multiple observation sequences where the causal state variables interact. Specifically, a hidden discrete state $\tilde{\mathrm{S}}_{t,k}^{(1)}$ depends on its previous state $\tilde{\mathrm{S}}_{t-1,k}^{(1)}$ and the prior state of another hidden chain $\tilde{\mathrm{S}}_{t-1,k}^{(2)}$. Similarly, a continuous hidden state $\tilde{\mathrm{X}}_{t,k}^{(1)}$ is affected by its previous state $\tilde{\mathrm{X}}_{t-1,k}^{(1)}$ and the preceding state of another hidden chain $\tilde{\mathrm{X}}_{t-1,k}^{(2)}$.



(a)

(b)

Fig. 2. Graphical representations of the proposed method: (a) GDBN, represents the offline phase (AAV's perception). (b) Active-GDBN, is the online deployment phase (AAV's active inference). The model emphasizes the continuous causality of neuronal message passing at discrete and continuous coupled-states, which enable the AAV to estimate the posterior across hidden states.

Fig. 2-b illustrates the online active inference phase, during which the AAV is required to determine a joint decision across all actions. This involves identifying the optimal combination of subchannel assignment $b_{k,n}$, power allocation $p_{k,n}$, and AAV position at each time slot $q(t)$ to maximize the overall sum rate. The model allows the AAV to map each discrete variable to a continuous variable, enabling the evolution of constant neuronal message passing and belief updating at different hierarchical levels. As indicated in Fig. 2(a) and Fig. 2(b), the blue arrows represent descending messages or signals from discrete to continuous states, constituting the AAV's prior states, while the red arrows signify ascending messages, which encompass the AAV's future states. Essentially, both prior and future states are continuously represented over time as new observations become available. The green arrows connect previous states to present states, indicating the temporal evolution and causality of the proposed model. These arrows capture successive state transitions and dependencies, allowing the AAV to estimate posterior beliefs efficiently.

D. The Perceptual Learning Process (Offline Training)

During the training phase, the AAV starts sensing the radio space to learn the diverse vocabularies representing the different models, namely, the noise, PU, and the superimposed SUs.

Note that the superimposed SUs encode the interdependencies among all possible subchannel assignments, power allocations, and the AAV's trajectory. In the initial learning phase, the AAV is equipped with an initial model based on the Unscented Kalman Filter (UKF) [11], which presumes that the external states evolve with static rules and depend on (23) for predicting the continuous states, such that $\mathrm{U}_{\tilde{\mathrm{S}}^{(e)}_{t,k}} = 0$ [21]. As a result, the AAV's brain generates initial errors called generalized errors (GEs) from the received sensory inputs.

We examine I separate observations $\{\tilde{\mathrm{Z}}_t^{(i)}\}_{t=1}^T$, each dependent upon the hidden environmental state sequence across both discrete and continuous states $\{\tilde{\mathrm{S}}_t^{(i)}\}_{t=1}^T$, $\{\tilde{\mathrm{X}}_t^{(i)}\}_{t=1}^T$, $i = \{1, \ldots, I\}$. These hidden state variables engage in interactions within discrete and continuous levels. The transition probabilities associated with the state vectors are defined as follows:

$$\Pr\left(\tilde{\mathrm{S}}_t|\tilde{\mathrm{S}}_{t-1}\right) = \Pr\left(\tilde{\mathrm{S}}_t^{(1)}, \ldots, \tilde{\mathrm{S}}_t^{(I)}|\tilde{\mathrm{S}}_{t-1}^{(1)}, \ldots, \tilde{\mathrm{S}}_{t-1}^{(I)}\right), \quad (25)$$

$$\Pr\left(\tilde{\mathrm{X}}_t|\tilde{\mathrm{X}}_{t-1}\right) = \Pr\left(\tilde{\mathrm{X}}_t^{(1)}, \ldots, \tilde{\mathrm{X}}_t^{(I)}|\tilde{\mathrm{X}}_{t-1}^{(1)}, \ldots, \tilde{\mathrm{X}}_{t-1}^{(I)}\right). \quad (26)$$

As illustrated in Fig. 2(a), in this scenario (where $I = 2$), correspond to the AAV's trajectory and the SU's mobility at each time slot.

We used an unsupervised clustering technique known as Growing Neural Gas (GNG) [65] for training the coupled GDBN. GNG is a type of artificial neural network that can automatically learn and model the underlying structure of the input data through a cooperative and incremental learning process. The GNG takes as input the GEs data and generates discrete state clusters along with a set of generalized state clusters at the continuous level.

E. Active Inference Phase

We define the relationship between the AAV and the radio environment as a six-element tuple ($\tilde{\mathbf{S}}_\mathbf{t}$, $\tilde{\mathbf{X}}_\mathbf{t}$, $\boldsymbol{\mathcal{A}}$, $\mathbf{T}_{\boldsymbol{\tau}}^{\boldsymbol{pu}}$, $\boldsymbol{\Pi}_{\boldsymbol{\tau}}^{\boldsymbol{a}}$, $\tilde{\boldsymbol{Z}}_{\boldsymbol{t}}$), where $\tilde{\mathbf{S}}_\mathbf{t}$ and $\tilde{\mathbf{X}}_\mathbf{t}$ are hidden environmental states, including noise, PUs, and SUs. $\boldsymbol{\mathcal{A}} = \{\mathcal{A}^{[\mathrm{f}]}, \mathcal{A}^{[p]}, \mathcal{A}^{[u]}\}$ denotes the AAV's action space, comprising the optimal subchannel assignment, power allocation, and AAV trajectory. $\boldsymbol{T}_{\boldsymbol{\tau}}^{\boldsymbol{pu}}$ represents the temporal transition model of the PU. $\boldsymbol{\Pi}_{\boldsymbol{\tau}}^{\boldsymbol{a}}$ denotes the active inference probabilistic table that encodes the state-action pairs, while $\tilde{\mathbf{Z}}_{\boldsymbol{t}}$ represents the observations (sensory signals).

1) Initialization: We initialize three distinct matrices that the AAV utilizes for making online decisions. $\boldsymbol{\Pi}_{\boldsymbol{\tau}}^{[\mathbf{f}]}$ is a time-varying matrix that encodes the probabilistic interactions between states and the AAV's discrete actions, which is subchannel assignment. Also, $\boldsymbol{\Pi}_{\boldsymbol{\tau}}^{[\mathbf{p}]}$ and $\boldsymbol{\Pi}_{\boldsymbol{\tau}}^{[\mathbf{u}]}$ represent time-varying matrices that encode the probabilistic relationships between states and the AAV's continuous actions and power allocation, respectively,

2) Action Selection Process: At the start of every time slot, the AAV transitions into a particular state, characterized by its present location and the current channel conditions. Initially, the AAV selects random discrete actions since all the possible discrete actions have equal probability ($\frac{1}{K}$) of being selected and thus, the AAV selects initial continuous power actions defined as $A_{t-1}^{[\mathrm{p}]} = A_0^{[p]}$. The performed action in $A_{t-1}^{[\mathrm{f}]}$ determines the next environmental states (discrete and continuous states), $\tilde{\mathrm{S}}_{t,k}$, $\tilde{\mathrm{X}}_{t,k}$ which are described by $\Pr(\tilde{\mathrm{S}}_{t,k}|\tilde{\mathrm{S}}_{t-1,k}, A_{t-1}^{[\mathrm{f}]})$ and $\Pr(\tilde{\mathrm{X}}_{t,k}|\tilde{\mathrm{X}}_{t-1,k}, A_{t-1}^{[p]})$. In the subsequent episodes, the AAV adapts the action selection procedure by predicting the future states of the PU in accordance with $\boldsymbol{T}_{\boldsymbol{\tau}}^{\boldsymbol{pu}}$ and avoiding the sub-channels with a high probability of being occupied by the PU.

