Communication-Enabled Multi-Agent Decentralised Deep Reinforcement Learning to Optimise Energy-Efficiency in UAV-Assisted Networks

Babatunji Omoniwa, Boris Galkin, and Ivana Dusparic

Abstract—Unmanned Aerial Vehicles (UAVs) are increasingly deployed to provide wireless connectivity to static and mobile ground users in situations of increased network demand or points-of-failure in existing terrestrial cellular infrastructure. However, UAVs are energy-constrained and experience the challenge of interference from nearby UAV cells sharing the same frequency spectrum, thereby impacting the system’s energy efficiency (EE). We aim to address research gaps that focus on optimising the system’s EE using a 2D trajectory optimisation of UAVs serving only static ground users, and neglect the impact of interference from nearby UAV cells. Unlike previous work that assume global spatial knowledge of ground users’ location via a central controller that periodically scans the network perimeter and provides real-time updates to the UAVs for decision making, we focus on a realistic decentralised approach suitable in emergencies. Thus, we apply a decentralised Multi-Agent Reinforcement Learning (MARL) approach that maximizes the system’s EE by jointly optimising each UAV’s 3D trajectory, number of connected static and mobile users, and the energy consumed, while taking into account the impact of interference and the UAVs’ coordination on the system’s EE in a dynamic network environment. To address this, we propose a direct collaborative Communication-Enabled Multi-Agent Decentralised Double Deep Q-Network (CMAD–DDQN) approach. The CMAD–DDQN is a collaborative algorithm that allows UAVs to explicitly share knowledge by communicating with its nearest neighbours based on existing 3GPP guidelines. Our approach is able to maximise the system’s EE without hampering performance gains in the network. Simulation results show that the proposed approach outperforms existing baselines in term of maximising the systems’ EE without degrading coverage performance in the network. The CMAD-DDQN approach outperforms the MAD-DDQN that neglects direct collaboration among UAVs, the MAD-Agent Deep Deterministic Policy Gradient (MADDPG) and random policy approaches that consider a 2D UAV deployment design while neglecting interference from nearby UAV cells by about 15%, 65% and 85%, respectively.

Index Terms—Energy efficiency, deep reinforcement learning, UAV base stations, multi-agent system.

I. INTRODUCTION

It is envisaged that machine-to-machine (M2M) connections will grow 2.4-fold, from 6.1 billion in 2018 to 14.7 billion by 2023 [1]. Unmanned Aerial Vehicles (UAVs) can play a vital role in supporting the Internet of Things (IoT) networks by providing ubiquitous connectivity to static and mobile ground devices [2]. For instance, the deployment of UAVs to provide wireless connectivity to ground users is gaining significant research attention [3]–[12]. UAVs’ deployment can complement cellular networks by accommodating for the projected growth of connected things. Specifically, UAVs’ adjustable altitude and mobility make them suitable candidates for flexible deployment as aerial base stations in the event of increased network demand, points-of-failure in existing terrestrial infrastructure, or emergencies [5]. However, it is challenging to conserve the energy of UAVs during prolonged coverage tasks, considering their limited on-board battery capacity. UAVs may deplete energy during propulsion for flying and hovering, and during communication [6]. To derive the full benefit of UAV deployments, recent research have focused on addressing some main challenges, they include: the 3D trajectory optimisation [5], [8], [11], [14], energy efficiency (EE) optimisation [5]–[4], [15], and coverage optimisation [3], [4]. As energy-constrained UAVs fly in the sky, they may encounter interference from nearby UAV cells or other access points sharing the same frequency band, thereby affecting the system’s EE [12]. Several research contributions has been made to optimise the EE of UAVs deployed to serve ground users, however, many of such works neglect the impact of interference on the system’s performance.

Compared with a terrestrial cellular communication network, channel modeling for an airborne, UAV-assisted wireless system is more challenging due the mobility and direct line-of-sight (LoS) communication link from nearby UAVs [16]. Crucially, UAVs require robust strategies to provide ubiquitous wireless coverage to static and mobile ground users in this dynamic environment. Unlike previous work that assume global spatial knowledge of ground users’ location through a central controller that periodically scans the network perimeter and provides real-time updates to the UAVs for decision making, we focus on a decentralised approach suitable in emergency scenarios where there may be service outage due to failure in the controller, or loss of UAVs’ control packets due traffic congestion in the network. Moreover, in such scenarios it is difficult to keep track of the location of all ground users in real-time. To simplify the model, recent approaches that optimise the system’s EE consider a 2D trajectory optimisation design of UAVs serving static users in an interference-free network environment. However, these approaches are impractical in realistic deployment of UAVs serving as aerial base stations.
Nevertheless, UAVs require robust strategies to optimise their flight trajectory while providing coverage to ground users in a dynamic environment. Multi-Agent Reinforcement Learning (MARL) has been shown to perform well in decision-making tasks in such a dynamic environment [3, 4, 11, 12]. To improve the performance of the decentralised control, several methods have been studied [26, 28]. In this work, we adopt a MARL approach and propose a direct collaborative Communication-enabled Multi-Agent Decentralised Double Deep Q-Network (CMAD–DDQN) algorithm to maximise the system’s EE by optimising the 3D trajectory of each UAV, the energy consumed and the number of connected static and mobile ground users over a series of time-steps, while taking into account the impact of interference from nearby UAV cells.

In our previous work [5], we considered a decentralised MARL where there was no direct collaboration among UAVs and other agents are treated as a part of the environment, with the reward of each agent reflecting the coverage performance in its neighbourhood. However, this approach ignores the potential benefit of direct collaboration among agents. Moreover, finding a globally optimal solution for agents with partial information is known to be intractable [27]. As an extension to our prior work [3], we leverage agents’ capability to communicate with neighbours to maximise the system’s EE by jointly optimising the number of connected ground users and the energy consumption in the network. The incorporation of collaborative algorithm into MARL can allow the agents to assist each other in filling the knowledge gaps by exchanging information that could improve the decision making of UAVs over a series of time-steps [29]. However, several real-time applications place considerable restrictions on communication, especially in terms of both throughput and latency. Nevertheless, communication has extensively been used to address non-stationarity issues in multi-agent learning process [25].

