



Doctoral Thesis in Electrical Engineering

# Modelling and Lagrangian control of mixed traffic: platoon coordination, congestion dissipation and state reconstruction

MLADEN ČIČIĆ

# Modelling and Lagrangian control of mixed traffic: platoon coordination, congestion dissipation and state reconstruction

MLADEN ČIČIĆ

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Friday the 12th of March 2021, at 4:00 p.m. in Q2, Malvinas väg 10, Stockholm.

Doctoral Thesis in Electrical Engineering  
KTH Royal Institute of Technology  
Stockholm, Sweden 2021

© Mladen Čičić

ISBN 978-91-7873-782-6  
TRITA-EECS-AVL-2021:13

Printed by: Universitetservice US-AB, Sweden 2021

## Abstract

Traffic congestion is a constantly growing problem, with a wide array of negative effects on the society, from wasted time and productivity to elevated air pollution and reduction of safety. The introduction of connected, autonomous vehicles enables a new, Lagrangian paradigm for sensing and controlling the traffic, by directly using connected vehicles inside the traffic flow, as opposed to the classical, Eulerian paradigm, which relies on stationary equipment on the road. By using control methods specifically tailored to the Lagrangian paradigm, we are able to influence the traffic flow even if the penetration rate of connected vehicle is low. This allows us to answer one of the central impending questions of the traffic control using emerging technologies: How can we influence the overall traffic by using only a small portion of vehicles that we can control directly?

Traffic phenomena such as moving bottlenecks and stop-and-go waves are particularly pertinent to Lagrangian traffic control, and therefore need to be captured in traffic models. In this thesis we introduce the influence of these phenomena into the cell transmission model, multi-class cell transmission model, and tandem queueing model. We also propose a transition system model based on front tracking, which captures the relevant phenomena, and show under which conditions it corresponds to the Lighthill-Whitham-Richards model. Moving bottlenecks are introduced as a moving zone in which a reduced flux function describes the traffic flow, and their influence on the surrounding traffic is given by solving the Riemann problems at the flux function boundaries. Stop-and-go waves are introduced by constraining the wave speed of rarefaction, resulting in constant stop-and-go wave propagation speed and discharging flow lower than the road capacity, which is consistent with the empirical observations.

We use the proposed traffic models to design control laws that address three problems: platoon merging coordination, congestion reduction, and traffic state reconstruction. We study the case when two trucks are closing the distance and merging into a platoon on a public road, and propose an optimal control algorithm which accounts for the mutual influence between the trucks and the surrounding traffic. The proposed control law minimizes the total fuel consumption of the trucks, and improves the reliability of platooning. Then, we consider two forms of the congestion reduction problem: stationary bottleneck decongestion, and stop-and-go wave dissipation. In both cases, connected vehicles are used as moving bottlenecks to restrict the traffic flow enough to let the congestion dissipate. By applying these control laws, the throughput of the road is increased and the total travel time of all vehicles is reduced. Finally, we generalize the stop-and-go wave dissipation problem by dropping the assumption that the full traffic state is known, and instead propose traffic state reconstruction algorithms which use local measurements originating from the connected vehicles. We show that the proposed control laws can also be implemented using the reconstructed traffic state. In this case, as the number of available connected vehicles increases, the control performance approaches the full-information control case.

## Sammanfattning

Trafikstockning är ett ständigt växande problem, med ett brett utbud av negativa effekter på samhället, från bortkastad tid och produktivitet till ökade mängd luftföroreningar och minskning av säkerhet. Införandet av uppkopplade, autonoma fordon möjliggör ett nytt, Lagrangianskt paradigms för att styra och mäta trafiken, genom att direkt använda uppkopplade fordon inuti trafikflödet, i motsats till det klasiska, Euleriska paradigmet, som är beroende på stillastående utrustning på vägen. Genom att använda kontrollmetoder som är anpassad för Lagrangian-paradigmet kan vi påverka trafikflödet även om marknadsintrång av uppkopplade fordon är låg. Detta gör det möjligt för oss att besvara en av de centrala överhängande frågorna om trafikkontrollen med framväxande teknik: Hur kan vi påverka den totala trafiksituationen genom att direkt kontrollera en liten del av fordonen?

Vissa trafikfenomen som rörliga flaskhalsar och stop-and-go-vågor är särskilt relevanta för Lagrangian trafikstyrning, och måste därför modelleras. I denna avhandling introducerar vi påverkan av dessa fenomen i cellöverföringsmodellen, flerklasscellöverföringsmodellen, och tandemkömodellen. Vi föreslår även en övergångssystemmodell baserad på front-tracking-metoden, som beskriver relevanta fenomen, och visar under vilka förhållanden den motsvarar Lighthill-Whitham-Richards-modellen. Rörliga flaskhalsar introduceras som en rörlig zon där en reducerad flödesfunktion beskriver trafikflödet, och deras inflytande på trafiken beräknas genom att lösa Riemann-problemen vid flödesfunktioners gränser. Stop-and-go-vågor introduceras genom att begränsa sällsynthetens våghastighet, som resulterar i konstant stop-and-go-vågens hastighet och utflöde som är lägre än vägkapaciteten, vilket överensstämmer med de empiriska observationerna.

Vi använder de föreslagna trafikmodellerna för att utforma kontrollalgar som hanterar tre problem: koordinering av fordonstågsammanfogning, minskning av trafikstockningar och uppskattning av trafiktillstånd. Vi studerar fallet när två lastbilar närmar sig varandra och sammanfogar till en fordonståg på allmän väg, och föreslår en optimal kontrollalgoritm som tar hänsyn till interaktionen mellan lastbilarna och den omgivande trafiken. Den föreslagna kontrollalgoritmen minimerar den totala bränsleförbrukningen för lastbilarna och förbättrar pålitligheten av fordonstågkörning. Sedan granskar vi två former av problem med minskning av trafikstockningar: stationär flaskhalsavlastning och stop-and-go-vågskingring. I båda fallen används uppkopplade fordon som rörliga flaskhalsar för att begränsa trafikflödet så att trängseln upplösas. Genom att tillämpa dessa kontrollalgar ökar vägens genomströmning och den totala restiden för alla fordon minskas. Slutligen, generaliserar vi stop-and-go-vågskingringsproblem genom att släppa antagandet att hela trafiktillståndet är känt, och istället föreslår trafiktillståndsuppskattningalgoritmer som använder lokala mätningar från de uppkopplade fordonen. Vi visar att de föreslagna kontrollalgarorna kan även implementeras med hjälp av det uppskattade trafiktillståndet. I detta fall, när antalet tillgängliga uppkopplade fordon ökar, blir kontrollprestationer nästan lika bra som när det fullständiga trafiktillståndet är känt.

*To my family and friends*

*Mojoj porodici i prijateljima*

# Acknowledgements

When starting my doctoral studies, I did not expect that the contents of this thesis, nor the circumstances in which I am defending it, would be as they turned out to be in the end. And yet, I am grateful for everything that I have experienced, discovered, learned, struggled with, and overcome during the course of my studies, as well as for everyone that I have met, befriended, and collaborated with. I truly feel that I have grown as a person and as a professional thanks to everything and everyone that I have encountered over these years.

First and foremost, I would like to express my deepest gratitude to my supervisor Karl Henrik Johansson for all the support, feedback, guidance, inspiration, as well as the opportunity to work on interesting research topics, that he provided. All of the freedom to choose my research path that I was given, although somewhat daunting and intimidating initially, has ultimately proven worthwhile and rewarding. I have learned much from you, not only academically, and over time I understand more and more just how good your advice was. I would also like to thank my co-supervisor, Jonas Mårtensson for all the enjoyable collaboration on courses and projects.

Special thanks to Li Jin for being a great host for my research visit at NYU Tandon School of Engineering, as well as for our successful collaboration. Along with him, big thanks also goes to my co-authors, Matthieu Barreau, Giulia Piacentini, Igor Mikolášek, Kuo-Yun Liang, Xi Xiong, Amr Ibrahim, as well as Saurabh Amin, Antonella Ferrara, Dip Goswami and Twan Basten. Working with you proved not only fruitful, but also fun and gratifying in more ways than one. I would like to thank Miroslav Krstić, for kindly accepting to serve as my opponent, and to Daniel Work, Paola Goatin, and Bart de Schutter, for agreeing to be on the committee for the defence of this thesis. I am also very grateful to Matthieu Barreau, Apostolos Rikos, Miguel Campos Pinto Coelho de Aguiar, and Alexander Johansson, for proofreading and providing feedback on parts of this thesis, and to Xiaoliang Ma for acting as an advance reviewer for me for the second time, as well as for agreeing to act as a substitute, together with Erik Jenelius.

It was truly a pleasure working with everyone at the department, and in the oCPS and WASP programmes, and I am very grateful to all my past and current colleagues for making it such a cordial, relaxed and homey environment. I am also thankful to everyone from the administrative staff that made sure that everything was working smoothly. Special thanks go to Jezdimir Milošević, for

unknowingly paving the way for me, David Umsonst, for organizing a significant portion of social activities, Vahan Petrosyan, for hospitality and introduction to the vibrant Armenian culture and cuisine, Marina Rantanen Mod  er, for the numerous oCPS trips, Joakim Bj  rk, for the outdoor activities, as well as to In  s De Miranda De Matos Louren  o, Christos Verginis, Rijad Ali  i  , Sebastian van de Hoef, Valerio Turri, Dominik Baumann, Rui Oliveira, Joana Filipa Gouveia Fonseca, Robert Marczuk Bereza-Jarocinski, the rest of my D&D group, and everybody else — far too many to name all here — who participated in lunch discussions, barbecues, hikes, boardgame nights, after-works, conference trips, and other parts of my doctoral life. All of these experiences have made this time much more enjoyable and pleasant.

The research leading to this thesis has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Sk  łodowska-Curie grant agreement No 674875, VINNOVA, the Swedish Research Council, the Swedish Foundation for Strategic Research, Knut and Alice Wallenberg Foundation, and KAUST OSR-2019-CoE-NEOM-4178.5. I am grateful for their support. I also appreciate the opportunities provided to me as an affiliated doctoral student of the Wallenberg AI, Autonomous Systems and Software Program.

Last but not least, a heartfelt thanks goes to my family and friends. First, to my parents, Milorad and Nata  a, and my sister Ana, for giving me a strong foundation to build upon, and unwavering support to keep what was built standing. I am also thankful to my friends back home, especially Spasoje, to whom I am indebted for acting continually as both a host and a rearguard. Finally, I want to thank Astrid, for bearing with me and brightening up a significant portion of my doctoral life.

*Mladen   i  i  ,*  
Stockholm, March 2021

# Contents

<b>Contents</b>	<b>vii</b>
<b>List of symbols</b>	<b>ix</b>
<b>I Introduction and background</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Motivation . . . . .	4
1.2 Problem formulation . . . . .	9
1.3 Thesis outline and contributions . . . . .	15
<b>2 Background</b>	<b>23</b>
2.1 Intelligent transportation systems . . . . .	23
2.2 Platoons in traffic . . . . .	26
2.3 Mixed traffic models . . . . .	28
2.4 Lagrangian sensing and actuation . . . . .	30
2.5 Summary . . . . .	33
<b>II Modelling</b>	<b>35</b>
<b>3 Basic traffic models</b>	<b>37</b>
3.1 Microscopic traffic models . . . . .	37
3.2 Lighthill-Whitham-Richards model . . . . .	42
3.3 Cell transmission model . . . . .	48
3.4 Tandem queueing model . . . . .	51
3.5 Second-order traffic models . . . . .	53
3.6 Summary . . . . .	54
<b>4 Traffic model extensions</b>	<b>55</b>
4.1 Traffic phenomena of interest . . . . .	55
4.2 Moving bottlenecks in LWR and CTM . . . . .	57
4.3 Multi-class CTM . . . . .	67
4.4 Tandem queueing model with platoons . . . . .	81
4.5 Front tracking transition system model . . . . .	90

4.6	Summary	109
<b>III Control</b>		<b>111</b>
<b>5</b>	<b>Platoon merging coordination</b>	<b>113</b>
5.1	Vehicle model and deviations from the nominal speed	114
5.2	Platoon merging distance prediction based on the road grade	116
5.3	Energy-optimal platoon merging in traffic	126
5.4	Summary	138
<b>6</b>	<b>Bottleneck decongestion</b>	<b>139</b>
6.1	Bottleneck decongestion problem	140
6.2	Bottleneck decongestion with stochastic traffic breakdown	142
6.3	Platoon coordination for deterministic bottleneck decongestion	152
6.4	Summary	169
<b>7</b>	<b>Single stop-and-go wave dissipation</b>	<b>171</b>
7.1	Single stop-and-go wave dissipation problem	172
7.2	Moving bottleneck control	175
7.3	Accumulated moving bottlenecks control	184
7.4	Summary	191
<b>8</b>	<b>Reconstruction-based multiple stop-and-go wave dissipation</b>	<b>193</b>
8.1	Multiple stop-and-go wave dissipation	193
8.2	Basic traffic state reconstruction and probe vehicle selection	196
8.3	Reconstruction-based control with model learning	201
8.4	Summary	218
<b>IV Conclusions</b>		<b>219</b>
<b>9</b>	<b>Summary and future research</b>	<b>221</b>
9.1	Summary	221
9.2	Future research directions	223
<b>Bibliography</b>		<b>225</b>

# List of symbols

$\rho$	Traffic density
$n$	Number of vehicles in a cell or a queue
$q$	Traffic flow
$v$	Traffic or individual vehicle speed
$u$	Reference vehicle speed
$t$	Continuous time
$t^T$	Discrete time instant with time step $T$
$x$	Position
$\mathcal{V}$	Average density-dependent traffic speed
$Q$	Flux function (fundamental diagram)
$\sigma$	Flux function breakpoint (critical density for triangular $Q$ )
$V$	Flux function slope (free flow speed for triangular $Q$ )
$P$	Jam (maximum) traffic density
$W$	Wave speed bounds (congestion wave speed for triangular $Q$ )
$z$	Wave front position
$\lambda$	Wave front propagation speed
$L$	Cell length
$T$	Time step
$g$	Inter-vehicular gap
$\gamma$	Road grade
$\alpha$	Capacity drop severity
$\beta$	Bottleneck severity
$\xi$	Platoon or moving bottleneck
$\psi$	Stop-and-go wave



## Part I

# Introduction and background



# Chapter 1

## Introduction

INTRODUCING new communication, sensing, and control technologies into vehicles and transportation infrastructure offers new possibilities to solve the ever-growing traffic congestion problem. Even a small share of connected collaborating autonomous vehicles on the roads can enable the traffic control centres to directly influence the traffic situation, without the need for building additional road infrastructure, or installing traffic control equipment. The omnipresence of mobile phones, and widespread use of navigation apps is already affecting the traffic flows in notable ways. With these prerequisites fulfilled, a new approach to traffic control is becoming available, Lagrangian<sup>1</sup> traffic control.

In the Lagrangian traffic control paradigm, we rely on sensing and actuation executed inside the traffic flows, provided by the participating vehicles themselves, as opposed to relying on stationary equipment like the classical, Eulerian traffic control does. This paradigm offers us a way to utilize the resources that become available over the course of the slow, partial or complete, transition from human-driven to self-driving vehicles, and is especially important for the intermediate phase when human and autonomous traffic coexist on the roads. It is therefore important to understand and model the mutual influence the individual vehicles and the overall traffic have on each other. This understanding enables us to better predict how some vehicles of interest, such as heavy-duty vehicles and platoons, as well as connected and autonomous vehicles, will actually be able to move in traffic, as well as to develop control strategies that exploit this interdependence. This is the focus of this thesis.

Naturally, in order to achieve best results, a combination of Eulerian and Lagrangian methods is preferable, using all available resources. However, while numerous Eulerian methods have been proposed and implemented over the years,

---

<sup>1</sup>The notion of Lagrangian traffic control is based on the Lagrangian flow specifications from the classical field theories. In the Lagrangian approach, the flow is described through the motion of the observer following a flow particle as it moves through space and time. As opposed to this approach, the Eulerian specification of the flow is based on a stationary observer, describing the flow through a specific location.

due to technical and partially legislative limitations, the Lagrangian approach has not yet seen practical use, except in case of some traffic state estimation based on trajectories of mobile phones in vehicles. Therefore, the control and traffic state reconstruction methods proposed in this thesis should not be seen as a replacement for the classical traffic control, but as an alternative approach which complements the well-established methods already in use.

The outline of this chapter is as follows. In Section 1.1 we motivate why using individual vehicles to control the traffic is a promising approach. In Section 1.2 we formulate the problems this work addresses. Lastly, Section 1.3 gives an overview of this thesis, its contents and contributions.

## 1.1 Motivation

Traffic congestion has been a growing problem for at least as long as there have been cars. Barring exceptional circumstances, the likes of which we have been experiencing in recent times, its gravity can only be expected to increase in the long term [1]. The negative effects of traffic congestion are not limited to wasting road user's time in traffic jams, leading to decreased reliability, efficiency and quality of life; additionally, it also poses a safety hazard, since it both stresses and frustrates the drivers and increases the risk for collisions due to stop-and-go traffic and low headways. Finally, traffic congestion also leads to an increase in fuel consumption [2], and as a direct consequence, greenhouse gas emissions, contributing to air and noise pollution. In the European Union in 2017, transportation accounted for 29.7% of total CO<sub>2</sub> emissions, and about 26.3% of total greenhouse gas emissions, of which the share of road transportation was around 71.7% [3], as shown in Figure 1.1. Similar figures are actual in the United States [4], with an even larger share of road transportation, around 82%. It is clear that if the goal of reducing the greenhouse gas emissions by 90% by 2050 [5] is to be attained, new, more efficient traffic management systems that would reduce congestion are a necessary part of the solution.

The advent of connected and autonomous vehicles (CAVs) promises to change the way we think of traffic forever. This technology has the potential to greatly impact virtually all facets of traffic [6], including significantly increasing traffic safety, reducing congestion and fuel consumption, and increasing the efficiency of freight transport, to name but a few. A number of major car manufactures promised to have fully autonomous cars in highway driving scenarios in the early 2020s [7]. These promises have since proven to be overly optimistic, and Level 5 autonomy, with full driving automation requiring no human input, remains a pipe dream for the time being. However, major technological advances have been made on Level 3 and 4 autonomy, with partial and conditional automation requiring some human participation, and we can expect the penetration rate of vehicles equipped with these technologies to steadily increase in the coming years [8].

However, these beginning phases of the introduction of CAVs into highway traffic

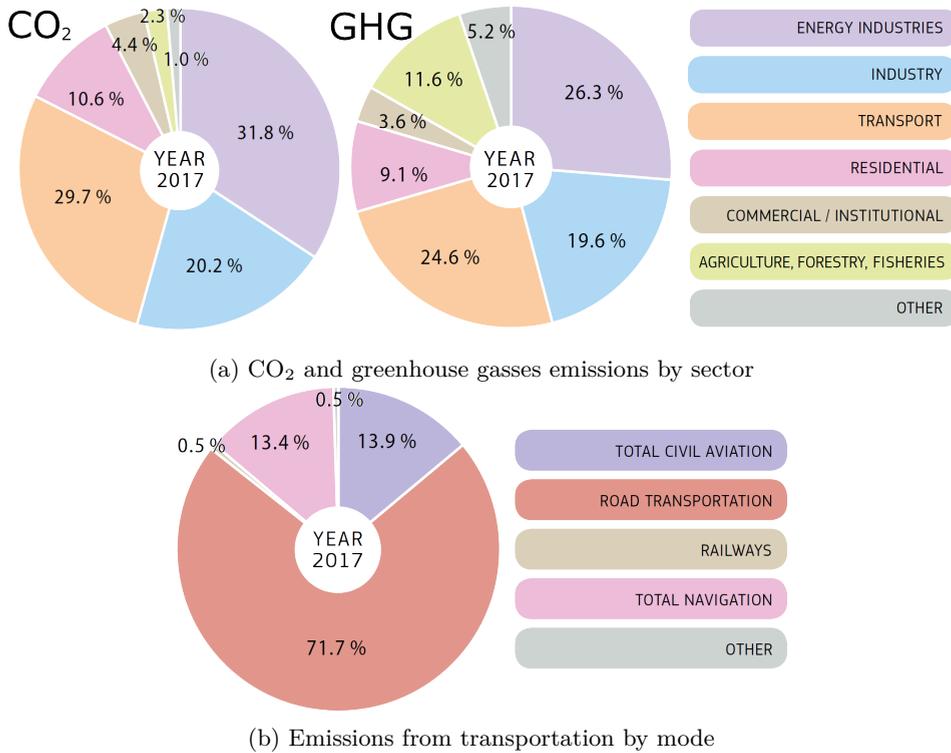


Figure 1.1: CO<sub>2</sub> and greenhouse gases emissions in the European Union in 2017. Source: EU transport in figures 2019 [3]

strike a delicate balance between the benefits and detriments that they bring, as it will take decades for the market penetration rate to become high enough for most of the benefits to become significant [9]. For example, reservation-based intersection control mechanisms [10] vastly outperforms conventional traffic light control with average delays that are two to three hundred times lower—but only in case all vehicles are autonomous. Including even a small number of human drivers would lead to a sharp performance deterioration, and traditional traffic lights become a preferred strategy if the portion of human drivers is significant. Shladover et al. [11] estimate that we can achieve 90% higher lane capacity if we have an 80% market penetration rate of cooperative adaptive cruise control (CACC), but with low market penetration rates around 10%, the increase in capacity drops to only 1%. While long-term, we may expect significant throughput and capacity increases once CAVs reach a high market penetration rate [12, 11], the low-to-mid-term effects are likely to be adverse, due to increased demand and overly conservative autonomous driving behaviour [13, 14].

Fortunately, the presence of even a small number of directly controllable CAVs on the road gives us new potentials for traffic sensing and control, with unprece-



Figure 1.2: Illustration of connected vehicles. Vehicles communicate with each other (V2V) and with the infrastructure (V2I).

mented level of detail, which we might use to improve the traffic flow and offset their negative effects. Much before autonomy, we can expect vehicles to start communicating with other vehicles and with the infrastructure (See Figure 1.2) [15]. In some sense, through ubiquitous mobile phones, vehicles have been indirectly connected in some way for a long time, and this communication can be used to acquire traffic measurements [16]. Even today, routing and navigation apps are widely used, and have started to affect the traffic in various ways [17]. It is not hard to envisage a further development of this concept, where an app would be used by the traffic authorities to also assign reference speed or other controls to drivers via their phones, as shown in Figure 1.3, which would execute those commands for potential monetary compensation. Therefore, in order to successfully deal with the transition period, we have to answer the following question: How can we improve the overall traffic by using only a small portion of vehicles that we have direct control over? This notion of Lagrangian control of the macroscopic traffic conditions by acting at a microscopic level is the centrepiece of emerging traffic control strategies, and the focus of this thesis.

Automated heavy-duty vehicle platooning, shown in Figure 1.4, is expected to enter the traffic at an accelerated pace [18]. Since these vehicles are typically moving slower than the rest of the traffic, they naturally act as moving bottlenecks. In the future, we can expect fleet management systems to employ some centralized



Figure 1.3: A traffic control loop closed via an app. The traffic control centre uses the trajectories of the participating app-connected vehicles, and uses it to estimate the traffic situation. Then the control centre issues reference speed to the drivers through the app. The participating drivers are rewarded monetarily, or with services.

remote control over vehicles, using vehicle-to-infrastructure (V2I) communication to enable advanced route planning [19]. Since in addition, these vehicles would send their status to the fleet management system and receive reference speed profiles to follow, this makes them an ideal candidate for in-flow traffic actuators. It is therefore important to understand how controlling these vehicles can affect the traffic around them, as well as how the traffic conditions could cause these vehicles to be delayed, thus enabling us to better plan their trajectories.

If the traffic volume exceeds the road capacity, there is little that can be done to reduce the congestion other than rerouting some of the traffic to other roads [20]. However, if the traffic density is close to the critical density, the variability in the traffic flow may cause a traffic breakdown to occur at some bottleneck. After the breakdown, the traffic efficiency is further reduced, making it hard to return to free flowing traffic, even if the incoming traffic flow is low enough for the bottleneck to handle. In particular, arrival of large groups of vehicles, such as truck platoons, is likely to cause congestion at the bottlenecks. It is therefore important to distribute the arrivals of long platoons of heavy-duty vehicles in time, in order to minimize the effect that they have on the traffic. Ideally, if the arrival of platoons can be postponed until the time when the density of the remainder of the traffic is low enough, or we are able to actively control the traffic flow to ensure that this is the case, we may avoid causing the traffic breakdown altogether.

It is also possible for the actions of individual drivers to cause congestion to emerge, even without a stationary bottleneck. For example, an aggressive lane change can force the driver in front of which the lane changing vehicle cut in to brake. This braking in turn forces the driver behind to brake harder, and the



Figure 1.4: Heavy-duty vehicles in a platoon. Source: Scania

disturbance propagates upstream, amplified until the point some car is forced to come to a full stop, and a so-called phantom traffic jam, or stop-and-go wave, shown in Figure 1.5, is formed. Individual vehicles can, under some circumstances, also help prevent or dissipate congestion waves. The notion of “jam-busting” or “jam-absorbing” driving techniques has been discussed, not only in research [21, 22], but also in media [23]. These techniques offer guidelines to the drivers, for example to leave a large gap in front of them and drive at the average speed of the surrounding traffic. By doing this, the driver is less likely to be forced to break aggressively if another car cuts in front of them, while also leaving enough space for cars to change lanes and move towards the exit lane or merge into the mainstream from merging lanes. Although this strategy focuses on vehicle interaction at a microscopic level, we may think of extending this approach to a higher layer of control, and devise macroscopic traffic control methods that use individual directly controlled vehicles.

Finally, the first step to implementing any type of traffic control method is to know what the situation in traffic is. The more sensors there are on the roads, the clearer the picture of the traffic state can be. If fixed sensors are not available to measure the traffic conditions at some location, we need to reconstruct and estimate the traffic state, based on the available measurements from elsewhere in the network. It is therefore immensely advantageous to be able to use the connected vehicles themselves as a type of mobile sensors, travelling inside the traffic flow. Even just reporting the locations of individual vehicles, like the vast majority of



Figure 1.5: A phantom traffic jam in effect. Vehicles at a standstill are circled in red. Source: YouTube, Author:Joe Miroe

GPS-enabled mobile phones that the occupants of the vehicle carry do, can be used to get a good picture of the overall traffic state, as was shown in [16], and now widely implemented by Google and other companies. Utilization of more detailed trajectory data, as well as potentially accessing the measurements of the on-board sensors of connected vehicles, can only provide a more comprehensive picture of the traffic conditions, allowing us to first understand the dynamics of the traffic flow better, and also to propose, design, and implement novel approaches to traffic control.

## 1.2 Problem formulation

The focus of this thesis is on modelling and control of connected passenger and heavy-duty vehicles, both individually and in platoons, as well as the rest of the traffic around them. Through understanding, modelling and utilizing the interaction between these two types of traffic, we are able to predict the trajectories of some specific vehicles, and regulate the overall traffic flow through directly controlling a small subset of connected vehicles. This allows us to better select the speed profiles

of heavy-duty vehicles catching up on the road to form a platoon, minimizing their total fuel spent. Furthermore, the influence over the traffic flow that we can exert allows us to improve the efficiency of road utilization, using Lagrangian traffic control. We are able to decongest bottlenecks and keep them in free flow, negating the capacity drop that happens once congestion builds up. Furthermore, we are able to dissipate stop-and-go waves, harmonizing the traffic flow and increasing the throughput of the road. In the rest of this section, we will present these problems in more detail.

### 1.2.1 Platoon merging coordination problem

Consider the simplest platoon merging problem, illustrated in Figure 1.6. Assume a potential platooning pair, driving along the common stretch of road, has been identified by a platooning coordinator at a higher decision layer. These two vehicles adjust their speeds so that the leader (the vehicle farther ahead,  $\xi = 1$ ) drives slower than the follower ( $\xi = 2$ ), so that they would meet and merge into a platoon at some point. Their state at time  $t$  is given by their positions  $x_\xi(t)$  and their speeds  $v_\xi(t)$ , so we have that  $x_1(t) > x_2(t)$  and  $v_1(t) < v_2(t)$ .

Starting at some time  $t_0$ , we can write the dynamics of the vehicles can thus be written

$$\begin{aligned}\dot{x}_1(t) &= v_1(t), \\ \dot{x}_2(t) &= v_2(t).\end{aligned}$$

The distance between the vehicles is  $d(t) = x_1(t) - x_2(t)$ . We say that the vehicles merge into a platoon at time  $t_m$  if the distance between them is lower than some predefined distance  $d_m$  for the first time,

$$t_m = \min \{t > t_0 | d(t) \leq d_m\},$$

and the position of the merge as the position of the follower vehicle at the time of the merge,

$$x_m = x_2(t_m) = x_1(t_m) - d_m.$$

In the simplest case, vehicles attempt to drive at some constant desired speeds  $u_1$  and  $u_2$  until they have successfully merged into a platoon, and then proceed

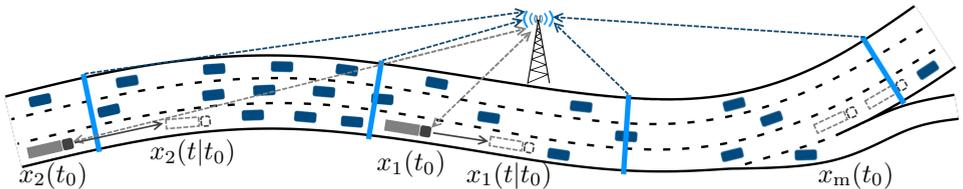


Figure 1.6: Platoon merging problem.

together driving at speed  $u_p$ . However, due to disturbances, the actual vehicle speed will deviate from the reference values, and we need to compensate. There are two parts of this problem that we study. Given constant reference speeds, in case we can predict  $v_1(t)$ ,  $t \geq t_0$ , and  $v_2(t)$ ,  $t \geq t_0$ , we may calculate the predicted time and place where the platoon merging will occur. Conversely, if we can control  $u_1(t)$  and  $u_2(t)$  within some range, we instead want to find the reference vehicle merging speeds that are optimal with regard to some metric. Both parts are covered in Chapter 5.

### 1.2.2 Congestion reduction problem

The ultimate goal of traffic control is minimization of some traffic performance metric, most often the total time spent (TTS) of all vehicles on the considered road segment. Denoting by  $n(t)$  the total number of vehicles on the segment, including vehicles queueing at its entrance, at time  $t$ , we may write

$$\text{TTS} = \int_{t_0}^{t_{\text{end}}} n(t) dt,$$

where  $t_0$  is the initial, and  $t_{\text{end}}$  the final considered time. The number of vehicles will evolve as

$$\dot{n}(t) = q^{\text{in}}(t) - q^{\text{out}}(t),$$

where  $q^{\text{in}}(t)$  is the aggregate inflow, and  $q^{\text{out}}(t)$  the aggregate outflow from the segment at time  $t$ . Given some initial number of vehicles  $n(t_0)$ , the number of vehicles at time  $t$  is

$$n(t) = n(t_0) + \int_{t_0}^{t_{\text{end}}} (q^{\text{in}}(t) - q^{\text{out}}(t)) dt.$$

If the initial number of vehicles present  $n(t_0)$  and the arrival of new vehicles  $q^{\text{in}}(t)$  cannot be controlled, minimizing TTS is equivalent to maximizing the early outflow from the road segment:

$$\begin{aligned} & \text{maximize} && J_{\text{cr}} \\ & \text{subject to} && \text{traffic dynamics} \end{aligned} \tag{1.1}$$

where  $J_{\text{cr}}$  is the cost function for the congestion reduction problem,

$$J_{\text{cr}} = \int_{t_0}^{t_{\text{end}}} (t - t_{\text{end}}) q^{\text{out}}(t) dt, \tag{1.2}$$

We use whatever form of actuation is available to shape the local traffic flows and affect the internal traffic state of the considered segment to achieve this.

In particular, in this thesis we will be solving two instances of this problem: bottleneck decongestion problem and stop-and-go wave dissipation problem.

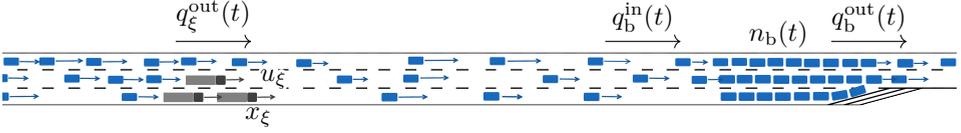


Figure 1.7: Bottleneck decongestion problem. The inflow to the bottleneck  $q_b^{in}(t)$  is regulated so that the bottleneck is kept in free flow by controlling the overtaking flow of moving bottlenecks  $q_\xi^{out}(t)$ .

### Bottleneck decongestion problem

The most straightforward, and most common, type of congestion reduction problem is bottleneck decongestion, shown in Figure 1.7. We focus on the stretch of road upstream of a stationary bottleneck, and the outflow from the road segment coincides with the outflow of the bottleneck,  $q^{out}(t) = q_b^{out}(t)$ , which depends on the bottleneck state:

$$q_b^{out}(t) = \begin{cases} \min\{q_b^{in}(t), q_b^{\max}\}, & m_b(t) = \text{FF}, \\ q_b^{\text{dis}}, & m_b(t) = \text{CD}. \end{cases}$$

Typically, when the bottleneck is in free flow,  $m_b(t) = \text{FF}$ , its capacity is higher than when it is congested,  $m_b(t) = \text{CD}$ , due to the capacity drop phenomenon,  $q_b^{\text{dis}} < q_b^{\max}$ . Therefore, it is desirable to regulate the traffic flow reaching the bottleneck to keep the bottleneck in free flow mode.

This can be achieved by controlling the speed  $u_\xi(t)$  and formation, which in turn determines the maximum overtaking flow  $q_\xi^{\text{cap}}(t)$ , of platoons in order to shape the flow of vehicles arriving at the bottleneck. The bottleneck decongestion problem can in this case be written as

$$\begin{aligned} & \underset{\substack{u_\xi(\cdot) \in [u^{\min}, u^{\max}] \\ q_\xi^{\text{cap}}(\cdot) \in [q^{\text{lo}}, q^{\text{hi}}]}}{\text{minimize}} && J_{\text{cr}} \\ & \text{subject to} && \begin{aligned} & \text{controlled vehicles dynamics and constraints} \\ & \text{stationary bottleneck mode transitions} \\ & \text{traffic dynamics with moving bottlenecks.} \end{aligned} \end{aligned}$$

In Chapter 6, we tackle this problem, with both stochastic and deterministic stationary bottleneck mode transitions.

### Stop-and-go wave dissipation problem

Another type of congestion reduction problem that we study here is stop-and-go wave dissipation, as shown in Figure 1.8. A stop-and-go wave consists of a zone of higher traffic density, with vehicles leaving from its downstream end at some

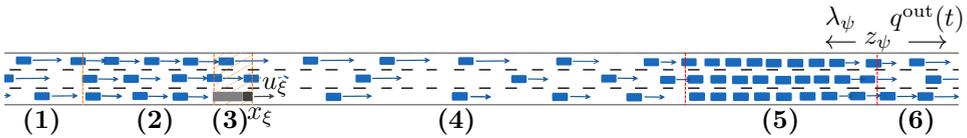


Figure 1.8: Congestion wave dissipation problem. The road can be split into six zones: (1) unaffected oncoming vehicles, (2) congestion upstream of the moving bottleneck, (3) moving bottleneck zone, (4) “starvation” zone downstream of the moving bottleneck, (5) traffic jam, and (6) vehicles discharging from the traffic jam.

discharging rate lower than the road capacity, and new vehicles arriving at its upstream end. The position of the downstream end of the stop-and-go wave is denoted  $z_\psi(t)$ , and it propagates upstream at some constant speed  $\lambda_\psi$ . The width of this zone of congested traffic shrinks or grows according to the difference of outflow and inflow.

In order to return the road to free flow, we use some connected vehicles that we have control over. The position of CAV  $\xi \in \Xi$  is denoted  $x_\xi(t)$ , and its speed can be controlled within some limits,  $u_\xi(t) \in [u^{\min}, u^{\max}]$ . If the controlled vehicle is slower than the surrounding traffic, it acts as a moving bottleneck and limits the traffic flow that can go past it, thus creating a zone of lower traffic density between it and the traffic jam (“starvation” zone), and also delaying some of the traffic flow that reaches the traffic jam. By doing this, a congestion is formed upstream of the moving bottleneck. However, if the vehicle is suitably controlled, this congestion will be less severe and harmful than the one in the congestion wave.

The stop-and-go wave dissipation problem consists of finding the optimal speeds for controllable vehicles, so that the stop-and-go waves are dissipated with minimum interruption to the traffic flow,

$$\begin{aligned} & \underset{u_\xi(\cdot) \in [u^{\min}, u^{\max}]}{\text{minimize}} && J_{\text{cr}} \\ & \text{subject to} && \text{controlled vehicles dynamics and constraints} \\ & && \text{traffic dynamics with stop-and-go waves and moving bottlenecks.} \end{aligned}$$

In Chapter 7, we focus on the simplest case of stop-and-go wave dissipation, with a single stop-and-go wave and a single controlled moving bottleneck, either available a priori, or created by accumulating connected vehicles. Then, in Chapter 8, we generalize this problem to the case when we have multiple stop-and-go waves and multiple controlled moving bottlenecks.

### 1.2.3 Traffic state reconstruction and model learning problem

If the full traffic density profile  $\rho(x, t)$  is not known, we can attempt to reconstruct it using local measurements, as shown in the example of real-time traffic state reconstruction in Figure 1.9. We assume that the connected vehicles  $\xi$  can communicate

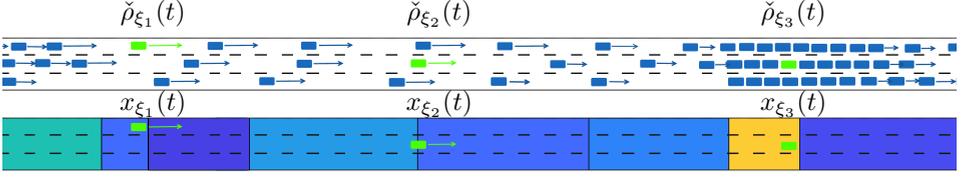


Figure 1.9: Traffic state reconstruction and model learning problem, showing the real traffic conditions top and reconstructed traffic state below. The traffic state estimate  $\hat{\rho}(x, t)$  (shown color-coded with brighter colors indicating denser traffic) is reconstructed according to the local traffic density measurements of each connected vehicle  $\check{\rho}_{\xi}(t)$ , shown in bright green. Other traffic measurements are used for model learning.

noisy measurements of the local traffic conditions,

$$\check{\rho}_{\xi}(t) = \rho(x_{\xi}(t), t) + \delta_{\xi}^o(t),$$

where  $\delta_{\xi}^o(t)$  represents the measurement noise.

Assuming the traffic evolves according to some first-order traffic model, we may use the past measurements of all available connected vehicles  $\xi \in \Xi$  over some time frame  $[t_0 - T, t_0]$  to find the traffic state reconstruction  $\hat{\rho}(x, t_0)$  which best fits the traffic model,

$$\begin{aligned} & \underset{\hat{\rho}(\cdot, t_0)}{\text{minimize}} && \text{traffic model deviation} \\ & \text{subject to} && \check{\rho}_{\xi}(t) = \hat{\rho}(x_{\xi}(t), t) + \delta_{\xi}^o(t), \quad \xi \in \Xi, \quad t \in [t_0 - T, t_0], \\ & && \hat{\rho}(x, t_0 - T) = \hat{\rho}_{\text{init}}(x), \end{aligned}$$

given some initial traffic state estimate at time  $t_0 - T$ ,  $\hat{\rho}_{\text{init}}(x)$ . In particular, we are interested in finding a traffic state reconstruction update rule that calculates the current state estimate  $\hat{\rho}(x, t)$  based on some past estimate  $\hat{\rho}(x, t - T)$ , and the traffic measurements  $\check{\rho}_{\xi}(t)$ ,  $\xi \in \Xi$  in time interval  $[t - T, t]$ .

Furthermore, if the parameters of the traffic model are not known, we may use additional traffic measurements, such as the local measurements of the traffic speed  $\check{v}_{\xi}(t)$ , traffic flow  $\check{q}_{\xi}(t)$ , or of the flow overtaking the connected vehicle  $\check{\omega}_{\xi}(t)$  to estimate the traffic model of the chosen form. In Chapter 8, both aspects of the problem are discussed, and methods for state reconstruction and model learning are proposed. The resulting state and model estimates are used for control. First, we focus only on traffic state reconstruction, with the traffic model known, and then on simultaneous traffic state reconstruction and model learning.

### 1.3 Thesis outline and contributions

In this section, we provide an overview of the thesis. The thesis consists of four parts:

- Introductions and Background,
- Modelling,
- Control,
- Conclusions.

This chapter is a component of the first part of the thesis. We describe each chapter's content and contribution, and indicate the publications on which they are based.

## Chapter 2 Background

In this chapter, we provide the background for the topics covered in the thesis. We discuss how the traffic congestion problem is addressed in the literature, and how the introduction of new technologies can change this. In Section 2.1, we review the classical traffic control methods, as a part of a wider overview of the techniques and technologies in the field of intelligent transportation systems (ITS). In particular, connected and autonomous vehicles are a relatively new technology that is likely to revolutionize the transportation sector. Next, in Section 2.2 we focus on one particular piece of the ITS, truck platooning. We specifically focus on the mutual influence that the platooning heavy-duty vehicles and the surrounding traffic have on each other, and discuss dynamic real-time platoon coordination. In order to be able to study these effects in depth, as well as be able to use truck platoons and other connected vehicles for traffic control, in Section 2.3 we give an overview of traffic modelling where multiple classes of vehicles can be incorporated. These different classes can, e.g., account for different vehicle types, such as passenger cars and trucks, or different levels of autonomy, such as human-driven, cruise-control-equipped, and autonomous. Then, in Section 2.4, we discuss the emerging field of Lagrangian traffic sensing and actuation. We give an overview about how we can use traffic measurements and trajectory data that the connected vehicles communicate with the infrastructure in order to get a picture of the traffic situation, as well as use these vehicles to actuate the desired control actions by communicating commands back to them. Finally, the content of the chapter is briefly summarized in Section 2.5.

## Part II Modelling

### Chapter 3 Basic traffic models

The second part of this thesis consists of different traffic models and their extensions. In this chapter, we start discussing the traffic models which will be used in the remainder of the thesis. Here we present the basic microscopic and macroscopic traffic models, that are well-known in the literature, which will be used as the basis for extensions. We start the chapter by discussing microscopic traffic models, in Section 3.1, as an introduction, and then transition to macroscopic traffic models. We introduce the Lighthill-Whitham-Richards (LWR) model in Section 3.2, and discuss its components, properties, and solutions. Next, the cell transition model (CTM) is introduced in Section 3.3, as a discrete counterpart of the LWR model. Then, in Section 3.4 we discuss the tandem queueing model, as an even higher level of abstraction, disregarding the spatial aspects of traffic. Finally, some prominent second-order traffic models that fall outside of the scope of this thesis are mentioned and outlined in Section 3.5, and the chapter is concluded with a summary in Section 3.6.

### Chapter 4 Traffic model extensions

In this chapter, we present the contributed extensions and modifications to the basic models given in the previous chapter, as well as introduce novel models. These extended models will be used to design and test control laws in the third part of this thesis. We open the chapter by discussing the relevant traffic phenomena in Section 4.1, which need to be captured in order to be able to study the Lagrangian traffic control, but are commonly not properly handled in the simpler classical traffic models. Namely, moving bottlenecks, capacity drop, and stop-and-go waves are discussed. We then first introduce the moving bottlenecks to the LWR model in Section 4.2.1, and in Section 4.2.2 propose an extension of the CTM that includes their influence. Next, another extension of the CTM is given in Section 4.3, the multi-class CTM. In it, the CTM is generalized in order to be able to handle different classes of vehicles (e.g. human-driven and autonomous) differently, as well as to capture the influence of platoons, moving bottlenecks, capacity drop, and stop-and-go waves. Then, in Section 4.4, the tandem queueing model is revisited and extended to allow for moving queues, which will correspond to slow-moving platoons acting as moving bottlenecks. This model is validated against microscopic simulations. Finally, the front tracking transition system model (FTTSM) is presented in Section 4.5, introducing the notion of wave-speed-bounded solutions in order to model the behaviour of stop-and-go waves, and modelling moving bottlenecks in a similar way as a zone with different flux function. The contributions of this chapter are summarized in Section 4.6.

This chapter is based on the author's work on traffic models in the following publications:

- M. Čičić and K. H. Johansson, “Traffic regulation via individually controlled automated vehicles: a cell transmission model approach,” in *21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, US, 2018
- M. Čičić and K. H. Johansson, “Energy-optimal platoon catch-up in traffic in moving bottleneck framework,” in *18th European Control Conference (ECC)*, Napoli, Italy, 2019, pp. 3674–3679
- M. Čičić and K. H. Johansson, “Stop-and-go wave dissipation using accumulated controlled moving bottlenecks in multi-class CTM framework,” in *IEEE 58th Conference on Decision and Control (CDC)*, Nice, France, 2019, pp. 3146–3151
- M. Čičić, X. Xiong, L. Jin, and K. H. Johansson, “Coordinating vehicle platoons for highway bottleneck decongestion and throughput improvement,” *IEEE Transactions on Intelligent Transportation Systems*, 2021, accepted, also presented at Transportation Research Board 99th Annual Meeting
- M. Čičić, M. Barreau, and K. H. Johansson, “Numerical investigation of traffic state reconstruction and control using connected automated vehicles,” in *IEEE 23rd Intelligent Transportation Systems Conference (ITSC)*, Rhodes, Greece, 2020, pp. 1–6
- M. Čičić and K. H. Johansson, “Front tracking transition system model for traffic state reconstruction and control,” in preparation

## Part III Control

### Chapter 5: Platoon merging coordination

The third part of this thesis deals with traffic and platooning control, and is opened by discussing platoon merging coordination. In this chapter, we address the problem of coordinating the process of vehicles catching up and forming a platoon while driving on a highway. First, we present the vehicle model in Section 5.1, and consider the reasons why the vehicles might have to deviate from their set reference speeds. Then, in Section 5.2 we discuss predicting the vehicle trajectories during the merging phase, and using these predicted trajectories to estimate when and where the vehicles will merge into a platoon. Here, we use the data obtained from an experiment to learn the vehicles’ control law and dynamics, and use the information about the road grade to achieve better prediction of platoon merging position. Second, in Section 5.3 we calculate optimal merging speeds for the vehicles attempting to form a platoon in traffic. We consider the mutual influence between the controlled vehicles and the traffic using CTM augmented with moving bottlenecks, and calculate energy-optimal merging speed pairs for the vehicles forming a platoon. The chapter is concluded and its contents summarized in Section 5.4.