3) Perception and Joint State-Prediction: Once the AAV has executed its combined set of actions, based on its flight trajectory, allocated subchannels, and power levels, the AAV can predict perfectly the effect of its actions by utilizing a M-MJPF [63]. The AAV also senses and observes the other unselected sub-channels concurrently to determine their states (occupied or unoccupied) to improve future decisions.

The M-MJPF adopts a switching strategy by applying Particle filtering (PF) for estimations and updates at the discrete level, and Kalman filtering (KF) for estimations and updates of weights at the continuous level. Using the dynamics of causal relationships, we can differentiate a top-down inference from a bottom-up inference. The temporal top-down inter-slice predictive messages $\pi(\tilde{\mathrm{X}}_{t,k})$ and $\pi(\tilde{\mathrm{S}}_{t,k})$ is a function of the knowledge obtained in the dynamic model. The bottom-up intra-slice inference comprises of ascending transmitted messages $\lambda(\tilde{\mathrm{X}}_{t,k})$ and $\lambda(\tilde{\mathrm{S}}_{t,k})$ using the likelihood function toward the discrete state. The continuous level prediction is based on discrete states. For every propagated particle in the discrete form, a KF is initiated to estimate the corresponding continuous state. The PF propagates L particles with equal weights using a proposal density coded in the transition matrix Π_k. The posterior probability related with $\tilde{\mathrm{X}}_{t-1,k}$ is stated as $\pi(\tilde{\mathrm{X}}_{t,k}) = \Pr(\tilde{\mathrm{X}}_{t,k}, \tilde{\mathrm{S}}_{t,k}|\tilde{\mathrm{Z}}_{t-1,k})$. As the AAV obtains new observations, predictive messages are propagated bottom-up to update the AAV's belief at various levels. As a result, the posterior is updated with:

$$\Pr\left(\tilde{\mathrm{X}}_{t,k}, \tilde{\mathrm{S}}_{t,k}|\tilde{\mathrm{Z}}_{t,k}\right) = \pi\left(\tilde{\mathrm{X}}_{t,k}\right)\lambda\left(\tilde{\mathrm{X}}_{t,k}\right), \quad (27)$$

where $\lambda(\tilde{\mathrm{X}}_{t,k}) = \Pr(\tilde{\mathrm{Z}}_{t,k}|\tilde{\mathrm{X}}_{t,k})$. Moreover, the likelihood message $\lambda(\tilde{\mathrm{S}}_{t,k})$ is employed to update particle weights W given by:

$$W_t^l = W_t^l \lambda\left(\tilde{\mathrm{S}}_{t,k}\right), \quad (28)$$

4) Abnormality Measurements and Action Evaluation: The continuous state error measurement quantifies the divergence between two signals arriving at node $\tilde{\mathrm{X}}_{t,k}$, represented by $\pi(\tilde{\mathrm{X}}_{t,k})$ and $\lambda(\tilde{\mathrm{X}}_{t,k})$. This metric evaluates how well the observation aligns with the predictions by calculating the Kullback-Leibler (KL) divergence, as described in (21). A higher KL value implies a larger discrepancy between the actual observations and the model's predictions, indicating higher errors. Conversely, a smaller value signifies closer agreement between the observations and predictions, suggesting a more accurate prediction with a lower error.

5) Updating the AAV's Actions and Incremental Learning: The AAV can perceive the radio space through observations (i.e., sensory signals) and modify the radio space through actions. During exploitation, the AAV chooses actions that

Algorithm 2 Online Active Inference (Resource Allocation)

```
1: Initialization: Initialize the $K \times K$ matrix ($\mathrm{T}_{\tau}^{pu}$) that encodes the likely transitions of the PUs between the different sub-channels,
2: Initialize the time-changing transition matrix $\boldsymbol{\Pi}_{\boldsymbol{\tau}}^{[\mathrm{f}]} \in \mathbb{R}^{K, C_{\mathrm{f}}}$ that encodes the probability dependencies of states and the AAV's discrete actions,
3: Initialize the time-changing transition matrices $\boldsymbol{\Pi}_{\boldsymbol{\tau}}^{[\mathrm{p}]} \in \mathbb{R}^{1}$ and $\boldsymbol{\Pi}_{\boldsymbol{\tau}}^{[\mathrm{u}]}$ that encode the probability dependencies of states and the AAV's continuous power actions and mobility, respectively.
4: Initialize the particle filter (PF) and Kalman filter (KF) parameters, such as the number of particles, proposal density, observation model, state transition matrix, and noise covariance matrices.
5: Action selection: Discrete actions are selected according to:
   $$A_{t-1}^{[\mathrm{f}]} = \begin{cases} \mathrm{randint}\Big(1, N_f\Big) \text{ with probability } \frac{1}{K} \rightarrow 1^{st} \text{ iteration} \\ \mathrm{argmax}_{\tilde{\mathrm{S}}_{t-1,k}, \mathrm{T}_{\tau}^{pu}}\left(\tilde{\mathrm{S}}_{t-1,k}\right) \pi\left(A_{t-1}^{[\mathrm{f}]}\right) \rightarrow \text{successive} \\ \text{iterations}. \end{cases}$$
   Continuous actions are selected following:
   $$A_{t-1}^{[\mathrm{p}]} = \begin{cases} A_0^{[p]} \rightarrow 1^{st} \text{ iteration} \\ A_{t-1}^{[p]} + \beta \mathbb{E}\left(\left|\tilde{\mathcal{E}}_{A_{t-1}^{[p]}}\right|\right) \rightarrow \text{successive iterations}. \end{cases}$$
6: Perception and joint state-prediction: Propagate L particles using the PF based on the transition matrix and proposal density given by: $(\tilde{\mathrm{S}}_{t,k}^{l}, W_t^{l}) \sim (\Pi_k, \frac{1}{L})$.
7: For each propagated particle in the discrete form, initiate a KF to estimate the corresponding continuous state.
8: Abnormality measurements and belief updating: Compute the continuous state abnormality according to (21).
9: Update the KF and PF parameters based on the updated GS posterior and particles' weights to improve the estimation accuracy according to (27) and (28).
10: Exploit the generalized errors (GEs) in each episode to learn new actions that minimize the prediction errors.
11: Repeat steps 5 to 10 for each episode until convergence.
```

minimize free energy based on its current model and observations. For exploration, it selects actions that minimize expected free energy, which represents the updated (new) model arising from the chosen action. Subsequently, the AAV incrementally updates its POMDP model to incorporate the newly learned actions.

Algorithm 2 illustrates the detailed steps of the proposed algorithmic architecture involving active inference. Furthermore, the proposed data-driven solution operates primarily with minimal online signaling since it leverages on the learned generative model to perform online resource allocation. During the offline data generation phase, we considered signaling information such as CSI and AAV trajectory data. This model is learned during training and requires no additional signaling. The main advantage of the proposed method is that online signaling is substantially reduced as the agent encodes the state dynamics during training. However, while online signaling is not completely eliminated, essential signaling may be required for more complex and dynamic scenarios, which we summarized in Table II.

