Queue-based Eco-Driving at Roundabouts with Reinforcement Learning

Anna-Lena Schlamp[1], Werner Huber[2], Stefanie Schmidtner[3]

***Abstract*—We address eco-driving at roundabouts in mixed traffic to enhance traffic flow and traffic efficiency in urban areas. The aim is to proactively optimize speed of automated or non-automated connected vehicles (CVs), ensuring both an efficient approach and smooth entry into roundabouts. We incorporate the traffic situation ahead, i.e. preceding vehicles and waiting queues. Further, we develop two approaches: a rule-based and an Reinforcement Learning (RL) based eco-driving system, with both using the approach link and information from conflicting CVs for speed optimization. A fair comparison of rule-based and RL-based approaches is performed to explore RL as a viable alternative to classical optimization. Results show that both approaches outperform the baseline. Improvements significantly increase with growing traffic volumes, leading to best results on average being obtained at high volumes. Near capacity, performance deteriorates, indicating limited applicability at capacity limits. Examining different CV penetration rates, a decline in performance is observed, but with substantial results still being achieved at lower CV rates. RL agents can discover effective policies for speed optimization in dynamic roundabout settings, but they do not offer a substantial advantage over classical approaches, especially at higher traffic volumes or lower CV penetration rates.**

I. INTRODUCTION

In the course of climate change, reducing pollutant emissions is becoming increasingly important. In 2019, road transport was responsible for 26% of all CO2 emissions in the European Union [1], emphasizing the need for innovative measures on optimizing traffic efficiency in order to decisively contribute to reducing emissions in urban areas [2]. Automated driving (AD) is one innovative technology that provides great potential for improving traffic efficiency [3]. Introducing innovative algorithms for automated vehicle (AV) control allows driving behavior to be adapted to traffic conditions at an early stage. This holds the promise of liquefying traffic by minimizing stop-and-go behavior, idling times and accelerations. Taking advantage of infrastructure-to-vehicle (I2V) and vehicle-to-vehicle (V2V) communications offers additional improvements by providing connected automated vehicles (CAVs) early access to traffic-related information [4]. When it comes to automated driving, there are various challenges especially posed by roundabouts as merging and driving behavior solely depends on road users' interactions [5], [6]. Automated vehicles depend on a wide range of information on the environment and surrounding road users for being able to fully take over the driving task [3]. But due to the roundabouts' high dynamics and interaction density, perception and maneuver planning in approaching and merging can be a complex task. This paper proposes a proactive vehicle control approach to alleviate the challenges of automated driving at roundabouts. The question of how to eco-friendly approach a roundabout and by that ensure a more efficient and safe merging into the roundabout is described. A classic rule-based and a Reinforcement Learning (RL) based approach are compared.

[1]Anna-Lena Schlamp is with the Institute AImotion Bavaria, Technische Hochschule Ingolstadt, 85049 Ingolstadt, Germany anna-lena.schlamp@thi.de

[2]Werner Huber is with the CARISSMA Institute of Automated Driving (C-IAD), Technische Hochschule Ingolstadt, 85049 Ingolstadt, Germany werner.huber@thi.de

[3]Stefanie Schmidtner is with the Institute AImotion Bavaria, Technische Hochschule Ingolstadt, 85049 Ingolstadt, Germany stefanie.schmidtner@thi.de

A. Eco-Driving at Roundabouts

Given the potential of connected automated driving as well as the challenges posed by roundabouts, a lot of research is being done to develop CAV control systems at roundabouts. Considering traffic efficiency, there have been many studies addressing the coordination of CAVs at roundabouts in a fully automated setting. They primarily aim at improving stop-and-go behavior and waiting times when entering and by that optimizing traffic efficiency as a second effect. By means of different techniques, such as forming clusters [7] or determining sequence and speed trajectories [8], [9], the vehicles are steered in a way that their driving behavior, i.e. speed or acceleration, is optimized. Optimization starts at a distance of 200 meters [7], 275 meters [9] or 300 meters [8] from the roundabout merging point. In all studies, a reduction in travel and waiting times, fuel consumption and emissions is observed depending on the setting and the traffic demand.

But these studies suffer from a number of limitations. They have shown that waiting times cannot be avoided completely by coordination [9], indicating that queues can still occur in front of the roundabout merging point. However, there is no research being done so far taking into account queues or vehicles driving ahead, even though studies on eco-driving at signalized intersections demonstrated that significant improvements can be achieved by that (see next paragraph). Optimal speeds or accelerations are derived by analytical optimization. Machine Learning (ML) based methods such as Reinforcement Learning are not yet applied, although its potential has already been demonstrated in eco-driving strategies at signalized intersections (see Section I-C). In addition, previous studies only focus on optimizing the entry into roundabouts in terms of waiting times and stops. Fuel consumption and emissions during approach are not explicitly considered in the optimization process. Even accelerating towards the roundabout is encouraged if a cluster or a gap can be caught, which counteracts the aspect of traffic efficiency. Optimization starts at a maximum of 300 meters ahead, yet strategies for efficient driving at signalized intersections suggest considering the approach to minimize accelerations at an early stage. The focus of previous works is on coordinated driving behavior in a fully automated setting. In their settings, a full CAV penetration rate is actually necessary to achieve significant improvements [3], [8]. These approaches are therefore of limited applicability to settings of mixed traffic.

