
Cooperative Adaptive Cruise Control for Truck Platooning using Reinforcement Learning

Hoseinali Borhan (SCPD student)¹
Max Sobol Mark (Project Mentor)²

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 requirements 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 for this application.

1. Introduction

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 be-



Figure 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.

tween 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 (Hucho & Sovran, 1993). Moreover, several studies such as (Al Alam et al., 2010; Humphreys & Bevely, 2016; Smith et al., 2014) 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 (Shladover et al., 2018; Smith et al., 2014; Tsugawa et al., 2011) between trucks, or 2) setting a constant reference time gap between trucks (Gouy et al., 2014; Turri et al., 2016; Alam et al., 2015). To realize this, several control methods have been utilized in platooning before such as PID controllers (McAuliffe et al., 2018; Ploeg et al., 2011; Davila et al., 2013; Nieuwenhuijze et al., 2012; Lidstrom et al., 2012) and optimal control methods such as Model Predictive Control (J. Liu & Borhan, 2019; Kianfar et al., 2012; Bu et al., 2010; Jovanovic & Bamieh, 2005; Alam et al., 2015).

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

^{*}Equal contribution ¹Cummins Technical Center, Columbus, IN, USA ²Stanford University. Correspondence to: Hoseinali Borhan <hborhan@gmail.com>.

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 utilized and compared in terms of convergence rate and performance. The performance requirements include distance target tracking, and enforcing constraints on acceleration, minimum distance between trucks, and speed. Robustness to new operating conditions are considered and compared as an important requirement for vehicle automation. The environment model includes a model of 2 trucks with acceleration as input and speed and distance as outputs. More details about DDPG and TD3 algorithms are available in (Fujimoto & Mege, 2018), (Mathworks, 2020b), (Lillicrap, 2016), and (Mathworks, 2020a).

2. 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 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

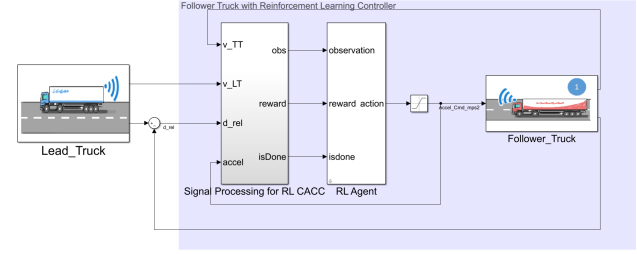


Figure 2. Developed Truck Platooning Model in Simulink with RL Agent

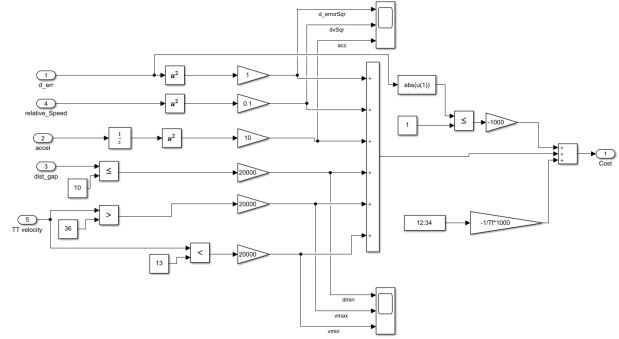


Figure 3. Cost Model (note that the calculated cost is multiplied by -1 to compute reward based on an action of RL algorithm and states/observation of the system)

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).

3. 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

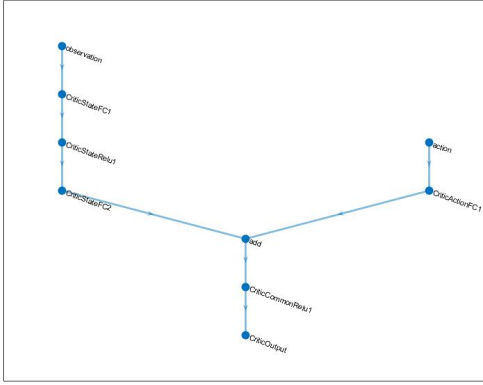


Figure 4. Critic Network Structure

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 source codes.

4. 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.