Multi-agent learning is challenging in itself, requiring agents to learn their policies while taking into account the consequences of the actions of others. The authors in [4, 8, 9] proposed a multi-agent deep deterministic policy gradient (MADDPG) approach to improve the system’s EE as UAVs hover at fixed altitudes while providing coverage to static ground users in an interference-free network environment. This problem becomes even more challenging in an interference-limited network environment, where the interference from nearby UAV cells impacts the system’s EE. Hence, we propose a direct collaborative Communication-Enabled Multi-Agent Decentralised Double Deep Q-Network (CMAD–DDQN) approach where each agent relies on its local observations, as well as the information it receives from its nearby UAVs for decision making. The communicated information from the nearby UAVs will contain the number of connected ground users, instantaneous energy value, and distances from nearby UAVs in each time-step. We propose an approach where each agent executes actions based on state information. We assume a two-way communication link among neighboring UAVs [30]. Although the 3GPP system provides a methodology to set up and optimise neighbour relations with little or no human intervention [31] and to allow a 3rd party to request and obtain real-time monitored status information (e.g., position, communication link status, power consumption) of a UAV [30], to the best of our knowledge this work is first to investigate the impact of collaborations on the system’s EE using the communication mechanism based on the existing 3GPP standard [31]. This paper has three main contributions given as follows:

- We propose a direct collaborative CMAD–DDQN approach that relies on local observations from each UAV and the explicitly-communicated information from its neighbours for decision-making. We adopt a collaborative algorithm based on an existing 3GPP standard [31] that allows agents to collaborate by exchanging information with their nearest neighbours to improve the system’s EE by jointly optimising each UAV’s 3D trajectory, the number of connected ground users, and the energy consumed by the UAVs in a shared dynamic environment.

- We conduct simulations based on a realistic model of the agent’s environment, taking into consideration the dynamic and interference-limited nature of the wireless environment. We consider a real-world deployment of static and mobile end-users in an area of Dublin, Ireland. We leverage mobility models widely used in literature [32, 35], the random walk (RW), random waypoint (RWP) and the Gauss–Markov (GMM) mobility models.

- We evaluated the proposed CMAD-DDQN approach by comparing it with the MAD-DDQN [3] that ignores direct collaboration among UAVs, the MADDPG [4] that considers a 2D UAV deployment design while neglecting interference from nearby UAV cells, and the random policy. Results show that our proposed approach is able to maximise the total system’s EE while jointly optimising the 3D trajectory, number of connected users, and the energy consumed by the UAVs serving ground users under a strict energy budget.

The remainder of this work is organised as follows. In Section II, we present related work. The environment model is provided in Section III. We discuss the proposed decentralised MARL approach for EE optimisation in Section IV. In Section V, we present the simulation setup and evaluation plan. We discuss and analyse results in Section VI. Section VII concludes the paper and outlines future directions.

II. RELATED WORK

Energy efficiency (EE) optimisation in UAV-assisted networks have been studied recently. The works [15, 17, 23] proposed classical optimisation techniques to optimise the EE of a single UAV deployed to provide wireless service to static ground users. A similar technique was used in a relay scenario [24] to jointly optimise the energy and trajectory of UAVs while transferring information from source ground users to corresponding destination ground users. In [18], an iterative algorithm was proposed to optimise the trajectory of a fixed-wing UAV base station deployed at fixed altitudes while optimising the transmit power in each iteration. However, these models may not be applicable in larger geographical areas where multiple UAVs are deployed to serve ground
users. Recent research focus on optimising EE in multi-UAV networks [5]–[9]. An iterative algorithm was proposed in [8] to minimise the energy consumption of UAV base stations providing coverage to static ground users. Game theory was proposed in [7] to optimise the system’s EE while maximising the ground area covered by the UAVs irrespective of the presence of ground users. However, this work may only be suitable in scenarios with unlimited energy budget or cost of UAVs deployment. Furthermore, these works rely on a ground controller that supports the decision making of the UAVs, hence making emergency deployment impractical due to the significant amount of information exchanged between the UAVs and the controller. Moreover, tracking ground user locations at each time-step may be difficult. In [21], a classical optimisation method was used to minimise the energy consumption of static ground users by optimizing the UAVs’ trajectory. As energy-constrained UAVs fly in the sky, they may encounter interference from nearby UAV cells or other access points sharing the same frequency band, thereby affecting the system’s EE [12].

The adoption of machine learning to solve complex multi-UAV deployment problems is gaining research attention [10]. Specifically, multi-agent reinforcement learning (MARL) approaches have been used in several works to optimise the system’s EE. Our prior work applied a distributed Q-learning approach [5] to optimise the energy utilisation of UAVs providing coverage to ground users without taking into account the system’s EE. To solve this problem, a deep reinforcement learning (DRL) approach [12] was proposed in our recent work for intelligent UAV cellular user to base station association, allowing a UAV flying over an urban area to intelligently connect to underlying base stations. In our prior work [13], a DRL-based approach was proposed to optimise the EE of fixed-winged UAVs that move in circular orbits and are typically not able to hover like the rotary-winged UAVs. Moreover, the attention was on UAVs deployed to provide connectivity to static ground users. The distributed DRL work in [14] improved on the centralised approach in [19], where all deployed UAVs are controlled by a single autonomous agent. The authors in [14], [8], [9] proposed a deep deterministic policy gradient (DDPG) approach to optimise the system’s EE as the UAVs hover at fixed altitudes while serving static ground users in an interference-free network environment. Although the approaches [4], [8], [9] improve the coverage performance of UAVs, they focus on the 2D trajectory optimisation of the UAVs serving static ground users.

This paper extends the decentralised MARL approach proposed in [3], where each agent relies on locally sensed information and makes decisions based on implicitly provided neighbour connectivity information reflected in the agent’s reward function [26] (i.e., no communication mechanism was provided on how agents can collaborate to optimise the system’s EE in this dynamic network environment). We observed in our previous work [3] that as the number of deployed UAVs the network increases, a significant drop in the system’s EE was observed. This motivates us to investigate novel collaborative techniques that improve the system’s EE. Hence, we present a collaborative CMAD–DDQN approach, where each agent makes decisions based on its local observation and direct interaction via existing 3GPP guidelines [30], [31] with its nearest neighbours.