F. Complexity Analysis

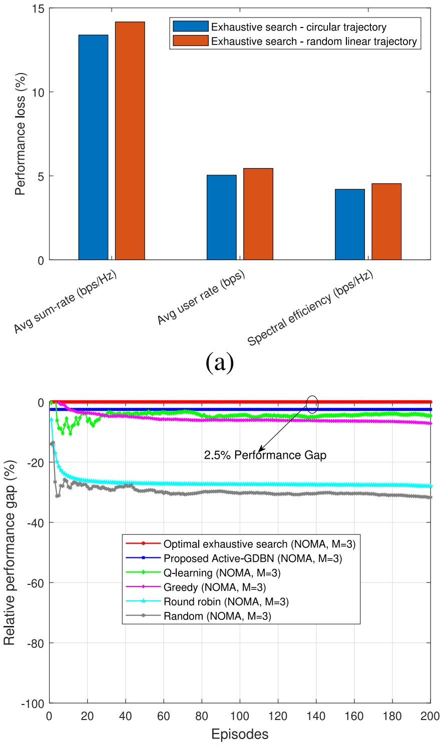
The complexity of the proposed Active-GDBN arises primarily from estimating the M-MJPF. The switching strategy adopted by M-MJPF introduces a complexity of $\mathcal{O}(K^2N)$, where K represents the number of switching observation models, and N denotes the number of particles. Updates to the hidden states and the covariance matrix, which are based on incoming new data, have a computational complexity of $\mathcal{O}(N^2)$, at each time step. This implies that computational complexity can be minimized by reducing the number of propagated particles and the sample size; however, this may compromise the model's accuracy. It is important to note that in our scenario, the discrete level consists of a finite number of states. Therefore, unlike in continuous spaces, where a large number of particles may be required for accurate representation, a smaller number of particles is sufficient to accurately represent the posterior with low complexity.

IV. Simulation Results

In this section, we present simulation results to demonstrate the performance of our proposed Active-GDBN. We consider a cell with a AAV in the center, a PBS, and SUs randomly distributed within a cell radius of 500 *m*. The SUs are actively communicating with the AAV, seeking resources. To reduce complexity, we assumed a single PU. The total time of flight is set to $T = 60$ s, such that $T/L = 1$ s, $V_{\max} = 30\mathrm{m/s}$. Similar to [46], the flight radius is 158*m*. The other network parameters are summarized in Table I. We select default simulation settings using existing literature that examined similar wireless networks [46]. Our simulation results are provided using Intel core i7-12700 CPU with 2.90 GHz frequency, 16 GB of Random Access Memory (RAM), and a windows-11 64-bit operating system. The experiments are simulated using MATLAB R2023b. Some parameters of the GNG (Growing Neural Gas) clustering algorithm network are defined, including the maximum number of neurons, the learning rate for the winning node, and that of its topological neighbors, defined as ε_b and ε_n respectively.

A. Data Generation

We conduct an exhaustive search across all possible combinations of subchannel assignments and power allocations, considering both AAV circular and random linear motion. To generate all possible combinations, we use nested loops and the nchoosek function in MATLAB. For each combination, we calculate the sum rate across all time slots. If a higher sum rate is achieved, we update the optimal subchannel assignment, power allocation, accordingly. The optimal solution of the subchannel assignment and power allocation is stored along with the optimal cumulative sum rate from which the AAV will learn the Generative model. Fig. (3)-(a) shows the performance comparison of the exhaustive search for both circular and random linear AAV trajectories in terms of average sum rate, spectral efficiency, and average user rate. As indicated, the random linear trajectory demonstrates superior performance compared to the circular trajectory across all metrics, due to its random environment and greater coverage diversity. The random nature enables it to find more globally optimal user points for power allocation and subchannel assignment, which are not limited by a fixed circular path. Additionally, the diverse coverage areas allow the random linear trajectory to explore suitable spots with favorable channel conditions for multiple NOMA users. Given its superiority over the circular trajectory, we adopt the generalized random linear AAV trajectory for the subsequent simulation results.



(a)

(b)

Fig. 3. Performance comparison of exhaustive search for both circular and randomly linear AAV trajectories (a), Performance gap of the proposed Active-GDBN with exhaustive search and baseline methods (b), when $M = 5$, $H = 100$ m, $P_{max} = 20$ Watts.

TABLE I
NETWORK PARAMETERS OF THE PROPOSED SCHEME

UAV height H	100 m [66]
Modulation scheme of PU	BPSK
Modulation scheme of SUs	QPSK
Path loss model	3GPP [49]
Noise power	-174 dBm/Hz
System Bandwidth, B_w	1.4 MHz
Number of sub-channels K	6
Power budget of SUs P_{max}	20 W [67]
Number of SUs multiplexed per sub-channel M	$M = [2, 3, 4, 5, 6]$, $M = 1$ (OMA)
Power difference threshold P_{th}	1
Learning rate of GNG clustering	0.01

TABLE II
ONLINE SIGNALING PROCESS OF ACTIVE-GDBN

Stage	**Type of Signaling**	**Description**	**Frequency**
Initialization	Network State	channel conditions, user positions, power limits	Per episode
Online Active-GDBN	CSI	Updated UAV-user channel gains	Per episode
Online Active-GDBN	User Feedback	SINR, achievable data rates	Per episode
Online Active-GDBN	UAV Status	Current position and velocity	Per episode
Optimization	Resource Allocation	Subchannel assignment and power decisions	Per episode
Model Update	Updated Parameters	New model weights	Per episode

TABLE III
NETWORK PARAMETERS OF DL

No. of sub-channels K	6
No. of DNN layers	5
No. of epochs	200
Learning rate	0.001
No. of training data	$10,000$
Mini-batch size	32
Parameter M	3

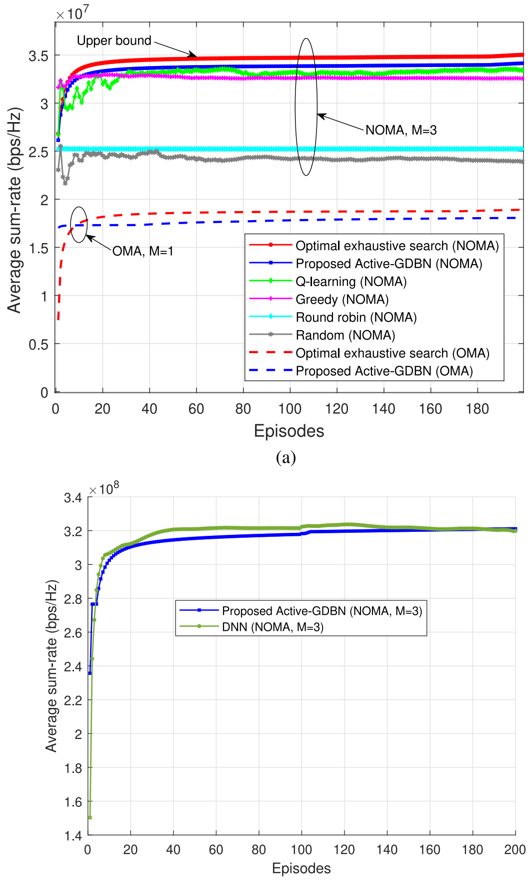
Fig. (3)-(b) indicates the near-optimal relative performance gap of the proposed Active-GDBN compared to the optimal solution. In the proposed Active-GDBN, the agent continuously infers belief states and learns a generative model, which it uses to encode its knowledge of the wireless environment. This enables the agent to make appropriate decisions regarding power allocation and subchannel assignment. The proposed solution achieves a 2.5% performance gap near the optimal solution.

