Battery Electric Vehicle Control Strategy for String Stability based on Deep Reinforcement Learning in V2V Driving

20XX-01-XXXX

Battery Electric Vehicle Control Strategy for String Stability based on Deep Reinforcement Learning in V2V driving

Author, co-author (Do NOT enter this information. It will be pulled from participant tab in MyTechZone)

Affiliation (Do NOT enter this information. It will be pulled from participant tab in MyTechZone)

Abstract

This works presents a Reinforcement Learning (RL) agent to implement a Cooperative Adaptive Cruise Control (CACC) system that simultaneously enhances energy efficiency and comfort, while also ensuring string stability. CACC systems are a new generation of ACC which systems rely on the communication of the so-called ego-vehicle with other vehicles and infrastructure using V2V and/or V2X connectivity. This enables the availability of robust information about the environment thanks to the exchange of information, rather than their estimation or enabling some redundancy of data. CACC systems have the potential to overcome one typical issue that arises with regular ACC, that is the lack of string stability. String stability is the ability of the ACC of a vehicle to avoid unnecessary fluctuations in speed that can cause traffic jams, dampening these oscillations along the vehicle string rather than amplifying them. In this work, a real-time ACC for a Battery Electric Vehicle, based on a Deep Reinforcement Learning algorithm called Deep Deterministic Policy Gradient (DDPG), has been developed, aiming at maximizing energy savings, and improving comfort, thanks to the exchange of information on distance, speed and acceleration through the exploitation of vehicle-to-vehicle technology (V2V). The aforementioned DDPG algorithm is also designed in order to achieve the string stability. It relies on a multi-objective reward function that is adaptive to different driving cycles. The simulation results show how the agent can obtain energy savings up to 11% comparing the first following vehicle and the Lead on standard cycles and good adaptability to driving cycles different from the training one.

Introduction

In the last decades the automotive field has faced increasingly steep challenges with respect to different industry needs, such as safety, comfort and energy consumption. Advanced Driver Assistance Systems (ADASs) are a set of technological developments that can increase road safety, reducing for example the number of accidents [1]. In the context of automated driving systems, vehicle platooning represents an enabler technology for both safety and energy savings. The latter has been investigated over the last decade as it represents a potential application in reducing the greenhouse gas emissions, since the road transport sector represents the major contributor [2]. A platoon of vehicles is a string of vehicles that move together at a certain safety distance. Different platooning projects from different countries were completed in the last 20 years investigating different systems, as reported in the extensive review in [3,4], covering different vehicles, type of infrastructures and sensors. The focus of these projects was to investigate the potential energy consumption reduction and the road capacity increase, which is a key challenge of the transportation sector. The advantages linked to the drag reduction is a key aspect, investigated by many researchers as done by [5]. These works reveal that heavy-duty vehicles (HDV) are the most suitable for platooning systems, exploiting the benefits from the slipstream effect due to the impact on surface area. On the downside, driving close to the previous vehicle leads to a safety concern, but this problem can be addressed directly through the selection of the proper spacing policy [6].

To fulfil platooning cooperative task cooperative adaptive cruise control (CACC) frameworks demonstrated to be more suitable than classic Adaptive Cruise Control (ACC), since they exploit more accurate information from the whole string of vehicles such as distance, velocity, and accelerations. The exchange of data relies on the V2X or V2V technologies. One of the advantages of the CACC is that it can guarantee the string stability, which is a well-known issue in regular ACC. For example, [7] investigated string stability by developing a delay-based spacing policy.

Different cooperative strategies have been implemented in the literature. For example, a real-time controller based on the same principle underlying the equivalent consumption minimization strategy was developed by [8]. Moreover, in several works model predictive control (MPC) [9,10] has been suggested as one of the most promising solutions to CACC problems. In the last years, the innovations made in the field of reinforcement learning (RL) have allowed researchers to achieve surprising results in energy management problems [11].

Recently, deep RL algorithms have been applied to this field to overcome the hindrance of dimensionality, due to the discretization of large and continuous state and action spaces, typical of Q-learning and DP. The main goal of this paper is to show the effectiveness of the DDPG, a deep RL (DRL) algorithm, for optimal acceleration control of electric vehicles (EV) in enhancing comfort conditions while also achieving energy savings and string stability. Truck platooning is the subject of this analysis; thus, the platoon system has been modelled to include the aerodynamics effects on drag resistance. The paper is divided as follows: in the first part the vehicle platoon model and RL architecture are presented. Then, simulation results of a platoon with the lead vehicle following a FPT75 and a WLTP driving cycle are shown and conclusions are drawn.