This chapter is based on the following publications:

- M. Čičić, K.-Y. Liang, and K. H. Johansson, “Platoon merging distance prediction using a neural network vehicle speed model,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 3720–3725, 2017
- M. Čičić and K. H. Johansson, “Energy-optimal platoon catch-up in traffic in moving bottleneck framework,” in *18th European Control Conference (ECC)*, Napoli, Italy, 2019, pp. 3674–3679

## Chapter 6: Bottleneck decongestion

In this chapter, we study the problem of minimizing the total travel time on a road segment upstream of a stationary bottleneck. First, the bottleneck decongestion problem is analysed with generic type of switching between the free-flow and congested mode of the bottleneck in Section 6.1, demonstrating the benefit of keeping the bottleneck decongested. Next, in Section 6.2 we propose a reactive control scheme to deal with stochastic traffic breakdown at the stationary bottleneck, using created controlled moving bottlenecks to decongest it. Then, in Section 6.3 the simpler, deterministic case is considered, and prediction-base platoon coordination control scheme is designed and analysed. Both control approaches are tested in simulations, with the FTSM as the simulation model for the stochastic traffic breakdown case, and tandem queueing model with moving bottlenecks as the prediction model, and multi-class CTM as the simulation model for the deterministic traffic breakdown case. Finally, the content of the chapter is summarized in Section 6.4.

This chapter is based on the control parts of the following publications:

- M. Čičić, I. Mikolášek, and K. H. Johansson, “Front tracking transition system model with controlled moving bottlenecks and probabilistic traffic breakdowns,” in *The 21st IFAC World Congress*, 2020
- M. Čičić, X. Xiong, L. Jin, and K. H. Johansson, “Coordinating vehicle platoons for highway bottleneck decongestion and throughput improvement,” *IEEE Transactions on Intelligent Transportation Systems*, 2021, accepted, also presented at Transportation Research Board 99th Annual Meeting

## Chapter 7: Single stop-and-go wave dissipation

In this chapter, we study the simplest case of stop-and-go wave dissipation using controlled moving bottlenecks, assuming there is a single stop-and-go wave present, and that we use a single moving bottleneck. First, in Section 7.1, the stop-and-go wave problem is discussed in more detail, and its dissipation is first discussed using general traffic flow control framework, and then using controlled moving bottlenecks. Then, in Section 7.2, we tackle the simplest case, when there is already an available controlled moving bottleneck upstream of the stop-and-go wave, and

the control action consists of setting the reference speed for it such that the stop-and-go wave is dissipated as soon as possible, without unnecessary delays for the overall traffic. Finally, in Section 7.3 we generalize this approach to the case when a controlled moving bottleneck taking up a sufficient number of lanes is not available a-priori, but instead needs to first be created, by accumulating controllable connected vehicles at some point, and then used to actuate the traffic flow. Both control approaches are tested in simulations, using CTM with moving bottlenecks as the simulation model for the first one, and multi-class CTM as the simulation model for the second one. We close the chapter by summarizing its contributions and results in Section 7.4.

This chapter is based on the control parts of the following publications:

- M. Čičić and K. H. Johansson, “Traffic regulation via individually controlled automated vehicles: a cell transmission model approach,” in *21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, US, 2018
- M. Čičić and K. H. Johansson, “Stop-and-go wave dissipation using accumulated controlled moving bottlenecks in multi-class CTM framework,” in *IEEE 58th Conference on Decision and Control (CDC)*, Nice, France, 2019, pp. 3146–3151

## Chapter 8: Reconstruction-based multiple stop-and-go wave dissipation

In this chapter, we generalize the control problem studied in the previous one. Instead of dissipating a single stop-and-go wave using a single controlled moving bottleneck, whether a priori available or accumulated, in Section 8.1, we now study dissipating multiple, randomly arising stop-and-go waves, using multiple, randomly arriving connected vehicles that can act as controlled moving bottlenecks. Furthermore, we no longer assume that the traffic state is known, but instead rely on the local traffic measurements collected by the connected vehicles to first estimate and reconstruct the traffic density profile, and then use this reconstructed traffic state to calculate control actions for the connected vehicles. First, in Section 8.2, we present a simple traffic state reconstruction method, using traffic measurements from a dynamically selected subset of all available connected vehicles. Then, in Section 8.3, a more complex method is presented, relying on the front tracking transition system model. In addition to estimating the traffic state, we may also use these measurements to learn the traffic model, enabling us to design the whole control loop relying solely on the connected vehicles. Finally, in Section 8.4, we conclude the chapter.

This chapter is based on the control part of the following publications:

- M. Čičić, M. Barreau, and K. H. Johansson, “Numerical investigation of traffic state reconstruction and control using connected automated vehicles,” in

*IEEE 23rd Intelligent Transportation Systems Conference (ITSC)*, Rhodes, Greece, 2020, pp. 1–6

- M. Čičić and K. H. Johansson, “Front tracking transition system model for traffic state reconstruction and control,” in preparation

## Part IV Conclusion

### Chapter 9: Summary and future research

Finally, in this chapter we conclude the thesis, summarizing and discussing the results in Section 9.1, and outline some future and ongoing work in Section 9.2, indicating some possible directions in which this work can be extended.

### Publications

In summary, the majority of this thesis is based on the work presented in the following publications:

- M. Čičić, K.-Y. Liang, and K. H. Johansson, “Platoon merging distance prediction using a neural network vehicle speed model,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 3720–3725, 2017
- M. Čičić and K. H. Johansson, “Traffic regulation via individually controlled automated vehicles: a cell transmission model approach,” in *21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, US, 2018
- M. Čičić and K. H. Johansson, “Energy-optimal platoon catch-up in traffic in moving bottleneck framework,” in *18th European Control Conference (ECC)*, Napoli, Italy, 2019, pp. 3674–3679
- M. Čičić and K. H. Johansson, “Stop-and-go wave dissipation using accumulated controlled moving bottlenecks in multi-class CTM framework,” in *IEEE 58th Conference on Decision and Control (CDC)*, Nice, France, 2019, pp. 3146–3151
- M. Čičić, I. Mikolášek, and K. H. Johansson, “Front tracking transition system model with controlled moving bottlenecks and probabilistic traffic breakdowns,” in *The 21st IFAC World Congress*, 2020
- M. Čičić, X. Xiong, L. Jin, and K. H. Johansson, “Coordinating vehicle platoons for highway bottleneck decongestion and throughput improvement,” *IEEE Transactions on Intelligent Transportation Systems*, 2021, accepted, also presented at Transportation Research Board 99th Annual Meeting
- M. Čičić, M. Barreau, and K. H. Johansson, “Numerical investigation of traffic state reconstruction and control using connected automated vehicles,” in

*IEEE 23rd Intelligent Transportation Systems Conference (ITSC)*, Rhodes, Greece, 2020, pp. 1–6

- M. Čičić and K. H. Johansson, “Front tracking transition system model for traffic state reconstruction and control,” in preparation

The order of the author names reflects the workload, where the first author has the most important contribution. In all listed publications, all authors were actively involved in formulating the problems, developing the solutions, evaluating the results, and writing the paper.

Apart from the listed publications that were included in this thesis, the author has also been involved in the following publications, that are not a part of the thesis, but inspired some of the presented work:

- L. Jin, M. Čičić, S. Amin, and K. H. Johansson, “Modeling the impact of vehicle platooning on highway congestion: A fluid queuing approach,” in *Proceedings of the 21st International Conference on Hybrid Systems: Computation and Control (part of CPS Week)*. ACM, 2018, pp. 237–246
- G. Piacentini, M. Čičić, A. Ferrara, and K. H. Johansson, “VACS equipped vehicles for congestion dissipation in multi-class CTM framework,” in *18th European Control Conference (ECC)*, Napoli, Italy, 2019, pp. 2203–2208
- A. Ibrahim, M. Čičić, D. Goswami, T. Basten, and K. H. Johansson, “Control of platooned vehicles in presence of traffic shock waves,” in *IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, 2019, pp. 1727–1734
- L. Jin, M. Čičić, K. H. Johansson, and S. Amin, “Analysis and design of vehicle platooning operations on mixed-traffic highways,” *IEEE Transactions on Automatic Control*, 2020
- J. Liu, M. Barreau, M. Čičić, and K. H. Johansson, “Learning-based traffic state reconstruction using probe vehicles,” in *16th IFAC Symposium on Control in Transportation Systems*, Lille, France, 2021, submitted



## Chapter 2

# Background

THIS chapter provides some background on a number of topics relevant to the rest of the thesis. Since the thesis deals with control of automated vehicles and traffic, it falls within the broad scope of intelligent transportation systems (ITS). Platoons of passenger and heavy-duty vehicles are of particular interest, due to the extensive effect they have on the rest of the traffic, and the fact that they are likely one of the first formations of connected vehicles to be controllable from the infrastructure. In order to be able to do any type of traffic control using these special connected vehicles, we first need to understand how they interact with the rest of the traffic, necessitating models for mixed traffic situations. Finally, the central point of this thesis is exploiting the connected vehicles as Lagrangian sensors and actuators, providing us with new opportunities for traffic control. These topics are discussed in this chapter.

### 2.1 Intelligent transportation systems

The ITS use a wide variety of advanced technologies to improve safety, efficiency and performance of the transportation system, with reducing traffic congestion as one of its main goals. Maintaining free flow with high traffic density leads to an increased throughput, which reduces congestion levels and total travel time of all vehicles. However, controlling the traffic flow when the demand is close to road capacity is a very challenging task, since this is a metastable situation, and small perturbations can cause traffic breakdown and emergence of congestion. Apart from having a plethora of negative effects, traffic congestion is also known to reduce the efficiency of the traffic flow, and thus causing more traffic congestion to build up. Once traffic breakdown happens, the capacity of the part of the road where it arises will be reduced [37, 38, 39, 40, 41, 42], until the built-up congestion dissipates. This congestion remains anchored at some bottleneck, or propagates upstream as a stop-and-go wave, also known as wide moving jam or phantom jam [43, 44, 45, 46, 47, 48], leading to a significant reduction in throughput. Since vehicles need to decelerate or



(a) Ramp metering signal

Source: Wikimedia, Author:Patriarca12, CC-BY

(b) Variable speed limit signs

Source: geograph.org.uk, Author: David Dixon, CC-BY

Figure 2.1: Traditional traffic control methods.

come to a full stop when entering the stop-and-go wave, and afterwards accelerate as they leave it, their fuel consumption is increased, while the safety and comfort are decreased.

Conventionally, the problem of traffic congestion would be tackled either from the supply side, by increasing the capacity of the road network, adding more lanes or new links and routes, or from the demand side, by reducing the number of vehicles that use the road network, using various policy approaches such as congestion pricing [49] and other incentive schemes [50]. The first approach to congestion mitigation typically requires costly investment in transportation infrastructure, and it may produce undesirable side-effects such as induced demand [51, 52], as well as have a negative impact on the environment. Building new infrastructure is therefore a decreasingly attractive solution, and it is desirable to attempt to solve this problem by other, more cost-efficient, means. The second approach, attempting to reduce demand by congestion pricing and other similar means, is often controversial and difficult to implement from the political and legislative side, and its effectiveness largely depends on how the revenue is spent [53].

Apart from excess demand, traffic congestion can also be caused by poor traffic management [54], or conflicts between traditional and new traffic management methods [55]. Traditional traffic control methods had been successfully used since the sixties, with constant albeit slow improvements. Ramp metering [56, 57, 58, 59, 60], where traffic flow entering a highway from an on-ramp is regulated in order to keep the mainstream in free flow, is one of the prominent examples. The control action is either based on local measurements in the area close to the on-ramp, or is coordinated by some centralized controller in order to regulate multiple on-ramps in a wider area. Variable speed limits [61, 62, 63, 64] are another traffic control method, although they have historically been used primarily to improve

traffic safety through speed harmonization [65, 66]. By regulating the mainstream traffic flow such that some of the traffic flow is delayed, we are able to improve the efficiency by e.g. preventing off-ramps blockage, decongesting bottlenecks, or dissolving stop-and-go waves [67]. Even better results can be achieved by combined use of these measures [68, 69, 70]. In larger networks, where multiple viable routes connect some origin-destination pairs, route guidance [71, 72] at bifurcations along with providing information about queue lengths or travel times can also reduce congestion and delays.

These control strategies rely on traffic monitoring infrastructure consisting of inductive loops detectors, cameras, radars, etc. [73], and communicate or enforce their control actions by use of traffic lights or variable-message signs. When dealing with stationary, recurrent bottlenecks, such as on-ramps or lane drops, we may know in advance where the congestion will arise, thus allowing for pre-emptive installation of the necessary fixed equipment where it is required. However, if a non-recurrent bottleneck, such as a work zone, traffic incident etc., arises elsewhere, where such fixed equipment is not available, we will not be able to apply any of these control strategies to reduce congestion. Indeed, it is possible that the congestion will not be detected at all, until it propagates to a section of the road that is instrumented with traffic sensors. Although the cost of installing additional equipment is much lower than the cost of expanding or building new roads, it is not negligible, and we cannot expect the required equipment to be available wherever it is needed, thus limiting the applicability of these methods. While these relatively old traffic control strategies are still an important part of the broader, newer trend of ITS [74, 75], the focus has been steadily shifting towards using new technologies, and towards using the information from the vehicles themselves.

In order to compensate for scarcity of traffic sensors, as well as to improve the quality of the traffic state estimation, we may employ some filtering technique, such as the extended Kalman filter [76, 77, 78], using available traffic measurements. Apart from measurements coming from fixed traffic sensors, such as inductive loops, mobile sensors play an increasingly important role, using the trajectories and potential local measurements of connected vehicles as input. While connected vehicles, in the sense of connected and automated vehicles (CAVs) [79, 80], are still rare on the roads, even today virtually all vehicles are networked in some way via the mobile phones of their occupants. The vast amounts of location data collected from GPS-enabled mobile phones by Google, for example, has proven to be an invaluable resource [81, 82]. A number of different algorithms for traffic state estimation and reconstruction using mobile sensor data from connected probe vehicles have been proposed [83, 16, 84, 85], and their development is currently a very active field of research.

The proliferation of mobile phones has had another effect on the traffic flow, as the navigation apps have started to affect the routing decisions that drivers make in a significant way. Before navigation apps, route guidance could only be disseminated using variable message signs, requiring the installation of this equipment upstream of important diverges [86, 71]. Dynamic real-time routing indications are

now widely available to the drivers, via their mobile phones, although the effect of this availability of information is not necessarily positive for the traffic flow overall [87, 88, 72].

The effect that the “intelligent” vehicles will have on the traffic is not limited to their potential to be used as probe vehicles. Vehicles with varying degrees of autonomy, such as those equipped with adaptive cruise control (ACC) and collaborative ACC (CACC) [89], and platooning systems [19], are already partaking in the traffic flow, albeit still in low numbers. By virtue of being able to drive more smoothly or with lower headways [11, 32], as well as to stabilize the traffic flow in their lane [90, 91, 89, 92], these vehicles are able to improve the efficiency of the traffic.

A specific case of Lagrangian sensing and traffic state reconstruction, as well as the potential for using these CAVs as Lagrangian actuators for the macroscopic traffic, will be discussed in more detail in Section 2.4.

## 2.2 Platoons in traffic

With heavy-duty vehicle platooning [19] progressing persistently towards becoming a commonplace technology [18], studying and understanding the impact it will have on the overall traffic is becoming increasingly important. There have been a number of projects working towards realising truck platooning [93, 94], and this technology is slowly transitioning from academia to industry. Traditionally, platooning was primarily regarded as means of reducing the air drag acting on the vehicles [95, 96], and thus the fuel consumption, as well as having the potential to greatly reduce the work load on drivers [97]. Apart from these benefits, truck platooning is also expected have a positive impact on traffic efficiency through reducing the headways between vehicles [11, 98], alleviating the adverse effect trucks have on the traffic [99]. There has been much work done on controlling the vehicles inside a platoon [100, 101, 102, 103], and this technology has already been demonstrated in field tests [104, 105, 106].

However, 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 [107]. One identified problem pertains to the interaction between truck platoons and passenger cars close to on- and off-ramps, and bottlenecks in general [105, 108]. There is concern that long platoons might block access to an off-ramp, or from an on-ramp, forcing drivers to slow down excessively or cut into a platoon, resulting in significant disturbances for both the platoon and the rest of 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 [109] and macroscopic [32] frameworks.

Another important question that needs to be answered is how platoons should be formed, and how to make decisions on which vehicles should platoon with which

other vehicles. The simplest solution is the so-called spontaneous platoon formation [110], where heavy-duty vehicles form platoons if they find themselves in the vicinity of each other in the course of their trips, but do not otherwise change their plans in order to facilitate platooning. While this approach does lead to some fuel savings, compared to the case when no platoons are formed, it is severely limited by availability of suitable platooning partners. Therefore, some level of active platooning coordination is needed in order to fully realise the potential benefits.

Most proposed large-scale platooning solutions involve a layered control architecture [19]. On the higher layer, platooning coordinator plans the transport assignments and optimizes vehicle routes, including identifying and managing potential platoons. On the middle layer, vehicles receive their routes and generate their speed profiles, which the lower layer control is tasked to follow. Dynamic planning strategies have been proposed, with platooning coordinator matching and organizing vehicles into platoons [111]. Selected vehicles receive jointly fuel-optimal speed profiles and routes, and by following them, merge into a platoon and drive together for some time. However, this also means that, since the participating vehicles will have to deviate from their own optimal speed profiles, attempting to form a platoon entails higher fuel consumption during the merging phase. The hope is to offset this effect by fuel savings during the time the vehicles drive in the platoon [112]. If the platoon merging is delayed due to some unpredicted disturbance [113], or if the vehicles fail to merge into a platoon, the net energy consumption could be much higher than expected, potentially leading to more fuel being spent compared to the case when the vehicles would continue driving at their individual optimal speeds. It is therefore important to have a good prediction of when the platoon merging will be completed, so as to be able to calculate predicted energy savings and make a better informed decision on whether to attempt to form a platoon at all. This problem was studied in [114] and [115], as well as in [116]. In these papers, the authors did consider the influence of traffic, but did not study how to compensate for it.

An alternative to en route platoon formation is platoon formation at hubs [117, 118], where vehicles meet at hubs like freight terminals, parking areas, etc., wait for potential platooning candidates to arrive, and then continue the trip together as a platoon. The advantage of this approach is that it causes less disruption to the rest of the traffic, and can utilize mandatory truck driver rest periods. The downside is that it requires the vehicles to stop and wait for each other, limiting its flexibility, although the general framework can be extended to the case of en-route platoon formation if road segments themselves are considered as “hubs”, and driving slower considered as “waiting”.

If all of the vehicles in the platoon belong to the same fleet, the fact that the follower vehicles benefit more from platooning than the leader vehicle is not an issue. However, in case vehicles from multiple fleets are platooning, the leader vehicle will typically need to be incentivised in order to accept the deviation to its plan. Designing market-based systems to facilitate these exchanges and match-making is an open field of research [119, 120, 121].

Due to their large size and the existence of fleet management infrastructure, truck platoons are an ideal candidate for moving bottleneck control. Since they consist of heavy, slow-moving vehicles, truck platoons will act as moving bottlenecks with or without external control, and we may use the communication channels already in place to send centrally computed reference speeds and other control actions [19]. This way, using platoons as in-flow actuators for controlling the traffic flow, we are able to mitigate the negative effects trucks have on the traffic, and even improve the overall traffic situation. Apart from these positive effects on the traffic, truck platoons may improve the situation for themselves as well, leading to potentially less delay, smoother speed profiles, as well as increased travel time reliability. Furthermore, if the trucks are equipped with some sensors that can measure the overtaking traffic flow, they can act as probe vehicles and provide data for traffic state estimation. For these reasons, as well as in order to understand the influence that the surrounding traffic has on the platooning, traffic models which include the influence of special vehicles acting as moving bottlenecks are required.

### 2.3 Mixed traffic models

In order to represent, predict, and analyse the dynamics of traffic flow, including the interactions between some specific vehicles and the rest of the traffic, we need an appropriate model, that is complex enough to be able to capture all the relevant phenomena, but tractable enough that it can be used for the chosen purpose [122]. The behaviour of “intelligent” or autonomous vehicles, heavy-duty vehicles and platoons can significantly differ from the behaviour of human-driven passenger cars, at least in some specific scenarios, which motivates considering them as different flows. After the requiem [123] and subsequent resurrection [124] of the second-order traffic flow models, as the difference between the behaviours of different vehicle types becomes more pronounced, a specific type of traffic models that deal with mixed traffic have arisen. For an overview on different types of traffic models, the reader is referred to [125, 126, 127], and to [128] for a historical perspective.

The idea of treating different types of vehicles separately dates back to the traffic assignment problem [129], with the trucks and passenger vehicles as the most common choice for different vehicle classes. However, with the autonomous vehicles playing a more and more significant role in the traffic, classifying the vehicles by their driving behaviour is becoming more common, e.g. dividing the vehicles into the human-driven ones and ones enabled with various levels of autonomy. Even without considering autonomy, vehicles could be classified using the heterogeneous driving behaviour of the drivers, potentially explaining some known phenomena such as the reverse-lambda nature of the fundamental diagrams [130]. Even if all vehicles have the same driving behaviour, we might want to classify them based on their destination or route, in order to be able to correctly model their behaviour at off-ramps and diverges without using splitting ratios and similar statistical parameters.

The simplest way of incorporating multiple classes of vehicles is by directly assigning different driving behaviour to individual vehicles, like one can do in microscopic models [131]. In these models, the traffic evolution is described through the longitudinal (car-following) and lateral (lane-changing) behaviour of each single vehicle. This high level of complexity allows microscopic traffic models to replicate real-life traffic conditions with high fidelity, at the cost of requiring numerous parameters to be properly calibrated. Car-following (or follow-the-leader) models describe how drivers follow the preceding vehicle (leader). Some of the most well-known car-following models are Gazis-Herman-Rothery model [132], Gipps model [133] and Intelligent Driver Model [134, 135], along with its improved variant [125]. Recently, efforts have been made to model the car-following behavior by using artificial neural networks [136], specifically aimed at capturing some emergent phenomena such as stop-and-go waves. There exist many commercial traffic simulators that use microscopic traffic models, e.g. Aimsun [137] and PTV Vissim [138], as well as open-source traffic simulators such as SUMO [139]. Different car-following parameters can be assigned to different classes of vehicles [140], for example, by setting different reaction times to human drivers, ACC-enabled and CACC-enabled vehicles in Improved Intelligent Driver Model, as was done in [91].

Modelling different classes of vehicles in macroscopic traffic models is more challenging, and there are numerous different approaches to choosing the parameters that will differ across the classes. In [141] the authors present a generalization of the Lighthill-Whitham and Richards (LWR) traffic flow model, with different classes having different free flow speeds, and [142] gives a multi-class gas-kinematic traffic model. The model from [143] distinguishes between different classes of traffic by allowing for different reaction times for each (with automated vehicles having a shorter reaction time than human-driven vehicles), leading to congestion wave speed that depends on the ratio of automated vehicles in traffic. Conversely, in [144], different classes of traffic are allowed separate fundamental diagrams, and traffic flow is allocated between them based on how much space they take on the road. While in [144] the space the vehicle take was a constant, in [145] a model is proposed where the personal car equivalent of heavy vehicles is dynamically depending on the speeds of the vehicles. The model given in [146] captures the overtaking and creeping behaviour, where small vehicles are able to advance even though larger vehicles are not moving. Various control laws for multi-class traffic have been proposed, including VSL [147], boundary control [148], etc. Some other notable macroscopic multi-class traffic models include those presented in [149] and [150]. A specific type of multi-class models stems from the interpretation of the generalized Aw-Rascle-Zhang model [124], or generic second order traffic flow model [151], where the second conserved quantity is related to the fraction of autonomous or otherwise specific vehicles, or some other driver attribute [152, 153, 154]. Such models have been used, for example, in the framework of traffic state estimation [155].

While multi-class traffic models are appropriate for modelling the interaction of different vehicle classes, they are often, in their basic form, unable to describe the

effect individual vehicles, or formations of vehicles such as platoons, can have on the rest of the traffic. One notable way a single vehicle can affect the overall traffic is by acting as a moving bottleneck. If a vehicle moves slower than the surrounding traffic, it affects the traffic flow by limiting the number of vehicles that can pass it, as witnessed experimentally and empirically [156, 157, 158]. In order to model this effect, we may impose some additional local constraints on the traffic flow in the area close to the slow-moving vehicle, or consider the part of the road with the moving bottleneck to be described by a different flux function. This gives rise to a class of coupled ODE-PDE models [159, 160, 161, 162, 163], that yield themselves readily to Lagrangian control of the traffic flow, which will be discussed in the following section, and the majority of this thesis.

## 2.4 Lagrangian sensing and actuation

The inflexibility of classical traffic control methods, due to their reliance on stationary signalling and messaging equipment can be mitigated by using new technologies. We may expect to have the possibility to communicate control actions directly to the vehicles, either through an in-car advisory system in case of human drivers [164], or as commands to connected automated vehicles [165]. However, the performance of most control algorithms that rely on CAVs suffers from low market penetration rates, which is the situation that we are likely to have on the roads for quite some time [9]. For example, in case variable message signs are not available at the location of interest, we may attempt to actuate variable speed limit controls by directly communicating it to the vehicles, but in case of low market penetration rate, this would amount to the case when we have a low compliance rate, significantly reducing the effectiveness of the control method [166]. Therefore, there is an acute need to develop control approaches that can be used in this intermediate period when CAVs become available, but not yet in large numbers.

Lagrangian approaches seem particularly well suited for this case. As opposed to the Eulerian traffic control paradigm, where the actuation and sensing happens at the boundaries of some segments of the road, the Lagrangian paradigm focuses on what happens inside the traffic flow, centred on some specific vehicles. Notable examples of Eulerian mainstream traffic control include boundary control of various types (e.g. explicit state feedback [167], PI control [168] or backstepping [169]), and variable speed limits [67, 63]. The advantage of Lagrangian traffic control is that, since the traffic flow is controlled using mobile actuators in form of CAVs, it can be applied anywhere on the road, including uninstrumented areas away from recurrent bottlenecks. Even if initially there are no usable CAVs in the area where control action should be applied, some CAVs could be re-routed and coordinated to first accumulate in the area of interest, and then actuate the traffic flow. This approach in particular is exploited in [26] and Section 7.3.

One setting where the Lagrangian approach comes naturally is oscillation dampening, using autonomous vehicles to stabilize the traffic flow. This problem is related



Figure 2.2: A rolling roadblock on the M40 motorway in England, implemented by a Highways England Traffic Officer. Source: Wikimedia, author: Chris McKenna CC BY-SA

to platoon string stability [103, 170], where all vehicles in a platoon are considered to be controlled, with the distinction that only a small subset of vehicles would be CAVs, and the rest would be human-driven vehicles [171, 172]. A sizeable body of work has been dedicated to dissipating stop-and-go waves in ring road setting, demonstrated experimentally in [46]. Stabilization of the flow on a ring road was studied theoretically in [173, 174, 175], and also demonstrated in experiments [176]. Assuming no overtaking, a reinforcement learning approach to bottleneck decongestion was proposed in [177]. These control laws are strictly speaking not Lagrangian in the same sense as is considered here, since the rest of the traffic flow is also described as individual vehicles instead by a conservation law, but the mechanism of their actuation is similar to what would be used in macroscopic Lagrangian control.

The idealised setting of ring roads, or a string of vehicles serves to demonstrate how CAVs can improve the traffic situation in their own lane, assuming there is no overtaking. However, in real-life applications, stabilizing a single lane of a multi-lane highway might be insufficient. In this case, we need to consider the effect of the CAVs, acting as Lagrangian actuators, on the whole traffic flow, in a macroscopic setting. A form of Lagrangian actuation, the so-called rolling roadblock, or traffic break [178], shown in Figure 2.2, is already used to restrict the flow of traffic in some hazardous conditions, though this is typically done for safety, not in order to increase the efficiency of the traffic flow. In case a less severe traffic control action is applied, with only some of the lanes blocked by CAVs, this approach can be called moving bottleneck control, and is one of the most direct types of Lagrangian

actuation of traffic flow. Similar jam-absorption driving strategies have also been proposed in the literature [21, 22].

While moving bottlenecks are usually seen as detrimental to traffic efficiency, the prospect of controlling them for traffic regulation has attracted some attention lately. If we are able to control the speed of a moving bottleneck, and potentially its formation, changing how many lanes it obstructs, we are also able to control and shape the traffic flow that is overtaking it, within some range. This approach effectively emulates ramp metering or variable speed control, achieving a similar type of regulation without the need for additional fixed equipment. In [179], the authors show that a controlled moving bottleneck can reduce fuel consumption of the vehicles in traffic at a bottleneck. A similar idea was explored in [180], using a model-predictive control approach, and then in [181], using multiple controlled moving bottlenecks. An optimal traffic control problem, again minimizing fuel consumption, was addressed in [182], and a fast algorithm based on an extension of the semi-analytical solutions to arbitrary Hamilton-Jacobi equations is used in [162]. This control approach is used throughout this thesis, and in [24, 25, 26, 27, 28, 31, 29].

Apart from actuation, recently there has been a resurgence of interest in the Lagrangian approach to traffic control for sensing and traffic state estimation and reconstruction. This approach has its roots in the moving observer method [183, 184], where the traffic flow and speed is calculated based on the trajectories of some “floating vehicles”, which overtake as many vehicles as it was overtaken by. These methods have been established for a long time, but did not see broad usage, due to the need to have vehicles with drivers and observers, and the complexity of the set-up, driving up the costs. Over time, there have been developments of this method, using data acquired from taxis and vehicles equipped with GPS receivers [185, 83]. However, with the proliferation of GPS-equipped mobile phones came the availability of vast amounts of data that can be used for this purpose [16, 186], reinvigorating this field of research.

There has been much work done on traffic state estimation over the years [187]. In [188], traffic state estimation was done based on the observed spacing and position data of probe vehicles. The traffic state of mixed flow of human-driven vehicles and CAVs was estimated using probe vehicle data and some stationary sensors in [78]. A combination of stationary sensors and floating car data was used in [189, 190, 191]. The theoretical treatment of the problem of reconstructing the traffic state using only traffic data along the trajectories of probe vehicles is given in [192, 193]. In addition to providing information about the traffic state, these local measurements may be used to learn the models governing the behaviour of the background traffic and the influence of the CAVs on it. Additionally, the use of local data allows us to adapt to any changes in these models, e.g. due to changing weather conditions [194]. The problem of traffic state reconstruction and model learning based on the local data acquired from the CAVs is discussed in Chapter 8, and in [28, 29].

## 2.5 Summary

In this chapter we have provided a number of references relevant to the scope of the thesis. A broad overview of the relevant topics includes traffic congestion and related phenomena, classical traffic control and other elements of the ITS, heavy-duty vehicle platooning, traffic models that capture the interaction of different vehicle classes, and finally, Lagrangian sensing and actuation for traffic control. All of the mentioned topics are actively being researched, and are likely to garner even more interest in the future.



**Part II**

**Modelling**



## Chapter 3

# Basic traffic models

MODELLING the traffic flow has been an active area of research for over half a century. Before discussing the extended traffic models that will be used for traffic control, in this chapter we present some basic microscopic and macroscopic traffic models, which will be used as the foundation for these extensions. For a more in-depth view on traffic models, the reader is referred to [127]. Even though microscopic traffic models are not used directly, we briefly introduce them to provide an explanation for the macroscopic traffic modelling. Macroscopic models have the advantage of being comparatively easy to simulate and analyse, which makes them suitable for traffic control design. The two most common macroscopic traffic models, the Lighthill-Whitham-Richards (LWR) model and the cell transmission model (CTM), are presented. Finally, the tandem queueing model can be seen as the simplification of the CTM, abstracting away the spatial description of the traffic state, and instead focusing only on the queue lengths at some bottlenecks.

### 3.1 Microscopic traffic models

As opposed to macroscopic traffic models, which deal with the aggregate characteristics of traffic flows, microscopic traffic flow models describe the behaviour of each individual vehicle separately. The resulting complexity enables these models to capture practically all traffic phenomena that happen in human-driven, autonomous, or mixed traffic, but also makes them computationally expensive, and hard to analyse and calibrate. Regardless, microscopic traffic models have seen wide application in commercial and open source traffic simulators, such as Aimsun, PTV Vissim, and SUMO. These traffic simulation engines are widely understood to be a good representation of the real traffic when properly calibrated.

The motion of the vehicles is split into the longitudinal and lateral component. The lateral motion [195], i.e. lane-changing, is arguably more difficult to model than the longitudinal, car-following motion, since it depends on the intentions and subjective preferences of individual drivers, and is typically governed by

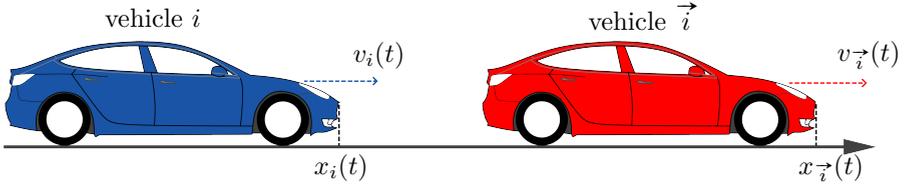


Figure 3.1: A car-following vehicle pair, with vehicle  $i$  as the follower.

some heuristics. Here we focus more on the car-following behaviour, since this aspect can be more readily related to the macroscopic traffic behaviour, and assume that lane-changing happens so that the traffic lanes are efficiently utilized.

### 3.1.1 Car-following models

Considering a single lane on the road, the longitudinal state equations for vehicle  $i$  can be written as

$$\begin{aligned}\dot{x}_i(t) &= v_i(t), \\ \dot{v}_i(t) &= a_i(t),\end{aligned}$$

with  $x_i(t)$  as the position of the vehicle,  $v_i(t)$  its velocity, and  $a_i(t)$  its acceleration. We denote the index of the leader vehicle immediately downstream of vehicle  $i$  as  $\vec{i}$ ,  $x_{\vec{i}}(t) - l_i > x_i(t)$  for all  $i$ , where  $l_i$  is the length of vehicle  $i$ . The driving behaviour of vehicle  $i$  typically depends on the motion of its leader vehicle  $\vec{i}$ , governed by some car-following model.

One of the simplest commonly used car-following models is the Gazis-Herman-Rothery, or Stimulus-Response model [132], with the acceleration of vehicle  $i$  given by

$$a_i(t) = cv_i^m(t) \frac{v_{\vec{i}}(t-T) - v_i(t-T)}{(x_{\vec{i}}(t-T) - l_{\vec{i}} - x_i(t-T))^p},$$

where  $c$ ,  $p$ ,  $m$ , and  $T$  are the model parameters to be calibrated. Although simple and capable of capturing most of the behaviour of drivers in congestion, this model does not provide a reasonable representation of drivers' behaviour in free flow, since the vehicles will adjust their speed according to the speed of the preceding vehicle even if it is very far away.

Another widely used car-following model is the so-called Intelligent Driver Model (IDM) [134, 135], and its modification, the Improved Intelligent Driver Model (IIDM) [125]. In IIDM, the acceleration is given by

$$a_i(t) = a^{\max} \left( 1 - \left( \frac{v_i(t)}{v_i^*(t)} \right)^{\delta_2} - \left( \frac{g_i^*(t)}{g_i(t)} \right)^{\delta_1} \right),$$

where  $v_i^*(t)$  is the desired maximum velocity of the vehicle,  $a^{\max}$  the maximum acceleration, and  $\delta_1$  and  $\delta_2$  positive parameters. The gap between vehicle  $i$  and the vehicle  $\vec{i}$  that it is following is  $g_i(t) = x_{\vec{i}}(t) - x_i(t)$  and the desired gap  $g_i^*(t)$  is given by

$$g_i^*(t) = g_i^{\min} + \max \left\{ 0, v_i(t)\tau_i + \frac{v_i(t)(v_i(t) - v_{\vec{i}}(t))}{2\sqrt{a_i^{\max}b_i}} \right\},$$

where  $\tau_i$  is the reaction time and  $b_i$  the desired vehicle deceleration. In order to allow for smoother deceleration when  $v_i(t) > v_i^*(t)$ , the IIDM modifies the acceleration of the vehicle to

$$a_i(t) = a_i^*(t) \left( 1 - \left( \frac{g_i^*(t)}{g_i(t)} \right)^{\delta_1 \frac{a_i^{\max}}{|a_i^*(t)|}} \right),$$

where the reference acceleration of vehicle  $i$  is given by

$$a_i^*(t) = \begin{cases} a_i^{\max}, & g_i(t) \leq g_i^*(t) \\ a^{\max} \left( 1 - \left( \frac{v_i(t)}{v_i^*(t)} \right)^{\delta_2} \right), & g_i(t) > g_i^*(t), v_i(t) \leq v_i^*(t) \\ -a^{\max} \left( 1 - \left( \frac{v_i^*(t)}{v_i(t)} \right)^{\delta_2} \right), & g_i(t) > g_i^*(t), v_i(t) > v_i^*(t). \end{cases}$$

Effectively, if  $g_i(t) \geq g_i^*(t)$ , the vehicle accelerates to its desired speed, and if  $g_i(t) < g_i^*(t)$ , the vehicle controls its speed so that the gap is stabilized around its desired value. Reaction time  $\tau_i$  and minimum acceptable gap  $g_i^{\min}$  depend on the characteristics of individual drivers. The vehicle speed equilibrium  $a(t) = 0$  is achieved in two cases,

- $g_i(t) > g_i^*(t)$  and  $v_i(t) = v_i^*(t)$ , corresponding to free flow, or
- $g_i(t) = g_i^*(t)$  and  $v_i(t) = v_{\vec{i}}(t)$ , corresponding to congestion.

In free flow, the inter-vehicular distances are large enough and vehicles can achieve their desired speeds, regardless of the speed of other vehicles. In congestion, the speed of each vehicle is determined by the speed of the vehicle in front of it, and collectively, all vehicle speeds will depend on the traffic density. The space that vehicle  $i$  takes on the road, under congested equilibrium conditions, is  $g_i^{\min} + l_i + v_{\vec{i}}\tau_i$ , and if all vehicles have the same characteristics,  $g_i^{\min} = g^{\min}$ ,  $l_i = l$ ,  $\tau_i = \tau$ , and constant desired speed  $v_i^*(t) = V$ , the equilibrium traffic speed and density are linked by

$$\rho = \frac{1}{g^{\min} + l + v\tau}, \quad v \leq V.$$

We denote the jam density (traffic density if  $v = 0$ ) as  $P$ ,

$$P = \frac{1}{g^{\min} + l},$$

and the critical density (traffic density if  $v = V$ ) as  $\sigma$ ,

$$\sigma = \frac{1}{g^{\min} + l + V\tau} = \frac{P}{1 + PV\tau}.$$

Consequently, we may write the dependence of the equilibrium speed on the equilibrium density as

$$\mathcal{V}(\rho) = \begin{cases} V, & \rho \leq \sigma, \\ \frac{1}{\tau} \left( \frac{1}{\rho} - \frac{1}{P} \right), & \rho > \sigma, \end{cases}$$

Defining the equilibrium flow as the product of the equilibrium density and its corresponding speed  $Q(\rho) = \mathcal{V}(\rho)\rho$ , the IIDM yields a flow-density relation, also known as the fundamental diagram,

$$Q(\rho) = \begin{cases} V\rho, & \rho \leq \sigma, \\ W(P - \rho), & \rho > \sigma, \end{cases}$$

defined for  $\rho \in [0, P]$ , and with  $W$  denoting the congestion wave speed

$$W = \frac{V\sigma}{P - \sigma} = \frac{1}{P\tau}.$$

Note that all of the parameters of the car-following model can also be taken as different for each vehicle, allowing us to represent a variety of different driving behaviours. The aggregate traffic parameters that can be derived from these microscopic traffic models are also used in characterising the macroscopic traffic models.

### 3.1.2 Heterogenous car-following

So far, the case when the driving behaviour of all vehicles is homogeneous was analysed. We now briefly consider the situation when there are multiple different classes of vehicles with different driving behaviour. For conciseness, we focus on the case when there are two classes of vehicles,  $a$  and  $b$ , but the same rationale can be extended to an arbitrary number of different classes.

First, let each of the classes have a different reaction time,  $\tau^a$  and  $\tau^b$ , and the same free flow speed  $V$ . For example, if class  $a$  vehicles are automated and class  $b$  vehicles human driven, we can expect to have  $\tau^a < \tau^b$  [91], and the free flow speed will depend on the speed limit of the road. If the traffic consists solely of vehicles of one class  $\kappa$ , the critical density and congestion wave speed are

$$\sigma^\kappa = \frac{P}{1 + PV\tau^\kappa}, \quad W^\kappa = \frac{1}{P\tau^\kappa}.$$

Then the aggregate critical density of the mixed traffic is

$$\sigma(r) = \frac{P}{1 + PV(r\tau^a + (1-r)\tau^b)},$$

where  $r$  is the ratio of class  $a$  vehicles in the total number of vehicles, i.e. the critical density will range from  $\sigma^b$  to  $\sigma^a$  depending on which vehicle class is more prevalent.

The case when each vehicle class has a different desired speed,  $V^a$  and  $V^b$ , is much harder to analyse. In the trivial case, when there is a single lane and no overtaking is allowed, the vehicles with the lower desired speed simply force the other class of vehicles to match their speed. If there are  $N_L$  lanes available,  $N_L > 1$ , the traffic situation will depend strongly on the lane-changing behaviour of the vehicles of different classes, and we need to make some assumptions in order to make the analysis tractable. Note that the critical density  $\sigma$  and other parameters of the traffic flow assume there is a single lane, thus the critical density of the road with  $N_L$  lanes would be  $N_L\sigma$ . Let  $\tau^a = \tau^b$ ,  $V^a > V^b$ , and assume that in light traffic, the vehicles of different classes arrange themselves so that the road is utilized efficiently, i.e. the slow vehicles do not spread to the faster lanes unless the whole road is congested and therefore do not conflict with the fast vehicles. Then, in case of light traffic, if

$$\left\lceil \frac{\rho^a}{\sigma^a} \right\rceil + \left\lceil \frac{\rho^b}{\sigma^b} \right\rceil \leq N_L, \quad (3.1)$$

where  $\rho^a = r\rho$  and  $\rho^b = (1-r)\rho$ , free flow can be maintained with each class of vehicles occupying  $\lceil \rho^k/\sigma^k \rceil$  lanes, with

$$v^a = V^a, \quad v^b = V^b. \quad (3.2)$$

Conversely, in case of heavy traffic, if

$$\rho \geq \frac{P}{1 + V^b P \tau}, \quad (3.3)$$

there will be congestion regardless of the ratio of class  $a$  vehicles, and the speed of both classes of vehicles is

$$v^a = v^b = \frac{1}{\tau} \left( \frac{1}{\rho} - \frac{1}{P} \right). \quad (3.4)$$

Between these two regimes, if  $\rho$  and  $r$  satisfy neither (3.1) nor (3.3), the equilibrium speeds  $v^a$  and  $v^b$  will also depend on the lane-changing behaviour and on the ratio of vehicles of each class. We can expect  $v^a$  and  $v^b$  to lie somewhere between (3.2) and (3.4). Note that if the number of lanes  $N_L$  is large, we can expect this effect to be less noticeable, since the vehicles of different classes will be able to distribute the lanes among them more efficiently. However, in case there is a large number of lanes, other traffic phenomena originating from lane-changing behaviours can arise, such as forced lane changes in order to position the vehicle in the right lane to exit the road, etc.

### 3.2 Lighthill-Whitham-Richards model

The oldest macroscopic traffic model is the Lighthill-Whitham-Richards (LWR) model [196, 197]. Although it originated in the 50s, this model, and its numerous extensions, are still widely used. The model consists of a conservation law, i.e. a first-order nonlinear partial differential equation

$$\partial_t \rho(x, t) + \partial_x q(x, t) = 0.$$

The traffic density  $\rho(x, t)$  is the conserved quantity, and  $q(x, t) = \rho(x, t)v(x, t)$  represents the traffic flow, with  $v(x, t)$  as the traffic speed. In first-order macroscopic traffic models, such as the LWR model, the traffic flow is directly determined by the traffic density,

$$q(x, t) = Q(\rho(x, t)),$$

or equivalently, the traffic flow is given by the hydrodynamic equation,

$$q(x, t) = \rho(x, t)v(x, t),$$

where the traffic speed  $v(x, t)$  is a function of the traffic density,

$$v(x, t) = \mathcal{V}(\rho(x, t)).$$

The LWR model can thus be written

$$\partial_t \rho(x, t) + \partial_x Q(\rho(x, t)) = 0, \tag{3.5}$$

or equivalently,

$$\partial_t \rho(x, t) + \partial_x (\rho(x, t)v(x, t)) = 0, \tag{3.6}$$

with  $v(x, t) = \mathcal{V}(\rho(x, t))$ . The function  $Q(\rho)$  is known as the fundamental diagram, or the traffic flux function.

An alternative approach to solving the LWR model is through using the cumulative vehicle counts, i.e. the number of vehicles that have passed through some point  $x$  until time  $t$ , instead of traffic densities as the state of the system. This cumulative vehicle count is also known as the Moskowitz function  $M(x, t)$ , and can be related to the traffic density and traffic flow through

$$\begin{aligned} \rho(x, t) &= -\partial_x M(x, t), \\ q(x, t) &= \partial_t M(x, t). \end{aligned} \tag{3.7}$$

Substituting (3.7) into (3.5) yields the Hamilton-Jacobi PDE

$$\partial_t M(x, t) - Q(\partial_x M(x, t)) = 0, \tag{3.8}$$

where the flux function  $Q$  is the Hamiltonian. This alternative formulation of the LWR has some advantages, since it can make use of well-established numerical methods for finding the solution to (3.8). Fast Lax-Hopf algorithms for this purpose have been proposed [198, 199], even including the influence of moving bottlenecks [162].

### 3.2.1 Flux function

Many different forms of the fundamental diagrams are used in the literature [201], but probably the two most commonly used ones are the Greenshields [202] flux function and Newell-Daganzo [203, 204] (triangular or trapezoidal) flux function.

In Greenshields fundamental diagram, traffic speed is given by

$$\mathcal{V}(\rho) = V \left(1 - \frac{\rho}{P}\right),$$

where  $V$  denotes the free flow speed, at which the vehicles would be travelling if the rest of the traffic did not affect them. Using this expression to model average traffic speed, makes the traffic flow  $Q(\rho) = \rho\mathcal{V}(\rho)$  a parabolic function,

$$Q(\rho) = V \left(\rho - \frac{\rho^2}{P}\right),$$

defined for  $\rho \in [0, P]$ , where  $P$  is the maximum, jam traffic density at which the vehicles are bumper-to-bumper and stop moving. This flux function is shown in Figure 3.2a. Since this function is once continuously differentiable, (3.5) is a hyperbolic conservation law. This allows us to use the broad body of literature that deals with such systems (see, for example [205, 206]). However, while this flux function is used extensively in theory, it does not reflect the traffic flow on highways particularly well.