B. Benchmark Schemes

To evaluate the performance of our proposed algorithm, we compare it against various benchmark schemes. These include the exhaustive search optimization algorithm, which serves as an upper-bound optimal solution [2]; deep learning (DL) [68]; Q-learning [69], a model-free reinforcement learning (RL) algorithm effective in unknown and complex environments; and the greedy, round-robin, and random schemes, presented as feasible strategies. In all cases, we adapt the algorithms for resource allocation in a AAV-based uplink NOMA scenario. In the DL study, we train a deep neural network (DNN) offline to learn the nonlinear input of the exhaustive search optimization, which includes subchannel assignment and power allocation based on AAV position in each time slot. The aim is to map the signal at the receiver with the equivalent transmitted signal during the online stage. For this work, the DNN has five layers: the input layer, the long short-term memory layer (LSTM), which is a type of recurrent neural network (RNN) for exploiting time dependencies in data, a fully connected layer, the softmax function, and the classification layer. The main network parameters are presented in Table II. Furthermore, with Q-learning, by employing the formulation of the Markov Decision Process (MDP), the AAV is capable of finding an optimal policy to satisfy the maximum sum rate of users. During the learning process, the environmental state of the users is assumed to be randomly changing from one time slot to the next. This implies that, given the positions of users, the AAV learns to adapt optimally using the epsilon-greedy optimization policy. The AAV has knowledge of the positions and states of SUs. Thus, in each episode, the AAV computes a cumulative reward equivalent to the sum rate of SUs.

C. Convergence of Proposed Scheme

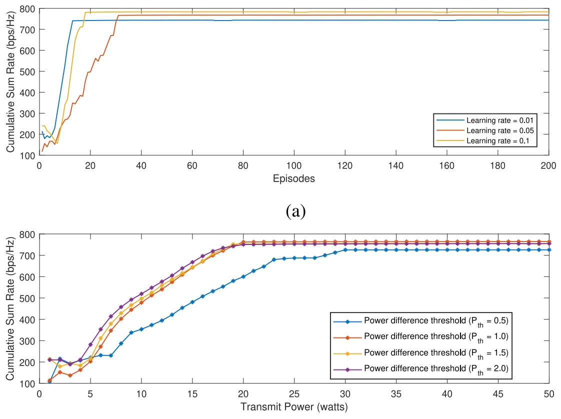
To assess the performance of our algorithm, we evaluate and compare its convergence. As shown in Fig. 4-(a), the Active-GDBN framework achieves near-optimal performance to the upper bound (exhaustive search) and converges to a stable sum rate more quickly than the Q-learning and greedy algorithms, also outperforming other baselines. Unlike the exhaustive search strategy, which requires exploring every possible combination to obtain the optimal solution, the Active-GDBN agent utilizes a learned probabilistic generative model to predict and select the optimal configuration with reduced computational complexity. The AAV dynamically adapts to changes in the environment by continuously updating its predictions (beliefs) based on new sensory observations. The inability to avoid undesired states results in a strong influence



(a)

(b)

Fig. 4. Convergence comparison of the proposed Active-GDBN with exhaustive search, baseline methods, and OMA (a), with DNN (b), when $M = 3$, $H = 100$ m, $P_{max} = 20$ Watts.



(a)

(b)

Fig. 5. Performance analysis of proposed Active-GDBN: (a) Cumulative sum rate across various GNG learning rates, (b) Impact of varying P_{th}. In both cases, $M = 5$, $V_{max} = 30\ m/s$, $P_{max} = 20$ Watts.

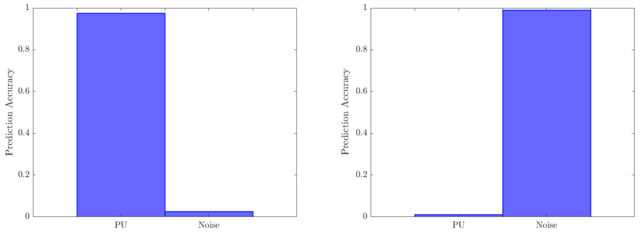
of negative rewards; as a result, Q-learning requires more training episodes to achieve faster and stable convergence. The greedy heuristic, though fast and computationally affordable, performs locally optimal decisions in each episode and does not accumulate knowledge over time, leading to lower performance. Similarly, the inherent lack of online adaptability and continuous feedback-action mechanisms results in weaker sum rate performance for the round-robin and random schemes. Furthermore, we compare the proposed method and the optimal solution with OMA. The results provide a detailed assessment, highlighting the advantages of NOMA over OMA. Fig. 4-(b) demonstrates the convergence performance of the proposed Active-GDBN with the DNN model. The optimality gap between the DNN model and the proposed model is almost negligible.

D. Impact of Learning Rate and Power Difference Threshold (P_{th})

We use the GNG unsupervised clustering algorithm to learn a dynamic topology of our sensory input with variable neurons. The learning rate is used to update the weight of each neuron and affects the performance of the Active-GDBN algorithm. The learning rate controls how quickly the Active-GDBN adapts to the sensory input. From Fig. 5-(a), it is clear that choosing a lower learning rate generates faster convergence, but at the expense of a lower sum rate. On the other hand, choosing a higher learning rate generates a higher sum rate, but at the cost of a slower convergence time. We also observed that the higher learning rate exhibits noisy signal oscillations. As revealed in Fig. 5-(b), the power difference threshold is critical for SIC decoding. A small value of (P_{th}) can cause SIC decoding failure [70]. So, finding the best power difference threshold between the superimposed signals is essential for successful SIC decoding. We evaluated the sum rate versus the transmit power of the SUs for the variable (P_{th}). It is apparent that as the transmit power of the SUs increases, the sum rate increases to a certain level and converges because the power control of the SUs does not exceed 20 W. We also observed that if the power difference threshold is less than or equal to a minimum threshold value (i.e., 0.5), the SIC decoding of the signals may fail, and the AAV will not be able to separate the signals since the superimposed signal constellation points overlap. Furthermore, as the power difference threshold approaches the value of 1, the performance of the sum rate increases significantly. However, when we increased the value to 1.5 and 2.0, there was no significant improvement in the performance of the sum rate. Thus, to obtain the optimal sum rate value, we set $(P_{th} = 1)$ for all simulation settings.

