Cooperative Adaptive Cruise Control Based on Reinforcement Learning for Heavy-Duty BEVs

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 September 2023, accepted 8 November 2023, date of publication 9 November 2023,
date of current version 16 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3331827

RESEARCH ARTICLE

Cooperative Adaptive Cruise Control Based on Reinforcement Learning for Heavy-Duty BEVs

MATTEO ACQUARONE, (Member, IEEE), FEDERICO MIRETTI, (Member, IEEE), DANIELA MISUL, AND LUCA SASSARA

Dipartimento Energia "Galileo Ferraris," Center for Automotive Research and Sustainable Mobility (CARS@Polito), Politecnico di Torino, 10129 Turin, Italy

Corresponding author: Federico Miretti (federico.miretti@polito.it)

ABSTRACT This study proposes a novel approach for cooperative adaptive cruise control (CACC) based on the twin delayed deep deterministic policy gradient algorithm (TD3) for heavy duty battery electric vehicles (BEVs). CACC is an advanced driver assistance systems (ADAS) that exploits vehicle connectivity to bring new advantages to cruise control technologies. The TD3 algorithm, which is a deep reinforcement learning (DRL) algorithm, was selected because it is currently at the forefront of the state of the art for problems with continuous states and actions. Furthermore, compared to state-of-the-art techniques, such as linear MPC, a DRL approach is more effective in dealing with highly nonlinear objectives. This enables us to explicitly model the effect of air drag reduction in the ego vehicle, which positively affects energy savings. The air drag reduction characteristic was modeled through experimental data from a previous work. At the same time, driving comfort was also optimized with respect to the reference driving cycle, chosen as the HHDDT driving cycle. Three different types of spacing strategies have been investigated that involve minimum time headway and time-to-collision (TTC) to study the safety guarantee of the algorithm, particularly when facing critical and unexpected situations such as sudden hard braking. The results achieved show how the Ego vehicle can reduce energy consumption by up to 19.8% without the comfort worsening with respect to the preceding vehicle, still guaranteeing safe driving conditions, when considering the spacing strategy based only on TTC, developed to obtain the highest air drag reduction.

INDEX TERMS Cooperative adaptive cruise control, reinforcement learning, TD3 algorithm, heavy-duty vehicle, air drag reduction, BEV.

I. INTRODUCTION

Advanced driver assistance systems (ADAS) are playing an increasingly important role in supporting the driver to create safer and more efficient driving conditions. Among all ADAS, adaptive cruise control (ACC) is a system that provides consistent aid, especially in highway mobility, guaranteeing safety by minimizing the possible risk of collision due to variations in the speed of the vehicle in front, automatically adjusting the vehicle velocity and maintaining the correct spacing. Theoretically, this type of system also makes it possible to optimize road throughput, increasing its capacity and reducing traffic congestion. However, it was found in practice that the current generation of ACC systems does not guarantee the so-called string stability of a vehicle platoon and can therefore lead to an actual decrease in traffic capacity [1].

To overcome these issues, new *cooperative* adaptive cruise control (CACC) systems are being proposed that exploit vehicle-to-vehicle (V2V) connectivity, which can provide additional safety and robustness guarantees and introduce the possibility of concretely improving traffic flow stability [2], [3], [4], [5].

A. REVIEW OF EXISTING ACC AND CACC SYSTEMS

In its simplest version, an ACC aims to maintain a constant temporal gap between the two vehicles, using common PID or LQ controllers [6], [7]. However, most of the ACC and CACC solutions, such as [8], [9], [10], and [9] are based

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

on model predictive control (MPC), which uses a dynamic model of the system to calculate the optimal control signal through an online optimization process. MPC is widely used for automotive control systems because of its reliability and solid theoretical foundations and because of the possibility to explicitly deal with hard constraints. Recently, particular attention has been paid to nonlinear MPC as an effective control method for ACC/CACC problems [11], [12], since it allows one to directly consider also energy-saving features, which depend on strong nonlinear dynamics.

On the other hand, an alternative approach that is rapidly gaining traction in control applications is reinforcement learning (RL) [13]. RL, and more specifically deep reinforcement learning (DRL), a combination of RL with deep learning, can also be particularly suitable for dealing with continuous control problems that may be difficult to model [14], [15]. It has been demonstrated how DRL can achieve comparable or enhanced performance with respect to more common optimal control strategies regarding an ACC problem, especially in terms of computational costs and in the presence of high-dimensional and uncertain environments [16].

Recently, several ACC solutions that exploit a RL-based control strategy have been proposed: for example, [17] proposed a solution aimed at improving comfort and safety using a deep deterministic policy gradient (DDPG) algorithm [18], a widely used RL method in many ACC solutions due to the possibility of considering a continuous action space. Another relevant study [19] proposed a supervised actor–critic approach, combining the benefits of supervised learning with RL.

Similarly, several CACC systems that use V2V communication and at the same time exploit control strategies based on RL have been introduced. For example, [20] proposed a CACC solution using a control strategy based on the policy gradient algorithm, while [21] developed an RL-based car-following strategy with the aim of damping possible traffic oscillations and improving energy consumption. A CACC based on supervised RL was designed and validated in [22], while the solution proposed by [23] takes advantage of a model-based DRL control strategy. Other solutions are instead based on controlling entire platoons of several vehicles, resulting in more complex multiagent strategies [24], [25].

B. RESEARCH GAP AND PROPOSED SOLUTION

In any case, most ACC and CACC solutions are designed for standard light-duty vehicles rather than heavy-duty vehicles (HDVs). However, HDVs make up a considerable share of highway traffic, as they are indispensable for regional and long-distance freight transport. According to [26], energy demand from heavy-duty trucking is expected to increase by more than 50% between 2000 and 2040.

Although electrification may contribute to mitigate the GHG intensity of freight transport, this sector will remain in the near future very challenging to electrify and there is a growing need to explore alternative solutions that can reduce unnecessary energy consumption. To this end, one promising solution is platooning [27], [28], [29], a technique that allows to reduce energy consumption by exploiting the reduction in aerodynamic drag resulting from cruising at a reduced intervehicular distance. In practice, platooning can be achieved using ACC or CACC systems and leads to improved fuel economy. Many works have already demonstrated the potential benefits of platooning for heavy-duty vehicles [30], [31], [32]. The solutions proposed by [33], [34], and [35] are all examples of CACC in which the reduction in air drag is explicitly considered for energy savings purposes.

However, there is a poor availability of solutions that also factor in passenger comfort in conjunction with the energy-saving benefits of platooning, which could be detrimental to user experience. Furthermore, very few CACC solutions for heavy duty vehicles are already available that take advantage of a DRL control approach [36], [37]. For these reasons, in this work, a novel CACC solution for heavy-duty vehicles based on a DRL control strategy is proposed, simultaneously focusing on energy savings and comfort improvement. As shown later in this article, this is particularly relevant as these two are conflicting objectives; therefore, if passenger comfort is not explicitly taken into account in its development, a CACC system can lead to unacceptable behavior.

Clearly, designing such a CACC system poses a complex multi-objective problem driven by non-linear dynamics. For this reason, we propose a novel implementation based on the recently introduced twin delayed deep deterministic policy gradient (TD3) algorithm [38], instead of the more commonly used deep deterministic policy gradient algorithm (DDPG). To test our solution, we developed a case study simulating a platoon of two heavy-duty vehicles, with equal characteristics, considering highway scenarios.

Two notable aspects of our proposed solution are the multi-objective reward function that was developed for the RL agent and the employment of an experimental air drag reduction characteristic. The proposed reward function explicitly accounts for energy saving and comfort and also employs a two-term spacing policy based on time headway and time to collision. As discussed in Sec. VI, this spacing policy proved to be superior in our simulation results with respect to simpler spacing policies. Another important aspect is in the inclusion of a term associated to the air drag reduction obtainable as a function of intervehicular distance, which is a novel solution in RL-based ACC systems. This additional term improves the ability of the agent to achieve energy efficient operation.

Summarizing, the CACC system proposed in this work aims to address the previously mentioned research gap with respect to coordinated optimization of energy saving and passenger comfort by using an advanced RL algorithm (TD3) that leverages an explicit non-linear model of the achievable

air drag reduction and V2V connectivity and a spacing policy based on time headway and time to collision.