Newell-Daganzo flux function is a piecewise linear function (see Figure 3.2b), and is given by

$$Q(\rho) = \min(V\rho, q^{\max}, W(P - \rho)),$$

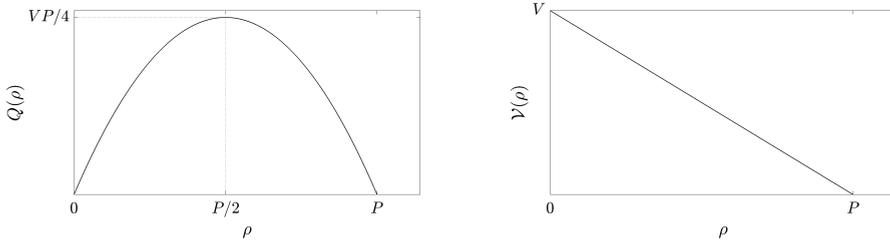
defined for  $\rho \in [0, P]$ , where  $W$  is the backward congestion wave propagation speed (i.e., the negative slope in congested mode,  $\rho > \sigma$ ) and  $q^{\max}$  some maximum traffic flow. We denote by  $\sigma$  the critical density at which  $V\sigma = W(P - \sigma)$ , and take  $q^{\max} = V\sigma$ , so that

$$Q(\rho) = \begin{cases} V\rho, & 0 \leq \rho \leq \sigma, \\ W(P - \rho), & \sigma < \rho \leq P, \end{cases} \quad (3.9)$$

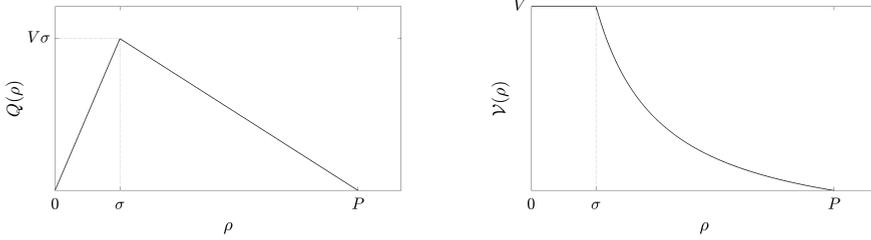
and the traffic speed dependence on traffic density is

$$\mathcal{V}(\rho) = \begin{cases} V, & 0 \leq \rho \leq \sigma, \\ W \left(\frac{P}{\rho} - 1\right), & \sigma < \rho \leq P. \end{cases}$$

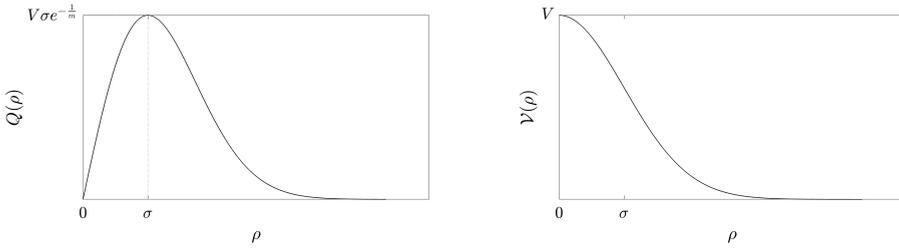
Note that Newell-Daganzo flux function is not continuously differentiable, but an arbitrary smoothed version of it is, so (3.5) will be a limit case of a hyperbolic conservation law. Although very simple, this flux function can represent the highway traffic flow reasonably well, and its parameters can easily be estimated from the traffic measurements.



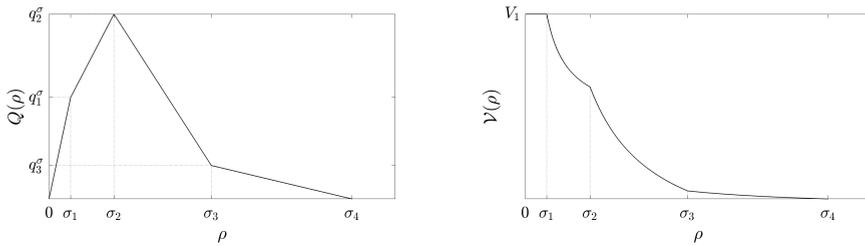
(a) Greenshields



(b) Newell-Daganzo



(c) FD from [200]



(d) Piecewise-linear approximation of the FD from [200]

Figure 3.2: Various flux functions and their corresponding speed-density relations.

Another commonly used flux function is the one proposed in [200], given through the traffic speed-density relation

$$\mathcal{V}(\rho) = V \exp\left(-\frac{1}{m} \left(\frac{\rho}{\sigma}\right)^m\right), \quad (3.10)$$

with the fundamental diagram thus given by

$$Q(\rho) = \rho V \exp\left(-\frac{1}{m} \left(\frac{\rho}{\sigma}\right)^m\right), \quad (3.11)$$

as shown in Figure 3.2c. The additional parameter  $m$  can be used to tune the shape of the fundamental diagram. This fundamental diagram has seen wide use, notably in the METANET model [207].

The piecewise-linear form of the Newell-Daganzo flux function can be very convenient for some numerical solution methods, such as front tracking. More general piecewise-linear flux functions, with more than one breakpoint, can also be handled with little increase in complexity. In practice, any smooth flux function can be arbitrarily closely approximated by a piecewise-linear function, and we may use

$$Q(\rho) = \begin{cases} V_1\rho, & 0 \leq \rho \leq \sigma_1, \\ Q(\sigma_1) + V_2(\rho - \sigma_1), & \sigma_1 < \rho \leq \sigma_2, \\ \vdots & \vdots \\ Q(\sigma_{i-1}) + V_i(\rho - \sigma_{i-1}), & \sigma_{i-1} < \rho \leq \sigma_i, \\ \vdots & \vdots \\ Q(\sigma_m) + V_m(\rho - \sigma_m), & \sigma_m < \rho \leq \sigma_{m+1}, \\ 0, & \rho > \sigma_{m+1}, \end{cases} \quad (3.12)$$

where  $Q(\sigma_i) > 0$ ,  $i = 1, \dots, m$ , and  $Q(\sigma_m) + V_m(\sigma_{m+1} - \sigma_m) = 0$ , as a fundamental diagram. An example of such fundamental diagram, approximating (3.11) is shown in Figure 3.2d.

In the remainder of this section, we focus on the simplest case of the LWR model, with triangular flux function (3.9). The general piecewise-function (3.12) will be discussed in more detail in Section 4.5, along with the procedure for finding the solution to (3.5) with (3.12) using front tracking.

### 3.2.2 Riemann problem

Since the flux function (3.9) is piecewise linear, finding the entropy solution to (3.5) will include solving Riemann problems at the possible discontinuities in  $\rho(x, t)$ , either due to the initial conditions, or arising as shocks. The Riemann problem is a Cauchy problem (problem of finding a solution to a PDE given initial conditions) in the particular case when the initial conditions are given as

$$\rho(x, 0) = \begin{cases} \rho_-, & x < 0, \\ \rho_+, & x > 0. \end{cases} \quad (3.13)$$

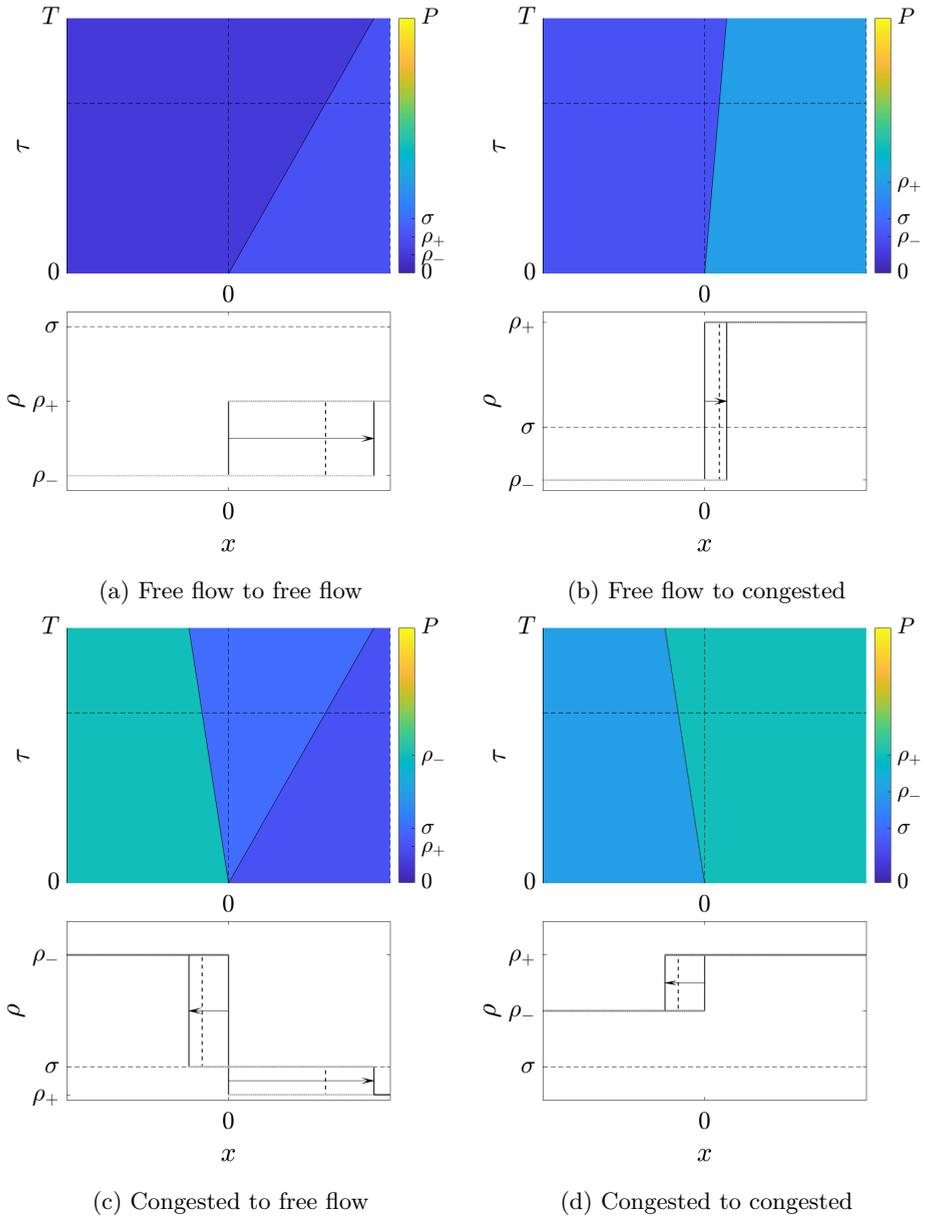


Figure 3.3: Riemann problem solutions for the four cases of initial conditions. The solutions for densities  $\rho(x, t)$  are shown color-coded, with space on x-axis and time on y-axis, and on lower figures, we show three snapshots of the solution with space on x-axis and traffic density on y-axis. The direction in which the wavefronts travel over time is indicated with an arrow.

This corresponds to assuming we have a sudden jump in traffic density at  $x = 0$  and looking at the evolution of traffic density around it. The solution of this problem will be a self-similar function of the form  $\rho(x, t) = f(x/t)$ .

If  $\rho_- = \rho_+$ , the initial conditions are not discontinuous, and the solution will stay constant,  $\rho(x, t) = \rho_- = \rho_+$ . Otherwise, the solution can either be a shock or a rarefaction wave. A shock is a solution of the form

$$\rho(x, t) = f(x/t) = \begin{cases} \rho_-, & x/t < \lambda, \\ \rho_+, & x/t > \lambda, \end{cases}$$

where  $\lambda = \Lambda(\rho_-, \rho_+)$  is the Rankine-Hugoniot transition speed between  $\rho_-$  and  $\rho_+$ ,

$$\Lambda(\rho_-, \rho_+) = \frac{Q(\rho_+) - Q(\rho_-)}{\rho_+ - \rho_-}.$$

We call such a moving discontinuity a wave front or simply a front. A rarefaction wave is a solution of the form

$$\rho(x, t) = f(x/t) = \begin{cases} \rho_-, & x/t \leq a, \\ r(x/t), & a \leq x/t < b, \\ \rho_+, & x/t > b, \end{cases}$$

where  $r(x/t)$  is a monotonic function related to the lower convex envelope of  $Q$  if  $\rho_- < \rho_+$  or the upper concave envelope of  $Q$  if  $\rho_- > \rho_+$ .

In the particular case of Newell-Daganzo flux function (3.9), the solution to the Riemann problem ((3.5), (3.13)) for  $\rho_- \neq \rho_+$  will consist of one or two wave fronts radiating from the discontinuity, depending on the density upstream of the discontinuity,  $\rho_-$  and downstream,  $\rho_+$ . In case we have congested traffic upstream  $\rho_- > \sigma$ , and free flow downstream  $\rho_+ \leq \sigma$ , the solution will be a rarefaction wave, with two wave fronts,

$$\rho(x, t) = \begin{cases} \rho_-, & x < -Wt \\ \sigma, & -Wt < x < Vt, \\ \rho_+, & x > Vt, \end{cases}$$

as shown in Figure 3.3c. Otherwise, the solution will be a shock, consisting of one wavefront,

$$\rho(x, t) = \begin{cases} \rho_-, & x < \Lambda(\rho_-, \rho_+)t, \\ \rho_+, & x > \Lambda(\rho_-, \rho_+)t, \end{cases}$$

corresponding to Figure 3.3a, Figure 3.3b or Figure 3.3d, depending on  $\rho_-$  and  $\rho_+$ . We will be using the style of the upper figures to describe the evolution of solutions in space and time.

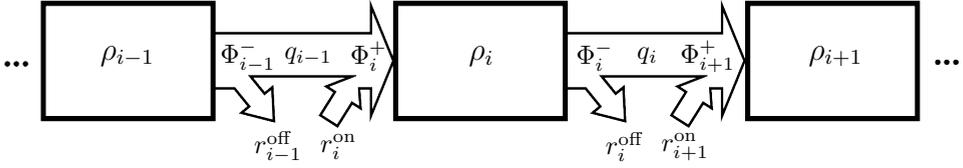


Figure 3.4: A representation of traffic flows in CTM.

The overall solution to (3.5),(3.9) for general piecewise constant initial conditions

$$\rho(x, 0) = \begin{cases} \rho_0(t), & x \leq X_1, \\ \rho_1(t), & X_1 \leq x < X_2, \\ \vdots & \vdots \\ \rho_i(t), & X_i \leq x < X_{i+1}, \\ \vdots & \vdots \\ \rho_N(t), & X_N \leq x < X_{N+1}, \\ \rho_{N+1}(t), & X_{N+1} \leq x, \end{cases}$$

can be acquired by solving a composite Riemann problem, i.e., solving a Cauchy problem with piecewise constant initial conditions through solving Riemann problems for all discontinuities in initial conditions  $X_i$ , evolving the solutions in time until some wave fronts originating from these interfaces collide, and then solving the new Riemann problems that thus appear. This procedure is known as front tracking, and since the flux function  $Q$  is piecewise-linear, it will yield exact solutions for all  $t > 0$ . In case the flux function is not piecewise-linear, front tracking can still be applied, but the flux function needs to be approximated with a piecewise-linear function. A more general approach will be discussed in Section 4.5.

### 3.3 Cell transmission model

Consider a stretch of highway with length  $L_i$ , between positions  $X_i$  and  $X_{i+1}$ ,  $L_i = X_{i+1} - X_i$ . We can describe the evolution of the number of vehicles inside this “cell”,  $n_i$  according to the conservation law

$$\dot{n}_i(t) = q_i^{\text{in}}(t) - q_i^{\text{out}}(t),$$

where  $q_{\text{in}}(t)$  is the flow into the segment and  $q_{\text{out}}(t)$  the flow from the segment, dependent on the surrounding traffic conditions. If these cells are connected so that  $q_i^{\text{out}}(t) = q_{i+1}^{\text{in}}(t)$ , and we follow the update of the numbers of vehicles in each cell at discrete time instants  $t^T$  with time step  $T$ ,  $t = t^T T$ , we arrive at the original formulation of the cell transmission model (CTM) [203, 208]. We will instead follow the traffic density  $\rho_i = n_i/L_i$ .

### 3.3.1 The basic CTM

The evolution of traffic density  $\rho_i$  in cell  $i$  is given by

$$\rho_i(t^T + 1) = \rho_i(t^T) + \frac{T}{L_i} (\Phi_i^+(t^T) - \Phi_i^-(t^T)), \quad i = 1, \dots, N. \quad (3.14)$$

Here  $N$  is the number of cells, and  $\Phi_i^+(t^T)$  and  $\Phi_i^-(t^T)$  are the average total flow during one time step into cell  $i$ , and out of cell  $i$ , respectively, given by

$$\begin{aligned} \Phi_i^+(t^T) &= q_{i-1}(t^T) + r_i^{\text{on}}(t^T), \\ \Phi_i^-(t^T) &= q_i(t^T) + r_i^{\text{off}}(t^T), \end{aligned} \quad (3.15)$$

where  $q_i$  is the flow from cell  $i$  to cell  $i + 1$ ,  $r_i^{\text{on}}(t^T)$  is the flow entering cell  $i$  from on-ramps, and  $r_i^{\text{off}}(t^T)$  the flow exiting cell  $i$  via off-ramps (see Figure 3.4). We require  $T$  to be short enough so that the Courant-Friedrichs-Lewy condition for numerical stability is satisfied,  $V \leq L/T$ .

For now, we will assume that all cells are of same length,  $L_i = L$  and consider a section that has no on- or off-ramps,  $r_i^{\text{on}}(t^T) = r_i^{\text{off}}(t^T) = 0, i = 1, \dots, N$ . The CTM then reduces to

$$\rho_i(t^T + 1) = \rho_i(t^T) + \frac{T}{L_i} (q_{i-1}(t^T) - q_i(t^T)), \quad i = 1, \dots, N. \quad (3.16)$$

We can determine  $q_i(t^T)$  as a minimum between a ‘‘demand’’ (sending) function of cell  $i$  and ‘‘supply’’ (receiving) function of cell  $i + 1$ ,

$$q_i(t^T) = \min(D_i(t^T), S_{i+1}(t^T)), \quad (3.17)$$

where

$$D_i(t^T) = \min(V_i \rho_i(t^T), q_i^{\text{max}}), \quad (3.18)$$

$$S_i(t^T) = \min(W_i(P_i - \rho_i(t^T)), q_i^{\text{max}}). \quad (3.19)$$

In order to make the model consistent with the LWR model with Newell-Daganzo flux function, we set the congestion wave speeds to  $W_i = V_i \frac{\sigma_i}{P_i - \sigma_i}$ , so that  $V_i \sigma_i = W_i(P_i - \sigma_i)$ , and set  $q_i^{\text{max}} = V_i \sigma_i$ .

The simplest way of defining the inflow from the on-ramps is to and prioritize the mainstream traffic, by setting the inflow from the on-ramps to

$$r_i^{\text{on}}(t^T) = \min\{\phi_{r,i}(t^T), S_i(t^T) - q_{i-1}(t^T)\},$$

assuming the traffic that potentially could not enter the road at time step  $t^T$  does not queue, and instead heads elsewhere. The desired on-ramp inflow  $\phi_{r,i}(t^T)$  is given as an external input. The off-ramp traffic flow can be defined using a splitting ratio of the traffic flow that leaves the road via the off-ramp  $R_i(t^T)$ ,

$$r_i^{\text{off}}(t^T) = \frac{R_i(t^T)}{1 - R_i(t^T)} q_i(t^T),$$

in which case the demand of cell  $i$  is reduced to

$$D_i(t^T) = \min \left\{ (1 - R_i(t^T)V_i\rho_i(t^T), q_i^{\max} \right\},$$

We can handle the boundaries of the model by separately defining either the flow into the first cell  $q_0(t^T)$  and out of the last cell  $q_N(t^T)$ , or boundary traffic densities  $\rho_0(t)$  and  $\rho_{N+1}(t^T)$ .

Although in its original formulation [203] it was conceived as a distinct model, the CTM has been shown [209] to be equivalent to a Godunov (finite volume) discretization of the LWR model, assuming all parameters are the same for all cells,  $\sigma_i = \sigma$ ,  $P_i = P$ ,  $V_i = V$ ,  $W_i = W$ . Godunov discretization of (3.5) corresponds to taking piecewise constant initial conditions for  $\rho(x, t)$ ,

$$\rho(x, t^T T) = \begin{cases} \rho_0(t^T), & x \leq X_1, \\ \rho_1(t^T), & X_1 \leq x < X_2, \\ \vdots & \vdots \\ \rho_i(t^T), & X_i \leq x < X_{i+1}, \\ \vdots & \vdots \\ \rho_N(t^T), & X_N \leq x < X_{N+1}, \\ \rho_{N+1}(t^T), & X_{N+1} \leq x, \end{cases}$$

where  $\rho_0(t^T)$  and  $\rho_{N+1}(t^T)$  are the boundary conditions,  $\rho(x, t^T T) = \rho_0(t^T)$ ,  $x < X_1$ , and  $\rho(x, t^T T) = \rho_{N+1}(t^T)$ ,  $x > X_N + L$ , solving the initial values problem for time up to  $t = (t^T + 1)T$ . We can then obtain average cell traffic density at the next time step  $\rho_i(t^T + 1)$  by averaging the solution of the PDE  $\rho(x, (t^T + 1)T)$  over the interval corresponding to each cell,  $(X_i, X_{i+1})$ . The reader is referred to [206] for more details on handling hyperbolic conservation laws.

### 3.3.2 CTM with a general flux function

While the triangular flux function (3.9) is the standard choice for CTM, it is also possible to use a generic flux function  $Q(\rho)$ , with little modification to the model. This model can be seen as a Godunov discretization of the LWR model. If the flux function  $Q(\rho)$  is concave, the critical density  $\sigma$  is defined as the density for which the flux function is maximized,

$$\sigma = \arg \max_{\rho} Q(\rho).$$

Then the demand and supply functions can also be written as a function of minimum and maximum, respectively, of  $\rho$  and  $\sigma$ ,

$$\begin{aligned} D_i(t^T) &= Q_i(\min(\rho_i(t^T), \sigma_i)), \\ S_i(t^T) &= Q_i(\max(\rho_i(t^T), \sigma_i)), \end{aligned} \tag{3.20}$$

where the flux function is possibly different for different cells.

For example, with the flux function (3.11), we may use an approximation for the demand and supply functions

$$D_i(t^T) = \begin{cases} V\rho_i(t^T) \exp\left(-\frac{1}{m} \left(\frac{\rho_i(t^T)}{\sigma_i}\right)^m\right), & \rho_i(t^T) \leq \sigma_i, \\ V\sigma_i \exp(-1), & \rho_i(t^T) > \sigma_i, \end{cases}$$

$$S_i(t^T) = \begin{cases} V\sigma_i \exp(-1), & \rho_i(t^T) \leq \sigma_i, \\ V\rho_i(t^T) \exp\left(-\frac{1}{m} \left(\frac{\rho_i(t^T)}{\sigma_i}\right)^m\right), & \rho_i(t^T) > \sigma_i. \end{cases}$$

Note that flux function (3.11), though unimodal, is not concave and has an inflection point in  $\sigma(m+1)^{\frac{1}{m}}$ , but the resulting Godunov discretization would be the same.

### 3.4 Tandem queueing model

Queueing models have long been used for modelling traffic, mostly at signalized intersections [210]. Many different types of queueing models have been proposed over the years [211], dealing with various scenarios. Tandem fluid queueing model is another type of models used to capture the behaviour of congested traffic networks. This model can be seen as an abstraction of the CTM that focuses on the queuing behaviour at bottlenecks [32]. Abstracting away the spatial component of congestion yields a much simpler model, which is still capable of dealing with the majority of considered problems.

In queueing models, the traffic situation is described by the numbers of vehicles  $n_i(t)$  queueing at some bottlenecks  $i$ . The evolution of these queue lengths is given by

$$\dot{n}_i(t) = q_i^{\text{in}}(t) - q_i^{\text{out}}(t), \quad (3.21)$$

where  $q_i^{\text{in}}(t)$  is the traffic flow arriving at the queue, and  $q_i^{\text{out}}(t)$  the flow discharging from the queue. The outflow discharging from a queue depends on the queue length,

$$q_i^{\text{out}}(t) = \begin{cases} \min\{q_i^{\text{cap}}, q_i^{\text{in}}(t)\}, & n_i(t) = 0, \\ q_i^{\text{dis}}, & n_i(t) > 0, \end{cases}$$

where  $q_i^{\text{cap}}$  is the maximum capacity of the bottleneck, and  $q_i^{\text{dis}} \leq q_i^{\text{cap}}$  is the discharging flow from the congested bottleneck, which can be equal to the capacity or lower, due to capacity drop. The inflow to a queue can either be exogenously determined, or it can correspond to the outflow of another queue that is upstream of the considered one,

$$q_i^{\text{in}}(t) = q_{\overleftarrow{i},i}^{\text{out}}(t) + q_i^{\text{r}}(t), \quad (3.22)$$

where  $q_i^{\text{r}}(t)$  represents the net flow from on-ramps and to off-ramps upstream of queue  $i$ . Here the set of all queues immediately upstream of queue  $i$  is written  $\overleftarrow{i}$ ,

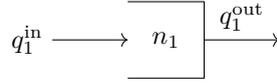


Figure 3.5: A single queueing server

and we denote summation over a set of queues by writing that set in the subscript,

$$q_{\overleftarrow{i},i}^{\text{out}}(t) = \sum_{k \in \overleftarrow{i}} q_{k,i}^{\text{out}}(t),$$

with  $q_{k,i}^{\text{out}}(t)$  representing the flow from queue  $k$  to queue  $i$ . This way, we can represent merging flows from multiple queues in  $\overleftarrow{i}$  to a single one,  $i$ . We also allow flows to diverge, with the outflow from one queue splitting into inflows to multiple downstream queues,

$$q_{i,j}^{\text{out}}(t) = R_{i,j} q_i^{\text{out}}(t), \quad j \in \overrightarrow{i},$$

where  $\overrightarrow{i}$  is the set of queues immediately downstream of queue  $i$ , and  $R_{i,j}$  are the splitting ratios that determine the portion of the outflow from queue  $i$  that is bound for queue  $j$ , with  $R_{i,\overrightarrow{i}} = 1$  and  $q_{i,\overrightarrow{i}}^{\text{out}}(t) = q_i^{\text{out}}(t)$ .

The model described in the previous subsection assumes that the outflow from one queue immediately appears as the inflow to the other queue, and that the queue lengths can grow to an arbitrary value without affecting the behaviour of the rest of the network, which is, of course, not in line with the physical reality. Instead, we may explicitly include the delays by replacing  $q_{\overleftarrow{i}}^{\text{out}}(t)$  in (3.22) with

$$q_{\overleftarrow{i}}^{\text{out}}(t) = \sum_{k \in \overleftarrow{i}} q_k^{\text{out}}(t - \theta_{k,i}),$$

where  $\theta_{i,j}$  are the individual delays from queue  $i$  to queue  $j$ , applied as time shift. The second important modification is limiting the maximum queue length to some maximum buffer capacity  $n_i(t) \leq N_i$ . Once the queue length reaches this limit  $n_i(t) = N_j$ , queue  $i$  will not be able to accept any additional inflow, which will limit the outflow of the queues immediately upstream,

$$q_i^{\text{out}}(t) = \min \left\{ \{q_i^{\text{f}}(t)\} \cup \left\{ \frac{q_j^{\text{out}}(t)}{R_{i,j}} \mid j \in \overrightarrow{i}, n_j = N_j \right\} \right\},$$

$$q_i^{\text{f}}(t) = \begin{cases} \min\{q_i^{\text{cap}}, q_i^{\text{in}}(t)\}, & n_i(t) = 0, \\ q_i^{\text{dis}}, & n_i(t) > 0. \end{cases}$$

We use  $q_i^{\text{out},N}(t)$  in place of  $q_i^{\text{out}}(t)$  in (3.21). Note that all of the diverging flows

$$q_{i,j}^{\text{out},N}(t) = R_{i,j} q_i^{\text{out},N}(t), \quad j \in \overrightarrow{i},$$

are now limited by the most severe restriction from queues downstream from queue  $i$ , whether or not the other queues can accept more traffic flow, modelling the reduction of flow due to the congestion spillback. This gives rise to, e.g., reduction of total outflow and efficiency when off-ramps are blocked by the congestion on the mainstream.

### 3.5 Second-order traffic models

While in this thesis we deal exclusively with first-order macroscopic traffic models, where the traffic speed is directly determined by the traffic density, for completeness we give a brief overview of some of the more commonly used second-order models in this section.

One of the most widely acclaimed macroscopic traffic model is METANET [200, 207]. Like the CTM, METANET is a model that is discretized in space and time, and it shares the conservation of traffic density, (3.14) and (3.15), with  $q_i(t^T) = \rho_i(t^T)v_i(t^T)$ , but instead of the traffic speed  $v_i(t^T)$  being directly determined by the traffic density  $v_i(t^T) = \mathcal{V}(\rho_i(t^T))$ , it evolves dynamically,

$$\begin{aligned} v_i(t^T + 1) = & v_i(t^T) + \frac{T}{\tau} (\mathcal{V}(\rho_i(t^T)) - v_i(t^T)) + \frac{T}{L_i} v_i(t^T) (v_{i-1}(t^T) - v_i(t^T)) \dots \\ & \dots - \frac{\nu T (\rho_{i+1}(t^T) - \rho_i(t^T))}{\tau L_i (\rho_i(t^T) + \chi)} - \delta_{\text{on}} T \frac{v_i(t^T) r_i^{\text{on}}(t^T)}{L_i (\rho_i(t^T) + \chi)} \end{aligned}$$

where  $\tau$ ,  $\nu$ ,  $\chi$ , and  $\delta_{\text{on}}$  are the parameters of the model. Here, the summands represent the old traffic speed in cell  $i$ , relaxation to the equilibrium speed  $\mathcal{V}(\rho_i(t^T))$ , convection of the traffic speed from the upstream cell, anticipation of the denser traffic in the downstream cell  $\rho_{i+1}(t^T)$ , and the influence of the slower traffic arriving from the on-ramps, respectively. The equilibrium speed is given by (3.10), where  $m$  and  $\sigma$  are additional model parameters.

Another currently widely used second-order model is the Aw-Rascle-Zhang (ARZ) model [124, 212], introduced to solve some inconsistencies [123] with the older second-order models. This model can be seen as an extension of the LWR model (3.6), with the traffic speed determined by

$$\partial_t (v(x, t) + p(\rho(x, t))) + v(x, t) \partial_x (v(x, t) + p(\rho(x, t))) = \frac{\mathcal{V}(\rho(x, t)) - \rho(x, t)}{\tau},$$

where  $p(\rho(x, t))$  is a smooth and increasing function modelling the traffic ‘‘pressure’’,

$$p(\rho(x, t)) = \mathcal{V}(0) - \mathcal{V}(\rho(x, t)).$$

A convenient form of the ARZ model is the so-called Generalized ARZ model [153], where the pressure term is incorporated into the new state  $y(x, t) = v(x, t) + p(\rho(x, t))$ , yielding

$$\partial_t y(x, t) + v(x, t) \partial_x y(x, t) = \frac{\mathcal{V}(0) - y}{\tau},$$

where  $v(x, t) = \mathcal{V}(\rho(x, t)) - \mathcal{V}(0) + y(x, t)$ .

Other types of properties, such as the proportion of CAVs in the traffic, can also be used in place of  $y(x, t)$ , resulting in a similar model, with the traffic speed  $v(x, t)$  given as a function of  $\rho(x, t)$  and  $y(x, t)$ . This can be seen as a way of introducing two classes of vehicles into the traffic model, which will be further discussed in Section 4.3.

### 3.6 Summary

There are many models we may use to capture the behaviour of highway traffic. Microscopic models are easy to understand, since they deal with individual vehicles and can easily be visualized, but are numerically expensive to simulate. Due to their relative simplicity, macroscopic traffic models are widely used for traffic control design. We described the two most well-known macroscopic traffic models, the CTM and the LWR model. These two models were shown to be equivalent, since the CTM can be seen as a discretization of LWR. A further simplification of the CTM, focusing on the queues at the bottlenecks, the tandem queueing model, is also presented. Finally, we outline some prominent second-order traffic models, which were not directly used in this thesis, but are presented for completeness.

## Chapter 4

# Traffic model extensions

WHILE widely used, the basic macroscopic traffic models described in the previous chapter fail to capture some of the traffic phenomena that are relevant to Lagrangian traffic control, and the problems that are studied in this thesis. Therefore, new models, or suitable extensions of the existing ones, are necessary. Namely, we need the models to be able to capture the influence of

- Platoons and moving bottlenecks,
- Capacity drop and traffic breakdown at stationary bottlenecks, and
- Stop-and-go waves,

while still being tractable and suitable for control design.

In this chapter, we propose model extensions and new models that can capture the influence of these traffic phenomena, which will be used for control design in the control part of this thesis. After introducing these phenomena in Section 4.1, first, in Section 4.2 we introduce moving bottlenecks to the LWR model and to the CTM. Then, in Section 4.3 we present a multi-class extension of the CTM, which handles all of the relevant phenomena. Next, in Section 4.4 the tandem queueing model is extended to be able to model platoons and moving bottlenecks as moving queueing servers. Finally, in Section 4.5 we propose the front tracking transition system model (FTTSM).

### 4.1 Traffic phenomena of interest

Firstly, we need to incorporate the influence of vehicles moving slower than the rest of the traffic, either alone or driving together with some other vehicles in formation such as a platoon. A slow-moving vehicle in traffic forces faster moving vehicles to overtake it, or be held up behind it, restricting the road capacity at the slow-moving vehicle's position. We call this slow-moving vehicle or platoon a moving bottleneck. Broadly speaking, there are two ways moving bottlenecks can be modelled, as shown

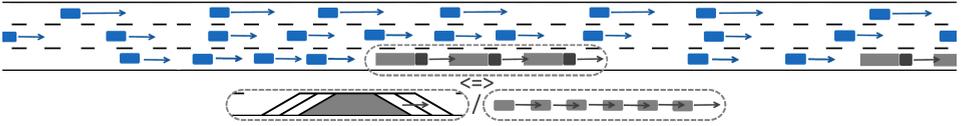


Figure 4.1: A platoon of trucks acting as a moving bottleneck and two ways of representing this phenomenon: either by reducing the capacity of the road, or by virtually adding vehicles to the road.

in Figure 4.1. The first approach is to consider it a moving traffic flow constraint, i.e. a reduction of road capacity at the moving bottleneck's position, and not explicitly count the slow vehicle acting as a moving bottleneck a part of the overall traffic density. The second approach is to represent the slow moving vehicles through their density, and model interactions between vehicles moving at different speeds in some cell-based model.

Secondly, except for the tandem queueing model, which can exhibit capacity drop if  $q_i^{\text{dis}} < q_i^{\text{cap}}$ , the discharging flow at a stationary bottleneck in the basic traffic models is the same as the bottleneck capacity, which is contrary to what is empirically observed. There exist many different ways of implementing this phenomenon in cell-based macroscopic traffic models [38], and capacity drop is commonly implemented by modifying the supply or demand function, so that the capacity is reduced as the traffic density increases. Modelling capacity drop in first-order PDE traffic model is much more challenging, and has to be done by either creating a temporary zone with a different flux function with reduced capacity, or by defining appropriate additional constraints on the solutions of the Riemann problem at the interface between two different flux functions. Furthermore, the mechanism of traffic breakdown needs to be defined, i.e. how the state of the stationary bottleneck changes from free flow to congested with capacity drop, and vice versa. The simpler option is to define this transition deterministically, even though empirical observations indicate that the transition happens stochastically, based on the behaviour of individual drivers. Stochastic traffic breakdown is therefore an important phenomenon that is essential in some applications.

Finally, the traffic phenomenon of stop-and-go waves, also known as wide moving jams, phantom traffic jams, or traffic waves, needs to be included. These stop-and-go waves are coupled with the capacity drop phenomenon, and therefore reduce the capacity of the road, and consequently also its achievable throughput. Even though in cell-based traffic models, the existence of stop-and-go waves follows from the capacity drop, additional steps are required in order to model them correctly. Namely, due to spatial discretization, the downstream boundary of the stop-and-go waves will not be precisely defined in the traffic model, and therefore needs to be tracked independently, as an additional state. Furthermore, even if the spatial discretization is done with very high resolution, requiring a large number of short cells, the diffusion that is inherent in the model will smooth the traffic density profile

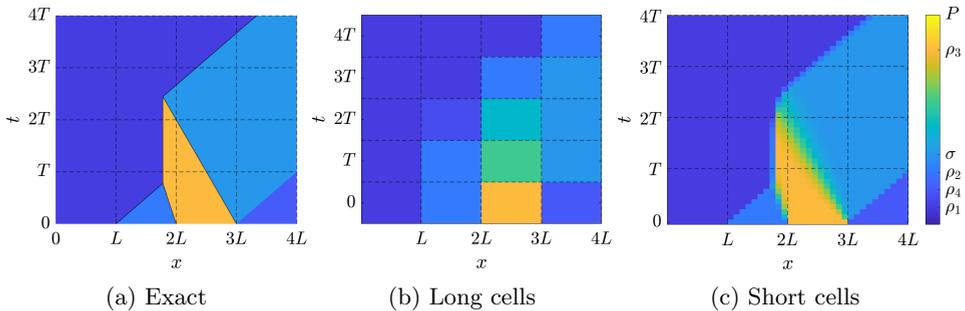


Figure 4.2: Example of the influence of spatial discretization and diffusion on modelling stop-and-go waves. The exact solution in (a) is given in the front tracking transition system model.

around the boundary between the stop-and-go wave and the discharging traffic flow, thus reducing the capacity drop effect. These two effects are demonstrated in Figure 4.2. Conversely, in first-order PDE traffic models, stop-and-go waves can be modelled by imposing additional constraints on the wave speed, defining a new type of weak solution. This approach is further justified by empirical observations that stop-and-go waves typically propagate upstream with constant velocity. Therefore, additional steps are needed to properly model the influence of stop-and-go waves in macroscopic traffic models.

## 4.2 Moving bottlenecks in LWR and CTM

In this section we first introduce the influence of moving bottlenecks to the LWR model, through imposing a moving constraint on the traffic flow in the zone of the moving bottleneck, and discussing the Riemann problems that arise from its treatment. Since the CTM corresponds to Godunov discretization of the LWR model [209], we introduce moving bottlenecks to CTM by locally solving a composite Riemann problem, consisting of the cells adjacent to the one with the moving bottleneck. Then, using a Godunov-like scheme, we average the resulting solution after one time step to acquire the traffic flow updates to be used in the model. Here we consider the simplest, triangular flux function, which allows for a clear presentation of the results. A more general flux function is discussed in Section 4.5 in the framework of the FTTSM.

### 4.2.1 Moving bottlenecks in the LWR model

Consider a vehicle (or a platoon of vehicles) in traffic, at position  $x_\xi(t)$ , moving at speed  $u_\xi(t)$  that is lower than the speed of the surrounding traffic  $\mathcal{V}(\rho(t, x_\xi(t)+))$ , and thus acting as a moving bottleneck. We can model this phenomenon by im-

posing additional constraints on the LWR model. This yields a PDE-ODE strongly coupled system [159, 161], with the traffic conditions evolution described by a scalar conservation law with a moving flux constraint, and the motion of the moving bottleneck described by an ordinary differential equation,

$$\begin{aligned} \partial_t \rho(x, t) + \partial_x Q(\rho(x, t)) &= 0, \\ Q(\rho(x_\xi(t), t)) - \dot{x}_\xi(t) \rho(x_\xi(t), t) &\leq q_\beta^{\max}(\dot{x}_\xi(t)), \\ \dot{x}_\xi(t) &= \min(u_\xi(t), \mathcal{V}(\rho(x_\xi(t)+, t))). \end{aligned}$$

Here  $x_\xi(t)$  is the position of the moving bottleneck,  $u_\xi(t)$  its desired speed and  $q_\beta^{\max}$  maximum flow past the bottleneck. The maximum flow is obtained by studying the problem in the reference frame of the moving bottleneck,

$$q_\beta^{\max}(\dot{x}_\xi(t)) = \max_{0 \leq \rho \leq P} Q_\xi(\rho) - \dot{x}_\xi(t) \rho,$$

as shown in Figure 4.3b.

This formulation is equivalent to using a different flow model in the zone of the moving bottleneck. Let the flux function in this zone,  $Q_\xi(\rho)$ , be of the same form as (3.9), with different parameters,

$$Q_\xi(\rho) = \begin{cases} V_\xi \rho, & \rho \leq \sigma_\xi, \\ W_\xi(P_\xi - \rho), & \rho > \sigma_\xi, \end{cases}$$

where  $W_\xi = W V_\xi / V$ ,  $V_\xi \sigma_\xi = W_\xi(P_\xi - \sigma_\xi)$ . An example of flux functions  $Q(\rho)$  and  $Q_\xi(\rho)$  is shown in Figure 4.3a.

In order to model the capacity reduction in presence of a bottleneck, we introduce a new parameter  $\beta \in [0, 1]$  that describes the severity of the bottleneck. The critical density  $\sigma$  and jam density  $P$  are reduced to

$$\sigma_\xi = \sigma(1 - \beta),$$

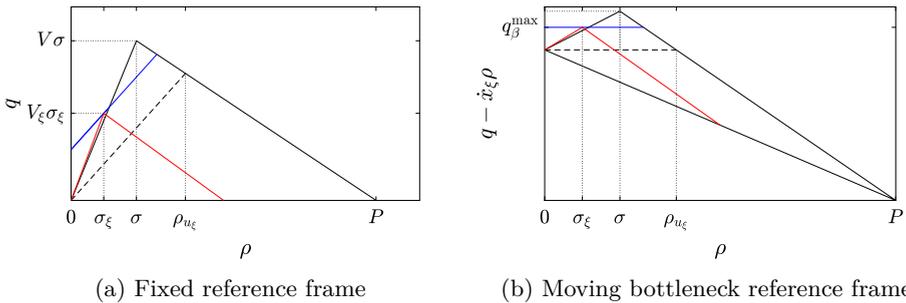


Figure 4.3: Flux functions  $Q(\rho)$  (solid black) and  $Q_\xi(\rho)$  (red) for  $V_\xi > V$  in fixed and moving bottleneck reference frame.

$$P_\xi = P(1 - \beta).$$

Since it depends on the behaviour of drivers, the bottleneck severity parameter  $\beta$  would in general have to be experimentally determined. However, a good approximation is to take  $\beta$  to be close to the portion of the road that the moving bottleneck takes. For example, if one of two lanes is blocked, we can set  $\beta = 0.5$ , or we might choose a somewhat higher value to capture additional “friction” effects due to induced lane changes.

Furthermore, we allow the free flow speed at the position of the bottleneck  $V_\xi$  to differ from the free flow speed elsewhere, possibly even as a function of  $u_\xi(t)$ . This enables us to model the overtaking behaviour in more detail, with  $V_\xi > V$  indicating eagerness, and  $V_\xi < V$  indicating reluctance to overtake. In order to simplify the solutions of the Riemann problems, we assume that  $u_\xi(t) \leq V_\xi \leq (V - u_\xi(t)\beta)/(1 - \beta)$ . As with  $\beta$ ,  $V_\xi$  depends on driver behaviour, and would have to be experimentally determined. We make the standing assumption that  $V_\xi$  is constant,  $V_\xi > V$  and  $u_\xi(t) \leq V/\beta - V_\xi(1 - \beta)/\beta$ . If we let  $V_\xi$  depend on  $u_\xi(t)$ , we can use this function  $V_\xi(u_\xi(t))$  to capture any bottleneck diagram [157], i.e. a curve whose tangents with slope  $u_\xi(t)$  we intersect with  $Q(\rho)$  to obtain traffic densities upstream and downstream of a moving bottleneck.

To model the influence the moving bottleneck has on the surrounding traffic, we solve two Riemann problems, one for its head (downstream end) and one for its tail (upstream end). We denote the traffic density upstream of the bottleneck as  $\rho_-$  and downstream  $\rho_+$ , and the traffic density in the bottleneck zone as  $\rho_\xi$ .

First, we define the projection functions

$$r_f(\rho_\xi, u_\xi) = \frac{Q_\xi(\rho_\xi) - u_\xi \rho_\xi}{V - u_\xi}, \quad (4.1)$$

$$r_c(\rho_\xi, u_\xi) = \frac{WP - Q_\xi(\rho_\xi) + u_\xi \rho_\xi}{W + u_\xi}, \quad (4.2)$$

that return the traffic density of the projection of the point  $(\rho_\xi, Q_\xi(\rho_\xi))$  with slope  $u_\xi$  onto the free flow ( $\rho < \sigma$ ) and congested ( $\rho > \sigma$ ) part of  $Q(\rho)$ , respectively. Conversely, the inverse of (4.1) and (4.2) returns projections of the point  $(\rho, Q(\rho))$  with slope  $u_\xi$  onto the free flow and congested part of  $Q_\xi(\rho)$ , respectively,

$$r_f^{-1}(\rho, u_\xi) = \frac{Q(\rho) - u_\xi \rho}{V_\xi - u_\xi}, \quad (4.3)$$

$$r_c^{-1}(\rho, u_\xi) = \frac{W_\xi P_\xi - Q(\rho) + u_\xi \rho}{W_\xi + u_\xi}. \quad (4.4)$$

Effectively, these projections are the intersections between the flux function  $Q(\rho)$ , and a line with slope  $u_\xi$  originating from  $(\rho_\xi, Q_\xi(\rho_\xi))$ , in case of (4.1) and (4.2), and between the flux function  $Q_\xi(\rho)$ , and a line with the same slope originating from  $(\rho, Q(\rho))$ , in case of (4.3) and (4.4). It is easy to check that  $r_{f,c}^{-1}(r_{f,c}(\rho, u_\xi), u_\xi) = \rho$

and  $r_{f,c}(r_{f,c}^{-1}(\rho, u_\xi), u_\xi) = \rho$ . Note that for  $V_\xi = V$  and  $\rho_\xi \leq \sigma_\xi$ ,  $r_f(\rho_\xi, u_\xi) = \rho_\xi$ , regardless of  $u_\xi$ .