B. Queue-based Eco-Driving at Signalized Intersections

There is a wide range of established eco-driving systems for approaching signalized intersections. A well-known concept in urban traffic research is the Green Light Optimized Speed Advisory (GLOSA) system [10], which relies on rule-based calculations of vehicle motion (see Eq. 1). Information on signal phase and timing (SPAT) is used to early recommend an optimal speed to connected vehicles (CVs) or CAVs that allows them to efficiently approach and pass a signalized intersection at green phase (for more details see e.g. [11]). Several papers, such as [10], [12], reveal beneficial effects on traffic flow and traffic efficiency. It is shown that waiting and travel times, fuel consumption and emissions are reduced with increasing GLOSA penetration rate and traffic demand. With regard to the ideal activation distance, the studies have revealed that optimization needs to be done at least 350 meters ahead of the traffic light. Best results are obtained at higher distances of around 500 meters, depending on the setting [10], [13].

Classic GLOSA only receives SPAT information for the next traffic light and ignores the traffic situation ahead like preceding vehicles or waiting queues. As a consequence, the speed advisory can be counteracted by the traffic situation, causing unexpected stops or accelerations as well as minor improvements in terms of traffic efficiency. To avoid this, in [14] for instance, the traffic density is included into the GLOSA algorithm. But the results lack in accuracy as the traffic density is not directly related to the waiting time. Besides, in our previous work [11], we extended the classic GLOSA approach by including queue information, i.e. queue length and queue dissipation time. Further improvements in traffic flow and traffic efficiency compared to classic GLOSA were achieved by this approach.

C. Reinforcement Learning and Eco-Driving

Beyond the application of rule-based formulas derived from classic driving physics, finding an optimal speed can also be seen as an optimization problem and thus be formalized as a Markov Decision Process (MDP) [15]. With respect to eco-driving at roundabouts, i.e. the coordination of CAVs, recent studies deal with formulating the problem and solving it analytically. Besides employing analytical methods, the use of Machine Learning and in particular Reinforcement Learning has proven its potential across various optimization domains [16]. A so-called agent is tasked to solve a problem that is formalized as a MDP. To achieve this, the agent learns to make decisions by interacting with its environment and by maximizing rewards, i.e. feedback on its actions. Exploiting the capabilities of neural networks as function approximators for policy representation, a powerful framework called Deep Reinforcement Learning (DRL) is established. Within this framework, there are three different categories of algorithms trying to find the optimal policy: Q-learning, policy-based learning, and actor-critic learning, which combines the two [16]. Given its potential, RL has already been successfully applied in the context of GLOSA. Actor-critic learning has been proven in several studies to be suitable for learning optimal speeds or accelerations [17], [18], [19]. As combining two separate components, an actor that learns to make decisions by selecting actions and a critic that evaluates these actions to provide feedback on their quality, it allows for stable and efficient learning in complex and high-dimensional environments. For example, in [17], a Twin-Delayed Deep Deterministic Policy Gradient (TD3) agent was implemented and trained. Results have shown that the RL-based GLOSA achieves significant improvements compared to a GLOSA reference system and a human driver. However, the GLOSA reference system was not designed according to well-established algorithms, but was rather a proprietary development. For this reason, in our previous work [11], we performed a fair comparison of classic rule-based and RL-based GLOSA in a common comparable environment. It has been demonstrated that an RL agent can successfully learn an optimal speed for approaching signalized intersections, but does not offer a substantial advantage over classic GLOSA. A similar approach should also be possible for eco-driving at roundabouts, but has not been tested so far. Compared to signalized intersections, roundabouts include higher traffic dynamics and less predictable entrance gaps. A Machine Learning approach might here be even more beneficial compared to a classic approach.

D. Contributions

This paper aims to address the limitations of previous roundabout eco-driving studies, i.e. coordination approaches, mentioned in section I-A, by presenting a novel eco-driving strategy for automated and non-automated connected vehicles at roundabouts. It fills the gap of considering the traffic situation ahead, i.e. preceding vehicles and waiting queues, when optimizing the speed. In addition, it faces the challenge of applying RL to speed optimization problems when approaching roundabouts. A fair comparison of rule-based and RL-based eco-driving at roundabouts in a common comparable environment is being performed. As a secondary feature, speed is optimized well in advance by incorporating the approach link on a length of 500 meters. Thus, we address two target criteria: the reduction of stop-and-go behavior on entering and the minimization of accelerations during approach. In this way, both an efficient approach and a smooth entry into roundabouts should be ensured. Recent work has focused on settings with full CAV penetration rate to achieve a more fluent traffic behavior, i.e. assume a fully automated setting [7], [8], [9]. In contrast, we aim for applicability of our approach to mixed traffic with automated and non-automated CVs. We evaluate different penetration rates of CVs to find out whether significant results can still be obtained at lower CV penetration rates. There are three main contributions of this study:

(1) An eco-driving system for approaching and entering a roundabout is developed, which integrates information on vehicle queues ahead and optimizes speed at an early stage.
(2) A Reinforcement Learning based roundabout eco-driving system incorporating queue information is designed. In addition, a fair comparison of rule-based and RL-based approaches in a common comparable environment is performed.
(3) The role of traffic volume and CV penetration rate is examined considering both rule-based and RL-based eco-driving at roundabouts.