5. Robustness Assessment

One of the key expectations of a control system is robustness to variations in operating conditions or model parameters. To assess the robustness of the trained agent, different driving conditions are simulated and assessed. The performance of DDPG and TD3 learning are compared. The driving conditions to assess the robustness of the algorithms include:

- Initial conditions of the distance and also the distance target.
- High-fidelity model simulation with real-world traffic data: the trained agent will be integrated in a high fidelity truck model and tested for a route with traffic

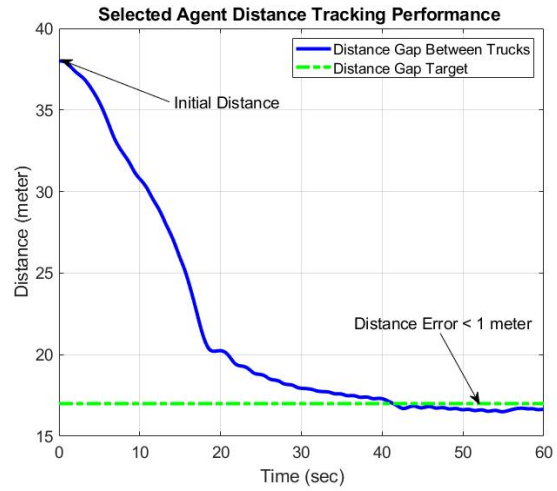


Figure 5. Distance is converging to the target

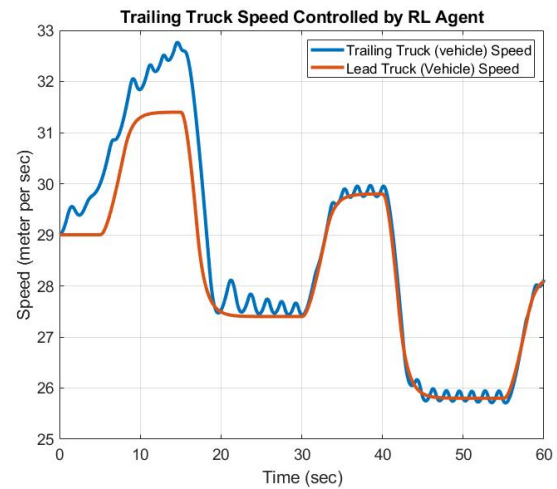


Figure 6. The trailing truck speed versus lead truck change in speed

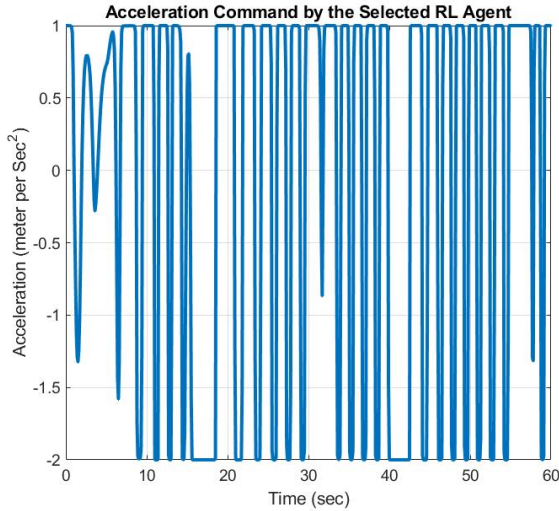


Figure 7. The acceleration command by the agent during the episode after training

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 8 and 9.

5.1. 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 10 and 11. 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.

5.2. 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 8 and 9.