II. VEHICLE MODEL

The vehicle model considered in this work is a quasi-static, backward-facing powertrain simulation model for a heavy-duty battery electric vehicle (BEV). The discrete longitudinal dynamics equation of the vehicle is based on Newton's second law:

$$\begin{cases} s(k+1) = s(k) + v(k)\delta t \\ v(k+1) = v(k) + a(k)\delta t \\ a(k) = \dfrac{1}{m_v}\left(F_{traction}(k) - F_{aero}(k) - F_{roll}(k)\right) \end{cases} \tag{1}$$

where s, v and a are respectively the travelled distance, the velocity and the acceleration of the vehicle, m_v is the vehicle mass and δt is the discrete time interval. $F_{roll} = c_{roll} m_v g$ is the simplified rolling resistance force, with c_{roll} corresponding to the rolling coefficient and g to gravity acceleration, $F_{aero}(k) = \frac{1}{2}\rho_{air} c_d(k) m_v A v(k)^2$ is the aerodynamic force to which the vehicle is subject, where ρ_{air} is the air density, c_d is the air drag coefficient and A is the frontal area of the vehicle, and $F_{traction}(k)$ is the traction force at the wheels. The vehicle is powered by a 372 kW electric motor, which characteristic are reported in Table 1, together with the other vehicle's main parameters.

Combining Eq. (1) with the dynamic equation of the motor, the relation that associate the acceleration of the vehicle to the motor torque $T_m(k)$ can be expressed as:

$$T_m(k) = \frac{(F_{roll} + F_{aero}(k))r_{wheel} + I_{m,eq}\dfrac{a(k)}{r_{wheel}}}{\tau_{diff}} \tag{2}$$

where $I_{m,eq}$ is the equivalent inertia at the motor output, r_{wheel} is the wheel radius and τ_{diff} is the differential ratio.
The total power requested to the battery is therefore the sum of three terms:

$$P_{batt}(k) = P_m(k) + P_{loss}(k) + P_{aux} \tag{3}$$

where the motor power is calculated from the motor torque net of the gear efficiency (η_{gear}) and the final drive efficiency (η_{fd}). The power related to motor losses $P_{loss}(k)$ due to friction, hysteresis, and parasitic effects depends instead on the actual motor efficiency $\eta_m(k)$, while P_{aux} is the average power absorbed by the auxiliary devices.

The motor is supplied by a 346 kWh battery, modeled using an equivalent Thevenin circuit for which the Open Circuit Voltage $V_{OC}(k)$ and the equivalent internal resistance $R_{int}(k)$ values depends on the actual State of Charge $SOC(k)$ level. The SOC variation at each time step is calculated including also the battery nominal capacity Q_{nom} following Eq. (4):

$$SOC(k+1) = SOC(k) - \frac{V_{OC}(k) - \sqrt{V_{OC}(k)^2 - 4R_{int}(k)P_{batt}(k)}}{2R_{int}(k)Q_{nom}} \tag{4}$$

TABLE 1. Vehicle parameters.

Parameter	Value
Vehicle's average mass:	$m_v = 12864$ kg
Vehicle's frontal area:	$A = 8.9$ m^2
Vehicle's longitudinal length:	$l_v = 9$ m
Wheel radius:	$r_{wheel} = 0.5715$ m
Rolling coefficient:	$c_{roll} = 0.005$
Gear ratio:	$\tau_{gear} = 1$
Differential ratio:	$\tau_{diff} = 19.74$
Equivalent inertia at the motor:	$I_{m,eq} = 4200$ kg m^2
Gear efficiency:	$\eta_{gear} = 0.95$
Final drive efficiency:	$\eta_{fd} = 1$
Undisturbed air drag coefficient:	$c_{d,inf} = 0.57$
Air density:	$\rho_{air} = 1.2$ kg m^{-3}
Motor power:	371 kW
Motor maximum torque:	150 Nm
Motor maximum angular velocity:	1500 rad/s
Average auxiliary power:	$P_{aux} = 500$ W
Battery total energy:	$E_{tot} = 346$ kWh
Battery nominal capacity:	$Q_{nom} = 693$ Ah
Maximum SOC:	$SOC_{max} = 95\%$
Minimum SOC:	$SOC_{min} = 20\%$

The main battery parameters are reported in Table 1. For the sake of clarity, it is important to highlight that the vehicle dynamics alone is sufficient to calculate the correct action through the RL algorithm, while the e-machine and battery models are only used to assess the energy efficiency of the developed CACC.

A. SAFETY PARAMETERS

The RL agent, which represents the follower (or *ego*) vehicle, is trained with the aim to mainly reduce energy consumption by minimizing the intervehicular gap, thus reducing the air drag coefficient, and at the same time improve driving comfort, a feature commonly associated with vehicle acceleration and jerk [39], with respect to the leading truck, also thanks to the consideration of a variable time headway in the control strategy.

In the CACC framework, the ego vehicle must respect safety conditions that are strictly related to two main kinematic quantities, the time headway and the time to collision (TTC), as defined in Eq. (5) and Eq. (6) respectively.

$$h(k) = \frac{s_{Lead}(k) - s_{Ego}(k) - l_v}{v_{Ego}(k)} \tag{5}$$

$$TTC(k) = \frac{s_{Lead}(k) - s_{Ego}(k) - l_v}{v_{Ego}(k) - v_{Lead}(k)} \tag{6}$$

where s_{Lead}, s_{Ego}, v_{Lead}, v_{Ego} are respectively the travelled distances and the longitudinal velocities of the leading and ego vehicles, while l_v is the length of the leading truck. The time headway simply corresponds to the temporal distance between the two vehicles, while the TTC is a measure of the time it will take for two vehicles to collide, given their current positions and velocities [40]. TTC explicitly quantifies the risk of rear-end collision at a certain time, and is commonly considered as a traffic safety indicator, including possible

related variants [41], [42]. Both time headway and TTC have been taken into account in different ways, resulting in three different adopted strategies.

V2V communication has been exploited instead with the objective of obtaining more precise values of the leader's velocity and acceleration in fast time, allowing the ego truck to act promptly to their possible variations and thus to follow more closely the leader, also improving string stability [2], [43]. Anyway, for simplicity, in this work the communication between the two vehicles is considered ideal, without taking into consideration possible communication delays or communication losses that may be present in real case scenarios.

B. VARIABLE AIR DRAG COEFFICIENT

Heavy-duty vehicles have a larger frontal area compared to commercial light duty vehicles and are consequently subject to much larger aerodynamic forces. For this reason, they are particularly sensitive to the trial effect that arises from the presence of the preceding lead vehicle in an adaptive cruise control problem. The impact of reduced air resistance can be taken into account considering a variable air drag coefficient, depending on the intervehicular gap [44]: intuitively, the air drag coefficient decreases with distance, and consequently the aerodynamic forces are lower. Therefore, a reduced gap leads the ego vehicle to be subject to a lower resistance force.

Several relations between the air drag coefficient and the intervehicular gap have been investigated in previous works regarding heavy-duty vehicles [28]. In the proposed solution, the relation between c_d and the distance between leader and ego vehicle is based on [31] and it has been obtained experimentally:

$$\begin{cases} c_d(k) = c_{d,inf} \dfrac{a_3 g^3 + a_2 g^2 + a_1 g + a_0}{b_3 g^3 + b_2 g^2 + b_1 g + b_0} & \text{if } g < G_0 \\ c_d(k) = c_{d,inf} & \text{otherwise} \end{cases} \tag{7}$$

a_i, b_i and G_0 are parameters calculated specifically for an heavy-duty vehicle, while ρ_{air} is the air density and $g(k) = s_{Lead}(k) - s_{Ego}(k) - l_v$ is the intervehicular distance, corresponding to the space between the front end of the ego vehicle and the rear end of the Lead vehicle. The relationship between $c_d(k)$ and $g(k)$ is shown graphically in Fig. 1.

The above relation has been calculated for a reference velocity of 100 km/h, and for this reason it can be considered only for highway scenarios. Anyway, the ego vehicle is much less affected by the aerodynamic influence if the velocity is low. Hence, the air drag coefficient reduction has a negligible impact when considering urban driving cycles.

III. REINFORCEMENT LEARNING AND ALGORITHM SETTINGS

A. SELECTION OF THE ALGORITHM

Reinforcement learning (RL) has recently gained increasing importance in control problems characterized by complex and difficult-to-model systems. In the RL framework, an *agent* is trained to select actions based on a set of observations which are used to define the current state of the system [45], [46]. After choosing an action, the agent also receives a *reward*; when deployed, the agent attempts to maximize both the instantaneous reward and future rewards.

