Truck Platooning with Cooperative Adaptive Cruise Control using Reinforcement Learning

Hoseinali Borhan*
Stanford University
hborhan@stanford.edu
hoseinali.borhan@cummins.com

***Abstract*—Recent advancements in vehicle connectivity and advanced driver-assistance systems allow for more efficient driving in automated driving applications on the road. The practice of truck platooning utilizes following distances as small as a few meters of each vehicle in a string to benefit from slipstream effects and reduce aerodynamic drag. By this, fuel economy is then improved in the vehicles. In this project, a cooperative adaptive cruise controller is developed for 2 truck platooning using reinforcement learning techniques. A reward function is defined to meet the platooning operation including distance target tracking, and enforcing constraints on acceleration, minimum distance between trucks, and speed. The environment model includes a model of 2 trucks with acceleration as input and speed and distance as outputs. The speed of the lead truck in platoon is communicated to the trailing truck through wireless DSRC communication. In this study, reinforcement learning (RL) is applied to develop a cooperative adaptive cruise control for truck platooning application. Deep Deterministic Policy Gradient (DDPG) and Twin-Delayed Deep Deterministic Policy Gradient (TD3) methods are applied and compared to develop the control agent. These methods are recommended for applications with continuous actions as the platooning control in this study. Matlab and simulink are used for coding and simulation.**

***Index Terms*—truck platooning, cooperative adaptive cruise control, reinforcement learning, DDPG, TD3**

I. Introduction and Related Works

The effect of short range distance keeping between trucks on drag force, and consequently fuel economy, has been studied for several years. With the recent advancements in vehicle connectivity and advanced driver-assistance systems (ADAS), it is feasible to develop a coordination scheme between vehicles to form safe platoons on roads. As different levels of automation are considered for potential commercialization, there is still a need to explore new control methods to make the development, calibration and performance of the control system more efficient.

The overall control system objective is to maintain a close distance between vehicles in a platoon while still guaranteeing safety. If vehicles driving in a platoon keep a short distance from each other, drafting effects occur and the drag coefficient can be reduced by 8% to 50% - depending on their position in the platoon [1]. Moreover, several studies such as [2]–[4] suggest that even in distances longer than 50 meters between two trucks, there is still a considerable amount of reduction in drag and, consequently, increase in fuel economy. The close range driving has been achieved through either: 1) setting a constant reference distance [4]–[6] between trucks, or 2) setting a constant reference time gap between trucks [7]–[9]. To realize this, several control methods have been utilized in platooning before such as PID controllers [10]–[14] and optimal control methods such as Model Predictive Control [9], [15]–[18].

Fig. 1. Connected cruise control to form a two-truck platoon. Here, trailing vehicles benefit from drafting effects to reduce aerodynamic drag and thus save on energy.

The above classic control methods either rely on tuning of a fixed structure control design as in PID control or an embedded model of the system with real-time optimization as in model predictive control. While these control algorithms have solid theoretical background from stability to robustness performance, they require control theory expertise and knowledge of the system dynamics and characteristics to design the feedback control system. In this study, the machine learning based methods from reinforcement learning (RL) are applied for the truck platooning control development. Model-free, online, off-policy reinforcement learning methods including DDPG and TD3 are studied and compared for this application. Deep neural network models are trained as the function approximator for the actor-critic models. More details about DDPG and TD3 algorithms are available in [19], [20], [21], and [22].

II. Model Development

As shown in Figure 2, the model of the trucks in platooning operation are developed in Matlab Simulink. Matlab and Simulink are one of the main tools used in automotive industry for model-based control system engineering, software development, and auto code generation for embedded applications like ECU (engine control unit). The model includes the longitudinal vehicle dynamics to calculate vehicle velocity and position based on acceleration command. Higher fidelity

* I am a SCPD student doing my graduate certificate in artificial intelligence. I work as a technical project lead of control systems at Cummins Inc. This document is to report the project progress for CS221-Artificial Intelligence: Principles and Techniques course at Stanford university.

