# Reinforcement Learning for Improved UAV-based Integrated Access and Backhaul Operation

Nikita Tafintsev†, Dmitri Moltchanov†, Meryem Simsek⋆, Shu-ping Yeh⋆,
Sergey Andreev†, Yevgeni Koucheryavy†, and Mikko Valkama†
†Tampere University, Tampere, Finland
⋆Intel Corporation, Santa Clara, CA, USA.

***Abstract*—There is a strong interest in utilizing commercial cellular networks to support unmanned aerial vehicles (UAVs) to send control commands and communicate heavy traffic. Cellular networks are well suited for offering reliable and secure connections to the UAVs as well as facilitating traffic management systems to enhance safe operation. However, for the full-scale integration of UAVs that perform critical and high-risk tasks, more advanced solutions are required to improve wireless connectivity in mobile networks. In this context, integrated access and backhaul (IAB) is an attractive approach for the UAVs to enhance connectivity and traffic forwarding. In this paper, we study a novel approach to dynamic associations based on reinforcement learning at the edge of the network and compare it to alternative association algorithms. Considering the average data rate, our results indicate that the reinforcement learning methods improve the achievable data rate. The optimal parameters of the introduced algorithm are highly sensitive to the donor next generation node base (DgNB) and UAV IAB node densities, and need to be identified beforehand or estimated via a stateful search. However, its performance nearly converges to that of the ideal scheme with a full knowledge of the data rates in dense deployments of DgNBs.**

## I. Introduction

The unmanned aerial vehicles (UAVs), known as drones, are becoming popular across numerous applications. The recent forecasts indicate that the UAV market will grow tenfold to $40 billion over the coming years [1]. Prominent UAV applications include newsgathering (photography and videography), inspection, delivery, and various industrial applications.

The telecom sector is also interested in leveraging UAVs [2]. They are inherently mobile and thus highly dependent on wireless connectivity to support their communication needs. These include authentication, authorization, control communication for UAVs operation, and payload data transmission of the applications onboard the UAVs (e.g., high bandwidth video streaming). Mobile networks can provide flexible communication means to support various UAV use cases from low latency to high bandwidth scenarios [3]. Further, emerging fifth generation (5G) networks will empower UAV connectivity by introducing ultra-reliable low-latency communications (URLLC). However, specific enhancements should be made to optimize UAV connectivity with 5G and beyond cellular systems to improve performance, agility, and flexibility while maintaining adequate performance for ground users [4].

As interest in UAVs increases, the standardization community and telecom industry are attempting to improve the support for UAVs. There have been standardization works in 3GPP that enhance wireless capabilities of UAVs via cellular systems. In this context, [5] reviews the UAV functionality as well as defines potential new service-level requirements related to the payload and the operation of radio access nodes onboard UAVs.

Another area of interest for the leveraging of UAVs is aerial relays. In such deployments, UAVs employ wireless backhauling that is initially supported by 5G networks and act as intermediate nodes for integrated access and backhaul (IAB) operation. By utilizing UAVs as IAB nodes, named in this work UAV IAB nodes, significant enhancements can be made to the connectivity, coverage, and capacity of wireless networks [6]–[9]. Further, due to their mobility, UAVs operating as IAB nodes can optimize the number of hops and enhance the topology flexibility [10]. The 3GPP finalized the Study Item on IAB [11], where the feasibility of IAB over 5G infrastructure was justified. Additionally, this study identified and evaluated potential solutions for the efficient operation of the target system for the frequencies of up to 100 GHz. IAB is currently being standardized for 3GPP Rel-16 with expected completion by Q1 2020.

The research works on wireless backhauling proliferated recently. The study in [12] discussed millimeter-wave (mmWave) access and backhaul features and capabilities. In particular, the authors analyzed the requirements and implementation challenges of mmWave systems. They concluded that mmWave networks are more effective compared to the current cellular technologies. The authors in [13] compared mmWave-based IAB deployments with existing architectures. They summarized the 3GPP standardization activities on IAB and estimated the performance of IAB networks for various applications and data traffic types. In [14], the authors investigated traffic forwarding strategies in mmWave-capable IAB systems. They compared the performance of various path selection schemes for efficient communications. The performance of UAV access was evaluated analytically in [15], [16] by showing that the usage of mmWave technology allows achieving an extremely high density of supported UAVs.

