Coordinating Vehicle Platoons for Highway Bottleneck Decongestion and Throughput Improvement

Mladen Čičić, Xi Xiong, Li Jin, *Member, IEEE*, and Karl Henrik Johansson, *Fellow, IEEE*

***Abstract*— Truck platooning is a technology that is expected to become widespread in the coming years. Apart from the numerous benefits that it brings, its potential effects on the overall traffic situation need to be studied further, especially at bottlenecks and ramps. Assuming we can control the platoons from the infrastructure, they can be used as controlled moving bottlenecks, actuating control actions on the rest of the traffic, and potentially improving the throughput of the whole system. In this work, we use a tandem queueing model with moving bottlenecks as a prediction model to calculate control actions for the platoons. We use platoon speeds and formations as control inputs, and design a control law for throughput improvement of a highway section with a stationary bottleneck. By postponing and shaping the inflow to the bottleneck, we are able to avoid capacity drop, which significantly reduces the total time spent of all vehicles. We derived the estimated improvement in throughput that is achieved by applying the proposed control law, and tested it in a simulation study, with multi-class cell transmission model with platoons used as the simulation model, finding that the median delay of all vehicles is reduced by 75.6% compared to the uncontrolled case. Notably, although they are slowed down while actuating control actions, platooned vehicles experience less delay compared to the uncontrolled case, since they avoid going through congestion at the bottleneck.**

***Index Terms*— Bottleneck decongestion, Lagrangian traffic control, tandem queueing model, vehicle platooning.**

I. Introduction

WITH truck platooning progressing towards becoming a commonplace technology [1], understanding the impact it will have on the overall traffic is becoming increasingly important. Apart from its traditional primary purpose of providing potential fuel savings through air drag reduction [2], truck platooning is also expected have a positive impact on traffic efficiency through reducing the headways between vehicles [3], alleviating the adverse effect trucks have on the traffic [4]. Although there have been numerous field tests of truck platooning in real traffic [5], insufficient emphasis has been put on understanding how these platoons affect the behaviour of other vehicles on the road; thus the possible drawbacks of this technology are not yet fully understood [6].

One identified problem pertains to the interaction between truck platoons and passenger cars close to on- and off-ramps, and bottlenecks [7]. There is concern that long platoons might block access to an off-ramp, or from an on-ramp, resulting in significant disturbances for the traffic. Furthermore, the arrival of platoons can cause traffic breakdown at a bottleneck, causing reduction of throughput due to the capacity drop phenomenon. Recently, there have been efforts to address this problem in microscopic [8] and macroscopic [9] frameworks. In this paper, we are focusing on applying a new type of macroscopic control, using the truck platoons as actuators.

Bottleneck decongestion has long been tackled by classical traffic control measures, such as ramp metering [10] and variable speed limits [11]. However, these control methods require additional equipment to be installed upstream of the bottleneck, which limits their flexibility, especially for handling non-recurrent bottlenecks, such as work zones, incidents etc. as it is not reasonable to assume the required equipment would be available wherever such a bottleneck arises.

With the introduction of connected autonomous vehicles (CAVs) to the highways, new opportunities for sensing [12] and actuation [13] of the traffic are becoming available. While variable speed limits control benefits from the introduction of CAVs, its performance is significantly diminished when the controllable vehicles are only a small portion of all traffic [14]; thus a different control paradigm is needed. Lagrangian traffic control, where we use a subset of vehicles that can be controlled directly from the infrastructure as actuators, is lately garnering some attention [15], [16]. This approach, with actuator vehicles acting as controlled moving bottlenecks, can achieve a similar type of regulation as the classical traffic control, without the need for additional stationary equipment.

Due to their size and the existence of fleet management systems, truck platoons are ideal candidates for moving bottleneck

Manuscript received 23 March 2020; revised 17 October 2020 and 13 April 2021; accepted 18 May 2021. Date of publication 24 June 2021; date of current version 8 July 2022. This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant 674875, in part by the UM-SJTU Joint Institute, in part by the J. Wu and J. Sun Foundation, in part by the C2SMART University Transportation Center, NYU Tandon School of Engineering, in part by VINNOVA within the FFI Program under Contract 2014-06200, in part by the Swedish Research Council, in part by the Swedish Foundation for Strategic Research, and in part by the Knut and Alice Wallenberg Foundation. The work of Mladen Čičić and Karl Henrik Johansson was supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP). The Associate Editor for this article was N. Bekiaris-Liberis. *(Corresponding author: Mladen Čičić.)*

Mladen Čičić and Karl Henrik Johansson are with the Division of Decision and Control Systems, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden (e-mail: cicic@kth.se; kallej@kth.se).

Xi Xiong is with the C2SMART University Transportation Center, Department of Civil and Urban Engineering, NYU Tandon School of Engineering, Brooklyn, NY 11201 USA (e-mail: xi.xiong@nyu.edu).

Li Jin is with the School of Electronic Information and Electrical Engineering, University of Michigan-Shanghai Jiao Tong University Joint Institute, Shanghai 200240, China, and also with the NYU Tandon School of Engineering, Brooklyn, NY 11201 USA (e-mail: li.jin@sjtu.edu.cn).

Digital Object Identifier 10.1109/TITS.2021.3088775

TABLE I
SUITABILITY OF VARIOUS TRAFFIC MODEL FAMILIES

Model	Mov. bot.	Cap. drop	Ctrl. des.	Tractable
Microscopic	✓	✓	✗	✗
LWR	✓	✗	✗	✓
CTM	✓	✓	✓	✗
Queueing	✗	✓	✗	✓
Proposed	✓	✓	✓	✓

control. Since they consist of slow-moving vehicles, truck platoons act as moving bottlenecks with or without external control, which can be exploited for traffic control. This way, using platoons as actuators for regulating the traffic flow, we are able to mitigate the negative effects trucks have on the traffic, and even improve the overall traffic situation.

To this end, we need a suitable control-oriented prediction model to use for control design and calculation that is:

1) able to model platoons and moving bottlenecks,
2) able to model the capacity drop phenomenon,
3) conducive to control design, and
4) tractable when predicting over a long horizon.

We give an overview of which commonly used classes of traffic models fulfil each of these requirements in Table I. For a summary of state-of-the-art models for mixed-autonomy traffic, see [17]. Microscopic traffic models offer the highest level of detail, are able to capture all the complex behaviour, and allow for a straightforward representation of trucks and platoons [18], making them good simulation models. However, the complexity of simulating individual vehicles makes them untractable and hard to use for control design. First-order PDE traffic models, such as the Lighthill-Whitham-Richards (LWR) model, offer a consistent way of introducing moving bottlenecks [19], but are ill-suited for modelling capacity drop, even though fast algorithms for solving them exist [20], [21]. Various extensions of the Cell Transmission Model (CTM) have been proposed to tackle modelling moving bottlenecks [16] and capacity drop [22], as well as moving bottleneck control [16], [23], but require the use of short spatial and temporal discretization steps to describe the traffic with high resolution, necessitating a very high number of states and prediction steps. These models offer exact spatial characterization of congested areas, but if we can predict the evolution of queue lengths at stationary and moving bottlenecks, these details prove unnecessary in control design. Tandem queueing models like the fluid [9] and point queues [24] focus on queue lengths, and are very computationally simple as a result. It is also simple to model capacity drop in this framework, but in their basic form, these models do not consider moving bottlenecks which we intend to use as actuators.

Therefore, the main contribution of this work is in extending the tandem fluid queueing model to represent controlled moving bottlenecks in a way that is conducive to control design. Using the proposed prediction model, we design a control law for bottleneck decongestion using randomly arriving platoons as actuators, with their speed and formation as control inputs. We conduct stability analysis of the closed-loop system, and derive estimates for the achieved improved throughput.

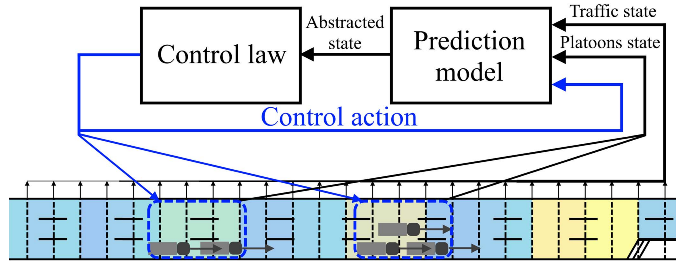


Fig. 1. Schematic representation of the control loop. We use the current traffic state and state of controllable platoons to calculate the control actions and improve the traffic situation.

The designed control law is tested in simulations on a road segment with one on-ramp and one off-ramp upstream of a bottleneck, and shown to achieve a significant reduction in total time spent, with the median delay of all vehicles reduced by 75.6%, compared to the case with no control.

The paper is structured as follows. In Section II, we discuss the overall control problem and propose a system architecture for solving it using connected vehicles. Next, in Section III, we present the simulation and prediction models that will be used. Then, in Section IV, we use the proposed prediction model to design control laws for improving the road throughput, and in Section V give a closed-loop stability analysis, as well as estimates on achieved throughput. Section VI describes the simulation set-up and results, and finally, in Section VII we conclude and discuss the results.

II. Lagrangian Traffic Control System

The general problem that we address in this paper is reduction of Total Time Spent (TTS) of all vehicles on the considered road segment. Typically, the flow of the whole road segment is constrained by some bottlenecks' capacity, which is further reduced once these bottlenecks get congested due to capacity drop. Therefore, maximizing the flow through the most severe bottleneck by decongesting it and keeping it in free flow will be required for minimizing TTS. More specifically, we study bottleneck decongestion using randomly arriving platoons as actuators, with their speed and formation as control inputs. We consider a group of vehicles, controlled to travel at the same speed in close proximity, to be a platoon. A representation of the control loop can be seen in Figure 1.

In this work we focus on a single stationary bottleneck, with one on-ramp and one off-ramp upstream of its location acting as disturbances to the traffic flow. Note that such segment can be used as a building block for a more complex road network. The traffic state is assumed to be known, and can be measured and observed using stationary sensors or connected vehicles. By communicating with platoon p, we may change its reference speed u_p and formation, which in turn affects the surrounding traffic by limiting the overtaking traffic flow to q_p^{cap} (the overtaking flow will be reduced if the platoon splits and drives side by side, instead of taking only one lane). In general, reducing u_p and taking multiple lanes makes the platoon act as a more severe moving bottleneck, and we assume that we can control the overtaking flow limit q_p^{cap} in some range. Therefore, we may use the platoons as controlled

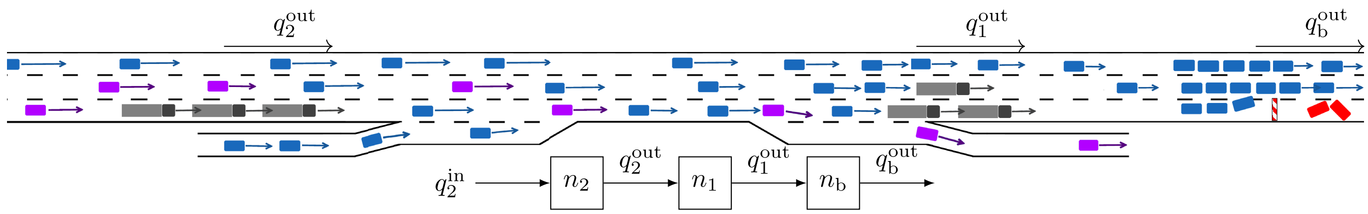


Fig. 2. Queues corresponding to stationary bottleneck n_{b}, downstream platoon n_1, and upstream platoon n_2. The overtaking flow of the downstream platoon q_1^{out} is limited to one lane, and the overtaking flow of the upstream platoon q_2^{out} to two lanes of traffic. Both the inflow from the on-ramp and the outflow to the off-ramp factor in the inflow to the downstream platoon queue q_1^{in}.

moving bottlenecks to control the inflow to the stationary bottleneck in order to decongest it. Using the proposed prediction model, which accounts for the effects of the control actions, we calculate optimal control that minimizes the total time spent by first restricting the inflow to the stationary bottleneck as much as possible, until its queue is dissipated, and then keeping its flow as close as possible to its capacity.