II. Implementation of Algorithms

A. Gym Environment

To be able to apply RL algorithms to an optimization problem, a suitable environment must be created. Doing so, we refer to the environment presented in our previous work [11]. Relying on a standardized interface called gym [20], it enables the use of pre-implemented RL algorithms and libraries like [21]. A microscopic traffic simulation, specifically the Simulation of Urban MObility (SUMO) [22], forms the core component of our gym environment. It allows the agent to have numerous interactions in order to find a satisfactory policy. Integrating it with the gym framework, we define a state representation S to be extracted from the SUMO simulation, which serves as input to the RL algorithm's neural network. In order to establish the agent's optimization objective, a reward function R is defined, relying on information obtained from SUMO and ultimately yielding to a single numerical value. In each iteration of the so-called RL-loop, the agent determines an action, represented by a continuous value within the permissible speed range, based on the current policy and the state information from the previous time step. During each simulation step, i.e. every second, the current speed advisory is employed and the current reward is incrementally accumulated. To be able to perform a fair comparison of rule-based and RL-based eco-driving, we also embed the rule-based algorithm into the gym environment structure. In each simulation step, it therefore receives a state and a reward, denoted as 1, as input, uses the state to internally run various calculations and ultimately returns a speed advisory as an action. Our work is based on a real urban single-lane roundabout which is part of a test field for connected automated driving located in Ingolstadt, Germany. As can be seen in Fig. 1, it has four links, south and east with double lanes, and a shortcut that can be used to go from south to east directly without having to pass the roundabout. The roundabout with its links is introduced into the SUMO simulation according to real dimensions. Random traffic is generated, consisting of conventional passenger cars following the standard SUMO vehicle model [23]. Each possible route is equally distributed among all vehicles, with them being randomly inserted into

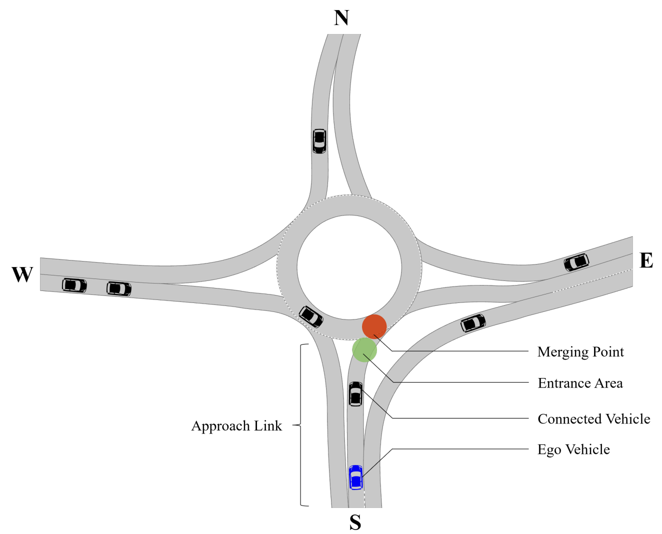

Fig. 1: Visualization of the urban roundabout located in Ingolstadt with the ego vehicle's approach link, entrance area (green) and merging point (red).

the traffic network at a speed of 50 km/h. Despite focusing on this particular roundabout, our approaches are generic and therefore also applicable to other roundabouts.

In this initial work, our approaches are intended to be applied to one single ego vehicle that is connected or connected automated and directly implements the given speed advisory. Always following the same route, it drives towards the roundabout on the left-hand lane of the southbound approach link and moves on to the northern link. Overtaking the ego vehicle is prohibited, as this can affect the results due to the varying number of vehicles ahead. Optimization starts at the point the ego vehicle reaches a distance of 500 meters to the roundabout merging point.

B. Rule-based Eco-Driving with Queue Information

Assuming that all vehicles are connected and share their information such as route, position and speed, we generate status plans for the relevant sectors of conflicting traffic, i.e. merging point (see Fig. 1, red) and entrance area (see Fig. 1, green) to be passed by the ego vehicle. They indicate whether the corresponding sector is free or occupied by other vehicles. To do so, we iterate over all vehicles at the beginning of the simulation. Considering the roundabout's geometric dimensions, we derive a maximum possible speed of 30 km/h for driving in it. Consequently, we assume that vehicles are blocking the merging point for one second. German road traffic regulations specify that a safety distance of at least one second between two consecutive vehicles must be maintained in urban areas with speeds below 50 km/h [24]. For that reason, we set the point of time in the status plan at which a vehicle passes the merging point as well as the second before and after to occupied. This results in an overall occupancy duration of three seconds per vehicle. In addition, a second status plan is created to cover queuing vehicles at the entrance area. Again including a safety distance, the status is set to be occupied for two seconds at the time a vehicle with a speed of less than 15 km/h blocks the entrance area. This addresses slow-moving vehicles shortly before entering the roundabout. Considering a vehicle waiting for a longer period of time until being able to enter, it ensures the entrance area to be occupied for the total waiting and clearing time. If a queue of several vehicles has already been formed, this procedure further guarantees that the status is set to occupied until the queue has cleared. In this way, stopping as well as braking of the ego vehicle caused by other vehicles or queues ahead can be avoided in all cases.

Building upon the classic GLOSA concept [10], a speed optimization strategy for roundabouts is developed, which is summarized in Algorithm 1. The queue-based optimal speed is determined by executing the speed optimization algorithm in two iterations, taking into account both factors: occupancy of the merging point and occupancy of the entrance area. As input, the state representation defined in our gym environment is used, extracting the following information from the simulation:

(1) The current speed v of the ego vehicle.
(2) The distance d to the entrance area.
(3) The occupancy status of the entrance area at arrival.
(4) The status transition time t_{free} from occupied to free with regard to the entrance area.
(5) The distance d to the merging point.
(6) The occupancy status of the merging point at arrival.
(7) The status transition time t_{free} from occupied to free with regard to the merging point.