One of the most critical challenges in performance optimization of aerial networks is that the environment may be varying continuously, i.e., the available association points change over time as UAVs travel over the area of interest. On the other

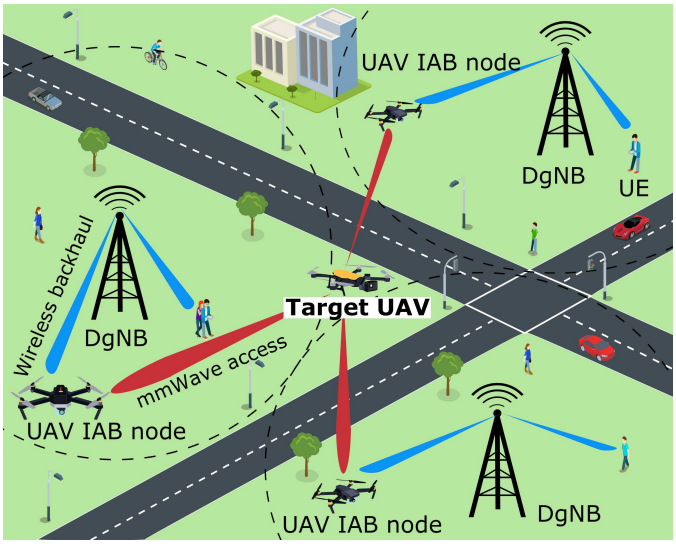



Fig. 1. Usage of UAV IAB nodes for data traffic offloading.

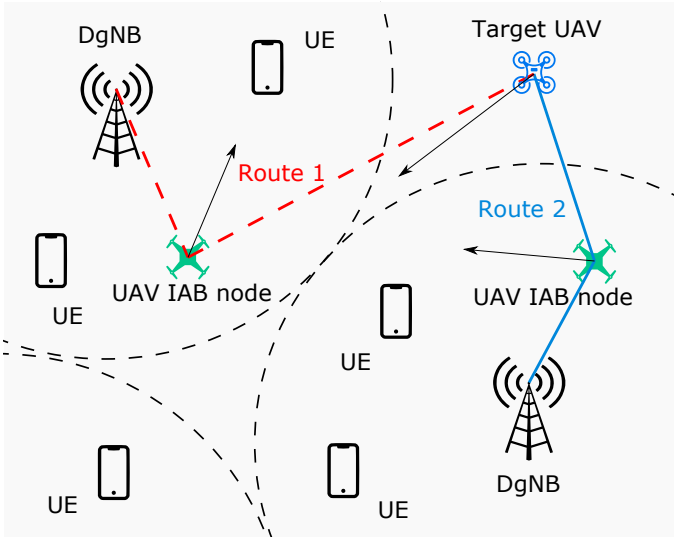



Fig. 2. Illustration of route selection for the target UAV.

hand, most of the modern optimization techniques, including machine learning approaches, have primarily been developed for a stationary "fully explorable" environment, wherein an algorithm may learn sufficiently over a limited time. In this paper, we tailor the reinforcement learning methods to the case of a constantly changing environment. We then proceed to investigate its performance by comparing them with a set of semi- and fully-dynamic approaches. The main contributions of this work are:

- we show that in the noise-limited regime, i.e., where extremely directional antennas are employed at the access and backhaul interfaces, a straightforward scheme based on the Reference Signal Receive Power (RSRP) measurements of the access links provides near-optimal performance;
- we compare the proposed algorithms and demonstrate that the reinforcement learning approach allows improving the performance by approximately $50\%$ across the wide range of donor next generation node base (DgNB), connected to a core network through a wired interface, and UAV IAB node densities;
- we demonstrate that the optimal exploration rate $\epsilon$ in reinforcement learning, which is involved in deciding upon the trade-off between exploiting and exploring, is sensitive to the density of DgNBs and UAV IAB nodes and has to be adjusted online.

The rest of this paper is organized as follows. In Section II, we introduce our system model. Next, we present association schemes and detail our performance evaluation campaign in Section III. Numerical results are reported in Section IV. Conclusions are drawn in the last section.

## II. System Model

In this section, we introduce our system model by specifying components including the deployment, propagation, antenna design, associations, and resource allocation. We also define the metric of interest.

### A. Deployment Considerations

We consider a square geographical area with width $w$. We assume that the area is covered by terrestrial DgNBs that have wired fiber backhaul connectivity with the core network and operate in 28 GHz band over the channel bandwidth of $W$ (Fig. 1). Additionally, the considered scenario encompasses mmWave-capable IAB nodes, i.e., the moving UAVs that may be leveraged as relays for the data transmission. They maintain wireless connections with DgNBs and other UAVs acting as user equipment (UE). We assume that the backhaul links from UAV IAB nodes to DgNBs and the fronthaul links from UAVs to UAV IAB nodes are mmWave.

In our model, we concentrate on one target UAV entering the area of interest. We expect that this UAV is used in a critical high data rate service, where it needs to continuously transmit data traffic and thus be associated with the core network. For example, this may be a surveillance UAV that captures a high-resolution video and streams it over the cellular infrastructure. Therefore, it generates heavy data traffic and requires high throughput for the transmission. For the sake of exposition, we assume that the target UAV enters from a corner of the area and moves in a diagonal direction.

In the considered deployment area, the flying altitude $h_U$ and the speed of the UAVs $v_U$ are constant, and the UAVs travel solely in the horizontal plane. We also require that the UAV altitude exceeds the DgNB height $h_D$. The process of UAVs entering the area is assumed to be homogeneous Poisson with the intensity of $\lambda$. A fraction of UAVs is expected to act as the UAV IAB nodes by advertising this capability over the air interface according to [11].