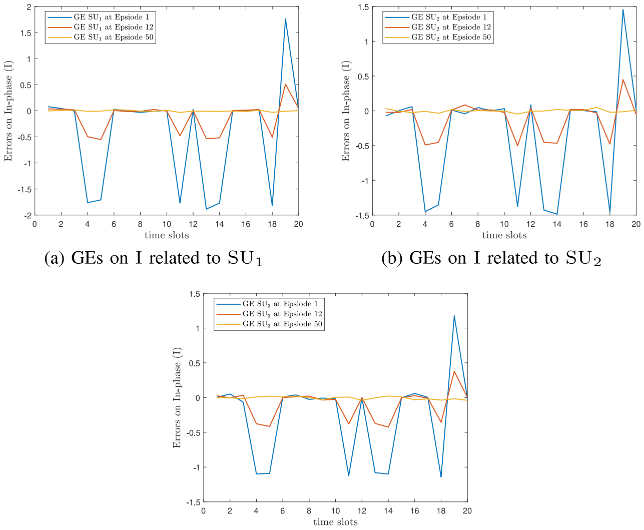
E. Prediction Accuracy of Sub-Channel Occupancy

Fig. 6 shows the average prediction accuracy of sub-channel occupancy (i.e., occupied and vacant sub-channels). The AAV relies on the learned vocabularies representing the PU and noise models during the perceptual learning phase to decide whether the sub-channel is vacant or occupied by PU. During the online deployment stage, the AAV performs multiple predictions based on the two vocabularies (i.e., PU and noise models) and generates two abnormality indicators. Then, the AAV evaluates the model that produces the minimum abnormality. So, if the PU model generates the minimum abnormality, this means that the current observations obtained by sensing a certain sub-channel support the PU model and so that sub-channel is occupied by the PU. Likewise, if



(a) (b)

Fig. 6. Average prediction accuracy over: (a) Occupied sub-channels, (b) Vacant sub-channels.



(a) GEs on I related to SU$_1$

(b) GEs on I related to SU$_2$

(c) GEs on I related to SU$_3$

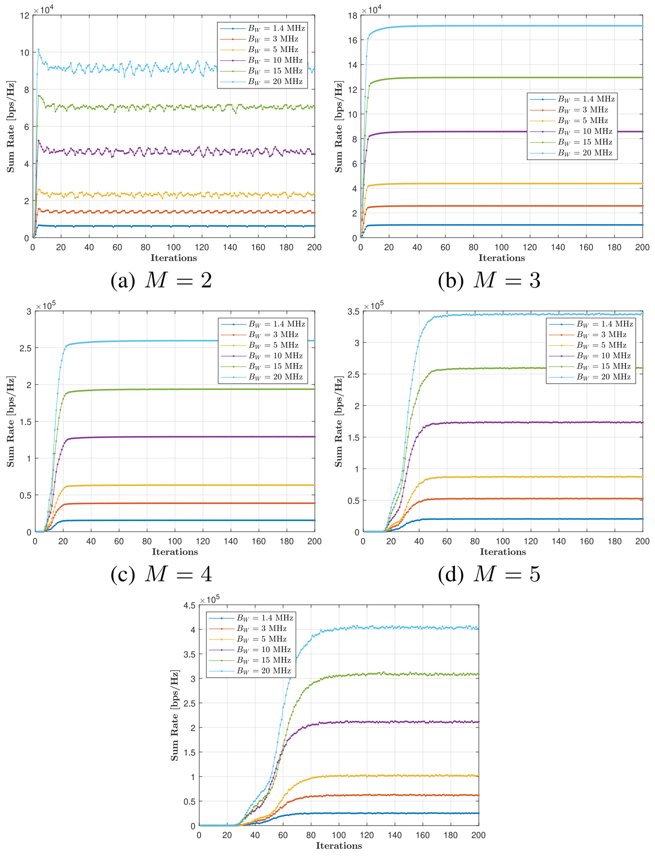
Fig. 7. In-phase prediction errors of dynamic power allocation for each SU.

the noise model generates the minimum abnormality, the AAV can deduce that the predictions generated by the noise model match the current observations and so that sub-channel is vacant. As shown in Figs. 6-(a) and (b), the AAV can accurately predict the PU and noise occupancy with almost 100% accuracy.

F. Generalized Errors (GEs) of Dynamic Power Allocation

Fig. 7 shows the GEs of dynamic power allocation on the in-phase related to each SU signal (SU1, SU2, and SU3). Three SUs are chosen as examples for validation. The AAV can reach the optimal power allocation by exploiting the GEs.

During the first episode, the AAV's actions produced a high number of errors and abnormalities as indicated by the blue colours. The AAV exploits the errors in each episode to update its actions by adapting the power values to reach the targeted value of zero. As illustrated in Figs. 7-(a), 7-(b), and 7-(c), We can see that in 1 episode, the errors are high; in 12 episodes (red), the errors start to minimize towards zero; and in 50 episodes (orange), the errors are minimized to almost zero at convergence.



(a) $M = 2$

(b) $M = 3$

(c) $M = 4$

(d) $M = 5$

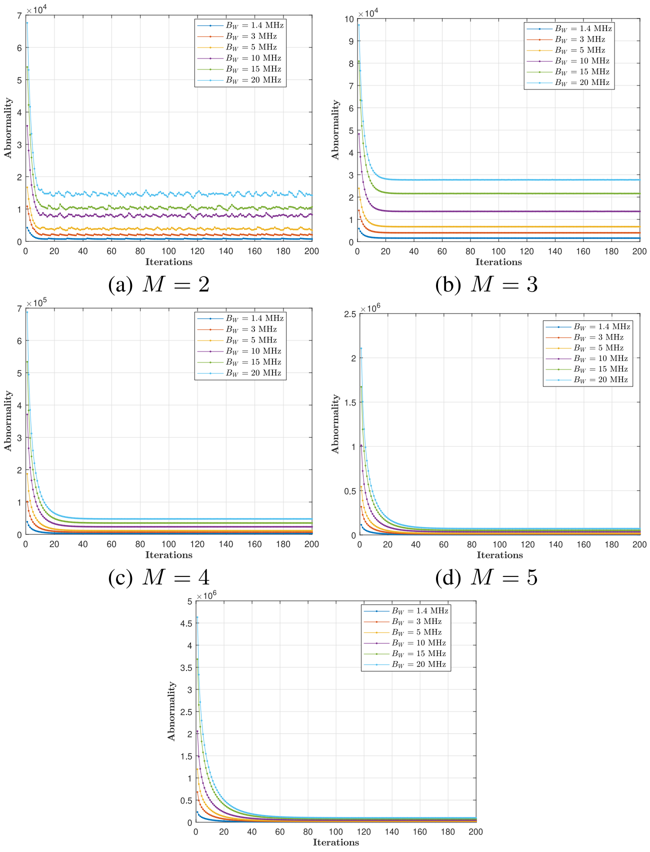
(e) $M = 6$

Fig. 8. Convergence behaviour of the proposed Active-GDBN in maximizing the achievable sum-rate with different bandwidth (B_w) and number of SUs (M) multiplexed on the same sub-channel.

G. Impact of the Number of Subchannels on the Performance and the Achievable Number of Served Users

The Active-GDBN algorithm's convergence behavior is illustrated in Fig. 8. Its objective is to maximize the achievable sum-rate, taking into account different bandwidths and the number of SUs (i.e., M) that can be multiplexed on a sub-channel. The proposed approach shows flexibility, as the time required to reach convergence remains relatively constant, even with an increase in the number of subchannels. This is possible because the algorithm can allocate resources to multiple users in parallel and uses multiple generalized errors to correct the initial allocations made by the AAV. However, as the number of SUs to be multiplexed on the same subchannel increases, the time required to reach convergence also increases. This is because the proposed method needs more time to regulate the power values allocated to the multiplexed users on a certain subchannel. Therefore, the more users there are to multiplex, the more time the algorithm requires to reach the preferred power values.