Many ACC solutions based on RL ([17], [47]) use a particular model-free and off-policy method based on policy gradient that exploit the actor-critic architecture, called *deep deterministic policy gradient* (DDPG) introduced by [18]. DDPG has become one of the most popular RL choice in this kind of control applications due to the possibility to work in both continuous action and state space. Another advantage of DDPG is that it works well in a noisy environment: indeed, a noisy environment is explicitly wanted for exploration making the algorithm off-policy. DDPG is based also on the use of target networks and experience replay, two techniques already introduced in DQN, that allow, respectively, to make the learning process more stable and avoid correlations between samples, basing the learning process on independent data.

The proposed solution is based on a particular variant of DDPG, called *twin delayed deep deterministic policy gradient* (TD3), introduced by [38]. TD3 algorithm has basically the same structure and characteristics as DDPG, but it also considers some few additional features that address the issue of overestimation bias, typical of deep Q-learning and also present in DDPG, which can lead to sub-optimal policies. In particular, the algorithm tries to minimize the overestimation effect that may arise from the recursive formulation of the updating step using a double critic. The two critics estimate both Q values, but only the lower of them is used in the subsequent update step, mitigating the problem. TD3 also faces the problem of high variance that may slow the learning process [13], due to a policy regularization due to the addition of a clipped noise to the target action, and a less frequent policy update, reducing the accumulating error, thus improving the stability and the performance of the algorithm. Considering its peculiarities, the TD3 algorithm is particularly suitable for a CACC problem, since it takes advantage of all the benefits of the DDPG algorithm but at the same time addresses the common problems of overestimation and high variance.

B. ALGORITHM SETUP

The states considered in this work are the lead vehicle acceleration $a_{Lead}(k)$ and speed $v_{Lead}(k)$, the speed of the ego vehicle $v_{Ego}(k)$ and the time headway $h(k)$. All these quantities are related to the objectives of CACC, which are the vehicle's acceleration and jerk, the TTC, the time headway, and the air drag coefficient. The control action calculated by the algorithm is instead associated with the acceleration of the ego vehicle, similarly to many other CACC solutions ([17], [21]), and is necessarily scaled in order to match the chosen acceleration physical range.

Regarding the characteristics of the actor/critic network and the training process, the settings were mostly the same as considered in the original algorithm evaluation test [38].

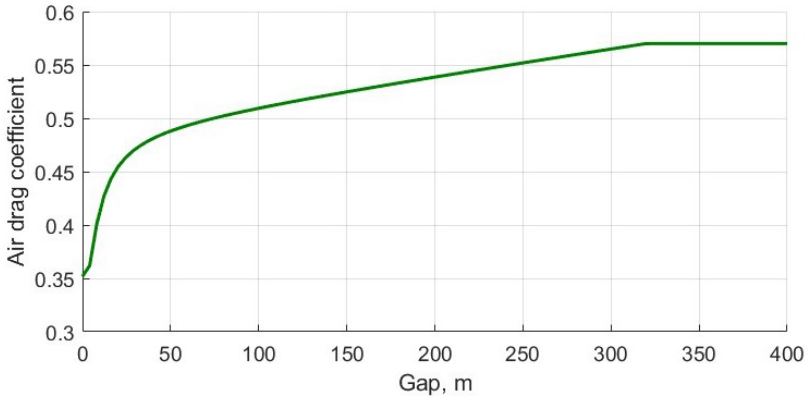

FIGURE 1. Drag coefficient as function of the gap.

TABLE 2. Algorithm parameters.

Parameter	Value
# hidden layers for the actor/critic networks:	2
Hidden layers size for the actor/critic networks:	64 nodes
Target networks delay:	$\tau = 0.005$
Discount factor:	$\gamma = 0.99$
Learning starts:	1000 steps
Learning rate:	10^{-4}
Exploration noise:	$\mathcal{N}(0, 0.1)$
Target policy smoothing noise:	$\mathcal{N}(0, 0.2)$ clipped to $(-0.5, 0.5)$
Replay buffer size:	500000 transitions
Mini-batch size:	32 transitions
Actor's update frequency:	$d = 2$ steps

The only differences are in the size of the actor and critic networks, in the replay buffer and mini-batch sizes, and in the learning rate value, which has been lowered from the moment that it showed more stable convergence results for this specific type of problem. The algorithm parameters are summarized in Table 2.

IV. REWARD FUNCTION DEVELOPMENT

As discussed in Sec. III, the formulation of the reward function plays a fundamental role in meeting the identified controller objectives and constraints. In particular, for cruise control applications, the reward function must:

- Implement a spacing policy, which is an expression of the desired spacing between two consecutive vehicles at steady state operation [48].
- Favor desirable behavior in terms of other control objectives, such as comfort and energy savings.
- Strongly penalize unacceptable behavior in terms of safety, which are to be considered as hard constraints.

With respect to these requirements, RL applications have so far focused on formulating reward functions based on a desired IVD or time headway [20], [49], and then possibly terms related to velocity, acceleration, and/or jerk to account for comfort [16], [21], [22], [24]. Some ([17]) authors have also considered time-to-collision as a safety threshold.

In contrast, our approach introduces a reward function that directly considers features related to comfort and energy efficiency (jerk, acceleration, and air drag) as well as a spacing policy including both time headway and time-to-collision. Furthermore, the term associated with air drag is evaluated by explicitly modeling the air drag coefficient as a function of the IVD, as described in Sec. II-B, which is based on experimental data.

More in detail, the proposed reward function consists in a partial weighted sum of terms having the general formulation:

$$r_{tot}(k) = \frac{w_{jerk} r_{jerk}(k) + w_{acc} r_{acc}(k) + w_{drag} r_{drag}(k)}{w_{jerk} + w_{acc} + w_{drag}} + \\ + r_h(k) + r_{TTC}(k) \tag{8}$$

where $r_{jerk}(k)$ and $r_{acc}(k)$ are terms related to jerk and acceleration, two quantities typically associated to comfort, while $r_{drag}(k)$ is a air drag reduction term that represent the energy-saving feature. These three terms can assume values between $[-1;\ \ 1]$ and are weighted and then normalized using three corresponding weighting factors, which gives the possibility of considering a trade-off between comfort and reduction in energy consumption, the two main objectives of the proposed solution.

While the first part of the equation is devoted to comfort and energy-saving features, the last two terms of the reward function, $r_h(k) \in [-1;\ \ 0]$ and $r_{TTC}(k) \in [-1;\ \ 0]$, related, respectively, to the time headway and TTC, are essentially used as soft constraints to ensure safety conditions in different ways, depending on the adopted strategy. Differently from the other terms, they never give positive rewards to the agent, since their aim is to act only when the ego vehicle approaches a possible episode failure and, in all other cases, not interfere with the other reward terms.

A. TIME HEADWAY PENALTY TERM

The time headway penalty term $r_h(k)$ is necessary to maintain the temporal gap between the two vehicles in a certain reasonable range. Instead of defining a fixed desired target for the time headway, in this work the time headway is left free to vary between two bounds. Possible advantages of a variable time headway considering heavy-duty vehicles have been discussed in [50]. In this work, the consideration of a variable time headway has the main purpose to give the agent more freedom in satisfying the comfort and air drag reduction goals.

Specifically, if $h(k)$ is kept between the two desired values, a null penalty is applied. Otherwise, if these limits are exceeded, a linear penalty is considered according to equation (9):

$$\begin{cases} r_h(k) = -\dfrac{h(k) - h_{low,des}}{h_{min} - h_{low,des}} & \text{if } h(k) < h_{low,des} \\ r_h(k) = -\dfrac{h(k) - h_{high,des}}{h_{max} - h_{high,des}} & \text{if } h(k) > h_{high,des} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where $h_{max,des}$ is the time headway desired upper bound, considered always equal to 2 s; $h_{min,des}$ is the time headway desired lower bound, and can assume different values depending on the specific strategy; $h_{min} = 0$ s is the minimum time headway limit, which corresponds to a crash with the Lead vehicle; $h_{min} = 4$ s is the maximum time headway limit that, if exceeded, it can be deemed that the ego vehicle has lost the leader.

The maximum and minimum limits are instead related to a failure of the algorithm: if these limits are reached, a severe overall penalty of $r_{tot} = -100$ is given to the agent strongly discouraging it to somehow approach this headway values in the future.