III. SYSTEM MODEL

We consider a set of static and mobile ground users ξ located in a given area as in [3]. Each user $i \in \xi$ at time $t$ is located in the coordinate $(x^t_i, y^t_i) \in \mathbb{R}^2$. In this work, we assume connectivity service outages from the existing terrestrial infrastructure due to disasters or increased network load. As such, a set $N$ of quad-rotor UAVs are deployed within the area to provide wireless coverage to the ground users. As an extension to our prior work in [3], we assume that the UAVs are able exchange state information by communicating with near-by neighbours as shown in Figure 1.

A. Wireless channel model

A UAV $j \in N$ providing wireless coverage to ground users at time $t$ is located in the coordinate $(x^t_j, y^t_j, h^t_j) \in \mathbb{R}^3$. Without loss of generality, a guaranteed line-of-sight (LOS) channel condition is assumed, due to the aerial positions of the UAVs in the sky. Each user $i \in \xi$ in time $t$ can be connected to a single UAV $j \in N$ which provides the strongest downlink signal-to-interference-plus-noise-ratio (SINR). An expression for the SINR at time $t$ is given by [5],

$$\gamma^t_{i,j} = \frac{\beta P(d^t_{i,j})^{-\alpha}}{\sum_{z \in \chi_{i,m}} \beta P(d^t_{i,z})^{-\alpha} + \sigma^2},$$

where $\alpha$ and $\beta$ are the path loss exponent and attenuation factor that characterises the wireless channel, respectively. $\sigma^2$ is the power of the additive white Gaussian noise at the receiver, $d^t_{i,j}$ is the distance between the $i$ and $j$ at time $t$, $\chi_{i,m} \subset N$ is the set of interfering UAVs. $z$ is the index of an interfering UAV in the set $\chi_{i,m}$. $P$ is the transmit power of the UAVs. We model the mobility of mobile users using the random walk (RW), random waypoint (RWP) and the Gauss–Markov (GMM) mobility models [34], [35], which allows users to dynamically change their positions. UAVs are expected to optimise their flight trajectory to provide ubiquitous connectivity to users which may be static or mobile. Given a channel bandwidth $B_w$, using Shannon’s equation the receiving data rate of a ground user can be expressed [12],

$$R^t_{i,j} = B_w \log_2(1 + \gamma^t_{i,j}).$$
B. Connectivity model

We consider an interference-limited system where coverage is affected by the SINR. Thus, the connectivity score of a UAV \( j \in N \) at time \( t \) is calculated as [4],

\[
C^t_j = \sum_{i \in \xi} w^t_j(i),
\]

(3)

where \( w^t_j(i) \in [0, 1] \) represents whether user \( i \) is connected to UAV \( j \) at time \( t \). \( w^t_j(i) = 1 \) if \( \gamma^t_{i,j} = \gamma^t_{i,j} > \gamma_{th} \), otherwise \( w^t_j(i) = 0 \), where \( \gamma_{th} \) is the SINR predefined threshold. Likewise \( R^t_{i,j} = 0 \) if user \( i \) is not connected to UAV \( j \). In a variety of network service scenarios, including disasters, it is desirable to have nearly all ground users connected fairly to the available UAVs. As such, we define geographical fairness using the Jain’s fairness index as [4], [5], [8],

\[
f^t_l = \left( \frac{\sum_{j \in N} C^t_j(j)}{N \sum_{j \in N} C^t_j(j)} \right)^2
\]

(4)

C. Energy consumption model

During a flight operation, UAV \( j \in N \) at time \( t \) expends energy \( e^t_j \). A UAVs’ total energy \( \epsilon_C \) is expressed as the sum in propulsion \( \epsilon_P \) and communication \( \epsilon_C \) energies, \( \epsilon_C = \epsilon_P + \epsilon_C \).

Since \( \epsilon_C \) is practically much smaller than \( \epsilon_P \), i.e., \( \epsilon_C \ll \epsilon_P \), we ignore \( \epsilon_C \). A closed-form analytical propulsion power consumption model for a rotary-wing UAV at time \( t \) is given as [5],

\[
P(t) = \kappa_0 \left( 1 + \frac{3V^2}{U_{\text{tip}}^2} \right) + \kappa_i \left( 1 + \frac{V^4}{4u^4_0} + \frac{V^2}{2v^2_0} \right)^{\frac{1}{2}} + \frac{\rho}{2} \nu s A V^3,
\]

(5)

where \( \kappa_0 \) and \( \kappa_i \) are the UAVs’ flight constants (e.g., rotor radius or weight), \( U_{\text{tip}} \) is the rotor blade’s tip speed, \( v_0 \) is the mean hovering velocity, \( \nu \) is the drag ratio, \( s \) is the rotor solidity, \( A \) is the rotor disc area, \( V \) is the UAVs’ speed at time \( t \), and \( \rho \) is the air density. In particular, we take into account the basic operations of the UAV, such as, hovering and acceleration. Therefore, we can derive the average propulsion power over all time-steps as \( \frac{1}{T} \sum_{t=1}^{T} P(t) \), and the total energy consumed by UAV \( j \) at time \( t \) is given as [5],

\[
e^t_j = \delta_t \cdot P(t),
\]

(6)

where \( \delta_t \) is the duration of each time-step. The EE of UAV \( j \) can be expressed as the ratio of the data throughput and the energy consumed in time-step \( t \), expressed as [5],

\[
\eta^t_j = \frac{\sum_{i \in \xi} R^t_{i,j}}{e^t_j}.
\]

(7)

IV. MULTI-AGENT REINFORCEMENT LEARNING APPROACH FOR ENERGY EFFICIENCY OPTIMISATION

In this section, we formulate the problem and propose a CMAD-DDQN algorithm to improve the trajectory of each UAV in a manner that maximises the total system’s EE.