First, an optimal speed for passing the entrance area is calculated with the help of Algorithm 1. If the entrance area is free upon arrival, the ego vehicle continues at the maximum possible speed v_{free} for status free. If being occupied, the algorithm calculates an optimal speed v_{opt} at which the ego vehicle can pass the entrance area smoothly without stopping at the time of status transition. Doing so, basic rules of motion are used, given by Eq. (1). Being important to comply with the permitted speed limits, it is also ensured that the output speed is not lower than the minimum permitted vehicle control speed to not present a traffic obstruction driving too slow. The outcome is then used to iteratively execute the algorithm a second time, now including the merging point distance, occupancy status and the calculated optimal entrance speed. Replacing the variable of the current speed by the optimal speed defined for the entrance area, it is checked whether the merging point can also be reached in a free status when driving at the entrance speed. If this is not the case, a lower optimal speed is calculated in order to reach both the entrance area and the merging point in a free status.

$$v_{opt} = \frac{2 * d}{t} - v \tag{1}$$

Algorithm 1 Speed Optimization

1: Calculate time of arrival t_a based on distance d and current speed v
2: Check status at t_a
3: **if** FREE **then**
4: Continue trip
5: Optimal speed v_{opt} = Maximum possible speed v_{free} for status free
6: **else if** OCCUPIED **then**
7: Get status transition time t_{free} from occupied to free starting from the time of arrival t_a
8: Calculate optimal speed v_{opt} for t_{free} + t_a
9: **end if**
10: Output speed v_{out} = max(v_{opt}, v_{min}) & min(v_{opt}, v_{max})

C. RL Agent Design

The Soft Actor Critic (SAC) algorithm is well-suited for speed optimization due to its effectiveness in handling complex continuous control problems. It integrates actor-critic methods with entropy regularization to promote exploration and robust learning. By maximizing both expected rewards and entropy, a measure of uncertainty or randomness, SAC achieves a balance between exploration and exploitation. Having successfully applied SAC to GLOSA optimization problems in our previous work [11], it is also employed to the present roundabout optimization problem. Similarly, a Multi-Layer Perceptron (MLP) is utilized as the network architecture for both the actor and critic components. Each component contains two hidden layers, each with 256 neurons. The agent's output is represented by a continuous value within the range of -1 to 1. Thus, we map the interval $[v_{\text{min}}; v_{\text{max}}]$ to the corresponding speed advisory. To ensure algorithm performance is unaffected by implementation specifics, we adopt the SAC implementation outlined in [21].

A multi-objective optimization is aimed by using a linear combination of three components as reward function R. One component, denoted as r_{waiting}, involves assigning a negative reward when the vehicle is stopped. In addition, the deviation r_{diff} between the current speed advisory and the previous one is considered to mitigate fluctuating speed advisories. However, r_{diff} is excluded from the calculation for the first action after simulation begins, as the agent should not be penalized for adjusting its speed in this instance. As a third component r_{speed}, we incorporate the difference between v_{max} and the speed advisory to encourage the agent selecting the highest possible speed and minimizing an increase in travel time. Empirically determined coefficients of 1 for r_{waiting}, 0.05 for r_{diff}, and 0.05 for r_{speed} have proven to be effective in roundabout speed optimization. The total reward for one simulation step is then calculated as follows:

$$r_f = 1 * r_{\text{waiting}} + 0.05 * r_{\text{diff}} + 0.05 * r_{\text{speed}} \tag{2}$$

As a state representation S, the agent receives the same information as used by the rule-based eco-driving algorithm described in Section II-B. A simple encoding scheme, representing a free status as 1 and occupancy as 0, is used to enable the neural network processing the status information upon arrival (see state representation (3) and (6)). If the status is free upon arrival, instead of status transition time the value 1 is assigned for state representation (4) and (7).

III. EVALUATION

A. Evaluation Setup

After implementation, we evaluated both approaches, rule-based and RL-based eco-driving, using the roundabout traffic simulation described in Section II-A. Performing a generic evaluation, settings differ in the modeled traffic volume with {600, 800, 1000, 1200, 1400} vehicles per hour (veh/h) entering the network. The upper limit of 1400 veh/h corresponds to the empirical determined capacity of our roundabout scenario. In addition, different CV penetration rates of {20, 40, 60, 80, 100} percent were examined. Our approach is tested on one vehicle – the ego vehicle. It needs to be a CV for the approach, but can also be a CAV. Its maximum allowed speed is 50 km/h, as usual in urban areas. Besides, we set the minimum permitted speed when controlling the ego vehicle to 25 km/h, half the maximum speed according to [10], to not create traffic obstructions by driving too slow. In order to investigate which type of vehicle benefits most from our eco-driving system, we evaluate trips with both an internal combustion engine (ICE) vehicle and a battery electric vehicle (BEV) vehicle. To train the RL algorithm, we expose the ego vehicle to different settings by randomly starting into the simulation at varying traffic volumes. The training involves 20×10^4 iterations. For evaluation of both approaches, we initiate simulations at 30 randomly selected time points for each traffic volume and CV penetration rate. During these evaluation trips, performance metrics such as energy and fuel consumption, CO2 emissions, travel and waiting time as well as number of stops are recorded to test the various configurations. In addition, at each traffic volume, we analyze the rate of trips that require an optimization of the speed trajectory. That means the percentage of trips where merging point or entrance area are not reachable in a free status, i.e. the ego vehicle is forced to stop if not adjusting its speed. Further, we calculate the proportion of fully optimized trips where stops are eliminated. Both rates are measured in relation to the total number of trips, which allows for direct comparison. Regardless of whether optimization is performed or not, the performance metrics are averaged over all 30 evaluation trips, even including those with +/-0% improvement. Given a strong performance dependency on the trips' random traffic scenario, several evaluations on different sets of 30 random trips were performed for each setting to verify the results' robustness, showing comparable outcomes.