Platoon Model and Control Algorithm

Platoon and Vehicle Model

In this section the vehicle and platoon models are introduced. The latter is composed by a number N of heavy-duty electric vehicles. The first vehicle in the platoon is the lead vehicle, while the others are the followers or ego vehicles. Each vehicle is modelled individually in a simulation environment that was developed in Simulink, considering the longitudinal dynamics and the powertrain. The vehicles' position, velocity, and acceleration (x_i, v_i, a_i) are all measured in a fixed reference frame. The powertrain and vehicle model relates the torque command, which is controlled by the RL agent, to the vehicle's acceleration, considering the e-machines' speed and torque constraints as well as the single-speed transmission of the model and vehicle's resistive load.

The resistive load for vehicle i, $F_{\mathrm{res},i}$, is:

$$F_{\mathrm{res},i} = m_i\, g\, \sin(\alpha) + m_i\, g\, f_{0,i} \cos(\alpha) + \left(\frac{1}{2}\rho\, c_{x,i}\left(d_{i-1,i}\right) A_{f,i}\right) v_i^2, \quad (1)$$

where m_i is the vehicle mass, α is the road slope, $f_{0,i}$ is the rolling resistance coefficient, ρ is the air density, $A_{f,i}$ is the vehicle frontal area, and $c_{x,i}$ is the drag coefficient. For the lead vehicle, $c_{x,0}$ is a constant (equal to the undisturbed drag coefficient), whereas for the ego vehicles $c_{x,i}$ is a function of the inter-vehicle distance $d_{i-1,i}$.

The longitudinal dynamics and the powertrain model then relate the torque command $T_{\mathrm{m},i}$ to the vehicle's acceleration, neglecting various driveline efficiencies for ease of notation, according to:

$$m_{\mathrm{tot}_i} a_i = \frac{T_{\mathrm{m},i}\tau_i}{r_{w,i}} - F_{\mathrm{res},i}\,, \quad (2)$$

where $m_{tot,i}$ is the equivalent total mass, accounting for translating and rotating mass and inertia, $r_{w,i}$ is the wheel radius, and τ_i represents the overall transmission ratio. In the test case presented in this work, all vehicles have the same parameters.

In addition, the e-machine model evaluates the electrical power using a steady-state efficiency map and the battery model relates this to the state of charge (SOC) dynamics using an internal resistance model:

$$I_{\mathrm{b}} = \frac{V_{\mathrm{oc}} - \sqrt{V_{\mathrm{oc}}^2 - 4\, R_{\mathrm{b}}\, P_{\mathrm{b}}}}{2\, R_{\mathrm{b}}}, \quad (3)$$

$$\dot{SOC} = \frac{I_{\mathrm{b}}}{Q_{\mathrm{b}}}. \quad (4)$$

Here I_{b}, V_{oc}, R_{b}, P_{b} and Q_{b} are respectively the battery current, open-circuit voltage, internal resistance, power, and maximum capacity (in ampere-seconds).

As stated in Eq. 1, the drag coefficient was modelled as a function of the inter-vehicle distance. In particular, the following semi-empirical relationship, derived by [12], was used to characterize the effect of the inter-vehicular distance on the drag reduction ΔC_x:

$$\Delta C_x(\%) = \left(1 - \frac{a_{3,i}d^3 + a_{2,i}d^2, a_{1,i}d + a_{0,i}}{b_{3,i}d^3 + b_{2,i}d^2 + b_{1,i}d + b_{0,i}}\right) \cdot 100 \quad (5)$$

Where d represent the relative distance at which the vehicle is influenced by the reduction of the drag resistances, while $a_{n,i}$ and $b_{n,i}$ are the experimental coefficient obtained by [12] based on empirical data. Figure 1 shows the trend of the drag reduction coefficient with respect to the inter-vehicle distance for a platoon composed by 4 HDVs. In our simulation framework, followers 1 and 2 are considered as middle vehicles of a platoon. As expected, the two middle and the trail vehicles are the most impacted by drag reduction.