The Riemann problems for the moving bottleneck boundaries can be written as

$$\begin{aligned} \partial_t \rho + \partial_x (Q_\pm(\rho, x, t)) &= 0, \\ Q_\pm(\rho, x, t) &= \begin{cases} Q_-(\rho), & x < u_\xi t, \\ Q_+(\rho), & x > u_\xi t, \end{cases} \\ \rho(x, 0) &= \begin{cases} \rho_-, & x < 0, \\ \rho_+, & x > 0. \end{cases} \end{aligned}$$

Example solutions to Riemann problems for moving bottleneck head and tail for all traffic density cases are given in Figure 4.4.

Consider first the Riemann problem for the moving bottleneck head. In this case, we have  $Q_-(\rho) = Q(\rho)$ ,  $Q_+(\rho) = Q_\xi(\rho)$ ,  $\rho_- = \rho_\xi$ , and the Riemann problem corresponding to it is

$$\begin{aligned} \partial_t \rho + \partial_x (Q_+(\rho, u_\xi, x, t)) &= 0, \\ Q_+(\rho, u_\xi, x, t) &= \begin{cases} Q_\xi(\rho), & x < u_\xi t, \\ Q(\rho), & x > u_\xi t, \end{cases} \\ \rho(x, 0) &= \begin{cases} \rho_\xi, & x < 0, \\ \rho_+, & x > 0. \end{cases} \end{aligned}$$

We control the movement of the bottleneck, so the transition speed between the zones with different flux functions has to be equal to its speed  $u_\xi$ . The Rankine-Hugoniot condition for the discontinuity,  $u_\xi(\rho_+ - \rho_\xi) = Q(\rho_+) - Q_\xi(\rho_\xi)$ , can only hold for  $\rho_+ = r_f(\rho_\xi, u_\xi)$  or  $\rho_+ = r_c(\rho_\xi, u_\xi)$  (equivalently,  $\rho_\xi = r_f^{-1}(\rho_+, u_\xi)$  or  $\rho_\xi = r_c^{-1}(\rho_+, u_\xi)$ ). In this case, the entropy solution is simply

$$\rho(x, t) = \begin{cases} \rho_\xi, & x < u_\xi t, \\ \rho_+, & x > u_\xi t. \end{cases}$$

Otherwise, the entropy solution will, depending on  $\rho_\xi$  and  $\rho_+$ , have one or two additional wavefronts, moving at Rankine-Hugoniot transition speed

$$\Lambda(\rho_-, \rho_+) = \frac{Q(\rho_+) - Q(\rho_-)}{\rho_+ - \rho_-}, \quad \Lambda_\xi(\rho_-, \rho_+) = \frac{Q_\xi(\rho_+) - Q_\xi(\rho_-)}{\rho_+ - \rho_-}.$$

These solutions are:

- If  $\rho_\xi \leq \sigma_\xi$  and  $\rho_+ < r_c(\rho_\xi, u_\xi)$ ,

$$\rho(x, t) = \begin{cases} \rho_\xi, & x < u_\xi t, \\ r_f(\rho_\xi, u_\xi), & u_\xi < x < \Lambda(r_f(\rho_\xi, u_\xi), \rho_+)t, \\ \rho_+, & x > \Lambda(r_f(\rho_\xi, u_\xi), \rho_+)t. \end{cases}$$

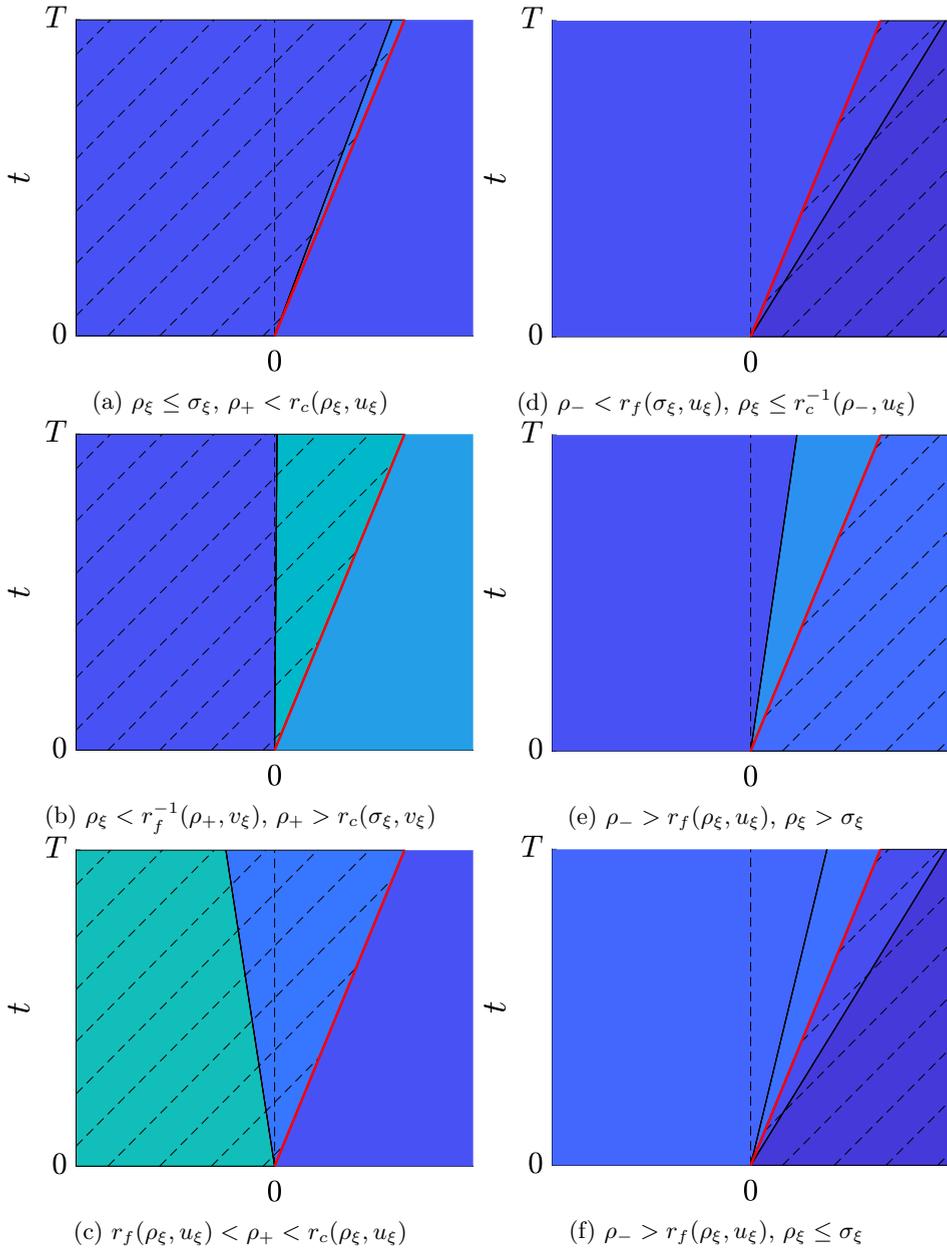


Figure 4.4: Solutions for all cases of Riemann problems for moving bottleneck head (a-c) and tail (d-f). Denser traffic is shown in warmer colours, and the bottleneck zone is shown hatched.

- If  $\rho_\xi < r_f^{-1}(\rho_+, v_\xi)$  and  $\rho_+ > r_c(\sigma_\xi, v_\xi)$ ,

$$\rho(x, t) = \begin{cases} \rho_\xi, & x < \Lambda_\xi(\rho_\xi, r_c^{-1}(\rho_+, v_\xi))t, \\ r_c^{-1}(\rho_+, v_\xi), & \Lambda_\xi(\rho_\xi, r_c^{-1}(\rho_+, v_\xi))t < x < v_\xi t, \\ \rho_+, & x > v_\xi t, \end{cases}$$

where  $v_\xi = \min(u_\xi, \mathcal{V}(\rho_+))$ . If  $\mathcal{V}(\rho_+) < u_\xi$ , the speed of both the platoon head and tail are set to  $\mathcal{V}(\rho_+)$ .

- If  $r_f(\rho_\xi, u_\xi) < \rho_+ < r_c(\rho_\xi, u_\xi)$ ,

$$\rho(x, t) = \begin{cases} \rho_\xi, & x < W_\xi t, \\ \sigma_\xi, & W_\xi t < x < u_\xi t, \\ r_f(\rho_\xi, u_\xi), & u_\xi t < x < Vt, \\ \rho_+, & x > Vt. \end{cases}$$

We have a similar situation for the Riemann problem for the moving bottleneck tail (upstream end). Now,  $Q_-(\rho) = Q_\xi(\rho)$ ,  $Q_+(\rho) = Q(\rho)$  and  $\rho_+ = \rho_\xi$ , and the Riemann problem is

$$\begin{aligned} \partial_t \rho + \partial_x (Q_-(\rho, x, t)) &= 0, \\ Q_-(\rho, u_\xi, x, t) &= \begin{cases} Q(\rho), & x < u_\xi t, \\ Q_\xi(\rho), & x > u_\xi t, \end{cases} \\ \rho(x, 0) &= \begin{cases} \rho_-, & x < 0, \\ \rho_\xi, & x > 0. \end{cases} \end{aligned}$$

Again, we have three cases of the entropy solution, depending on  $\rho_\xi$  and  $\rho_+$ :

- If  $\rho_- < r_f(\sigma_\xi, u_\xi)$  and  $\rho_\xi \leq r_c^{-1}(\rho_-, u_\xi)$ ,

$$\rho(x, t) = \begin{cases} \rho_-, & x < u_\xi t, \\ r_f^{-1}(\rho_-, u_\xi), & u_\xi < x < \Lambda(\rho_-, \rho_\xi)t, \\ \rho_\xi, & x > \Lambda(\rho_-, \rho_\xi)t. \end{cases}$$

- If  $\rho_- > r_f(\rho_\xi, u_\xi)$  and  $\rho_\xi > \sigma_\xi$ ,

$$\rho(x, t) = \begin{cases} \rho_-, & x < \Lambda(\rho_-, r_c(\rho_\xi, u_\xi))t, \\ r_c(\rho_\xi, u_\xi), & \Lambda(\rho_-, r_c(\rho_\xi, u_\xi))t < x < u_\xi t, \\ \rho_\xi, & x > u_\xi t. \end{cases}$$

- If  $\rho_- > r_f(\rho_\xi, u_\xi)$  and  $\rho_\xi \leq \sigma_\xi$ ,

$$\rho(x, t) = \begin{cases} \rho_\xi, & x < Wt, \\ r_c(\sigma_\xi, u_\xi), & Wt < x < u_\xi t, \\ \sigma_\xi, & u_\xi t < x < V_\xi t, \\ \rho_\xi, & x > V_\xi t. \end{cases}$$

### 4.2.2 CTM with moving bottlenecks

Having described the effect of the moving bottleneck in the LWR model framework, we can now apply a similar Godunov-like scheme to calculate the effects of the moving bottleneck on traffic flows of adjacent cells. If  $X_i \leq x_\xi(t^T) < X_i + L$ , where  $x_\xi(t^T)$  is the position of the moving bottleneck at discrete time instant  $t^T$ , and  $X_i$  is the position of the upstream boundary of cell  $i$ , the moving bottleneck is in cell  $i$  and  $i_\xi(t^T) = i$ . For compactness, we will omit writing the time step for all CTM-related variables wherever the time step is obvious. We may write the resulting flows as

$$q_i = \min(V\rho_{i_\xi}, V\sigma, W(P - \rho_{i_\xi+1})) + \Delta q_{\xi,i}.$$

Assume the moving bottleneck in cell  $i_\xi$  is the only one in the road stretch considered. Since it only affects traffic flows through the interfaces of the cell it is currently in (i.e. from cell  $i_\xi - 1$  to  $i_\xi$  and from cell  $i_\xi$  to  $i_\xi + 1$ ), we have  $\Delta q_{\xi,i} = 0$  for all  $i \notin \{i_\xi - 1, i_\xi\}$ . Therefore the resulting model will be the same as the already described standard CTM (3.16)–(3.17) for  $i \neq i_\xi, i \neq i_\xi - 1$ .

In order to correctly describe the behaviour of the moving bottleneck and the effect it has on the surrounding traffic, we need to augment the cell transmission model with three additional states: the position of the moving bottleneck  $x_\xi(t^T)$ , the traffic density directly upstream of it  $\rho_{\xi-}(t^T)$  and the traffic density in the moving bottleneck zone  $\rho_\xi(t^T)$ . The second and third additional states are necessary in order to properly model the flow of traffic overtaking the bottleneck [161], effectively splitting the cell  $i$  into three parts. We will keep  $\rho_i(t^T)$  as a state and instead, calculate the traffic density downstream of the bottleneck so that

$$\rho_i(t^T) = \frac{(x_\xi(t^T) - l_\xi - X_{i_\xi})\rho_{\xi-}(t^T) + l_\xi\rho_\xi(t^T) + (X_{i_\xi+1} - x_\xi(t^T))\rho_{\xi+}(t^T)}{L},$$

where  $l_\xi$  is the length of the bottleneck in question. If there are multiple bottlenecks fully or partially in the same cell, splitting the cell and calculating traffic densities in its different segments is done in a similar way, starting from the cell's upstream end and calculating traffic densities towards its downstream end so that they still average to  $\rho_i(t^T)$ .

We obtain  $\Delta q_{\xi,i_\xi-1}(t^T)$  and  $\Delta q_{\xi,i_\xi}(t^T)$ , as well as updates  $x_\xi(t^T+1)$ ,  $\rho_{\xi-}(t^T+1)$  and  $\rho_\xi(t^T+1)$  by solving the composite Riemann problem

$$\begin{aligned} \partial_t \rho + \partial_x (Q(\rho, x, t)) &= 0, \\ Q(\rho, x, t) &= \begin{cases} Q(\rho), & x \notin (\chi_\xi(t) - l_\xi, \chi_\xi(t)), \\ Q_\xi(\rho), & x \in (\chi_\xi(t) - l_\xi, \chi_\xi(t)), \end{cases} \\ \dot{\chi}_\xi(t) &= \min(u_\xi, \mathcal{V}(\rho(t, \chi_\xi(t)_+))), \end{aligned} \tag{4.5}$$

with initial conditions

$$\rho(x, 0) = \begin{cases} \rho_{i_\xi-1}, & x < X_{i_\xi}, \\ \rho_{\xi-}, & X_{i_\xi} < x < x_\xi(t^T) - l_\xi, \\ \rho_\xi, & x_\xi(t^T) - l_\xi < x < x_\xi(t^T), \\ \rho_{\xi+}, & x_\xi(t^T) < x < X_{i_\xi+1}, \\ \rho_{i_\xi+1}, & x > X_{i_\xi+1}, \end{cases} \quad (4.6)$$

$$\chi_\xi(0) = x_\xi(t^T),$$

for  $t$  up to  $T$ . The solution is easily obtained through front tracking, successively solving Riemann problems as described in the previous section. An example of a solution is shown in Figure 4.5. The position of the moving bottleneck is updated as  $x_\xi(t^T + 1) = \chi_\xi(T)$ .

If there are multiple moving bottlenecks, we either solve (4.5) with initial conditions (4.6) for each one if there is at least one cell between them, or include both of them in a larger composite Riemann problem. For example, for the situation shown in Figure 4.6, the initial conditions would be

$$\rho(x, 0) = \begin{cases} \rho_{i_\xi-1}, & x < X_{i_\xi}, \\ \rho_{\xi-}, & X_{i_\xi} < x < x_{\xi_1} - l_{\xi_1}, \\ \rho_{\xi_1}, & x_{\xi_1} - l_{\xi_1} < x < x_{\xi_1}, \\ \rho_{\xi_1+}, & x_{\xi_1} < x < X_{i_{\xi_1}+1}, \\ \rho_{\xi-2}, & X_{i_{\xi_1}+1} < x < x_{\xi_2} - l_{\xi_2}, \\ \rho_{\xi_2}, & x_{\xi_2} - l_{\xi_2} < x < x_{\xi_2}, \\ \rho_{\xi_2+}, & x_{\xi_2} < x < X_{i_{\xi_1}+2}, \\ \rho_{i_\xi+2}, & x > X_{i_{\xi_1}+2}. \end{cases}$$

In this case, one moving bottleneck is in cell  $i_{\xi_1}$  and a second one in cell  $i_{\xi_1} + 1 = i_{\xi_2}$ , so we need to include cells  $i_{\xi_1} - 1$  through  $i_{\xi_1} + 2$  into the problem. Here we assumed that the moving bottlenecks are truck platoons, and that the follower catches up with the leader and merges into one platoon. If the tail of one and head of another bottleneck collide, we say that those two bottlenecks have merged, and take their speed to be the speed of the leader bottleneck.

Formally, we may write the updated traffic model that incorporates moving bottlenecks as

$$\begin{aligned} \rho_i(t^T + 1) &= \rho_i(t) + \frac{T}{L} (q_{i-1}(t^T) - q_i(t^T)), \\ q_i(t^T) &= \min(V\rho_i(t^T), V\sigma, W(P - \rho_{i+1}(t^T))) + \Delta q_{\xi,i}(t^T), \\ \begin{bmatrix} \Delta q_\xi(t^T) \\ x_\xi(t^T + 1) \\ \rho_{\xi-}(t^T + 1) \\ \rho_\xi(t^T + 1) \end{bmatrix} &= \mathcal{P}(\rho(t^T), x_\xi(t^T), u_\xi(t^T), \rho_{\xi-}(t^T), \rho_\xi(t^T)), \end{aligned} \quad (4.7)$$

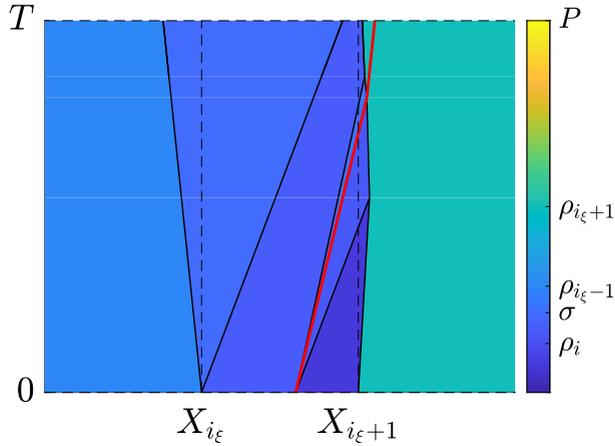


Figure 4.5: Front tracking solution example for  $t \in [0, T]$  and  $l_\xi \approx 0$ . Note that the moving bottleneck slows down when it enters the dense traffic from cell  $i_{\xi+1}$ .

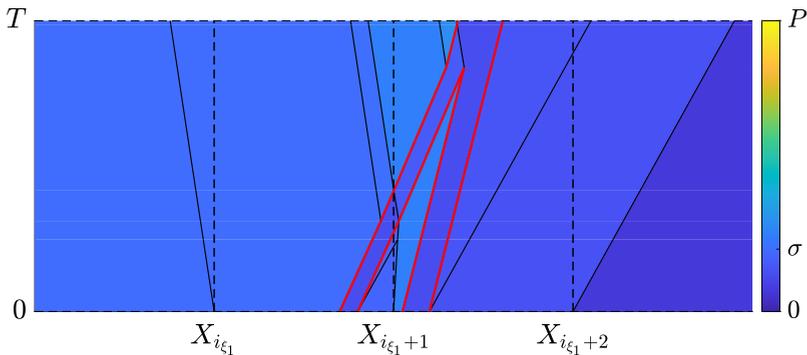


Figure 4.6: Front solution example of two platoons merging. Note that the follower platoon slows down when it enters the denser traffic from cell  $i_{\xi_1} + 1$ , originating from the leader platoon. The two moving bottlenecks merge into one before  $t = T$ .

where by  $\mathcal{P}$  we encapsulate the procedure of calculating  $\Delta q_\xi$  and new values for  $x_\xi$ ,  $\rho_{\xi-}$  and  $\rho_\xi$  from the solution of (4.5) with initial conditions (4.6) at  $t = T$ .

For ease of presentation, consider the case where there is only one moving bottleneck. Then we may calculate the updates  $\Delta q_{\xi, i_\xi}(t^T)$  as

$$\Delta q_{\xi, i_\xi}(t^T) = \frac{1}{T} \int_{X_{i_\xi+1}}^{X_{i_\xi+2}} \rho(x, T) - \bar{\rho}(x, T) dx, \quad (4.8)$$

and  $\Delta q_{\xi, i_{\xi}-1}(t^T)$  as

$$\Delta q_{\xi, i_{\xi}-1}(t^T) = \frac{1}{T} \int_{X_{i_{\xi}-1}}^{X_{i_{\xi}}} \rho(x, T) - \tilde{\rho}(x, T) dx. \quad (4.9)$$

Here by  $\tilde{\rho}(x, T)$  we denote the solution of the composite Riemann problem with no moving bottlenecks and for initial conditions

$$\tilde{\rho}(x, 0) = \begin{cases} \rho_{i_{\xi}-1}, & x < X_{i_{\xi}}, \\ \rho_{i_{\xi}}, & X_{i_{\xi}} < x < X_{i_{\xi}+1}, \\ \rho_{i_{\xi}+1}, & x > X_{i_{\xi}+1}. \end{cases}$$

Since this solution can be expressed explicitly, integrals of  $\tilde{\rho}(x, T)$  can easily be calculated as

$$\begin{aligned} \frac{1}{T} \int_{X_{i_{\xi}+1}}^{X_{i_{\xi}+2}} \tilde{\rho}(x, T) dx &= \max(\Lambda(\min(\rho_{i_{\xi}}, \sigma), \rho_{i_{\xi}+1}), 0) (\rho_{i_{\xi}+1} - \min(\rho_{i_{\xi}}, \sigma)), \\ \frac{1}{T} \int_{X_{i_{\xi}-1}}^{X_{i_{\xi}}} \tilde{\rho}(x, T) dx &= \min(\Lambda(\rho_{i_{\xi}-1}, \max(\rho_{i_{\xi}}, \sigma)), 0) (\rho_{i_{\xi}-1} - \max(\rho_{i_{\xi}}, \sigma)). \end{aligned}$$

Finally, the new position of the bottleneck is

$$x_{\xi}(t^T + 1) = \chi_{\xi}(T), \quad (4.10)$$

and the new traffic density upstream of it

$$\rho_{\xi-}(t^T + 1) = \frac{\int_{X_{i_{\xi}(t^T+1)}}^{\chi_{\xi}(T)-l_{\xi}} \rho(x, T) dx}{\chi_{\xi}(T) - l_{\xi} - X_{i_{\xi}(t^T+1)}}, \quad (4.11)$$

and inside the bottleneck zone

$$\rho_{\xi}(t^T + 1) = \frac{\int_{X_{i_{\xi}(t^T+1)}}^{\chi_{\xi}(T)} \rho(x, T) dx}{l_{\xi}}. \quad (4.12)$$

In case we have merging moving bottlenecks, we also need to keep track of their number and their lengths.

To summarize, the model we propose is an extended version of CTM (3.16)–(3.17), which can be written as (4.7). Traffic flow updates for cells adjacent to

the moving bottleneck are calculated according to (4.8) and (4.9). To properly model the dynamics of the moving bottleneck, we require adding two additional states (4.10), (4.11) and (4.12), whose updates are obtained from the solution of the composite Riemann problem. The proposed model is simple and tractable, as well as consistent with the PDE moving bottleneck traffic models. This approach also allows extensions to other PDE traffic models and different traffic phenomena, such as police cars, as well as enables traffic control design using the speed of the moving bottleneck as control variable and cell traffic densities as measurements.

### 4.3 Multi-class CTM

Multi-class traffic models are often inspired by the introduction of autonomous and connected vehicles, but can also provide a useful tool for capturing uncertainties in the flow model. Introducing two (or more) classes of traffic, for example the 'rabbits' (fast and aggressive drivers) and the 'slugs' (slow and timid drivers), or human-driven and autonomous vehicles, can also be seen as a simple way of defining a second-order (or higher-order) traffic model, where the conserved quantities would be the total traffic density and the ratio of vehicles of one class in it. The multi-class cell transmission model (MCCTM) presented here is a variant of the model used in [33] and [26], similar to the model introduced in [150], with proportional priority allocated to all vehicle classes.

We first describe the base multi-class cell transmission model, with homogeneous road geometry and no specific traffic phenomena, and then proceed to extend it. We discuss extending the model to cover non-trivial traffic network structures, with on- and off-ramps, merges, and diverges. Then, we introduce platoons and moving bottlenecks, followed by capacity drop and stop-and go waves. Finally, we generalize the multi-class CTM to use a general flux function, and summarize the model.

#### 4.3.1 The base model

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  $t$  will be expressed in terms of passenger car equivalents, and denoted  $\rho_i^\kappa(t^T)$ . We will denote the aggregate variables over multiple classes in some set by denoting that set in superscript, e.g. the aggregate traffic density of all classes  $\mathcal{K}$  in cell  $i$  is denoted

$$\rho_i^{\mathcal{K}}(t^T) = \sum_{\kappa \in \mathcal{K}} \rho_i^\kappa(t^T).$$

We allow each of the classes to have a distinct reference free flow speed  $U_i^\kappa(t^T) \leq V_i$  in every cell and varying in time, where  $V_i$  is the overall maximum vehicle speed for the cell. In practice, we use  $U_i^\kappa(t^T)$  to capture some richer behaviour not covered by the base model, like platoons stop-and-go waves, as well as to apply the control action to the classes of vehicles we have control over.

Consider a highway stretch consisting of  $N$  cells. Same as in the single-class case (3.14), the evolution of cell traffic densities for each class is given by

$$\rho_i^\kappa(t^T + 1) = \rho_i^\kappa(t^T) + \frac{T}{L_i} (\Phi_i^{+\kappa}(t^T) - \Phi_i^{-\kappa}(t^T)), \quad (4.13)$$

where  $T$  is the time step,  $L_i$  the length of cell  $i$ , and  $\Phi_i^{+\kappa}(t^T)$  and  $\Phi_i^{-\kappa}(t^T)$  the flows into and out of the cell  $i$  at time  $t$ , respectively. Assuming that the cells are of the same length,  $L_i = L$ , and that there are no on- and off-ramps, (4.13) simplifies to

$$\rho_i^\kappa(t^T + 1) = \rho_i^\kappa(t^T) + \frac{T}{L} (q_{i-1}^\kappa(t^T) - q_i^\kappa(t^T)),$$

where  $q_i^\kappa(t^T)$  is the mainstream traffic flow from cell  $i$  to cell  $i + 1$  immediately downstream, given by

$$q_i^\kappa(t^T) = \min\{D_i^\kappa(t^T), S_i^\kappa(t^T)\}.$$

We discuss the case when there are on- and off-ramps in Section 4.3.2.

In contrast with the classical CTM (3.17), in multi-class CTM, the demand and supply functions of each class,  $D_i^\kappa(t^T)$  and  $S_i^\kappa(t^T)$ , respectively, will now also depend on vehicles of other classes, since they are sharing the same road. Therefore, we write these functions in terms of the portion of aggregate demand and supply, written  $D_i^\mathcal{K}(t^T)$  and  $S_i^\mathcal{K}(t^T)$ , that is allotted to each vehicle class,

$$\begin{aligned} D_i^\kappa(t^T) &= \frac{d_i^\kappa(t^T)}{d_i^\mathcal{K}(t^T)} D_i^\mathcal{K}(t^T), \\ S_i^\kappa(t^T) &= \frac{\rho_i^\kappa(t^T)}{\rho_i^\mathcal{K}(t^T)} S_i^\mathcal{K}(t^T). \end{aligned} \quad (4.14)$$

Note that the allocation of aggregate demand and supply is distributed according to different properties. The supply available to cell  $i$  is allocated to vehicles in it, proportionally with their traffic density  $\rho_i^\kappa(t^T)$ . Conversely, the demand of cell  $i$  is allocated to vehicles in it, proportionally with  $d_i^\kappa(t^T)$ , defined as

$$d_i^\kappa(t^T) = \min\{U_i^\kappa(t^T), V_i\} \rho_i^\kappa(t^T), \quad (4.15)$$

where  $U_i^\kappa(t^T)$  is the reference free flow speed of class  $\kappa$  vehicles, representing maximum demand of each vehicle class if other classes were absent.

Since the desired gap between two vehicles in most car-following models grows with increase of their speeds, as discussed in Section 3.1, this choice can be seen as allocating aggregate demand to vehicle classes proportionally with how much space they take on the road, assuming  $U_i^\kappa(t^T) \geq U^{\min}$ , where  $U^{\min}$  is such that the desired gap is significantly larger than vehicle length. For example, if  $U_i^a(t^T) = 2U_i^b(t^T)$  and  $2\rho_i^a(t^T) = \rho_i^b(t^T)$ , we have  $d_i^a(t^T) = d_i^b(t^T)$ , and half of the lanes will be allocated to vehicles of class  $a$  and  $b$  each. Here we assume the lanes can be approximately

continuously distributed between vehicles driving at different speeds, which will only be satisfied if the number of lanes is larger than the number of different free flow speeds  $U_i^\kappa(t^T)$ , since otherwise the slower vehicles will block the faster ones, as discussed in Subsection 3.1.2.

We write the aggregate demand and supply functions

$$D_i^\mathcal{K}(t^T) = \min \{d_i^\mathcal{K}(t^T), q_i^{\max}(t^T)\}, \quad (4.16)$$

$$S_i^\mathcal{K}(t^T) = \min \{W_{i+1}(P_{i+1} - \rho_{i+1}^\mathcal{K}(t^T)), q_{i+1}^{\max}(t^T)\}, \quad (4.17)$$

where  $q_i^{\max}(t^T)$  is the capacity of cell  $i$ , given by

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

and cell parameters  $V_i$ ,  $W_i$ ,  $\sigma_i$  and  $P_i$  are the free flow speed, congestion wave speed, critical density and jam density of cell  $i$ , respectively. Note that  $S_i^\mathcal{K}(t^T)$  does not depend on  $\rho_i^\mathcal{K}(t^T)$ , since it does not represent the supply of cell  $i$ , but instead depends on  $\rho_{i+1}^\mathcal{K}(t^T)$ , since it represents the supply *available to traffic* in cell  $i$ .

In further text, when no cell number is indicated, we use the same cell length  $L$  and maximum free flow speed  $V = V_i$ , and take  $W = V \frac{\sigma}{P - \sigma}$  yielding a triangular fundamental diagram,

$$Q(\rho) = \min \{V\rho, W(P - \rho)\}. \quad (4.18)$$

Furthermore, we adopt  $L = VT$ , at the limit of the Courant-Friedrichs-Lewy condition. It is easy to verify that in case we only have one class  $\mathcal{K} = \{a\}$  and  $U_i^a(t^T) = V$ , expressions (4.16) and (4.17) simplify to (3.18) and (3.19). The cell capacity depends on the free flow speeds of each class  $U_i^\kappa(t^T)$ , as well as on the share of vehicles of each class in the cell, and is lower or equal to the maximum value  $q_i^{\max}(t^T) \leq V_i \sigma_i$ . Note that here we assume that the capacity depends only on the share of each vehicle class, but not their aggregate density, implicitly assuming that the capacity drop phenomenon is absent. We discuss including capacity drop into the model in Subsection 4.3.4.

### 4.3.2 On- and off-ramps, merges, and diverges

Although a large number of phenomena of interest happen in mainstream traffic flow, where on- and off-ramps are either not present or not relevant, the behaviour of traffic at merges, diverges, and on- and off-ramps needs to be described if we want to study any non-trivial network configuration. The simplest extension is to introduce the flows into cells from on-ramps and out of cells cells to off-ramps by setting  $\Phi_i^{+\kappa}(t^T) = q_{i-1}^\kappa(t^T) + r_i^{\text{on},\kappa}(t^T)$  and  $\Phi_i^{-\kappa}(t^T) = q_i^\kappa(t^T) + r_i^{\text{off},\kappa}(t^T)$ , yielding

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

where  $r_i^{\text{on},\kappa}(t^T)$  is the inflow and  $r_i^{\text{off},\kappa}(t^T)$  the outflow of each vehicle class from a potential on-ramp and to a potential off-ramp, respectively. An example of such traffic flows is given in Figure 4.7.

Unless we assume that all inflow from on-ramps enters the road regardless of its traffic condition, a part of vehicles entering the road might have to queue if their entry gets blocked by congestion in the cell where they merge. We model the evolution of these queues  $n_{r,i}^\kappa(t^T)$ , for on-ramps in cell  $i$ , with

$$n_{r,i}^\kappa(t^T + 1) = n_{r,i}^\kappa(t^T) + (\phi_{r,i}^\kappa(t^T) - r_i^\kappa(t^T)) T,$$

where  $\phi_i^\kappa(t^T)$  is the inflow of class  $\kappa$  vehicles arriving at the on-ramp. There are multiple ways of modelling the actual on-ramp flow  $r_i^{\text{on},\kappa}(t^T)$ , ranging from fully prioritizing the mainstream flow,

$$\begin{aligned} r_i^{\text{on},\kappa}(t^T) &= \min \{ D_{r,i}^\kappa(t^T), S_{r,i}^\kappa(t^T) \}, \\ D_{r,i}^\kappa(t^T) &= \phi_{r,i}^\kappa(t^T) + \frac{n_{r,i}^\kappa(t^T - 1)}{T}, \\ S_{r,i}^\kappa(t^T) &= \frac{\phi_{r,i}^\kappa(t^T)T + n_{r,i}^\kappa(t^T - 1)}{\phi_{r,i}^\kappa(t^T)T + n_{r,i}^\kappa(t^T - 1)} \min \{ S_i^\kappa(t^T) - q_{i-1}^\kappa(t^T), q_{r,i}^{\text{on}} \}, \end{aligned}$$

to fully prioritizing the on-ramp flow,

$$r_i^{\text{on},\kappa}(t^T) = \phi_{r,i}^\kappa(t^T) + \frac{n_{r,i}^\kappa(t^T - 1)}{T}.$$

Here we denote the capacity of the on-ramp in cell  $i$  by  $q_{r,i}^{\text{on}}$ . A good balance is to prioritize the mainstream flow, but increase the capacity and critical density of the cell into which the on-ramps merge, due to the presence of a merging lane.

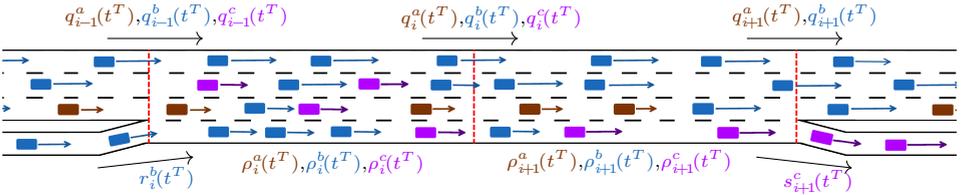


Figure 4.7: An example of three-class traffic flows in two cells. Vehicle classes  $a$ ,  $b$ , and  $c$  are colour-coded. Cell  $i$  receives traffic of all three classes from cell  $i - 1$  and class  $b$  vehicles from an on-ramp. Class  $a$  and  $b$  vehicles are mainstream-bound, and will leave cell  $i + 1$  and enter cell  $i + 2$ , whereas class  $c$  vehicles are off-ramp-bound and will leave cell  $i + 1$  via the off-ramp.

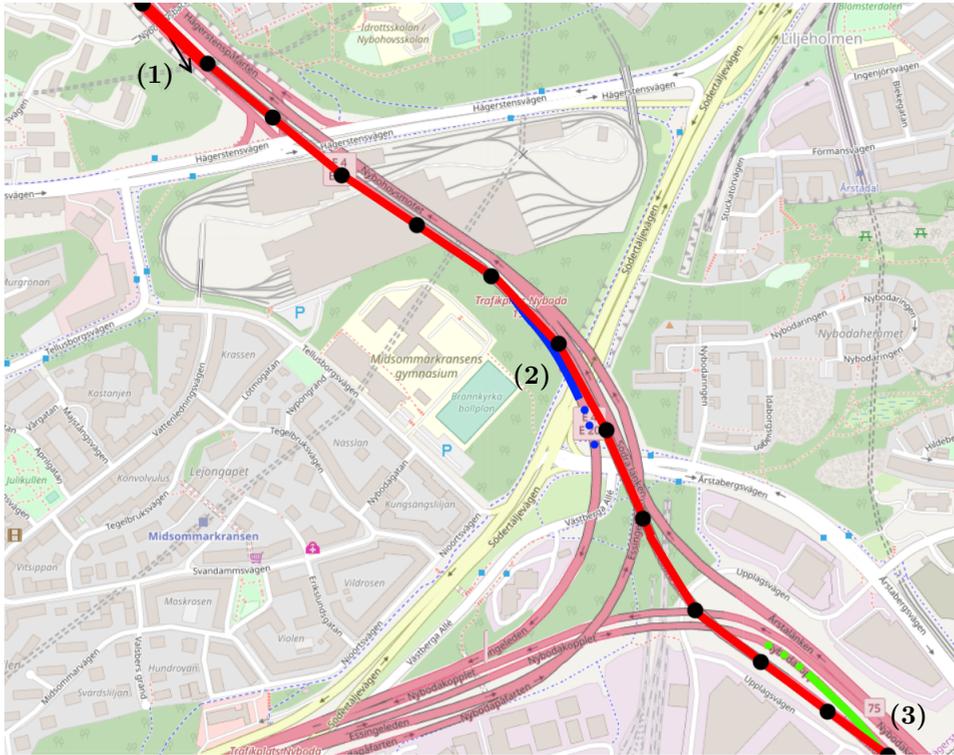


Figure 4.8: Map of Trafikplats Nyboda with an overlay illustrating an arrangement of cells, off-ramps, diverges and merges for modelling traffic originating from Eslingeleden and bound for Södra länken eastwards. The mainstream is shown in red, with black circles indicating cell boundaries. At (1) there is an off-ramp, at (2) a portion of the mainstream diverges westwards (shown in blue), and at (3) the mainstream traffic merges with the traffic from another link (shown in green). Imagery taken from OpenStreetMap.

This way, queueing will only occur if congestion from some bottleneck downstream propagates upstream and blocks the on-ramp.

An interesting benefit of using the multi-class CTM is that it gives us a way of precisely defining flows of off-ramps or diverging links. Instead of assuming that a fraction of all vehicles leaves the mainstream, we can now distinguish vehicles with different destinations as members of different classes. Let  $i$  be a cell with an off-ramp where vehicles of classes  $\mathcal{K}_{r,i} \subset \mathcal{K}$  exit the mainstream. We may then write

$$r_i^{\text{off},\kappa}(t^T) = \begin{cases} \min \left\{ d_i^\kappa(t^T), S_{i+1}^\kappa(t^T), \frac{\rho_i^\kappa(t^T)}{\rho_i^{\mathcal{K}_{r,i}}(t^T)} q_{r,i}^{\text{off}} \right\}, & \kappa \in \mathcal{K}_{r,i}, \\ 0, & \kappa \notin \mathcal{K}_{r,i}, \end{cases}$$

where  $q_{r,i}^{\text{off}}$  is the capacity of the off-ramp. Finally, we also update  $D_i^\kappa(t^T)$  accordingly,

$$D_i^\kappa(t^T) = \begin{cases} 0, & \kappa \in \mathcal{K}_{r,i}, \\ d_i^\kappa(t^T) \min \left\{ 1, \frac{q_i^{\max}(t^T)}{d_i^\kappa(t^T)} \right\}, & \kappa \notin \mathcal{K}_{r,i}. \end{cases}$$

While on- and off-ramps act as external sources and sinks of traffic flow, respectively, the case of roads merging or diverging needs to be studied separately. One such case is shown in Figure 4.8. In case we study a road network, where each link  $l$  has  $N_l$  cells, we will instead consider the traffic density of each class  $\kappa$ , cell  $i$ , and link  $l$ , separately as  $\rho_{i,l}^\kappa(t^T)$ . The system dynamics within links are unchanged, but the inflow to the first cell and outflow from the last cells need to be redefined. We define operator  $\vec{\ell}$  as a map to a set of links immediately downstream of all links in set  $\ell$ , and operator  $\overleftarrow{\ell}$  as a map to a set of links immediately upstream of all links in set  $\ell$ . Then, the class  $\kappa$  inflow to the first cell of link  $l$  is given as

$$q_{0,l}^\kappa(t^T) = q_{N_{\overleftarrow{\ell}}, \overleftarrow{\ell}}^\kappa(t^T),$$

where the aggregate quantity over multiple links in some set is denoted by writing that set in subscript,

$$q_{N_{\overleftarrow{\ell}}, \overleftarrow{\ell}}^\kappa(t^T) = \sum_{k \in \overleftarrow{\ell}} q_{N_k, k}^\kappa(t^T).$$

The class  $\kappa$  outflow from the last cell of link  $l$  is, conversely, given by modifying the supply available to vehicles in cell  $N_l$  of link  $l$ ,

$$S_{N_l, l}^\kappa(t^T) = \frac{\rho_{N_l, l}^\kappa(t^T)}{\rho_{N_{\overrightarrow{\ell}}, \overrightarrow{\ell}}^\kappa(t^T)} \min \left\{ W_{1, \overrightarrow{\ell}} \left( P_{1, \overrightarrow{\ell}} - \rho_{1, \overrightarrow{\ell}}^\kappa(t^T) \right), q_{1, \overrightarrow{\ell}}^{\max}(t^T) \right\},$$

and we assume the supply is distributed proportionally over all cells and all links that merge with the same links that link  $l$  merges with, i.e. links that are upstream of the links that are downstream from link  $l$ .

### 4.3.3 Platoons and moving bottlenecks

Considering multiple classes of vehicles moving at different speeds on the road allows us to separately treat specific formations of vehicles, such as vehicle platoon, by designating a vehicle class to represent them. However, although it is often driven by the need to classify automated and human-driven vehicles separately, multi-class CTM in its basic form is not suitable for modelling the behaviour of platoons if they move slower than the rest of the traffic. Namely, the traffic density profile of a platoon will become smoothed and spread to the neighbouring cells, due to the diffusion present in spatially discretized traffic models when the traffic speed is lower than the maximum allowed speed per Courant-Friedrichs-Lewy condition. This way we lose the information about the exact position of the platoon boundaries.

One way of dealing with this problem is to allow cell interfaces to move [213] and have these interfaces coincide with the boundaries of the platoons. However, the information about these boundary position can often be encoded in traffic density of some classes, as will be described in the remainder of this subsection.

Let there be some number of platoons, denoted by unique identifiers  $\xi \in \Xi$  on the road under consideration, and let each platoon  $\xi$  have reference speed  $u_\xi(t^T) \in [U^{\min}, U^{\max}]$ , with  $U^{\max} < V$ . We denote the position of the platoon head (downstream end)  $x_\xi(t^T)$ , and the platoon length  $l_\xi(t^T)$ , so the position of the platoon tail (upstream end) will be  $x_\xi(t^T) - l_\xi(t^T)$ . Here, class  $a$  will consist of platooned vehicles, and class  $b$  of the background human-driven traffic. Since vehicle platoons consist of a number of vehicles driving together as a single unit with constant regulated spacing, their traffic density profile is determined by  $x_\xi(t^T)$ ,  $l_\xi(t^T)$ , and their reference density  $\rho_\xi^*(t^T)$ ,

$$\rho_i^a(t^T) = \rho_{\xi,i}^{*a}(t^T) = \begin{cases} \rho_\xi^*(t^T) \frac{X_{i_\xi^t(t^T)+1} - x_\xi(t^T) + l_\xi(t^T)}{L}, & i = i_\xi^t(t^T), \\ \rho_\xi^*(t^T), & i_\xi^t(t^T) < i < i_\xi^h(t^T), \\ \rho_\xi^*(t^T) \frac{x_\xi(t^T) - X_{i_\xi^h(t^T)}}{L}, & i = i_\xi^h(t^T), \end{cases} \quad (4.19)$$

where  $i_\xi^h(t^T) = \lceil x_\xi(t^T)/L \rceil$  is the number of the cell where the platoon head is, and  $i_\xi^t(t^T) = \lceil (x_\xi(t^T) - l_\xi(t^T))/L \rceil$  the number of the cell where the platoon tail is, with operator  $\lceil x \rceil$  denoting the ceiling of  $x$ . Elsewhere, where there are no platoons, we expect  $\rho_i^a(t^T) = 0$ . Note that we require the length of the platoon to satisfy  $l_\xi(t^T) \geq 2L$ , so that at least one cell will have  $\rho_i^a(t^T) = \rho_\xi^*(t^T)$ . The platoon position update after  $T$  will be

$$x_\xi(t^T + 1) = x_\xi(t^T) + \min \left\{ u_\xi(t^T), \frac{q_{i_\xi^h(t^T)}^{\mathcal{K}}(t^T)}{\rho_{i_\xi^h(t^T)}^{\mathcal{K}}(t^T)} \right\} T,$$

and class  $a$  traffic densities need to be updated accordingly.