B. Rule-based Eco-Driving Performance

The rule-based eco-driving system considering queue information produces substantial improvements compared to the baseline without speed optimization (see Table I, left and Fig. 2). Best results are obtained at a traffic volume of 1200 veh/h. Being forced to stop in 80% of all trips, 37% can be fully optimized in a way that stops are eliminated. Average energy consumption is reduced by about 11%, fuel consumption along with emissions are lowered by 6%. Travel time decreases by 2%, waiting time and stops by 61% and 48% respectively. Close to capacity, i.e. at a traffic volume of 1400 veh/h, a drawback becomes clear in all metrics except BEV energy consumption. However, improvements in waiting times and stops remain at a high level of 56% and 48%. Investigations with 600 and 800 veh/h reveal only minor improvements in fuel economy and travel time, resulting in a growing algorithm performance up to a traffic volume of 1200 veh/h. Nevertheless, already at low traffic volumes, waiting times and stops can be substantially reduced and the majority of optimizable trips can be fully optimized.

The more vehicles circulate at the roundabout, the more the ego vehicle's driving behavior is affected. Correspondingly, the rate of optimizable trips in which the ego vehicle is forced to stop increases with increasing traffic volume. When applying speed optimization, two categories of optimizable trips can be differentiated: (1) fully optimizable trips, in which waiting times and stops can be completely avoided by optimization, and (2) partially optimizable trips, in which stops cannot be eliminated, but at least waiting times can be reduced. Partially optimizable trips result from the minimum permitted speed restriction to half of the maximum speed as it forces the ego vehicle to approach the occupied roundabout at 25 km/h instead of driving even slower to catch a gap. Higher traffic volumes increase the occurrence of trips of category (2) due to intense occupancy, which means that fewer trips can be fully optimized compared to the rate of optimizable trips. Thus, at 1400 veh/h, stops still occur in 53% of trips but cannot be reduced to a greater extent, which causes the drawback in performance. Being part of our contributions to optimize speed already during approach, i.e. from a distance of 500 meters, we verified its benefits by examining the system's performance when starting optimization at distances of {100, 200, 300, 400, 500} meters to the roundabout. The improvements compared to the baseline decline with decreasing optimization distance, leading to best results being achieved at the highest investigated distance of 500 meters. Previous studies used an optimization distance of 200 to 300 meters. Our results show that enlarging the distance up to 400 meters leads to more than twice as good improvements in comparison to 200 meters.

C. RL-Agent Performance

Applying the RL-based roundabout eco-driving system, results demonstrate that the RL agent is capable of finding a meaningful policy. Similar to the rule-based approach, improvements increase with increasing traffic volume and best results are obtained at 1200 veh/h (see Table I, right). The agent achieves an average reduction in waiting time of 37% and in stops of 32%. Emissions decrease by about 3% while energy consumption is reduced by 7%. Travel time is shortened by about 1%. Similar to the rule-based approach, a high rate of trips is fully optimized, i.e. stops are eliminated in 30% of all trips. Again, a drawback close to capacity becomes apparent.

Compared to the classic rule-based eco-driving approach, the BEV experiences a disadvantage across all traffic volumes since the agent cannot maintain an approach speed precisely

TABLE I: Rule-based and RL-based queue algorithm improvement in performance metrics for different traffic volumes [veh/h], averaged over 30 evaluation trips.

Performance Metric	Rule-based Eco-Driving					RL-based Eco-Driving				
	600	800	1000	1200	1400	600	800	1000	1200	1400
Rate of Optimizable Trips	33.3%	40.0%	63.3%	80.0%	86.7%	33.3%	40.0%	63.3%	80.0%	86.7%
Rate of Fully Optimized Trips	33.3%	33.3%	33.3%	36.7%	33.3%	33.3%	30.0%	23.3%	30.0%	16.7%
BEV Energy Consumption	-2.7%	-3.8%	-9.3%	-11.3%	-11.9%	-2.2%	-3.1%	-5.2%	-7.1%	-7.5%
Fuel Consumption & CO2 Emissions	-0.7%	-1.2%	-2.2%	-5.6%	-2.2%	-1.1%	-1.4%	-1.0%	-3.4%	-2.7%
Travel Time	+0.1%	-0.1%	-0.4%	-1.7%	-0.6%	+1.3%	+0.8%	+0.5%	-1.4%	-1.4%
Waiting Time	-33.3%	-34.4%	-51.4%	-60.7%	-56.1%	-33.3%	-32.7%	-30.9%	-36.5%	-27.5%
Number of Stops	-33.3%	-35.6%	-39.4%	-48.3%	-47.8%	-33.3%	-31.1%	-25.6%	-31.7%	-16.6%