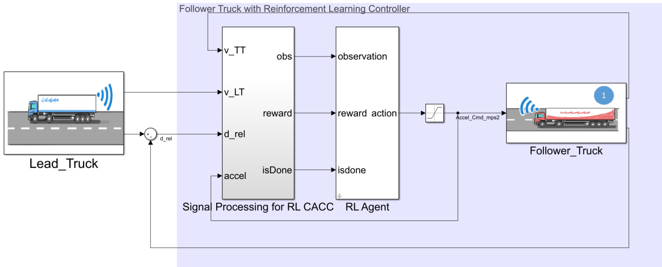

Fig. 2. Developed Truck Platooning Model in Simulink with RL Agent

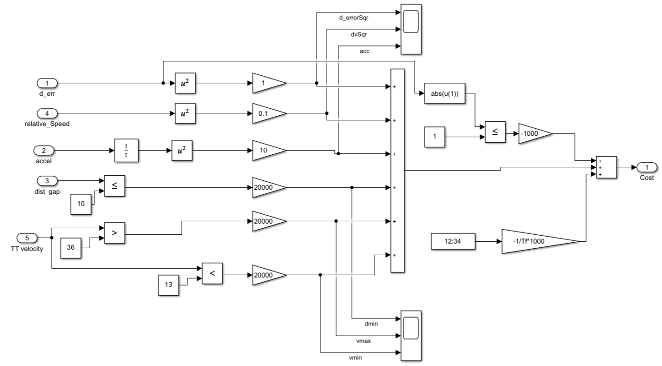

Fig. 3. Cost Model (note that the calculated cost is multiplied by -1 to compute reward based on an action of RL algorithm and states/observation of the system)

models to include engine and powertrain dynamics can be integrated in the system in the same framework. The governing differential equations of the system are solved by Simulink to predict the states and calculate reward/cost based on a given agent/controller action. The agent model is calculating the acceleration commands using reinforcement learning algorithms. Model-free, online, off-policy reinforcement learning agents including DDPG and TD3 are applied and compared as shown in the model.

The observations (states) for the learning agent include the distance (or gap) between the vehicles d (measured by radar sensor), the relative velocity $\delta v = v_{LT} - v_{TT}$ (lead truck speed minus trailing truck speed), the speed of trailing truck v_{TT} and the speed of lead truck v_{LT}. The trailing truck is controlled by the RL agent to follow the lead truck at the desired distance target d_r with minimum distance constraint enforced for safety during transient $d > d_{min}$. In practice, the speed of the lead truck is communicated to the trailing truck through vehicle to vehicle DSRC communication. As shown in Figure 3, the reward function is modeled to reduce the tracking error and to maintain a safe distance to the lead truck. Defining the reward function is one of the key steps in the RL design. This requires technical expertise and some preliminary simulation iterations to make sure the agent is rewarded for the right decisions. Besides tracking error control objectives, the agent is penalized for acceleration/deceleration commands (to improve fuel consumption). Furthermore, the agent is rewarded if the truck velocity and distance remain within a limit (to help the agent to stay within safe operation range), if distance is less than 1 meter (to reduce the steady-state tracking error) and if simulation time of episode is closer to the final time of simulation (to help agent to find solutions to complete the duty cycle).

Besides the states and reward model, an $IsDone$ function is added to stop the simulation during learning process if the distance or velocity are outside the feasible range (for example if distance is less than zero where accident happens).

III. Critic and Actor Models

Neural network models are designed for the critic and actor functions. The critic network structure is shown in Figure 4. For TD3 algorithms, the same critic neural network model structure is used for both critics. This model structure is also used for DDPG. Neural network model is also used for actor function in both DDPG and TD3 with the same structure. More details of the model structure and the options used for optimization, learning rate, regularization and other parameters are referred to the submitted source codes.