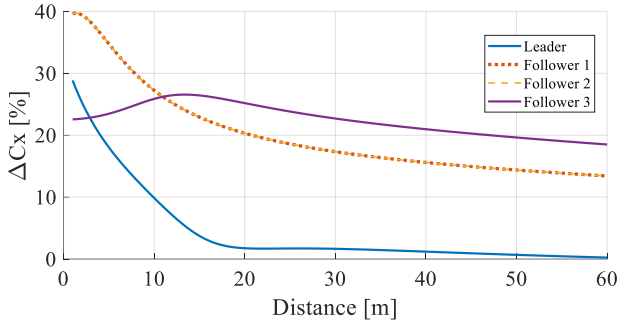

Figure 1. Drag coefficient as a function of the inter-vehicle distance.

The inter-vehicle distance and its derivative are defined as:

$$\begin{cases} d_{i-1,i} = x_{,i-1} - x_i - l_{i-1} \\ \dot{d}_{i-1,i} = v_{i-1} - v_i \end{cases} \quad (6)$$

The vehicle platooning is designed to achieve the advantages due to the implementation of a CACC. In the present paper a platoon of vehicles is designed so that the lead vehicle presents a PID logic following a specific mission, while the follower vehicles are controlled by their own control units, that exploit the information of the sensor measures (i.e. the velocity v_i and the inter-vehicle distance $d_{i-1,i}$), and those available thanks to the V2V communication, (i.e. leader acceleration). The control unit computes the torques and transfers the signals to the actuators.

Control Algorithm

In this section, we describe the Reinforcement Learning agent that was implemented in this work. The role of a Reinforcement Learning agent is to evaluate and execute decisions, also called actions. Each action is evaluated through a reward function that expresses its effectiveness and in achieving the control. RL agents learn through trial and error, leading the trained agent to achieve an optimal policy in an external environment, that can be described by an observable state *s*. The agent takes an action *a* that is evaluated through a reward *r* which can also depend on the observed state. This interaction between agent and environment is the learning source of the agent. The objective of an RL agent is maximizing the cumulative reward over the whole mission. Comparing RL with supervised learning the

biggest difference is that RL does not require data before starting the training process whose duration is set as the maximum number of runs, called episodes. For this work, the latter was set to E_{max} = 2000.

DDPG Algorithm

The deep deterministic policy gradient (DDPG) algorithm [13] estimates the action-value $Q(s,a)$, defined as discounted sum of rewards through deep neural networks (DNNs), that work as function approximators, addressing the problem of variables discretization that affects tabular Q-learning. Moreover, it is a model-free approach; hence, it does not require a prediction model of the system as, for instance, in model predictive control (MPC).

DDPG is an actor-critic algorithm characterized by a continuous action space and an observation space that can be either continuous or discrete. Since it is a deep RL algorithm, it is more suitable for large action and state spaces (as is the case for the control of vehicle platoons) with respect to more classical RL approaches. The agent uses four simple feed-forward neural networks (function approximators): two actors μ and two critics Q characterized by θ^{μ} and θ^{Q} weights. Each net contains three hidden layers with 56 neurons, using the rectified linear unit (ReLU) activation function. The observed state of the environment is the input of the actor networks, whose output is the action; the latter and the state are the input of the critic nets that estimate the Q values.

The use of two networks for actor and for critic is necessary for the stability of Q-values. A target actor μ_t and a target critic Q_t are characterized by weights θ^{μ_t} and θ^{Q_t}, at first initialized equal to θ^{μ} and θ^{Q} respectively, and updated for each training episode using a smoothing method for target update:

$$\theta^{\mu_t} = \tau\theta^{\mu} + (1-\tau)\theta^{\mu_t}, \tag{6}$$

$$\theta^{Q_t} = \tau\theta^{Q} + (1-\tau)\theta^{Q_t}, \tag{7}$$

where $\tau \ll 1$ is a smoothing factor. This was set to $\tau = 10^{-3}$, which is a common value in literature [13] to strike a reasonable balance between stability and speed of the training process (the extent to which the target actor and critic networks are updated).

To balance exploration and exploitation, an Ornstein-Uhlenbeck action noise model is used, to enhance exploration. The default noise mean value is 0, while standard deviation is here set to 0.6. To push toward the exploitation as the training process progresses, we set the standard deviation decay rate parameter to $1 * 10^{-5}$.