B. TIME-TO-COLLISION PENALTY TERM

The time-to-collision is a quantity strictly related to safety, as it quantifies the collision risk between two vehicles at a specific time instant. For this reason, the additional penalty term $r_{TTC}(k)$ based on TTC has been introduced in the total reward definition. Some ACC solutions based on RL that considers the TTC in the reward function uses a logarithmic variability ([17], [47]) below a certain safety threshold under which the risk of collision is considered high. This limit is usually chosen equal to 4 seconds [51]. However, in [52] it is highlighted how a limit of 4 seconds may lead to false alarms, suggesting a threshold of 3 seconds. Basing on these considerations, a linear penalty has been introduced, according to the following equation:

$$\begin{cases} r_{TTC}(k) = 0 & \text{if } TTC(k) \geq TTC_{lim} \\ r_{TTC}(k) = -1 & \text{if } TTC(k) \leq TTC_{lim,low} \\ r_{TTC}(k) = \dfrac{TTC(k) - TTC_{low}}{TTC_{lim} - TTC_{lim,low}} & \text{otherwise} \end{cases} \tag{10}$$

where $TTC_{lim} = 4$ s and $TTC_{lim,low} = 3$ s. This type of variability proved to discourage the ego vehicle to exceed the TTC threshold maintaining safe driving conditions, and, differently from a logarithmic penalty, it allows to severely penalize the agent even before the TTC reaches values close to zero.

C. JERK REWARD TERM

The acceleration rate of change is an important quantity that is directly related to driver and passengers' comfort. Reduce the jerk can lead to more comfortable driving conditions [53], and for this reason, the reward term $r_{jerk}(k)$ is introduced with the aim to encourage the agent to provide, when possible, smooth acceleration profiles. The reward term has the following expression:

$$\begin{cases} r_{jerk}(k) = 1 & \text{if } |j(k)| \leq j_{min} \\ r_{jerk}(k) = 1 - 2\dfrac{|j(k)| - j_{min}}{j_{max} - j_{min}} & \text{if } j_{min} < |j(k)| < j_{max} \\ r_{jerk}(k) = -1 & \text{if } |j(k)| \geq j_{max} \end{cases} \tag{11}$$

where $j_{min} = 1$ m/s^3 is the limit under which the jerk, in absolute value, is considered optimal in terms of comfort [54], while $j_{max} = 10$ m/s^3 is a maximum limit that should be never overcome, if possible.

D. ACCELERATION REWARD TERM

Also the acceleration of the ego vehicle, similarly to its rate of change, can be associated to comfort features. Unnecessary oscillations of the acceleration, especially with high magnitude, can lead to uncomfortable rides [39] if compared to driving conditions with slow changes in the velocity profile. For this reason, the agent is rewarded with positive values when, if possible, provides low acceleration values. The acceleration reward term $r_{acc}(k)$ significantly influences the ego vehicle behavior depending on its formulation and, for this reason, several attempt has been done investigating the best reward definition. While a linear reward definition makes the agent more sensible also to accelerations with very small magnitude, leading to a more unstable training process, a quadratic formulation [55] proved to be more tolerant with a relatively wide range of small acceleration values but at the same time more stringent with medium and high acceleration peaks. Consequently, the chosen reward formulation has a quadratic variability:

$$\begin{cases} r_{acc}(k) = 1 - 2\left(\dfrac{a(k)}{a_{max}(k)}\right)^2 & \text{if } a(k) < a_{max}(k) \\ r_{acc}(k) = -1 & \text{if } a(k) \geq a_{max}(k) \end{cases} \tag{12}$$

where a_{max} is chosen as the minimum value between the peak acceleration of the reference highway driving cycle (0.8 m/s^2) and the maximum acceleration that the vehicle can produce depending on the motor power at each time step.

E. AIR DRAG COEFFICIENT REDUCTION REWARD TERM

Finally, the reward term $r_{drag}(k)$ is introduced to achieve energy savings thanks to the air drag reduction. This reward term indirectly encourages the agent to follow the leading vehicle with the smallest gap possible, maximizing the air drag coefficient reduction, which is modeled as a function of IVD. The chosen reward formulation is based on a linear variability between two air drag coefficients limits:

$$\begin{cases} r_{drag}(k) = 1 & \text{if } \bar{c}_d \leq \bar{c}_{d,min} \\ r_{drag}(k) = -1 & \text{if } \bar{c}_d \geq \bar{c}_{d,max} \\ r_{drag}(k) = 1 - 2\left(\dfrac{\bar{c}_d - \bar{c}_{d,min}}{\bar{c}_{d,max} - \bar{c}_{d,min}}\right) & \text{otherwise} \end{cases} \tag{13}$$

where $\bar{c}_d$ is the air drag coefficient normalized with respect to the undisturbed one ($c_{d,inf}$). $\bar{c}_{d,min}$ and $\bar{c}_{d,max}$ are the normalized air drag coefficients calculated according to Eq. (7), considering the mean gaps at time headways $h_{low,des}$ and $h_{high,des}$.

V. ADOPTED STRATEGIES

In this work, three different spacing strategies have been investigated involving minimum time headway and TTC as safety parameters with the aim of showing the different related behaviors and the entity of reduction of energy consumption.

A. STRATEGY BASED ON MINIMUM TIME HEADWAY ONLY (H STRATEGY)

The first solution considered is based only on minimum time headway as safety parameter, chosen equal to 0.5 seconds. Hence, the time headway is let free to vary between the two desired values equal to $h_{min,des} = 0.5\ s$ and $h_{high,des} = 2\ s$, as also considered in [56]. At this stage, the TTC is not used in the reward function to ensure safety.

B. STRATEGY BASED ON TIME-TO-COLLISION ONLY (TTC STRATEGY)

As already discussed before, the TTC, differently from time headway, quantifies at every time instant the potential collision risk between the two vehicles, a peculiarity that makes it a perfect candidate for a penalty term that aims at safety. In the second spacing solution, the TTC penalty term is introduced in the reward expression as a safety feature, but at the same time it is not considered any lower bound for the time headway. The objective is to minimize as much as possible the intervehicular gap in order to strongly reduce the energy consumption, but at the same time guaranteeing safety conditions thanks to the presence of the TTC penalty term.

C. STRATEGY BASED ON BOTH MINIMUM TIME HEADWAY AND TIME-TO-COLLISION (H-TTC STRATEGY)

The use of TTC as a safety parameter without any lower bound on the time headway leads occasionally to very close gap between the vehicles. It has been proven by [57] that in case of communication loss the effectiveness of cooperation drops and the distance that ensures a secure following in a CACC problem inevitably increases. Moreover, [58] investigated the relation between time headway and TTC and has highlighted how the use of time headway can help to prevent the approaching to critical TTC values. For these reasons, a third approach based on the combination of minimum time headway and TTC allows to consider a double check regarding safety, increasing the safety guarantees. The minimum time headway limit has been reduced to $h_{min,des} = 0.25\ s$, thanks to the presence of the TTC as an additional safety parameter, with the objective of utilizing a reduced gap for energy-saving purposes. The consideration of both minimum time headway and TTC make this approach the best candidate for a possible implementation among the presented spacing strategies, taking into account safety guarantees, energy-saving performance, and the possible presence of real-case issues like communication delays.

VI. RESULTS

A. SIMULATION SETUP

The CACC problem has been discretized in 0.1 s: at each time step, the physical information regarding the ego vehicle is calculated from the action provided by the TD3 algorithm. The ego vehicle starts each episode with an instantaneous velocity equal to that of the leading vehicle, and with an initial time headway of 1 s. Moreover, both vehicles are supposed to start an episode with a SOC equal to 80%. The ego truck is trained until the cumulative reward reaches convergence, which is considered achieved when it does not vary over 5% for at least 100 episodes.

The vehicle is subject to some physical limitations that must be respected in all simulations. More specifically, it cannot exceed a maximum velocity equal to $v_{max} = 30\,m/s$, while it is not allowed to reach negative velocity values, to restrict the problem to a classical car-following problem. Acceleration has a limited asymmetric range $-3\ m/s^2 \leq a(k) \leq 2\ m/s^2$, cautiously below performance criteria [59].