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

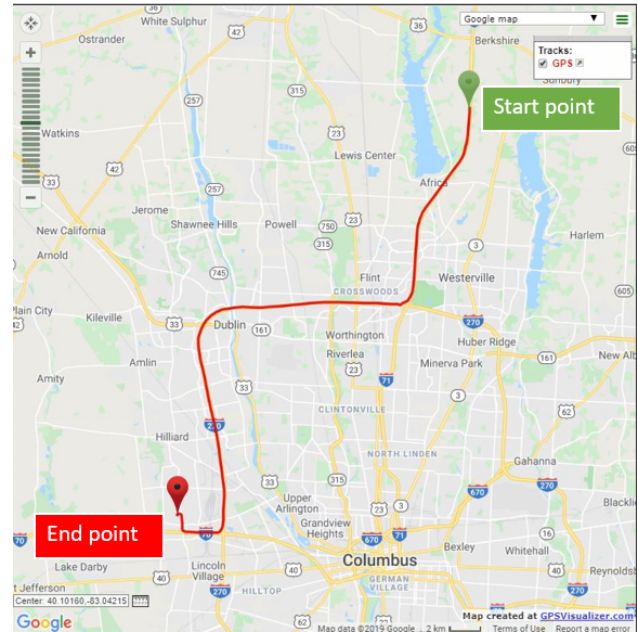


Figure 8. Selected Route for Traffic Impact

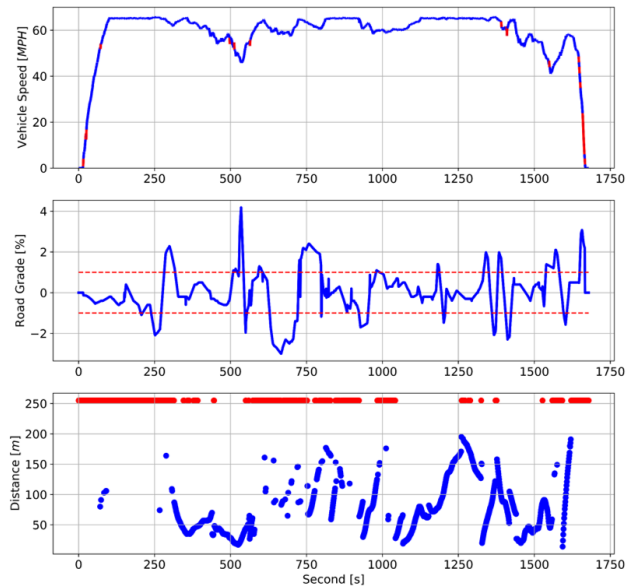


Figure 9. Logged Data from a Truck Running on the Route with Traffic

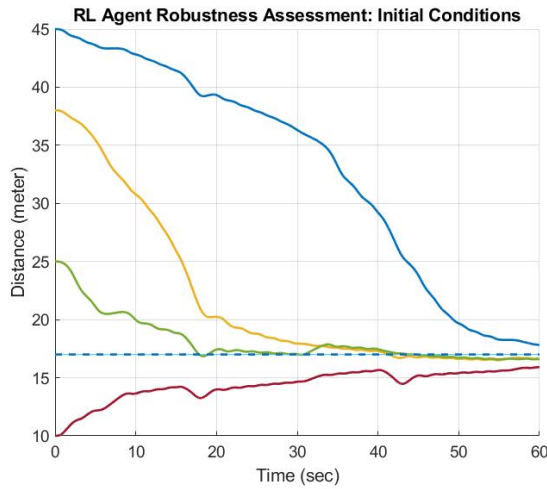


Figure 10. Agent response as the initial condition for distance is changed.

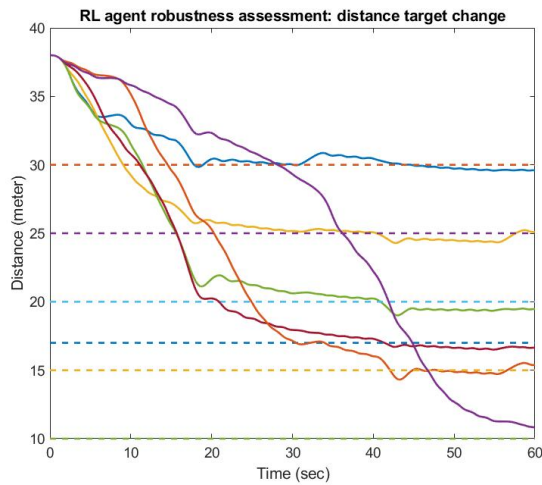


Figure 11. Agent response as distance target is changed

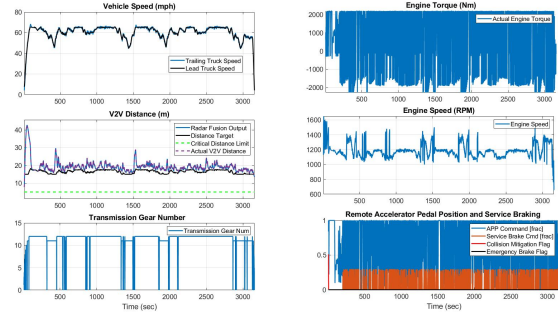


Figure 12. Agent Performance on High Fidelity Truck Model with the Real-World Testing Scenario under Highway Traffic Conditions

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 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 as the future steps of this study.

6. 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 overestimation bias issue of DDPG is one of the reasons for the improvement as described in (Fujimoto & Mege, 2018).

7. 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

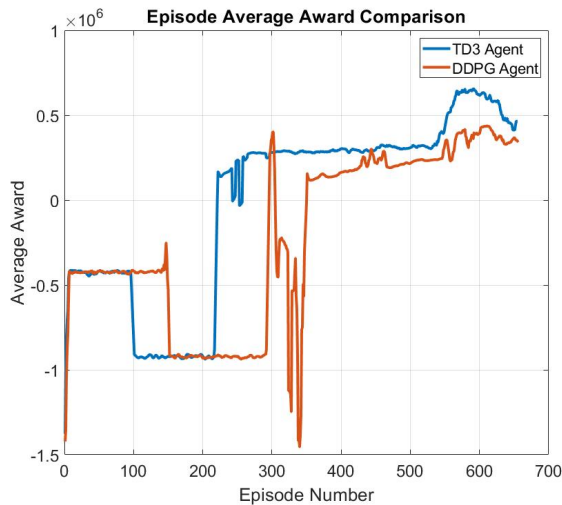


Figure 13. DDPG versus TD3 learning performance

the requirements.