Another important option in the DDPG algorithm is the experience replay memory. For every training iteration, a batch of *n* random arrays (s, a, r, s') are sampled from the replay memory and are used to train the critic and actor networks through the respective loss functions L_c and L_a:

$$L_c = \frac{1}{n}\sum_{i=1}^{n}(y - Q(s,a|\theta^{Q}))^2, \tag{8}$$

$$L_a = \frac{1}{n}\sum_{i=1}^{n} Q(s,\mu(s).) \tag{9}$$

Control strategy

The present work scenario comprehends a platoon of N vehicles travelling in a longitudinal-only dynamics (or hypothetically on a straight and flat road). The follower vehicles receive the information about the other vehicles. As already mentioned, the strong assumption made at this stage is that all the information is instantaneously available and without any error. The state variables, chosen among the available data are the time gap between the i^{th} vehicle and its preceding one, the time to collision of the same two vehicles, their velocities and the acceleration of the leading vehicle of the platoon.

$$\begin{cases} t_h^i \\ TTC^i \\ v_{ego}^i \\ v_{ego}^{i-1} \\ a_{lead} \end{cases}.$$

The main objective of the agent is to enhance energy savings, safety, comfort, and the stability of the string of vehicles, satisfying inter-vehicle distance limits. In this work the spacing policy does not rely on the measured distance (in meters) but on the time, called time headway, during which the front bumper of the preceding vehicle and the front bumper of the following vehicle pass a fixed position on the road. In order to achieve these objectives, the following reward function is used:

$$r = \frac{w_{th}r_{th} + w_{TTC}r_{TTC} + w_{acc}r_{acc}}{w_{th} + w_{TTC} + w_{acc}}, \tag{10}$$

where r_{th}, r_{TTC} and r_{acc} are respectively reward functions for time headway, time to collision and acceleration, while w_{th}, w_{TTC} and w_{acc} are the weights of the respective rewards. This multi-objective reward function is composed as the weighted sum of three rewards, each with value saturated between -1 and 1, and thus have value oscillations in the same range. Different attempts have been made, and the following values were ultimately retained:

$$\begin{cases} w_{th} = 0.5 \\ w_{TTC} = 0.5 \\ w_{acc} = 0 \end{cases} \text{ if } a_{lead} \geq 0.5\frac{m}{s^2} \text{ or } a_{lead} \leq -0.5\frac{m}{s^2},$$

$$\begin{cases} w_{th} = 0.25 \\ w_{TTC} = 0.25 \\ w_{acc} = 0.5 \end{cases} \text{ if } -0.5\frac{m}{s^2} < a_{lead} < 0.5\frac{m}{s^2}.$$

The following distribution allows to focus on safety, through time headway mainly, at higher accelerations, and on comfort at lower acceleration, when the acceleration is more likely to present frequent small changes. Since we are dealing with a fully electric vehicle the main advantages in terms of energy consumption derive from drag-distance correlation and regeneration during braking. For this reason, since the reward defined aims at optimizing the distance, through the time headway and time to collision terms, and the acceleration, no term directly regarding SOC is added. During training, when the distance between two consecutive vehicles becomes too high or zero, or the speed of the ego vehicle is negative the episode is stopped, the reward is set at the most negative value, in order to penalize such a condition. In general, each term of the reward function is built in

order to penalize bad actions and enhance the good ones. To achieve this the reward terms are shaped as follows:

$$r_{th} = 1 - 2^{sign(th_{desired} - th)} \cdot (th_{desired} - th)^2, \quad (11)$$

$$r_{TTC} = 1 + \frac{TTC - TTC_{critic}}{TTC_{critic} - 4}, \quad (12)$$

$$r_{acc} = 1 - |acc|. \quad (13)$$

With this reward definition, the platoon of vehicles is able to achieve stability, comfort, energy savings and safety with good adaptability to different driving cycles.

The spacing policy affects the string stability, as mentioned in [6]. The time headway reward term was used as the main factor in meeting the car-following objective. The acceleration reward term was then added to mitigate steep accelerations and therefore improve comfort and, to a minor extent, energy savings. Finally, reward term based on TTC was added to improve the training phase in specific situations: specifically, when the ego vehicles proceed at low speed and the action chosen by the agent led to lower time headways than the desired one. The addition of the TTC term proved in practice to improve convergence during training.