The velocity profile of the preceding vehicle is chosen as the cruise portion of a standard driving cycle, the Heavy Heavy-Duty Diesel Truck (HHDDT) cycle [60], with the aim of obtaining replicable and comparable results. At first, a restricted section of the HHDDT cycle corresponding to its first 400 seconds has been used in order to investigate clearly the behaviors and benefits of each strategy. Then, the three strategies have also been tested on the full HHDDT cycle to investigate their behavior during a complete driving mission. Finally, the algorithm has also been tested on a modified small portion of the reference driving cycle, including a hard braking that corresponds to a maximum permitted constant deceleration of $-3\ m/s^2$ applied for 4.5 s, with the objective of investigating its approach to critical situations.

B. GENERAL OBSERVATIONS

The consideration of the air drag reduction as an energy-saving feature in the reward structure leads the ego vehicle to reduce the intervehicular distance with the aim to exploit

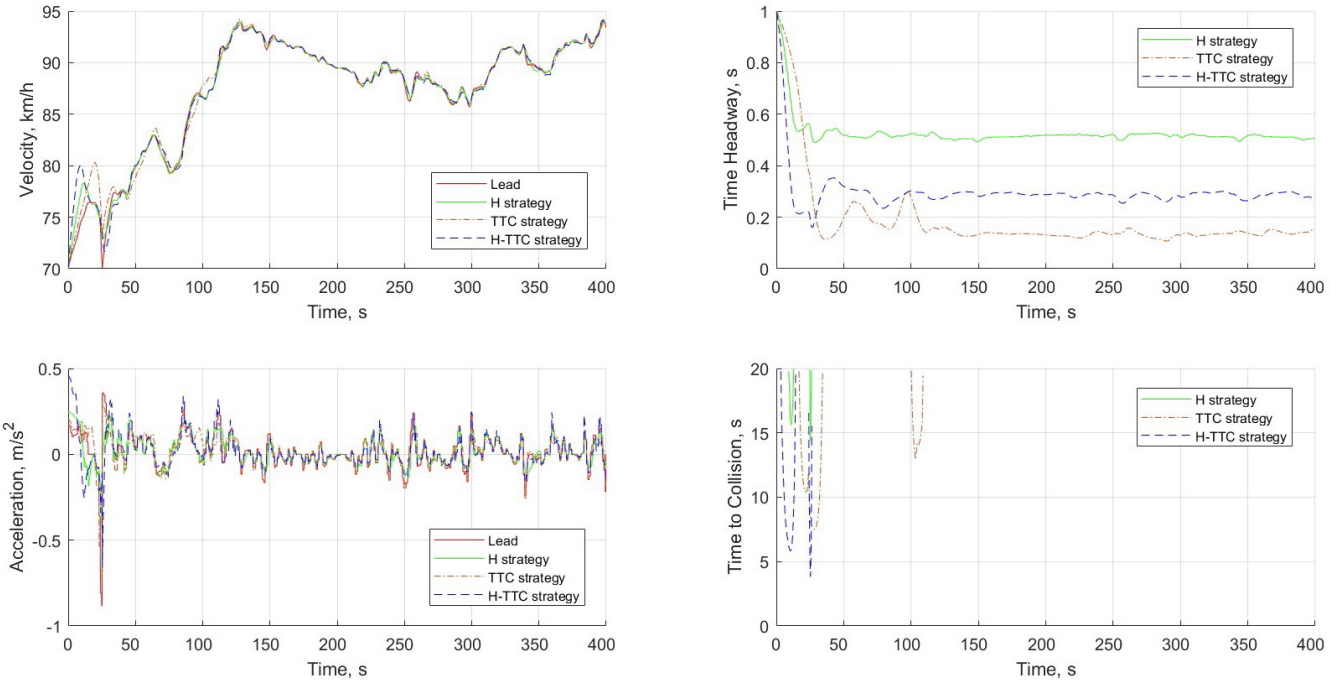

FIGURE 2. Results for the portion of HHDDT driving cycle.

the aerodynamic effect of the vehicle in front. The ego vehicle has the tendency to close the gap with the leader until it is allowed by the safety parameters defined in the adopted spacing strategy. All three strategies respect the safety guarantees for the whole driving cycle, without exceeding the safety limitations, also when considering the full HHDDT driving cycle. For all the three different spacing strategies, the proposed solution also proved to reach good results in improving at the same time the driving comfort with respect to the leader's driving cycle, reducing the acceleration and jerk signals thanks to the consideration of a variable time headway, which provides an additional degree of freedom in meeting the comfort goals. The entity of the acceleration and jerk reductions in RMS terms for the three spacing strategies, together with their relative obtained SOC savings, are resumed in Table 3. On the other hand, the behaviors of each strategy considering the portion of the HHDDT driving cycle are shown in Fig. 2, while the results of all the different strategies considering the full reference driving cycle are reported in Table 3.

C. H STRATEGY

Regarding the H strategy, since a minimum time headway is given as a safety parameter, the ego truck tends to quickly get closer to the Lead vehicle and then to correctly maintain the time headway toward the defined minimum desired limit of 0.5 s, as can be seen in the bottom right graph in Fig. 2. The H strategy results to be the most conservative approach, with a fairly high TTC for almost the majority of the driving mission, as can be seen in the bottom right graphic in Fig. 3, even if it is not directly considered in

TABLE 3. Performance of the three strategies. Acceleration and jerk here refer to reduction in the respective RMS values.

		HHDDT portion	HHDDT full	Braking
SOC saving	H	13.36%	4.38%	14.02%
	TTC	19.6%	19.82%	19.8%
	H-TTC	16.69%	18.15%	15.17%
Acceleration	H	15.86%	2.65%	7.68%
	TTC	6.3%	15.81%	13.35%
	H-TTC	4.71%	1.98%	0.75%
Jerk	H	28.05%	15.25%	10.04%
	TTC	9.89%	36.35%	10.82%
	H-TTC	19.17%	1.46%	0.11%

the reward function, confirming the fact that a higher time headway helps to prevent low TTC values [58]. Moreover, the comfort improvement is significant, as can be noticed in Table 3. Anyway, this strategy leads to the lowest energy saving compared to the other two, due to a minor reduction in the aerodynamic coefficient, with a maximum SOC reduction of the 14.38% in the full HHDDT driving cycle with respect to the leading vehicle consumption.

D. TTC STRATEGY

On the contrary, the TTC strategy, without any lower bound on the time headway, leads to reduce in a significant way the intervehicluar space showing excellent energy savings, without decreasing comfort and never running into concrete risks of collision keeping the TTC always above the safety threshold (Fig. 2, bottom right). The TTC strategy proved to reach the highest energy savings among the three spacing

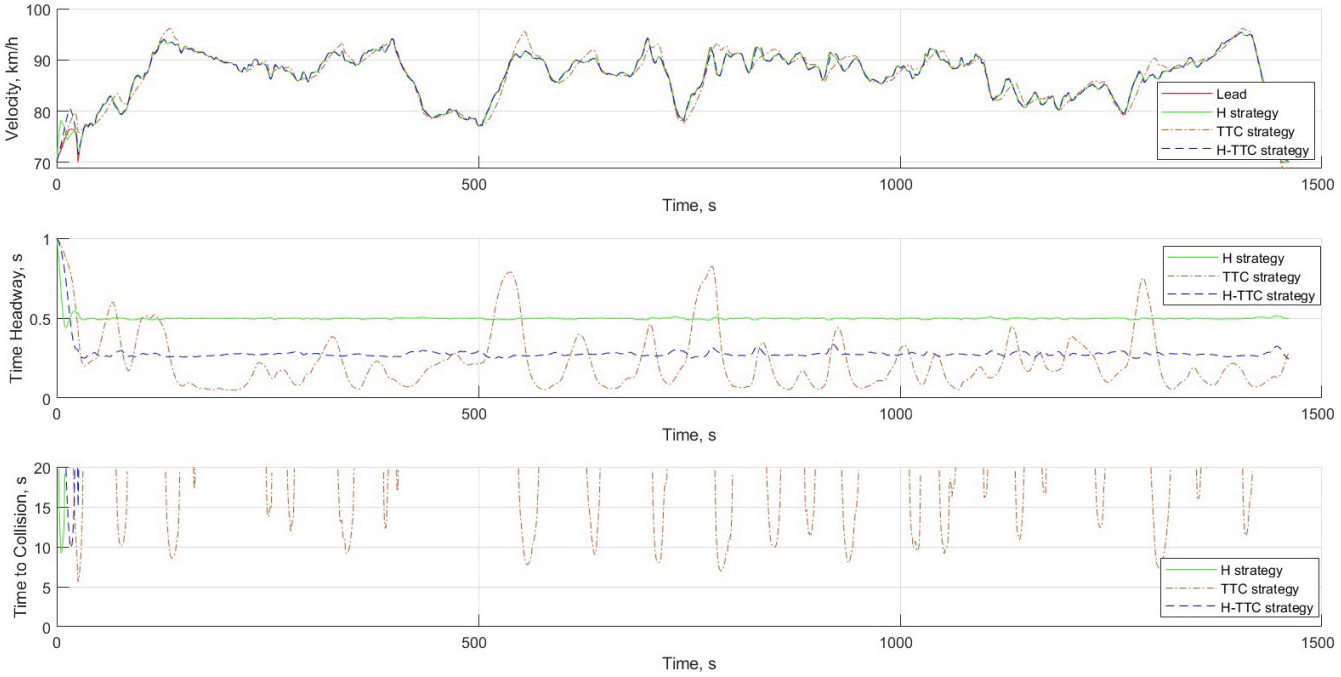

FIGURE 3. Results for the full HHDDT driving cycle.