A. Problem Formulation

Our objective is to maximise the total system’s EE by jointly optimising its 3D trajectory, number of connected users, and the UAVs’ energy consumed while deployed to serve ground users under a strict energy budget. Maximisation of the number of connected users \( C^t_j \) will maximise the total amount of data \( \sum_{i \in \xi} R^t_{i,j} \) the UAV \( j \) will deliver in time-step \( t \) which, for a given amount of consumed energy \( e^t_j \), will also maximise the total EE \( \eta_{tot} \). Therefore, we formulate the optimisation problem as [3],

\[
\begin{align}
\max_{\forall j \in N: x^t_j, \ y^t_j, \ h^t_j, \ e^t_j, \ c^t_j} & \eta_{tot} = \frac{\sum_{t=1}^{T} \sum_{j \in N \in \xi} R^t_{i,j}}{\sum_{i=1}^{T} \sum_{j \in N} e^t_j} \\
\text{s.t.} & \quad \gamma^t_{i,j} \geq \gamma_{th}, \quad \forall u^t_j(i) \in [0, 1], \ i, j, t, \\
& \quad e^t_j \leq e_{\text{max}}, \quad \forall j, t, \ \ (8a) \\
& \quad x_{\text{min}} \leq x^t_j \leq x_{\text{max}}, \quad \forall j, t, \ \ (8b) \\
& \quad y_{\text{min}} \leq y^t_j \leq y_{\text{max}}, \quad \forall j, t, \ \ (8c) \\
& \quad h_{\text{min}} \leq h^t_j \leq h_{\text{max}}, \quad \forall j, t, \ \ (8d)
\end{align}
\]

where \( x_{\text{min}}, y_{\text{min}}, h_{\text{min}} \) and \( x_{\text{max}}, y_{\text{max}}, h_{\text{max}} \) are the minimum and maximum 3D coordinates of \( x, y \) and \( h \), respectively, \( e_{\text{max}} \) is the maximum UAV energy level. The possibility of interference is significantly increased as multiple wireless transmitters sharing the same frequency band are in close proximity to one another. In particular, the complexity of the problem (8a) increases as more UAVs are deployed in a shared wireless environment. Hence it is difficult to obtain the optimal cooperative strategies that improves the system’s EE while completing the coverage tasks under dynamic settings. This is often because UAVs may exhibit a selfish behaviour and pursue the goal of improving their individual EE while maximising the coverage and energy utilization, rather than the collective goal of maximising the system’s EE. In such cases, cooperative MARL approaches may be suitable when there is conflict between the individual and collective interests of UAVs. Deep RL has been proven to perform well in decision-making tasks in this kind of dynamic environment [22]. In our prior work [3] where UAVs have no mechanism to communicate, the system’s EE degraded as the number of UAVs in the network increased. Hence, we adopt a cooperative communication-enabled deep MARL approach to solve the system’s EE optimisation problem.

B. Cooperative Communication-Enabled Multi-Agent Decentralised Double Deep Q-Network (CMAD-DDQN)

It is expected that UAVs will be legally required to broadcast their telemetry information for safety reasons, which involves sharing their coordinates, UAV identification, flight plans (or rather velocity and direction, for security and privacy reasons), vehicle type [31]. This communication can be done through standardized 3GPP sidelink communication (enabling device-to-device (D2D) communications without going through the network infrastructure). Hence, we propose a cooperative
CMAD-DDQN approach that relies on a communication mechanism among neighbouring UAVs for improved system performance. In the scenario we consider, each agent’s reward reflects the coverage performance in its neighbourhood. As seen in Figure 2, each UAV is controlled by a Double Deep Q-Network (DDQN) agent that aims to maximise the system’s EE by jointly optimising its 3D trajectory, number of connected ground users, and the energy consumed by the UAVs. We assume that as the agents interact with each other in a shared and dynamic environment, they may observe learning instabilities due to conflicting policies from other agents. Algorithm 1 shows the DDQN for Agent $j$ with direct collaboration with its neighbours. Agent $j$ follows an $\epsilon$-greedy policy by executing an action $a$ in its present state $s$ after which it transits to a new state $s'$ and receives a reward that reflects the coverage performance in its neighbourhood as given in [3]. Furthermore, the DDQN procedure described on line 23–31 optimises the agent’s decisions. We explicitly define the states, actions, and reward as follows [3]:

- **State space**: We consider the three-dimensional (3D) position of each UAV [11], the UAV's connectivity score, the UAV's instantaneous energy level, maximum of 6 closest neighbor distances using a defined communication mechanism, the neighbour connectivity score, and neighbour instantaneous energy consumed at time $t$, expressed as a tuple,

  $$ (x_t, y_t, z_t, x_{max}, h_t) $$

- **Action space**: At each time-step $t \in T$, each UAV executes an action by changing its direction along the 3D coordinates. Unlike a closely related work and evaluation baseline [4], we discretise the agent’s actions following the design from [3, 5] and [11], as follows: $\{+x, 0, 0\}, \{-x, 0, 0\}, \{0, +y, 0\}, \{0, -y, 0\}, \{0, 0, +z\}, \{0, 0, -z\}$ and $\{0, 0, 0\}$. Our rationale to discretise the action space was to ensure that the agents quickly adapt and converge to an optimal policy.

- **Reward**: The goal of the agent is to learn a policy that maximises the system’s EE by jointly maximising the ground users connectivity while minimising the total UAVs energy consumption. Hence, we introduce a shared cooperative factor $\bar{\mu}$ to shape the reward formulation of each agent $j$ in each time-step $t \in T$ given as [3],

  $$ R_j^t = \begin{cases} \bar{\mu} + \omega + 1, & \text{if } C_j^t > C_j^{t-1} \\ \bar{\mu} + \omega, & \text{if } C_j^t = C_j^{t-1} \\ \bar{\mu} + \omega - 1, & \text{otherwise} \end{cases} $$

  (9)

where $C_j^t$ and $C_j^{t-1}$ are the connectivity score in present and previous time-step, respectively. $\omega = \frac{e_j^t - e_j^{t-1}}{e_j^t + e_j^{t-1}}$, where $e_j^t$ and $e_j^{t-1}$ are the instantaneous energy consumed by agent $j$ in present and previous time-step, respectively. To enhance cooperation, we assign each agent a ‘+1’ incentive from its neighbourhood via a function $\bar{\mu}$ only when the overall connectivity score, which is the total number of connected users by UAVs in its locality in the present time-step $C_i^t$ exceeds that in the previous time-step $C_i^{t-1}$, otherwise the