Results

In this section, simulation results obtained by the trained agent are described. The agent was trained in a car-following scenario, with one lead vehicle and one ego vehicle. The training cycle was a segment of the FTP75 driving cycle (specifically, the time interval [605s, 1022s]), saturated at a minimum velocity of 2 m/s to avoid reverse movement. The choice has been made in order to take a representative segment of the driving cycle while reducing the training time. Once the agent was trained, it was deployed and used as the controller for all follower vehicles. In particular, a platoon of four vehicles was tested (1 leader and 3 followers) over the training cycle, the WLTP class 2 cycle and the FTP75 complete cycle.

Figure 2 shows the velocity of all vehicles on a portion of the WLTP cycle. These results show that all following vehicles can track the speed profile of their preceding vehicle while gradually attenuating the peaks present in the leader vehicle profile, which indicates a visibly stable behaviour for the platoon. As expected, a similar behaviour is also obtained in the training cycle, as shown by the results in Figure 3.

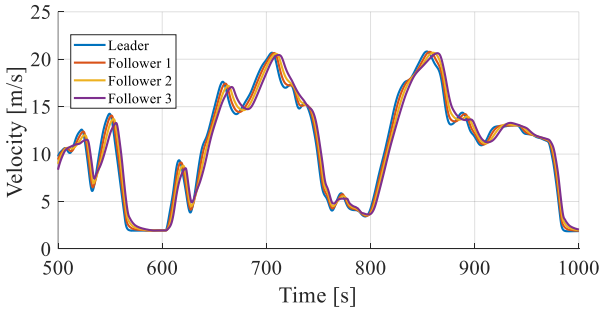

Figure 2. Velocity profiles over the WLTP cycle in the time interval [500s,1000s].

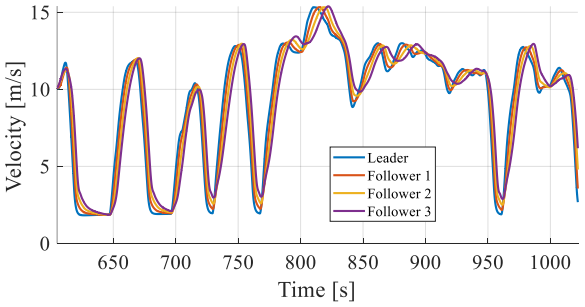

Figure 3. Velocity profiles over the training cycle (the time interval [605s,1022s] of the FTP75 cycle).

To further analyse the stability of the string of vehicles, the inter-vehicular distance of the vehicles in the platoon is represented in Figure 4. In fact, as stated by [14], several definitions of string stability exist, all considering the amplification of oscillation of a signal. Considering the oscillations in velocity, Figures 2 and 3 show an attenuation under most conditions. Considering instead the inter-vehicle distance, Figure 4 for shows a less stable behaviour: in some of the peaks the inter-vehicle distance from one vehicle to its leader is slightly decreasing along the platoon, while at other times the distance is slightly increasing. However, this increase is always very small.

It is useful to remark that in DRL-based control, there are no theoretical guarantees for string stability as the trained agent is essentially a black box. Instead, [15] proposed an *empirical* measure of string stability that can be evaluated by considering attenuations in the norm of acceleration values (the ratios identified as dampening ratio $d_{p,i}$):

$$d_{p,i} = \frac{\left|\left|a_i^t\right|\right|_2}{\left|\left|a_0^t\right|\right|_2} = \frac{\left(\sum_{t=0}^{K}|a_i^t|^2\right)^{\frac{1}{2}}}{(\sum_{t=0}^{K}|a_0^t|^2)^{\frac{1}{2}}}, \quad (14)$$

and ensuring that $d_{p,i} \leq 1, \forall i \in [1,2,\dots,N]$. Here, a_i^t is the i-th vehicle acceleration at timestep t and K is the total number of timesteps, and index 0 refers to the leading vehicle.

According to this definition, we achieved: $d_{p,1} = 0.933$; $d_{p,2} = 0.871$; $d_{p,3} = 0.807$ which is compatible with the requirement that $d_{p,i} \leq d_{p,i-1}$. While this is not a substitute for a theoretical guarantee of string stability, it can still be considered a relevant metric if combined with the considerations of the previous paragraphs.