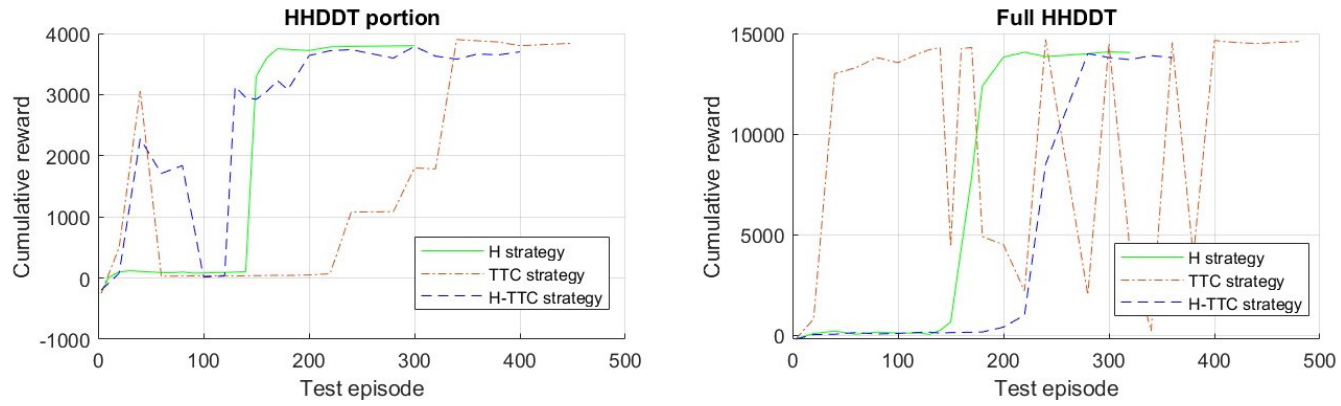

FIGURE 4. Cumulative reward trend.

strategies, obtaining up to the 19.82% of SOC saving with respect to the leader in the entire HHDDT cycle thanks to an average time headway of 0.23 *s*, still improving at the same time the driving comfort due to the reduction of acceleration and jerk, obtaining satisfying reductions in terms of the RMS values which are reported in Table 3. The TTC has the tendency to be lower compared to the one in the H strategy, but the imposed safety limitation of 4 *s* is never exceeded, therefore, without ever reaching a concrete risk of collision.

E. H-TTC STRATEGY

The hybrid H-TTC strategy can be considered as a compromise between the previous two, as it considers both TTC and minimum time headway, this time chosen as 0.25 *s*, as safety parameters. The combination of these two features leads to obtain a high SOC saving with respect to the leader (up to 18.15% over the whole HHDDT cycle) and at the same time generally high TTC values during the whole cycle. Unlike the TTC strategy, the ego vehicle never gets too close to the leader, avoiding the occurrence of very short gaps. All these characteristics make this strategy the best candidate among the other proposed spacing strategies for a possible implementation.

F. CUMULATIVE REWARD TREND

The training progress in terms of cumulative reward is shown in Fig. 4. While the H strategy converges stably after a relatively low number of episodes, the TTC one has the most unstable trend and needs more attempts to perform correctly. The learning process of the H-TTC strategy is instead comparable with that of the H strategy, even if it is slightly slower.

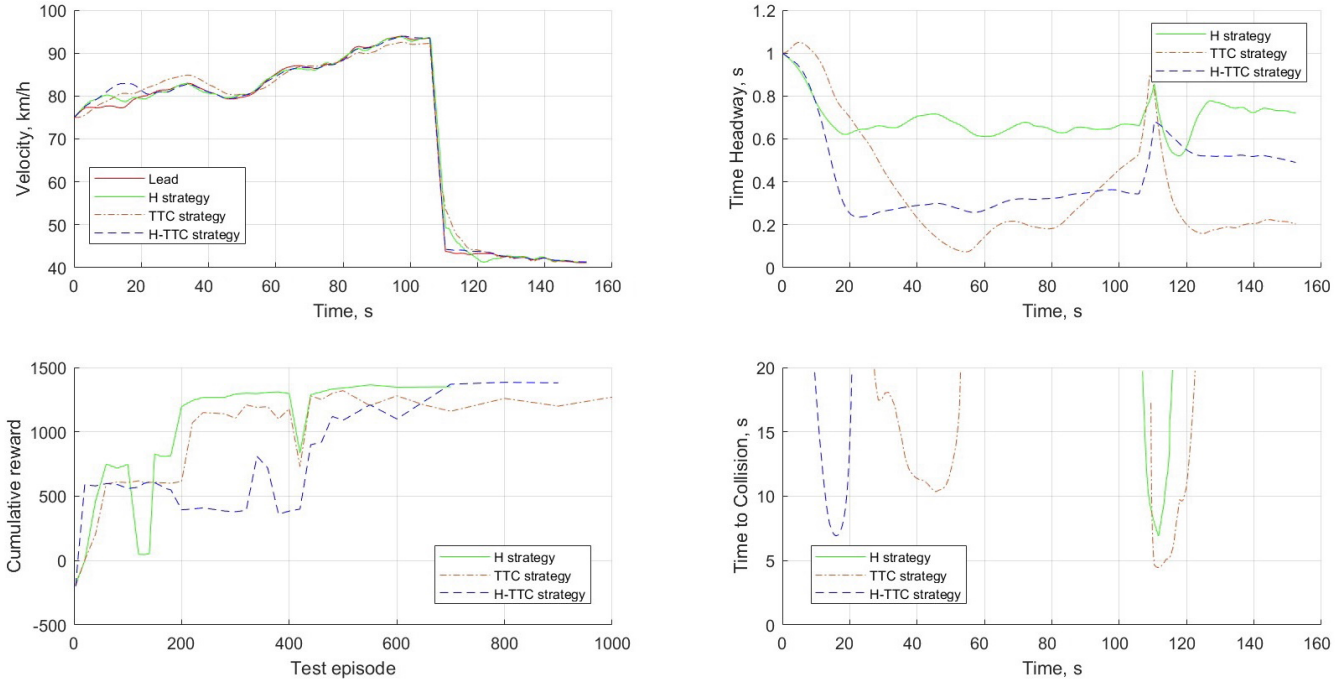

FIGURE 5. Braking test.

G. BRAKING TEST

Standard driving cycles are representative of common driving attitudes and are suitable references for vehicle testing activities, typically with respect to emissions or fuel consumption. Anyway, they usually do not contemplate critical or particular unexpected situations that may arise from real driving scenarios, like, for example, hard braking due to traffic slowdowns or obstacles on the road. For this reason, in order to investigate the algorithm's behavior also in front of these particular situations, the braking driving cycle defined in Sec. VI-A has been used. The reason of this test is to simulate a critical situation, and compare the different safety approaches of the three spacing strategies to this specific problem, and their possible chanches of success. As resumed in Fig. 5, all three spacing strategies proved to perform correctly the braking without colliding with the leading vehicle, respecting the corresponding safety parameters. Also, the TTC strategy, which is critical since it leads to the smallest average gap between the two vehicles, proved to successfully face the braking maneuver, if correctly trained. An interesting observation to note is that all three strategies have a quite stable learning process, including the TTC one unlike for the standard driving cycle (Fig. 5, bottom left), demonstrating good adaptability by the algorithm even in critical situations.