**Algorithm 1 Double Deep Q-Network (DDQN) for Agent j with Direct Collaboration with its Neighbours**

1. Input: UAV3DPosition, ConnectivityScore, InstantaneousEnergyConsumed, UAVNeighboirDistances, NeighboursConnectivityScore, NeighboursInstantaneousEnergyConsumed ∈ S and Output: Q-values corresponding to each possible action (+x, 0, 0), (−x, 0, 0), (0, +y, 0), (0, −y, 0), (0, 0, +z), (0, 0, −z), (0, 0, 0) ∈ A_{j}.
2. for all a ∈ A_{j} and s ∈ S do:
3. θ – initial network parameters, \(\theta^–\) – copy of \(\theta\), \(N_i\) – maximum size of replay buffer, \(N_p\) – batch size, \(N^-\) – target replacement frequency.
4. \(s \leftarrow \) initial state.
5. 1500 ← maxStep.
6. while goal notReached and Agent alive and maxStep not reached do:
7. \(s \leftarrow\) MapLocalObservationToState(Env).
8. \(a \leftarrow \) DeepQnetwork.SelectAction(s).
9. \(r \leftarrow\) Agent executes action in state s.
10. \(a, s, s', r \leftarrow\) Map sensed observations to new state s'.
11. if \(a, s, s', r\) is inputted in the replay memory then:
12. \(\theta \leftarrow\) TrainQnetwork(s, a, s', r, \(\epsilon\)-greedy).
13. \(\theta^– \leftarrow\) Replace target parameters \(\theta^-\) every \(N^-\) step.
14. endwhile.

\[\mathcal{U} = \begin{cases} +1, & \text{if } C_i^\epsilon > C_{i-1}^\epsilon \\ -1, & \text{otherwise.} \end{cases}\]  

(10)

**C. DDQN Implementation**

The neural network (NN) architecture of Agent j’s DDQN shown in Figure 2 comprises of a 23-dimensional state space input vector, densely connected to 2 layers with 128 and 64 nodes, with each using a rectified linear unit (ReLU) activation function, leading to an output layer with 7 dimensions. Our decentralised approach assume agents to be independent learners while relying on the direct collaboration with nearby UAVs. We follow the analysis presented in [19]. The computational complexity of the NN architecture used in the CMAD-DDQN and MAD-DDQN [3] approaches is approximately \(O(D_{x}K_{W})\) with an average response time of 5.6 ms, while that of a closely related work and evaluation baseline [4] (MADDPG) is approximately \(O(D_{x}K_{W} + O((D_{a} + D_{K})K_{W}))\) with an average response time of 7.4 ms, where \(D_{x}\) is the dimension of the state space, \(D_{a}\) is the dimension of the action space, \(K\) is the number of layers, and \(W\) is the number nodes in each hidden layer.

During the training phase and given the state information as input, Agent j trains the main network to improve its decisions by yielding Q-values that match to each possible action as output. The maximum Q-value obtained is a determinant to the action the agent executes. At each time-step Agent j observes its present state \(s\) and updates its trajectory by selecting an action \(a\) according to its policy. Following its action in time-step \(t\), Agent j observes a reward \(r\) which is defined in [9], and transits to a new state \(s'\) [33]. The information \((s, a, r, s')\) is inputted in the replay memory as shown in Figure 2. Agent j now samples the random mini-batch from the replay memory and uses the mini-batch to get \(y_j\). The optimisation is performed with \(L(\theta)\) and \(\theta\) updated accordingly. The target Q-network updates the parameters \(\theta^-\) with the same parameters \(\theta\) of the main network in every 100th time-step. We set the memory size was set to 10,000 for the training, while using a mini-batch size of 1024. We perform the optimisation using a variant of the stochastic gradient descent called RMSprop to minimise the loss following the methodology described in [20] Chapter 4. The learning rate is set to 0.0001 and the discount factor of 0.95 was applied. Our Q-networks were trained by running multiple episodes, and at each training step the \(\epsilon\)-greedy policy is used to have a balance between exploration and exploitation [20]. In the \(\epsilon\)-greedy policy, the action is randomly selected with \(\epsilon\) probability, whereas the action with the largest action value is selected with a probability of \((1 - \epsilon)\) [33]. The initial value of \(\epsilon\) was set to 1 and linearly decreased to 0.01.

**V. EVALUATION**

**A. Simulation Settings**

The problem formulation described in Section IV is simulated in Python programming language, using the “Pytorch” library. Simulation parameters are presented in Table I. We deploy a varying number of UAVs ranging from 2 to 14 to serve both static and mobile ground users in a 1000×1000 m² area as shown in the environment on Figure 2. We perform 2000 runs of Monte-Carlo (MC) trials over trained episodes.

**B. Baselines**

In this work, we compare the effectiveness of the proposed CMAD-DDQN approach against the following baselines:
The random policy, where UAVs choose their flight directions and travel distances randomly at each time-step \( t \).

The MADDPG \([4]\) approach that considers a 2D trajectory optimisation while neglecting interference from nearby UAV cells. Here, the action space for each agent is continuous.

The MAD-DDQN \([3]\) approach that considers a 3D trajectory optimisation and the interference from nearby UAV cells while neglecting communication among UAVs in a shared network environment.

### C. Metrics

We considered the following metrics for performance evaluation:

- The total energy efficiency \( \eta_{tot} \) which is defined as the ratio of the total throughput and the total energy consumed given as \( \frac{\text{Total Throughput}}{\text{Total Energy}} \) \([3], [13]\).
- The number of connected ground users \([3], [5]\).
- The total energy consumed by UAVs \([3], [5]\).
- The geographic fairness index which reflects the QoS level of ground users served by UAVs from the initial time-step to the current time-step given as \( \frac{\text{Throughput of all users}}{\text{Energy Consumption of all users}} \).

### D. Mobility scenarios

Due to the difficulty in obtaining non-sparse and temporal mobility traces, we adopt 3 mathematical-based mobility models. These models are extensively used in the ad-hoc networks literature to depict realistic mobility patterns of ground devices \([34], [35]\).

1) Random walk mobility model (RW): The RW was developed to mimic the stochastic behaviour of mobile ground devices \([35]\). In RW, the ground devices can change their speed and direction \( \theta(t) \) randomly and uniformly distributed in the range \([0, 2\pi]\) in each time-step \( t \) with zero pause time \([36]\). Each movement occurs either in constant time interval or in a constant travel distance.

2) Random way point mobility model (RWP): The RWP is a more realistic mobility model used ad-hoc networks with an introduction of device pause times between changes in direction and/or speed \([36]\). In this model, the travel distance varies in each time-step \( t \) \([35]\).

3) Gauss–Markov Mobility Model (GMM): The GMM was designed to adapt to different levels of randomness via one tuning parameter. Initially each ground device is assigned a current speed and direction. At each time-step, movement occurs by updating the speed and direction of each device \([35]\).