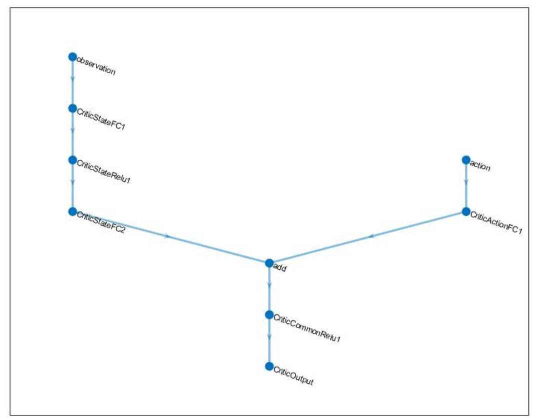

Fig. 4. Critic Network Structure

IV. Results

The DDPG and TD3 algorithms are applied to the truck platooning models as described above. The target distance is set to 17 meter. The lead truck speed is changing (representing the operation in traffic) and the goal of the agent is to follow the lead truck during this transient operation at the desired distance target. The initial distance between trucks are set to 38 meter, larger than the target distance. The results of the trained agent based on TD3 algorithm are shown in Figures 5, 6 and 7. It is observed that the agent is successfully able to meet the performance requirements. DDPG agent is also capable to meet the requirements after training is completed.