III. MODELS

In this section, we present the traffic models that will be used for simulation and prediction. The prediction model, tandem queueing model with moving bottlenecks, will be used for control design and analysis. The simulation model, multi-class CTM, is discussed briefly along with how it is connected to the prediction model. Control actions will be calculated using the prediction model, and then applied on the more detailed simulation model.

A. Simulation Model

In order for a traffic model to be suitable for use as a simulation model for the control problem we study, it needs to satisfy the first two requirements outlined in Table I. In case of real-world application, the actual traffic fills the role of the "simulation" model. Here we choose to use a multi-class extension of CTM as the simulation model, due to its relative simplicity and ease of use, while still able to represent the influence of platoons as moving bottlenecks, and capacity drop. This model is a variant of the one used in [25] and [23], extended to handle on- and off-ramps and multiple platoons, and its full formulation is given in the Appendix.

We assume that the current traffic density profile of the simulation model $\rho(x,t)$ at current time is known and available to the prediction model. We further link the prediction model by either analytically deriving or estimating from data the following parameters: free flow speed V, minimum enforceable platoon speed $U_{\min}$, stationary bottleneck capacity $q_{\text{b}}^{\text{cap}}$ and discharging flow $q_{\text{b}}^{\text{dis}}$, minimum enforceable overtaking flow at the moving bottlenecks Q^{lo}, maximum enforceable overtaking flow at the moving bottleneck that is lower than the stationary bottleneck capacity Q^{hi}, and average splitting ratio at the off-ramps R_k. The general road geometry, including the positions of the stationary bottleneck X_{b} and of the on- and off-ramps X_k^{r}, is also assumed to be known.

If the considered simulation model is deterministic with a constant free flow speed V for all vehicles everywhere on the road, and the road is in free flow initially, then the only place where congestion can emerge is at the bottlenecks. Where the road is in free flow and without on- and off-ramps, the traffic density profile propagates downstream at speed V, and we have $\rho(x,t+\theta) = \rho(x-V\theta,t)$. Therefore, the evolution of the full traffic state can be described using only the initial traffic density and the evolution of the queue lengths at the bottlenecks. We propose a model that exploits this fact in the following subsection.

B. Queueing Prediction Model

In this work, we study the situation when there is a single bottleneck at the downstream end of the considered stretch of highway, and want to predict its outflow based on the control action we chose for the platoons. Apart from this stationary bottleneck, platoons themselves can act as moving bottlenecks, since they will be moving slower than the rest of the traffic. We propose modelling this highway stretch using a queuing-based model, with queue length at the stationary bottleneck n_{b} and queue lengths at the platoons n_p, $p = 1, \ldots, \Pi$ as the only states. An example of a traffic situation with its corresponding queuing representation is shown in Figure 2, and an illustration of the derivation of the proposed model is given in Figure 3.

Since this model is used for predicting the evolution of traffic after sometime t_0, we assume that the current traffic situation is fully known and use it to predict the evolution of system states. We enumerate the platoons that are on the considered highway segment at $t = t_0$, $p = 1, \ldots, \Pi$, and denote their position at that time x_p, with $x_1 > x_2 > \ldots > x_\Pi$. Without loss of generality, we may set $t_0 = 0$, have t represents the prediction time shift after t_0, and write the current traffic density profile $\rho(x, t_0)$ as just $\rho(x)$.

The evolution of the queue at the bottleneck is given by

$$\dot{n}_{\text{b}}(t) = q_{\text{b}}^{\text{in}}(t) - q_{\text{b}}^{\text{out}}(t), \tag{1}$$

where the inflow and the outflow are

$$q_{\text{b}}^{\text{in}}(t) = q_{\text{b}}^{u}(t) + q_{\text{b}}^{V}(t), \tag{2}$$

$$q_{\text{b}}^{\text{out}}(t) = \begin{cases} q_{\text{b}}^{\text{in}}(t), & q_{\text{b}}^{\text{in}}(t) \le q_{\text{b}}^{\text{cap}} \wedge n_{\text{b}}(t) = 0, \\ q_{\text{b}}^{\text{dis}}, & q_{\text{b}}^{\text{in}}(t) > q_{\text{b}}^{\text{cap}} \vee n_{\text{b}}(t) > 0. \end{cases} \tag{3}$$

Typically, due to capacity drop, the discharge rate of the queue at the bottleneck $q_{\text{b}}^{\text{dis}}$ will be lower than its capacity $q_{\text{b}}^{\text{cap}}$, $q_{\text{b}}^{\text{dis}} < q_{\text{b}}^{\text{cap}}$. The first part of the inflow to the queue

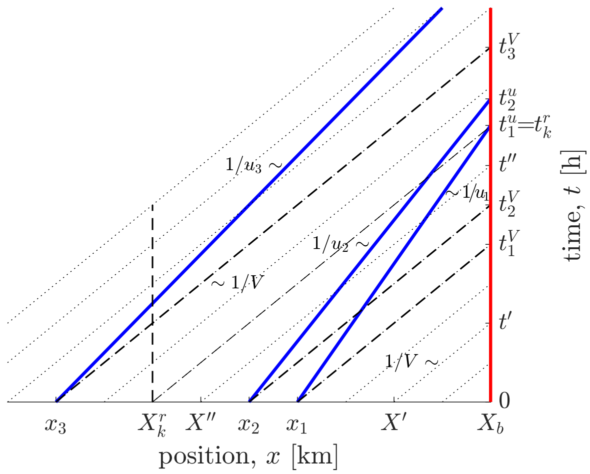


Fig. 3. Illustration of the queueing model. The dotted lines represent free flow propagation. Platoon trajectories are shown in blue. At $t = t'$, inflow to the bottleneck is $q_{\mathrm{b}}^{\mathrm{in}}(t') = V\rho(X')$. At $t = t''$, we have $q_{\mathrm{b}}^{\mathrm{in}}(t'') = \tilde{q}_1^{\mathrm{out}}(t'')$, and inflows to the platoons $q_1^{\mathrm{in}}(t'') = \tilde{q}_2^{\mathrm{out}}(t'')$, and $q_2^{\mathrm{in}}(t'') = V\rho(X'')$. Ramp k will affect $q_2^{\mathrm{in}}(t)$ for $\frac{X_{\mathrm{b}}-X_k^{\mathrm{r}}}{V} < t \leq t_2^u$, $q_3^{\mathrm{in}}(t)$ while $x_3^u(t) \geq X_k^{\mathrm{r}}$ and $t < t_3^u$, and $q_{\mathrm{b}}^{\mathrm{in}}(t)$ for the rest of time.

at the bottleneck $q_{\mathrm{b}}^{\mathrm{in}}(t)$ models the arrival of the platooned vehicles,

$$q_{\mathrm{b}}^{u}(t) = \begin{cases} V\sigma_l, & t_p^u \leq t \leq t_p^u + \dfrac{l_p}{V},\ p = 1, \ldots, \Pi, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

travelling at speed $u_p < V$, where $t_p^u = \frac{X_{\mathrm{b}}-x_p}{u_p}$ is the time at which platoon p reaches the bottleneck, X_{b} is the position of the stationary bottleneck, l_p is the length of platoon p, and σ_l is the traffic density of the platoon taking a single lane. The second part models the arrival of background traffic,

$$q_{\mathrm{b}}^{V}(t) = \begin{cases} q_p^{\mathrm{out}}\left(\dfrac{x_p + Vt - X_{\mathrm{b}}}{V-u_p}\right), & \max\left\{t_p^V, t_{p-1}^u\right\} \leq t \leq t_p^u, \\ & p = 1, \ldots, \Pi, \\ V\rho(X_{\mathrm{b}}-Vt), & \text{otherwise}, \end{cases}$$

travelling at free flow speed V, where $t_p^V = \frac{X_{\mathrm{b}}-x_p}{V}$. We assume that the speed of each platoon u_p is constant during the prediction horizon, and such that there is no platoon merging prior to reaching the bottleneck, $t_{p-1}^u > t_p^u$.

Under these assumptions, we define the evolution of the queue at each of the platoons $p = 1, \ldots, \Pi$ as

$$\dot{n}_p(t) = \frac{V-u_p}{V}\left(q_p^{\mathrm{in}}(t) - q_p^{\mathrm{out}}(t)\right), \quad 0 \leq t \leq t_p^u,$$

which is defined until time t_p^u, when the platoon reaches the bottleneck and their queues merge,

$$n_{\mathrm{b}}(t_p^u+) = n_{\mathrm{b}}(t_p^u) + n_p(t_p^u). \tag{5}$$

The outflow and inflow are defined as

$$q_p^{\mathrm{out}}(t) = \begin{cases} q_p^{\mathrm{in}}(t), & q_p^{\mathrm{in}}(t) \leq q_p^{\mathrm{cap}}(t) \wedge n_p(t) = 0, \\ q_p^{\mathrm{dis}}(t), & q_p^{\mathrm{in}}(t) > q_p^{\mathrm{cap}}(t) \vee n_p(t) > 0, \end{cases}$$

$$q_p^{\mathrm{in}}(t) = \begin{cases} q_{p+1}^{\mathrm{out}}\left(\dfrac{(V-u_p)t - x_p + x_{p+1}}{V-u_{p+1}}\right), & t > \dfrac{x_p - x_{p+1}}{V-u_p}, \\ V\rho(x_p - (V-u_p)t), & t \leq \dfrac{x_p - x_{p+1}}{V-u_p}, \end{cases}$$

where we assume that the queue dissipates at rate equal to its capacity $q_p^{\mathrm{dis}}(t) = q_p^{\mathrm{cap}}(t)$, and allow $q_p^{\mathrm{cap}}(t)$ to vary in time and be used as a control input.

The model can be simplified by adopting a coordinate transfer $\tau_p = \frac{x_p - X_{\mathrm{b}} + Vt}{V-u_p}$, for each platoon, which yields

$$\frac{\mathrm{d}n_p(t(\tau_p))}{\mathrm{d}\tau_p} = q_p^{\mathrm{in}}(t(\tau_p)) - q_p^{\mathrm{out}}(t(\tau_p)), \quad t_p^V \leq \tau_p \leq t_p^u$$

with $t = \frac{(V-u_p)\tau_p + X_{\mathrm{b}} - x_p}{V}$. Taking $\tilde{n}_p(\tau_p) = n_p(t(\tau_p))$, $\tilde{q}_p^{\mathrm{in}}(\tau_p) = q_p^{\mathrm{in}}(t(\tau_p))$, and $\tilde{q}_p^{\mathrm{out}}(\tau_p) = q_p^{\mathrm{out}}(t(\tau_p))$, we may write

$$\dot{\tilde{n}}_p(t) = \tilde{q}_p^{\mathrm{in}}(t) - \tilde{q}_p^{\mathrm{out}}(t), \quad t_p^V \leq t \leq t_p^u \tag{6}$$

for each $p = 1, \ldots, \Pi$. The inflow to the queue at the bottleneck and at platoons can now be simplified to

$$q_{\mathrm{b}}^{V}(t) = \begin{cases} \tilde{q}_p^{\mathrm{out}}(t), & \max\left\{t_p^V, t_{p-1}^u\right\} \leq t \leq t_p^u, \\ V\rho(X_{\mathrm{b}} - Vt), & \text{otherwise}, \end{cases}$$

$$\tilde{q}_p^{\mathrm{in}}(t) = \begin{cases} \tilde{q}_{p+1}^{\mathrm{out}}(t), & t_{p+1}^V < t < t_p^u, \\ V\rho(X_{\mathrm{b}} - Vt), & t \leq t_{p+1}^V, \end{cases}$$

and the outflow from the platoon becomes

$$\tilde{q}_p^{\mathrm{out}}(t) = \begin{cases} \tilde{q}_p^{\mathrm{in}}(t), & \tilde{q}_p^{\mathrm{in}}(t) \leq \tilde{q}_p^{\mathrm{cap}}(t) \wedge \tilde{n}_p(t) = 0, \\ \tilde{q}_p^{\mathrm{cap}}(t), & \tilde{q}_p^{\mathrm{in}}(t) > \tilde{q}_p^{\mathrm{cap}}(t) \vee \tilde{n}_p(t) > 0. \end{cases} \tag{7}$$