### B. Propagation and Antenna Modeling

To characterize the mmWave propagation, we utilize the 3GPP urban macro (UMa) model [17]. Accordingly, the path loss measured in dB is given by

$$PL(f_c, d) = 28.0 + 22\log_{10} d + 20\log_{10} f_c, \quad (1)$$

where $d$ is the 3D distance, $f_c$ is the carrier frequency in GHz.

To model antenna array systems, we employ planar uniform rectangular antenna arrays with $N_V \times N_H$ elements [18]. The efficiency of the antenna array depends on the element spacing and their weights, and is described by the Hadamard product as $\widetilde{W} = VW$, where $V$ is the phase shift due to array placement and $W$ is the weighting factor. The phase shift $V$ is

$$V = [v_{1,1}, \ldots, v_{1,N_V}, \ldots, v_{N_H,1}, \ldots, v_{N_H,N_V}]^T, \quad (2)$$

where $v_{m,n}$ are the steering matrix components

$$v_{m,n} = e^{i2\pi\left[(n-1)\ \frac{d_V}{\lambda}\ \cos(\theta)+(m-1)\frac{d_H}{\lambda}\ \sin(\theta)\sin(\phi)\right]}, \quad (3)$$

where $\theta \in (0, \pi)$ is the elevation angle, $\phi \in (-\pi, \pi)$ is the azimuth angle, $d_V$ and $d_H$ are the element spacing along $y$-axis and $z$-axis, respectively, and $\lambda$ is the wavelength.

The weighting factor $W$ is provided by

$$W = [w_{1,1}, \ldots, w_{1,N_V}, \ldots, w_{N_H,1}, \ldots, w_{N_H,N_V}]^T, \quad (4)$$

where $w_{m,n}$ are the weighting components

$$w_{m,n} = \frac{e^{i2\pi\left[\frac{(n-1)d_V}{\lambda}\sin(\theta_{et})-\frac{(m-1)d_H}{\lambda}\cos(\theta_{et})\sin(\phi_{es})\right]}}{\sqrt{N_H N_V}}, \quad (5)$$

where $\theta_{et}$ is the electrical down-tilt steering and $\phi_{es}$ is the electrical horizontal steering. The weighting factor can provide control of side lobe levels as well as perform horizontal and vertical steering. Its phase is dependent on the required steering angles and the element spacing. In our implementation, it is also assumed that the amplitude of the weighting vector is identical for each radiation element.

Overall, the composite array radiation pattern in dB is

$$A_A(\theta, \phi) = A_E(\theta, \phi) + \\ + 10\log_{10}\left[1 + \rho\Big(\Big|\sum_{m=1}^{N_H}\sum_{n=1}^{N_V} w_{m,n} v_{m,n}\Big|^2 - 1\Big)\right], \quad (6)$$

where $A_E(\theta, \phi)$ is a single element radiation pattern, $\rho$ is the correlation coefficient between the patterns of elements.

### C. Traffic, Associations, and Resource Allocation

In our study, we concentrate on a single UAV that travels across the area of interest. We assume the full-buffer traffic model, where the UAV always has data to transmit, i.e, it occupies the entire set of provided resources. Furthermore, along with its flight through the area, UAV is allowed to associate either with UAV IAB nodes or with DgNBs according to the association schemes introduced in Section III.

To reflect realistic deployment conditions, we assume that the fixed amount of bandwidth, $W$, is available to the overall system. This bandwidth is then divided between DgNBs access, IAB backhaul, and IAB access links as follows. We expect that terrestrial UEs receive a higher priority in service and occupy $Y(t)$ share of the resources, where $t$ is the time step. The remaining set of the resources is divided equally between the backhaul links of UAV IAB nodes that are currently associated with DgNB. Here, with each UAV IAB node receives $(W - Y(t))/N(t)$, where $N(t)$ is the number of currently associated UAV IAB nodes. The same set of resources is available at UAV IAB nodes to the target UAV interface, i.e, the resources are used for both backhaul and access links. We require that a UAV IAB node is able to support at most one UAV. Therefore, if the current UAV IAB node is associated with one UAV, the latter receives exactly $(W - Y(t))/N(t)$ share of bandwidth. Note that there might be interference between different entities in the system as DgNBs do not perform explicit coordination in their bandwidth allocations. We explicitly take it into account as discussed in Section III.

To capture the resource utilization at DgNBs by terrestrial UEs, we model the current amount of occupied resources at DgNBs by using the covariance-stationary autoregressive process of order 1, AR(1), $\{Y(t), t = 0, 1, \ldots\}$. Recall that this process is defined by [19] as

$$Y(t) = \phi_0 + \phi_1 Y(t-1) + \epsilon(t), t = 0, 1, \ldots, \quad (7)$$

where $\phi_0$ and $\phi_1$ are constants, $\{\epsilon(t), t = 0, 1, \ldots\}$ are independent and identically distributed (i.i.d.) random variables having zero-mean Normal distribution with variance $\sigma_\epsilon^2$.