Note that simply setting  $U_i^a(t^T) = u_\xi(t^T)$  in cells where the platoon is would not be sufficient, since it would not maintain crisp boundaries of the platoon, as some vehicles would diffuse to the next cell. For example, for a two cells long platoon travelling at  $u_\xi = V/2$ , we would have

$$\begin{aligned} \rho_i^a(0) &= \rho_\xi^*, & \rho_{i+1}^a(0) &= \rho_\xi^*, & \rho_{i+2}^a(0) &= 0, & \rho_{i+3}^a(0) &= 0, \\ \rho_i^a(1) &= \frac{\rho_\xi^*}{2}, & \rho_{i+1}^a(1) &= \rho_\xi^*, & \rho_{i+2}^a(1) &= \frac{\rho_\xi^*}{2}, & \rho_{i+3}^a(1) &= 0, \\ \rho_i^a(2) &= \frac{\rho_\xi^*}{4}, & \rho_{i+1}^a(2) &= \frac{3\rho_\xi^*}{4}, & \rho_{i+2}^a(2) &= \frac{3\rho_\xi^*}{4}, & \rho_{i+3}^a(2) &= \frac{\rho_\xi^*}{4}, \end{aligned}$$

whereas the correct behaviour would be

$$\rho_i^a(2) = 0, \quad \rho_{i+1}^a(2) = \rho_p, \quad \rho_{i+2}^a(2) = 0, \quad \rho_{i+3}^a(2) = 0.$$

Assuming the other vehicle classes have no influence on the platooned vehicles, which holds if the road downstream of the platoon is in free flow, we can ensure that  $\rho_i^a(t^T)$  adheres to the desired traffic density profile (4.19) by setting  $U_i^a(t^T)$  to

$$U_i^a(t^T) = \begin{cases} V, & i_{\bar{\xi}}^-(t^T) \leq i < i_{\xi}^t(t^T), \\ V \min \left\{ 1, \max \left\{ 0, \frac{\rho_{\xi, i+1}^{*a}(t^T+1) - \frac{V-U_{i+1}^a(t^T)}{V} \rho_{i+1}^a(t^T)}{\rho_i^a(t^T)} \right\} \right\}, & i_{\xi}^t(t^T) \leq i < i_{\xi}^h(t^T), \\ V \min \left\{ 1, \frac{\rho_{\xi, i+1}^{*a}(t^T+1)}{\rho_i^a(t^T)} \right\}, & i = i_{\xi}^h(t^T), \\ 0, & i_{\xi}^h(t^T) < i \leq i_{\xi}^+(t^T), \end{cases} \quad (4.20)$$

in the area where the platoon is, denoted by  $i \in \{i_{\bar{\xi}}^-(t^T), \dots, i_{\xi}^+(t^T)\}$ . Since the reference traffic density profile  $\rho_{\xi, i}^{*a}(t^T)$  depends only on the position  $x_{\xi}(t^T)$  and reference density  $\rho_{\xi}^*(t^T)$  of platoon  $\xi$ , it is easy to calculate its one step prediction  $\rho_{\xi, i}^{*a}(t^T + 1)$ . Note that in case  $\rho_i^a(t^T) = \rho_{\xi}^*(t^T)$ ,  $i_{\xi}^t(t^T) < i < i_{\xi}^h(t^T)$ , we have  $U_i^a(t^T) = u_{\xi}(t^T)$ ,  $i_{\xi}^t(t^T) < i < i_{\xi}^h(t^T)$ . Even if the initial class  $a$  density profile differs from the reference, by applying (4.20) it will converge to (4.19) as long as  $u_{\xi}(t^T) < V$ . If a single platoon is present on the road, we have  $i_{\bar{\xi}}^-(t^T) = 1$ ,  $i_{\xi}^+(t^T) = N$ , otherwise  $i_{\bar{\xi}}^-(t^T)$  and  $i_{\xi}^+(t^T)$  are set to the cells at the halfway point between the platoons, i.e. if platoon  $\xi_2$  is upstream of platoon  $\xi_1$ , we have  $i_{\xi_2}^+(t^T) = \lceil (x_{\xi_2}(t^T) + x_{\xi_1}(t^T) - l_{\xi_1}(t^T)) / (2L) \rceil$ ,  $i_{\xi_1}^-(t^T) = i_{\xi_2}^+(t^T) + 1$ . If two platoons come close to each other,  $i_{\xi_2}^h(t^T) + 1 = i_{\xi_1}^t(t^T)$ , we consider the two platoons to have merged into a single one.

A platoon moving at speed slower than the rest of the traffic will act as a moving bottleneck to the rest of the traffic. Let the remainder of traffic consist of human-driven vehicles of class  $b$ ,  $U_i^b(t^T) = V$ , and  $\sigma_i = \sigma$ ,  $P_i = P$ ,  $W_i = W = V\sigma / (P - \sigma)$ ,  $L = VT$ . Assume a platoon stretches at least over two cells,

$$\rho_{i_{\xi}^h(t^T)-2}^a(t^T) = \rho_{i_{\xi}^h(t^T)-1}^a(t^T) = \rho_{\xi}^*(t^T) < \sigma, \quad (4.21)$$

and that cell  $i_{\xi}^h(t^T)$  is in free flow,  $\rho_{i_{\xi}^h(t^T)}^a(t^T) + \rho_{i_{\xi}^h(t^T)}^b(t^T) < \sigma$ . Then we will have

$$\begin{aligned} q_{i_{\xi}^h(t^T)-1}^b(t^T) &= V \rho_{i_{\xi}^h(t^T)-1}^b(t^T), \\ \rho_{i_{\xi}^h(t^T)-1}^b(t^T + 1) &= q_{i_{\xi}^h(t^T)-2}^b(t^T) / V. \end{aligned} \quad (4.22)$$

If the platoon is travelling at constant speed lower than the free flow speed,  $u_{\xi}(t^T) < V$ , and the class  $b$  inflow at the beginning of the road is constant,  $q_0^b(t^T) = \phi^b$ , the traffic density profile of both classes in cells  $i_{\xi}^h(t^T) - 1$  and  $i_{\xi}^h(t^T) - 2$  will reach a stationary state after some time, when

$$\begin{aligned} \rho_{i_{\xi}^h(t^T)-2}^b(t^T) &= \rho_{i_{\xi}^h(t^T)-1}^b(t^T), \\ q_{i_{\xi}^h(t^T)-2}^b(t^T) &= q_{i_{\xi}^h(t^T)-1}^b(t^T). \end{aligned} \quad (4.23)$$

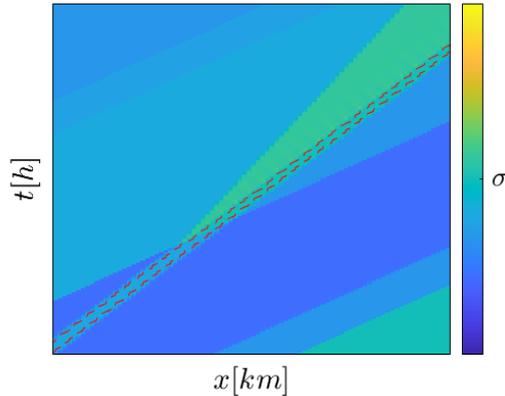


Figure 4.9: A platoon acting as a moving bottleneck implemented in multi-class CTM, shown outlined by dashed red lines. As the denser traffic reaches the platoon, congestion is formed in its wake.

Then, from (4.14) we have

$$q_{i_{\xi}^b(t^T)-2}(t^T) \leq \frac{\rho_{i_{\xi}^b(t^T)-2}^b(t^T)}{\rho_{i_{\xi}^a(t^T)-2}^a(t^T) + \rho_{i_{\xi}^b(t^T)-2}^b(t^T)} W \left( P - \rho_{i_{\xi}^a(t^T)-1}^a(t^T) - \rho_{i_{\xi}^b(t^T)-1}^b(t^T) \right),$$

and, substituting (4.21), (4.22), and (4.23), we may write

$$\rho_{i_{\xi}^b(t^T)-1}^b(t^T) \leq \frac{\rho_{i_{\xi}^b(t^T)-1}^b(t^T)}{\rho_{\xi}^*(t^T) + \rho_{i_{\xi}^b(t^T)-1}^b(t^T)} \frac{\sigma}{P - \sigma} \left( P - \rho_{\xi}^*(t^T) - \rho_{i_{\xi}^b(t^T)-1}^b(t^T) \right),$$

which simplifies to

$$\begin{aligned} \rho_{i_{\xi}^b(t^T)-1}^b(t^T) &\leq \sigma - \rho_{\xi}^*(t^T), \\ q_{i_{\xi}^b(t^T)-1}^b(t^T) &\leq V (\sigma - \rho_{\xi}^*(t^T)). \end{aligned} \quad (4.24)$$

Therefore, we confirm that the traffic flow overtaking a platoon with density  $\rho_{\xi}^*(t^T)$  is limited to  $V (\sigma - \rho_{\xi}^*(t^T))$ , which is consistent with PDE moving bottleneck models with triangular flux function. If  $\phi^b < V (\sigma - \rho_{\xi}^*(t^T))$ , the class  $b$  traffic will flow by the platoon unimpeded. Otherwise, the platoon acts as a moving bottleneck, and congestion will build up in its wake.

The effects of platoons and other vehicles acting as moving bottlenecks can also be modelled directly, through appropriate choice of  $U_i^{\kappa}(t^T)$ , without explicitly tracking the traffic density of this vehicle class. Let  $\xi$  now represent the moving

bottlenecks with speed  $u_\xi(t^T)$ , which restricts the flow at its position by scaling the flux function  $Q(\rho)$  down to

$$Q^{\beta_\xi(t^T)}(\rho) = (1 - \beta_\xi(t^T))Q\left(\frac{\rho}{1 - \beta_\xi(t^T)}\right),$$

where  $\beta_\xi(t^T) \in [0, 1)$  indicates the severity of the moving bottleneck, with  $\beta_\xi(t^T) = 0$  corresponding to not obstructing the flow at all, and  $\beta_\xi(t^T) \approx 1$  corresponding to fully blocking the road, similarly to how moving bottlenecks were handled in Section 4.2. The position of the moving bottleneck evolves according to  $x_\xi(t^T + 1) = x_\xi(t^T) + u_\xi(t^T)T$ , and the cell that contains it is  $i_\xi^z(t^T) = \lceil x_\xi(t^T)/L \rceil$ . Then the traffic density profile in cells around the moving bottleneck should follow

$$\rho_i^{\mathcal{K}}(t^T) = \rho_{\xi, i}^{*\mathcal{K}}(t^T) = \begin{cases} \rho_\xi^c(t^T), & i = i_\xi^z(t^T) - 1, \\ \rho_\xi^c(t^T) - \frac{X_{i_\xi^z(t^T)} - x_\xi(t^T)}{L}(\rho_\xi^c(t^T) - \rho_\xi^d(t^T)), & i = i_\xi^z(t^T), \\ \rho_\xi^d(t^T), & i = i_\xi^z(t^T) + 1, \end{cases}$$

where the densities of the congestion in the wake of the moving bottleneck  $\rho_\xi^c(t^T)$ , and of the overtaking flow  $\rho_\xi^d(t^T)$ , are

$$\rho_\xi^c(t^T) = \frac{WP - (V - u_\xi(t^T))(1 - \beta_\xi(t^T))\sigma}{W + u_\xi(t^T)}, \quad (4.25)$$

$$\rho_\xi^d(t^T) = (1 - \beta_\xi(t^T))\sigma. \quad (4.26)$$

This can be achieved by setting the free flow speed for all classes to

$$U_i^{\mathcal{K}}(t^T) = \begin{cases} V \min \left\{ 1, \frac{\rho_{\xi, i}^{*\mathcal{K}}(t^T + 1)}{\rho_i^{\mathcal{K}}(t^T)} \right\}, & i = i_\xi^z(t^T), \\ V \min \left\{ 1, \max \left\{ 0, \frac{\rho_{\xi, i}^{*\mathcal{K}}(t^T + 1) - \frac{V - U_{i+1}^{\mathcal{K}}(t^T)}{V} \rho_{i+1}^{\mathcal{K}}(t^T)}{\rho_i^{\mathcal{K}}(t^T)} \right\} \right\}, & i_\xi^z(t^T) - i \in \{1, 2\}. \end{cases} \quad (4.27)$$

One can verify that this has the same effect on the rest of the traffic as a platoon with reference density  $\rho_\xi^*(t^T) = \beta_\xi(t^T)\sigma$  travelling at the same speed  $u_\xi(t^T)$ .

#### 4.3.4 Capacity drop and stop-and-go waves

Out of many ways of modeling capacity drop in first-order traffic models [38], we chose to capture it as a linear reduction of capacity, as in [214]. We model the capacity drop phenomenon by modifying the cell capacity  $q_i^{\max}(t^T)$  depending on the traffic density in the cell. Denoting by  $\alpha$  the maximum capacity drop ratio under jam traffic density, we write

$$q_i^{\max}(t^T) = \min \left\{ \frac{\sum_{\kappa \in \mathcal{K}} d_i^\kappa(t^T) \frac{V_i P_i \sigma_i U_i^\kappa(t^T)}{(P_i - \sigma_i) U_i^\kappa(t^T) + V_i \sigma_i}}{d_i^\kappa(t^T)}, W_i \frac{V_{i+1} \sigma_{i+1}}{V_i \sigma_i} (P_i - (1 - \alpha) \sigma_i - \alpha \rho_i^{\mathcal{K}}(t^T)) \right\}. \quad (4.28)$$

Substituting  $\rho_i^{\mathcal{K}}(t^T) = P_i$  into the second operand of minimum in (4.28), and assuming  $U_i^{\mathcal{K}}(t^T) = V_i$ , we can see that in that case

$$q_i^{\max}(t^T) = \min \{V_i\sigma_i, V_{i+1}\sigma_{i+1}(1 - \alpha)\}.$$

Consider a road with homogeneous free flow speed  $V_i = V$  and a stationary bottleneck in cell  $i + 1$ , with a significantly lower critical density  $\sigma_{i+1} < \sigma_i$ , corresponding to the maximum capacity of  $q_{i+1}^{\max} \leq V\sigma_{i+1}$ . Due to the capacity drop phenomenon, in case of excess demand at the bottleneck, its capacity will be decreased once it becomes congested. A congestion of density  $\rho^c$  will be formed in cell  $i$ , with the density of discharging traffic being  $\rho^d$ . The congestion density  $\rho^c$  can be calculated from  $W_i(P_i - \rho^c) = W_i \frac{\sigma_{i+1}}{\sigma_i} (P_i - (1 - \alpha)\sigma_i - \alpha\rho^c)$ , so that

$$\rho^c = \frac{P_i(\sigma_i - \sigma_{i+1}) + (1 - \alpha)\sigma_i\sigma_{i+1}}{\sigma_i - \alpha\sigma_{i+1}}. \quad (4.29)$$

We calculate the discharge density from  $V\rho^d = W_i(P_i - \rho^c)$ :

$$\rho^d = \frac{\sigma_i\sigma_{i+1}(1 - \alpha)}{\sigma_i - \alpha\sigma_{i+1}} < \sigma_{i+1},$$

and the discharging flow from the congested bottleneck will be lower than the maximum capacity,  $V\rho^d \leq V\sigma_{i+1}$ .

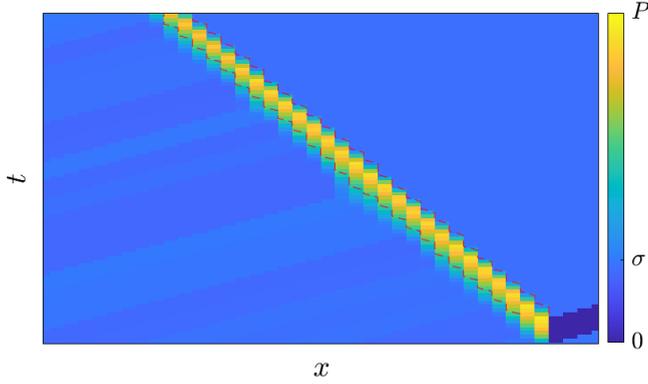
If an active bottleneck is by some means removed from the road, the built up congestion will begin to dissipate, propagating upstream in the process, as a stop-and-go wave. As was the case with platoons discussed in the previous subsection, modelling the capacity drop of stop-and-go waves also suffers from diffusion inherent in spatially discretized traffic models. For example, consider a homogeneous road where a traffic light between cells  $i$  and  $i + 1$  is blocking the traffic until time  $t = 0$ . Starting with  $t = 1$ , the traffic light turns green and the accumulated congestion will begin dissipating, and the traffic density  $\rho_i^{\mathcal{K}}(t^T)$  will evolve with traffic flows determined by (4.17), where  $q_i^{\max}(t^T)$  includes capacity drop (4.28), as

$$\begin{aligned} \rho_{i-1}^{\mathcal{K}}(0) &= P, & \rho_i^{\mathcal{K}}(0) &= P, & \rho_{i+1}^{\mathcal{K}}(0) &= 0, \\ \rho_{i-1}^{\mathcal{K}}(1) &= P, & \rho_i^{\mathcal{K}}(1) &= P - (1 - \alpha)\sigma, & \rho_{i+1}^{\mathcal{K}}(1) &= (1 - \alpha)\sigma, \\ \rho_{i-1}^{\mathcal{K}}(2) &= P - \frac{(1 - \alpha)\sigma^2}{P - \sigma}, & \rho_i^{\mathcal{K}}(2) &= P - (1 - \alpha)\left(2\sigma - \frac{(1 - \alpha)\sigma^2}{P - \sigma}\right), & \rho_{i+1}^{\mathcal{K}}(2) &= (1 - \alpha)\left(\sigma + \frac{\alpha\sigma^2}{P - \sigma}\right), \end{aligned}$$

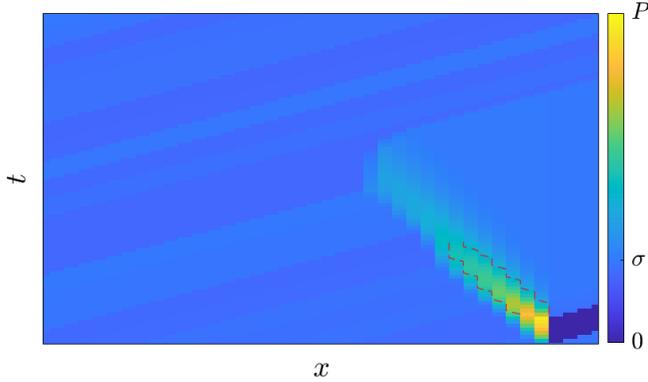
whereas the correct behaviour would be

$$\rho_{i-1}^{\mathcal{K}}(2) = P, \quad \rho_i^{\mathcal{K}}(2) = P - 2(1 - \alpha)\sigma, \quad \rho_{i+1}^{\mathcal{K}}(2) = (1 - \alpha)\sigma.$$

Essentially, capacity drop in cell  $i$  becomes diminished as soon as the wavefront of the downstream end of the dissipating congestion starts propagating upstream through the cell. Furthermore, due to diffusion, vehicles in cell  $i - 1$  were able to



(a) A stop-and-go wave if the diffusion is properly handled



(b) A stop-and-go wave if the diffusion is not properly handled

Figure 4.10: The effect of diffusion on stop-and-go waves.

start moving before the downstream end of dissipating congestion reached them. A comparison of the correct behaviour of the stop-and-go wave and its premature dissipation due to diffusion is shown in Figure 4.10.

Therefore, we use a similar technique as the one employed in the previous subsection to eliminate the effect of diffusion on stop-and-go waves. Let there be some number of stop-and-go waves, denoted by unique identifiers  $\psi \in \Psi$  on the road under consideration, characterised by their congestion traffic densities  $\rho_\psi^c(t^T)$ , and with positions of their downstream wavefronts  $z_\psi(t^T)$ . According to (4.28), the density of the traffic flow dissipating from this stop-and-go wave is

$$\rho_\psi^d(t^T) = \frac{W_{i_\psi^z}(t^T) \sigma_{i_\psi^z}(t^T)+1}{V_{i_\psi^z}(t^T) \sigma_{i_\psi^z}(t^T)} \left( P_{i_\psi^z}(t^T) - (1 - \alpha) \sigma_{i_\psi^z}(t^T) - \alpha \rho_\psi^c(t^T) \right), \quad (4.30)$$

where  $i_\psi^z(t^T) = \lceil z_\psi(t^T)/L \rceil$  is the number of the cell where the downstream wavefront of stop-and-go wave  $\psi$  is. The congestion traffic density of stop-and-go waves is updated every time step to

$$\rho_\psi^c(t^T + 1) = \max \left\{ \rho_\psi^c(t^T), \rho_{i_\psi^z(t^T)}^{\mathcal{K}}(t^T), \rho_{i_\psi^z(t^T)-1}^{\mathcal{K}}(t^T) \right\}.$$

This stop-and-go wave propagates upstream with speed  $\lambda_\psi(t^T) < 0$ ,

$$z_\psi(t^T + 1) = z_\psi(t^T) \lambda_\psi(t^T) T,$$

and the traffic density profile in cells around its downstream wavefront should follow

$$\rho_i^{\mathcal{K}}(t^T) = \rho_{\psi,i}^{*\mathcal{K}}(t^T) = \begin{cases} \rho_\psi^c(t^T), & i = i_\psi^z(t^T) - 1, \\ \rho_\psi^c(t^T) - \frac{X_{i_\psi^z(t^T)} - z_\psi(t^T)}{L} (\rho_\psi^c(t^T) - \rho_\psi^d(t^T)), & i = i_\psi^z(t^T), \\ \rho_\psi^d(t^T), & i = i_\psi^z(t^T) + 1. \end{cases}$$

This can be achieved by setting the free flow speed in the vicinity of the stop-and-go wave for all classes to

$$U_i^{\mathcal{K}}(t^T) = \begin{cases} V_i \min \left\{ 1, \frac{\rho_{\psi,i}^{*\mathcal{K}}(t^T+1)}{\rho_i^{\mathcal{K}}(t^T)} \right\}, & i = i_\psi^z(t^T), \\ V_i \min \left\{ 1, \max \left\{ 0, \frac{\rho_{\psi,i}^{*\mathcal{K}}(t^T+1) - \frac{V_i - U_{i+1}^{\mathcal{K}}(t^T)}{V_i} \rho_{i+1}^{\mathcal{K}}(t^T)}{\rho_i^{\mathcal{K}}(t^T)} \right\} \right\}, & i_\psi^z(t^T) - i \in \{1, 2\}. \end{cases} \quad (4.31)$$

The stop-and-go wavefront propagation speed  $\lambda_\psi(t^T)$  is given by

$$\rho_\psi^c(t^T) + \frac{\sigma_{i_\psi^z(t^T)}}{P_{i_\psi^z(t^T)} - \sigma_{i_\psi^z(t^T)}} \left( P_{i_\psi^z(t^T)} - \rho_\psi^c(t^T) \right) - \rho_\psi^d(t^T) = \rho_\psi^c + \frac{\lambda_\psi(t^T)}{V_{i_\psi^z(t^T)} (\rho_\psi^c(t^T) - \rho_\psi^d(t^T))},$$

which, substituting (4.30), yields

$$\lambda_\psi(t^T) = V_i \frac{(1-\alpha)\sigma_i\sigma_{i+1} + P_i(\sigma_i - \sigma_{i+1}) - \rho_\psi^c(t^T)(\sigma_i - \alpha\sigma_{i+1})}{(1-\alpha)\sigma_i\sigma_{i+1} - P_i\sigma_{i+1} + \rho_\psi^c(t^T)(P_i - \sigma_i + \alpha\sigma_{i+1})}, \quad i = i_\psi^z(t^T). \quad (4.32)$$

In case the road is homogeneous,  $\sigma = \sigma_{i_\psi^z(t^T)} = \sigma_{i_\psi^z(t^T)+1}$ ,  $P = P_{i_\psi^z(t^T)} = P_{i_\psi^z(t^T)+1}$ ,  $V = V_{i_\psi^z(t^T)} = V_{i_\psi^z(t^T)+1}$ , (4.32) simplifies to

$$\lambda_\psi(t^T) = -V \frac{(1-\alpha)\sigma}{P - (1-\alpha)\sigma},$$

and is independent of  $\rho_\psi^c(t^T)$ . Note that substituting (4.29) into (4.32) yields  $\lambda_\psi(t^T) = 0$ , i.e. if congestion accumulates at a stationary bottleneck, it will not start propagating upstream as a stop-and-go wave, but instead remains at the bottleneck, as expected.

New stop-and-go waves are created whenever the outflow of some cell is limited by the capacity drop,

$$q_i^{\mathcal{K}}(t^T) = W_i \frac{V_{i+1} \sigma_{i+1}}{V_i \sigma_i} (P_i - (1 - \alpha) \sigma_i - \alpha \rho_i^{\mathcal{K}}(t^T)),$$

and there are no stop-and-go waves already present in cell  $i$ . The downstream wavefront position of the thus newly created stop-and-go wave is set to  $z_\psi(t^T) = X_{i+1}$ , and its congestion traffic density is set to  $\rho_\psi^{\mathcal{K}} = \rho_i^{\mathcal{K}}(t^T)$ . Conversely, a stop-and-go wave is considered to be dissipated once both  $\rho_{i^z_\psi}^{\mathcal{K}}(t^T) \leq \sigma_{i^z_\psi}(t^T)$  and  $\rho_{i^z_\psi-1}^{\mathcal{K}}(t^T) \leq \sigma_{i^z_\psi-1}(t^T)$ , and its identifier is then removed from  $\Psi$ .

### 4.3.5 Multi-class CTM with a general flux function

While the triangular fundamental diagram (4.18) is widely used due to its simplicity and ease of calibration, many other fundamental diagram types exist. The multi-class CTM presented in this section was initially developed with triangular fundamental diagram in mind, but can be extended to admit an arbitrary flux function  $Q_i(\rho)$ , with speed-density relation  $Q_i(\rho) = \mathcal{V}_i(\rho)\rho$ . We define the critical density of aggregate traffic  $\sigma_i$  as the traffic density for which  $Q_i(\rho)$  is maximized,

$$\sigma_i = \arg \max_{\rho \geq 0} Q(\rho), \quad q_i^{\max} = Q_i(\sigma_i),$$

and the maximum free flow speed  $V_i$  is given as the maximum slope of  $Q_i(\rho)$ ,

$$V_i = \max_{\rho \geq 0} Q'(\rho),$$

which for the vast majority of commonly used flux functions happens at zero,  $V_i = Q'(0)$ . For each vehicle class  $\kappa$ , we also define a separate critical density  $\sigma_i^\kappa(t^T)$  for which  $\min\{Q_i(\rho), U_i^\kappa(t^T)\rho\}$  is maximized,

$$\sigma_i^\kappa(t^T) = \arg \max_{\rho \geq 0} \min\{Q(\rho), U_i^\kappa(t^T)\rho\}, \quad q_i^{\max, \kappa}(t^T) = Q_i(\sigma_i^\kappa(t^T)),$$

The majority of equations stays the same as defined in the base model in Subsection 4.3.1, and the redefined equations are listed in what follows. First, the maximum demand of each vehicle class if other classes were absent (4.15) is now given by

$$d_i^\kappa(t^T) = \min \left\{ U_i^\kappa(t^T), \mathcal{V}(\underline{\rho}_i^\kappa(t^T)) \right\} \rho_i^\kappa(t^T),$$

where  $\underline{\rho}_i^\kappa(t^T) = \min\{\rho_i^\kappa(t^T), \sigma_i^\kappa(t^T)\}$ . The aggregate demand function remains unchanged, but the aggregate supply now becomes

$$S_i^\mathcal{K}(t^T) = \min \left\{ Q_i(\bar{\rho}_{i+1}^\mathcal{K}(t^T)), q_{i+1}^{\max}(t^T) \right\} = \min \left\{ \mathcal{V}_i(\bar{\rho}_{i+1}^\mathcal{K}(t^T)) \bar{\rho}_{i+1}^\mathcal{K}(t^T), q_{i+1}^{\max}(t^T) \right\},$$

where  $\bar{\rho}_{i+1}^{\mathcal{K}}(t^T) = \max\{\rho_{i+1}^{\mathcal{K}}(t^T), \sigma_i\}$ , and the capacity of cell  $i$  is redefined to

$$q_i^{\max}(t^T) = \frac{\sum_{\kappa \in \mathcal{K}} d_i^{\kappa}(t^T) q_i^{\max, \kappa}(t^T)}{d_i^{\mathcal{K}}(t^T)}.$$

We may also generalize the modelling of moving bottlenecks and stop-and-go waves, by redefining  $\rho_{\xi}^c(t^T)$ ,  $\rho_{\xi}^d(t^T)$ ,  $\rho_{\psi}^c(t^T)$ , and  $\rho_{\psi}^d(t^T)$ . For modelling moving bottlenecks,  $\rho_{\xi}^c(t^T)$  and  $\rho_{\xi}^d(t^T)$  are given as the maximum and the minimum solution of

$$Q(\rho) = Q^{\beta_{\xi}(t^T)}(\rho^{\tau}) + u_{\xi}(t^T)(\rho - \rho^{\tau}),$$

respectively, where  $\rho^{\tau}$  is the traffic density at which a line with slope  $u_{\xi}(t^T)$  is tangent to the reduced flux function  $Q^{\beta_{\xi}(t^T)}(\rho)$ ,

$$\rho^{\tau} = \arg \max_{\rho} Q^{\beta_{\xi}(t^T)}(\rho) - u_{\xi}(t^T)\rho,$$

instead of by (4.25) and (4.26). For modelling stop-and-go waves, we now need to specify a wavefront propagation speed  $\lambda_{\psi}(t^T)$ . Then, for  $\rho_{\psi}^c(t^T)$  such that  $Q'(\rho_{\psi}^c) < \lambda_{\psi}(t^T)$ ,  $\rho_{\psi}^d(t^T)$  is given as the minimum solution of

$$Q(\rho_{\psi}^d(t^T)) = Q(\rho_{\psi}^c(t^T)) + \lambda_{\psi}(t^T)(\rho_{\psi}^d(t^T) - \rho_{\psi}^c(t^T)),$$

instead of by (4.30), and a stop-and-go wave is considered to be dissipated once  $Q'(\rho_{\psi}^c(t^T)) \geq \lambda_{\psi}(t^T)$ .

#### 4.4 Tandem queueing model with platoons

The base tandem queueing model presented in Section 3.4 can already represent the capacity drop phenomenon, and is indeed the simplest model that is able to do this. However, including moving traffic phenomena into the model proves to be much more difficult. Consider the traffic situation shown in Figure 4.11. The lane drop at the downstream end of the shown road segment is acting as an obvious stationary bottleneck, which can be modelled as a queueing server. If the trucks, shown in grey, would not be moving, they too would represent stationary bottlenecks for the rest of the traffic, and the whole system could be described as a tandem queue with some delay and  $n_b(t)$  corresponding to the lane drop bottleneck, and  $n_1(t)$  and  $n_2(t)$  corresponding to queues at the trucks. However, if the truck platoons are moving along the road, they will affect the rest of the traffic in a different way, and the delays between these queues would vary in time. In this section we extend the tandem queueing model to allow for queues at moving bottlenecks. We then present a coordinate transform that simplifies the model, eliminating delay between the queues. Finally, the proposed model is validated against the microscopic simulations executed in SUMO and multi-class CTM.

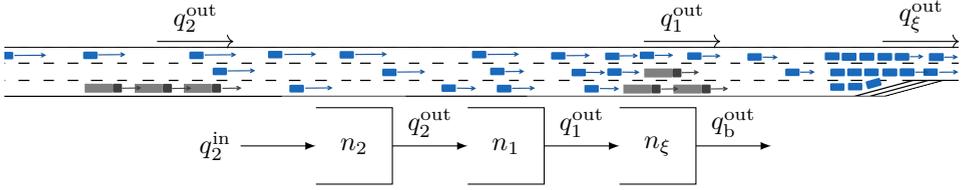


Figure 4.11: Queues corresponding to stationary and moving bottlenecks. The stationary bottleneck corresponds to  $n_b$ , the downstream platoon to  $n_1$  and the upstream platoon to  $n_2$ . The overtaking flow of the downstream platoon  $q_1^{\text{out}}$  is limited to one lane of traffic,  $q_1^{\text{cap}} = V\sigma_l$ , where  $\sigma_l$  is the critical density of a single highway lane, and the overtaking flow of the upstream platoon  $q_2^{\text{out}}$  is limited to two lanes of traffic,  $q_2^{\text{cap}} = V2\sigma_l$ .

#### 4.4.1 Moving bottlenecks with a single stationary bottleneck

Consider a highway stretch with a single stationary bottleneck at position  $X_b$  at its downstream end. We denote the queue at this bottleneck  $n_b(t)$ . Apart from this stationary bottleneck, at time  $t_0$ , there is a number of platoons  $\xi \in \Xi(t_0)$  potentially acting as moving bottlenecks, moving in the traffic flow, and we denote their position  $x_\xi(t)$ . We may enumerate the truck platoons so that  $x_1(t_0) > x_2(t_0) > \dots > x_\Xi(t_0)$ , i.e. the platoon that is farther along the road has a lower index than the ones upstream of it. In this case, assuming the platoons do not overtake each other, the platoon immediately downstream of platoon  $\xi$  is  $\bar{\xi} = \xi - 1$ , if  $\xi \neq 1$ , where platoon  $\xi = 1$  is the first platoon immediately upstream of the stationary bottleneck at time  $t_0$ . Similarly, the platoon immediately upstream of platoon  $\xi$  is  $\bar{\xi} = \xi + 1$ , if  $\xi \neq \Xi(t_0)$ , where platoon  $\xi = \Xi(t_0)$  is the upstream-most platoon in  $\Xi(t_0)$  at time  $t_0$ . The queues at the platoons will be denoted  $n_\xi(t)$ , and together with  $n_b(t)$ , they constitute the state of the system.

The evolution of the queue at the stationary bottleneck is given by

$$\dot{n}_b(t) = q_b^{\text{in}}(t) - q_b^{\text{out}}(t),$$

where the outflow is

$$q_b^{\text{out}}(t) = \begin{cases} \min\{q_b^{\text{in}}(t), q_b^{\text{cap}}\}, & n_b(t) = 0, \\ q_b^{\text{dis}}, & q_b^{\text{in}}(t) > n_b(t) > 0. \end{cases}$$

As is standard, due to capacity drop, the discharge rate of the queue at the bottleneck  $q_b^{\text{dis}}$  will be lower than its capacity  $q_b^{\text{cap}}$ ,  $q_b^{\text{dis}} < q_b^{\text{cap}}$ .

Let the free flow speed of the background traffic be  $V$  everywhere on the road. Then the traffic flow originating from position  $x$  at time  $t_0$  reaches the stationary bottleneck after  $(X_b - x)/V$ , unless it is blocked by some other bottleneck. We denote by  $t_\xi^V(t_0)$  the time when traffic flow originating from the position of platoon

$\xi$  at some time  $t_0$  would reach the stationary bottleneck,

$$t_\xi^V(t_0) = t_0 + \frac{X_b - x_\xi(t_0)}{V}. \quad (4.33)$$

Assuming that the platoons are moving at their individual constant reference speeds  $u_\xi \leq V$ , we define the time  $t_\xi^u(t_0)$ , at which they reach the stationary bottleneck

$$t_\xi^u(t_0) = t_0 + \frac{X_b - x_\xi(t_0)}{u_\xi}. \quad (4.34)$$

Note that since we assume that  $x_\xi(t) = x_\xi(t_0) + (t - t_0)u_\xi(t_0)$ ,  $t_\xi^u(t) = t_\xi^u(t_0)$  and  $t_\xi^u(t)$  does not depend on time, so we omit writing  $(t_0)$  in further text.

We also assume that the platoons do not merge before reaching the stationary bottleneck,  $t_1^u < t_2^u < \dots < t_\Xi^u$ , which constrains their admissible reference speeds. Note that the assumption about the constant speed for each platoon is only significant until  $t_\xi^u$  when that platoon reaches the stationary bottleneck, as it then leaves the road segment under consideration.

The inflow to the queue at the stationary bottleneck  $q_b^{\text{in}}(t)$  is given by

$$q_b^{\text{in}}(t) = q_b^u(t) + q_b^V(t),$$

and it consists of two parts that travel at different speeds. The first part,  $q_b^u(t)$ , models the part of the flow that originates from the arrival of the platooned vehicles,

$$q_b^u(t) = \begin{cases} V\sigma_l, & t_\xi^u \leq t \leq t_\xi^u + \frac{l_\xi}{V}, \xi \in \{1, \dots, \Xi\}, \\ 0, & \text{otherwise,} \end{cases}$$

where  $l_\xi$  is the length of platoon  $\xi$  when it is taking a single lane. We assume that each platoon will approach the bottleneck taking up one lane, thus its density will be equal to the critical density per lane  $\sigma_l$ . The second part consists of the background traffic travelling at free flow speed  $V$ ,

$$q_b^V(t) = \begin{cases} q_1^{\text{out}} \left( \frac{x_1(t_0) + Vt - X_b}{V - u_1} \right), & t < t_1^u, \\ q_\xi^{\text{out}} \left( \frac{x_\xi(t_0) + Vt - X_b}{V - u_\xi} \right), & t_{\xi-1}^u \leq t < t_\xi^u, \xi \in \{2, \dots, \Xi\}, \end{cases}$$

originating from the discharging flow of the queue at one of the platoons, i.e., the traffic flow overtaking one of the platoons. If the contribution of platooned vehicles to the queue at the stationary bottleneck is negligible, we may set  $q_b^u(t) = 0$  and  $q_b^{\text{in}}(t) = q_b^V(t)$ .

The evolution of the queue at platoon  $\xi$  is given by

$$\dot{n}_\xi(t) = \frac{V - u_\xi}{V} (q_\xi^{\text{in}}(t) - q_\xi^{\text{out}}(t)), \quad 0 \leq t \leq t_\xi^u,$$

for  $\xi = 1, \dots, \Xi$ , which is defined until time  $t_\xi^u$ , when the platoon reaches the bottleneck and their queues merge,

$$n_b(t_\xi^u +) = n_b(t_\xi^u) + n_\xi(t_\xi^u).$$

The outflow from a queue at a platoon is defined the same way as with the queue at the stationary bottleneck,

$$q_\xi^{\text{out}}(t) = \begin{cases} \min\{q_\xi^{\text{in}}(t), q_\xi^{\text{cap}}(t)\}, & n_\xi(t) = 0, \\ q_\xi^{\text{dis}}(t), & n_\xi(t) > 0, \end{cases}$$

for  $\xi = 1, \dots, \Xi$ , except here we assume  $q_\xi^{\text{dis}}(t) = q_\xi^{\text{cap}}(t)$ , and allow  $q_\xi^{\text{cap}}(t)$  to vary in time and be used as a control input. The inflow to the platoons is given by

$$q_\xi^{\text{in}}(t) = q_{\xi+1}^{\text{out}} \left( \frac{(V - u_\xi)t - x_\xi + x_{\xi+1}}{V - u_{\xi+1}} \right),$$

for  $\xi = 1, \dots, \Xi - 1$ , and the inflow to platoon  $\Xi$  is given as an external input.

In case there are on- and off-ramps, their influence can be added to  $q_b^V(t)$  and  $q_\xi^{\text{in}}(t)$ . Denoting by  $q_k^r(t)$  the inflow from an on-ramp (if  $q_k^r(t) > 0$ ), or outflow to an off-ramp (if  $q_k^r(t) < 0$ ), at position  $X_k^r$ ,  $k \in K$ , we may write

$$\begin{aligned} q_b^V(t) &= q_b^{V \setminus r}(t) + \sum_{k \in K_b(t)} q_k^r \left( t - \frac{X_b - X_k^r}{V} \right), \quad (4.35) \\ q_b^{V \setminus r}(t) &= \begin{cases} q_1^{\text{out}} \left( \frac{x_1(t_0) + Vt - X_b}{V - u_1} \right), & t < t_1^u, \\ q_\xi^{\text{out}} \left( \frac{x_\xi(t_0) + Vt - X_b}{V - u_\xi} \right), & t_{\xi-1}^u \leq t < t_\xi^u, \xi \in \{2, \dots, \Xi\}, \end{cases} \\ K_b(t) &= \left\{ k \in K \mid t \notin \left[ t_\xi^V \left( \frac{X_k^r - x_\xi(t_0) + u_\xi t_0}{u_\xi} \right), t_\xi^u \right], \xi \in \{1, \dots, \Xi\} \right\} \end{aligned}$$

and for the inflow to the queue at platoons,

$$\begin{aligned} q_\xi^{\text{in}}(t) &= q_\xi^{\text{in} \setminus r}(t) + \sum_{k \in K_\xi(t)} q_k^r \left( t - \frac{X_k^r - x_\xi(t)}{V} \right), \\ q_\xi^{\text{in} \setminus r}(t) &= q_{\xi+1}^{\text{out}} \left( \frac{(V - u_\xi)t - x_\xi(t_0) + x_{\xi+1}(t_0)}{V - u_{\xi+1}} \right), \\ K_\xi(t) &= \left\{ k \in K \mid x_{\xi+1} \left( t - \frac{x_\xi(t) - X_k^r}{V} \right) \leq X_k^r \right\}. \end{aligned}$$

Here,  $K_d(t)$  are sets of indices of all on- and off-ramps with positions  $X_k^r < X_b$  between the bottleneck (in case of  $K_b(t)$ ) or platoon  $\xi$  (in case of  $K_\xi(t)$ ), and the place where their inflows would originate from. Note that  $q_k^r(t)$  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  $n_\xi$  at the time when the platoon reaches it,

$$n_\xi(t+) = n_\xi(t) - n_\xi^{r,k}(t), \quad x_\xi(t) = X_k^r,$$

and the part of the queue  $n_\xi(t)$  that leaves the highway,  $n_\xi^{r,k}(t)$ , depends on the length of the queue at the platoon and the ratio of off-ramp-bound vehicles in it.

In order to initialize the model at some time  $t = t_0$ , we define the initial conditions  $n_b(t_0)$  and  $n_\xi(t_0)$ ,  $\xi = 1, \dots, \Xi$ . We also need to know the past values of outflows from the platoons  $q_\xi^{\text{out}}(t)$ , and the past flows of any on- and off-ramps present on the road  $q_k^r(t)$ . We need to know  $q_\xi^{\text{out}}(t)$  for time  $t_0 - \tau_{0,\xi}(t_0) \leq t < t_0$ , where  $\tau_{0,\xi}(t_0)$  is given by

$$\tau_{0,\xi}(t_0) = \begin{cases} \frac{X_b - x_\xi(t_0)}{V - u_\xi}, & \xi = 1, \\ \frac{x_{\xi-1}(t_0) - x_\xi(t_0)}{V - u_\xi}, & \xi \in \{2, \dots, \Xi\}, \end{cases}$$

and  $q_k^r(t)$  for time  $t_0 - \tau_{0,k}^r(t_0) \leq t < t_0$ , where  $\tau_{0,k}^r(t_0)$  is given by

$$\tau_{0,k}^r(t_0) = \begin{cases} \frac{X_b - X_k^r}{V}, & x_1(t_0) > X_k^r, \\ \frac{x_\xi(t_0) - X_k^r}{V}, & x_{\xi+1}(t_0) \leq X_k^r < x_\xi(t_0), \xi \in \{1, \dots, \Xi - 1\}, \\ \frac{x_\Xi(t_0) - X_k^r}{V}, & X_k^r < x_\Xi(t_0). \end{cases}$$

In case we want to initialize the model using the traffic density profile at time  $t_0$ ,  $\rho(x, t_0)$ , we may do so by setting  $n_b(t_0) = 0$ ,  $n_\xi(t_0) = 0$ ,  $\xi = 1, \dots, \Xi$ , and

$$q_\xi^{\text{out}}(t) = V\rho(x_\xi(t_0) + (V - u_\xi)(t_0 - t), t_0), \quad \xi = 1, \dots, \Xi, \quad (4.36)$$

with the ramp flows  $q_k^r(t) = 0$ ,  $t \leq t_0$ . Note that (4.36) may yield unrealistic values for overtaking flows, as the flows from the ramps would in reality have played a role in shaping  $\rho(x, t_0)$ , but this will have no effect on the state of the system for  $t > t_0$ .

#### 4.4.2 Coordinate transformation for simplifying delay

It is apparent that the introduction of moving bottlenecks to the tandem queueing model significantly increases its complexity, even in case their speeds are constant. However, in this case, the model can be greatly simplified by adopting a coordinate transformation

$$\begin{aligned} \tau_\xi &= \frac{x_\xi(t_0) - X_b - u_\xi t_0 + Vt}{V - u_\xi}, \\ t &= \frac{V - u_\xi}{V} \tau_\xi + \frac{X_b - x_\xi(t_0) + u_\xi t_0}{V}, \end{aligned}$$

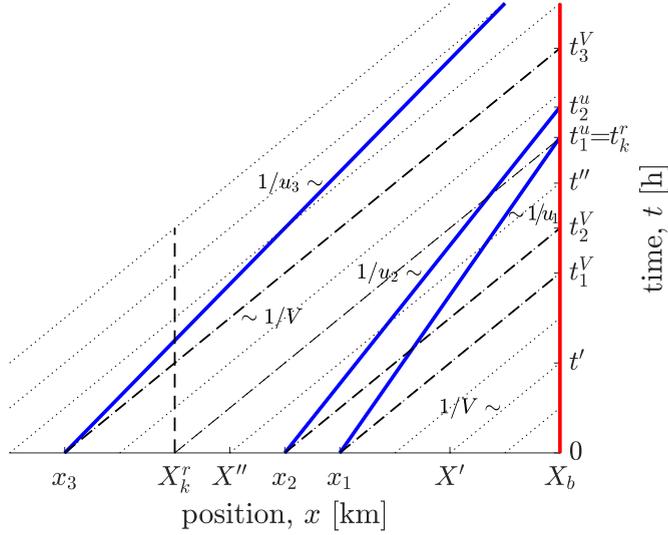


Figure 4.12: Illustration of the queueing model with  $t_0 = 0$ . The dotted lines represent free flow propagation. Platoon trajectories are shown in blue. As shown in the figure, at  $t = t'$ , inflow to the bottleneck is  $q_b^{\text{in}}(t') = V\rho(X')$ . At  $t = t''$ , inflow to the bottleneck is  $q_b^{\text{in}}(t'') = \tilde{q}_1^{\text{out}}(t'')$ , and inflows to the platoons  $q_1^{\text{in}}(t'') = \tilde{q}_2^{\text{out}}(t'')$ , and  $q_2^{\text{in}}(t'') = V\rho(t'')$ . Ramp  $k$  will affect  $q_2^{\text{in}}(t)$  for  $\frac{X_b - X_k^r}{V} < t \leq t_2^u$ ,  $q_3^{\text{in}}(t)$  while  $x_3^u(t) \geq X_k^r$  and  $t < t_3^u$ , and  $q_b^{\text{in}}(t)$  for the rest of time.

for each platoon  $\xi$ , which yields

$$\frac{dn_\xi(t(\tau_\xi))}{d\tau_\xi} = q_\xi^{\text{in}}(t(\tau_\xi)) - q_\xi^{\text{out}}(t(\tau_\xi)), \quad t_\xi^V(t_0) \leq \tau_\xi \leq t_\xi^u.$$

An illustration of the derivation of the coordinate-transformed model is given in Figure 4.12.