The influence of on- and off-ramps can be added to $q_{\mathrm{b}}^V(t)$ and $\tilde{q}_p^{\mathrm{in}}(t)$. Denoting $q_k^r(t)$ the inflow from an on-ramp (if $q_k^{\mathrm{r}}(t) > 0$), or outflow to an off-ramp (if $q_k^{\mathrm{r}}(t) < 0$), we have

$$q_{\mathrm{b}}^{V}(t) = q_{\mathrm{b}}^{V\setminus\mathrm{r}}(t) + \sum_{k \in K_o^d(t)} \tilde{q}_k^{\mathrm{r}}(t),$$

$$q_{\mathrm{b}}^{V\setminus\mathrm{r}}(t) = \begin{cases} \tilde{q}_p^{\mathrm{out}}(t), & \max\left\{t_p^V, t_{p-1}^u\right\} \leq t \leq t_p^u, \\ V\rho(X_{\mathrm{b}} - Vt), & \text{otherwise}, \end{cases}$$

$$K_o^d(t) = \begin{cases} K_p^b(t), & \max\left\{t_p^V, t_{p-1}^u\right\} \leq t \leq t_p^u, \\ K_p^b(t), & \text{otherwise}, \end{cases} \tag{8}$$

and for the inflow to the queue at platoons,

$$\tilde{q}_p^{\mathrm{in}}(t) = \tilde{q}_p^{\mathrm{in}\setminus\mathrm{r}}(t) + \sum_{k \in K_o^d(t)} \tilde{q}_k^{\mathrm{r}}(t),$$

$$\tilde{q}_p^{\mathrm{in}\setminus\mathrm{r}}(t) = \begin{cases} \tilde{q}_{p+1}^{\mathrm{out}}(t), & t > t_{p+1}^V, \\ V\rho(X_{\mathrm{b}} - Vt), & t \leq t_{p+1}^V, \end{cases}$$

$$K_o^d(t) = \begin{cases} K_{p+1}^p(t), & t > t_{p+1}^V, \\ K_p^p(t), & t \leq t_{p+1}^V. \end{cases} \tag{9}$$

Here, $\tilde{q}_k^{\mathrm{r}}(t) = q_k^{\mathrm{r}}\left(t - \frac{X_{\mathrm{b}} - X_k^{\mathrm{r}}}{V}\right)$, and $K_o^d(t)$ are sets of indices k of all on- and off-ramps with $X_k^{\mathrm{r}} < X_{\mathrm{b}}$ between the bottleneck or platoon p, and the place where their inflows originate from,

$$K_p^b(t) = \left\{ k \,\middle|\, x_p^u(t) < X_k^{\mathrm{r}} \le X_{\mathrm{b}}, t \ge t_k^{\mathrm{r}} \right\},$$
$$K_\rho^b(t) = \left\{ k \,\middle|\, X_{\mathrm{b}} - Vt < X_k^{\mathrm{r}} \le X_{\mathrm{b}}, t \ge t_k^{\mathrm{r}} \right\},$$
$$K_{p+1}^p(t) = \left\{ k \,\middle|\, x_{p+1}^u(t) < X_k^{\mathrm{r}} \le x_p^u(t), t \ge t_k^{\mathrm{r}} \right\},$$
$$K_\rho^p(t) = \left\{ k \,\middle|\, X_{\mathrm{b}} - Vt < X_k^{\mathrm{r}} \le x_p^u(t), t \ge t_k^{\mathrm{r}} \right\},$$

where $t_k^{\mathrm{r}} = \frac{X_{\mathrm{b}} - X_k^{\mathrm{r}}}{V}$ and we define $x_p^u(t)$ as

$$x_p^u(t) = \frac{u_p V t + V x_p - u_p x_p}{V - u_p}.$$

Note that $q_k^{\mathrm{r}}(t)$ will depend on the local traffic conditions around X_k^{r} at time t. Furthermore, since a portion of the queue at the platoon will also leave the road via the off-ramp, we reduce $\tilde{n}_p$ at the time when the platoon reaches it,

$$\tilde{n}_p(t+) = \tilde{n}_p(t) - n_p^{\mathrm{r},k}(t), \quad x_p^u(t) = X_k^{\mathrm{r}}, \tag{10}$$

and the part of the queue $\tilde{n}_p(t)$ that leaves the road, $n_p^{\mathrm{r},k}(t)$, depends on the ratio of off-ramp-bound vehicles in it.

In summary, the proposed model consists of $\Pi+1$ states, whose evolution is described by (1) and (6). Inflow to the bottleneck is given by (2), and consists of the background traffic travelling at free flow speed (8), and the platoons (4). Outflow from the bottleneck is (3), and there are discontinuous jumps in this state triggered by the arrival of platoons at the bottleneck, (5). For each platoon queue, inflow is given by (9), outflow by (7), and there is a discontinuous jump in the state when the platoon passes an off-ramp, (10). The model can be described as a tandem queuing system, with saturation and hysteresis, time-varying structure and jumps.

C. Validation

Finally, we validate the two used models, multi-class CTM and the tandem queueing model with moving bottlenecks, against microscopic traffic simulation done in SUMO, using an appropriate example scenario. Traffic density profiles in multi-class CTM and in SUMO (reconstructed according to vehicle trajectories) are shown in Figure 4. We study a 4 km stretch of road with a lane drop bottleneck at $X_{\mathrm{b}} = 3.75$ km, indicated by the vertical dashed red line. At the beginning of simulation, dense traffic enters the road, followed by sparser traffic and two controllable platoons, initially taking one lane. Once dense traffic reaches the bottleneck, congestion starts building up. At $t = 144$ s, both platoons are slowed down and commanded to take two lanes. This expedites the dissipation of the congestion at the bottleneck, and the platoons go back to taking one lane at $t = 216$ s, allowing the congestion that built up behind them to dissipate. Both times are indicated by horizontal dashed red lines.

Finally, in Figure 5 we show the comparison between the simulated queue length profiles, and the queue length prediction made using the proposed queueing prediction model, exhibiting similar behaviour. The prediction is made at time $t = 144$ s, using currently available traffic density data from the multi-class CTM simulation. The queue at the bottleneck grows at first, and is then dissipated by the platoons' control action. The queues at the platoons grow while they take two lanes, from $t = 144$ s to $t = 216$ s, and then decrease once they return to single lane formation. The congestion behind the first platoon does not get fully discharged, so it gets transferred to the queue at the bottleneck around $t = 230$ s. The discrepancies between the three queue profiles are mostly due to the difference in queue length definitions, using traffic density thresholds in case of multi-class CTM, and speed thresholds in case of SUMO, as well as due to stochasticity in lane-changing behaviour in case of the SUMO simulation.

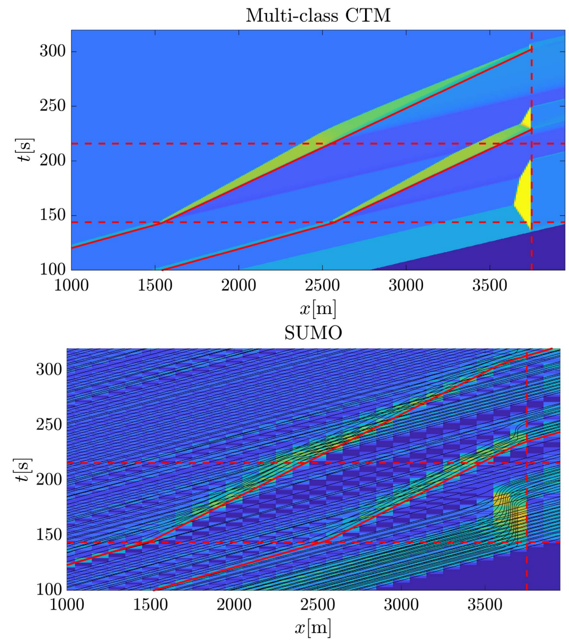


Fig. 4. Traffic density profiles comparison. Warmer colours represent denser traffic, thick red lines are the trajectories of platoons, and thin black lines are the trajectories of other individual vehicles.

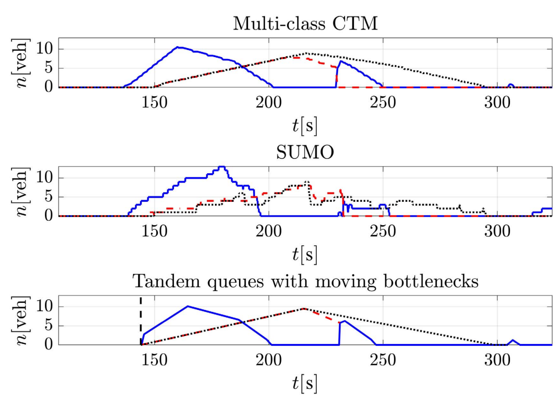


Fig. 5. Queue lengths comparison. Queue at the stationary bottleneck is shown in blue, queue at the first platoon in dashed red, and queue at the second platoon in dotted black.

IV. Control Design

Having defined the prediction model for the traffic system, in this section we will formulate a prediction-based control law for improving the throughput. We are looking to maximize the outflow from the bottleneck, which in case there are no off-ramps corresponds to minimizing the total travel time. In case there are off-ramps the total outflow of the mainstream and of the off-ramps needs to be maximized instead. We first consider the case when there are no on- or off-ramps and then extend the control to include on- and off-ramps.

As control inputs, we use the moving bottleneck speed $u_p(t)$, controlled by changing the reference speed of the platoons, and the overtaking flow limit $q_p^{\text{cap}}(t)$, controlled by changing the formation of the platoons, i.e. how many lanes they occupy. We are thus able to first help dissipate the congestion at the stationary bottleneck, by restricting the flow as much as possible, and then dissipate the congestion in the wake of the moving bottleneck, by reducing the moving bottleneck severity while making sure the stationary bottleneck remains in free flow. The proposed control laws will be described in the remainder of the section.

A. Platoon-Actuated Not Aware of On- or Off-Ramps

The control objective, maximizing the throughput, i.e., the outflow $q_{\text{b}}^{\text{out}}$, can be achieved by keeping $n_{\text{b}} = 0$ and $q_{\text{b}}^{\text{in}} = q_{\text{b}}^{\text{cap}}$. Additionally, we require that the queue at the platoon is already discharged when the platoon reaches the bottleneck, $n_p(t_p^u) = 0$. Therefore we employ control law

$$\tilde{q}_p^{\text{cap}}(t) = \begin{cases} q^{\text{ref}}(t), & n_{\text{b}}(t) = 0 \wedge t \geq t_{p-1}^u, \\ \tilde{q}_{p-1}^{\text{cap}}(t), & \tilde{n}_{p-1}(t) = 0 \wedge t < t_{p-1}^u, \\ Q^{\text{lo}}, & \text{otherwise}, \end{cases} \tag{11}$$

where the reference flow $q^{\text{ref}}(t)$ can be externally determined. For maximizing the throughput, we set $q^{\text{ref}}(t) = Q^{\text{hi}} - q_b^u(t)$, taking the largest admissible $Q^{\text{hi}} \leq q_{\text{b}}^{\text{cap}}$. In order to compute the current $q_p^{\text{cap}}(t) = \tilde{q}_p^{\text{cap}}(t_p^V)$ for all platoons, we need to predict n_{b} until t_{Π}^V, which requires calculating $q_{\Pi}^{\text{cap}}(0)$ and $q_p^{\text{cap}}(t)$ for $0 \leq t \leq \min\left\{t_p^u, t_{\Pi}^V\right\}$.