The parameters of AR(1) models are given as in [20]

$$\begin{cases} \phi_1 = K_X(1), \\ \phi_0 = \mu_X(1 - \phi_1), \\ \sigma_\epsilon^2 = \sigma_X^2(1 - \phi_1^2), \end{cases} \quad (8)$$

where $K_X(1)$, $\mu_X$, and $\sigma_X^2$ are lag-1 autocorrelation coefficients, mean, and variance of the amount of the occupied resources, respectively.

As one may observe, the proposed DgNBs loading model is straightforward and allows to account for gradual changes in the expected load generated by the terrestrial UEs. It also allows to abstract the amount of resources occupied by the terrestrial UEs at the access mechanism of DgNBs, while simplifying our evaluation methodology.

### D. Metric of Interest

In this paper, we concentrate on the ergodic achievable capacity of the target UAV. Let $R(t)$ be the instantaneous data rate provided to it. The metric of interest is then defined as

$$R = \lim_{t\to\infty} R(t), \quad (9)$$

where the instantaneous data rate, $R(t)$, is calculated as

$$R(t) = B(t)\log(1 + S(t)), \quad (10)$$

where $B(t)$ is a share of bandwidth available for UAV IAB node and $S(t)$ is the signal-to-interference-plus-noise ratio (SINR).

## III. Algorithms and Simulations

In this section, we first propose the association algorithms for the UAV connectivity. Then, we proceed to specify details of the simulation environment and performance assessment procedures.

### A. Association Algorithms

Owning to the inherently dynamic nature of the considered environment, we introduce and assess several association schemes, including semi-dynamic and fully-dynamic ones, having different types of information taken into account.

*1) Ideal Case:* This scheme is a theoretical abstraction of the ideal association rules that maximizes the instantaneous data rate of the UAV at all times by assuming perfect knowledge of the system state. Accordingly, UAV chooses the best association point among associated nodes at any given instant of time. We focus on access link selection, assuming fixed backhaul topology formation. In practice, this scheme can be facilitated by providing the following timely knowledge to the target UAV: (i) channel quality of backhaul and access links, (ii) current amount of resources available at the backhaul and access links, and (iii) bandwidth of DgNBs. Having these parameters, for any given time instant $t$, the target UAV may determine an association point that maximizes its ergodic rate.

*2) RSRP-based Associations:* In this case, the decision on the choice of an association point is taken based on the RSRP measurements of the UAV's access links to either DgNBs or UAV IAB nodes. Note that in this scheme, the UAV does not account for the backhaul links nor for the set of resources available at DgNBs or UAV IAB nodes. This scheme represents a naive approach, where the target UAV attempts to maintain the quality of its access link, by making selections solely based on RSRP.

*3) Classification-based Associations:* This strategy employs multiple steps to perform node selection. The goal is to permit the UAV to select the nodes, which may potentially be better in terms of the instantaneous data rate while having worse RSRP. In this approach, upon entering the area of interest, the target UAV assesses the detected UAV IAB connectivity options and DgNBs to perform their classification. To rank the nodes, the density-based spatial clustering of applications with noise (DBSCAN) algorithm is employed [21]. It is a density-based non-parametric clustering algorithm, which groups together the points that are packed closely and marks as outliers the points that lie alone in low-density regions. At the next step, the UAV chooses a random UAV IAB node or DgNB from the best group, where it is determined by an arithmetic average of the RSRP values of each node of the group. In addition, when a new UAV IAB node/DgNB appears or leaves the operating area of the target UAV, the classification and selection procedure is repeated. Therefore, as one may observe, the algorithm is semi-dynamic in nature, and reevaluates the connectivity options solely at the connectivity changing time steps.

*4) $\epsilon$-greedy:* This approach introduces a dynamic reinforcement learning based method. Reinforcement learning is a computational approach and one of the three basic machine learning principles, alongside with supervised learning and unsupervised learning. It is different from other methods by its emphasis on learning by an agent from direct interactions with its surrounding environment. In particular, this approach does not require supervision or detailed models of the environment. In this work, we employ the multi-armed bandit (MAB) learning scheme [22]. It is named by its similarity to a slot machine, or "one-armed bandit," except that it has several levers instead of one. The goal of the method is to maximize the aggregated reward via repeated learning iterations.

The main challenge of the problem at hand is to strike a balance between sufficient exploration of the environment and exploitation of the optimal action. In this scheme, to find an association point, we leverage a conventional approach to balance the exploitation-exploration trade-off, named $\epsilon$-greedy algorithm. Particularly, the best node is selected for the proportion $1-\epsilon$ of the trials, and a node is selected at random for the proportion $\epsilon$. The estimated value of action $a$ at time step $t$ is calculated according to the experience by averaging the corresponding rewards $R_t$ associated with the target action

$$Q_t(a) = \frac{\sum_{i=1}^{t-1} R_i \cdot \mathbb{1}_{A_i=a}}{\sum_{i=1}^{t-1} \mathbb{1}_{A_i=a}}, \tag{11}$$

where $\mathbb{1}$ is a binary indicator function.