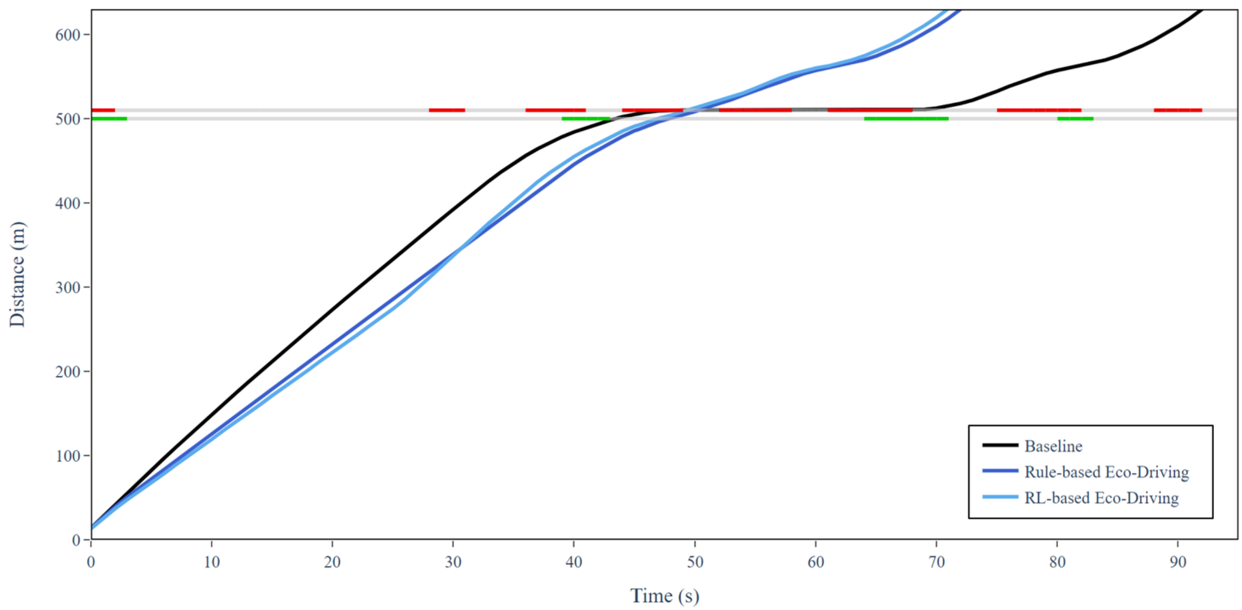

Fig. 2: Trajectories of the baseline (black), rule-based (dark blue) and RL-based (light blue) eco-driving with queue information from one sample trip at a traffic volume of 1200 veh/h. The horizontal lines correspond to occupancy status of entrance area (lower line) and merging point (upper line), with the sectors' color indicating occupancy by other vehicles.

over 500 meters which notably affects energy consumption. At low traffic volumes of 600 and 800 veh/h, the RL agent achieves comparable results in all metrics, with only a minor drawback in energy consumption and travel time. This is because at low traffic volumes, primarily trips of category (1), i.e. fully optimizable trips with stop elimination, occur and the agent in these performs similar to the rule-based approach (see Fig. 2). However, in comparison, the RL-based performance decreases as the traffic volume increases. The decline is more pronounced in terms of waiting time and stops than in energy, emissions and travel time. At higher traffic volumes, the rule-based eco-driving system achieves up to twice as good improvements in waiting time and stops than the RL-based approach. This is because the agent experiences a disadvantage in trips of category (2) where the minimum speed restriction prevents a full elimination of stops, as it does not constantly approach the roundabout with exactly the minimum permitted speed. The third component of the reward function presents a limiting factor in this regard as it encourages the agent to perform less pronounced speed reductions and thus minimize an increase in travel time. Training the agent without considering this reward component produces improvements in waiting time and stops for trips of category (2), but deteriorates the performance in terms of travel time, energy consumption and emissions in non-optimizable trips and trips of category (1), especially at lower traffic volumes. This creates a trade-off between vehicle efficiency and stop-and-go reduction across different traffic volumes.

D. Impact of CV Penetration Rate

In previous evaluations, a full CV penetration rate was always assumed when applying the eco-driving algorithms, meaning that all vehicles share their information. Thus, the occupancy status plans for merging point and entrance area were generated with maximum reliability. To examine the system's real-world applicability, we investigate whether substantial improvements can also be achieved even at lower CV penetration rates. Different CV penetration rates are analyzed by randomly modeling non-connected vehicles providing no information. This means that these vehicles are not included in generating the status plans, which causes uncertainty in occupancy information for both merging point and entrance area. Applying both approaches, rule-based and RL-based eco-driving, best results are achieved at full CV penetration rate of 100%, as can be seen in Table II (left) exemplarily for a traffic volume of 1200 veh/h. As the penetration rate decreases, the

TABLE II: Rule-based and RL-based queue algorithm improvement in performance metrics for different CV penetration rates [%] at a traffic volume of 1200 veh/h, averaged over 30 evaluation trips.

Performance Metric	Rule-based Eco-Driving					RL-based Eco-Driving				
	20	40	60	80	100	20	40	60	80	100
Rate of Optimizable Trips	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%
Rate of Fully Optimized Trips	3.3%	3.3%	16.7%	30%	36.7%	3.3%	3.3%	10.0%	16.7%	30.0%
BEV Energy Consumption	-2.3%	-3.2%	-6.4%	-10.6%	-11.3%	-1.1%	-1.0%	-2.9%	-5.0%	-7.1%
Fuel Consumption & CO2 Emissions	-1.7%	-0.1%	-3.5%	-5.5%	-5.6%	-0.9%	-0.6%	-2.0%	-2.4%	-3.4%
Travel Time	-0.7%	+0.1%	-1.1%	-1.7%	-1.7%	+0.1%	-0.1%	-1.1%	-1.4%	-1.4%
Waiting Time	-14.9%	-15.5%	-32.8%	-56.5%	-60.7%	-7.9%	-11.7%	-20.9%	-27.4%	-36.5%
Number of Stops	-8.9%	-7.2%	-22.2%	-40.0%	-48.3%	-6.4%	-7.5%	-13.6%	-21.3%	-31.7%