Assuming this control law is applied, we set the speed of each platoon so that $n_p(t_p^u) = 0$ and $n_{\text{b}}(t) = 0, t_p^c \leq t \leq t_p^u$, with minimum t_p^c such that $t_p^c \geq \max\{t_p^V, t_{p-1}^{u_{p-1}} + \frac{l_{p-1}}{V}\}$. This is achieved when

$$\tilde{n}_p(t_p^u) = \tilde{n}_p(t_p^c) + \int_{t_p^c}^{t_1^u} \tilde{q}_1^{\text{in}}(t)\mathrm{d}t - Q^{\text{hi}}(t_1^u - t_1^c) = 0. \tag{12}$$

For $p = 1$, in case it is known that $t_2^V < t_1^u$, (12) simplifies to

$$\tilde{n}_1(t_1^u) = \tilde{n}_1(t_2^V) + Q^{\text{lo}}(t_1^u - t_1^c) - Q^{\text{hi}}(t_1^u - t_1^c) = 0,$$
$$u_1 = \frac{\left(Q^{\text{hi}} - Q^{\text{lo}}\right)(X_{\text{b}} - x_1)}{\tilde{n}_1(t_2^V) + \left(Q^{\text{hi}} - Q^{\text{lo}}\right) t_1^c},$$

since we can explicitly calculate

$$\tilde{n}_1(t_2^V) = \int_{t_1^V}^{t_2^V} V\rho(X_{\text{b}} - Vt, 0)\mathrm{d}t - Q^{\text{lo}}(t_2^V - t_1^V).$$

Otherwise, u_p is calculated by solving (12) numerically, and can be obtained as a by-product of iterating the prediction steps for n_{b} and $\tilde{n}_p$. We may calculate u_p by initializing it to

$$u_p^{(0)} = \min\left\{U^{\max}, u_{p-1}\frac{X_{\text{b}} - x_p}{X_{\text{b}} - x_{p-1} + l_{p-1}}\right\},$$

and then decrease it until either $u_p = U^{\min}$ or (12) is satisfied. This also ensures that u_p is constrained to be within the range

$$U^{\min} \leq u_p \leq \min\left\{U^{\max}, u_{p-1}\frac{X_{\text{b}} - x_p}{X_{\text{b}} - x_{p-1} + l_{p-1}}\right\},$$

which is required if there is no platoon merging.

B. Platoon-Actuated Aware of On- or Off-Ramps

Consider now the case when there are on- or off-ramps. In order to exactly predict the evolution of queues, we need to know the ramp flows $\tilde{q}_k^{\text{r}}(t)$ in advance, which is very hard to ensure. Therefore, we use the predicted ramp flows instead.

If ramp k is an on-ramp, we can replace the actual ramp flow with its average $\hat{q}_k^{\text{r}} = \bar{q}_k^{\text{r}}$, which can be determined statistically. If ramp k is an off-ramp, we can assume that a constant ratio of vehicles R_k leave the road via the off-ramp,

$$\hat{q}_k^{\text{r}}(t) = -R_k\Big(\tilde{q}_k^{\text{in,r}}(t) + \sum_{l \in K_o^{\text{r},k}(t)} \tilde{q}_l^{\text{r}}(t)\Big),$$

$$\tilde{q}_k^{\text{in,r}}(t) = \begin{cases} q_{\text{b}}^V(t), & x_1^u(t) < x_k^{\text{r}} < X_{\text{b}} \\ \tilde{q}_{p+1}^{\text{out}}(t), & x_{p+1}^u(t) < x_k^{\text{r}} < x_p^u(t) \end{cases}$$

$$K_o^{\text{r},k}(t) = \begin{cases} \left\{l | x_1^u(t) < x_l^{\text{r}} < x_k^{\text{r}}\right\}, & t > t_1^V, x_k^{\text{r}} < x_{p-1}^u(t) \\ \left\{l | x_p^u(t) < x_l^{\text{r}} < x_k^{\text{r}}\right\}, & x_k^{\text{r}} < x_{p-1}^u(t), p > 1 \\ \left\{l | X_{\text{b}} - Vt < x_l^{\text{r}} < x_k^{\text{r}}\right\}, & \text{otherwise} \end{cases}$$

depending on the origin of the flow to off-ramp k at time t.

The portion of queue at platoon p that remains after the platoon has passed the off-ramp k can be estimated to be

$$\tilde{n}_p(t_p^{\text{r},k}+) = (1 - R_k)\tilde{n}_p(t_p^{\text{r},k}), \quad x_p^u(t_p^{\text{r},k}) = X_k^{\text{r}},$$

and we may now apply a control law similar to the one derived for the case when there are no on- and off-ramps. We modify (11) to take into account the fact that there might be some off-ramps $k \in K^*$ whose flow we do not want to obstruct. Since it is not possible to selectively allow the off-ramp-bound traffic to pass without also releasing the mainstream-bound traffic, we will only allow unrestricted flow towards those off-ramps by setting $\tilde{q}_p^{\text{cap}} = Q^{\text{hi}}$ if there are other platoons downstream that are regulating the inflow to the bottleneck,

$$\tilde{q}_p^{\text{cap}}(t) = \begin{cases} q^{\text{ref}}(t), & n_{\text{b}}(t) = 0 \wedge t \geq t_{p-1}^u, \\ Q^{\text{hi}}, & K_p^{p-1*}(t) \neq \emptyset \wedge t < t_{p-1}^u, \\ \tilde{q}_{p-1}^{\text{cap}}(t), & K_p^{p-1*}(t) = \emptyset \wedge \tilde{n}_{p-1}(t) = 0 \wedge t < t_{p-1}^u, \\ Q^{\text{lo}}, & \text{otherwise}, \end{cases} \tag{13}$$

where $K_p^{p*}(t) = K_{p+1}^p(t) \cap K^*$.

The platoon speeds are obtained in the course of predicting the queue evolution, as described in the previous subsection.

V. Analysis

In order to understand the effects and limitations this control law will have in realistic situations, we first study it in a simplified, idealised setting. In simulations the inflow of background traffic will vary in time, taking random values within some range, and platoons arrive with exponentially distributed gaps. Here, we first assume constant background traffic inflow $Q^{\text{in}}(t) = Q^{\text{in}}$ and periodic platoon arrivals, with period τ_π, and each platoon consisting of n_π passenger car equivalents, and later allow the inflow and gaps between two platoons vary within some range. In this section, we derive:

1) Exact limits on the maximum initial excess congestion for which the uncontrolled and controlled systems are stable, assuming constant inflow and periodic platoon arrivals,
2) The number of controlled platoons required to fully dissipate the congestion at a stationary bottleneck and return the road to the unperturbed free flow state, and
3) An estimate of throughput given varying inflow and gap between platoons, i.e., the average inflow for which we decongest the bottleneck with a predefined probability.

The stationary bottleneck has capacity $q_{\text{b}}^{\text{cap}}$, which is reduced to $q_{\text{b}}^{\text{dis}}$ in case there is capacity drop. We study the case when the bottleneck is initially congested. If there is no queue at the platoon and it arrives at a bottleneck in free flow, the platoon passes through without causing traffic breakdown. Otherwise, its vehicles are added to the bottleneck queue.

In summary, the system that we study in this section is

$$n_{\text{b}}(t_1^V) = \mu_0, \quad \dot{n}_{\text{b}}(t) = q_{\text{b}}^{\text{in}}(t) - q_{\text{b}}^{\text{out}}(t), \tag{14}$$

$$q_{\text{b}}^{\text{in}}(t) = \begin{cases} \tilde{q}_p^{\text{out}}, & \max\left\{t_p^V, t_{p-1}^u\right\} \le t \le t_p^u, \\ Q^{\text{in}}(t), & \text{otherwise,} \end{cases} \tag{15}$$

$$q_{\text{b}}^{\text{out}}(t) = \begin{cases} q_{\text{b}}^{\text{in}}(t), & q_{\text{b}}^{\text{in}}(t) \le q_{\text{b}}^{\text{cap}} \wedge n_{\text{b}}(t) = 0, \\ q_{\text{b}}^{\text{dis}}, & q_{\text{b}}^{\text{in}}(t) > q_{\text{b}}^{\text{cap}} \vee n_{\text{b}}(t) > 0, \end{cases} \tag{16}$$

$$\tilde{n}_p(t_p^V) = 0, \ \dot{\tilde{n}}_p(t) = \tilde{q}_p^{\text{in}}(t) - \tilde{q}_p^{\text{out}}(t), \ t_p^V < t < t_p^u, \tag{17}$$

$$\tilde{q}_p^{\text{in}}(t) = \begin{cases} \tilde{q}_{p+1}^{\text{out}}, & t_{p+1}^V < t < t_{p+1}^u, \\ Q^{\text{in}}(t), & t \le t_{p+1}^V, \end{cases} \tag{18}$$

$$\tilde{q}_p^{\text{out}}(t) = \begin{cases} \tilde{q}_p^{\text{in}}(t), & \tilde{q}_p^{\text{in}}(t) \le \tilde{q}_p^{\text{cap}}(t) \wedge \tilde{n}_p(t) = 0, \\ \tilde{q}_p^{\text{dis}}, & \tilde{q}_p^{\text{in}}(t) > \tilde{q}_p^{\text{cap}}(t) \vee \tilde{n}_p(t) > 0, \end{cases} \tag{19}$$

$$n_{\text{b}}(t_p^u+) = \begin{cases} n_{\text{b}}(t_p^u) + \tilde{n}_p(t_p^u) + n_\pi, & n_{\text{b}}(t_p^u) + \tilde{n}_p(t_p^u) > 0, \\ 0, & n_{\text{b}}(t_p^u) + \tilde{n}_p(t_p^u) = 0, \end{cases} \tag{20}$$

for $p = 1, \ldots, \Pi$, where $\tilde{q}_p^{\text{cap}}(t) \in [Q^{\text{lo}}, Q^{\text{hi}}]$ is set by control law (11). The platoon speed $u_p \in [U^{\min}, V]$ determines the time when the platoon reaches the bottleneck t_p^u.

A. Constant Inflow and Periodic Platoon Arrivals

We study the stability of the queue at the bottleneck under conditions of constant inflow and periodic platoon arrivals for different initial bottleneck queue lengths. First, in case no control is applied, i.e. $u_p = V$, $t_p^u = t_p^V$, and $\tilde{q}_p^{\text{cap}} = Q^{\text{hi}}$, the system under consideration simplifies to (14)–(16) and (20) with $\tilde{n}_p(t_p^u) = 0$. The system is stable if

$$Q^{\text{in}} + \frac{n_\pi}{\tau_\pi} < q_{\text{b}}^{\text{dis}},$$

i.e., if the total inflow is less than the bottleneck dissipating flow, the queue will dissipate regardless of its initial length.

If the platoons can be controlled, we are able to extend the range of Q^{in} for which the system is stable. In this case, it is of interest to study what is the maximum initial queue length μ_0 for which the system is stable for a given Q^{in}. The length of the considered road segment is ℓ and a platoon moving at speed u_k traverses it and reaches the bottleneck after $\tau_k^u = \frac{\ell}{u_k}$. Assuming the first platoon enter the road at time $t = 0$, we define the initial queue length $\mu_0 = n_{\text{b}}(\ell/V)$ as the queue length at the bottleneck at the time when the overtaking flow from the platoon reaches it. We say that there is μ_0 excess congestion to be dissipated, i.e. μ_0 vehicles need to be delayed in order for the bottleneck to return to free flow. For the first platoon entering the road segment, the entire congestion will be located at the bottleneck, and for subsequent platoons, the initial excess congestion μ_k will be distributed between the bottleneck and the platoons that entered the road previously.