*5) Upper Confidence Bound (UCB):* The $\epsilon$-greedy solution indiscriminately selects actions without preference among those that are nearly greedy or particularly uncertain. To avoid such inefficient exploration, we follow the UCB approach, where the target UAV favors exploration of actions with a strong potential to reach the optimal value. Actions are selected by taking into account both how close their estimates are to the maximum and the uncertainties in those estimates as provided by

$$A_t = \arg\max_a \left[ Q_t(a) + c\sqrt{\frac{\ln t}{N_t(a)}} \right], \tag{12}$$

where $c > 0$ controls the degree of exploration and $N_t(a)$ denotes the number of times that action $a$ has been selected prior to time $t$.

### B. Simulation Environment

*1) Implementation:* We implemented the proposed system using our dedicated simulation environment that captures the details of the system model specified in Section II. The simulation engine is based on the discrete-event method [23] and is written in Python with multi-threaded implementation for parallel execution.

We capture the effects of interference and time-dynamics as the two most critical aspects of the implemented modeler. Since the system operates with the shared bandwidth for DgNBs and UAV IAB nodes, the simulation environment

explicitly recreates interference between these interfaces with the help of the radiation patterns detailed in Section II by accounting for the orientation of active transmissions. When there are $M$ links, each using $k_i$, $i = 1, 2, \ldots, M$, fraction of radio resources that is currently assigned to the target UAV, the overall interference is computed as

$$I = \sum_{i=1}^{M} \frac{k_i P_{T_i} G_{T,i}(\alpha, \beta) G_R(\alpha, \beta)}{10^{2 \log_{10} f_c + 2.2 \log_{10} d + 2.8}}, \quad (13)$$

where $P_{T_i}$ is the transmit power, $G_{T,i}(\alpha, \beta)$ is the transmit gain of interferer $i$ towards the target UAV, $G_R(\alpha, \beta)$ is the receive gain of the target UAV towards interferer $i$, and $k_i$ is the bandwidth sharing coefficient. The denominator is the path loss value in linear scale as specified in (1).

The antenna alignment between the target UAV and the currently serving DgNB or UAV IAB node is assumed to be ideal owing to the periodic beamforming procedure. Further, the SINR and Shannon capacity are computed to obtain the instantaneous data rate. In the simulation model, the time discretization step is set independently of the UAV speed and is taken as 1 ms that corresponds to the standard New Radio (NR) frame duration. The initial share of radio resources utilized by the terrestrial UEs at DgNBs is set to be randomly and uniformly distributed in $(0, 1)$.

*2) Data Collection and Analysis:* To apply standard statistical methods and ensure that the obtained samples contain i.i.d. observations, we employed a combination of sampling and data aggregation techniques. Particularly, to gather the simulation data, we sampled the state of the system every 100 seconds of simulation time. Further, we applied the batch means data aggregation technique to remove the remaining residual correlation in the data [24]. The results were tested using the Ljung-Box portmanteau test for residual correlations [25]. As a result, the data that passed the test was used for calculating the metric of interest.

To demonstrate accurate point estimates of the sought metrics, we employed our simulation procedure with a given accuracy. Accordingly, we did not initially limit the sample size thus performing in-simulation assessment of accuracy by using the interval estimates [26]. The target accuracy was set to $x \pm 0.05x$, where $x$ is the point estimate. To achieve this type of results in a limited deployment size, we considered a long simulation run that repeats the passing of the target UAV through the area on interest. This procedure leads to the point estimates in the next section.

## IV. Numerical Results

In this section, we evaluate the performance of the discussed schemes. We start by highlighting the importance of interference for the considered scenario and then proceed with evaluating the considered association algorithms. The default system parameters are provided in Table I.

### A. Effect of Interference

We begin our analysis by assessing the effect of interference on the performance of the considered system. Fig. 3 illustrates