### VI. RESULTS

In this section, we present experimental results based on the settings earlier discussed. We consider the presence of both static and mobile ground devices deployed within the geographical area to depict a realistic dynamic settings and capture the distribution of ground users. Unless stated otherwise, we leveraged location data in the Drumcondra South A area of Dublin with coordinates around 53° 22' 9" N, 6° 14' 45" W \([37]\) along with synthetic data. For mobility of devices, we adopt different mobility scenarios that depicts realistic mobility patterns since there was difficulty in obtaining non-sparse and temporal mobility traces around the Dublin area.

A. Evaluating the impact of the number of deployed UAVs on the evaluation metrics while comparing the proposed approach with state-of-the-art approaches

To observe how the proposed CMAD-DDQN approach performs while deploying varying numbers of UAVs in Figure \([3]\) we compare the proposed CMAD-DDQN approach with baselines to evaluate the impact of different number of deployed UAVs on the EE, ground users connectivity and total energy consumption. Since we focus on comparing the EE values rather than showing their absolute values, we normalise the EE values with respect to the mean values of the proposed CMAD-DDQN approach. Figure \([3a]\) shows the plot of the normalised EE versus the number of deployed UAVs serving ground users. From Figure \([3a]\) we observe that as more UAVs are deployed, the EE decreases in all approaches possibly because the system becomes more unstable with more UAVs, decreasing the throughput as interference increases, and also takes longer to converge. However, the CMAD-DDQN approach outperforms the MAD-DDQN, MADDPG, and random policy approaches by approximately 15%, 65% and 85%, respectively. The proposed CMAD-DDQN approach on the other hand begins to outperform the MAD-DDQN approach only after the deployment of 8 UAVs. However, the communication overhead on the CMAD-DDQN and MAD- DDQN is given as \( 6m + 6q \) and \( 6q \), respectively, where \( m \) is the unit state-based message \(< e_i^t, e_j^t, x_j, y_j, z_j >\) and \( q \) is the unit reward-based message \(< e_i^t, e_j^t >\). From Figure \([3]\) the communication overhead in the CMAD-DDQN approach results in a slight performance improvement in the evaluation metrics as the number of deployed UAVs are increased.

Figure \([3b]\) shows the plot of the number of connected users versus the number of deployed UAVs while comparing our proposed CMAD-DDQN approach with the baselines. From Figure \([3b]\) we observe a marginally better performance by the MADDPG approach over the CMAD-DDQN and MAD- DDQN approaches in maximising the number of connected ground users by about 0.5% and 2%, respectively. However, the slight performance gain by the MADDPG comes at a huge computational training cost which is 8 times higher than the CMAD-DDQN and MAD-DDQN approaches. On the other hand, the random policy performed worst among the approaches in reducing connection outages, emphasizing the relevance of strategic decision making in MARL problems. Figure \([3c]\) illustrates the plot of the total energy consumed versus the number of deployed UAVs serving ground users, and clearly shows that the MAD-DDQN and CMAD-DDQN approaches significantly minimises the total energy consumed by all UAVs as compared to the other baselines. Although the MADDPG approach performs better in terms of improving the number of connected users than the random policy, the approach trades energy consumption for improved coverage of ground users. Figure \([3d]\) shows the plot of the geographical
Figure 3. Impact of number of deployed UAVs on the UAVs’ EE, number of connected ground users, fairness, and total energy consumption under dynamic network conditions with 400 ground users deployed in a 1 km$^2$ area. Results are averaged over 2000 runs of MC trials.

fairness versus the number of deployed UAVs serving ground users. The graph show that the fairness improves as the number of UAVs are increased over all approaches. We observed better performance in the fairness index for the MADDPG approach than the CMAD-DDQN and MAD-DDQN approaches when 10 or lesser number of UAVs are deployed. As 12 and 14 UAVs are deployed, the proposed CMAD-DDQN approach outperforms the other baselines in terms of fairness.

B. Evaluating the impact of the various mobility models on the evaluation metrics while comparing the proposed approach with state-of-the-art approaches

In a dynamic scenario where users may be mobile, the UAVs’ locations need to be adjusted in such a way as to improve system performance. In Figure 4, we compare the proposed CMAD-DDQN approach with baselines to evaluate the impact of the various mobility models on the EE, number of connected ground users, the geographical fairness and total energy consumption when 8 UAVs are deployed to serve ground users in a 1 km$^2$ area. Figure 4a shows the plot of the normalised EE versus the mobility models. The ground users mobility models considered are the Static, Gauss Markov Mobility (GMM), Random Walk (RW), and the Random Way-Point (RWP) models. Over all deployment of ground users using these mobility models, the proposed CMAD-DDQN approach outperformed the MAD-DDQN, MADDPG and Random Policy approaches in term of maximising the system’s EE by about 15%, 75% and 85%, respectively. Figure 4b show how the various mobility models impact the number of connected users while comparing the proposed CMAD-DDQN approach with the baselines. In all mobility models considered, the MADDQN approach performed closely to the proposed CMAD-DDQN approach. However, the MADDQN approach experience very good coverage performance, it had a larger variance than the CMAD-DDQN approach. Our proposed CMAD-DDQN approach converged to a significantly better average over multiple experimental runs.

Figure 4c show the plot of the total energy consumed
versus the mobility models while comparing the performance of our proposed CMAD-DDQN with the baselines. The random policy consumed the most amount of energy over all mobility models examined. The CMAD-DDQN approach consumes lesser amount of energy in the static scenario than in the GMM, RW and RWP by about 25%, 20% and 15%, respectively. Although, the MADDPG approach performed well in improving the number of connections, it performed poorly in minimizing the total energy consumed. Figure 4(c) shows the plot of the geographical fairness versus the mobility models. The CMAD-DDQN approach performed better than the MAD-DDQN approach and random policy but worse than the MADDPG approach. We observed that all approaches performed slightly better in the static scenario, which implies that decision-making in the dynamic scenario is worse over all approaches.

C. Evaluating the impact of the number of deployed UAVs on the evaluation metrics while comparing the mobility model

To find the optimal number of deployed UAVs in a given coverage area in Figure 5, we compare the proposed CMAD-DDQN approach with baselines to evaluate the impact of the number of deployed UAVs on the system’s EE, ground users outage and total energy consumption while varying the mobility models. Figure 5(b) shows the plot of the number of connected users versus the number of deployed UAVs while varying the mobility models. We observe improve connections as the number of UAVs are increase. Intuitively, more UAVs access points are able to cover more ground users irrespective of the mobility scenario. However, we observe that the static scenario presents us with more number of connected ground users. We see the plot of the total energy consumed versus the number of deployed UAVs in Figure 5(c) while Figure 5(d) shows the plot of the geographical fairness versus the number of deployed UAVs. As more UAVs are deployed to improve
coverage, we observe increased geographical fairness and energy consumption.