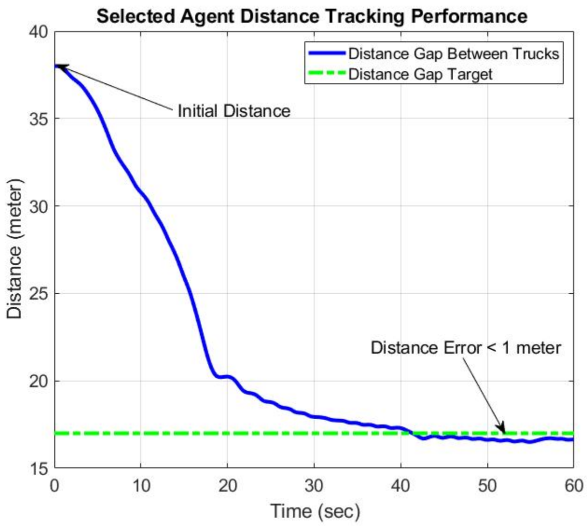

Fig. 5. Distance is converging to the target

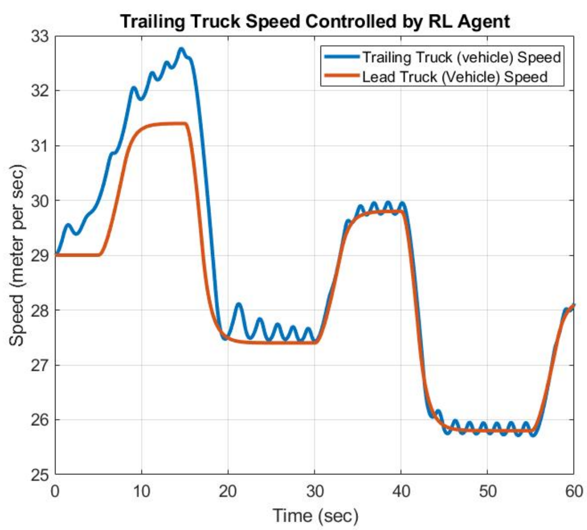

Fig. 6. The trailing truck speed versus lead truck change in speed

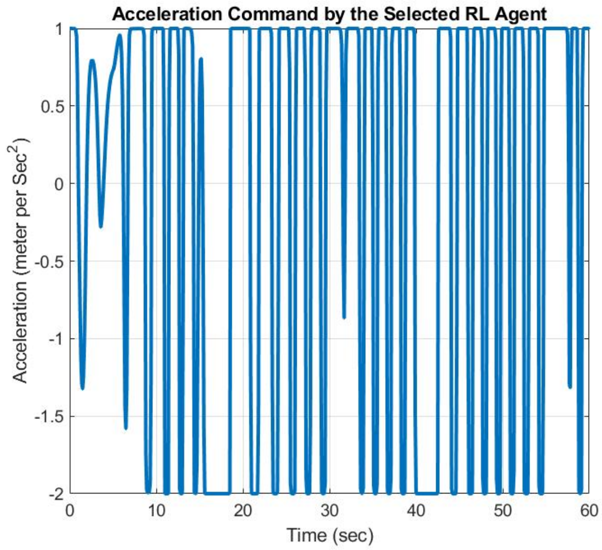

Fig. 7. The acceleration command by the agent during the episode after training

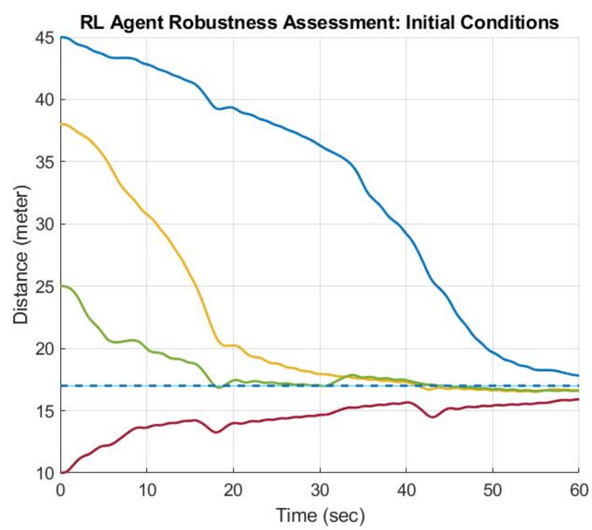

Fig. 8. Agent response as the initial condition for distance is changed.

V. Robustness Assessment

A. Initial condition and distance target variation impact

One of the key expectation of a control system is robustness to variation in operating conditions or model parameters. To assess the robustness of the trained agent, the initial conditions of the distance and also the distance target are changed without retraining of the agent. The results are shown in Figures 8 and 9. The results indicate that although the agent was not trained with different initial conditions or for different distance targets, it is still robust enough to meet the requirements.

B. High-fidelity model simulation with real-world traffic data

In this section, the TD3 trained agent was integrated in a high fidelity truck model and tested for a route with traffic impact. The route data is collected from trucks operating on a fleet in Ohio. The route information and logged data from the truck are shown in Figures 10 and 11.

The speed profile of this route is used to impose speed variation on lead truck. The agent is controlling the trailing truck to follow the lead truck at a desired time-varying distance target. Note that the agent is not trained for this route and on this high fidelity model (still the same agent from last section is deployed). The results of the vehicle, engine, and transmission are shown in Figure 12. It can be seen that the agent is still capable to follow the lead truck. This performance is significant as it is showing robustness of the trained agent when it is applied on a high fidelity model of the truck and on a real-world duty cycle. While tracking performance is acceptable, the agent is aggressive in commanding engine