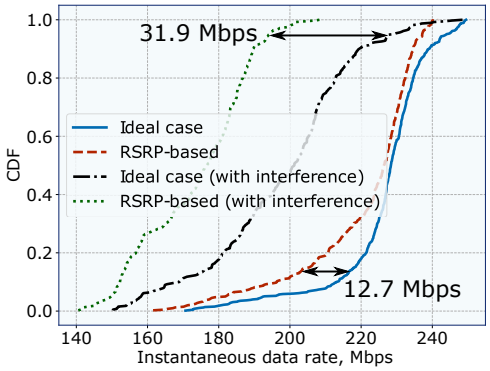


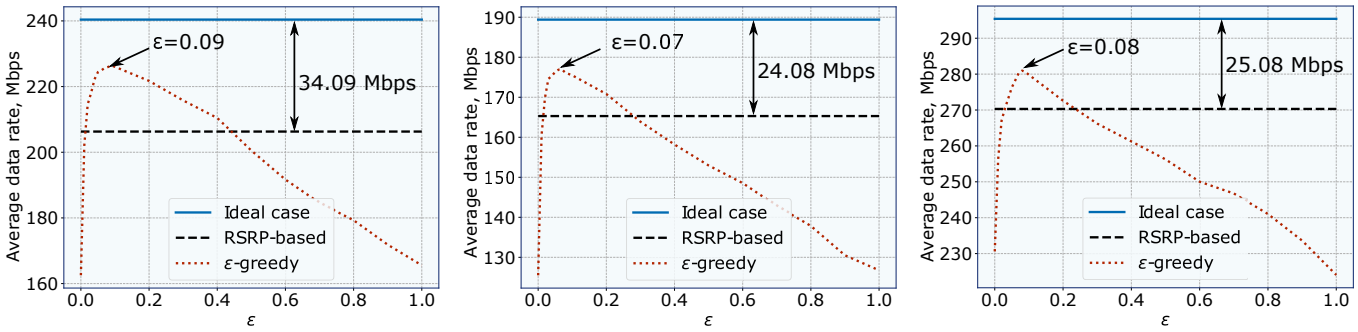

Fig. 3. CDF of instantaneous data rate of the target UAV.

the ergodic data rates for the ideal and RSRP-based schemes under the density of DgNBs $\zeta = 0.0005$ node/m$^2$, density of UAV IAB nodes $\lambda = 0.001$ node/m$^2$, UAV speed of $v = 10$ m/s, mean DgNB load by terrestrial users $\mu_Y = 0.5$. First, even in the presence of directional antennas at all the radio interfaces, the effect of interference is not negligible as the mean data rate for the ideal scheme decreases from 230 down to 197 Mbps. Another key observation is that the RSRP-based scheme displays results that are fairly close to those of the ideal scheme as the maximum deviation between the observed data rates is around 7 Mbps. These two conclusions hold across a wide range of system parameters. Recalling the results of [27], we may conclude that for extremely directional antennas, the naive RSRP-based scheme may prove to be sufficient for the data rate optimization at the UAV access interface.

Further, analyzing the results in Fig. 3, one may observe that for today's realistic antenna arrays, there is a large gap totalling up to 32 Mbps that can be explored by using more comprehensive algorithms. For different values of system parameters, it varies from 15% to approximately 30% of the data rate achieved with the ideal scheme. We now proceed by assessing whether the considered algorithms allow bridging this gap by exploiting the knowledge of dynamically changing instantaneous data rates with UAV IAB nodes.

TABLE I
Default System Parameters.

| Parameter | Value | Description |
|---|---|---|
| $w$ | 200 m | Deployment area |
| $W$ | 400 MHz | Bandwidth of mmWave DgNB |
| $f_c$ | 28 GHz | mmWave carrier frequency |
| $N_V \times N_H$ | $4 \times 4$ | Antenna array size |
| $\rho$ | 1 | Correlation coefficient |
| $h_U$ | 30 m | UAV altitude |
| $v_U$ | 10 m/s | UAV velocity |
| $h_D$ | 20 m | DgNB height |
| $\zeta$ | 0.0005 node/m$^2$ | Density of DgNBs |
| $\lambda$ | 0.001 node/m$^2$ | Density of UAV IAB nodes |
| $P_{TX}$ | 24 dBm | Transmit power |




(a) $\zeta$ = 0.00005 m$^{-2}$, $\lambda$ = 0.005 m$^{-2}$ (b) $\zeta$ = 0.00005 m$^{-2}$, $\lambda$ = 0.01 m$^{-2}$ (c) $\zeta$ = 0.00001 m$^{-2}$, $\lambda$ = 0.005 m$^{-2}$

Fig. 4. Average data rate as a function of $\epsilon$.

### B. Comparison of Association Schemes

We continue by analyzing the performance of the alternative association algorithms in Fig. 4, which illustrates the mean achievable data rate as a function of the coefficient $\epsilon$ for several values of DgNBs and UAV IAB nodes densities. First, observe that the maximum feasible data rate obtained by using the scheme with the full knowledge of the instantaneous rate varies drastically as a function of DgNBs and UAV IAB nodes densities. Similar fluctuations are observed for the RSRP-based scheme. However, the gap between these schemes is relatively similar and remains within 24 to 35 Mbps.

Analyzing the data presented in Fig. 4, we may infer that the optimal value of $\epsilon$ is insensitive to the density of DgNBs and UAV IAB nodes. Indeed, as one may see for $\zeta$ = 0.00005 m$^{-2}$ and $\lambda$ = 0.005 m$^{-2}$, the optimal value is 0.09, while for the drastically different densities of $\zeta$ = 0.00001 m$^{-2}$ and $\lambda$ = 0.005 m$^{-2}$, the optimal value is 0.08. Given these results, the proposed reinforcement learning algorithm allows exploiting up to 50% of the gap between the naive RSRP-based method and the ideal scheme with the full knowledge of the connectivity options.