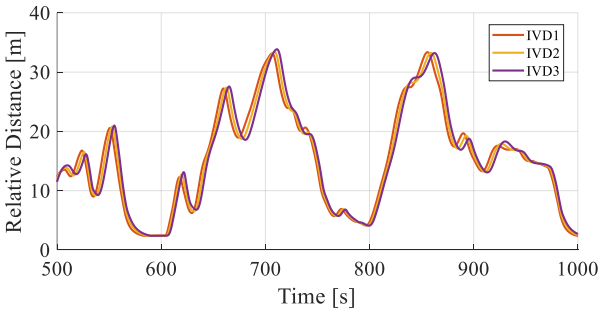

Figure 4. Inter-vehicular distance (IVD) profiles over the WLTP cycle in the time interval [500s,1000s].

For what concerns the safety requirements, Figure 5 shows that the time headway remains between 2s and 1s for all vehicles in the platoon, when considering the whole WLTP test cycle. As stated by [16], this is the most frequent situation in a car-following scenario. Nevertheless, even time headways lower than 1s can be still considered safe under some conditions such as at low speed when the situation ahead is easily predictable.

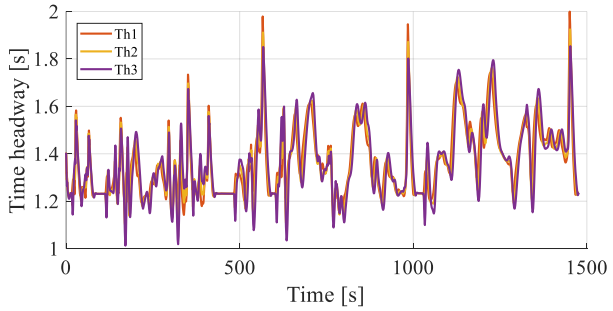

Figure 5. Time headway profiles over the WLTP cycle.

Figure 6 shows the time profiles of the battery state of charge for each vehicle. Clearly, all follower vehicles have a higher SOC than the lead vehicle. This is mainly attributable to two factors: the reduced air drag that the follower vehicles face with respect to the lead vehicle, and a reduction in the acceleration. This reduction can be inferred from Figure 7), which represents the root mean square (RMS) acceleration values for each vehicle. The fact that the difference in the final SOC between the lead and follower 1 vehicles is larger than the difference between the followers themselves is probably a reflection of the largest contribution attributable to air drag reduction with respect to the reduction in acceleration.

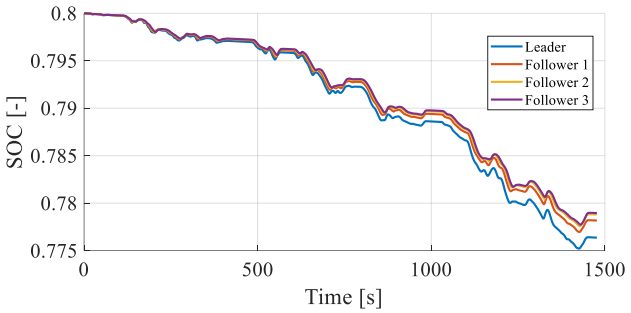

Figure 6. SOC profiles over WLTP cycle.

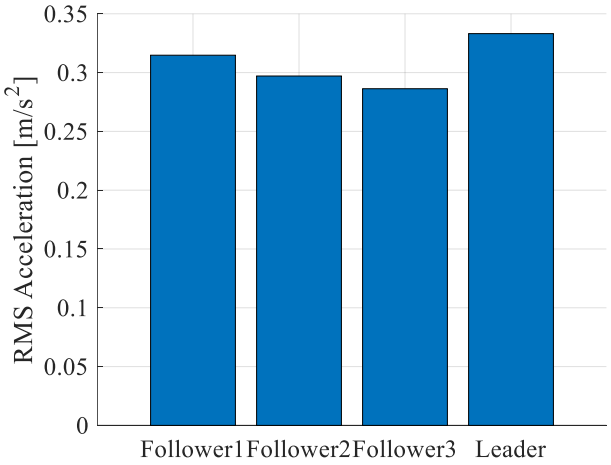

Figure 7. RMS Accelerations values over WLTP cycle

Finally, the agent's performance in comfort can be quantified by the reduction in jerk obtained by the follower vehicles with respect to the lead vehicle. Table 1 shows these reductions in terms of RMS: all follower vehicles in the platoon have lower jerk than the lead vehicle. Our results also showed reductions in peak jerk values (not represented here). The same table also reports the reduction in energy consumption of each follower vehicle with respect to the lead vehicle; this, together with the previous discussion related to the SOC profiles, confirms the effectiveness of the agent with respect to energy saving.