8. Future Works

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

- 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.

9. Source Codes

The link to models and codes is <https://github.com/hborhan/Reinforcement-Learning-cs234-project>

References

- Al Alam, A., Gattami, A., and Johansson, K. H. An experimental study on the fuel reduction potential of heavy duty vehicle platooning. In *13th International IEEE Conference on Intelligent Transportation Systems*, pp. 306–311. IEEE, 2010.
- Alam, A., Mårtensson, J., and Johansson, K. H. Experimental evaluation of decentralized cooperative cruise control for heavy-duty vehicle platooning. *Control Engineering Practice*, 38:11–25, 2015.
- Bu, F., Tan, H.-S., and Huang, J. Design and field testing of a cooperative adaptive cruise control system. In *Proceedings of the 2010 American Control Conference*, pp. 4616–4621. IEEE, 2010.
- Davila, A., Aramburu, E., and Freixas, A. Making the best out of aerodynamics: Platoons. Technical report, SAE Technical Paper, 2013.
- Fujimoto, Scott, H. v. H. and Mege, D. Addressing function approximation error in actor-critic methods. *ArXiv:1802.09477*, 2018.
- Gouy, M., Wiedemann, K., Stevens, A., Brunett, G., and Reed, N. 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*, 27:264–273, 2014.
- Hucho, W. and Sovran, G. Aerodynamics of road vehicles. *Annual review of fluid mechanics*, 25(1):485–537, 1993.
- Humphreys, H. and Bevely, D. Computational fluid dynamic analysis of a generic 2 truck platoon. Technical report, SAE Technical Paper, 2016.
- J. Liu, B. Pattel, A. S. D. E. H. and Borhan, H. Fuel efficient control algorithms for connected and automated line-haul trucks. In *2019 IEEE Conference on Control Technology and Applications (CCTA)*, pp. 730–737. IEEE, 2019.
- Jovanovic, M. R. and Bamieh, B. On the ill-posedness of certain vehicular platoon control problems. *IEEE Transactions on Automatic Control*, 50(9):1307–1321, 2005.
- Kianfar, R., Augusto, B., et al. Design and experimental validation of a cooperative driving system in the grand cooperative driving challenge. *IEEE transactions on intelligent transportation systems*, 13(3):994–1007, 2012.

- Lidstrom, K., Sjoberg, K., Holmberg, U., Andersson, J., Bergh, F., Bjade, M., and Mak, S. A modular CACC system integration and design. *IEEE Transactions on Intelligent Transportation Systems*, 13(3):1050–1061, 2012.
- Lillicrap, T. P. Continuous control with deep reinforcement learning. *International Conference on Learning Representations*, 2016.
- Mathworks. Mathworks rl toolbox: Ddpg agent training, 2020a. URL <https://www.mathworks.com/help/reinforcement-learning/ug/ddpg-agents.html>.
- Mathworks. Mathworks rl toolbox: Td3 agent training, 2020b. URL <https://www.mathworks.com/help/reinforcement-learning/ug/td3-agents.html>.
- McAuliffe, B., Lammert, M., et al. Influences on energy savings of heavy trucks using cooperative adaptive cruise control. Technical report, SAE Technical Paper, 2018.
- Nieuwenhuijze, M. R., van Keulen, T., Oncu, S., Bonsen, B., and Nijmeijer, H. Cooperative driving with a heavy-duty truck in mixed traffic: Experimental results. *IEEE Transactions on Intelligent Transportation Systems*, 13(3):1026–1032, 2012.
- Ploeg, J., Scheepers, B. T., Van Nunen, E., Van de Wouw, N., and Nijmeijer, H. Design and experimental evaluation of cooperative adaptive cruise control. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 260–265. IEEE, 2011.
- Shladover, S. E., Nowakowski, C., and Lu, X.-Y. 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.
- Smith, J., Mihelic, R., Gifford, B., and Ellis, M. Aerodynamic impact of tractor-trailer in drafting configuration. *SAE International Journal of Commercial Vehicles*, 7(2014-01-2436):619–625, 2014.
- Tsugawa, S., Kato, S., and Aoki, K. An automated truck platoon for energy saving. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4109–4114. IEEE, 2011.
- Turri, V., Besselink, B., and Johansson, K. H. Cooperative look-ahead control for fuel-efficient and safe heavy-duty vehicle platooning. *IEEE Transactions on Control Systems Technology*, 25(1):12–28, 2016.