The graph in Fig. 9 demonstrates how the proposed method manages to minimize the abnormality signals. Interestingly, we can see that the convergence behavior of the abnormality signals is opposite to that of the sum rate shown in Fig. 8. In other words, the proposed approach aims to maximize the sum rate while minimizing the abnormality. The abnormality

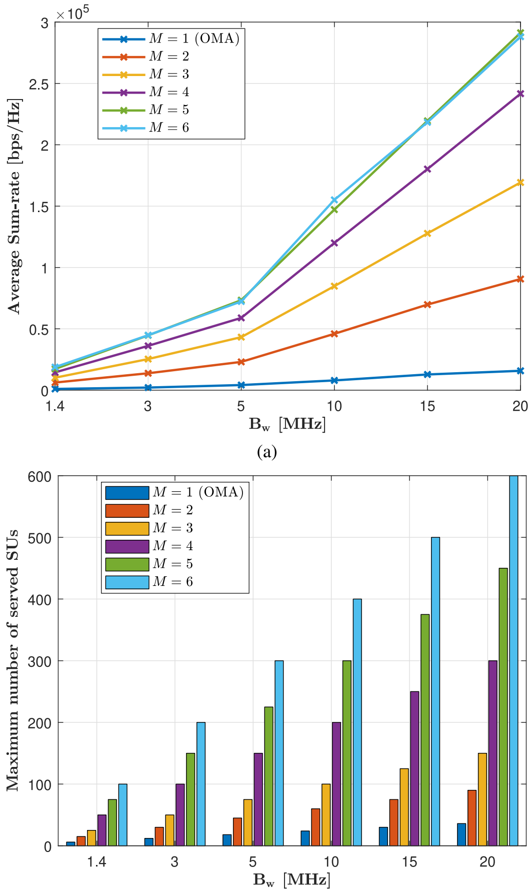


(a) $M = 2$ (b) $M = 3$

(c) $M = 4$ (d) $M = 5$

(e) $M = 6$

Fig. 9. Convergence behaviour of the proposed Active-GDBN in minimizing the abnormalities with different bandwidth (B_w) and number of SUs (M) multiplexed on the same sub-channel.



(a)

(b)

Fig. 10. Illustration of the achievable sum rate and number of served users when $M = 6$, $P_{max} = 20$ Watts: (a) The achievable sum-rate of the proposed Active-GDBN versus the number of subchannels (BWs), (b) The achievable number of served users versus the number of sub-channels.

signals are essential in evaluating the allocation policy adopted by the AAV through active-GDBN. Despite increasing the bandwidth, we observe that the convergence time of the abnormality signals remains constant due to the parallel evaluation of power allocation strategies among multiple users. However, the abnormality values increase with bandwidth, which leads to an increase in calculated errors. With increasing errors, the time to suppress those errors also increases to reach convergence.

Fig. 10-(a) illustrates the achievable sum rate for different numbers of users multiplexed on the same sub-channel versus the number of subchannels. As we increase the number of subchannels, the sum rate increases. Similarly, the sum-rate improves as we increase the number of SUs on the same sub-channel. The figure indicates that multiplexing $M = 5$ users on the same channel can achieve the highest sum rate. Attempting to multiplex more than $M = 5$ users would not improve the sum-rate. Therefore, the proposed approach is limited to a maximum of $M = 5$ secondary users on the same sub-channel. In Fig. 10-(b), we can observe the maximum number of SUs that can be served via different multiplexing mechanisms and a varying number of subchannels. As we increase the number of subchannels, the number of users served also increases. Furthermore, by multiplexing additional users on the same sub-channel, the count of served users can become double or even triple. It is crucial to note that these users can be served simultaneously using NOMA.

V. Conclusion

This paper proposes a novel data-driven framework based on active inference to jointly optimize subchannel assignment, power allocation, and AAV mobility in uplink AAV-aided cognitive NOMA systems. This joint objective function is usually difficult to solve analytically due to the complex time-varying dynamics of wireless networks. As a result, we transform the problem into prediction error minimization, employing a POMDP and generalized state space model to describe the time-changing radio environment, which we solve with Active-GDBN. By using the errors in each episode, the AAV is able to learn new actions that minimize future uncertainties (i.e., adapting its positions and resource allocation policies online to reach the target solution). The results demonstrate the feasibility of our proposed approach in achieving near-optimal sum rate performance relative to the optimal upper bound and surpassing other baselines. This study paves the way for more intelligent, explainable, and efficient resource allocation designs in future wireless networks, revealing the viability of active inference to resolve complicated optimization challenges in unknown and dynamic environments.

References

[1] A. Shahraki, M. Abbasi, M. Piran, and A. Taherkordi, "A comprehensive survey on 6G networks: Applications, core services, enabling technologies, and future challenges," 2021, *arXiv:2101.12475*.

[2] X. Wang, Y. Zhang, R. Shen, Y. Xu, and F.-C. Zheng, "DRL-based energy-efficient resource allocation frameworks for uplink NOMA systems," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7279–7294, Aug. 2020.

[3] Z. Lin, M. Lin, J.-B. Wang, T. de Cola, and J. Wang, "Joint beamforming and power allocation for satellite-terrestrial integrated networks with non-orthogonal multiple access," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 657–670, Jun. 2019.

[4] C. He, Y. Hu, Y. Chen, and B. Zeng, "Joint power allocation and channel assignment for NOMA with deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2200–2210, Oct. 2019.

[5] F. Obite, A. D. Usman, and E. Okafor, "An overview of deep reinforcement learning for spectrum sensing in cognitive radio networks," *Digit. Signal Process.*, vol. 113, Jun. 2021, Art. no. 103014.

[6] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, 3rd Quart., 2019.

[7] Z. Lin, M. Lin, T. de Cola, J.-B. Wang, W.-P. Zhu, and J. Cheng, "Supporting IoT with rate-splitting multiple access in satellite and aerial-integrated networks," *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11123–11134, Jul. 2021.

[8] K. An et al., "Exploiting multi-layer refracting RIS-assisted receiver for HAP-SWIPT networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 12638–12657, Oct. 2024.

[9] Y. Huang, W. Mei, J. Xu, L. Qiu, and R. Zhang, "Cognitive UAV communication via joint maneuver and power control," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7872–7888, Nov. 2019.

[10] T. Elmokadem and A. V. Savkin, "Towards fully autonomous UAVs: A survey," *Sensors*, vol. 21, no. 18, p. 6223, 2021.

[11] A. Krayani, A. S. Alam, L. Marcenaro, A. Nallanathan, and C. Regazzoni, "An emergent self-awareness module for physical layer security in cognitive UAV radios," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 2, pp. 888–906, Jun. 2022.

[12] C. S. Regazzoni, L. Marcenaro, D. Campo, and B. Rinner, "Multisensorial generative and descriptive self-awareness models for autonomous systems," *Proc. IEEE*, vol. 108, no. 7, pp. 987–1010, Jul. 2020.

[13] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for interference management," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5438–5453, Oct. 2018.

[14] M. Eisen, C. Zhang, L. F. O. Chamon, D. D. Lee, and A. Ribeiro, "Learning optimal resource allocations in wireless systems," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2775–2790, May 2019.