Defining the transformed-time platoon queues, and inflows and outflows from them,

$$\begin{aligned} \tilde{n}_\xi(\tau_\xi) &= n_\xi(t(\tau_\xi)), \\ \tilde{q}_\xi^{\text{in}}(\tau_\xi) &= q_\xi^{\text{in}}(t(\tau_\xi)), \\ \tilde{q}_\xi^{\text{out}}(\tau_\xi) &= q_\xi^{\text{out}}(t(\tau_\xi)), \end{aligned}$$

we may write

$$\dot{\tilde{n}}_\xi(t) = \tilde{q}_\xi^{\text{in}}(t) - \tilde{q}_\xi^{\text{out}}(t), \quad t \leq t_\xi^u$$

for each  $\xi = 1, \dots, \Xi$ . The inflow to the queue at the stationary bottleneck and at platoons can now be simplified to

$$q_b^{V \setminus r}(t) = \begin{cases} \tilde{q}_1^{\text{out}}(t), & t < t_1^u, \\ \tilde{q}_\xi^{\text{out}}(t), & t_{\xi-1}^u \leq t < t_\xi^u, \xi \in \{2, \dots, \Xi\}, \end{cases}$$

$$\tilde{q}_\xi^{\text{in} \setminus r}(t) = \tilde{q}_{\xi+1}^{\text{out}}(t),$$

and the outflow from each platoon becomes

$$\tilde{q}_\xi^{\text{out}}(t) = \begin{cases} \min \left\{ \tilde{q}_\xi^{\text{in}}(t), \tilde{q}_\xi^{\text{cap}}(t) \right\} & \tilde{n}_\xi(t) = 0, \\ \tilde{q}_\xi^{\text{cap}}(t), & \tilde{n}_\xi(t) > 0. \end{cases}$$

On- and off-ramps  $k \in K$  are included in the same way as when dealing in absolute time, by adding their influence to  $q_b^V(t)$  and  $\tilde{q}_\xi^{\text{in}}(t)$ . Denoting by  $\tilde{q}_k^r(t)$  the transformed-time ramp flow,

$$\tilde{q}_k^r(t) = q_k^r \left( t - \frac{X_b - X_k^r}{V} \right),$$

we may write

$$q_b^V(t) = q_b^{V \setminus r}(t) + \sum_{k \in \tilde{K}_b(t)} \tilde{q}_k^r(t),$$

$$\tilde{K}_b(t) = \left\{ k \in K \mid t \notin \left[ t_0 + \frac{X_k^r - x_\xi(t_0)}{u_\xi} + \frac{X_b - X_k^r}{V}, t_\xi^u \right], \xi \in \{1, \dots, \Xi\} \right\}$$

and for the inflow to the queue at platoons,

$$\tilde{q}_\xi^{\text{in}}(t) = \tilde{q}_\xi^{\text{in} \setminus r}(t) + \sum_{k \in \tilde{K}_\xi(t)} \tilde{q}_k^r(t),$$

$$\tilde{K}_\xi(t) = \{k \in K \mid \tilde{x}_{\xi+1}(t) \leq X_k^r\},$$

where  $\tilde{x}_\xi(t)$  is the position of platoon  $\xi$  at transformed-time  $t$ ,

$$\tilde{x}_\xi(t) = \frac{Vx_\xi(t_0) - u_\xi X_b}{V - u_\xi} + \frac{Vu_\xi}{V - u_\xi}(t - t_0).$$

Again, when a platoon passes an off-ramp, its queue  $\tilde{n}_\xi$  is reduced,

$$\tilde{n}_\xi(t+) = \tilde{n}_\xi(t) - \tilde{n}_\xi^{r,k}(t), \quad \tilde{x}_\xi(t) = X_k^r,$$

as a part of the vehicles leave the road and the queue via the off-ramp.

Initializing the model follows similar rules as in the case when absolute time was used. In order to initialize the model at absolute time  $t_0$ , we need initial conditions

$n_b(t_0)$  and  $\tilde{n}_\xi(t_\xi^V(t_0))$ ,  $\xi = 1, \dots, \Xi$ , in coordinate-transformed time. We need to know  $\tilde{q}_\xi^{\text{out}}(t)$  for coordinate-transformed time  $t_{\xi-1}^V(t_0) \leq t < t_1^V(t_0)$ , where we define  $t_0^V(t_0) = t_0$ , and  $\tilde{q}_k^r(t)$  for coordinate-transformed time  $t_{\xi_k}^V(t_0) \leq t < t_0 + \frac{X_b - X_k^r}{V}$ , where  $\xi_k$  is the number of the first platoon downstream of the ramp  $k$  at absolute time  $t_0$ ,  $x_{\xi+1}(t_0) < X_k^r \leq x_\xi(t_0)$ , or  $\xi_k = 0$  if  $X_k^r > x_1(t_0)$ . If the model is being initialization based on the traffic density profile at time  $t_0$ ,  $\rho(x, t_0)$ , we set  $\tilde{n}_b(t_0) = 0$ ,  $\tilde{n}_\xi(t_\xi^V(t_0)) = 0$ ,  $\xi = 1, \dots, \Xi$ , and

$$\tilde{q}_\xi^{\text{out}}(t) = V\rho(X_b - V(t - t_0), t_0), \quad \xi = 1, \dots, \Xi,$$

with the ramp flows  $\tilde{q}_k^r(t) = 0$ ,  $t \leq t_0 + \frac{X_b - X_k^r}{V}$ .

### 4.4.3 Validation

Finally, we validate the proposed tandem queueing model with moving bottlenecks against microscopic traffic simulation done in SUMO, and macroscopic simulation in multi-class CTM, using an appropriate example scenario. We study a 4 km stretch of road with a lane drop bottleneck at the 3.75 km mark. 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 time  $t = 144$  s, both platoons are slowed down and commanded to restrict the overtaking flow by taking two lanes. This causes the congestion at the bottleneck to be dissipated quicker, and the platoons go back to taking one lane at time  $t = 216$  s, allowing the congestion that built up behind them to dissipate.

Traffic density profiles in multi-class CTM and in SUMO are shown colour-coded in Figure 4.13. The bottleneck is indicated by the vertical dashed red line, and times  $t = 144$  s and  $t = 216$  s by horizontal dashed red lines. In case of the SUMO simulation, the traffic density is reconstructed according to vehicle trajectories, and resembles the traffic density expected from the multi-class CTM. Finally, in Figure 4.14 we show the comparison between the simulated queue length profiles, and the queue length prediction made using the proposed queueing prediction model. The prediction is made at time  $t = 144$  s, using currently available traffic density data from the multi-class CTM simulation. We can see that the queue lengths exhibit similar behaviour. 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 first platoon is not able to fully discharge the congestion that built up behind it, so this congestion 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.

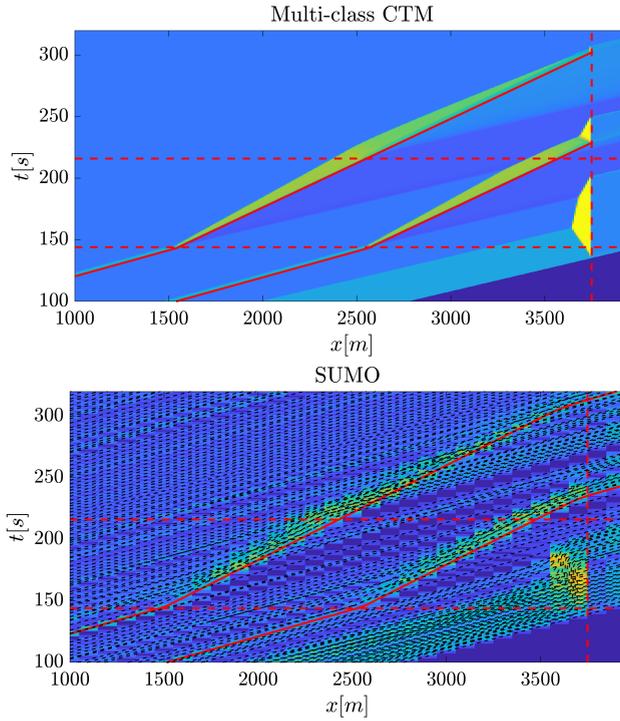


Figure 4.13: Traffic density profiles comparison. Warmer colours represent denser traffic, solid red lines are the trajectories of platoons, and dashed black lines are the trajectories of individual vehicles in SUMO.

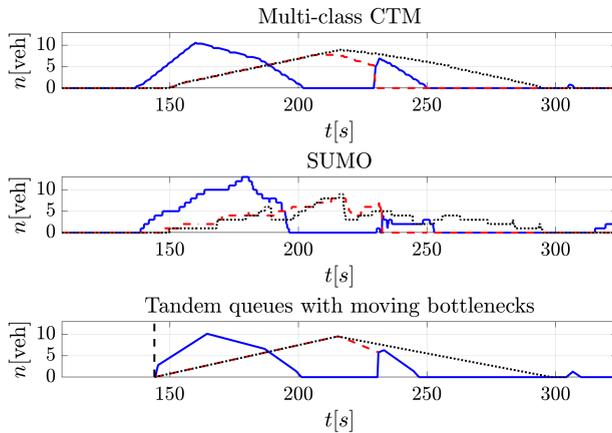


Figure 4.14: Queue lengths comparison. Queue at the stationary bottleneck is shown in blue, queue at the first platoon in dashed red, and queue at the second platoon in dotted black.

## 4.5 Front tracking transition system model

Due to the complexities of dealing with PDE traffic models, it can be difficult to find analytic solutions to the Cauchy problem of the LWR model for all but the simplest flux functions. Numerical approaches, such as the Godunov scheme, have been applied, resulting in approximation errors, and an unfavourable trade-off between the spatial resolution and the number of states. Efficient Lax-Hopf based grid-free solution algorithms have been developed to deal with this problem [198, 162], yielding an exact solution, but while they can model the influence of moving bottlenecks [161], they do not provide a straightforward way of implementing capacity drop or other similar phenomena. Therefore, in this section, we propose a cell-free modelling approach that can handle stop-and-go waves, models the influence of moving bottlenecks, can incorporate Lagrangian measurements, and has a structure that yields itself to on-line model learning, the front tracking transition system model [31, 29]. The model represents a formalization of the front tracking numerical method for solving conservation laws, given as a transition system. We choose to model the capacity drop due to stop-and-go waves by limiting the stop-and-go wave propagation speed, which is empirically observed to be approximately constant [44]. To that end, we define a new type of weak solution, the wave-speed-bounded solution, which breaks the entropy condition at the stop-and-go wave boundary.

### 4.5.1 Front tracking solutions of the LWR model with zones of different flux functions

We are interested in finding the entropy solution to

$$\partial_t \rho(x, t) + \partial_x Q(\rho(x, t), x, t) = 0, \quad (4.37)$$

with piecewise constant initial traffic density  $\rho(x, 0)$ ,

$$\rho(x, 0) = \begin{cases} \rho_1, & x < X_1, \\ \vdots & \\ \rho_i, & X_{i-1} < x < X_i, \\ \vdots & \\ \rho_{N+1}, & x > X_N, \end{cases} \quad (4.38)$$

and the flux function  $Q(\rho(x, t), x, t)$  given by

$$Q(\rho(x, t), x, t) = \begin{cases} Q_1(\rho(x, t)), & x < X_1^Q(t), \\ \vdots & \\ Q_j(\rho(x, t)), & X_{j-1}^Q(t) < x < X_j^Q(t), \\ \vdots & \\ Q_{N^Q+1}(\rho(x, t)), & x > X_{N^Q}^Q(t). \end{cases} \quad (4.39)$$

Here  $X_j^Q(t)$  are piecewise linear and continuous functions,  $\dot{X}_j^Q(t) = \Lambda_j^Q(t)$ , with piecewise constant  $\Lambda_j^Q(t)$ , and  $X_j^Q(t) \leq X_{j+1}^Q(t)$ . Effectively, the space is divided

into zones  $[X_j^Q(t), X_{j+1}^Q(t)]$  where the dynamics of the traffic are described by different flux functions  $Q_j(\rho(x, t))$ , e.g. zones where the road has different number of lanes. Each flux function  $Q_j(\rho)$ ,  $j = 1, \dots, N^Q$ , is a piecewise linear continuous function of the form

$$Q_j(\rho) = \begin{cases} V_{j,1}\rho, & 0 \leq \rho \leq \sigma_{j,1}, \\ Q_j(\sigma_{j,1}) + V_{j,2}(\rho - \sigma_{j,1}), & \sigma_{j,1} < \rho \leq \sigma_{j,2}, \\ \vdots & \\ Q_j(\sigma_{j,i-1}) + V_{j,i}(\rho - \sigma_{j,i-1}), & \sigma_{j,i-1} < \rho \leq \sigma_{j,i}, \\ \vdots & \\ Q_j(\sigma_{j,m_j}) + V_{j,m_j}(\rho - \sigma_{j,m_j}), & \sigma_{j,m_j} < \rho \leq \sigma_{j,m_j+1}, \\ 0, & \rho > \sigma_{j,m_j+1}, \end{cases} \quad (4.40)$$

with  $Q_j(\sigma_{j,m_j}) + V_{j,m_j}(\sigma_{j,m_j+1} - \sigma_{j,m_j}) = 0$ . We denote values of  $Q_j$  at the breakpoints  $q_{j,i}^\sigma = Q_j(\sigma_{j,i})$ , the set of breakpoints

$$\Sigma_{Q_j} = \{\sigma_{j,1}, \dots, \sigma_{j,m_j+1}\},$$

and the set of slopes between the breakpoints as

$$\mathbf{V}_{Q_j} = \{V_{j,1}, \dots, V_{j,m_j}\}.$$

The minimum and maximum slopes are  $V_{Q_j}^{\min} = \min \{\mathbf{V}_{Q_j}\}$  and  $V_{Q_j}^{\max} = \max \{\mathbf{V}_{Q_j}\}$ , respectively. Note that the maximum traffic speed  $v_{Q_j}^{\max}$  need not be the same as the maximum slope,  $v_{Q_j}^{\max} = \max_i q_{j,i}^\sigma / \sigma_{j,i} \leq V_{Q_j}^{\max}$ , although in practice, most flux functions are concave for  $\rho \in [0, \sigma_{Q_j}^{\max}]$ , where  $\sigma_{Q_j}^{\max} = \arg \max_{\rho} Q_j(\rho)$ . The set of all functions  $Q_j$  that satisfy these requirements is denoted  $\mathcal{Q}$ .

Front tracking has long been used as a method for finding entropy solutions to PDE traffic models [215, 161, 206]. It corresponds to solving a sequence of Riemann problems, to find the entropy solution for piecewise-constant initial conditions  $\rho(x, 0)$ , assuming piecewise-linear flux functions. The entropy solution to the LWR model (4.37), with flux functions (4.39), and initial conditions (4.38) is of the form

$$\rho(x, t) = \begin{cases} \rho'_1, & x < X'_1 + \lambda_1 t, \\ \vdots & \\ \rho'_i, & X'_{i-1} + \lambda_{i-1} t < x < X'_i + \lambda_i t, \\ \vdots & \\ \rho'_{N'+1}, & x > X'_{N'} + \lambda_{N'} t, \end{cases} \quad (4.41)$$

with  $\lambda_{i-1} \leq \lambda_i$  wherever  $X'_{i-1} = X'_i$ . Here  $\lambda_i$ ,  $i = 1, \dots, N'$  are the transition speeds, defined by the Rankine-Hugoniot condition

$$\lambda_i = \frac{Q(\rho'_{i+1}) - Q(\rho'_i)}{\rho'_{i+1} - \rho'_i}.$$

The solution consists of zones of constant density separated by fronts  $X'_i + \lambda_i t$  where we have a discontinuity in the density. This solution holds for  $t \in [0, \tau]$ , where  $\tau$  is the minimum time when two fronts collide,  $X'_{i-1} + \lambda_{i-1}\tau = X'_i + \lambda_i\tau$ , with  $\lambda_{i-1} > \lambda_i$ . To get the solution after that time, we solve a new composite Riemann problem for initial conditions  $\rho(\tau, x)$ , and by iterating this step, we can obtain an exact entropy solution  $\rho(x, t)$  to the initial value problem (3.5), (4.38) for any  $t$ . Due to Corollary 2.8 from [206], the front tracking method yields *exact* entropy solutions in case when the flux function is continuous and piecewise-linear and initial conditions piecewise constant, which is the case we consider here.

Since  $\rho(x, 0)$  is piecewise constant, we have  $\partial_x Q(\rho(x, t), x, t) = 0$  everywhere except at discontinuities of  $\rho(x, t)$ , and at discontinuities of  $Q(\rho(x, t), x, t)$ . Therefore, the process of finding (4.41) can be decomposed into finding the solutions to the Riemann problems at each discontinuity. A Riemann problem, is an initial value problem of (4.37) with

$$\rho(x, 0) = \begin{cases} \rho_-, & x < 0, \\ \rho_+, & x > 0, \end{cases} \quad (4.42)$$

and

$$Q(\rho, x, t) = \begin{cases} Q_-(\rho), & x < \Lambda t, \\ Q_+(\rho), & x > \Lambda t, \end{cases} \quad (4.43)$$

where  $\Lambda$  is given as a parameter of the problem. Note that the space coordinate was shifted so that the discontinuity is at  $x = 0$ .

If the discontinuity is only in  $\rho(x, t)$ , and  $Q_-(\rho) = Q_+(\rho)$ , this Riemann problem corresponds to the basic form of finding the entropy solution to the LWR model, which is well-known in the literature. In the following, we first discuss this case, and then extend the results by defining a new type of weak solution, the wave-speed-bounded solution, for which we impose bounds on the wavefronts that originate from the discontinuity. Note that this solution can violate the entropy condition, but it allows us to model some additional traffic phenomena. Finally, we discuss the solutions at the boundary between two different flux functions, when  $Q_-(\rho) \neq Q_+(\rho)$ .

### Entropy solution

When finding the entropy solution to Riemann problems (4.37), (4.42), (4.43), we need to calculate *lower convex envelope* or *upper concave envelope* of the flux function if  $\rho_- < \rho_+$  or  $\rho_- > \rho_+$ , respectively. We define these envelopes

$$\begin{aligned} \rho_- \tilde{Q}_Q^{\rho_+}(\rho) &= \begin{cases} \rho_- \check{Q}_Q^{\rho_+}(\rho), & \rho_- < \rho_+, \\ \rho_- \hat{Q}_Q^{\rho_+}(\rho), & \rho_- > \rho_+, \end{cases} \\ \rho_- \check{Q}_Q^{\rho_+}(\rho) &= \sup \left\{ q(\rho) : q(\rho) \leq Q(\rho), q \text{ convex}, \rho \in [\rho_-, \rho_+] \right\}, \end{aligned}$$

$${}^{\rho_-}\tilde{Q}_Q^{\rho_+}(\rho) = \inf \left\{ q(\rho) : q(\rho) \geq Q(\rho), q \text{ concave}, \rho \in [\rho_+, \rho_-] \right\},$$

on  $[\rho_{\min}, \rho_{\max}]$ ,  $\rho_{\min} = \min(\rho_-, \rho_+)$ ,  $\rho_{\max} = \max(\rho_-, \rho_+)$ . We follow this notation in further text, with  $\tilde{\phantom{x}}$  signifying  $\tilde{\phantom{x}}$  (lower convex envelope) if  $\rho_- < \rho_+$ , or  $\hat{\phantom{x}}$  (upper concave envelope) if  $\rho_- > \rho_+$ .

Note that  ${}^{\rho_-}\tilde{Q}_Q^{\rho_+}(\rho)$  also is a continuous piecewise-linear function on  $[\rho_{\min}, \rho_{\max}]$  and it can be defined in a similar way as (4.40),

$$\tilde{Q}(\rho) = \begin{cases} Q(\rho_{\min}) + \tilde{V}_0(\rho - \rho_{\min}), & \rho_{\min} \leq \rho \leq \tilde{\sigma}_1, \\ Q(\tilde{\sigma}_1) + \tilde{V}_1(\rho - \tilde{\sigma}_1), & \tilde{\sigma}_1 \leq \rho \leq \tilde{\sigma}_2, \\ \vdots & \\ Q(\tilde{\sigma}_{\tilde{m}}) + \tilde{V}_{\tilde{m}}(\rho - \tilde{\sigma}_{\tilde{m}}), & \tilde{\sigma}_{\tilde{m}} \leq \rho \leq \rho_{\max}, \end{cases}$$

omitting superscript  $\rho_-$  and  $\rho_+$  and subscript  $Q$  for better readability, and with  $\tilde{Q}(\rho_{\max}) = Q(\rho_{\max})$ . We write the column vector of slopes of such function  ${}^{\rho_-}\tilde{V}_Q^{\rho_+}$ , ordered from  $\tilde{V}_0$  to  $\tilde{V}_{\tilde{m}}$  for  $\rho_- < \rho_+$  or from  $\tilde{V}_{\tilde{m}}$  to  $\tilde{V}_0$  for  $\rho_- > \rho_+$ . All breakpoints of  $\tilde{Q}$ ,  $\tilde{\sigma}_i$ , are also breakpoints of  $Q$ ,  $\sigma_i$ , on  $[\rho_{\min}, \rho_{\max}]$ , and they can be determined using efficient convex hull algorithms. Finally, we denote the sorted (ascending if  $\rho_- < \rho_+$  and descending if  $\rho_- > \rho_+$ ) column vector of elements of  ${}^{\rho_-}\tilde{\Sigma}_Q^{\rho_+}$ , including  $\rho_-$  and  $\rho_+$ , as  ${}^{\rho_-}\tilde{\Sigma}_Q^{\rho_+}$ , and its length as  ${}^{\rho_-}\tilde{m}_Q^{\rho_+}$ . Same as with the envelopes  ${}^{\rho_-}\tilde{Q}_Q^{\rho_+}(\rho)$ ,  ${}^{\rho_-}\tilde{\Sigma}_Q^{\rho_+}$  will consist of breakpoints of the lower convex or upper concave envelope, depending on whether  $\rho_-$  or  $\rho_+$  is larger.

The solution to the Riemann problem is then given by

$$\rho(x, t) = \begin{cases} \rho_-, & x < e_1^\top \tilde{V}t, \\ e_1^\top \tilde{\Sigma}, & e_1^\top \tilde{V}t < x < e_2^\top \tilde{V}t, \\ \vdots & \\ e_{\tilde{m}}^\top \tilde{\Sigma}, & e_{\tilde{m}}^\top \tilde{V}t < x < e_{\tilde{m}-1}^\top \tilde{V}t, \\ \rho_+, & x > e_{\tilde{m}-1}^\top \tilde{V}t, \end{cases} \quad (4.44)$$

again omitting superscript and subscript symbols, and with  $e_i$  as the standard basis vector of appropriate dimensions with 1 as the  $i$ -th element and 0 elsewhere.

### Wave-speed-bounded solution

Depending on the shape of the flux function, the entropy solution may contain solutions that do not reflect the real-world traffic behaviour and phenomena. One such phenomenon are stop-and-go waves, caused by the difference between deceleration when entering congestion, and acceleration when leaving it. We can model this effect by enforcing bounds on wavefront speed, defining a new type of weak solution to the Riemann problem.

Let  $[\check{W}_-, \check{W}_+]$  and  $[\hat{W}_-, \hat{W}_+]$  be the admissible ranges of wavefront speeds for compression (if  $\rho_- < \rho_+$ ) and rarefaction (if  $\rho_- > \rho_+$ ), respectively. In case  $\check{W}_- \leq \min\{\mathbf{V}_Q\} \geq \hat{W}_-$  and  $\check{W}_+ \geq \max\{\mathbf{V}_Q\} \leq \hat{W}_+$ , the entropy solution will never

violate the wave speed bounds, and it thus coincides with the wave-speed bounded solution. Otherwise, we define  $\tilde{V}_- = \min^{\rho_-} \tilde{V}_Q^{\rho_+}$  and  $\tilde{V}_+ = \max^{\rho_-} \tilde{V}_Q^{\rho_+}$ , and if  $\tilde{V}_- < \tilde{W}_-$  or  $\tilde{V}_+ > \tilde{W}_+$ , we need to look for a solution that fulfils the wavefront speed constraints, while minimally violating the entropy condition. The wave-speed-bounded solution is again given by (4.44), with

$$\rho_- \tilde{\Sigma}_{Q,W}^{\rho_+} = \begin{cases} \left[ \begin{array}{c} \rho_- \\ \tilde{\rho}_-^W \\ \tilde{\Sigma}_Q^{\tilde{\rho}_+^W} \\ \rho_+ \end{array} \right]^T, & \tilde{V}_- < \tilde{W}_-, \tilde{V}_+ > \tilde{W}_+ \\ \left[ \begin{array}{c} \rho_- \\ \tilde{\rho}_-^W \\ \tilde{\Sigma}_Q^{\rho_+} \end{array} \right]^T, & \tilde{V}_- < \tilde{W}_-, \tilde{V}_+ \leq \tilde{W}_+ \\ \left[ \begin{array}{c} \rho_- \\ \tilde{\Sigma}_Q^{\tilde{\rho}_+^W} \\ \rho_+ \end{array} \right]^T, & \tilde{V}_- \geq \tilde{W}_-, \tilde{V}_+ > \tilde{W}_+ \\ \left[ \begin{array}{c} \rho_- \\ \tilde{\Sigma}_Q^{\rho_+} \end{array} \right]^T, & \tilde{V}_- \geq \tilde{W}_-, \tilde{V}_+ \leq \tilde{W}_+ \end{cases}$$

$$\rho_- \tilde{V}_{Q,W}^{\rho_+} = \begin{cases} \left[ \begin{array}{c} \tilde{W}_- \\ \tilde{\rho}_-^W \\ \tilde{V}_Q^{\tilde{\rho}_+^W} \\ \tilde{W}_+ \end{array} \right]^T, & \tilde{V}_- < \tilde{W}_-, \tilde{V}_+ > \tilde{W}_+ \\ \left[ \begin{array}{c} \tilde{W}_- \\ \tilde{\rho}_-^W \\ \tilde{V}_Q^{\rho_+} \end{array} \right]^T, & \tilde{V}_- < \tilde{W}_-, \tilde{V}_+ \leq \tilde{W}_+ \\ \left[ \begin{array}{c} \rho_- \\ \tilde{V}_Q^{\tilde{\rho}_+^W} \\ \tilde{W}_+ \end{array} \right]^T, & \tilde{V}_- \geq \tilde{W}_-, \tilde{V}_+ > \tilde{W}_+ \\ \left[ \begin{array}{c} \rho_- \\ \tilde{V}_Q^{\rho_+} \end{array} \right]^T, & \tilde{V}_- \geq \tilde{W}_-, \tilde{V}_+ \leq \tilde{W}_+ \end{cases}$$

where we define the densities

$$\begin{aligned} \tilde{\rho}_-^W &= \max. \rho \quad \text{s.t. } \rho < \rho_-, \frac{dQ(\rho)}{d\rho} \geq \hat{W}_-, \frac{Q(\rho) - Q(\rho_-)}{\rho - \rho_-} = \tilde{W}_-, \\ \tilde{\rho}_+^W &= \min. \rho \quad \text{s.t. } \rho > \rho_+, \frac{dQ(\rho)}{d\rho} \leq \hat{W}_+, \frac{Q(\rho) - Q(\rho_+)}{\rho - \rho_+} = \tilde{W}_+, \\ \tilde{\rho}_-^W &= \min. \rho \quad \text{s.t. } \rho > \rho_-, \frac{dQ(\rho)}{d\rho} \geq \hat{W}_-, \frac{Q(\rho) - Q(\rho_-)}{\rho - \rho_-} = \tilde{W}_-, \\ \tilde{\rho}_+^W &= \max. \rho \quad \text{s.t. } \rho < \rho_+, \frac{dQ(\rho)}{d\rho} \leq \hat{W}_+, \frac{Q(\rho) - Q(\rho_+)}{\rho - \rho_+} = \tilde{W}_+. \end{aligned}$$

Note that the resulting solution may be non-monotone in density, since  $\tilde{\rho}_-^W$  and  $\tilde{\rho}_+^W$  might lie outside of  $[\min\{\rho_-, \rho_+\}, \max\{\rho_-, \rho_+\}]$ , or might be differently ordered than  $\rho_-$  and  $\rho_+$ . The entropy condition will only be violated at the discontinuities between  $\rho_-$  and  $\tilde{\rho}_-^W$ , and  $\rho_+$  and  $\tilde{\rho}_+^W$ . In order to model stop-and-go waves with capacity drop, it is enough to use  $\tilde{W}_-$ .

### Flux function boundary solution

Stationary and moving bottlenecks are modelled by using different flux functions in different regions, and we study the evolution of traffic conditions around a bottleneck by solving a Riemann problem (4.37), (4.42), (4.43), with  $Q_-(\rho) \neq Q_+(\rho)$ . In order to find the weak solution, the propagation speed of the boundary between

two flux functions  $\Lambda$  needs to be defined first. Then, we find the solution that maximizes the flow over the boundary between two regions, while satisfying the Rankine-Hugoniot condition across the boundary, and yielding entropy solutions inside both regions. The solution with initial conditions (4.42) and flux function (4.43) can thus be split into two parts, consisting of Riemann problems between  $\rho_-$  and  $\rho'_-$ , and between  $\rho'_+$  and  $\rho_+$ . The solutions to these problems both need to be constrained to only have discontinuities inside the zones of  $Q_-$  ( $x < \Lambda t$ ) and  $Q_+$  ( $x > \Lambda t$ ), respectively. The Rankine-Hugoniot condition at the boundary between the two flux functions

$$Q_+(\rho'_+) - Q_-(\rho'_-) = \Lambda (\rho'_+ - \rho'_-),$$

can be rewritten in terms of the flow over the boundary  $\omega$ ,

$$\omega = Q_-(\rho'_-) - \Lambda \rho'_- = Q_+(\rho'_+) - \Lambda \rho'_+.$$

Boundary densities  $\rho'_-$  and  $\rho'_+$  both depend on  $\rho_-$ ,  $\rho_+$ ,  $Q_-$  and  $Q_+$ , and are given as optimizers of the optimization problem

$$\begin{aligned} & \underset{\rho'_-, \rho'_+}{\text{maximize}} && \omega \\ & \text{s.t.} && Q_+(\rho'_+) - Q_-(\rho'_-) = \Lambda(\rho'_+ - \rho'_-), \\ & && \rho_- \tilde{V}_{Q_-, W}^{\rho'_-} < \Lambda, \\ & && \rho'_+ \tilde{V}_{Q_+, W}^{\rho'_+} > \Lambda, \end{aligned} \tag{4.45}$$

so that the flow over the boundary  $\omega$  is maximized, under specified constraints.

For most simple flux functions used in practice, solving this maximization problem can be done explicitly. Since optimal  $\rho'_-$  and  $\rho'_+$  will always be such that either  $\rho'_- \in \Sigma_{Q_-} \cup \{\rho_-\}$  or  $\rho'_+ \in \Sigma_{Q_+} \cup \{\rho_+\}$ , the problem can be solved by forming a set of all possible pairings of  $(\rho'_-, \rho'_+)$  that satisfy the Rankine-Hugoniot condition, and then checking the second and third constraint for each of them, in order of descending boundary overtaking flow, so that the first pair to satisfy these constraints is the optimizer.

Another type of boundary that we consider is the one that arises when we force the density on one side to be equal to some externally defined value,  $\rho(x, t) = F_-$ ,  $x < \lambda t$  or  $\rho(x, t) = F_+$ ,  $x > \lambda t$ . We write  $F_{\pm} \neq \emptyset$  for those sides where the density is forced, and  $F_{\pm} = \emptyset$  where it is not forced. In this case, the dynamics of traffic on the forced side are ignored, and the forced traffic density instead acts as a boundary condition for the other side, and the solution is given as

$$\begin{aligned} & \underset{\rho'_+}{\text{minimize}} && |\rho'_+ - F_-| \\ & \text{s.t.} && Q_+(\rho'_+) - Q_+(F_-) = \lambda(\rho'_+ - F_-), \\ & && \rho'_+ \tilde{V}_{Q_+, W}^{\rho'_+} > \Lambda, \end{aligned}$$

if  $F_- \neq \emptyset$  and  $F_+ = \emptyset$ , i.e. the density upstream is forced, or

$$\begin{aligned} & \underset{\rho'_-}{\text{minimize}} && |\rho'_- - F_+| \\ & \text{s.t.} && Q_-(F_+) - Q_-(\rho'_-) = \lambda(F_+ - \rho'_-), \\ & && \rho_- \tilde{V}_{Q_-,W}^{\rho'_-} < \Lambda, \end{aligned}$$

if  $F_- = \emptyset$  and  $F_+ \neq \emptyset$  and the density downstream is forced. Alternatively, we may define special rules for handling the behaviour at the boundary between two flux functions, e.g. when a moving bottleneck moves from a zone of one capacity to a zone of different capacity.

We may also model other traffic phenomena by introducing additional constraints in (4.45). For example, capacity drop can be captured through adding constraint

$$Q_-(\rho'_-) \leq Q_+(\sigma_+^{\Lambda-\beta}) + (\Lambda - \beta)(\rho'_- - \sigma_+^{\Lambda-\beta}), \quad (4.46)$$

where  $\beta \geq 0$  is the capacity drop slope parameter, and  $\sigma_+^{\Lambda-\beta}$  is the point at which a line with slope  $\Lambda - \beta$  is tangent to flux function  $Q_+$ , given as

$$\sigma_+^{\Lambda-\beta} = \arg \max_{\rho \in \Sigma_{Q_+}} Q_+(\rho) + (\Lambda - \beta)\rho.$$

Setting  $\beta = 0$  corresponds to the case with no capacity drop. Note that if this constraint is imposed on (4.45), the optimal  $\rho'_-$  may also be given at the boundary of the new constraint (4.46), with equality instead of inequality. Parameter  $\beta$  has no physical meaning, and the severity of capacity drop will depend on the shapes of  $Q_-$  and  $Q_+$ .

### Solution example

We will demonstrate the procedure of finding the front tracking solution on an example with initial conditions  $\rho(x, 0) = \rho_-$ ,  $x < 0$  and  $\rho(x, 0) = 0$ ,  $x > 0$ , and flux function

$$Q(\rho, t, x) = \begin{cases} Q_-(\rho), & x < 0 \vee t > t', \\ Q_+(\rho), & x > 0 \wedge t < t', \end{cases}$$

where  $Q_{\pm}$  are parametrized by  $\Sigma_- = [\sigma_1^- \ \sigma_2^-]^\top$ ,  $\Sigma_+ = (\Sigma_-)/2$ ,  $V_{\pm} = [V_1 \ V_2]^\top$ , and the only restricting wave speed bound is  $\tilde{W}_- < V_2$ . This setup corresponds to the situation at a bottleneck removed at time  $t = t'$ . The numerical values of all parameters are given in Table 4.1.

The solution is shown in Figure 4.15. For  $0 \leq t \leq t'$ , the solution consists of the flux function boundary solution,

$$\rho(x, t) = \begin{cases} \rho_-, & x < \lambda_1 t, \\ \rho^c, & \lambda_1 t < x < 0, \\ \sigma_1^+, & 0 < x < V_1 t, \\ 0, & V_1 t < x, \end{cases}$$

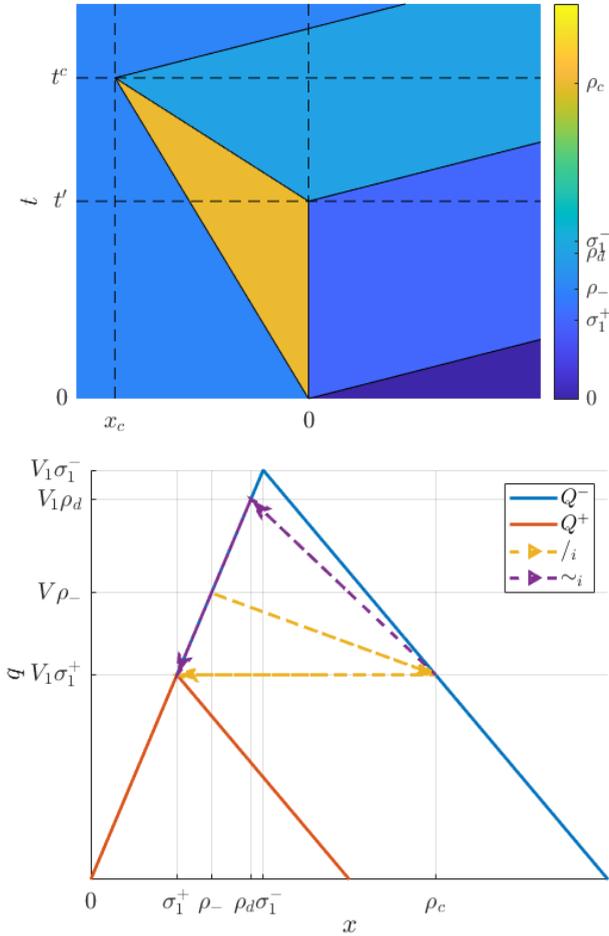


Figure 4.15: Example solution shown as a traffic density profile (top, traffic density  $\rho(x, t)$  is colour-coded, with warmer colours representing higher density), and on the flux functions (bottom).

$V_{\pm}$	$\Sigma_-$	$\hat{W}_-$	$\rho_-$	$\rho^c$	$\rho^d$	$\lambda_1$
$\begin{bmatrix} 1 \\ -0.5 \end{bmatrix}$	$\begin{bmatrix} 0.2 \\ 0.6 \end{bmatrix}$	-0.4	0.14	0.4	0.1857	-0.1538

Table 4.1: Parameters of the example solution.

with  $\lambda_1 = (Q_-(\rho^c) - Q_-(\rho_-))/(\rho^c - \rho_-)$ . Note that the congestion of density  $\rho^c$  starts accumulating at the bottleneck,  $Q_+(\sigma_1^+) - Q_-(\rho^c) = \sigma_1^+ - \rho^c$  at  $t = 0$ . Next, at time  $t'$ , the bottleneck is removed,  $Q_-$  becomes the global flux function,  $Q(\rho(x, t), x, t) = Q_-(\rho(x, t))$ ,  $t > t'$ , and the solution is

$$\rho(x, t) = \begin{cases} \rho_-, & x < \lambda_1 t, \\ \rho^c, & \lambda_1 t < x < \hat{W}_-(t - t'), \\ \rho^d, & \hat{W}_-(t - t') < x < V_1(t - t'), \\ \sigma_1^+, & V_1(t - t') < x < V_1 t, \\ 0, & V_1 t < x, \end{cases}$$

Note that since we have  $\rho^c \tilde{V}_{Q_-}^{\sigma_1^+} < \hat{W}_-$ , the wave-speed-bounded solution will differ from the entropy solution. Finally, the congestion is dissipated and the solution for  $t > t^c$  is

$$\rho(x, t) = \begin{cases} \rho_-, & x < x_c + V_1(t - t^c), \\ \rho^d, & x_c + V_1(t - t^c) < x < V_1(t - t'), \\ \rho^d, & V_1(t - t') < x < V_1 t, \\ 0, & V_1 t < x. \end{cases}$$

#### 4.5.2 Model formulation

The described procedure, with continuously changing solution between two composite Riemann problem solving instances and jumps resulting from them, lends itself to a transition system formulation. We follow the transition system formulation from the hybrid systems literature and given in [216]. The evolution of the front tracking solution to the scalar conservation is represented by the execution of the transition system given by the quadruple  $\mathcal{F} = (\mathcal{X}, \mathcal{X}_0, U, \rightarrow)$ . We call this transition system the front tracking transition system model (FTTSM), and will define it in this section.

#### States and initial states

The set of states  $\mathcal{X} = (N, t, \bar{z}, \bar{\rho}, \bar{Q}, W, \mathcal{R})$  is composed of:

- Number of active fronts:  $N \in \mathbb{N}$ ,  $N \leq N^{\max}$
- Time:  $t \in \mathbb{R}$
- Front positions:  $\bar{z} \in \mathbb{R}^{N^{\max}}$ ,  $z_i \leq z_{i+1}$  for  $i = 1, \dots, N$
- Traffic densities:  $\bar{\rho} \in \mathbb{R}_{\geq 0}^{N^{\max}+1}$
- Flux functions:  $\bar{Q} \in \mathcal{Q}^{N^{\max}+1}$ , where  $\mathcal{Q}$  is a set of flux functions
- Wavefront speed boundaries:  $W \in \mathbb{R}^4$ ,

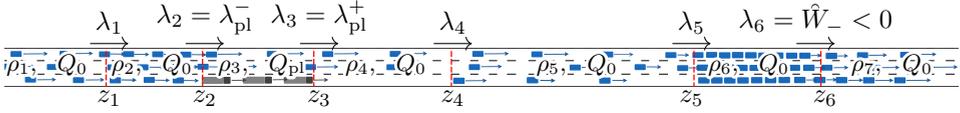


Figure 4.16: An illustration of the state of FTTSM. Flux function  $Q_0$  describes the behaviour of the background traffic, while flux function  $Q_{pl}$  describes the traffic flowing past a truck platoon,  $\xi_{pl} > \xi_0$ . The downstream end of a stop-and-go wave propagates upstream at speed defined by wavefront speed boundary  $\hat{W}_-$ .

- Set of rules:  $\mathcal{R}$ .

The four wavefront speed boundaries are written together as a quadruple,  $W = (\check{W}_-, \check{W}_+, \hat{W}_-, \hat{W}_+)$ . Each flux function can be described by a quintuple  $q = (V, \Sigma, \Lambda^\pm, \xi, F)$ , consisting of:

- Slopes and breakpoints:  $(V, \Sigma) \in \mathcal{L}$
- Boundary speeds:  $\Lambda^\pm \in \mathbb{R}^2$
- Identifier:  $\xi \in \mathbb{Z}$
- Forced density:  $F \in \mathbb{R} \cup \{\emptyset\}$ ,

where the set of allowed combinations of slopes and breakpoints is

$$\mathcal{L} = \{V \in \mathbb{R}^m, \Sigma \in \mathbb{R}_{>0}^{m+1} \mid \sigma_1 < \dots < \sigma_{m+1}, \dots, q_{Q,i}^\sigma \geq 0, i = 1, \dots, m, q_{Q,m+1}^\sigma = 0\},$$

$$q_{Q,i}^\sigma = q_{Q,i-1}^\sigma + V_i(\sigma_i - \sigma_{i-1}), i = 2, \dots, m+1, q_{Q,1}^\sigma = V_1\sigma_1,$$
(4.47)

without any further constraints on the shape of the flux function. A rule  $r \in \mathcal{R}$  consists of two parts, the premise  $r^\cdot = (Q_-^\cdot, Q_+^\cdot), Q_-^\cdot \neq Q_+^\cdot$ , and the consequent  $r^\cdot = (Q_-^\cdot, \Lambda^\pm), r = (r^\cdot, r^\cdot)$ . We require the rules to be unequivocal, i.e. no two rules in  $\mathcal{R}$  can have the same premise,  $r_i^\cdot = r_j^\cdot \iff i = j$ . An illustrations of the states of FTTSM is shown in Figure 4.16.

We denote the traffic flow given traffic density  $\rho$  and thus defined flux function  $q$  as  $q(\rho)$  and calculate it as (4.40), with  $V_i$  and  $\sigma_i$  given in  $V$  and  $\Sigma$ , respectively. Boundary speeds  $\Lambda^\pm = (\lambda^-, \lambda^+)$  represent the propagation speed of the upstream and downstream boundary of the region of  $q$ . The identifier  $\xi$  serves to differentiate flux functions, and is unique for each different flux function. We also use it to define the precedence when determining the propagation speed of the boundary between regions with different flux functions. Namely, considering a boundary between flux function  $q_i \neq q_{i+1}$ , the propagation speed is given by

$$\lambda_i = \begin{cases} \lambda_i^+, & \xi_i > \xi_{i+1}, \\ \lambda_{i+1}^-, & \xi_i < \xi_{i+1}. \end{cases}$$

We define the active fronts and densities

$$\begin{aligned} z &= [z_1 \quad \dots \quad z_N]^\top = [I_N \quad ; \quad \mathbb{0}_{n \times N^{\max} - N}] \bar{z} \\ \rho &= [\rho_1 \quad \dots \quad \rho_{N+1}]^\top = [I_{N+1} \quad ; \quad \mathbb{0}_{N+1 \times N^{\max} - N}] \bar{\rho}, \end{aligned}$$

to be all the states required to reconstruct the full traffic density profile  $\rho(x, t)$  at some time. Here  $\mathbb{0}_{m \times n}$  is a matrix of dimension  $m \times n$  with 0 for all elements. The maximum number of fronts  $N^{\max}$  can be taken large enough so that the number of active fronts never exceeds it, but effectively, the dimension of active states will vary as a part of the model dynamics. Furthermore, only the active fronts and densities, along with flux functions  $Q = [Q_1 \quad \dots \quad Q_{N+1}]^\top = [I_{N+1} \quad ; \quad \mathbb{0}_{N+1 \times N^{\max} - N}] \bar{Q}$ , will influence the behaviour of the system, so when describing the transitions, we only define their updates, and the inactive states may take arbitrary values.

The set of initial states  $\mathcal{X}_0$  can be the same as the set of all states, but in that case, we may be forced to take some number of Riemann transitions, described in the following section, at  $t = 0$ . This can be counteracted by imposing additional conditions on the set of initial states,

$$\begin{aligned} &(z_{j+1} > z_j) \vee ((z_{j+1} = z_j) \wedge (\lambda_j \leq \lambda_{j+1})), \quad j = 1, \dots, N-1, \quad (\star-) \\ &(\rho_j \neq \rho_{j+1}, \rho_j \tilde{\Sigma}_{Q_j, W}^{\rho_{j+1}} = [\rho_j \quad \rho_{j+1}]^\top), \quad \text{if } Q_j = Q_{j+1}, \quad j = 1, \dots, N, \quad (\star\sim) \\ &(\rho_j = \rho'_-, \rho_{j+1} = \rho'_+), \quad \text{if } Q_j \neq Q_{j+1}, \quad j = 1, \dots, N, \quad (\star /) \\ &\rho_j = F_j, \quad \text{if } F_j \neq \emptyset, \quad j = 1, \dots, N+1, \quad (\star!) \\ &(Q_{j+1} = Q_{\cdot}, \Lambda_{j+1}^\pm = \Lambda_{\cdot}^\pm), \quad \text{if } (\exists r \in \mathcal{R}) (r = (Q_j, Q_{j+1})), \quad j = 1, \dots, N, \quad (\star\mathcal{R}) \end{aligned}$$

where  $\rho'_-(\rho_j, \rho_{j+1}, Q_j, Q_{j+1})$  and  $\rho'_+(\rho_j, \rho_{j+1}, Q_j, Q_{j+1})$  are given as the optimizers from the solution described in Section 4.5.1. These conditions define the *admissible set* of states, and are also used to define guards of the transitions, as will be described in the following subsection.

Given the current state  $\mathcal{X} \in \mathcal{X}$  of the transition system, the density function  $\rho(x, t)$  can be reconstructed based on  $z_1, \dots, z_N$  and  $\rho_1, \dots, \rho_{N+1}$ ,

$$\rho(t, x) = \begin{cases} \rho_1, & x < z_1, \\ \vdots & \\ \rho_i, & z_{i-1} < x < z_i, \\ \vdots & \\ \rho_{N+1}, & x > z_N. \end{cases}$$

Note that we use notation  $\rho(x, t)$  for the reconstructed function, and  $\rho = [\rho_1 \dots \rho_{N+1}]^\top$  for the vector of traffic densities.

### Inputs and transitions

In this subsection, we will describe the various transitions that model the evolution of the transition system. For each of the transitions, the states that do not change

will be omitted from the description. The transitions are listed in order of increasing priority, i.e., we first present the transitions that can only be taken if no other transition can be taken, and end with transitions which do not depend on the state of the system, only on exogenous inputs. We use notation  $\circ$  to denote any transition. Inputs to the system are considered in the transition system framework, and they describe the passage of time and exogenous transitions.