We consider the case with flow values are arranged as

$$Q^{\text{lo}} < q_{\text{b}}^{\text{dis}} < Q^{\text{in}} < Q^{\text{in}} + \frac{n_\pi}{\tau_\pi} < Q^{\text{hi}} \le q_{\text{b}}^{\text{cap}}, \tag{21}$$

and the uncontrolled system is unstable.

The system is stable if $\mu_{k+1} < \mu_k$ until $\mu_k = 0$ for some k, i.e., if every subsequent platoon has less excess congestion to dissipate until the system returns to the unperturbed state. To find maximum μ_k for which this holds, we apply maximum control, i.e. $u_k = U^{\min}$ and maximum overtaking flow is Q^{lo} until the queue at the bottleneck is dissipated, which happens at $\tau_k^{\text{dis}} = \mu_k / \left(q_{\text{b}}^{\text{dis}} - Q^{\text{lo}}\right)$. Moving at minimum speed, a platoon will reach the bottleneck after $\tau^{\max} = \ell / U^{\min}$, so a necessary condition to be able to begin dissipating the congestion is that $\tau^{\max} > \tau_k^{\text{dis}}$, which yields

$$\mu_0 < \left(q_{\text{b}}^{\text{dis}} - Q^{\text{lo}}\right) \tau^{\max}.$$

The process of dissipating excess congestion can be split into two phases: saturation and recovery. In the saturation phase, maximum control action is applied and there is a queue at the platoons reaching the bottleneck. In the recovery phase, each subsequent platoon will have a higher speed, until the traffic returns to the unperturbed state.

In saturation phase, given $\mu_k \ge \mu_{Q^{\text{in}}}^{\text{sat}}$, the excess congestion left for platoon $k+1$ to dissipate will be

$$\mu_{k+1} = a\mu_k + b, \tag{22}$$

$$a = \frac{Q^{\text{hi}} - Q^{\text{lo}}}{q_{\text{b}}^{\text{dis}} - Q^{\text{lo}}} > 1, \tag{23}$$

$$b = \tau_\pi \left(Q^{\text{in}} - q_{\text{b}}^{\text{dis}}\right) + n_\pi - \tau^{\max} \left(Q^{\text{hi}} - q_{\text{b}}^{\text{dis}}\right) < 0, \tag{24}$$

where the minimum saturation phase excess congestion is

$$\mu_{Q^{\text{in}}}^{\text{sat}} = \frac{1}{a} \left(\left(Q^{\text{hi}} - Q^{\text{lo}}\right) \tau^{\max} - \left(Q^{\text{in}} - Q^{\text{lo}}\right) \tau_\pi \right). \tag{25}$$

Therefore, the excess congestion (22) will decrease if

$$\mu_k < \frac{b}{1-a}.$$

Starting with μ_0, we can calculate μ_k by recursing (22),

$$\mu_k = a^k\mu_0+\sum_{i=0}^{k-1} a^i b.$$

Note that since the discharging flow $q_{\mathrm{b}}^{\mathrm{dis}}$ by definition lies between Q^{lo} and Q^{hi}, we may represent it as

$$q_{\mathrm{b}}^{\mathrm{dis}} = \frac{a-1}{a}Q^{\mathrm{lo}}+\frac{1}{a}Q^{\mathrm{hi}},$$

where a is given by (23), and we have $\frac{a-1}{a} \in (0, 1)$ and $\frac{1}{a} \in (0, 1)$. We may regard a as a measure of capacity drop severity, with $a \approx 1$ indicating almost no capacity drop, and a high value of a indicating a severe capacity drop.

Given μ_0, the transition into the second phase of congestion dissipation happens after k^{sat} platoons, where k^{sat} is the lowest integer such that

$$a^{k^{\mathrm{sat}}}\mu_0+\sum_{i=1}^{k^{\mathrm{sat}}-1} a^i b \le \mu_{Q^{\mathrm{in}}}^{\mathrm{sat}},$$

The recovery phase is characterized by the lack of congestion at the stationary bottleneck, $n_{\mathrm{b}}(t) = 0$, i.e. all the congestion is in the queues at the platoons.

The minimum time when platoon k^{sat} can reach the bottleneck with no queue is

$$\tau_{k^{\mathrm{sat}}}^{u} = \frac{\mu_{k^{\mathrm{sat}}}}{q_{\mathrm{b}}^{\mathrm{dis}}-Q^{\mathrm{lo}}}+\tau_\pi\frac{Q^{\mathrm{in}}-Q^{\mathrm{lo}}}{Q^{\mathrm{hi}}-Q^{\mathrm{lo}}},$$

travelling at speed $u_{k^{\mathrm{sat}}} = \frac{\ell}{\tau_{k^{\mathrm{sat}}}^{u}}$. Since the stationary bottleneck will be in free flow, starting with $k = k^{\mathrm{sat}}$, μ_{k+1} follows

$$\mu_{k+1} = \mu_k-\tau_\pi\left(Q^{\mathrm{hi}}-Q^{\mathrm{in}}\right),$$

until for some $k = k^{\mathrm{rec}}$ we have $\mu_{k^{\mathrm{rec}}} \le \tau_\pi\left(Q^{\mathrm{hi}}-Q^{\mathrm{in}}\right)$, after which the traffic returns to the unperturbed state. Given $\mu_{k^{\mathrm{sat}}}$, we may calculate k^{rec} by rounding up

$$k^{\mathrm{rec}} = \left\lceil \frac{\mu_{k^{\mathrm{sat}}}}{\tau_\pi\left(Q^{\mathrm{hi}}-Q^{\mathrm{in}}\right)}\right\rceil.$$

The dynamics of μ_k through both phases of congestion dissipation can be jointly described as

$$\mu_{k+1} = \begin{cases} a\mu_k+b, & \mu_k \ge \mu_{Q^{\mathrm{in}}}^{\mathrm{sat}},\\ \mu_k-\tau_\pi\left(Q^{\mathrm{hi}}-Q^{\mathrm{in}}\right), & \tau_\pi\left(Q^{\mathrm{hi}}-Q^{\mathrm{in}}\right) \le \mu_k < \mu_{Q^{\mathrm{in}}}^{\mathrm{sat}},\\ 0, & \mu_k < \tau_\pi\left(Q^{\mathrm{hi}}-Q^{\mathrm{in}}\right),\end{cases}$$

with a and b given by (23) and (24).

We summarize the analysis in this proposition:

Proposition 1: Assuming constant inflow $Q^{\mathrm{in}}(t) = Q^{\mathrm{in}}$, periodic arrival of platoons with period τ_π and ordering of flow values (21)*, the queue length $n_{\mathrm{b}}(t)$ of system* (14)–(20) *controlled by control law* (11) *is stable and will remain zero after sometime t, if the initial queue length satisfies*

$$\mu_0 < \frac{b}{1-a},$$

where a and b are given by (23) *and* (24)*, respectively. Furthermore, if this condition is satisfied, the system returns to the unperturbed state with $n_{\mathrm{b}}(t) = 0$ and $\tilde{n}_p(t) = 0$ after platoon k^{rec} reaches the bottleneck.*

Conversely, substituting $b = (1-a)\mu_0$ into (24), we may derive the maximum Q^{in} for which the system will be stable for a given μ_0,

$$Q_{\mu_0}^{\mathrm{in}} = q_{\mathrm{b}}^{\mathrm{dis}}-\frac{n_\pi}{\tau_\pi}+\frac{a-1}{a}(Q^{\mathrm{hi}}-Q^{\mathrm{lo}})\frac{\tau^{\max}}{\tau_\pi}-(a-1)\frac{\mu_0}{\tau_\pi}. \quad (26)$$

B. Varying Inflow and Platoon Arrivals

The uncertainty coming from the varying inflow of background traffic and random platoon arrivals can be modelled by adding another term to (22):

$$\mu_{k+1} = a\mu_k+b+\varepsilon_k, \quad (27)$$

where $\varepsilon_k = \varepsilon_k^{\tau_\pi}(Q^{\mathrm{in}}-q_{\mathrm{b}}^{\mathrm{dis}})+\tau_\pi\varepsilon_k^{Q^{\mathrm{in}}}+\varepsilon_k^{\tau_\pi}\varepsilon_k^{Q^{\mathrm{in}}}$ represents the disturbance, i.e. the aggregate deviation from the average queue length update, $\varepsilon_k^{\tau_\pi}$ is the difference of the gap between platoon $k-1$ and k from τ_π, and $\varepsilon_k^{Q^{\mathrm{in}}}$ is the difference of the average inflow from Q^{in} during that time. We may also write

$$\mu_k = a^k\mu_0+\sum_{i=0}^{k-1} a^{k-1-i}\left(\varepsilon_i+b\right).$$

Proposition 2: Assuming $|\varepsilon_k| < E < |b|$, if for any k we have

$$\mu_k < \frac{b+E}{1-a}, \quad (28)$$

with a given by (23) *and b by* (24)*, then system* (27) *is stable, and we are able to dissipate the congestion at the bottleneck. Conversely, if for any k we have*

$$\mu_k > \frac{b-E}{1-a}, \quad (29)$$

then system (27) *is unstable, and the queue at the bottleneck will grow unbounded.*

Consequently the conclusions about stability can be extended to system (14)–(20) if a suitable bound on uncertainty E can be derived.

For the initial excess congestion between these two values, $\frac{b+E}{1-a} < \mu_0 < \frac{b-E}{1-a}$, μ_k will almost surely satisfy either condition (28) or (29) for some k. Assuming uniformly distributed ε_k, with $\mathbb{E}\{\varepsilon_k\} = 0$, $\mathrm{Var}\{\varepsilon_k\} = \frac{E^2}{3}$, the probability of μ_k satisfying (29) (i.e., failing to decongest the bottleneck) closely follows the logistic curve depending on μ_0,

$$\mathcal{P}_{\mathrm{uns}}(\mu_0) \approx \left(1+\exp\left(\frac{\frac{b}{1-a}-\mu_0}{\frac{E}{4}}\right)\right)^{-1}$$

and the probability of μ_k satisfying (28) for some k (i.e., decongesting the bottleneck) is $\mathcal{P}_{\mathrm{sta}}(\mu_0) = 1-\mathcal{P}_{\mathrm{uns}}(\mu_0)$.

Finally, we may define the estimate of throughput of the controlled system as the maximum Q^{in} for which the control algorithm decongests the bottleneck with probability $\mathcal{P}^*$, given an appropriately chosen μ_0. This yields the bound on Q^{in},

$$Q^{\text{in}}_{\mu_0,\Delta} = q^{\text{dis}}_{\text{b}} - \frac{n_\pi}{\tau_\pi} + \frac{Q^{\text{hi}} - q^{\text{dis}}_{\text{b}}}{\tau_\pi}\left(\tau^{\max} - \frac{\mu_0 + \Delta}{q^{\text{dis}}_{\text{b}} - Q^{\text{lo}}}\right),$$
$$Q^{\text{in}}_{\mu_0,\Delta} = q^{\text{dis}}_{\text{b}} - \frac{n_\pi}{\tau_\pi} + \frac{a-1}{a}(Q^{\text{hi}} - Q^{\text{lo}})\frac{\tau^{\max}}{\tau_\pi} - (a-1)\frac{\mu_0 + \Delta}{\tau_\pi}, \tag{30}$$

where Δ is the measure of combined uncertainty in the system,

$$\Delta = \frac{E}{4}\log\left(\frac{\mathcal{P}^*}{1 - \mathcal{P}^*}\right).$$

Moreover, if we substitute $\Delta = 0$ into (30), we recover (26), i.e. the deterioration of $Q^{\text{in}}_{\mu_0,\Delta}$ due to the introduction of varying inflow and platoon arrivals is

$$Q^{\text{in}}_{\mu_0,\Delta} - Q^{\text{in}}_{\mu_0} = -(a-1)\frac{\Delta}{\tau_\pi}.$$

Note that this bound is only valid if the dissipation process starts in the saturation phase, $\mu_0 > \mu^{\text{sat}}_{Q^{\text{in}}_{\mu_0,\Delta}}$. Substituting (30) into (25), we find the minimum μ_0 for which this holds,

$$\mu^{\text{sat}}_0 = \frac{\left(Q^{\text{hi}} - Q^{\text{lo}}\right)\left(\tau^{\max} - \tau_\pi\right)}{a} + n_\pi + (a-1)\Delta,$$

in which case we have

$$Q^{\text{in}}_{\mu^{\text{sat}}_0,\Delta} = Q^{\text{hi}} - a\frac{n_\pi}{\tau_\pi} - a(a-1)\frac{\Delta}{\tau_\pi}. \tag{31}$$

VI. Simulation Results

In order to assess the performance of the proposed control laws, we conducted a number of simulation runs, results of which will be presented in this section. We demonstrate the control laws' effectiveness and show that we are able to eliminate 52.7% of the total delay due to congestion experienced by all vehicles on average, or 75.6% of the total delay by median, compared to the uncontrolled case.