VII. CONCLUSION

In this work, a novel CACC solution has been proposed for heavy duty vehicles based on RL using the TD3 algorithm, focusing on air drag reduction as an energy-saving feature and, at the same time, on the comfort improvement obtainable thanks to the reduction in acceleration and jerk values. Three different spacing strategies that involve minimum time headway and TTC as safety parameters have been investigated, testing their behaviors considering the HHDDT standard driving cycle. Finally, a braking test has been performed that evaluates the algorithm behavior also in front of critical scenarios. All three adopted spacing strategies proved to correctly satisfy the comfort, energy-saving and safety objectives defined in the reward function in all possible situations. The ego vehicle tends to correctly close the gap with the leader, within the imposed limits: while the H strategy is the most conservative one in terms of safety, the TTC strategy led to the smallest average gap without consequent collision risks, showing up to 19.82% of energy reduction with respect to the Lead truck. However, the TTC strategy, which occasionally leads to very narrow gaps, may be not feasible for a real-case application, if considering also possible communication and actuation delays. The H-TTC strategy stands between the other strategies as the best candidate for a possible implementation, since it allows to obtain excellent energy-saving benefits and at the same time allows to always maintain a certain margin of gap with the leader, mitigating the influence of the aforementioned delays. The simultaneous comfort improvement with respect to the reference driving cycle is significant, resulting in smoother velocity and acceleration profiles and consistent reductions of the jerk peaks.

Regarding possible future developments of the proposed solution, some simplifications and modifications can be made to the reward function and to each reward term to optimize the training process and obtain more stable results. As has

been done with the acceleration reward term, a study on possible nonlinear variabilities of the other reward terms can be performed with the aim of increasing the training performances. Moreover, a deep investigation about the possible effects of time delays or communication losses in the V2V data exchange between the two trucks, by explicitly modeling these phenomena in the simulation model, can be interesting in highlighting their influence on the three spacing strategies in real-case applications.

REFERENCES

[1] M. Shang and R. E. Stern, "Impacts of commercially available adaptive cruise control vehicles on highway stability and throughput," *Transp. Res. C, Emerg. Technol.*, vol. 122, Jan. 2021, Art. no. 102897.

[2] W. J. Schakel, B. van Arem, and B. D. Netten, "Effects of cooperative adaptive cruise control on traffic flow stability," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 759–764.

[3] S. E. Shladover, D. Su, and X.-Y. Lu, "Impacts of cooperative adaptive cruise control on freeway traffic flow," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2324, no. 1, pp. 63–70, Jan. 2012.

[4] V. Milanés, S. E. Shladover, J. Spring, C. Nowakowski, H. Kawazoe, and M. Nakamura, "Cooperative adaptive cruise control in real traffic situations," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 296–305, Feb. 2014.

[5] V. Milanés and S. E. Shladover, "Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data," *Transp. Res. C, Emerg. Technol.*, vol. 48, pp. 285–300, Nov. 2014.

[6] B. A. Guvenc and E. Kural, "Adaptive cruise control simulator: A low-cost, multiple-driver-in-the-loop simulator," *IEEE Control Syst. Mag.*, vol. 26, no. 3, pp. 42–55, Jun. 2006.

[7] P. Shakouri, A. Ordys, D. S. Laila, and M. Askari, "Adaptive cruise control system: Comparing gain-scheduling PI and LQ controllers," *IFAC Proc. Volumes*, vol. 44, no. 1, pp. 12964–12969, Jan. 2011.

[8] L.-H. Luo, H. Liu, P. Li, and H. Wang, "Model predictive control for adaptive cruise control with multi-objectives: Comfort, fuel-economy, safety and car-following," *J. Zhejiang Univ. Sci. A*, vol. 11, pp. 191–201, Feb. 2010.

[9] S. Li, K. Li, R. Rajamani, and J. Wang, "Model predictive multi-objective vehicular adaptive cruise control," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 3, pp. 556–566, May 2011.

[10] A. Musa, F. Miretti, and D. Misul, "MPC-based cooperative longitudinal control for vehicle strings in a realistic driving environment," SAE Tech. Paper 2023-01-0689, Apr. 2023.

[11] D. R. Lopes and S. A. Evangelou, "Energy savings from an eco-cooperative adaptive cruise control: A BEV platoon investigation," in *Proc. 18th Eur. Control Conf. (ECC)*, Jun. 2019, pp. 4160–4167.

[12] F. Ma, Y. Yang, J. Wang, Z. Liu, J. Li, J. Nie, Y. Shen, and L. Wu, "Predictive energy-saving optimization based on nonlinear model predictive control for cooperative connected vehicles platoon with V2V communication," *Energy*, vol. 189, Dec. 2019, Art. no. 116120.

[13] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.

[14] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel, "Benchmarking deep reinforcement learning for continuous control," in *Proc. Int. Conf. Mach. Learn.*, vol. 3, 2016, pp. 2001–2014.

[15] Y. Zhu, D. Zhao, and X. Li, "Using reinforcement learning techniques to solve continuous-time non-linear optimal tracking problem without system dynamics," *IET Control Theory Appl.*, vol. 10, no. 12, pp. 1339–1347, Aug. 2016.

[16] Y. Lin, J. McPhee, and N. L. Azad, "Comparison of deep reinforcement learning and model predictive control for adaptive cruise control," *IEEE Trans. Intell. Vehicles*, vol. 6, no. 2, pp. 221–231, Jun. 2021.

[17] M. Zhu, Y. Wang, Z. Pu, J. Hu, X. Wang, and R. Ke, "Safe, efficient, and comfortable velocity control based on reinforcement learning for autonomous driving," *Transp. Res. C, Emerg. Technol.*, vol. 117, Aug. 2020, Art. no. 102662.

[18] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. ICLR*, 2015, pp. 1–14.

[19] D. Zhao, B. Wang, and D. Liu, "A supervised actor-critic approach for adaptive cruise control," *Soft Comput.*, vol. 17, no. 11, pp. 2089–2099, Nov. 2013.

[20] C. Desjardins and B. Chaib-draa, "Cooperative adaptive cruise control: A reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1248–1260, Dec. 2011.

[21] X. Qu, Y. Yu, M. Zhou, C.-T. Lin, and X. Wang, "Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach," *Appl. Energy*, vol. 257, Jan. 2020, Art. no. 114030.

[22] S. Wei, Y. Zou, T. Zhang, X. Zhang, and W. Wang, "Design and experimental validation of a cooperative adaptive cruise control system based on supervised reinforcement learning," *Appl. Sci.*, vol. 8, no. 7, p. 1014, Jun. 2018.

[23] T. Chu and U. Kalabic, "Model-based deep reinforcement learning for CACC in mixed-autonomy vehicle platoon," in *Proc. IEEE 58th Conf. Decis. Control (CDC)*, Dec. 2019, pp. 4079–4084.

[24] M. Li, Z. Cao, and Z. Li, "A reinforcement learning-based vehicle platoon control strategy for reducing energy consumption in traffic oscillations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5309–5322, Dec. 2021.

[25] A. Peake, J. McCalmon, B. Raiford, T. Liu, and S. Alqahtani, "Multi-agent reinforcement learning for cooperative adaptive cruise control," in *Proc. IEEE 32nd Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2020, pp. 15–22.

[26] *Outlook for Energy: A Perspective to 2040*, ExxonMobil, Houston, TX, USA, 2019.

[27] A. A. Alam, A. Gattami, and K. H. Johansson, "An experimental study on the fuel reduction potential of heavy duty vehicle platooning," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 306–311.

[28] V. Turri, B. Besselink, and K. H. Johansson, "Cooperative look-ahead control for fuel-efficient and safe heavy-duty vehicle platooning," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 1, pp. 12–28, Jan. 2017.

[29] A. Capuano, M. Spano, A. Musa, G. Toscano, and D. A. Misul, "Development of an adaptive model predictive control for platooning safety in battery electric vehicles," *Energies*, vol. 14, no. 17, p. 5291, Aug. 2021.

[30] F. Browand, J. Mcarthur, and C. Radovich, "Fuel saving achieved in the field test of two tandem trucks," Inst. Transp. Stud., Berkeley, CA, USA, Tech. Rep. UCB-ITS-PRR-2004-20, Jan. 2004.

[31] A. A. Hussein and H. A. Rakha, "Vehicle platooning impact on drag coefficients and energy/fuel saving implications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1199–1208, Feb. 2022.

[32] S. Tsugawa, S. Kato, and K. Aoki, "An automated truck platoon for energy saving," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2011, pp. 4109–4114.