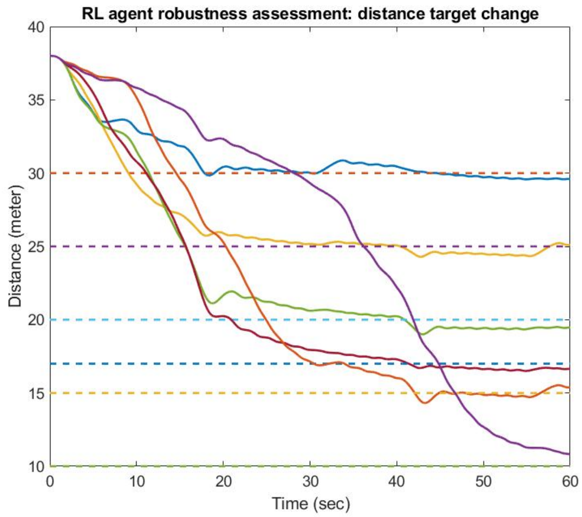

Fig. 9. Agent response as distance target is changed

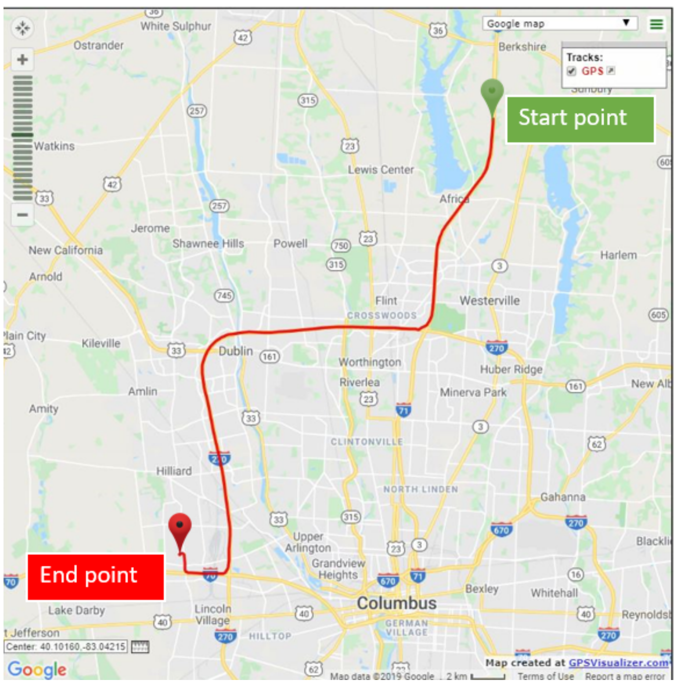

Fig. 10. Selected Route for Traffic Impact

and brake system to follow the lead truck comparing to a baseline control (the results of the baseline control cannot be shared due to confidential limitations). Retraining agent with higher cost on using acceleration command might improve the performance further. This is to be explored further in the next steps.

VI. DDPG and TD3 Learning Comparison

In Figure 13, the average award for each episode is compared for DDPG and TD3 algorithms. It is observed that TD3 is improving both learning speed and performance of DDPG for this application. Using 2 critic models in TD3 to reduce