**Passage of time transition  $\tau$ :** The first transition we describe is the passage of time, which describes the propagation of wavefronts between their interactions, or until the externally provided goal time  $t_{\text{end}}$ . This transition is taken if the state  $X$  is in the guard set:

$$\mathcal{X} \in \mathcal{G}_\tau = \{\mathcal{X} \in \mathcal{X} | (\star-), (\star\sim), (\star/), (\star!), (\star\mathcal{R}) \quad \tau \in [0, \tau^*]\},$$

i.e., the state is admissible (conditions  $(\star)$  hold). Traffic densities  $\rho$ , number of active states  $n$  and flux functions  $Q$  do not change in these transitions, so those will be omitted from the description. Only the wavefront positions of active states ( $i = 1, \dots, n$ ) are changed. We define this transition by

$$\begin{aligned} (t, z) &\xrightarrow{\tau} (t', z') \\ t' &= t + \tau, \quad z' = z + \Lambda\tau \end{aligned}$$

where  $\Lambda = [\lambda_1 \quad \dots \quad \lambda_N]^\top$ , and the wave speeds  $\lambda_i$  are given as

$$\lambda_i = \begin{cases} \frac{Q_{i+1}(\rho_{i+1}) - Q_i(\rho_i)}{\rho_{i+1} - \rho_i}, & \xi_i = \xi_{i+1}, \\ \lambda_i^+, & \xi_i > \xi_{i+1}, \\ \lambda_{i+1}^-, & \xi_i < \xi_{i+1}. \end{cases}$$

The maximum time shift  $\tau^*$  is the minimum of the time for which condition  $(\star-)$  is first violated,

$$\tau_z^* = \min \left\{ \frac{z_{i+1} - z_i}{\lambda_i - \lambda_{i+1}} \mid z_{i+1} \geq z_i, \lambda_i > \lambda_{i+1}, i = 1, \dots, N-1 \right\}$$

and the time to specified goal time  $\tau_{\text{end}}^* = t_{\text{end}} - t$ ,  $\tau^* = \max\{0, \min\{\tau_z^*, \tau_{\text{end}}^*\}\}$ . Note that if  $t \geq t_{\text{end}}$ ,  $X' = X$ .

**Front interaction transition  $-_i$ :** A front interaction transition is taken when two fronts interact (collide) and the state is in guard set

$$\mathcal{X} \in \mathcal{G}_{-i} = \left\{ \mathcal{X} \in \mathcal{X} \mid \neg(\star-)_i, (\star-)_j, j < i, (\star\sim), (\star/), (\star!), (\star\mathcal{R}) \right\},$$

where by  $\neg(\star-)_i$  we denote that the  $j = i$ -th condition in  $(\star-)$  is violated. For all transitions  $\circ \in \{-, \sim, /, !, \mathcal{R}\}$ , we write  $X \in \mathcal{G}_\circ$  if  $X \in \mathcal{G}_{\circ_i}$  for any  $i$ . In this case the position of wavefronts becomes equal,  $z_i = z_{i+1}$  while their distance is decreasing,

$\lambda_i > \lambda_{i+1}$ . The front interaction transition corresponds to deactivating one state,

$$\begin{aligned} (N, z, \rho, Q) &\xrightarrow{-i} (N', z', \rho', Q') \\ N' &= N - 1, \\ z' &= [z_1 \dots z_i \mid z_{i+2} \dots z_N]^\top, \\ \rho' &= [\rho_1 \dots \rho_i \mid \rho_{i+2} \dots \rho_{N+1}]^\top, \\ Q' &= [Q_1 \dots Q_i \mid Q_{i+2} \dots Q_{N+1}]^\top. \end{aligned}$$

If  $Q_i \neq Q_{i+2}$ , this transition is likely to cause condition  $(\star /)$  to be violated, and thus be followed by transition  $/_i$ . The Total Variation of the traffic density is nonincreasing through the transition,

$$\text{T.V.}(\rho') - \text{T.V.}(\rho) = |\rho_{i+2} - \rho_i| - |\rho_{i+1} - \rho_i| - |\rho_{i+2} - \rho_{i+1}| \leq 0.$$

**Internal Riemann transition  $\sim_i$ :** This transition results from solutions to Riemann problem given in Section 4.5.1, and it is taken when the state is in guard set

$$\mathcal{X} \in \mathcal{G}_{\sim_i} = \left\{ \mathcal{X} \in \mathcal{X} \mid \neg(\star \sim)_i, Q_i = Q_{i+1}, (\star \sim)_j, j < i, (\star /), (\star !), (\star \mathcal{R}) \right\},$$

The transition can be described by

$$\begin{aligned} (N, z, \rho, Q) &\xrightarrow{\sim_i} (N', z', \rho', Q') \\ N' &= N + m - 2, \quad m = \rho_i \tilde{m}_{Q_i, W}^{\rho_{i+1}} \\ z' &= [z_1 \dots z_{i-1} \mid z_i \mathbb{1}_{m-1}^\top \mid z_{i+1} \dots z_N]^\top, \\ \rho' &= [\rho_1 \dots \rho_{i-1} \mid \rho_i \tilde{\Sigma}_{Q_i, W}^{\rho_{i+1}} \mid \rho_{i+2} \dots \rho_{N+1}]^\top, \\ Q' &= [Q_1 \dots Q_{i-1} \mid Q_i \mathbb{1}_m^\top \mid Q_{i+2} \dots Q_{N+1}]^\top. \end{aligned}$$

Depending on  $\rho_i$  and  $\rho_{i+1}$ , the number of active states can decrease (if  $\rho_i = \rho_{i+1}$ ), increase, or stay the same.

**Boundary Riemann transition  $/_i$ :** This transition can occur at interfaces between zones with different flux functions and reflects the flux function boundary solution from Section 4.5.1. It is taken when the state is in guard set

$$\mathcal{X} \in \mathcal{G}_{/_i} = \left\{ \mathcal{X} \in \mathcal{X} \mid \neg(\star /)_i, Q_i \neq Q_{i+1}, (\star /)_j, j < i, (\star !), (\star \mathcal{R}) \right\}.$$

The transition can be described by

$$\begin{aligned} (N, z, \rho, Q) &\xrightarrow{/_i} (N', z', \rho', Q') \\ N' &= N + m_- + m_+ - 2, \quad m_- = \rho_i \tilde{m}_{Q_i}^{\rho'_-}, \quad m_+ = \rho'_{i+1} \tilde{m}_{Q_{i+1}}^{\rho_{i+1}} \\ z' &= [z_1 \dots z_{i-1} \mid z_i \mathbb{1}_{m_- + m_+}^\top \mid z_{i+1} \dots z_N]^\top, \\ \rho' &= [\rho_1 \dots \rho_{i-1} \mid \rho_i \tilde{\Sigma}_{Q_i}^{\rho'_-} \mid \rho'_{i+1} \tilde{\Sigma}_{Q_{i+1}}^{\rho_{i+1}} \mid \rho_{i+2} \dots \rho_{N+1}]^\top, \\ Q' &= [Q_1 \dots Q_{i-1} \mid Q_i \mathbb{1}_{m_-}^\top \mid Q_{i+1} \mathbb{1}_{m_+}^\top \mid Q_{i+2} \dots Q_{N+1}]^\top, \end{aligned}$$

where densities  $\rho'_- = \rho'_-(\rho_i, \rho_{i+1}, Q_i, Q_{i+1})$  and  $\rho'_+ = \rho'_+(\rho_i, \rho_{i+1}, Q_i, Q_{i+1})$  are obtained by solving the optimization problem (4.45), with  $\rho_- = \rho_i$ ,  $\rho_+ = \rho_{i+1}$ ,  $Q_- = Q_i$ , and  $Q_+ = Q_{i+1}$ .

**State forcing !<sub>i</sub>:** This transition ensures that the density matches the forced traffic density where  $F_i \neq \emptyset$ ; is thus taken when the state is in guard set

$$\mathcal{X} \in \mathcal{G}_{!_i} = \left\{ \mathcal{X} \in \mathcal{X} \mid \neg(\star!)_i, (\star!)_j, j < i, (\star\mathcal{R}) \right\},$$

and defined by

$$\rho'_i \stackrel{!_i}{\mapsto} (\rho') = \begin{cases} \rho_i, & F_i = \emptyset, \\ F_i, & F_i \neq \emptyset. \end{cases}$$

**Rule-based transition  $\mathcal{R}_i$ :** We take a rule-based transition if the  $i$ -th condition  $(\star\mathcal{R})$  is violated, i.e. when the premise of some rule  $r \in \mathcal{R}$  is satisfied,  $(Q_i, Q_{i+1}) = r^\cdot$ , and the consequent is not met,  $Q_{i+1} \neq Q_{i+1}^\cdot \vee \Lambda_{i+1}^\pm \neq \Lambda_{i+1}^\pm$ , and the state is in guard set

$$\mathcal{X} \in \mathcal{G}_{\mathcal{R}_i} = \left\{ \mathcal{X} \in \mathcal{X} \mid \neg(\star\mathcal{R})_i, (\star\mathcal{R})_j, j < i \right\},$$

The transition can be described by

$$\begin{aligned} (N, z, \rho, Q) &\xrightarrow{\mathcal{R}(r)_i} (N', z', \rho', Q') \\ N' &= N + 1, \\ z' &= [ z_1 \quad \dots \quad z_i \quad | \quad z_i \quad | \quad z_{i+1} \quad \dots \quad z_N ]^\top, \\ \rho' &= [ \rho_1 \quad \dots \quad \rho_i \quad | \quad \rho_i \quad | \quad \rho_{i+1} \quad \dots \quad \rho_{N+1} ]^\top, \\ Q' &= [ Q_1 \quad \dots \quad Q_i \quad | \quad Q_{i+1} \quad | \quad Q_{i+1} \quad \dots \quad Q_{N+1} ]^\top, \\ \Lambda_k^\pm &= \Lambda_{i+1}^\pm, \quad \xi_k = \xi_{i+1}, \end{aligned}$$

meaning that a zone with flux function  $Q_{i+1}^\cdot$  will be created, followed by potentially changing the boundary speeds of that flux function everywhere.

For example, consider a road segment described by  $Q_1$  for  $x < 0$  and  $Q_2$  for  $x > 0$ . Let a moving bottleneck moving at speed  $u_m$  be described by flux function  $Q_{m1}$  while it is in the  $Q_1$  zone, and by  $Q_{m2}$  while it is in the  $Q_2$  zone. Then the traversal of the moving bottleneck over  $x = 0$  may be modelled by rules

$$\begin{aligned} r_1 &= ((Q_{m1}, Q_2), (Q_{m2}, (0, u_m))), \\ r_2 &= ((Q_1, Q_{m2}), (Q_2, (0, 0))), \\ r_3 &= ((Q_2, Q_{m2}), (Q_{m2}, (u_m, u_m))), \end{aligned}$$

where  $\xi_{m2} > \xi_{m1}$ .

**State insertion**  $\vee(\rho_\vee, x_\vee)_i$ : State insertion is an exogenous transition, i.e. it can be taken for any  $\mathcal{X} \in \mathcal{X}$  given the appropriate external input. It consists of adding two fronts at position  $x_\vee$  downstream of front  $i$ , with  $z_i \leq x_\vee \leq z_{i+1}$ ,

$$\begin{aligned} (N, z, \rho, Q) &\xrightarrow{\vee(x_\vee)_i} (N', z', \rho', Q') \\ N' &= N + 2, \\ z' &= [ z_1 \quad \dots \quad z_i \quad | \quad x_\vee \quad x_\vee \quad | \quad z_{i+1} \quad \dots \quad z_N ]^\top, \\ \rho' &= [ \rho_1 \quad \dots \quad \rho_i \quad | \quad \rho_{i+1} \quad \rho_{i+1} \quad | \quad \rho_{i+1} \quad \dots \quad \rho_{N+1} ]^\top, \\ Q' &= [ Q_1 \quad \dots \quad Q_i \quad | \quad Q_{i+1} \quad Q_{i+1} \quad | \quad Q_{i+1} \quad \dots \quad Q_{N+1} ]^\top. \end{aligned}$$

It is only necessary to specify  $i$  if  $z_i = x_\vee$  or  $z_{i+1} = x_\vee$ , in order to disambiguate the ordering of wavefronts.

**Flux function transition**  $\mathcal{Q}(q, i, j)$ : Finally, flux function transition is another exogenous transition, which covers changes in flux functions in specific areas. The transition is defined as

$$\begin{aligned} (Q) &\xrightarrow{\mathcal{Q}(q, i, j)} (Q') \\ Q' &= [ Q_1 \quad \dots \quad Q_i \quad | \quad q \quad \dots \quad q \quad | \quad Q_{j+1} \quad \dots \quad Q_N ]^\top, \end{aligned}$$

with  $q \in \mathcal{Q}$  and  $j > i$ . Formally, this change has no immediate effect on any of the other states, but it is likely to cause  $(\star)$  to be violated.

### Solution example revisited

We revisit the example from Section 4.5.1 in order to demonstrate the use of the proposed transition system model. The initial state can be described as  $\mathcal{X}_0 = (N^0, 0, z^0, \rho^0, Q^0, W)$ , with  $N^0 = 1$ ,  $z^0 = 0$ ,  $\rho^0 = [\rho_- \ 0]^\top$ ,  $Q^0 = [Q_- \ Q_+]^\top$ , and the only restricting wave speed bound is  $\hat{W}_-$ . The flux functions are described as  $Q_\pm = (V, \Sigma_\pm, \Lambda^\pm, \xi_\pm, \emptyset)$ , where  $V = [V_1 \ V_2]^\top$ ,  $\Sigma_- = [\sigma_1^- \ \sigma_2^-]^\top$ ,  $\Sigma_+ = [\sigma_1^+ \ \sigma_2^+]^\top$ ,  $\Lambda^\pm = (0, 0)$ , and  $\xi_- > \xi_+$ .

Since the condition  $(\star /)$  is violated, the first transition that is taken is

$$\begin{aligned} (N^0, z^0, \rho^0, Q^0) &\xrightarrow{\perp} (N^{0'}, z^{0'}, \rho^{0'}, Q^{0'}), \\ N^{0'} &= 3, z^{0'} = 0_3, \rho^{0'} = [\rho_- \ \rho^c \ \sigma_1^+ \ 0]^\top, Q^1 = [Q_- \mathbb{1}_2^\top \ Q_+ \mathbb{1}_2^\top]^\top. \end{aligned}$$

Next, since the state is admissible, we take passage of time transition  $\tau$  until  $\tau = t'$ , to  $N^1 = N^{0'}$ ,  $z^1 = z^{0'} + \Lambda^{0'} t'$ ,  $\rho^1 = \rho^{0'}$ , when the bottleneck  $Q_+$  is removed,  $(Q^1) \xrightarrow{\mathcal{Q}(Q_-, 2, 4)} (Q^{1'})$ ,  $Q^{1'} = Q_+ \mathbb{1}_4$ , which now causes the condition  $(\star \sim)$  to be violated for  $i = 2$ , causing transition

$$\begin{aligned} (N^{1'}, z^{1'}, \rho^{1'}, Q^{1'}) &\xrightarrow{\sim 2} (N^{1''}, z^{1''}, \rho^{1''}, Q^{1''}) \\ N^{1''} &= 4, z^{1''} = [\lambda_1 t' \ 0 \ 0 \ V_1 t']^\top, \end{aligned}$$

$$\rho^{1''} = [\rho_- \ \rho^c \ \rho^d \ \sigma_1^+ \ 0]^\top, \quad Q^{1''} = Q_- \mathbf{1}_5.$$

The state is thus made admissible again, and after another passage of time transition can be taken with  $\tau = t^c - t'$ , to  $N^2 = N^{1''}$ ,  $z^2 = z^{1''} + \Lambda^{1''}(t^c - t')$ ,  $\rho^2 = \rho^{1''}$ . Since  $\lambda_1 t^c = \hat{W}_-(t^c - t')$ , there is a front interaction  $-_1$  at  $t^c$ ,

$$\begin{aligned} (N^2, z^2, \rho^2, Q^2) &\xrightarrow{-_1} (N^{2'}, z^{2'}, \rho^{2'}, Q^{2'}), \\ N^{2'} &= 3, z^{2'} = [x_c \ V_1(t^c - t') \ V_1 t^c]^\top, \\ \rho^{2'} &= [\rho_- \ \rho^d \ \sigma_1^- \ 0]^\top, Q^{2'} = Q_- \mathbf{1}_4. \end{aligned}$$

After this transition, we reach a steady state where the only possible transition is passage of time with arbitrary  $\tau \geq 0$ .

### 4.5.3 Existence, uniqueness and non-Zenoness of the solutions

In this section we study the properties of the solutions to the FTTSM, defined in this section, and how they correspond to the wave-speed-bounded solutions of the LWR with piecewise-linear flux functions and piecewise-constant initial conditions. Throughout this subsection we assume that there are no rules defined for the system. For the FTTSM, we define Zeno behaviour as the existence of a solution with an infinite series of transitions that are not passage of time  $\mathcal{X} \xrightarrow{\tau(t_{\text{end}})} \mathcal{X}'$ , with  $t < t_{\text{end}}$ ,  $t < t'$ . The results are stated in the following lemmas and theorems.

**Lemma 1.** *The FTTSM is nonblocking and deterministic.*

*Proof.* All  $\mathcal{X} \in \mathcal{X}$  are in exactly one guard set, since guard sets  $\mathcal{G}_o$  form a partition of  $\mathcal{X}$ . Furthermore, all transitions  $\circ$  ensure that if  $\mathcal{X} \in \mathcal{X}$  and  $\mathcal{X} \xrightarrow{\circ} \mathcal{X}'$  then  $\mathcal{X}' \in \mathcal{X}$ . Therefore there exists a unique solution of the FTTSM for every initial state  $\mathcal{X}_0 \in \mathcal{X}$ .  $\square$

This lemma establishes the basic properties of the FTTSM based directly on the guard sets of the transitions. Next, we study the behaviour of the FTTSM in zones where the flux function is homogeneous, and with constant boundary conditions.

**Lemma 2.** *If the state of FTTSM  $\mathcal{X} \in \mathcal{G}_{o_i}$ , where transition  $\circ_i$  is  $\sim_i$  if  $\mathcal{X} \xrightarrow{\sim_i} \mathcal{X}'$  or  $/_i$  if  $\mathcal{X} \xrightarrow{/_i} \mathcal{X}'$ , then  $\mathcal{X}' \notin \mathcal{G}_{\sim_j}$  and  $\mathcal{X}' \notin \mathcal{G}_{/_j}$  for  $j = i, \dots, i + N' - N$ .*

*Proof.* First, consider the case when  $\mathcal{X} \in \mathcal{G}_{\sim_i}$ . Based on the definition of the transition  $\sim_i$ , we have that

$$[\rho'_i \ \dots \ \rho'_{i+N'-N+1}]^\top = \rho^i \tilde{\Sigma}_{Q_i, W}^{\rho_{i+1}}.$$

Therefore, for  $k = 0, \dots, N' - N$ ,

$$[\rho'_{i+k} \ \rho'_{i+k+1}]^\top = \rho^{i+k} \tilde{\Sigma}_{Q_i, W}^{\rho'_{i+k+1}},$$

so condition  $(\star\sim)_{i+k}$  holds and  $\mathcal{X}' \notin \mathcal{G}_{\sim_j}$ ,  $j = i + k$ . Next, if  $\mathcal{X} \in \mathcal{G}_{/i}$ , based on the definition of the transition  $/i$ , we have that

$$\begin{aligned} [\rho'_i \cdots \rho'_{i+m_- - 1}]^\top &= \rho'_i \tilde{\Sigma}_{Q_i, W}^{\rho'_-}, \\ [\rho'_{i+m_-} \cdots \rho'_{i+N' - N + 1}]^\top &= \rho'_{i+m_-} \tilde{\Sigma}_{Q_{i+1}, W}^{\rho'_{i+1}}, \\ m_- &= \rho'_i \tilde{m}_{Q_i, W}^{\rho'_-}, \end{aligned}$$

and  $\rho'_\pm$  are the solutions to the optimization problem (4.45) for  $\rho_- = \rho_i$ ,  $\rho_+ = \rho_{i+1}$ ,  $Q_- = Q_i$ ,  $Q_+ = Q_{i+1}$ . Therefore, for  $k = 0, \dots, m_- - 2, m_-, \dots, N' - N$ ,

$$[\rho'_{i+k} \rho'_{i+k+1}]^\top = \rho'_{i+k} \tilde{\Sigma}_{Q'_{i+k}, W}^{\rho'_{i+k+1}},$$

so condition  $(\star\sim)_{i+k}$  holds and  $\mathcal{X}' \notin \mathcal{G}_{\sim_j}$ ,  $j = i + k$ . For  $i + m_- - 1$ , we have  $\rho'_{i+m_- - 1} = \rho'_-$ ,  $\rho'_{i+m_-} = \rho'_+$ ,  $Q'_{i+m_- - 1} = Q_i$ , and  $Q'_{i+m_-} = Q_{i+1}$ , so since  $\rho'_-$  and  $\rho'_+$  are the result of transition  $/i$ , condition  $(\star/)_{i+m_- - 1}$  holds and  $\mathcal{X}' \notin \mathcal{G}_{/i+m_- - 1}$ .  $\square$

**Theorem 1.** *Let  $\mathcal{X}_0$  be the initial state of the FTTSM with  $t = 0$ . Consider the zone between two flux function boundaries  $[z_-^*(t), z_+^*(t)]$ , described by flux function  $Q^*$ , and assume that for  $0 < t < T$ ,  $T > 0$ :  $z_-^*(t) < z_+^*(t)$ , boundary conditions  $\rho_{i_-}^*(t)$  and  $\rho_{i_+}^*(t+1)$  are constant, and conditions  $(\star-)_{i_-^*(t)}$  and  $(\star-)_{i_+^*(t)+1}$  hold. Then  $\rho(x, t')$ ,  $x \in [z_-^*(t'), z_+^*(t')]$ , given by  $\mathcal{X}$ , with  $t = t'$  is the unique wave-speed-bounded solution of the corresponding LWR model with initial conditions  $\rho(x, 0)$ ,  $x \in [z_-^*(0), z_+^*(0)]$ , given by  $\mathcal{X}_0$ , for  $0 < t' < T$ ,  $T > 0$ , and its Total Variation  $\text{T.V.}(\rho(\cdot, t'))$  is nonincreasing in  $t'$ .*

*Proof.* First, if the density is forced in the zone of flux function  $Q^*$ ,  $F^* \neq \emptyset$ , this holds trivially, with  $\rho(x, t') = F^*$  on  $[z_-^*(t'), z_+^*(t')]$  for  $t' > 0$ . Otherwise, after a finite number of transmissions  $\mathcal{X}_0 \xrightarrow{\circ_1} \dots \xrightarrow{\circ_K} \mathcal{X}'_0$ , the state  $\mathcal{X}'_0$  satisfies  $(\star)$  for  $i = i_-^*(0), \dots, i_+^*(0)$ . Here transitions  $\circ_k$  are:  $/i_-^*(0)$ ,  $/i_+^*(0)$ , and a finite number of transitions  $\sim_i$  and  $-_i$  for  $i_-^*(0) < i < i_+^*(0)$ , if the state enters their respective guard sets during the transitions from  $\mathcal{X}_0$  to  $\mathcal{X}'_0$ . Afterwards, due to Lemma 2, while  $t < T$ , the evolution of the state can be described with only transitions  $\tau(t_{\text{end}})$  and  $-_i$ , for which the Total Variation is nonincreasing.  $\square$

This theorem also holds in case all flux functions  $Q_i$  are identical, in which case  $[z_-^*(t), z_+^*(t)]$  corresponds to the entire space,  $[-\infty, \infty]$ . In case the wave-speed bounds are such that the entropy solution is identical to the wave-speed-bounded solution, this theorem corresponds to Corollary 2.8 from [206], for constant boundary conditions and inside a zone where the flux function is homogeneous. Next, we study the case of an infinitely narrow zone with different flux function, and show that this situation does not give rise to Zeno behaviour.

**Lemma 3.** *Let  $\mathcal{X}$  be the state of the FTTSM with  $z_{i_-} = z_{i_-+1} = \dots = z_{i_+}$  and no other wavefront positions equal, where  $Q_{i_-} = Q_L$ ,  $Q_i = Q_C$  for  $i = i_-+1, \dots, i_+$ ,  $Q_{i_++1} = Q_R$ ,  $Q_L \neq Q_C$ ,  $Q_C \neq Q_R$ , and the density in the zones of  $Q_L$  and  $Q_R$  is not forced,  $F_L = F_R = \emptyset$ . Then  $\mathcal{X}' \in \mathcal{G}_\tau$ , where  $\mathcal{X} \xrightarrow{\circ 1} \dots \xrightarrow{\circ K} \mathcal{X}'$ ,  $\circ k \in \{-i_k, \sim i_k, /i_k, !i_k\}$ ,  $k = 1, \dots, K$  with finite  $K$ .*

*Proof.* Without loss of generality, assume  $(\star \sim)_i$ ,  $(\star /)_i$ , and  $(\star !)_i$ , hold for  $i \notin [i_-, i_+]$ , which could always be achieved in a finite number of transitions. Firstly, if the density in the zone of  $Q_C$  is forced, after transitions  $!i$ ,  $i = i_- + 1, \dots, i_+$  inside the zone,  $/i_-$  and  $/i_+$  at both borders of the zone, and  $i_+ - i_- - 1$  instances of transition  $-i_{-+1}$ , the resulting state  $\mathcal{X}' \in \mathcal{G}_\tau$ . Otherwise, after up to two transitions  $/i$  and up to  $i_+ - i_- - 1$  transitions  $\sim i$ , the resulting state is  $\mathcal{X}^0$ , and it holds that either  $\lambda_{i_-^0} = \lambda_{i_-^0+1} = \dots = \lambda_{i_+^0}$  and  $\mathcal{X}^0 \in \mathcal{G}_\tau$ , satisfying the requirements of the lemma, or  $\mathcal{X}^0 \in \mathcal{G}_{-i}$  for some  $i_-^0 \leq i \leq i_+^0$ . We denote  $i_-^m$  and  $i_+^m$  based on state  $\mathcal{X}^m$  so that  $Q_{i_-^m}^m = Q_L$ ,  $Q_i^m = Q_C$  for  $i = i_-^m+1, \dots, i_+^m$ ,  $Q_{i_+^m+1}^m = Q_R$ . If  $\mathcal{X}^0 \in \mathcal{G}_{-i}$  for some  $i$ , after  $k^0$  front interaction transitions,  $1 \leq k^0 \leq i_+^0 - i_-^0 - 1$ , the resulting state  $\mathcal{X}^{0'}$  will either satisfy the requirements of the lemma, or  $\mathcal{X}^{0'} \in \mathcal{G}_{/i_-^{0'}} \cup \mathcal{G}_{/i_+^{0'}}$ . After transition  $/i_-^{0'}$  or  $/i_+^{0'}$  from  $\mathcal{X}^{0'}$  to  $\mathcal{X}^1$ , the process repeats with  $k^m$  front interaction transitions,  $1 \leq k^1 \leq i_+^1 - i_-^1 - 1$ , and with  $i_+^m < i_+^{m-1}$ ,  $m = 1, \dots, M$ , until state  $\mathcal{X}^{M'}$  satisfies the requirements of the lemma, with  $\lambda_{i_-^{M'}} \leq \lambda_{i_-^{M'}} \leq \dots \leq \lambda_{i_+^{M'}}$  if  $\lambda_{i_-^{M'}} \leq \lambda_{i_+^{M'}}$ , and  $Q_{i_-^{M'}}^{M'} = Q_L$ ,  $Q_{i_-^{M'}+1}^{M'} = Q_R$  if  $\lambda_{i_-^{M'}} > \lambda_{i_+^{M'}}$  (i.e. the zone described by flux function  $Q_C$  vanishes), or we have  $Q_{i_-^{M'}} = Q_L$ ,  $Q_{i_+^{M'}} = Q_C$ , and  $Q_{i_+^{M'}+1} = Q_R$ . Finally, if  $Q_{i_-^{M'}} = Q_L$ ,  $Q_{i_+^{M'}} = Q_C$ , and  $Q_{i_+^{M'}+1} = Q_R$ , the state satisfies the requirements of the lemma after up to three series of transitions, where the first and third series consist of a finite numbers of transitions  $/i_\pm^p$  or  $-i_\pm^p$ , and the second series consists of a finite number of transitions  $/i_\pm^p$  or  $-i_\pm^p$ , i.e. if the first transition is  $/i_-^{M'}$ , the second series of transitions will happen at the boundary between  $Q_C$  and  $Q_R$ , and if the first transition is  $/i_+^{M'}$ , the second series of transitions will happen at the boundary between  $Q_L$  and  $Q_C$ . Therefore, no infinite internal behaviour of the FTTSM, starting from  $\mathcal{X}$  and with no states in  $\mathcal{G}_\tau$  can exist, and after a finite number of transitions, the system reaches state  $\mathcal{X}' \in \mathcal{G}_\tau$ .  $\square$

Finally, we state the main result of this subsection.

**Theorem 2.** *There exists a unique, non-Zeno solution to the FTTSM, with initial state  $\mathcal{X}_0$  and a finite number of exogenous transitions, that defines a unique wave-speed-bounded solution of the LWR model  $\rho(x, t)$ , with corresponding piecewise-linear flux functions, corresponding piecewise-constant initial conditions, with the exogenous transitions applied in a corresponding way.*

*Proof.* Consider first the case when there are no exogenous transitions and passage of time transition is taken with an arbitrarily large  $t_{\text{end}}$ . Firstly, due to Lemma 1,

the solution to initial conditions  $\mathcal{X}_0$  from arbitrary  $t = t_0$  until arbitrary  $t_{\text{end}} > t_0$  exists and is unique. Due to Theorem 1, the only way Zeno behaviour can arise in FTSM is through interactions with flux function boundaries. Due to Lemma 2, the transitions at a flux function boundary can only happen once a wavefront reaches the boundary, and the wavefronts created by the flux function boundary transition travel away from the boundary. Therefore any kind of periodic behaviour would require interplay between two flux function boundaries. If the zone between two flux function boundaries is of non-zero length, the wavefronts originating from one boundary can only reach the other boundary after non-zero time, so the only case when Zeno behaviour could arise is if the length of the zone goes to zero. Therefore, due to Lemma 3, no Zeno behaviour can arise. If exogenous transitions are forced at times  $T_1, \dots, T_{N_{\text{exo}}}$ , the solution can be split into intervals  $[t_0, T_1), [T_1, T_2), \dots, [T_{N_{\text{exo}}}, t_{\text{end}}]$ . Since the exogenous transitions do not cause the output of the system to change, we may form  $\rho(x, t)$  out of pieces between two exogenous transitions, with each exogenous transition changing some part of the LWR model.  $\square$

### Modelling the influence of CAVs

We describe how CAVs acting as sensors and actuators can be modelled in the FTSM framework. We may add a CAV with identifier  $\xi$  to the model, at position  $x_\xi$  and moving at speed  $u_\xi$ , by first taking a transition  $\vee(x_\xi)_{i_-}$ , which creates two fronts downstream of front  $i_-$ ,  $z_{i_-} \leq x_\xi \leq z_{i_-+1}$ , followed by taking a transition  $\mathcal{Q}(q_\xi, i_- + 1, i_- + 2)$ , with  $q_\xi = (V_\xi, \Sigma_\xi, (u_\xi, u_\xi), \xi, F_\xi)$ . The choice of  $V_\xi$ ,  $\Sigma_\xi$  and  $F_\xi$  will depend on the role that the CAV has. In order for the CAVs' behaviour to be modelled correctly, we need the identifiers corresponding to each CAV to be greater than the identifier of the flux function describing the background traffic.

We assume that CAVs can act as both traffic sensors and traffic actuators. A CAV  $\xi$  acting as sensors will communicate its local measurements of traffic density  $\check{\rho}_\xi$ , traffic speed  $\check{v}_\xi$ , from which we can get  $\check{q}_\xi = \check{\rho}_\xi \check{v}_\xi$ . In case the CAV  $\xi$  is moving slower than the surrounding traffic, we may also measure the overtaking flow  $\check{\omega}_\xi$ . In general, the values that are directly related to some measurements will be denoted by caron ( $\check{\cdot}$ ). We model CAV  $\xi$  acting as an actuator by setting its reference speed  $u_\xi$ , which determines its actual speed together with the speed of the traffic immediately downstream,

$$\lambda_\xi^\pm = \min \{u_\xi, v_{\xi+}\},$$

$$v_{\xi+} = Q_{i_{\xi+}}(\rho_{i_{\xi+}})/\rho_{i_{\xi+}}, \quad (4.49)$$

$$i_{\xi+} = \min i, \text{ s.t. } z_i \geq x_\xi, Q_i \neq Q_\xi. \quad (4.50)$$

A vehicle moving slower than the rest of the traffic will act as a moving bottleneck, limiting the overtaking flow, which we may use for actuating traffic control. We model the effect a CAV acting as a moving bottleneck will have on the rest of the traffic by using  $V_\xi$  and  $\Sigma_\xi$  in the shape of a bottleneck diagram.

## 4.6 Summary

Moving traffic phenomena, such as moving bottlenecks and stop-and-go waves are known to be hard to represent in classical traffic models. Since dealing with these phenomena is a key part of Lagrangian traffic control, in this chapter, we presented extensions to the basic traffic models, as well as new traffic models that are able to capture their effects. While not moving, capacity drop is another traffic phenomenon that is included. The influence of moving bottlenecks is introduced to the LWR model through the addition of a flow constraint, or a zone with a different, reduced flux function. This approach is repeated in the CTM, exploiting the connection between the two models. Another extension of the CTM is also discussed, the multi-class CTM. By allowing the free flow speed of different classes to vary in time in each cell, we are able to describe the traffic phenomena of interest. This model is then simplified, abstracting away the spatial dimension, and focusing only on the queue lengths at stationary and moving bottlenecks, in the tandem queueing model. Finally, another numerical scheme for solving conservation laws is formalized into the FTTSM, and we discussed modelling the effects of stop-and-go waves through imposing wave speed bounds on the solutions of Riemann problems. These models will be used in the following chapters for simulation and control design and implementation.



**Part III**

**Control**



## Chapter 5

# Platoon merging coordination

TRUCK platooning has many benefits, such as reducing the air drag, and consequently, the fuel consumption. In the most general setting, the problem of platoon merging coordination may include an arbitrary number of vehicles traveling on a complex road network, with different constraints and objectives for each vehicle. Here we focus on the simplest case of the problem, considering two trucks located at different positions on the same road. The vehicle that is further along the road is called the leader vehicle, and the vehicle upstream of it is called the follower vehicle. We study the process of closing the distance between the two vehicles, merging into a platoon en route, by having the follower vehicle drive at a higher speed than the leader vehicle. During this merging phase, the combined fuel consumption of the two vehicles is higher than in case they did not attempt to merge into a platoon, since they have to deviate from their individual optimal speeds. This increase in fuel consumption is offset by the lower fuel consumption in the platooning phase, when the two vehicles experience reduced air drag due to the slipstream effect. Therefore, it is important for the platoon merging process to be made reliable and predictable, since any unforeseen delay in platoon merging causes a reduction in fuel savings. In particular, there are two sub-problems that we consider.

The first sub-problem pertains to predicting the vehicle trajectories during the merging phase. Here we assume that some control laws are governing the motion of the vehicles, and apply it to vehicle and environment models in order to predict when and where platoon merge will occur. We use the data from platoon merging experiments done on public roads combined with road grade information to learn the dynamic model of vehicles' motion, depending on the deviation of vehicles' speeds from their reference values and road grade at their position. These learned models are then used to predict the trajectories of the two vehicles during the merging phase.

The second sub-problem is to design an optimal control law for platoon merging. The case when only the road grade profile and engine power constraints are consid-

ered is referred in the literature as the look-ahead vehicle control problem [217, 218] In contrast to previous work on the look-ahead vehicle control problem, we take surrounding traffic into consideration when calculating energy-optimal catch-up speed pairs. The influence of the traffic is particularly challenging to account for, since the traffic conditions between the leader and follower vehicles is affected by both of their speeds. Moreover, the follower vehicle is likely to be affected by the traffic conditions once it reaches the congestion left in the wake of the slow moving leader vehicle.

We first give the model of vehicles used in this chapter in Section 5.1, and briefly describe the reasons why they deviate from their desired merging speeds. Then, we deal with the two sub-problems in detail in Sections 5.2 and 5.3.

## 5.1 Vehicle model and deviations from the nominal speed

The dynamics of vehicle  $\xi$  can be expressed as

$$\begin{aligned} m_\xi \dot{v}_\xi(t) &= F_{t,\xi}(t) - F_{b,\xi}(t) - F_{a,\xi}(v_\xi(t), d_\xi(t)) - F_{r,\xi}(\gamma(x_\xi(t))) - F_{g,\xi}(\gamma(x_\xi(t))), \\ \dot{x}_\xi(t) &= v_\xi(t) \end{aligned} \quad (5.1)$$

where  $x_\xi(t)$ ,  $v_\xi(t)$ , and  $m_\xi$  are the the vehicle's longitudinal position, the vehicle speed, and the vehicle mass, respectively. The reader is referred to [219] for more detailed vehicle models. The vehicle is actuated through controlling the traction force  $F_{t,\xi}(t)$ , and the braking force  $F_{b,\xi}(t)$ , either by a human driver or some form of cruise control. The following resistive forces are considered: roll resistance  $F_{r,\xi}(\gamma(x_\xi(t)))$  and aerodynamic drag  $F_{a,\xi}(v_\xi(t), d_\xi(t))$ . The roll resistance acting on vehicle  $\xi$  depends on the road grade at its position,

$$F_{r,\xi}(\gamma(x_\xi(t))) = -m_\xi g c_r \cos(\gamma(x_\xi(t))),$$

where  $c_r$  is the roll resistance coefficient, and  $g$  is the gravitational acceleration. The aerodynamic drag  $F_{a,\xi}(v_\xi(t), d_\xi(t))$  is a major component of the resistive force acting on a large vehicle, and it depends on the vehicle's speed  $v_\xi(t)$  and the distance to its preceding vehicle in a platoon  $d_\xi(t)$ . It is given by

$$F_{a,\xi}(v_\xi(t), d_\xi(t)) = c_a v_\xi^2(t) \phi_a(d_\xi(t)),$$

where  $c_a$  is a constant that depends on air density and the aerodynamic characteristics of the vehicle, and  $\phi_a(d)$  models the reduction in air drag based on the distance to the preceding vehicle if vehicle  $\xi$  is a follower in a platoon. The gravitational force acting on the vehicle in the opposite direction of its movement is given by

$$F_{g,\xi}(\gamma(x_\xi)) = m_\xi g \sin(\gamma(x_\xi)).$$

Consider a platoon of  $n_p$  vehicles that drive at constant speed  $v_p$ , and have a constant inter-vehicular distance  $d$ . Then the resistive force acting on the platoon

leader and followers is  $F_{a,l}(v_p) = c_a v_p^2 \phi_l$  and  $F_{a,f}(v_p) = c_a v_p^2 \phi_f$ , respectively, where  $\phi_l \approx 1$  is the air drag reduction ratio of the leader vehicle, and  $\phi_f = \phi_a(d) < 1$  is the air drag reduction of the follower vehicles in a platoon. The total aerodynamic resistance acting on the whole platoon can be written as  $F_{a,p}(v_p) = k_a v_p^2 \phi_p$ , where  $\phi_p = \phi_l + (n_p - 1)\phi_f$  is the total air drag reduction ratio due to platooning. Here we consider the case when two vehicles are merging into a platoon, with vehicle  $\xi = 1$  as the leader, and vehicle  $\xi = 2$  as the follower. Platooning always reduces the total air drag, i.e., air drag reduction ratio of two vehicles driving alone would be  $\phi_1 = \phi_2 = 1$ ,  $\phi_1 + \phi_2 = 2$ , while the total air drag coefficient of the platoon consisting of those vehicles would be approximately  $\phi_p \approx 1.7$ , assuming the inter-vehicular distance is 20 m [96].

Since the road grade is given as a function of the position, it can be beneficial to rewrite (5.1) to the form with position as the independent variable and time as dependent variable. Since  $v_\xi > 0$ ,  $x_\xi(t)$  is a bijection and we can write  $t_\xi(x)$  as its inverse, i.e., the time when vehicle  $\xi$  is at the position  $x$ . Then (5.1) is uniquely rewritten as

$$\begin{aligned} m_\xi v_\xi(x) \frac{dv_\xi(x)}{dx} &= F_{t,\xi}(t_\xi(x)) - F_{b,\xi}(t_\xi(x)) - F_{a,\xi}(v_\xi(x), d_\xi(x)) \dots \\ &\quad - F_{r,\xi}(\gamma(x)) - F_{g,\xi}(\gamma(x)), \\ \frac{dt_\xi}{dx} &= \frac{1}{v_\xi(x)}. \end{aligned} \tag{5.2}$$

Consider the simple platoon merging problem as formulated in Section 1.2.1. If both vehicles would be able to keep their desired speeds,  $v_1(t) = u_1$  and  $v_2(t) = u_2$  for  $0 \leq t \leq t_m$ , their merge time and position would be

$$\begin{aligned} t_m &= \frac{x_1(0) - x_2(0) - d_p}{u_2 - u_1}, \\ x_m &= x_2(0) + u_2 \frac{x_1(0) - x_2(0) - d_p}{u_2 - u_1}. \end{aligned} \tag{5.3}$$

However, even if the vehicles' desired speeds are constant, they will often be forced to deviate from them. Heavy vehicles will often need to reduce their speed in order to tackle even small uphill slopes, and they need to brake or coast on downhill slopes in order to keep speed within safe bounds. A comparison of speed deviation from the nominal for the two trucks of different weight is shown in Figure 5.1.

The surrounding traffic will also have an effect on the vehicles attempting to merge into a platoon. Although when traffic is in free flow, trucks typically drive at a lower speed than the passenger cars, entering congestion might force a truck to reduce its speed. Since the slow moving leader vehicle will act as a moving bottleneck, the congestion left in its wake will force the follower vehicle to deviate from its desired speed.

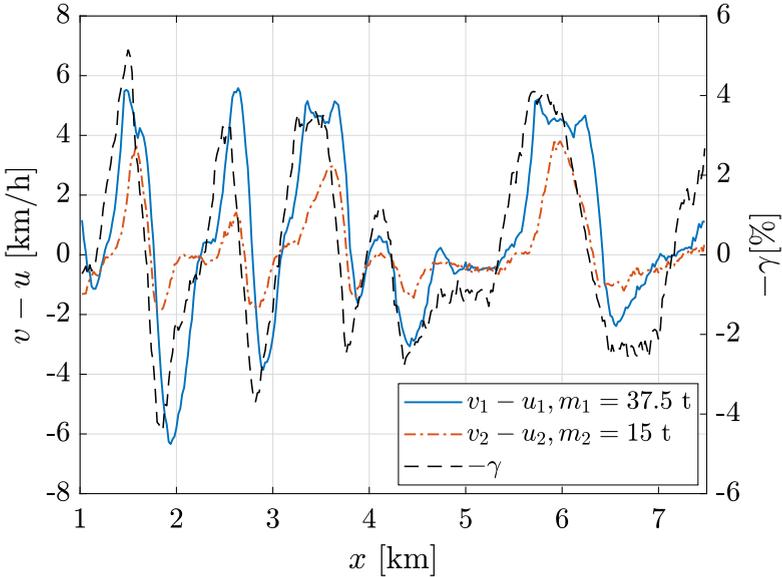


Figure 5.1: Deviation from the desired speed due to varying road grade for two trucks with different masses and negative road grade.

## 5.2 Platoon merging distance prediction based on the road grade

The first sub-problem considered in this chapter is predicting how long it will take two trucks to form a platoon while driving on a highway at set cruise speeds, considering only the influence of varying road grade. In order to model the influence of road grade on the vehicle speed, we can either use a cruise control model, if available, or identify the dependence from data. Here, we used the experimental data from [113] to train a neural network model that we then use to predict the evolution of vehicle speed. We then integrate the predicted vehicle speed profiles to calculate a prediction of when and where the platoon merge will occur.

A significant advantage of speed prediction based merge distance prediction is that it gives us a prediction of the vehicle positions during the whole merging phase. This means that a disturbance that will change the platoon merge time can be detected immediately, by comparing the current vehicle positions, acquired from the GPS system, with their predicted values. When such a disturbance is identified, the prediction can be recalculated taking into account the updated information. Additionally, the new information can be used to re-plan desired vehicle speed profiles in order to compensate for the disturbance.

The predicted positions of vehicles are

$$\begin{aligned}\frac{d\hat{x}_1(\tau|t)}{d\tau} &= \hat{v}_1(\tau|t), \\ \frac{d\hat{x}_2(\tau|t)}{d\tau} &= \hat{v}_2(\tau|t),\end{aligned}\tag{5.4}$$

where  $\hat{v}_\xi(\tau|t)$  is the predicted speed of vehicle  $\xi$ , and  $\tau > 0$  is the relative time for the prediction calculated at time  $t$ , i.e.,  $\hat{x}_\xi(\tau|t)$  is a prediction of  $x_\xi(t + \tau)$ . The predicted distance between the leader and the follower vehicle can be written  $\hat{d}(\tau|t) = \hat{x}_1(\tau|t) - \hat{x}_2(\tau|t)$ . The predicted platoon merge time and position are therefore given by

$$\begin{aligned}\hat{\tau}_m(t) &= \min \left\{ \tau \geq 0 \mid \hat{d}(\tau|t) \leq d_p \right\}, \\ \hat{x}_m(t) &= \hat{x}_2(\hat{\tau}_m(t)|t).\end{aligned}\tag{5.5}$$

We define the merge distance as the distance travelled by the follower vehicle until merging into a platoon,

$$\hat{d}_m(t) = \hat{x}_2(\hat{\tau}_m(t)|t) - x_2(t).$$

### 5.2.1 Vehicle speed prediction models

We first discuss two simple speed prediction models, and then give the neural network speed model in more detail. Finally, we describe how the proposed models were trained from data.

#### *Constant average speed models:*

The simplest vehicle speed prediction model assumes that both vehicles travel at constant speed  $\hat{v}_\xi = \bar{v}_\xi$ . The platoon merge time and distance predictions are given by

$$\begin{aligned}\hat{\tau}_m(t) &= \frac{d_0(t) - d_p}{\hat{v}_2 - \hat{v}_1}, \\ \hat{d}_m(t) &= \hat{v}_2 \frac{d_0(t) - d_p}{\hat{v}_2 - \hat{v}_1}.\end{aligned}\tag{5.6}$$

Here the speed of the leader and follower vehicles  $\bar{v}_\xi$  are given as average speeds during the merging phase of the simulation runs in the training set.