The simulations were executed on a 5 km long stretch of highway, illustrated by Figure 2, with an on-ramp around the 2 km mark, and an off-ramp around the 3 km mark. Most of the highway stretch has three lanes, corresponding to a critical density of $\sigma_- = 60$ veh/km and capacity of $Q^{\max}_- = 6000$ veh/h, with free flow speed of $V = 100$ km/h. There is a bottleneck caused by an accident 80 m upstream of the end of the considered stretch, with capacity of $Q^{\max}_+ = 4000$ veh/h. The capacity drop phenomenon is modelled with $\alpha = 0.4$, which causes the bottleneck capacity to be reduced to $Q^{\text{dis}}_+ = 3273$ veh/h, representing a 18.2% capacity drop.

The simulation model we used was the multi-class CTM, given in the Appendix, with three classes of traffic: class a consists of the platoons we control, class b is the mainstream-bound background traffic, and class c the off-ramp-bound background traffic. The arrival of class a vehicles is modelled as Poisson process with arrival rate of $\lambda = 81$ platoons/h, $\tau_\pi = 0.0123$ h. We assume that each platoon consists of 2 passenger car equivalents, corresponding to about three trucks due to shorter inter-vehicular gaps. The minimum platoon reference speed is set to $U^{\min} = 50$ km/h. The inflow of background traffic is assumed to be time-varying and uniformly distributed, changing every 14.4 seconds. At the beginning of the highway segment, the inflow of mainstream-bound background traffic takes values in $\phi^b_1(t) \sim \mathcal{U}(1000, 2000)$ veh/h, and the inflow of off-ramp bound traffic is $\phi^c_1(t) \sim \mathcal{U}(750, 1250)$ veh/h. All of the on-ramp traffic is assumed to be mainstream-bound, and is modelled as $\phi^b_{i_{\text{on}}}(t) \sim \mathcal{U}(900, 1500)$ veh/h.

With the parameters specified in previous paragraph, we may calculate an estimate of the throughput achievable by applying the presented control law. Using (31) with $E = \tau_\pi\left(\max\left(\phi^b_1 + \phi^b_{i_{\text{on}}}\right) - \mathbb{E}\left\{\phi^b_1 + \phi^b_{i_{\text{on}}}\right\}\right)$ and $\mathcal{P}^* = 0.9$, yielding $\Delta \approx 5.4$, we estimate that the throughput would be improved from $Q^{\text{in}}_{\text{unc}} + \frac{n_\pi}{\tau_\pi} = 3273$ veh/h to $Q^{\text{in}}_{\mu^{\text{c}}_0,\Delta} + \frac{n_\pi}{\tau_\pi} = 3513.2$ veh/h. Note that in deriving (31) we do not take into account the existence of the on-ramp.

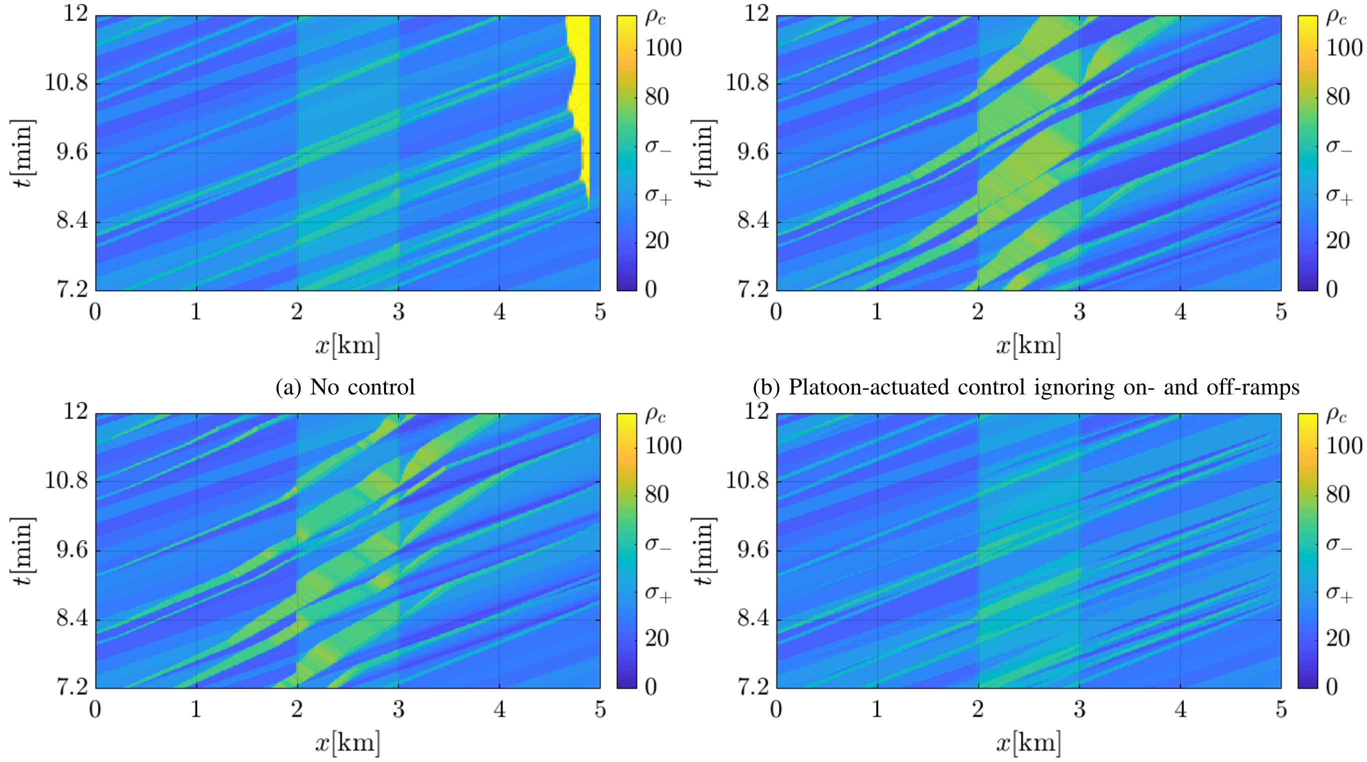
The duration of each simulation run is 2 hours, of which the background traffic inflow is halved for the first 3 minutes, in order to properly initialize the system, and for the last 12 minutes, in order to allow the traffic to return to free flow. Simulations are done with four cases of control:

(a) No control,
(b) Platoon-actuated control ignoring on- and off-ramps from Subsection IV-A, with $\tilde{q}^{\text{cap}}_p(t)$ given by (11),
(c) Platoon-actuated control accounting for on- and off-ramps from Subsection IV-B, with $\tilde{q}^{\text{cap}}_p(t)$ given by (13), and
(d) Ideally actuated control from the Appendix, with $U^b_i(t)$ given by (32), used as benchmark.

We demonstrate the effect these control laws have on the traffic on a part of a simulation run shown in Figure 6.

Consider the uncontrolled case shown in Figure 6a. Around time $t = 8.6$ min, the aggregate flow of platooned vehicles and background traffic arriving at the bottleneck exceeds bottleneck capacity. This causes a traffic breakdown, congestion is formed and bottleneck capacity is reduced. Therefore, even though the incoming flow is lower after $t = 9.2$ min, and does not exceed the original bottleneck capacity, it is not enough to dissipate the congestion at the bottleneck. Consequently, the throughput is reduced, the total time spent significantly increased, and the bottleneck will remain congested throughout the simulation run. In contrast, in the ideally actuated case shown in Figure 6d, a part of the mainstream-bound background traffic is delayed to allow the platoons to traverse the bottleneck without a traffic breakdown, maintaining free flow with throughput close to its theoretical maximum.

As shown in Figure 6b and Figure 6c, the performance of the two proposed control laws is similar. However, in case of the control from Section IV-A, the applied control is stronger than required, resulting in more congestion upstream of the off-ramp and lower efficiency. The control law from Section IV-B comes close to the ideal actuation case, with somewhat worse performance because it is unable to selectively affect mainstream-bound traffic, only has access to the average off-ramp splitting ratio, and requires delaying the platoons.



(a) No control

(b) Platoon-actuated control ignoring on- and off-ramps

(c) Platoon-actuated control taking on- and off-ramps into account

(d) Ideally actuated control

Fig. 6. An example comparing the four simulation cases. Traffic density is color-coded, with warmer color representing higher density.

TABLE II
AVERAGE AND MEDIAN TTS AND DELAY FOR EACH VEHICLE CLASS, AND ALL VEHICLES COMBINED

	Class a (platoons)		Class b (mainstream)		Class c (off-ramp)		Total	
	average	median	average	median	average	median	average	median
Case (a) TTS [veh h]	22.6	22.9	369.8	374.1	56.6	56.0	449.1	453.9
Case (b) TTS [veh h]	23.2	23.0	329.9	315.8	60.6	60.4	413.8	398.2
Case (c) TTS [veh h]	21.8	21.3	304.9	278.6	58.6	58.4	385.3	357.6
Case (d) TTS [veh h]	17.0	16.9	255.0	254.1	55.9	55.9	327.9	326.4
Delay 1, case (a) vs case (d) [%]	33.1	35.3	45.0	46.9	1.2	0.0	36.9	38.3
Delay 2, case (b) vs case (d) [%]	36.8	33.5	29.4	24.1	8.4	8.0	26.2	21.7
Delay 3, case (c) vs case (d) [%]	28.1	20.7	19.6	8.6	4.8	4.4	17.5	8.4

We executed 50 Monte Carlo simulations, with the same platoon arrival times and background traffic inflow profiles for each control case. The resulting average and median TTS and delay are shown in Table II, for each vehicle class, and for all vehicles combined. We define the delay as the percentage of the increase in TTS compared to the ideal actuation case, which is taken as a benchmark for minimum achievable TTS of each simulation run, and also show its box plots in Figure 7.

We can see that even by applying control that ignores the on- and off-ramps described in Section IV-A, we reduce the TTS by about 10% of the ideal TTS on average, with the median reduced by about 17%. This corresponds to eliminating 29.1% of the delay on average, or 43.7% by median. However, only the TTS of class b, the mainstream-bound background traffic, is reduced, while the TTS of other vehicles is even somewhat increased. This can be explained by the fact that the controller assumes that all vehicles are headed for the bottleneck, and will therefore delay the traffic too much, stalling the off-ramp-bound traffic which would otherwise be able to leave the highway unhindered. In spite of this inefficiency, and owing to the fact that vehicles of class b comprise the majority of the traffic, this control law is still able to preserve free flow and forestall capacity drop at the bottleneck, thus the overall TTS and delays are lower than in the uncontrolled case.