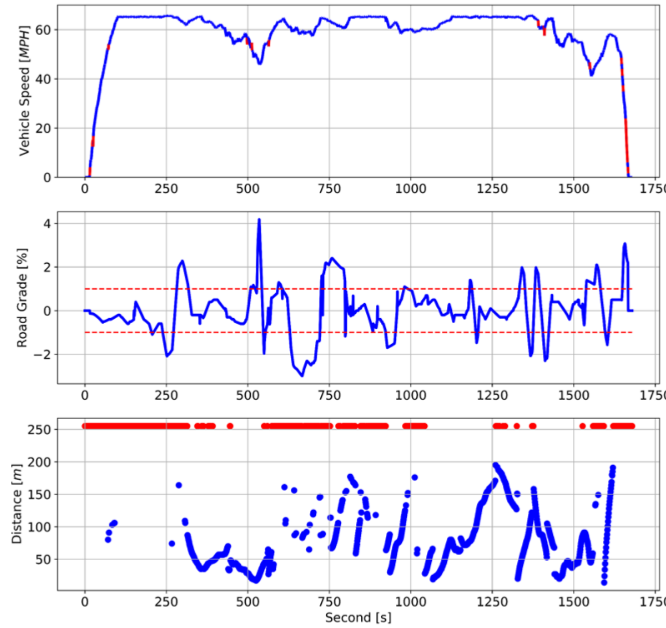

Fig. 11. Logged Data from a Truck Running on the Route with Traffic

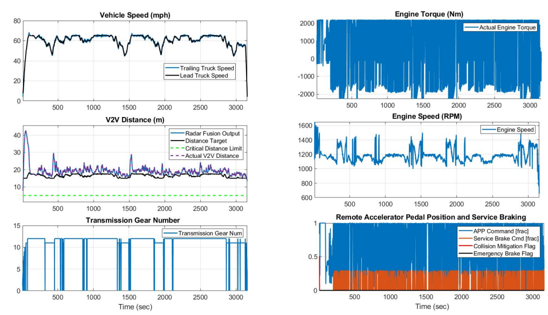

Fig. 12. Agent Performance on High Fidelity Truck Model with the Real-World Testing Scenario under Highway Traffic Conditions

overestimation bias issue of DDPG is one of the reasons for the improvement as described in [19].

VII. Summary

In this study, reinforcement learning (RL) is applied to develop a cooperative adaptive cruise control for truck platooning application. Deep Deterministic Policy Gradient (DDPG) and Twin-Delayed Deep Deterministic Policy Gradient (TD3) methods are applied and compared to develop the control agent. The results indicated better learning and performance results by TD3 comparing with DDPG. The trained agent could meet the performance requirements robustly even on a high fidelity model and a real-world test scenario. Engineering the reward function, defining the states and setting neural network architecture for actor and critic were key steps in achieving a successful agent to meet the requirements.

VIII. Next Steps

A few proposed next steps of the study are listed below:

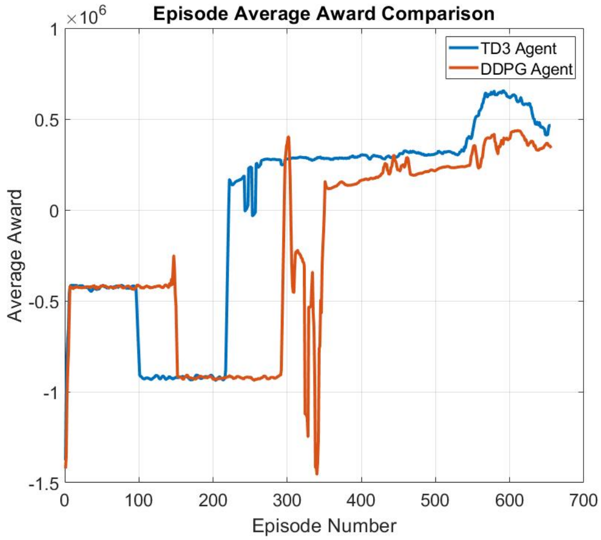

Fig. 13. DDPG versus TD3 learning performance

- Revising the reward model to reduce aggressive commanding of the agent and consequently fuel consumption improvement.
- Adding road grade to the observation states and train the agent again on a route with grade variation to assess performance on routes with road grade variation. Small grade variation has a significant impact on truck energy consumption due to high weight of these vehicles.
- Integrating fuel consumption, emission and drivability costs/reward to the reward function to train an agent to directly reduce fuel consumption and emission while meeting tracking and drivability requirements. This needs to be done with the high fidelity model of the trucks. Using parallel computation is important during training as simulation time is increased with higher fidelity models.
- Hyper parameter tuning of TD3 and DDPG algorithms to see the impact of different critic-actor settings on the performance.
- Testing the trained agent on an embedded control module in real-time.

References

[1] W. Hucho and G. Sovran, "Aerodynamics of road vehicles," *Annual review of fluid mechanics*, vol. 25, no. 1, pp. 485–537, 1993.

[2] A. Al Alam, A. Gattami, and K. H. Johansson, "An experimental study on the fuel reduction potential of heavy duty vehicle platooning," in *13th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2010, pp. 306–311.

[3] H. Humphreys and D. Bevly, "Computational fluid dynamic analysis of a generic 2 truck platoon," SAE Technical Paper, Tech. Rep., 2016.

[4] J. Smith, R. Mihelic, B. Gifford, and M. Ellis, "Aerodynamic impact of tractor-trailer in drafting configuration," *SAE International Journal of Commercial Vehicles*, vol. 7, no. 2014-01-2436, pp. 619–625, 2014.