rate of fully optimized trips and the performance steadily deteriorate. This decline had been expected, as with lower rates, the ego vehicle may not be able to adjust its speed accordingly due to the unknown occupancy of merging point or entrance area, resulting in stops or accelerations being unavoidable. Nevertheless, results show that even at low penetration rates the eco-driving approaches achieve improvements compared to the baseline. With the rule-based eco-driving algorithm, substantial results can still be obtained with 80% CVs. Energy consumption, emissions and travel time are reduced to the same extent as with the 100% setting. In comparison, the improvements in waiting time and number of stops decline by a magnitude of one-tenth to one-fifth, but still emerging a notable reduction of 57% and 40% respectively. The performance of the RL-agent is notably affected to a greater extent. With 80% CVs, the observed improvements in energy and waiting time decrease by around one-fourth, emissions and stops by one-third. Hence, the agent demonstrates lower robustness to lower CV penetration rates, i.e. information availability on conflicting vehicles.

IV. DISCUSSION AND CONCLUSION

In this study, we developed a novel speed optimization based eco-driving system for efficient control of automated and non-automated CVs at roundabouts. We faced the challenge of incorporating the traffic situation ahead, i.e. preceding vehicles and waiting queues. In addition, we performed a fair comparison of rule-based and RL-based eco-driving in a common comparable environment in order to explore RL as a viable alternative to classical optimization at roundabouts. As a secondary feature, by including the approach link on a length of 500 meters, we aimed at optimizing speed well in advance to minimize accelerations during approach and reduce stop-and-go behavior on entry. To ensure applicability in settings of mixed traffic, we pursued the approach of non-coordinated driving behavior, eliminating the need for full CAV penetration rate. As our proposed eco-driving system consequently requires other vehicles to only be connected, we investigated whether even lower CV penetration rates can already produce substantial improvements.

Previous approaches suffer from the lack of information on vehicles ahead, thus being of limited applicability especially to higher traffic volumes. Being aware of preceding vehicles, our proposed system is able to avoid queue-induced stops and accelerations. Both rule-based and RL-based eco-driving therefore outperform the baseline without speed optimization in terms of energy and fuel consumption, emissions, travel and waiting time as well as number of stops. Comparing energy and fuel savings, the BEV benefits more from our proposed systems than the conventional ICE vehicle. Best results are achieved at high traffic volumes. However, a drawback near capacity is observed caused by restricting the minimum permitted speed to half of the maximum speed. Lower optimization distances degrade performance, leading to best results being obtained when optimization is initiated at a distance of 500 meters to the roundabout. Investigations with lower CV penetration rates reveal diminishing improvements of both approaches with decreasing percentage, but no significant deterioration. The decline is more pronounced for lower penetration rates, leading to still notable results with 80% of vehicles being connected.

In comparison, at low traffic volumes, the RL agent produces similar results as the rule-based approach. Nevertheless, at high traffic volumes, the rule-based eco-driving system significantly outperforms the RL-based approach, since it experiences a disadvantage in trips where the minimum speed restriction prevents a full elimination of stops, i.e. the roundabout should be approached at the minimum speed. This creates a trade-off between reducing waiting times and stops versus improving fuel economy and travel time over increasing traffic volumes. Thus, a linear reward function might not be sufficient in dynamic roundabout settings with short-term entrance gaps and higher stop probability despite optimization. Further, the RL agent is more affected by modeling a CV penetration rate in the simulation.

To conclude, this study demonstrates that RL agents can discover effective policies for speed optimization at roundabouts. Nevertheless, in settings without uncertainties, the agents do not demonstrate a substantial advantage over classical approaches, especially at higher traffic volumes or lower CV penetration rates. Incorporating information on vehicles ahead helps to reduce the negative impacts of queues on traffic efficiency, thus obtaining traffic adaptivity. The more vehicles in the network, the more the ego vehicle's driving behavior is affected and the higher the positive effects of our proposed eco-driving system. However, the algorithms' performance is limited at capacity limits. Considering real-world implementation, even low CV penetration rates are sufficient to achieve improvements.

This work is a first step towards RL-based roundabout eco-driving in mixed traffic environments. Future research should expand this approach by controlling multiple vehicles with the challenge of finding the optimum among numerous options. In this case, the rule-based approach is expected to reach its limits. Thus, the role of RL will be amplified through the development of a multi-agent system. The RL agent's design should be enhanced to achieve balanced improvements across different traffic volumes. Additionally, it is worth investigating if the agent's design can be further optimized to achieve favorable outcomes even at low CV rates. Further investigations in mixed traffic with a varying compliance rate and response time of human drivers as well as uncertainty in communication capabilities should be performed. The question of how to integrate vulnerable road users into optimization should also be addressed. Real-world experiments in mixed traffic should be conducted to quantify uncertainties, i.e. the gap between simulation and reality.

ACKNOWLEDGMENT

This study was partially funded by the German Federal Ministry for Digital and Transport (Bundesministerium für Digitales und Verkehr, BMDV) within the research project "5GoIng – 5G Innovation Concept Ingolstadt".

REFERENCES

[1] Europäisches Parlament. CO2-Emissionen von Pkw: Zahlen und Fakten. URL: https://www.europarl.europa.eu/topics/de/article/20190313STO31218/co2-emissionen-von-pkw-zahlen-und-fakten-infografik. (visited on 04/25/2024).