However, an unguided choice of $\epsilon$ may lead to severe performance degradation in terms of the mean achievable data rate. First, when $\epsilon$ increases from 0 up, the achievable data rate grows quickly towards its optimal value. Any further increase in $\epsilon$ leads to a slow degradation of the achievable data rate and eventually yields to a smaller data rate than in the case of the RSRP-based scheme. Therefore, the value of $\epsilon$ has been updated online, where the initial values can be selected within the range of 0.07 to 0.09. Similar conclusions hold for other system parameters.

To provide a direct comparison between all of the considered algorithms and study their response to a wider range of systems parameters, we illustrate the mean achievable data rate as a function of the density of DgNBs in Fig. 5 for the fixed density of UAV IAB nodes set to $\lambda$ = 0.005 m$^{-2}$ and the optimal values of $\epsilon$. We see that the mean achievable data rate increases exponentially, as the density of DgNBs in the area grows for all the alternative algorithms. As we observed in Fig. 4, the gap between the ideal and the RSRP-based schemes remains similar for all the densities of DgNBs. Proceeding with the analysis of data rate optimization algorithms, we learn that the semi-static classification-based scheme that identifies the set of the best DgNBs and UAV IAB nodes, performs similarly to the RSRP-based option at the times of connectivity changes, which offers no further benefit. Both reinforcement learning methods demonstrate significant improvements over the RSRP-based scheme, with marginal gain on top of the UCB approach for small values of the DgNB density. Conversely, a higher density of DgNBs in the area leads to a smaller gap between the considered reinforcement learning algorithms and the ideal scheme.

Finally, we assess the response of all algorithms as a function of the density of UAV IAB nodes in Fig. 6 for the fixed density of DgNBs with $\zeta$ = 0.00005 m$^{-2}$ and the optimal values of $\epsilon$. Here, one may observe that contrarily to Fig. 5, the gap between the ideal scheme with full knowledge of the instantaneous data rate and the naive RSRP-based approach increases as the density of UAV IAB nodes grows. The reason is that the higher density of UAV IAB nodes augments interference on the backhaul links. Analyzing the performance of the data rate optimization algorithms, the semi-dynamic classification-based approach shows no improvement over the naive RSRP-based option. Similarly to Fig. 5, both considered reinforcement learning schemes display similar performance and allow to bridge at least 50% of the available gap.

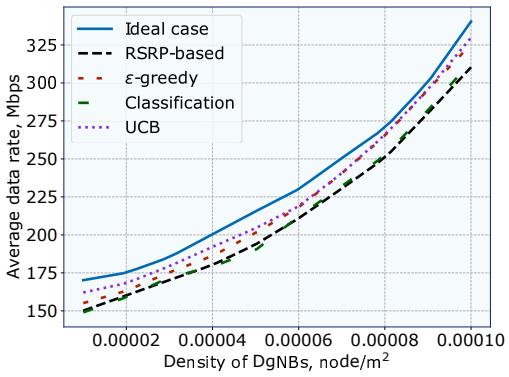



Fig. 5. Average data rate as a function of the density of DgNBs.

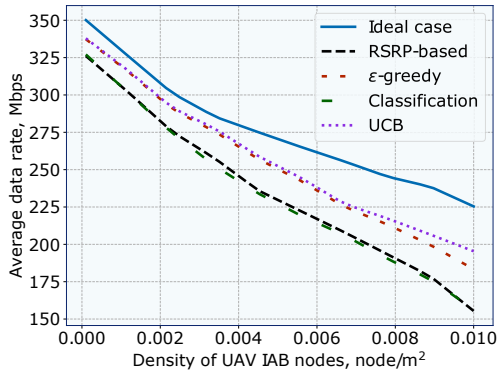



Fig. 6. Average data rate as a function of the density of UAV IAB nodes.

## V. Conclusion

The utilization of UAVs as moving relays with the recently introduced IAB feature is one of the promising ways to improve the data rate and coverage performance of beyond-5G systems. In our study, we account for the realistic deployment, propagation, antenna, and resource allocation models, to introduce several data rate optimization algorithms. These include semi- and fully-dynamic schemes, for which we investigate the performance in inherently non-stationary dynamic UAV-based IAB systems having mmWave radio technology at both IAB access and backhaul interfaces.

Our numerical results demonstrated that the considered reinforcement learning approach allows bridging up to 50% of the gap between the ideal scheme having the full knowledge of the instantaneous data rates and the straightforward RSRP-based association algorithm. However, the performance of the proposed solution heavily depends on the choice of its parameters, which is highly sensitive to the densities of DgNBs and UAV IAB nodes. Therefore, the parameters of this algorithm need to be provided in advance or learned online. Finally, we observed that the said algorithm nearly achieves the optimal data rate in dense DgNB deployments.