In contrast, with the control from Section IV-B, the TTS of both class a and class b vehicles is reduced, with the aggregate TTS lower by almost 20% of ideal TTS on average, or by almost 30% in median. This corresponds to eliminating 52.7% of the delay on average, or 75.6% by median. Even though the platoons will be delayed in order to actuate the control,

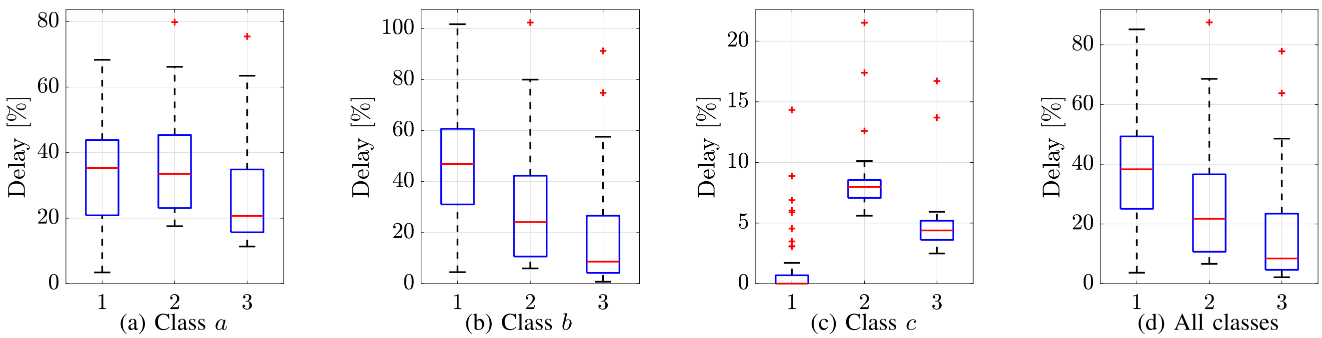


Fig. 7. Box plots showing the increase in TTS compared to the ideal actuation case.

their TTS will be lower, since they avoid the congestion at the bottleneck. This is especially important, since it shows that it is beneficial for the platooned vehicles to employ this control law, even if their goal is to minimize their own travel time. The increase of class c vehicles TTS is lower than with the previous control law. Overall, this control law comes very close to the ideal case, with the median delay being only 8.4%, and an average delay of 17.5%.

Note that while the proposed control laws achieve significant reduction of TTS, there is a number of outliers corresponding to unfavourable simulation runs when there are long gaps between two platoon arrivals. If this occurrence coincides with a higher demand of mainstream-bound background traffic, we will be unable to prevent the traffic breakdown, since there are no platoons available for actuation, resulting in a build-up of congestion and higher TTS.

VII. Conclusion

In this work, we proposed a tandem queueing model with moving bottlenecks, which we used as a prediction model for designing bottleneck decongestion control. We used platoons as actuators, with their speed and formation and control inputs, in order to keep the bottleneck in free flow, improve the throughput and reduce the TTS of all vehicles. The performance of these control laws was tested in multi-class CTM simulations, on a 5 km long stretch of highway upstream of a bottleneck, with one on- and one off-ramp. The achieved TTS using these control laws was compared to the case when no control is used, as well as with the case when we can fully control all vehicles individually. We demonstrated that applying the proposed control laws significantly reduces the TTS compared to the situation with no control, coming close to the performance of the ideal actuation case. Moreover, even the platooned vehicles, which are delayed in order to affect the rest of traffic, incur lower delays, since they avoid having to traverse the congestion at the bottleneck, making the proposed control beneficial for all traffic participants.

For future work, we are interested in expanding the scope by considering highway segments with multiple stationary bottlenecks regulated in cascade, as well as traffic networks. Theoretical bounds on the effect of the proposed control laws on TTS should also be derived, as well as tighter bounds on the achievable throughput. Finally, the proposed control law should also be tested on more complex simulation models, such as microscopic traffic simulators. In general, the influence of truck platoons on the traffic needs to be further investigated using both simulations and public road experiments.

Appendix
The Multi-class CTM

Consider a highway stretch consisting of N cells and let $\mathcal{K}$ be the set of vehicle classes. The traffic density of vehicles of class $\kappa \in \mathcal{K}$ in cell i at time t, $\rho_i^\kappa(t)$, evolves as

$$\rho_i^\kappa(t+1) = \rho_i^\kappa(t) + \frac{T}{L_i}\left(q_{i-1}^\kappa(t) - q_i^\kappa(t) + r_i^\kappa(t) - s_i^\kappa(t)\right),$$

where T is the time step, L_i the length of cell i, $r_i^\kappa(t)$ is the inflow and $s_i^\kappa(t)$ the outflow of each class from a potential on-ramp and to a potential off-ramp, respectively. The traffic flow of each class from cell i to cell $i+1$ is given by

$$q_i^\kappa(t) = \min\left\{D_i^\kappa(t), S_{i+1}^\kappa(t)\right\}.$$

Setting $q_0^\kappa(t) = 0$, the inflow to cell 1 is technically considered to come from an on-ramp.

Each vehicle class can have a distinct time-varying free flow speed $U_i^\kappa(t) \leq V_i$ in every cell, where V_i is the maximum vehicle speed for the cell. The demand and supply functions of each class $D_i^\kappa(t)$ and $S_i^\kappa(t)$ also depend on vehicles of other classes. Denoting the demands of each vehicle class if other classes were ignored as $d_i^\kappa(t) = U_i^\kappa(t)\rho_i^\kappa(t)$, we write the demand and supply allocated to each class

$$D_i^\kappa(t) = d_i^\kappa(t) \min\left\{1, \frac{Q_i(t)}{d_i^{\mathcal{K}}(t)}\right\},$$

$$S_i^\kappa(t) = \frac{\rho_{i-1}^\kappa(t)}{\rho_{i-1}^{\mathcal{K}}(t)} \min\left\{W_i(P_i - \rho_i^{\mathcal{K}}(t)), Q_i(t), F_{i-1}(t)\right\}.$$

Here, the cell capacity is given by

$$Q_i(t) = \frac{\sum_{\kappa\in\mathcal{K}} d_i^\kappa(t) \frac{V_i P_i \sigma_i U_i^\kappa(t)}{(P_i - \sigma_i)U_i^\kappa(t) + V_i \sigma_i}}{d_i^{\mathcal{K}}(t)} \leq V_i \sigma_i,$$

and cell parameters W_i, σ_i and P_i are the congestion wave speed, critical density and jam density of cell i, respectively. Capacity drop is modelled as a linear reduction of capacity,

$$F_i(t) = W_i \frac{\sigma_{i+1}}{\sigma_i}\left(P_i - (1-\alpha)\sigma_i - \alpha\rho_i^{\mathcal{K}}(t)\right),$$

where α is the maximum capacity drop ratio under jam traffic density $\rho_i^{\mathcal{K}}(t) = P_i$. We use the same cell length L and maximum free flow speed V for all cells, and take $W = V\frac{\sigma_i}{P_i - \sigma_i}$ yielding a triangular fundamental diagram.

Mainstream flow is prioritized, and the critical density is increased in cells close to on-ramps, where a merging lane is present. Therefore, vehicles entering the road will only queue at the on-ramp if congestion propagates upstream and blocks the on-ramp. We model the evolution of these queues $n_{\mathrm{r},i}^{\kappa}$, for on-ramps in cell i, with

$$
\begin{aligned}
n_{\mathrm{r},i}^{\kappa}(t+1) &= n_{\mathrm{r},i}^{\kappa}(t) + \big(\phi_i^{\kappa}(t) - r_i^{\kappa}(t)\big)\,T,\\
r_i^{\kappa}(t) &= \begin{cases} \phi_i^{\kappa}(t), & \kappa \in \mathcal{K}^{\mathrm{pr}},\\ \min\left\{D_{\mathrm{r},i}^{\kappa}(t), S_{\mathrm{r},i}^{\kappa}(t)\right\}, & \kappa \notin \mathcal{K}^{\mathrm{pr}}, \end{cases}\\
D_{\mathrm{r},i}^{\kappa}(t) &= \phi_i^{\kappa}(t) + \frac{n_{\mathrm{r},i}^{\kappa}(t-1)}{T},\quad S_{\mathrm{r},i}^{\kappa}(t) = \frac{n_{\mathrm{r},i}^{\kappa}(t)}{n_{\mathrm{r},i}^{\mathcal{K}\setminus\mathcal{K}^{\mathrm{pr}}}(t)}\,Q_{\mathrm{r},i}(t),\\
Q_{\mathrm{r},i}(t) &= \max\left\{0, \min\left\{S_i(t) - q_{i-1}(t), Q_{\mathrm{r},i}^{\max}\right\} - r_i^{\mathcal{K}^{\mathrm{pr}}}(t)\right\}.
\end{aligned}
$$

Here, $\phi_i^{\kappa}(t)$ is the inflow of class κ vehicles arriving at the on-ramp, $\mathcal{K}^{\mathrm{pr}}$ is the set vehicle classes that enter the road directly, and $Q_{\mathrm{r},i}^{\max}$ is the capacity of the on-ramp.

For cell i, where vehicles of classes $\mathcal{K}_{\mathrm{s},i} \subset \mathcal{K}$ exit the mainstream via an off-ramp, we may write

$$
s_i^{\kappa}(t) = \begin{cases} \min\left\{D_i^{\kappa}(t), S_{i+1}^{\kappa}(t), \dfrac{\rho_i^{\kappa}}{\rho_i^{\mathcal{K}_{\mathrm{s},i}}} Q_{\mathrm{s},i}^{\max}\right\}, & \kappa \in \mathcal{K}_{\mathrm{s},i},\\ 0, & \kappa \notin \mathcal{K}_{\mathrm{s},i}, \end{cases}
$$

where $Q_{\mathrm{r},i}^{\max}$ is the capacity of the off-ramp. Finally, we update

$$
D_i^{\kappa}(t) = \begin{cases} 0, & \kappa \in \mathcal{K}_{\mathrm{s},i},\\ d_i^{\kappa}(t)\min\left\{1, \dfrac{Q_i(t)}{d_i^{\mathcal{K}}(t)}\right\}, & \kappa \notin \mathcal{K}_{\mathrm{s},i}. \end{cases}
$$

Let there be Π platoons, and let platoon p move at speed $u_p(t) \in \left[U^{\min}, U^{\max}\right]$, with $U^{\max} < V$. We denote the position of the platoon downstream end $x_p(t)$, $x_1(t) > x_2(t) > \cdots > x_{\Pi}(t)$, and the reference density of platooned vehicles $\rho_p^*(t)$. Using vehicle class a for platooned vehicles and assuming $l_p(t) \geq 2\ L$, the platooned vehicles traffic density profile from cell $i_p^t(t) = \left\lceil \big(x_p(t) - l_p(t)\big)/L \right\rceil$ to cell $i_p^h(t) = \left\lceil x_p(t)/L \right\rceil$, where the upstream and downstream end of the platoon are, respectively, is given by

$$
\rho_i^a(t) = \begin{cases} 0, & i < i_{\Pi}^t(t),\\ \rho_p^*(t)\dfrac{X_{i_p^t(t)+1} - x_p(t) + l_p(t)}{L}, & i = i_p^t(t),\\ \rho_p^*(t), & i_p^t(t) < i < i_p^h(t),\\ \rho_p^*(t)\dfrac{x_p(t) - X_{i_p^h(t)}}{L}, & i = i_p^h(t),\\ 0, & i_p^h(t) < i < i_{p-1}^t(t), \end{cases}
$$

where $i_0^t = N+1$. The platoon position updates are $x_p(t+1) = x_p(t) + u_p(t)T$, and $\rho_i^a(t+1)$ needs to be updated accordingly, by setting

$$
\begin{aligned}
U_i^a(t) &= \begin{cases} V, & i < i_{\Pi}^t(t),\\ \psi_i^b(t), & i_p^t(t) \leq i < i_p^h(t),\\ \psi_i^h(t), & i = i_p^h(t),\\ 0, & i_p^h(t) < i < \frac{i_p^h(t) + i_{p-1}^t(t)}{2},\\ V, & \dfrac{i_p^h(t) + i_{p-1}^t(t)}{2} \leq i < i_{p-1}^t(t), \end{cases}\\
\psi_i^b(t) &= V\,\frac{V\rho_p^*(t) - \big(V - U_{i+1}^a(t)\big)\,\rho_{i+1}^a(t)}{V\rho_i^a(t)},\\
\psi_i^h(t) &= V\left(1 - \left(1 - \frac{u_p(t)}{V}\right)\frac{\rho_p^*(t)}{\rho_i^a(t)}\right).
\end{aligned}
$$

for $p = 1, \ldots, \Pi$. Furthermore, the traffic flow overtaking a platoon with density $\rho_p^*(t)$ is limited to $V\left(\sigma - \rho_p^*(t)\right)$, which is consistent with PDE moving bottleneck models.