Table 1. Summary of simulation results.

Train cycle	**Test cycle**	**Energy savings w.r.t. lead (%)**			**RMS jerk reduction w.r.t. lead (%)**		
		Ego1	***Ego2***	***Ego3***	***Ego1***	***Ego2***	***Ego3***
FTP75 (605-1022s)	**FTP75 (605-1022s)**	10.5	9.9	6.5	69.2	71.3	69.2
	FTP75	10.5	13.2	13.0	41.0	47.3	48.0
	WLTP class 2	11.0	13.8	13.9	25.3	26.7	18.8

Still considering Table 1, we now turn our attention to the differences between the three considered cycles. As previously mentioned, the training cycle was the time interval [605s, 1022s] of the FPT75 cycle, while the whole FTP75 cycle as well as the WLTP class 2 cycles, were used as test cycles. For all cycles, the trained agent proved to achieve significant energy savings and a comfort improvement.

The first noticeable difference is that the energy savings for the second and third followers in the training cycle appear to be worse than the test cycles. This can be explained by the characteristic of the training cycle, which is characterized in large part by a highly dynamic sequence of steep accelerations and braking phases. As a consequence, the last two vehicles in the platoon do not have enough time to exploit regenerative braking to the same extent as the leader.

Similar results were obtained and discussed in a work by [17] using an optimization algorithm based on dynamic programming. This difference is not present for the two test cycles, for which the energy savings are quite similar.

On the other hand, significant differences appear in the jerk values. In particular, the WLTP class 2 driving cycle shows the smallest savings. Rather than an issue with the training process, this was attributed to the fact that the jerk of the lead vehicle itself is much lower on this cycle (RMS 0.12 m/s^3) compared to the FTP75 cycle (RMS 0.23 m/s^3); as a result, there is a much lower margin for improvement.

Conclusions

In this paper, we addressed the problem of torque control for a platoon of fully electric vehicles following a lead vehicle on a straight road. The main objective of this work is to show the potential of DRL algorithms to achieve significant energy savings for the controlled vehicles (the platoon), enhancing comfort and safety conditions. The first main innovative contribution of this work is the use of a DDPG agent with the main objective of obtaining energy savings, still guaranteeing safe driving conditions. Moreover, the use of an adaptive multi-objective reward function able to achieve good results on different driving cycles without the necessity of tuning parameters is enabled thanks to the definition of a reward function that is not affected by the training driving cycle. The agent is trained and tested on different standard driving cycles such as FTP75 and WLTP, achieving good results and showing great adaptability also to conditions not seen during training. The main limitations of the work are related to the relatively simple driving scenario and to the ideal assumption of no communication delay. Possible future work might be building a higher-fidelity simulation model of the vehicle, allowing also steering actions and including errors or delays in the signals sent through the V2V communication.

References

1. Alam, A., Gattami, A., Johansson, K.H., and Tomlin, C.J., "Guaranteeing safety for heavy duty vehicle platooning: Safe set computations and experimental evaluations," *Control Engineering Practice* 24:33–41, 2014, doi:10.1016/j.conengprac.2013.11.003.

2. Vahidi, A. and Sciarretta, A., "Energy saving potentials of connected and automated vehicles," *Transportation Research Part C: Emerging Technologies* 95:822–843, 2018, doi:10.1016/j.trc.2018.09.001.

3. Bergenhem, C., Shladover, S., Coelingh, E., Englund, C., and Tsugawa, S., "Overview of Platooning Systems," 2012.

4. Tsugawa, S., Jeschke, S., and Shladover, S.E., "A Review of Truck Platooning Projects for Energy Savings," *IEEE Trans. Intell. Veh.* 1(1):68–77, 2016, doi:10.1109/TIV.2016.2577499.