## Acknowledgment

This work was supported by Intel Corporation and by the project 5G-FORCE.

## References

[1] Drone Industry Insights, "The Drone Market Report 2019," https://www.droneii.com/project/drone-market-report.

[2] Verizon, "How Verizon 5G Ultra Wideband is lifting drone technology to the next level," https://www.verizon.com/about/our-company/5g/how-verizon-5g-ultra-wideband-lifting-drone-technology-next-level.

[3] A. Fotouhi, H. Qiang, M. Ding, M. Hassan, L. G. Giordano, A. Garcia-Rodriguez, and J. Yuan, "Survey on UAV cellular communications: Practical aspects, standardization advancements, regulation, and security challenges," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3417–3442, fourthquarter 2019.

[4] M. Mozaffari, W. Saad, M. Bennis, Y. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2334–2360, thirdquarter 2019.

[5] "Enhancement for Unmaned Aerial Vehicles (Release 17)," 3GPP, TR 22.829, Sep. 2019.

[6] N. Tafintsev, D. Moltchanov, M. Gerasimenko, M. Gapeyenko, J. Zhu, S.-p. Yeh, N. Himayat, S. Andreev, Y. Koucheryavy, and M. Valkama, "Aerial access and backhaul in mmWave B5G systems: Performance dynamics and optimization," *IEEE Communications Magazine*, vol. 58, no. 2, pp. 93–99, February 2020.

[7] M. Gapeyenko, V. Petrov, D. Moltchanov, S. Andreev, N. Himayat, and Y. Koucheryavy, "Flexible and reliable UAV-assisted backhaul operation in 5G mmWave cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 11, pp. 2486–2496, 2018.

[8] A. Fouda, A. S. Ibrahim, I. Guvenc, and M. Ghosh, "UAV-based in-band integrated access and backhaul for 5G communications," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, Aug 2018, pp. 1–5.

[9] N. Tafintsev, M. Gerasimenko, D. Moltchanov, M. Akdeniz, S.-p. Yeh, N. Himayat, S. Andreev, Y. Koucheryavy, and M. Valkama, "Improved network coverage with adaptive navigation of mmWave-based drone-cells," in *2018 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–7.

[10] I. Bor-Yaliniz, M. Salem, G. Senerath, and H. Yanikomeroglu, "Is 5G ready for drones: a look into contemporary and prospective wireless networks from a standardization perspective," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 18–27, Feb. 2019.

[11] "Study on Integrated Access and Backhaul (Release 16)," 3GPP, TR 38.874, Jan. 2019.

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

[13] M. Polese, M. Giordani, T. Zugno, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "Integrated access and backhaul in 5G mmWave networks: Potentials and challenges," *arXiv preprint arXiv:1906.01099*, 2019.

[14] M. Polese, M. Giordani, A. Roy, D. Castor, and M. Zorzi, "Distributed path selection strategies for integrated access and backhaul at mmWaves," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018, pp. 1–7.

[15] R. Kovalchukov, D. Moltchanov, A. Samuylov, A. Ometov, S. Andreev, Y. Koucheryavy, and K. Samouylov, "Evaluating SIR in 3D millimeter-wave deployments: Direct modeling and feasible approximations," *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 879–896, 2018.

[16] ——, "Analyzing effects of directionality and random heights in drone-based mmWave communication," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 10 064–10 069, 2018.

[17] "Study on channel model for frequencies from 0.5 to 100 GHz (Release 16)," 3GPP, TR 38.901, Oct. 2019.

[18] "Study of Radio Frequency (RF) and Electromagnetic Compatibility (EMC) requirements for Active Antenna Array System (AAS) base station (Release 12)," 3GPP, TR 37.840, Jan. 2014.

[19] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[20] D. Moltchanov, "State description of wireless channels using change-point statistical tests," in *International Conference on Wired/Wireless Internet Communications*. Springer, 2006, pp. 275–286.

[21] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *KDD*, vol. 96, no. 34, 1996, pp. 226–231.

[22] M. E. Harmon and S. S. Harmon, "Reinforcement learning: A tutorial." Tech. Rep., 1997.

[23] E. Babulak and M. Wang, "Discrete event simulation: state of the art," in *Discrete Event Simulations*. IntechOpen, 2010.

[24] G. S. Fishman and L. S. Yarberry, "An implementation of the batch means method," *INFORMS Journal on Computing*, vol. 9, no. 3, pp. 296–310, 1997.

[25] H. Wong and S. Ling, "Mixed portmanteau tests for time-series models," *Journal of Time Series Analysis*, vol. 26, no. 4, pp. 569–579, 2005.

[26] H. Perros, "Computer simulation techniques," *The definitive introduction. North Carolina State University*, 2009.

[27] V. Petrov, M. Komarov, D. Moltchanov, J. M. Jornet, and Y. Koucheryavy, "Interference and SINR in millimeter wave and terahertz communication systems with blocking and directional antennas," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1791–1808, 2017.