[2] Das Umweltbundesamt. Klimaschutz im Verkehr. URL: https://www.umweltbundesamt.de/themen/verkehr-laerm/klimaschutz-im-verkehr. (visited on 02/16/2023).

[3] Umberto Montanaro et al. "Towards Connected Autonomous Driving: Review of Use-Cases". In: *Vehicle System Dynamics* Vol. 57 (July 2018), pp. 779–814. DOI : 10.1080/00423114.2018.1492142.

[4] Zaigham Mahmood. *Connected Vehicles in the Internet of Things*. Springer, 2020. Chap. Connected Vehicles in the IoV: Concepts, Technologies and Architectures.

[5] Muhammad Naveed and Björn Åstrand. "Predicting Agent Behaviour and State for Applications in a Roundabout Scenario Autonomous Driving". In: *Sensors* (Oct. 2019). DOI : 10.3390/s19194279.

[6] Bunyo Okumura et al. "Challenges in Perception and Decision Making for Intelligent Automotive Vehicles: A Case Study". In: *IEEE Transactions on Intelligent Vehicles* (Apr. 2016), pp. 20–32. DOI : 10.1109/TIV.2016.2551545.

[7] A. S. M. Bakibillah et al. "Bi-Level Coordinated Merging of Connected and Automated Vehicles at Roundabouts". In: *Sensors* (Sept. 2021). DOI : 10.3390/s21196533.

[8] Liuhui Zhao, Andreas Malikopoulos, and Jackeline Rios Torres. "Optimal Control of Connected and Automated Vehicles at Roundabouts: An Investigation in a Mixed-Traffic Environment". In: *15th IFAC Symposium on Control in Transportation Systems* (June 2018), pp. 73–78. DOI : 10.1016/$j.ifacol$.2018.07.013.

[9] Keke Long et al. "Optimization based trajectory planner for multilane roundabouts with connected automation". In: *Journal of Intelligent Transportation Systems* (Mar. 2022). DOI : 10.1080/15472450.2022.2046473.

[10] Konstantinos Katsaros et al. "Performance study of a Green Light Optimized Speed Advisory (GLOSA) Application Using an Integrated Cooperative ITS Simulation Platform". In: *International Wireless Communications and Mobile Computing Conference (IWCMC)* (July 2011). DOI : 10.1109/$IWCMC$.2011.5982524.

[11] Anna-Lena Schlamp et al. "User-Centric Green Light Optimized Speed Advisory with Reinforcement Learning". In: *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)* (Sept. 2023), pp. 3463–3470. DOI : 10.1109/$ITSC$57777.2023.10422501.

[12] Marie-Ange Lebre et al. "Real scenario and simulations on GLOSA traffic light system for reduced CO2 emissions, waiting time and travel time". In: *22nd ITS World Congress* (Oct. 2015).

[13] Tessa Tielert et al. "The impact of traffic-light-to-vehicle communication on fuel consumption and emissions". In: *2010 Internet of Things (IOT)* (Dec. 2010), pp. 1–8. DOI : 10.1109/IOT.2010.5678454.

[14] Vandung Nguyen et al. "An Efficient and Reliable Green Light Optimal Speed Advisory system for Autonomous Cars". In: *2016 18th Asia-Pacific Network Operations and Management Symposium (APNOMS)* (Oct. 2016), pp. 1–4. DOI : 10.1109/$APNOMS$.2016.7737260.

[15] Martin L. Puterman. *Markov Decision Processes*. Vol. 2. Handbooks in Operations Research and Management Science, 1990. Chap. 8, pp. 331–434.

[16] Volodymyr Mnih et al. "Asynchronous Methods for Deep Reinforcement Learning". In: *International Conference on Machine Learning*. 2016.

[17] Marius Wegener et al. "Automated eco-driving in urban scenarios using deep reinforcement learning". In: *Transportation Research Part C Emerging Technologies* Vol. 126 (May 2021). DOI : 10.1016/$j.trc$.2021.102967.

[18] Bo Liu et al. "Adaptive Speed Planning of Connected and Automated Vehicles Using Multi-Light Trained Deep Reinforcement Learning". In: *IEEE Transactions on Vehicular Technology* Vol. 71 (Apr. 2022), pp. 3533–3546. DOI : 10.1109/TVT.2021.3134372.

[19] Xiaoliang Zhang et al. "Eco-driving for Intelligent Electric Vehicles at Signalized Intersection: A Proximal Policy Optimization Approach". In: *6th International Conference on Information Science, Computer Technology and Transportation (ISCTT)* (2021), pp. 1–7.

[20] Greg Brockman et al. "OpenAI Gym". *In: ArXiv* (2016).

[21] Antonin Raffin et al. "Stable-Baselines3: Reliable Reinforcement Learning Implementations". In: *Journal of Machine Learning Research* Vol. 22.268 (2021), pp. 1–8.

[22] Michael Behrisch et al. "SUMO- Simulation of Urban MObility: An Overview". In: *The Third International Conference on Advances in System Simulation* (2011).

[23] Stefan Krauß, Peter Wagner, and Christian Gawron. "Metastable states in a microscopic model of traffic flow". In: *Physical Review E* Vol. 55.5 (May 1997), pp. 5597–5602. DOI : 10.1103/$PHYSREVE$.55.5597.

[24] Bußgeldkatalog 2024 (Thomas R.) *Sicherheitsabstand: Welchen Abstand schreibt die StVO vor?* URL: https://www.bussgeldkatalog.org/sicherheitsabstand/ (visited on 12/11/2023)