5. Humphreys, H. and Bevly, D., "Computational Fluid Dynamic Analysis of a Generic 2 Truck Platoon," 2016-01–8008, 2016, doi:10.4271/2016-01-8008.

6. Wu, C., Xu, Z., Liu, Y., Fu, C., Li, K., and Hu, M., "Spacing Policies for Adaptive Cruise Control: A Survey," *IEEE Access* 8:50149–50162, 2020, doi:10.1109/ACCESS.2020.2978244.

7. Gunter, G., Gloudemans, D., Stern, R.E., McQuade, S., Bhadani, R., Bunting, M., Delle Monache, M.L., Lysecky, R., Seibold, B., Sprinkle, J., Piccoli, B., and Work, D.B., "Are Commercially Implemented Adaptive Cruise Control Systems String Stable?," *IEEE Trans. Intell. Transport. Syst.* 22(11):6992–7003, 2021, doi:10.1109/TITS.2020.3000682.

8. Spano, M., Anselma, P.G., Musa, A., Misul, D.A., and Belingardi, G., "Optimal Real-Time Velocity Planner of a Battery Electric Vehicle in V2V Driving," *2021 IEEE Transportation Electrification Conference & Expo (ITEC)*, IEEE, Chicago, IL, USA, ISBN 978-1-72817-583-6: 194–199, 2021, doi:10.1109/ITEC51675.2021.9490121.

9. Stanger, T. and Del Re, L., "A model predictive Cooperative Adaptive Cruise Control approach," *2013 American Control Conference*, IEEE, Washington, DC, ISBN 978-1-4799-0178-4: 1374–1379, 2013, doi:10.1109/ACC.2013.6580028.

10. Musa, A., Miretti, F., and Misul, D., "MPC-Based Cooperative Longitudinal Control for Vehicle Strings in a Realistic Driving Environment," Detroit, Michigan, United States: 2023-01–0689, 2023, doi:10.4271/2023-01-0689.

11. Li, M., Cao, Z., and Li, Z., "A Reinforcement Learning-Based Vehicle Platoon Control Strategy for Reducing Energy Consumption in Traffic Oscillations," *IEEE Trans. Neural Netw. Learning Syst.* 32(12):5309–5322, 2021, doi:10.1109/TNNLS.2021.3071959.

12. Hussein, A.A. and Rakha, H.A., "Vehicle Platooning Impact on Drag Coefficients and Energy/Fuel Saving Implications," *IEEE Trans. Veh. Technol.* 71(2):1199–1208, 2022, doi:10.1109/TVT.2021.3131305.

13. Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D., Continuous control with deep reinforcement learning, 2019, doi:10.48550/arXiv.1509.02971.

14. Naus, G.J.L., Vugts, R.P.A., Ploeg, J., Van De Molengraft, M.J.G., and Steinbuch, M., "String-Stable CACC Design and Experimental Validation: A Frequency-Domain Approach," *IEEE Trans. Veh. Technol.* 59(9):4268–4279, 2010, doi:10.1109/TVT.2010.2076320.

15. Shi, H., Zhou, Y., Wu, K., Wang, X., Lin, Y., and Ran, B., "Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment," *Transportation Research Part C: Emerging Technologies* 133:103421, 2021, doi:10.1016/j.trc.2021.103421.

16. Vogel, K., "A comparison of headway and time to collision as safety indicators," *Accident Analysis & Prevention* 35(3):427–433, 2003, doi:10.1016/S0001-4575(02)00022-2.

17. Spano, M., Musa, A., Anselma, P.G., Misul, D.A., and Belingardi, G., "Battery Electric Vehicles Platooning: Assessing Capability of Energy Saving and Passenger Comfort

Improvement," *2021 AEIT International Conference on Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, IEEE, Torino, Italy, ISBN 978-88-87237-52-8: 1–6, 2021, doi:10.23919/AEITAUTOMOTIVE52815.2021.9662788.

Contact Information

Angelo Borneo, M.Eng.
Department of Energy "Galileo Ferraris" (DENERG)
Center for Automotive Research and Sustainable Mobility (CARS)
Politecnico di Torino, Corso Duca degli Abruzzi 24,
10129 Torino, Italy.
angelo.borneo@polito.it