[33] H. Long, A. Khalatbarisoltani, and X. Hu, "MPC-based eco-platooning for homogeneous connected trucks under different communication topologies," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2022, pp. 241–246.

[34] P. Smith and D. Bevly, "Analysis of on-road highway testing for a two truck cooperative adaptive cruise control (CACC) platoon," SAE Tech. Papers 2020-01-5009, Oct. 2019.

[35] B. McAuliffe, M. Lammert, X.-Y. Lu, S. Shladover, M.-D. Surcel, and A. Kailas, "Influences on energy savings of heavy trucks using cooperative adaptive cruise control," SAE Tech. Papers 2018-01-1181, Apr. 2018.

[36] S. Albeaik, T. Wu, G. Vurimi, F.-C. Chou, X.-Y. Lu, and A. M. Bayen, "Longitudinal deep truck: Deep longitudinal model with application to Sim2Real deep reinforcement learning for heavy-duty truck control in the field," *J. Field Robot.*, vol. 40, no. 2, pp. 306–329, 2022.

[37] S. Albeaik, T. Wu, G. Vurimi, F.-C. Chou, X.-Y. Lu, and A. M. Bayen, "Deep truck cruise control: Field experiments and validation of heavy duty truck cruise control using deep reinforcement learning," *Control Eng. Pract.*, vol. 121, Apr. 2022, Art. no. 105026.

[38] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 1–10.

[39] K. N. de Winkel, T. Irmak, R. Happee, and B. Shyrokau, "Standards for passenger comfort in automated vehicles: Acceleration and jerk," *Appl. Ergonom.*, vol. 106, Jan. 2023, Art. no. 103881.

[40] C Hydén, "Traffic conflicts technique: State-of-the-art," in *Traffic Safety Work with Video Processing*, vol. 37. Kaiserslautern, Germany: Universitat Kaiserslautern, 1996, pp. 3–14.

[41] M. M. Minderhoud and P. H. L. Bovy, "Extended time-to-collision measures for road traffic safety assessment," *Accident Anal. Prevention*, vol. 33, no. 1, pp. 89–97, Jan. 2001.

[42] M. Saffarzadeh, N. Nadimi, S. Naseralavi, and A. R. Mamdoohi, "A general formulation for time-to-collision safety indicator," *Proc. Inst. Civil Eng. Transp.*, vol. 166, no. 5, pp. 294–304, Oct. 2013.
[43] B. van Arem, C. J. G. van Driel, and R. Visser, "The impact of cooperative adaptive cruise control on traffic-flow characteristics," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 4, pp. 429–436, Dec. 2006.
[44] W. Hucho, *Aerodynamics of Road Vehicles*. London, U.K.: Butterworth-Heinemann, 1987.
[45] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015.
[46] C. J. C. H. Watkins and P. Dayan, "Technical note: Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.
[47] Y.-T. Yen, J.-J. Chou, C.-S. Shih, C.-W. Chen, and P.-K. Tsung, "Proactive car-following using deep-reinforcement learning," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6.
[48] C. Wu, Z. Xu, Y. Liu, C. Fu, K. Li, and M. Hu, "Spacing policies for adaptive cruise control: A survey," *IEEE Access*, vol. 8, pp. 50149–50162, 2020.
[49] Z. Li, T. Chu, I. V. Kolmanovsky, and X. Yin, "Training drift counteraction optimal control policies using reinforcement learning: An adaptive cruise control example," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 9, pp. 2903–2912, Sep. 2018.
[50] D. Yanakiev and I. Kanellakopoulos, "Variable time headway for string stability of automated heavy-duty vehicles," in *Proc. 34th IEEE Conf. Decis. Control*, Dec. 1995, pp. 4077–4081.
[51] R. Van der Horst and J. Hogema, *Time-to-Collision and Collision Avoidance Systems*. Groningen, The Netherlands: Univ. Groningen, 1994.
[52] S. J. Hirst and R. Graham, "The format and presentation of collision warnings," in *Ergonomics and Safety of Intelligent Driver Interfaces*. Boca Raton, FL, USA: CRC Press, 1997.
[53] M. Elbanhawi, M. Simic, and R. Jazar, "In the passenger seat: Investigating ride comfort measures in autonomous cars," *IEEE Intell. Transp. Syst. Mag.*, vol. 7, no. 3, pp. 4–17, Fall 2015.
[54] K. Czarnecki, "Automated driving system (ADS) high-level quality requirements analysis—Driving behavior comfort," Waterloo Intell. Syst. Eng. (WISE) Lab, Univ. Waterloo, Waterloo, CA, USA, Tech. Rep., Jul. 2018, doi: 10.13140/RG.2.2.19925.32483.
[55] M. Acquarone, A. Borneo, and D. A. Misul, "Acceleration control strategy for battery electric vehicle based on deep reinforcement learning in V2V driving," in *Proc. IEEE Transp. Electrific. Conf. Expo (ITEC)*, Jun. 2022, pp. 202–207.
[56] K. Laib, O. Sename, and L. Dugard, "String stable H_∞ LPV cooperative adaptive cruise control with a variable time headway," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 15140–15145, 2020.
[57] G. Sidorenko, J. Thunberg, K. Sjöberg, and A. Vinel, "Vehicle-to-vehicle communication for safe and fuel-efficient platooning," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 795–802.
[58] K. Vogel, "A comparison of headway and time to collision as safety indicators," *Accident Anal. Prevention*, vol. 35, no. 3, pp. 427–433, May 2003.
[59] J. R. Billing, "Performance limits of heavy trucks," Canadian Council Motor Transp. Administrator, Ottawa, ON, Canada, CCMTA Load Security Research Project, Tech. Rep. 20, Oct. 1998.
[60] F. Zhen, N. N. Clark, C. R. Bedick, M. Gautam, W. S. Wayne, G. J. Thompson, and D. W. Lyons, "Development of a heavy heavy-duty diesel engine schedule for representative measurement of emissions," *J. Air Waste Manag. Assoc.*, vol. 59, no. 8, pp. 950–959, 2009, doi: 10.3155/1047-3289.59.8.950.

MATTEO ACQUARONE (Member, IEEE) received the B.S. and M.S. degrees in mechanical engineering from Politecnico di Torino, Turin, Italy, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree in energetics. From October 2021 to July 2022, he was a Visiting Scholar with Stanford University, Stanford, CA, USA. He is the author of three peer-reviewed conferences and journal publications. His research interests include the development of artificial intelligence control algorithms for energy management strategies of hybrid electric vehicles and energy savings of connected vehicles.

FEDERICO MIRETTI (Member, IEEE) received the M.S. and Ph.D. degrees in energetics from Politecnico di Torino, Turin, Italy, in 2018 and 2022, respectively. He is currently a Postdoctoral Research Fellow with Politecnico di Torino. He is the author of 11 peer-reviewed conferences and journal publications. His research interest includes the design of optimization-based energy management strategies for hybrid-electric powertrains and autonomous and connected vehicles.

DANIELA MISUL received the degree (cum laude) in mechanical engineering from Politecnico di Torino, in October 1997. She is currently an Associate Professor with the Dipartimento Energia "Galileo Ferraris," Politecnico di Torino. She consists several publications in scientific production, covering the following research streams, such as SI engine technology for natural gas and bi-fuel ICEs; model-based control approaches for diesel ICEs; diagnostic and predictive 0D/1D tools for evaluating injection, combustion process, and emission formation in diesel and SI engines; in-cylinder fluid-dynamics and combustion (CFD) of ICEs; optimization of hybrid architectures; CFD modeling activities for turbomachinery; AI algorithms for xEVs control; and ADAS applications. She has been contributing to the coordination of the research activities within the Polito–Engine Research Center (PT-ERC) Research Group with specific reference to the scientific responsibility and coordination of some of the research contracts as well as to team building and fund-raising activities. She held the scientific responsibility of national and international research projects and ruled through partnership agreements with companies and/or public-private bodies, which are leaders in their sector. She has been recently coordinating activities related to the use of AI for smart applications in the automotive sector. She is also coordinating activities within the Center for Automotive Research and Sustainable Mobility@PoliTO (CARS@PoliTO) Group.

LUCA SASSARA received the B.S. degree in mechanical engineering and the M.S. degree in mechatronic engineering from Politecnico di Torino, Turin, Italy, in 2020 and 2022, respectively. He is currently an ADAS Functional Integrator with Politecnico di Torino.

Open Access funding provided by 'Politecnico di Torino' within the CRUI CARE Agreement