[5] S. E. Shladover, C. Nowakowski, and X.-Y. Lu, "Using cooperative adaptive cruise control (CACC) to form high-performance vehicle streams. definitions, literature review and operational concept alternatives," *UC Berkeley: California Partners for Advanced Transportation Technology*, 2018.

[6] S. Tsugawa, S. Kato, and K. Aoki, "An automated truck platoon for energy saving," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 4109–4114.

[7] M. Gouy, K. Wiedemann, A. Stevens, G. Brunett, and N. Reed, "Driving next to automated vehicle platoons: How do short time headways influence non-platoon drivers' longitudinal control?" *Transportation research part F: traffic psychology and behaviour*, vol. 27, pp. 264–273, 2014.

[8] V. Turri, B. Besselink, and K. H. Johansson, "Cooperative look-ahead control for fuel-efficient and safe heavy-duty vehicle platooning," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 1, pp. 12–28, 2016.

[9] A. Alam, J. Mårtensson, and K. H. Johansson, "Experimental evaluation of decentralized cooperative cruise control for heavy-duty vehicle platooning," *Control Engineering Practice*, vol. 38, pp. 11–25, 2015.

[10] B. McAuliffe, M. Lammert *et al.*, "Influences on energy savings of heavy trucks using cooperative adaptive cruise control," SAE Technical Paper, Tech. Rep., 2018.

[11] J. Ploeg, B. T. Scheepers, E. Van Nunen, N. Van de Wouw, and H. Nijmeijer, "Design and experimental evaluation of cooperative adaptive cruise control," in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2011, pp. 260–265.

[12] A. Davila, E. Aramburu, and A. Freixas, "Making the best out of aerodynamics: Platoons," SAE Technical Paper, Tech. Rep., 2013.

[13] M. R. Nieuwenhuijze, T. van Keulen, S. Oncu, B. Bonsen, and H. Nijmeijer, "Cooperative driving with a heavy-duty truck in mixed traffic: Experimental results," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1026–1032, 2012.

[14] K. Lidstrom, K. Sjoberg, U. Holmberg, J. Andersson, F. Bergh, M. Bjade, and S. Mak, "A modular CACC system integration and design," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1050–1061, 2012.

[15] A. S. D. E. H. J. Liu, B. Pattel and H. Borhan, "Fuel efficient control algorithms for connected and automated line-haul trucks," in *2019 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE, 2019, pp. 730–737.

[16] R. Kianfar, B. Augusto *et al.*, "Design and experimental validation of a cooperative driving system in the grand cooperative driving challenge," *IEEE transactions on intelligent transportation systems*, vol. 13, no. 3, pp. 994–1007, 2012.

[17] F. Bu, H.-S. Tan, and J. Huang, "Design and field testing of a cooperative adaptive cruise control system," in *Proceedings of the 2010 American Control Conference*. IEEE, 2010, pp. 4616–4621.

[18] M. R. Jovanovic and B. Bamieh, "On the ill-posedness of certain vehicular platoon control problems," *IEEE Transactions on Automatic Control*, vol. 50, no. 9, pp. 1307–1321, 2005.

[19] H. v. H. Fujimoto, Scott and D. Mege, "Addressing function approximation error in actor-critic methods," *ArXiv:1802.09477*, 2018.

[20] Mathworks, "Mathworks rl toolbox: Td3 agent training," 2020. [Online]. Available: https://www.mathworks.com/help/reinforcement-learning/ug/td3-agents.html

[21] T. P. Lillicrap, "Continuous control with deep reinforcement learning," *International Conference on Learning Representations*, 2016.

[22] Mathworks, "Mathworks rl toolbox: Ddpg agent training," 2020. [Online]. Available: https://www.mathworks.com/help/reinforcement-learning/ug/ddpg-agents.html