[15] D. Xu, X. Chen, C. Wu, S. Zhang, S. Xu, and S. Cao, "Energy-efficient subchannel and power allocation for HetNets based on convolutional neural network," in *Proc. IEEE 89th Veh. Technol. Conf.*, 2019, pp. 1–5.

[16] L. Kohoutová et al., "Toward a unified framework for interpreting machine-learning models in neuroimaging," *Nat. Protoc.*, vol. 15, no. 4, pp. 1399–1435, 2020.

[17] A. Tschantz, B. Millidge, A. K. Seth, and C. L. Buckley, "Reinforcement learning through active inference," 2020, *arXiv:2002.12636*.

[18] W. Zhang, Q. Wang, X. Liu, Y. Liu, and Y. Chen, "Three-dimension trajectory design for multi-UAV wireless network with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 600–612, Jan. 2021.

[19] K. J. Friston et al., "Reinforcement learning or active inference?" *PLOS ONE*, vol. 4, no. 7, 2009, Art. no. e6421.

[20] N. Sajid, P. Tigas, and K. Friston, "Active inference, preference learning and adaptive behavior," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1261, no. 1, 2022, Art. no. 12020.

[21] A. Krayani, A. S. Alam, L. Marcenaro, A. Nallanathan, and C. Regazzoni, "A novel resource allocation for anti-jamming in cognitive-UAVs: An active inference approach," *IEEE Commun. Lett.*, vol. 26, no. 10, pp. 2272–2276, Oct. 2022.

[22] K. Friston et al., "A free energy principle for the brain," *J. Physiol.-Paris*, vol. 100, nos. 1–3, pp. 70–87, 2006.

[23] K. J. Friston, N. Trujillo-Barreto, and J. Daunizeau, "DEM: A variational treatment of dynamic systems," *Neuroimage*, vol. 41, no. 3, pp. 849–885, 2008.

[24] M. T. Nguyen and L. B. Le, "NOMA user pairing and UAV placement in UAV-based wireless networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2019, pp. 1–6.

[25] Q. Wu, Y. Zeng, and R. Zhang, "Joint trajectory and communication design for multi-UAV enabled wireless networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2109–2121, Mar. 2018.

[26] M. Liu, G. Gui, N. Zhao, J. Sun, H. Gacanin, and H. Sari, "UAV-aided air-to-ground cooperative nonorthogonal multiple access," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2704–2715, Apr. 2020.

[27] Y. Sun, D. Xu, D. W. K. Ng, L. Dai, and R. Schober, "Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4281–4298, Jun. 2019.

[28] M. Alzenad, A. El-Keyi, and H. Yanikomeroglu, "3-D placement of an unmanned aerial vehicle base station for maximum coverage of users with different QoS requirements," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 38–41, Feb. 2018.

[29] A. V. Savkin and H. Huang, "Deployment of unmanned aerial vehicle base stations for optimal quality of coverage," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 321–324, Feb. 2019.

[30] W. Wang et al., "Joint Precoding optimization for secure SWIPT in UAV-aided NOMA networks," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5028–5040, Aug. 2020.

[31] Y. Gao, H. Tang, B. Li, and X. Yuan, "Robust trajectory and power control for cognitive UAV secrecy communication," *IEEE Access*, vol. 8, pp. 49338–49352, 2020.

[32] X. Jiang, Z. Wu, Z. Yin, Z. Yang, and N. Zhao, "Power consumption minimization of UAV relay in NOMA networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 666–670, May 2020.

[33] W. Mei and R. Zhang, "Uplink cooperative NOMA for cellular-connected UAV," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 644–656, Jun. 2019.

[34] Y.-F. Liu and Y.-H. Dai, "On the complexity of joint subcarrier and power allocation for multi-user OFDMA systems," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 583–596, Feb. 2014.

[35] X. You, C. Zhang, X. Tan, S. Jin, and H. Wu, "AI for 5G: Research directions and paradigms," *Sci. China Inf. Sci.*, vol. 62, pp. 1–13, Feb. 2019.

[36] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[37] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 93–99, Apr. 2019.

[38] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.

[39] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, 2018, pp. 298–303.

[40] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[41] T. Naous, M. Itani, M. Awad, and S. Sharafeddine, "Reinforcement learning in the sky: A survey on enabling intelligence in NTN-based communications," *IEEE Access*, vol. 11, pp. 19941–19968, 2023.

[42] B. K. S. Lima, R. Dinis, D. B. da Costa, R. Oliveira, and M. Beko, "User pairing and power allocation for UAV-NOMA systems based on multi-armed bandit framework," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13017–13029, Dec. 2022.

[43] P. K. Sharma and D. I. Kim, "UAV-enabled Downlink wireless system with non-orthogonal multiple access," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2017, pp. 1–6.

[44] M. Z. Hassan et al., "Interference management in cellular-connected Internet of Drones networks with drone-pairing and uplink rate-splitting multiple access," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 16060–16079, Sep. 2022.

[45] Z. Guan, S. Wang, L. Gao, and W. Xu, "Energy-efficient UAV communication with 3D trajectory optimization," in *Proc. 7th Int. Conf. Comput. Commun. (ICCC)*, 2021, pp. 312–317.

[46] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3747–3760, Jun. 2017.

[47] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2329–2345, Apr. 2019.

[48] F. Zhou, Y. Wu, R. Q. Hu, and Y. Qian, "Computation rate Maximization in UAV-enabled wireless-powered mobile-edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 1927–1941, Sep. 2018.

[49] "Enhanced LTE support for aerial vehicles," 3GPP, Sophia Antipolis, France, Rep. 36.777, 2018.

[50] T. Park, G. Lee, W. Saad, and M. Bennis, "Sum rate and reliability analysis for power-domain nonorthogonal multiple access (PD-NOMA)," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 10160–10169, Jun. 2021.

[51] L. Salaün, C. S. Chen, and M. Coupechoux, "Optimal joint subcarrier and power allocation in NOMA is strongly NP-hard," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2018, pp. 1–7.

[52] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[53] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ., 2005.

[54] M. Banagar and H. S. Dhillon, "3GPP-inspired stochastic geometry-based mobility model for a drone cellular network," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.

[55] F. Obite, A. Krayani, A. S. Alam, L. Marcenaro, A. Nallanathan, and C. Regazzoni, "Intelligent resource allocation for UAV-based cognitive NOMA networks: An active inference approach," in *Proc. IEEE Future Netw. World Forum (FNWF)*, 2023, pp. 1–7.

[56] P. C. Weeraddana, M. Codreanu, M. Latva-aho, A. Ephremides, and C. Fischione, "Weighted sum-rate maximization in wireless networks: A review," *Foundations Trends® Netw.*, vol. 6, nos. 1–2, pp. 1–163, 2012.

[57] S. K. Mahmud, Y. Liu, Y. Chen, and K. K. Chai, "Adaptive reinforcement learning framework for NOMA-UAV networks," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 2943–2947, Sep. 2021.

[58] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, "Optimal multiuser spectrum balancing for digital subscriber lines," *IEEE Trans. Commun.*, vol. 53, no. 12, pp. 2167–2167, May 2006.

[59] Z. Liu, X. Liu, V. C. M. Leung, and T. S. Durrani, "Energy-efficient resource allocation for dual-NOMA-UAV assisted Internet of Things," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3532–3543, Mar. 2023.

[60] K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, and G. Pezzulo, "Active inference and epistemic value," *Cogn. Neurosci.*, vol. 6, no. 4, pp. 187–214, 2015.