Consider a bottleneck at some position X_{b}, with the upstream critical density of σ_- and downstream σ_+, corresponding to the capacity of $q_{\mathrm{b}}^{\mathrm{cap}} = V\sigma_+$. Due to capacity drop, its capacity will be decreased once it becomes congested, with a congestion of density

$$
\rho_c = \frac{P_-(\sigma_- - \sigma_+) + (1-\alpha)\sigma_-\sigma_+}{\sigma_- - \alpha\sigma_+}
$$

formed, and the density of discharging traffic being

$$
\rho_d = \frac{\sigma_-\sigma_+(1-\alpha)}{\sigma_- - \alpha\sigma_+} < \sigma_+,
$$

given by $W(P_- - \rho_c) = W\frac{\sigma_+}{\sigma_-}(P_- - (1-\alpha)\sigma_- - \alpha\rho_c)$, and $V\rho_d = W(P_- - \rho_c)$. The discharging flow from the congested bottleneck is $q_{\mathrm{b}}^{\mathrm{dis}} = V\rho_d$.

As a benchmark for comparing the performance of the proposed control laws, we consider the ideal case, assuming we can fully control every class of traffic independently. We need to minimally delay the mainstream-bound background traffic so that the flow at the bottleneck never exceeds its capacity. This is equivalent to ensuring that the traffic density immediately upstream of the bottleneck $\rho_{i_{\mathrm{b}}}(t) \leq \sigma_+$ for all t, with $U_i^{\kappa}(t)$ as close as possible to V, which can be achieved by setting $U_{i_{\mathrm{b}}}^b(t) = V$ and

$$
\begin{aligned}
U_i^b(t) &= \min\left\{V, \max\left\{U_{\min}^b, \psi_i^b(t)\right\}\right\},\quad i = 1, \ldots, i_{\mathrm{b}}-1\\
\psi_i^b(t) &= \frac{V}{\rho_i^b(t)}\left(\rho_i^{b*}(t) - \frac{V - U_{i+1}^b(t)}{V}\rho_{i+1}^b(t)\right),\\
\rho_i^{b*}(t) &= \begin{cases} \sigma_+ - \rho_p^*, & \dfrac{X_{\mathrm{b}} - x_p(t)}{u_p(t)} < \dfrac{X_{\mathrm{b}} - (i-1)L}{V} < \dfrac{X_{\mathrm{b}} - x_p(t) + l_p(t)}{u_p(t)} + \dfrac{L}{V},\\ \sigma_+, & \text{otherwise}, \end{cases}
\end{aligned}
\tag{32}
$$

for $p = 1, \ldots, \Pi$. The mainstream-bound background traffic is thus controlled so that the total flow at the bottleneck, including the platoons, is close to the capacity.

References

[1] C. Bergenhem, S. Shladover, E. Coelingh, C. Englund, and S. Tsugawa, "Overview of platooning systems," in *Proc. 19th ITS World Congr.*, Vienna, Austria, 2012, pp. 1–8.

[2] C. Bonnet and H. Fritz, "Fuel consumption reduction in a platoon: Experimental results with two electronically coupled trucks at close spacing," SAE Tech. Paper 2000-01-3056, 2000.

[3] J. Lioris, R. Pedarsani, F. Y. Tascikaraoglu, and P. Varaiya, "Platoons of connected vehicles can double throughput in urban roads," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 292–305, Apr. 2017.

[4] S. Moridpour, E. Mazloumi, and M. Mesbah, "Impact of heavy vehicles on surrounding traffic characteristics," *J. Adv. Transp.*, vol. 49, no. 4, pp. 535–552, Jun. 2015.

[5] L. Aarts and G. Feddes, "European truck platooning challenge," in *Proc. Int. Symp. Heavy Vehicle Transp. Technol. (HVTT)*, Rotorua, New Zealand, 2016, pp. 15–18.

[6] A. K. Bhoopalam, N. Agatz, and R. Zuidwijk, "Planning of truck platoons: A literature review and directions for future research," *Transp. Res. B, Methodol.*, vol. 107, pp. 212–228, Jan. 2018.

[7] Rijkswaterstaat, "European truck platooning challenge 2016: Lessons learnt," Tech. Rep., 2016.

[8] A. Duret, M. Wang, and A. Ladino, "A hierarchical approach for splitting truck platoons near network discontinuities," *Transp. Res. Procedia*, vol. 38, pp. 627–646, Jan. 2019.

[9] L. Jin, M. Čičič, S. Amin, and K. H. Johansson, "Modeling the impact of vehicle platooning on highway congestion: A fluid queuing approach," in *Proc. 21st Int. Conf. Hybrid Syst., Comput. Control (CPS Week)*, Apr. 2018, pp. 237–246.

[10] Y. Wang, E. B. Kosmatopoulos, M. Papageorgiou, and I. Papamichail, "Local ramp metering in the presence of a distant downstream bottleneck: Theoretical analysis and simulation study," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2024–2039, Oct. 2014.

[11] M. Hadiuzzaman, T. Z. Qiu, and X.-Y. Lu, "Variable speed limit control design for relieving congestion caused by active bottlenecks," *J. Transp. Eng.*, vol. 139, no. 4, pp. 358–370, Apr. 2013.

[12] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobson, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 4, pp. 568–583, Aug. 2010.

[13] S. Cui, B. Seibold, R. Stern, and D. B. Work, "Stabilizing traffic flow via a single autonomous vehicle: Possibilities and limitations," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 1336–1341.

[14] M. Wang, W. Daamen, S. P. Hoogendoorn, and B. van Arem, "Connected variable speed limits control and car-following control with vehicle-infrastructure communication to resolve stop-and-go waves," *J. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 559–572, Nov. 2016.

[15] G. Piacentini, P. Goatin, and A. Ferrara, "Traffic control via moving bottleneck of coordinated vehicles," in *Proc. 15th IFAC Symp. Control Transp. Syst.*, 2018, vol. 51, no. 9, pp. 13–18.

[16] M. Čičić and K. H. Johansson, "Traffic regulation via individually controlled automated vehicles: A cell transmission model approach," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Maui, HI, USA, Nov. 2018, pp. 766–771.

[17] L. Jin, M. Čičić, K. H. Johansson, and S. Amin, "Analysis and design of vehicle platooning operations on mixed-traffic highways," *IEEE Trans. Autom. Control*, early access, Oct. 30, 2020, doi: 10.1109/TAC.2020.3034871.

[18] L. Liu, L. Zhu, and D. Yang, "Modeling and simulation of the car-truck heterogeneous traffic flow based on a nonlinear car-following model," *Appl. Math. Comput.*, vol. 273, pp. 706–717, Jan. 2016.

[19] M. L. Delle Monache and P. Goatin, "Scalar conservation laws with moving constraints arising in traffic flow modeling: An existence result," *J. Differ. Equ.*, vol. 257, no. 11, pp. 4015–4029, Dec. 2014.

[20] M. D. Simoni and C. G. Claudel, "A fast simulation algorithm for multiple moving bottlenecks and applications in urban freight traffic management," *Transp. Res. B, Methodol.*, vol. 104, pp. 238–255, Oct. 2017.

[21] M. Čičić, I. Mikolášek, and K. H. Johansson, "Front tracking transition system model with controlled moving bottlenecks and probabilistic traffic breakdowns," in *Proc. 21st IFAC World Congr.*, 2020, vol. 53, no. 2, pp. 14990–14996.

[22] M. Kontorinaki, A. Spiliopoulou, C. Roncoli, and M. Papageorgiou, "First-order traffic flow models incorporating capacity drop: Overview and real-data validation," *Transp. Res. B, Methodol.*, vol. 106, pp. 52–75, Dec. 2017.

[23] M. Čičić and K. H. Johansson, "Stop-and-go wave dissipation using accumulated controlled moving bottlenecks in multi-class CTM framework," in *Proc. IEEE 58th Conf. Decis. Control (CDC)*, Nice, France, Dec. 2019, pp. 3146–3151.

[24] W.-L. Jin, "Point queue models: A unified approach," *Transp. Res. B, Methodol.*, vol. 77, pp. 1–16, Jul. 2015.

[25] G. Piacentini, M. Čičić, A. Ferrara, and K. H. Johansson, "VACS equipped vehicles for congestion dissipation in multi-class CTM framework," in *Proc. 18th Eur. Control Conf. (ECC)*, Napoli, Italy, Jun. 2019, pp. 2203–2208.

Mladen Čičić received the B.Sc. and M.Sc. degrees in electrical engineering and computer science from the School of Electrical Engineering, University of Belgrade. He is currently pursuing the Ph.D. degree with the Division of Decision and Control Systems, KTH Royal Institute of Technology. He was a member of Marie Skłodowska Curie oCPS ITN, an Affiliate Student of Wallenberg AI, an Autonomous Systems and Software Program (WASP), and a Visiting Scholar at the C2SMART University Transportation Center, NYU Tandon School of Engineering. His research interest includes traffic control using mixed traffic models.

Xi Xiong received the B.Eng. degree in automotive engineering from Jilin University in 2012 and the M.S. degree in automotive engineering from Tsinghua University in 2015. He is currently pursuing the Ph.D. degree with the Department of Civil and Urban Engineering, New York University Tandon School of Engineering. He was a Student Researcher at the C2SMART Department of Transportation Center and the Center for Urban Science and Progress (CUSP). His research focuses on developing optimization and learning algorithms to improve transportation efficiency and resiliency.

Li Jin (Member, IEEE) received the B.Eng. degree in mechanical engineering from Shanghai Jiao Tong University in 2011, the M.S. degree in mechanical engineering from Purdue University in 2012, and the Ph.D. degree in transportation from the Massachusetts Institute of Technology in 2018. He was an Assistant Professor at the New York University Tandon School of Engineering, where he is currently a Research Assistant Professor. He is also a John Wu and Jane Sun Assistant Professor at the School of Electronic Information and Electrical Engineering, University of Michigan-Shanghai Jiao Tong University Joint Institute. His interests lie in control and optimization of smart and connected transportation systems. He was a recipient of the Ho-Ching and Hang-Ching Fund Award and the Schoettler Scholarship Fund.

Karl Henrik Johansson (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from Lund University. He is currently the Director of the Stockholm Strategic Research Area ICT–the Next Generation and a Professor with the School of Electrical Engineering, KTH Royal Institute of Technology. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests include networked control systems, cyber-physical systems, and applications in transportation, energy, and automation. He is a member of the IEEE Control Systems Society Board of Governors, the European Control Association Council, and the Royal Swedish Academy of Engineering Sciences. He is also a Distinguished Lecturer of IEEE. He has received several best paper awards and other distinctions, including a ten-year Wallenberg Scholar Grant, a Senior Researcher Position with the Swedish Research Council, the Future Research Leader Award from the Swedish Foundation for Strategic Research, and the Triennial Young Author Prize from IFAC.