In the case of 4 and 8 UAVs deployment as seen in Figure 5c, we observe that the RWP model consumes slightly more energy than other models, while the static scenario consumes lesser energy. The fairness index in all mobility scenarios even up when 12 UAVs are deployed to serve ground users. Figure 5a shows the plot of the normalised EE versus the number of deployed UAVs. We observe that as more UAVs are deployed, the system’s EE drops across all mobility models. The intuition behind that is the increased interference from neighbouring UAVs, that is, as more UAVs are deployed we see an increase in the energy consumed with little increase in the system’s throughput.

VII. CONCLUSION

In this work, we propose a cooperative Communication-Enabled Multi-Agent Decentralised Double Deep Q-Network (CMAD-DDQN) approach to optimise the energy efficiency (EE) of a fleet of UAVs serving static and mobile ground users in an interference-limited environment. As we increase the number of UAVs in the network, the system’s EE in the CMAD-DDQN approach outperforms existing baselines without degrading the coverage performance, energy utilisation, as well as fairness. The CMAD-DDQN approach guarantees quick adaptability and convergence in a shared and dynamic network environment. The CMAD-DDQN steadily converges faster than the MADDPG approach, hereby leading to better EE. We examine the robustness of the cooperative CMAD-DDQN approach over state-of-the-art approaches while varying various mobility models and observe consistent improvement in the system’s EE with minimally deployed number of UAVs. We demonstrate that the CMAD-DDQN approach significantly outperforms the random policy and state-of-the-art decentralised MARL solutions in terms of EE without degrading coverage performance in the network. Although the periodic exchange of information among agents can dramatically increase the entire system’s communication
overheads, it provides performance guarantee for convergence in most multi-agent systems as this one. Our future work will investigate if other cooperative methods can offer faster convergence under these dynamic scenarios. We also aim to investigate the impact of heterogeneous mobility models on the systems’ EE.

REFERENCES

[1] A. Kaloyxlos, A. Gavras, M. D. Camps Mur, M. Ghorashi and H. Hranica, AI and ML – Enablers for Beyond 5G Networks. Zenodo, Dec. 2020.
[2] B. Omoniwa, R. Hussain, M. A. Javed, S. H. Bouk and S. A. Malik, “Fog/Edge Computing-based IoT (FECIoT): Architecture, Applications, and Research Issues,” IEEE Internet of Things Journal, vol. 6, no. 3, pp. 4118-4149, June 2019.
[3] B. Omoniwa, B. Galkin and I. Dusparic, “Optimizing Energy Efficiency in UAV-Assisted Networks using Deep Reinforcement Learning,” IEEE Wireless Communications Letters, doi: 10.1109/LWC.2022.3167568.
[4] C. H. Liu, X. Ma, X. Gao and J. Tang, “Distributed Energy-Efficient Multi-UV Navigation for Long-Term Communication Coverage by Deep Reinforcement Learning,” IEEE Transactions on Mobile Computing, vol. 19, no. 6, pp. 1274-1285, June 2020.
[5] B. Omoniwa, B. Galkin and I. Dusparic, “Energy-Aware Optimization of UAV Base Stations Placement via Decentralized Multi-Agent Q-Learning,” 2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC), Jan. 2022, pp. 216-222.
[6] M. Mozaffari, W. Saad, M. Bennis and M. Debbah, “Mobile Unmanned Aerial Vehicles (UAVs) for Energy-Efficient Internet of Things Communications,” IEEE Transactions on Wireless Communications, vol. 16, no. 11, pp. 7574-7589, Nov. 2017.
[7] L. Ruan et al., “Energy-Efficient Multi-UAV Coverage Deployment in UAV Networks: A Game-Theoretic Framework,” China Communications, vol. 15, no. 10, pp. 194-209, Oct. 2018.
[8] L. Wang, K. Wang, C. Pan, W. Xu, N. Aslam and L. Hanzo, “Multi-Agent Deep Reinforcement Learning-Based Trajectory Planning for Multi-UAV Assisted Mobile Edge Computing,” IEEE Transactions on Cognitive Communications and Networking, vol. 7, no. 1, pp. 73-84, March 2021.
[9] C. H. Liu, Z. Chen, J. Tang, J. Xu and C. Piao, “Energy-Efficient UAV Control for Effective and Fair Communication Coverage: A Deep Reinforcement Learning Approach,” IEEE Journal on Selected Areas in Communications, vol. 36, no. 9, pp. 2059-2070, Sept. 2018.
[10] J. Cui, Y. Liu and A. Nallanathan, “Multi-Agent Reinforcement Learning-Based Resource Allocation for UAV Networks,” IEEE Transactions on Wireless Communications, vol. 19, no. 2, pp. 729-743, Feb. 2020.
[11] X. Liu, Y. Liu and Y. Chen, “Reinforcement Learning in Multiple-UAV Networks: Deployment and Movement Design,” IEEE Transactions on Vehicular Technology, vol. 68, no. 8, pp. 8036-8049, Aug. 2019.
[12] B. Galkin, E. Fonseca, R. Amer, L. A. DaSilva and I. Dusparic, “REQIBA: Regression and Deep Q-Learning for Intelligent UAV Cellular User to Base Station Association,” IEEE Transactions on Vehicular Technology, vol. 71, no. 1, pp. 5-20, Jan. 2022.
[13] B. Galkin, B. Omoniwa, and I. Dusparic, “Multi-Agent Deep reinforcement Learning For Optimising Energy Efficiency of Fixed-Wing UAV Cellular Access Points,” ICC 2022 - IEEE International Conference on Communications, arXiv:2211.02238.
[14] J. Lyu, Y. Zeng, R. Zhang and T. J. Lim, “Placement Optimization of UAV-Mounted Mobile Base Stations,” IEEE Communications Letters, vol. 21, no. 3, pp. 604-607, Mar. 2017.
[15] Y. Zeng, J. Xu and R. Zhang, “Energy Minimization for Wireless Communication With Rotary-Wing UAV,” IEEE Transactions on Wireless Communications, vol. 18, no. 4, pp. 2329-2345, April 2019.
[16] Q. Zhang, A. Ferdowsi, W. Saad and M. Bennis, “Conditional Generative Adversarial Networks (GANs) for Data-Driven Millimeter Wave Communications in UAV Networks,” IEEE Transactions on Wireless Communications.
[17] J. Yao and N. Ansari, “QoS-Aware Power Control in Internet of Drones for Data Collection Service,” IEEE Transactions on Vehicular Technology, vol. 69, no. 7, pp. 6649-6656, July 2019.
[18] X. Jing, J. Sun and C. Masouros, “Energy Aware Trajectory Optimization for Aerial Base Stations,” IEEE Transactions on Communications, vol. 69, no. 5, pp. 3352-3366, May 2021.