[61] Z. Chen et al., "Bayesian filtering: From Kalman filters to particle filters, and beyond," *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.

[62] P. M. Djuric et al., "Particle filtering," *IEEE Signal Process. Mag.*, vol. 20, no. 5, pp. 19–38, Sep. 2003.

[63] A. Krayani et al., "Self-learning Bayesian generative models for jammer detection in cognitive-UAV-radios," in *Proc. IEEE Global Commun. Conf.*, 2020, pp. 1–7.

[64] J. Pohle, R. Langrock, M. v. d. Schaar, R. King, and F. H. Jensen, "A primer on coupled state-switching models for multiple interacting time series," *Stat. Model.*, vol. 21, no. 3, pp. 264–285, 2021.

[65] I. J. Sledge and J. M. Keller, "Growing neural gas for temporal clustering," in *Proc. 19th Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.

[66] Y. Zheng and K.-W. Chin, "Joint trajectory and link scheduling optimization in UAV networks," *IEEE Access*, vol. 9, pp. 84756–84772, 2021.

[67] S. Wang, S. Cao, and R. Ruby, "Optimal power allocation in NOMA-based two-path successive AF relay systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, pp. 1–12, 2018.

[68] J. Thompson et al., "Deep learning for signal detection in non-orthogonal multiple access wireless systems," in *Proc. U.K./China Emerg. Technol. (UCET)*, 2019, pp. 1–4.

[69] Y. Huang, X. Mo, J. Xu, L. Qiu, and Y. Zeng, "Online maneuver design for UAV-enabled NOMA systems via reinforcement learning," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2020, pp. 1–6.

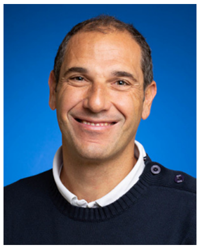
[70] Y. Tsuzuki, S. Shimamoto, and Z. Pan, "Evaluations of SIC by power difference in IM-NOMA," in *Proc. Wireless Days*, 2021, pp. 1–5.

Felix Obite (Member, IEEE) received the Postgraduate Diploma degree in electronics and telecommunication engineering from Ahmadu Bello University, Zaria, Nigeria, in 2012, the Master of Engineering degree in electrical, electronics, and telecommunication from Universiti Teknologi Malaysia in 2017, and the joint Ph.D. degree in interactive and cognitive environments, with a focus on wireless communication, from Queen Mary University of London and the University of Genoa, Italy, in October 2024. His doctoral research lies at the intersection of signal processing, information theory, explainable Bayesian inference, and neuroscience-inspired active machine learning, with a focus on emergent wireless communication systems. He also has research interests in reinforcement learning, federated learning, and semantic communication.

Ali Krayani (Member, IEEE) received the bachelor's degree in telecommunication engineering from the Politecnico di Torino, Italy, in 2014, the master's degree in telecommunication engineering from the University of Florence, Italy, in 2017, and the joint Ph.D. degree from the University of Genoa, Italy, and Queen Mary University of London, London, U.K., in April 2022.

He is an Assistant Professor with the Department of Electrical, Electronic, Telecommunications Engineering, and Naval Architecture (DITEN), University of Genoa. He has worked as a software engineer in various companies and was a Postdoctoral Research Fellow with DITEN, University of Genoa from 2021 to 2023. His current research interests include integrated sensing and communication, cognitive radios, AI-enabled radios, wireless communications (5G and 6G), AAV communications, physical layer security, IoT, semantic communications, AAV swarms, NOMA, federated learning, and artificial intelligence. In 2023, he received the Best Paper Award at the IEEE Wireless Communications and Networking Conference. He serves as a guest editor and a reviewer for several academic journals.

Atm S. Alam (Member, IEEE) is a Lecturer (Assistant Professor) with the School of Electronic Engineering and Computer Science, Queen Mary University of London, U.K. His research interests are in the areas of cognitive communications and computing, ML/AI in wireless communications, and emerging applications of cognitive communications and computing in verticals. With years of academic and research experience, he is adept at bridging research, innovations, and commercializations via digital transformations for a better world. He is a Fellow of Higher Education Academy, U.K.



Lucio Marcenaro (Senior Member, IEEE) is an Associate Professor of Telecommunications with the University of Genoa, Italy. He has 20 years of experience in image and video sequence analysis. He has authored about 130 technical papers on signal and video processing for computer vision. He is an Associate Editor of the IEEE Transactions on Image Processing and IEEE Transactions on Circuits and Systems for Video Technology, the Technical Program Co-Chair of 13th International Conference on Distributed Smart Cameras and the first IEEE International Conference on Autonomous Systems 2021, the Co-Organizer of the 2019 Summer School on Signal Processing, and the General Chair of the Symposium on Signal Processing for Understanding Crowd Dynamics. He is active within the IEEE Signal Processing Italy Chapter and was the Director of Student Services Committee from 2018 to 2021.

Arumugam Nallanathan (Fellow, IEEE) has been a Professor of Wireless Communications and the Head of the Communication Systems Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London since September 2017. He was with the Department of Informatics, King's College London from December 2007 to August 2017, where he was a Professor of Wireless Communications from April 2013 to August 2017 and a Visiting Professor from September 2017 till August 2020. He was an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore from August 2000 to December 2007. He published more than 700 technical papers in scientific journals and international conferences. His research interests include artificial intelligence for wireless systems, beyond 5G wireless networks, and Internet of Things.

Dr. Nallanathan is a co-recipient of the Best Paper Awards presented at the IEEE International Conference on Communications 2016, IEEE Global Communications Conference 2017, and IEEE Vehicular Technology Conference 2018. He is also a co-recipient of the IEEE Communications Society Leonard G. Abraham Prize in 2022. He has been selected as a Web of Science Highly Cited Researcher in 2016 and from 2022 to 2024. He received the IEEE Communications Society SPCE Outstanding Service Award in 2012 and the IEEE Communications Society RCC Outstanding Service Award in 2014. He was a Senior Editor of IEEE WIRELESS COMMUNICATIONS LETTERS and an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and IEEE SIGNAL PROCESSING LETTERS. He served as the Chair for the Signal Processing and Communication Electronics Technical Committee of IEEE Communications Society and the technical program chair and a member of technical program committees in numerous IEEE conferences. He is an IEEE Distinguished Lecturer.

Carlo Regazzoni (Senior Member, IEEE) is a Full Professor of Cognitive Telecommunications Systems with the Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture (DITEN), University of Genoa, Italy. He has been responsible for several national and EU-funded research projects. He is also the coordinator of international Ph.D. courses on interactive and cognitive environments involving several European universities. He has been a co-editor of four edited books (Kluwer) on intelligent video surveillance. He is an author/co-author of more than 100 articles in international scientific journals and more than 300 papers in peer-reviewed international conference proceedings.

Dr. Regazzoni served as the general chair for several conferences and an associate editor and a guest editor for many international technical journals. He served in many roles within the governance bodies of the IEEE Signal Processing Society (SPS), being the IEEE SPS Vice President Conferences from 2015 to 2017. He has been an Associate Editor of several international journals, including the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE SIGNAL PROCESSING LETTERS. He has been a Guest Editor of special issues in international journals, including the PROCEEDINGS OF THE IEEE and the *IEEE Signal Processing Magazine*.