[19] J. Hribar, A. Marinescu, A. Chiumento and L. A. DaSilva, “Energy Aware Deep Reinforcement Learning Scheduling for Sensors Correlated in Time and Space,” IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3114102.
[20] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, “An Introduction to Deep Reinforcement Learning,” Foundations and Trends in Machine Learning, vol. 11, no. 3-4, 2018.
[21] Y. Luo, W. Ding and B. Zhang, “Optimization of Task Scheduling and Dynamic Service Strategy for Multi-UAV-Enabled Mobile-Edge Computing System,” IEEE Transactions on Cognitive Communications and Networking, vol. 7, no. 3, pp. 970-984, Sept. 2021.
[22] M. Zhang, S. Fu and Q. Fan, “Joint 3D Deployment and Power Allocation for UAV-BSs: A Deep Reinforcement Learning Approach,” IEEE Wireless Commun. Lett., vol. 10, no. 10, pp. 2309-2312, Oct. 2021.
[23] M. Hua, Y. Wang, C. Li, Y. Huang and L. Yang, “Energy-Efficient Optimization for UAV-Aided Cellular Offloading,” IEEE Wireless Communications Letters, vol. 8, no. 3, pp. 769-772, June 2019.
[24] Z. Sun, D. Yang, L. Xiao, L. Cuthbert, F. Wu and Y. Zha, “Joint Energy and Trajectory Optimization for UAV-Enabled Relaying Network With Multi-Pair Users,” IEEE Transactions on Cognitive Communications and Networking, vol. 7, no. 3, pp. 939-954, Sept. 2021.
[25] D. Szer and F. Charpillet, “Improving coordination with communication in multi-agent reinforcement learning,” 16th IEEE International Conference on Tools with Artificial Intelligence, 2004, pp. 436-440.
[26] T. Ming, “Multi-Agent Reinforcement Learning: Independent versus Cooperative Agents,” Proceedings of the Tenth International Conference on Machine Learning (ICML 1993), San Francisco, CA, USA, pp. 330–337.
[27] D. Yagan and C. Tham, “Coordinated Reinforcement Learning for Decentralized Optimal Control,” 2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning, pp. 296-302.
[28] W. Kim, M. Cho and Y. Sung, “Message-Dropout: An Efficient Training Method for Multi-Agent Deep Reinforcement Learning,” AAAI’19/EAAI’19: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, Jan. 2019, pp. 6079–6086.
[29] K. Zhang, Z. Yang, T. Başar, “Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms,” In: Vamvoudakis K.G., Wan Y., Lewis F.L., Cansever D. (eds) Handbook of Reinforcement Learning and Control. Studies in Systems, Decision and Control, vol 325. Springer, Cham, 2021.
[30] https://portal.3gpp.org/desktopmodules/Specifications/Specificat1onDetails.aspx?specificationId=3557.
[31] https://www.3gpp.org/ftp/Specs/archive/32_series/32.511/
[32] E. Vinogradov, F. Minucci and S. Pollin, “Wireless Communication for Safe UAVs: From Long-Range Deconfliction to Short-Range Collision Avoidance,” IEEE Vehicular Technology Magazine, vol. 15, no. 2, pp. 895-905, June 2020.
[33] R. S. Sutton and A. G. Barto, Introduction to Reinforcement Learning, 1st ed. Cambridge, MA, USA: MIT Press, 1998.
[34] C. Song, T. Koren, P. Wang, and A. L. Barabasi, “Modelling the scaling properties of human mobility,” Nature Physics, vol. 6, no. 10, pp. 818–823, 2010.
[35] Camp, T et al (2002) A Selective Overview of Theories and Algorithms,” In: Vamvoudakis K.G., Wan Y., Lewis F.L., Cansever D. (eds)
Babatunji Omoniwa (Graduate Student Member, IEEE) received the bachelor’s degree in electrical engineering from Ahmadu Bello University, Zaria, Nigeria, in 2010, and the master’s degree in computer engineering from the COMSATS University Islamabad, Pakistan, in 2015. He is currently pursuing a Ph.D. degree in Artificial Intelligence for Smart Cities from the School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland. He was a Research Fellow with the National Mathematical Centre, Abuja, Nigeria. His research interests include Internet-of-Things, and the application of machine learning to wireless networks.

Boris Galkin (Member, IEEE) received the B.A.I. and M.A.I. degrees in 2014 in computer & electronic engineering and the Ph.D. degree in 2019 from Trinity College Dublin, Dublin, Ireland. He is currently a Research Fellow with CONNECT, Trinity College Dublin, Dublin, Ireland. In 2021, he was appointed a Science Foundation of Ireland Public Service Fellow. His research interests include unmanned aerial vehicles, heterogeneous networks, and the application of AI and machine learning to 5G and Beyond networks.

Ivana Dusparic received the B.Sc. degree from La Roche College, Pittsburgh, PA, USA, in 2003, and the M.Sc. and Ph.D. degrees from Trinity College Dublin, Dublin, Ireland, in 2005 and 2010, respectively. She is an Associate Professor with the School of Computer Science and Statistics, Trinity College Dublin. She was elected Fellow of Trinity College Dublin in 2022 and serves as a Co-Director of the Science Foundation of Ireland (CRTAI project). She is a funded investigator with CONNECT, the Telecommunications Research Centre funded by Science Foundation Ireland. Her expertise and research interests include the use of Artificial Intelligence techniques (reinforcement learning, multi-agent systems) to achieve autonomous optimization of large-scale heterogeneous infrastructures, applied to smart cities and sustainable urban mobility.