#### *Road grade moving average speed model:*

Better results can be obtained by modelling the vehicle speed deviation from its nominal value as a piecewise linear function of the moving average of road grade  $\bar{\gamma}(x)$ ,

$$\hat{v}_\xi^\gamma(x) = \begin{cases} u_\xi(1 + k_{\xi, \gamma_+} \bar{\gamma}(x)), & \bar{\gamma}(x) \geq 0 \\ u_\xi(1 + k_{\xi, \gamma_-} \bar{\gamma}(x)), & \bar{\gamma}(x) < 0. \end{cases}\tag{5.7}$$

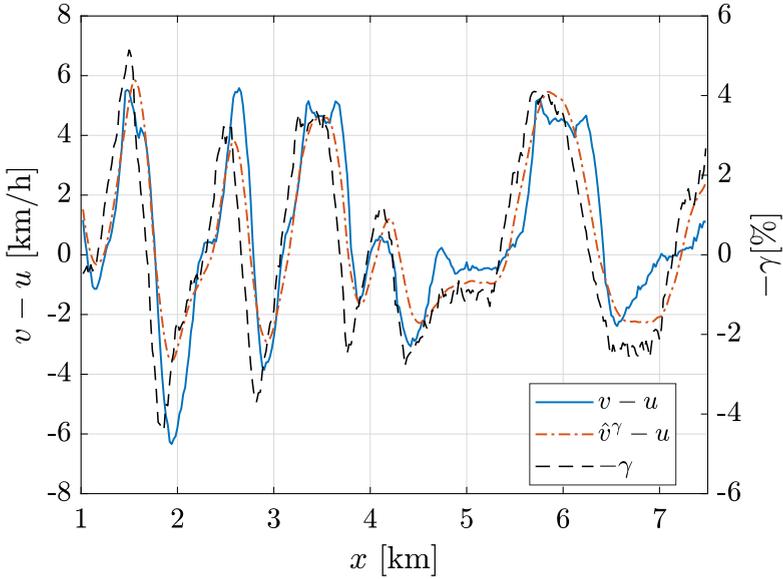


Figure 5.2: Actual and predicted deviation from the desired speed  $u$  using speed prediction model (5.7) and negative road grade.

where  $k_{\xi, \gamma_+}$  and  $k_{\xi, \gamma_-}$  are parameters that describe the influence of positive and negative  $\bar{\gamma}_k$ , respectively. An example of measured and modelled vehicle speeds are shown in Figure 5.2. The distance over which the moving average of the road grade is calculated is determined empirically.

### ***Neural network speed model:***

Finally, the neural network vehicle speed prediction model is based on learning the net propulsive force model and applying it in the dynamic equations (5.1) or (5.2). Grouping air drag resistance  $F_{a, \xi}$  and roll resistance  $F_{r, \xi}$  with the traction force  $F_{t, \xi}$  and braking force  $F_{b, \xi}$  into  $F_{p, \xi} = F_{t, \xi} - F_{b, \xi} - F_{a, \xi} - F_{r, \xi}$ , this net propulsive force can be treated as the control action of the cruise controller. Then, the speed dynamics of vehicle  $\xi$ , give in (5.1), simplifies to

$$m_{\xi} \dot{v}_{\xi} = F_{p, \xi} - m_{\xi} g \sin(\gamma(x_{\xi})).$$

We assume that  $F_{p, \xi}$  is a function of the deviation from the vehicle reference

speed, and road grade,  $F_{p,\xi}(v_\xi - u_\xi, \gamma(x_\xi))$ . The speed prediction can be written as

$$\begin{aligned}\frac{d\hat{v}_\xi(\tau|t)}{d\tau} &= \frac{F_{p,\xi}(\hat{v}_\xi(\tau|t) - u_\xi(t), \gamma(x_\xi(\tau|t)))}{m_\xi} - g \sin(\gamma(\hat{x}_\xi(\tau|t))), \\ \frac{d\hat{x}_\xi(\tau|t)}{d\tau} &= \hat{v}_\xi(\tau|t),\end{aligned}$$

with initial conditions  $\hat{x}_\xi(0|t) = x_\xi(t)$  and  $\hat{v}_\xi(0|t) = v_\xi(t)$ . Alternatively, taking position as the independent variable, we have

$$\begin{aligned}\frac{d\hat{v}_\xi(x|t)}{dx} &= \frac{1}{\hat{v}_\xi(x|t)} \left( \frac{F_{p,\xi}(\hat{v}_\xi(x|t) - u_\xi(t), \gamma(x))}{m_\xi} - g \sin(\gamma(x)) \right), \\ \frac{d\hat{\tau}_\xi(x|t)}{dx} &= \frac{1}{\hat{v}_\xi(x|t)},\end{aligned}$$

with initial conditions  $\hat{\tau}_\xi(x_\xi(t)|t) = 0$ , and  $\hat{v}_\xi(0|t) = v_\xi(t)$ . The function modelling the propulsive force

$$\frac{F_{p,\xi}(\hat{v}_\xi(x|t) - u_\xi(t), \gamma(x))}{m_\xi} = \hat{v}_\xi(x|t) \frac{d\hat{v}_\xi}{dx}(x|t) + g \sin(\gamma(x)) \quad (5.8)$$

can be learned from data. Using this model, we can predict vehicle speeds for the whole length of the road of interest.

The model is discretized into cells of equal length  $L$ , with upstream boundaries in  $X_i = X_0 + iL$ . The road cell length  $L$  is taken small enough to capture the dynamics of the system, but large enough so that  $L > Tv^{\max}$ , i.e., vehicles do not pass through cells of length  $L$  in less than  $T$ . Then, to each  $X_i$  along the vehicle trajectory, for both vehicles  $\xi$ , we assign

$$\begin{aligned}t_{\xi,i} &= \min \{t|x_\xi(t) > X_i\}, \\ v_{\xi,i} &= \frac{\int_{t_{\xi,i}}^{t_{\xi,i+1}} v_\xi(t) dt}{t_{\xi,i+1} - t_{\xi,i}},\end{aligned}$$

where  $t_{\xi,i}$  is the time vehicle  $\xi$  enters cell  $[X_i, X_{i+1}]$ , and  $v_{\xi,i}$  its average speed in the cell. The road grade  $\gamma_i$  is also taken as average road grade over the road cell  $[X_i, X_{i+1}]$ .

Discretizing (5.8) by integration and approximating  $\sin(\gamma) \approx \gamma$ , we have

$$\frac{F_{p,\xi}(v_{\xi,i-1} - u_\xi, \gamma_{i-1})}{m_\xi} = \frac{v_{\xi,i}^2 - v_{\xi,i-1}^2}{2L} + g\gamma_{i-1}, \quad (5.9)$$

which can be modelled using a feedforward neural network, as will be described in the following subsection. Once the model for  $F_{p,\xi}$  is available, the discrete prediction model for vehicle speed becomes

$$\hat{v}_{\xi,i+1} = 2L \sqrt{\hat{v}_{\xi,i}^2 + \frac{F_{p,\xi}(\hat{v}_{\xi,i} - u_\xi, \gamma_i)}{m_\xi} - g\gamma_i}, \quad (5.10)$$

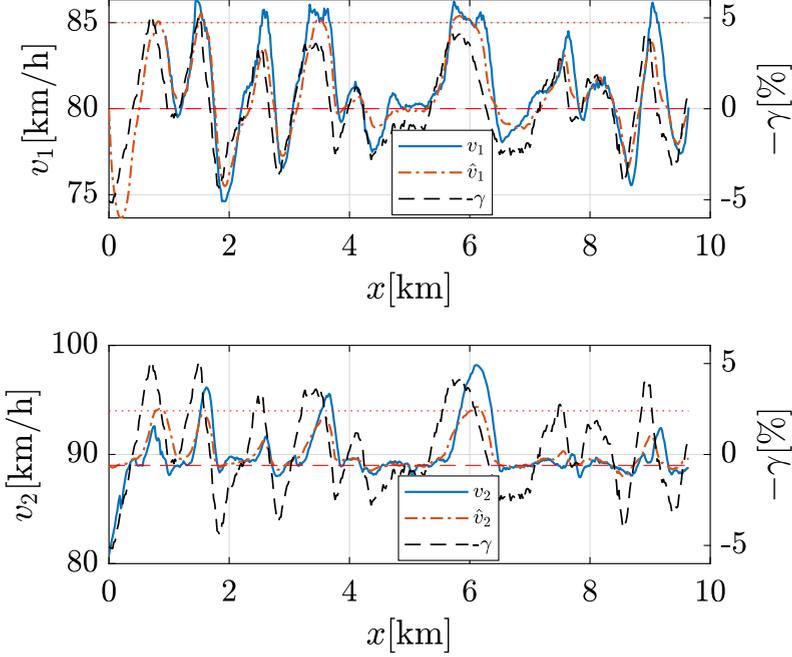


Figure 5.3: Predicted ( $\hat{v}$ ) and measured ( $v$ ) speed for the leader and the follower vehicle and negative road grade.

and we assume the speed predictions to be piecewise constant on each road segment,

$$\hat{v}_\xi(\tau|t) = \hat{v}_{\xi,i}, \quad \hat{t}_{\xi,i} \leq t + \tau < \hat{t}_{\xi,i+1}, \quad (5.11)$$

where

$$\hat{t}_{\xi,i+1} = \hat{t}_{\xi,i} + \frac{L}{\hat{v}_{\xi,i}}.$$

Finally, the vehicle speed prediction is calculated by initializing (5.10) with the current vehicle speed  $\hat{v}_{\xi,i_\xi(t)}(t) = v_\xi(t)$ , where  $i_\xi(t)$  is the cell the vehicle is in at time  $t$ ,  $X_{i_\xi(t)} \leq x_\xi(t) < X_{i_\xi(t)+1}$  and recursing (5.10). We then use (5.11) in (5.4) to calculate

$$\hat{x}_\xi(\tau|t) = \hat{v}_{\xi,i_\xi(t)-1} \hat{\tau}_{\xi,i_\xi(t)} + \sum_{i=i_\xi(t)}^{i_\xi(\tau|t)} (\hat{v}_{\xi,i} (\hat{\tau}_{\xi,i+1} - \hat{\tau}_{\xi,i})) + \hat{v}_{\xi,i_\xi(\tau|t)} (\tau - \hat{\tau}_{\xi,i_\xi(\tau|t)}),$$

where  $i_\xi(t)$  and  $i_\xi(\tau|t)$  are given by

$$X_{i_\xi(t)} \leq x_\xi(t) < X_{i_\xi(t)+1},$$

$$X_{i_\xi(\tau|t)} \leq x_\xi(\tau|t) < X_{i_\xi(\tau|t)+1}.$$

The platoon merge time and position predictions are calculated according to (5.5). A comparison between the measured speeds and the speed prediction acquired this way, for a part of a test run, is shown in Figure 5.3.

### 5.2.2 Speed prediction model training

In the experiments in [113], two HDVs were driving on an 11 km long stretch of public highway between Stockholm and Södertälje, namely between the Hallunda and Moraberg interchanges. Two standard Scania tractor trucks were used. The lead vehicle had a 480 hp engine and its total weight, including its trailer, was 37.5 tonnes. The follower vehicle had a 450 hp engine, had no trailer and weighed 15 tonnes. The road is fairly hilly, with road grades as high as  $\pm 5\%$ . The HDVs, initially apart, attempted to form a two-vehicle platoon by driving with different desired speed adaptive cruise control (ACC) settings. Three different desired speed pairs were considered,  $(u_1, u_2) = (75, 85), (75, 89)$  and  $(80, 89)$  km/h, where  $u_1$  is the reference speed of the leader vehicle and  $u_2$  of the follower. Downhill speed control was also active, with the offset of 5 km/h, allowing the vehicles to accelerate on downhill slopes and gain speed up to the set limit. The initial distance between the vehicles ranged from 400 m to 1300 m. The part of the experiment data that we used consist of periodical vehicle speed measurements and calculated distance between the vehicles, together with the information about road topography. Since we are primarily interested in the merging phase, we will consider the platoon merging completed when the distance between the vehicles is less than  $d_p = 80$ m, ignoring phenomena such as persistent drivers.

We used the vehicle speed data from the experiments to train the two proposed vehicle speed prediction models, the neural network approximation model (5.9)–(5.11) and the simple road grade moving average piecewise linear model (5.7). Roughly half of the experiment data was used for training and the rest was used for testing, and only the test runs which resulted in successful platoon formation were considered. Models for the leader and the follower vehicle speed prediction were trained independently.

Training the road grade moving average piecewise linear model consists of four linear regression equations of the form

$$\frac{v_{\xi,i} - u_\xi}{u_\xi} = k_{\xi,\gamma_\pm} \bar{\gamma}_i,$$

one for uphill ( $\gamma_+$ ) and one for downhill ( $\gamma_-$ ) slopes for each vehicle. Here, road grade is averaged over 400 meters and the calculated values of the regression parameters are  $k_{1,\gamma_+} = -1.28$ ,  $k_{1,\gamma_-} = -1.81$  for the leader and  $k_{2,\gamma_+} = -0.32$ ,  $k_{2,\gamma_-} = -0.73$  for the follower vehicle. The speeds of both vehicles are more affected by downhill slopes than uphill slopes, and this effect is more pronounced on the leader vehicle, since it is significantly heavier than the follower vehicle. Uphill

slopes have little effect on the follower vehicle speed, which could also be observed in experiment data (Figure 5.1).

Several structures of the neural network were tested, and best results were acquired using a neural network with two hidden layers with five and three nodes and hyperbolic tangent sigmoid activation functions. This neural network is shown in Figure 5.4. The output of the neural network is a nonlinear function of its inputs,  $y_\xi^{\text{out}}(j) = f_{w,\xi}(y_\xi^{\text{in}}(j))$ , parametrized by its weight matrices  $W_\xi^{(l)}$ ,  $l = 1, 2, 3$ , which are trained using a back-propagation algorithm. The  $j$ -th sample input and target data for both neural networks are

$$y_\xi^{\text{in}}(j) = [v_{\xi,j-1} - u_\xi \quad \gamma_{j-1}]^\top,$$

$$y_\xi^{\text{out}}(j) = \frac{v_{\xi,j}^2 - v_{\xi,j-1}^2}{2L} + g\gamma_{j-1}.$$

By adopting this simple model, we assume that the behaviour of the vehicles only depends on local road topography. This allows us to use this model on any road segment whose topography is represented in the training data. Since highways in general follow similar topographic guidelines, most highways should be covered, except for road segments with long uphill or downhill slopes, which were not present in the training data. To enable generalization to these road segments, more data would need to be collected by running more experiments on different roads.

The training data from all three desired speed pair scenarios ( $u_1, u_2$ ) was considered together, excluding data points if the distance between the vehicles is smaller than 200 m, vehicle speed differs from the goal speed by more than 10 km/h or the distance from the start is less than 200 m. These data points are excluded in order to avoid speed changes that occur during the final platoon merge maneuver or if the vehicle is forced to brake, as well as to give the follower vehicle enough time to reach its goal speed. Finally, to reduce computational effort, the trained

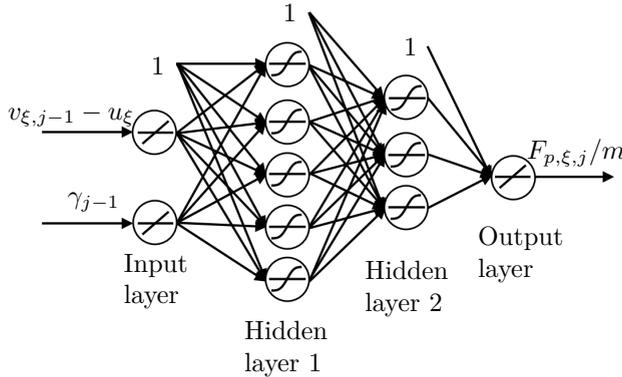


Figure 5.4: Structure of  $F_p/m$  neural networks.

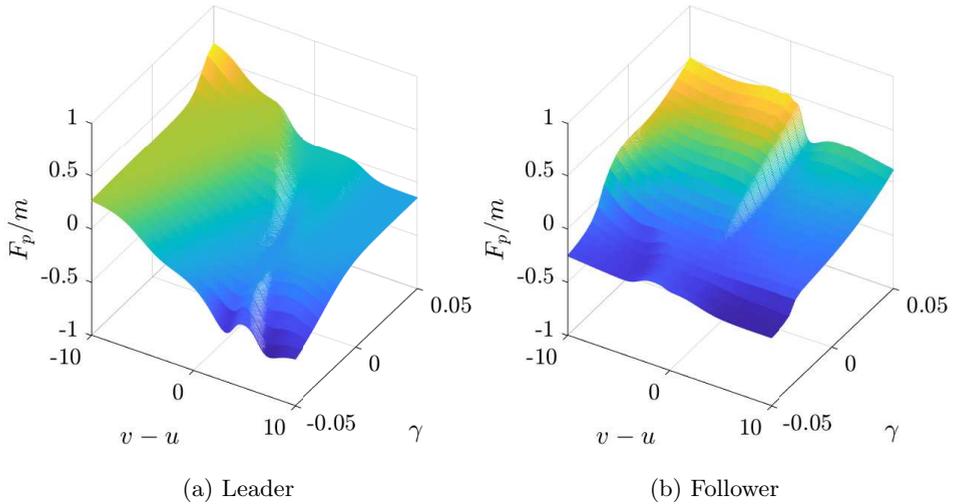


Figure 5.5:  $F_p/m$  as a function of  $v - v_{\text{ref}}$  and  $\gamma$  for the leader and the follower vehicle.

neural networks are implemented as look-up tables. Values of  $F_{p,\xi}/m_\xi$  are shown in Figure 5.5. In general, applied propulsive force will increase with road grade and vehicle speed deviation. This increase is faster around the origin ( $v_\xi \approx u_\xi, \gamma \approx 0$ ) and it gets slower for larger speed discrepancies and road grades because the engine power is limited.

### 5.2.3 Experimental results

Platoon merge distance prediction based on the two proposed vehicle speed prediction methods is evaluated using the test data set. The comparison is summed up in Table 5.1 and box plots of relative errors are shown in Figure 5.6. The relative error is defined as the ratio between the distance prediction error and the actual platoon merging distance,

$$e_m^{\text{rel}} = \frac{\hat{d}_m(t) - d_m(t)}{d_m(t)}.$$

Also shown are naive estimates according to (5.6), assuming vehicle speed is constant. We can see that the neural network based approach shows consistently better results, with the smallest root mean square error and standard deviation.

Once the future speed profile is predicted, it is easy to adopt some empirical criterion for recalculating the platoon merge distance predictions. This enables us to only recalculate speed profile predictions when the measured speed deviates from its predicted value due to some disturbances or model mismatch, instead of recal-

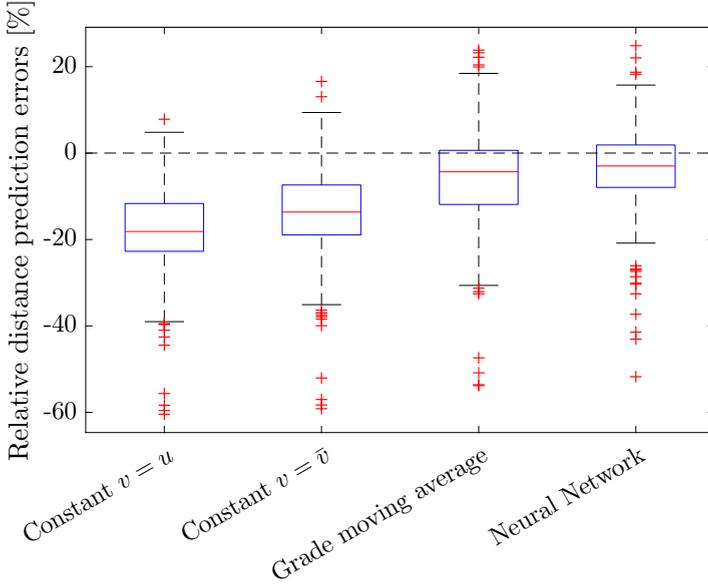


Figure 5.6: Box plots of relative platoon merge distance prediction errors.

Table 5.1: Comparison between the predicted merge distance errors for different speed prediction models.

	Constant $v = u$		Constant $v = \bar{v}$		Grade mov. avg.		Neural network	
	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
(75,85)	1492.65	704.58	1275.65	700.42	814.51	700.49	678.65	579.22
(75,89)	1386.28	948.91	1289.35	952.21	1060.23	956.22	865.83	829.03
(80,89)	1658.93	837.86	1287.53	870.10	975.07	861.52	835.46	786.28
Total	1516.22	855.45	1284.41	851.30	959.11	846.51	800.49	741.33

culating them periodically. The results of applying one such recalculation criterion for one test run are shown on Figure 5.7. Here, recalculations were done at most once per 400 m, when speed deviations are more than 3 km/h. The speed of the follower vehicle will be recalculated twice, once at  $x_2 = 600$  m and another time at  $x_2 = 1020$  m. We can see that recalculating the speeds improves the platoon merging distance prediction, from approximately 393 m (4.12% of the current remaining distance) at the start of the test run to 170 m (1.9%) after 600 m, and down to 70 m (0.8%) after another 420 m.

The neural network model predicts nominal vehicle speeds reasonably well in nominal conditions (Figure 5.3). However, the vehicles will often deviate from their nominal behaviour, resulting in larger discrepancies between the predicted and actual speed and causing outliers in merging distance prediction. Most often, we

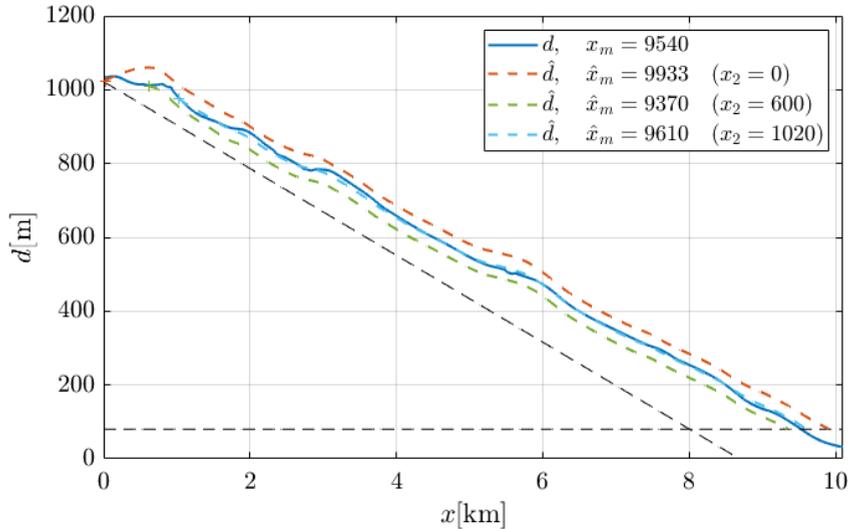


Figure 5.7: Recalculated merge distance predictions. The diagonal dashed black line shows the platoon merging phase if both vehicles would follow their reference speeds, and the horizontal dashed black line indicates distance  $d_p$ . The coloured dashed lines show the prediction of the distance between the vehicles, calculated when  $x_2 = 0$ ,  $x_2 = 600$  and  $x_2 = 1020$ .

cannot be sure what caused the deviation. In a number of test runs, the cruise control goal speeds were set wrong, and a vehicle drove slower or faster than intended. The nominal downhill speed control offset, set to 5 km/h, was exceeded in some test runs (clearly visible on Figure 5.3), and in some other test runs, the offset was reduced to 3 km/h. Apart from these situations, the traffic conditions are the most likely cause of larger deviations from nominal vehicle behaviour, especially when the nominal speed of the vehicles was close to the speed limit.

The box plots (Figure 5.6) show that the mean error for all methods is negative, i.e., all methods on average predict that the platoon will merge sooner than it actually does. The neural network speed model gives the smallest median and mean relative errors,  $-3\%$  and  $-4\%$ , respectively. In general, the influence of the surrounding traffic conditions on the trucks is hard to see from truck speed measurements when the speed of the truck is much lower than the average speed on the road. In the first test-scenario, the follower vehicle was driving with nominal speed of 85 km/h, while the speed limit on the road was 100 km/h, and the road grade was the main cause of its speed deviation. In two other scenarios, the influence of traffic conditions was much more apparent, resulting in larger root mean square errors (835.46 and 865.83 versus 678.65).

### 5.3 Energy-optimal platoon merging in traffic

The second sub-problem considered in this chapter is calculating energy-optimal speeds for vehicles attempting to merge to a platoon. These speeds will be given as reference speeds  $u_1$  and  $u_2$  that the vehicles will attempt to follow. We will assume these speeds are constant, unless something in the environment changes from the time they were calculated. Whereas in the previous section, the focus was solely on the merging phase, we now also have to take into account the platooning phase, during which the vehicles drive together and achieve fuel savings through air drag reduction.

We denote by  $X_f$  the end of the common road segment, i.e., the position at which the platoon will split, with vehicles continuing their separate ways afterwards. Without loss of generality, we will set  $d_p$  to zero in this section. Although in reality, vehicles might have some degree of flexibility with regard to timing, by only considering one fixed  $t_f$ , we ensure that the comparison between different pairs of speeds  $(u_1, u_2)$  is fair. In order to negate the ostensible energy saving by simply reducing the speed of a vehicle, in turn causing it to be delayed, we assume that both vehicles need to be at position  $X_f$  at some specified time  $t_f$ . This is trivially satisfied in case the vehicles did form a platoon, but even if we chose for the vehicles not to attempt to merge and form a platoon, we can use this assumption while calculating optimal speeds. While calculating the optimal merging speeds at time instant  $t$ , we will be using the relative coordinate system similar to the one used in the previous section, and

$$\begin{aligned}\tau_f(t) &= t_f - t, \\ d_f(t) &= X_f - x_2(t).\end{aligned}$$

For readability, we omit writing  $t$  wherever this time instant is irrelevant or obvious. If the calculated speeds are such that the vehicles merge into a platoon very close to  $X_f$ , we will know that is not beneficial to attempt forming a platoon, and the vehicles can proceed driving according to their own plans.

Another way of dealing with this issue is by including delay into the cost function. However, this would lead to a more complicated optimization problem and necessitate ad-hoc combination of two heterogeneous terms. To keep the optimization problem consistent and simple, we will therefore use the former approach and take  $t_f$  so that it satisfies the most stringent constraints the two vehicles have.

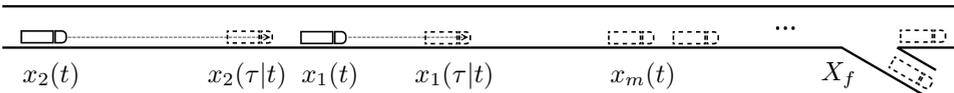


Figure 5.8: Platoon merging problem.

### 5.3.1 Energy-optimal merging problem

We focus on reducing the total work required to overcome the resistive forces acting on the vehicles. The three major external forces acting on vehicles are air drag, rolling resistance and gravity. Since we are generating reference speed profiles, we assume that road grade is zero. In reality, varying road grade will be handled by some form of look-ahead control. This assumption allows us to focus solely on reducing air drag, since the contribution of rolling resistance will be the same whether or not the vehicles adjust their speeds and attempt to merge into a platoon.

Ideally, the metric that we would like to use to evaluate the optimality of chosen merging speeds would be fuel consumption. However, getting accurate fuel consumption models can be very difficult, and will depend on the properties of the vehicles in question. Instead, we focus on reducing the total work required to overcome the resistive forces acting on the vehicles, which yields more general results.

Based on the air drag model described in 5.1, the cost function related to this component of the overall resistive force can therefore be written

$$J = \int_0^{\tau_m} v_1^3(\tau) + v_2^3(\tau) d\tau + \phi \int_{\tau_m}^{\tau_f} v_p^3(\tau) d\tau, \quad (5.12)$$

where by  $\phi$  we denote the total air drag reduction ratio of the platoon. In order for the vehicles to obey the timing and platoon merge constraints, we require that

$$\int_0^{\tau_m} v_2(\tau) - v_1(\tau) d\tau = d_0, \quad (5.13a)$$

$$\int_0^{\tau_m} v_1(\tau) d\tau + \int_{\tau_m}^{\tau_f} v_p(\tau) d\tau = d_f - d_0, \quad (5.13b)$$

$$\int_0^{\tau_m} v_2(\tau) d\tau + \int_{\tau_m}^{\tau_f} v_p(\tau) d\tau = d_f, \quad (5.13c)$$

$$u^{\min} \leq v_1 \leq u^{\max}, \quad (5.13d)$$

$$u^{\min} \leq v_2 \leq u^{\max}, \quad (5.13e)$$

$$u^{\min} \leq v_p \leq u^{\max}. \quad (5.13f)$$

Here constraint (5.13a) ensures that  $x_1(\tau_m) = x_2(\tau_m)$ , constraints (5.13b) and (5.13c) ensure that  $x_1(\tau_f) = X_f$  and  $x_2(\tau_f) = X_f$ , and constraints (5.13d), (5.13e) and (5.13f) give the admissible ranges for speeds  $v_1$ ,  $v_2$  and  $v_p$ .

Assume first that the vehicles were able to follow their desired speeds,  $v_\xi(\tau) = u_\xi$  and  $v_p(\tau) = u_p$ . We will denote the platoon merging time, position, and distance

for this case by  $\tau_{m_0}$ ,  $x_{m_0}$ , and  $d_{m_0}$ , respectively. Then, consistently with the conditions (5.13a)-(5.13c),  $\tau_{m_0}$  and  $x_{m_0}$  are given by (5.3), and the speed of the merged platoon is

$$v_p(\tau) = u_p = \frac{X_f - x_{m_0}}{\tau_f - \tau_{m_0}}.$$

Then the cost function (5.12) becomes

$$J_0(u) = (u_1^3 + u_2^3)\tau_{m_0} + \phi \frac{(X_f - x_{m_0})^3}{(\tau_f - \tau_{m_0})^2}.$$

This cost function is parametrized by  $d_0$ ,  $d_f$ ,  $\tau_f$  and  $\phi$ , and we seek to minimize it by choice of  $u_1$  and  $u_2$ .

Denote by  $u_{10} = \frac{d_f - d_0}{\tau_f}$  and  $u_{20} = \frac{d_f}{\tau_f}$  the constant speeds individual vehicles should keep in order to reach  $X_f$  at  $t_f$ , and assume this is possible without violating the constraints on minimum and maximum speed. Note that, although possible, it will never be beneficial for the leader to go faster than  $u_{10} \leq u^{\max}$  nor for the follower to go slower than  $u_{20} \geq u^{\min}$ . Therefore, we can further tighten the constraints to  $u^{\min} \leq u_1 \leq u_{10}$  and  $u_{20} \leq u_2 \leq u^{\max}$ . The minimization problem that we solve to calculate optimal  $u_1$  and  $u_2$  then becomes

$$\begin{aligned} & \underset{u_1, u_2}{\text{minimize}} && \frac{(u_1^3 + u_2^3)d_0}{u_2 - u_1} + \phi \frac{(d_f(u_2 - u_1) - u_2 d_0)^3}{(u_2 - u_1)(\tau_f(u_2 - u_1) - d_0)^2} \\ & \text{subject to} && u_{\min} \leq u_1 \leq u_{10} \\ & && u_{20} \leq u_2 \leq u_{\max} \end{aligned} \tag{5.14}$$

This is a convex problem and can easily be solved numerically.

However, since we assumed no interference from traffic and other extraneous factors that might render it impossible for the follower vehicle to maintain its optimal speed, we might see discrepancies in behaviour that will make this solution suboptimal. In Figure 5.9 we see the calculated optimal merging speeds and the average traffic speed upstream of the leader vehicle calculated in the moving bottleneck framework. The optimal speeds are given as a function of  $\phi$ , ranging from 1 (the follower vehicle in a platoon experiences no air drag) to 2 (platooning does not reduce air drag at all). We can see that even for  $\phi \approx 1.8$ , we have  $u_2 > \mathcal{V}(r_c(\sigma_\xi, u_1))$  ( $u_2$  greater than the average traffic speed upstream of the leader vehicle), so the follower vehicle will not be able to maintain its optimal merging speed in face of congestion caused by the leader, and this discrepancy will cause the actual platoon merge to occur later. Consequently, the energy savings will be lower than expected and suboptimal, further motivating including the traffic conditions in the optimization problem.

### 5.3.2 The influence of traffic

We assume we can split the initial traffic conditions between them into two zones,  $\rho(x, 0) \approx \rho_f$ ,  $x < x_c$  and  $\rho(x, 0) \approx \rho_c$ ,  $x > x_c$ , where  $x_c$  is the minimal  $x$  for which

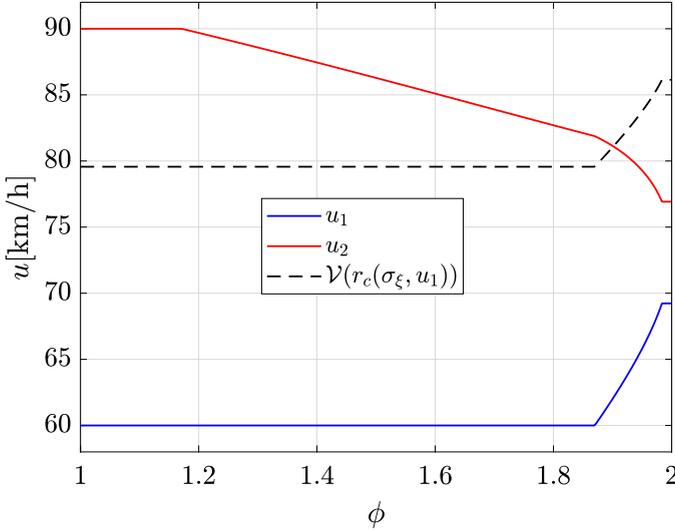


Figure 5.9: Optimal merging speeds for leader and follower vehicles and average traffic speed upstream of the leader vehicle.

$\mathcal{V}(\rho(x_c, 0)) < u^{\max}$ . We calculate  $\rho_f$  and  $\rho_c$  as average values of  $\rho$  on  $[x_2, x_c]$  and  $[x_c, x_1]$  respectively. If  $\mathcal{V}(\rho(x, 0)) \geq u_{\max}$  for all  $x \in [x_2, x_1]$ , we set  $x_c = x_1$  and  $\rho_c = \sigma$ .

The follower vehicle can only be slowed down when it enters the zone of density  $\rho_c$ , or the zone of density  $r_c(\sigma_\xi, u_1)$ , originating from the leader vehicle. In further text, we denote  $r_f(\sigma_\xi, u_2)$  as simply  $r_f$  and  $r_c(\sigma_\xi, u_1)$  as  $r_c$ .

The solution to thus described composite Riemann problem for  $\tau \approx 0$  is given by

$$\rho(x, t + \tau) = \begin{cases} r_f, & x < x_2 + \lambda_{r_f \rho_f} \tau, \\ \rho_f, & x_2 + \lambda_{r_f \rho_f} \tau < x < x_c + \lambda_{\rho_f \rho_c} \tau, \\ \rho_c, & x_c + \lambda_{\rho_f \rho_c} \tau < x < x_1 - W\tau, \\ r_c, & x_1 - W\tau < x, \end{cases}$$

where  $\lambda_{\rho_- \rho_+} = \Lambda(\rho_-, \rho_+)$  is the Rankine-Hugoniot transition speed, and  $W$  is the congestion wave speed, assuming triangular flux function. This solution is valid until the first front interaction, when either the zone of density  $\rho_f$  disappears,

$$x_2 + \lambda_{r_f \rho_f} \tau = x_c + \lambda_{\rho_f \rho_c} \tau \quad (5.15)$$

or the zone of density  $\rho_c$  disappears,

$$x_c + \lambda_{\rho_f \rho_c} \tau = x_1 - W\tau. \quad (5.16)$$

We denote the solution in  $\tau$  to (5.15)

$$\tau_{\rho_f}(u) = \frac{x_c - x_2}{\lambda_{r_f\rho_f} - \lambda_{\rho_f\rho_c}},$$

and the solution in  $\tau$  to (5.16)

$$\tau_{\rho_c}(u) = \frac{x_1 - x_c}{\lambda_{\rho_f\rho_c} + W}.$$

The times of following front interactions will be delineated by noting the order in which the zones of particular density vanished. For example,  $\tau_{\rho_c\rho_f}$  denotes the time at which zone of density  $\rho_f$  vanishes in second front interaction, after zone of density  $\rho_c$  vanished in first front interaction.

The rest of the front interaction times are given by

$$\begin{aligned} \tau_{\rho_f r_f} &= \tau_{\rho_f} \frac{\lambda_{r_f\rho_f} - \lambda_{r_f\rho_c}}{u_2 - \lambda_{r_f\rho_c}}, \\ \tau_{\rho_f\rho_c} &= \frac{d_0 + \tau_{\rho_f}(\lambda_{r_f\rho_c} - \lambda_{r_f\rho_f})}{\lambda_{r_f\rho_c} + W}, \\ \tau_{\rho_c\rho_f} &= \frac{d_0 - \tau_{\rho_c}(\lambda_{\rho_f r_c} + W)}{\lambda_{r_f\rho_f} - \lambda_{\rho_f r_c}}, \\ \tau_{\rho_c r_c} &= \tau_{\rho_c} \frac{\lambda_{\rho_f r_c} + W}{\lambda_{\rho_f r_c} - u_1}, \\ \tau_{t_1} = \tau_{\rho_f r_f\rho_c} &= \frac{d_0 + \tau_{\rho_f r_f}(v_{\rho_c}(u_2) - u_2)}{v_{\rho_c}(u_2) + W}, \\ \tau_{t_2} = \tau_{\rho_f\rho_c r_f} &= \frac{d_0 - \tau_{\rho_f\rho_c}(\lambda_{r_f r_c} + W)}{u_2 - \lambda_{r_f r_c}}, \\ \tau_{t_3} = \tau_{\rho_c\rho_f r_f} &= \tau_{\rho_c\rho_f} \frac{\lambda_{r_f\rho_f} - \lambda_{r_f r_c}}{u_2 - \lambda_{r_f r_c}}, \\ \tau_{\rho_c\rho_f r_c} &= \frac{d_0 + \tau_{\rho_c\rho_f}(\lambda_{r_f r_c} - \lambda_{r_f\rho_f})}{\lambda_{r_f r_c} - u_1} \end{aligned}$$

From the standpoint of cost function, there are four cases of traffic we need to consider based on the ordering of front interactions:

- **Case 0:** The follower is unaffected by traffic. This case typically happens in light traffic, when the zone of density  $\rho_c$  vanished first,  $\tau_{\rho_c} < \tau_{\rho_f}$ , and the zone of density  $r_c$  vanishes before the zone of density  $r_f$ , with either  $\tau_{\rho_c\rho_f r_c} < \tau_{\rho_c\rho_f r_f}$  or  $\tau_{\rho_c r_c} < \tau_{\rho_c\rho_f}$ .

This case was already discussed and corresponds to using  $J_0$  as the cost function.

- **Case 1:** The zone of density  $\rho_f$  vanishes first,  $\tau_{\rho_f} < \tau_{\rho_c}$ , then the zone of density  $r_f$ ,  $\tau_{\rho_f r_f} < \tau_{\rho_f \rho_c}$ . The follower vehicle first enters the zone of traffic of density  $\rho_c$ , and then of traffic density  $r_c$  at time  $\tau_{t_1}$ .
- **Case 2:** The zone of density  $\rho_f$  vanishes first,  $\tau_{\rho_f} < \tau_{\rho_c}$ , then the zone of density  $\rho_c$ ,  $\tau_{\rho_f r_f} > \tau_{\rho_f \rho_c}$ . The follower vehicle only enters the zone of traffic density  $r_c$  at time  $\tau_{t_2}$ .
- **Case 3:** The zone of density  $\rho_c$  vanishes first,  $\tau_{\rho_f} > \tau_{\rho_c}$ , and the zone of density  $r_f$  vanishes before the zone of density  $r_c$ , with  $\tau_{\rho_c \rho_f r_c} > \tau_{\rho_c \rho_f r_f}$  and  $\tau_{\rho_c r_c} > \tau_{\rho_c \rho_f}$ . The follower vehicle only enters the zone of traffic density  $r_c$  at time  $\tau_{t_3}$ .

The latter three cases are shown on Figure 5.10. We denote the speed the follower vehicle maintains in traffic of density  $\rho$ ,  $v_\rho(u_2) = \min(u_2, \mathcal{V}(\rho))$ . Finally, for traffic cases  $i = 1, 2, 3$ , the platoon merge will occur when the fronts corresponding to the leader and the follower vehicle intersect, at

$$\tau_{m_i} = \frac{d_0 - d_{t_i} + v_{r_c}(u_2)\tau_{t_i}}{v_{r_c}(u_2) - u_1}$$

$$x_{m_i} = x_1 + u_1\tau_{m_i}$$

where

$$d_{t_1} = d_0 - W\tau_{\rho_f r_f \rho_c},$$

$$d_{t_2} = u_2\tau_{\rho_f \rho_c r_f},$$

$$d_{t_3} = u_2\tau_{\rho_c \rho_f r_f},$$

are the positions where the follower vehicle enters the traffic zone of density  $r_c$ .

Under the stated assumptions, the cost function (5.12) can be written as

$$J(u) = \begin{cases} J_0(u), & \tau_{\rho_f} > \tau_{\rho_c}, (\tau_{\rho_c \rho_f r_c} < \tau_{\rho_c \rho_f r_f} \vee \tau_{\rho_c r_c} < \tau_{\rho_c \rho_f}), \\ J_1(u), & \tau_{\rho_f} < \tau_{\rho_c}, \tau_{\rho_f r_f} < \tau_{\rho_f \rho_c}, \\ J_2(u), & \tau_{\rho_f} < \tau_{\rho_c}, \tau_{\rho_f r_f} > \tau_{\rho_f \rho_c}, \\ J_3(u), & \tau_{\rho_f} > \tau_{\rho_c}, \tau_{\rho_c \rho_f r_f} < \tau_{\rho_c \rho_f r_c}, \tau_{\rho_c \rho_f} < \tau_{\rho_c r_c}, \end{cases}$$

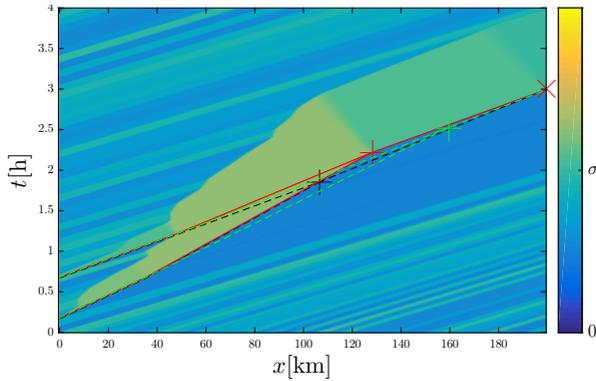
where

$$J_1(u) = u_1^3\tau_{m_1} + u_2^3\tau_{\rho_f r_f} + u_{\rho_c}^3(\tau_{t_1} - \tau_{\rho_f r_f}) + u_{r_c}^3(\tau_{m_1} - \tau_{t_1}) + \frac{(X_f - x_{m_1})^3}{(\tau_f - \tau_{m_1})^2},$$

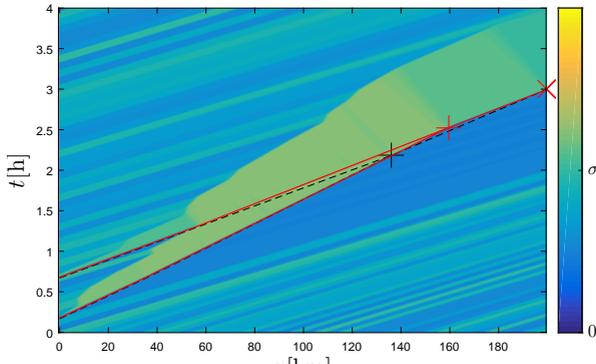
$$J_2(u) = u_1^3\tau_{m_2} + u_2^3\tau_{t_2} + u_{r_c}^3(\tau_{m_2} - \tau_{t_2}) + \frac{(X_f - x_{m_2})^3}{(\tau_f - \tau_{m_2})^2},$$

$$J_3(u) = u_1^3\tau_{m_3} + u_2^3\tau_{t_3} + u_{r_c}^3(\tau_{m_3} - \tau_{t_3}) + \frac{(X_f - x_{m_3})^3}{(\tau_f - \tau_{m_3})^2}.$$

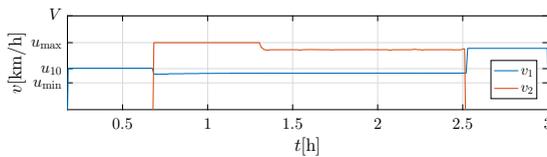
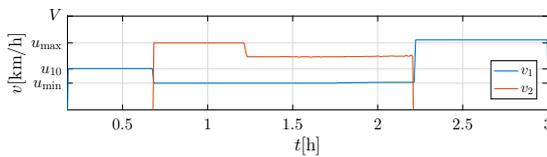




(a) Control law (2)

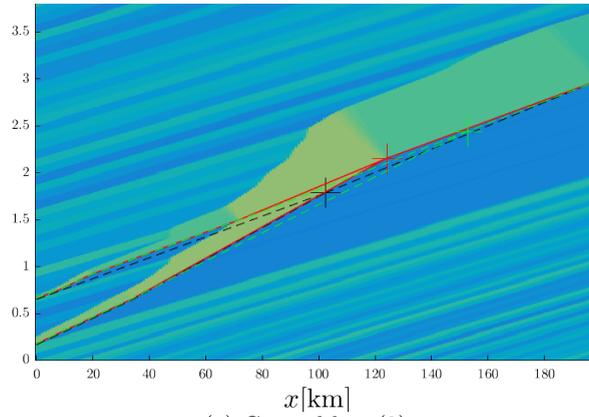


(b) Control law (3)

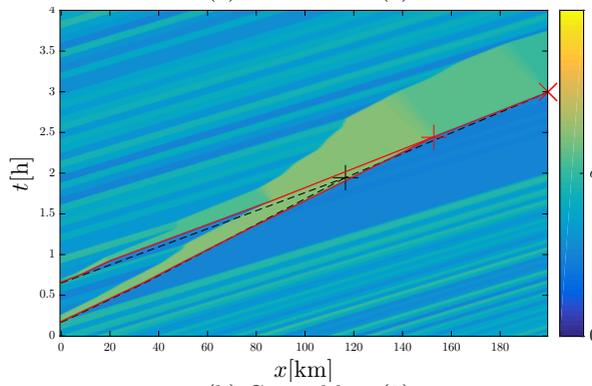


(c) Speed profiles under control law (2) (top) and (3) (bottom)

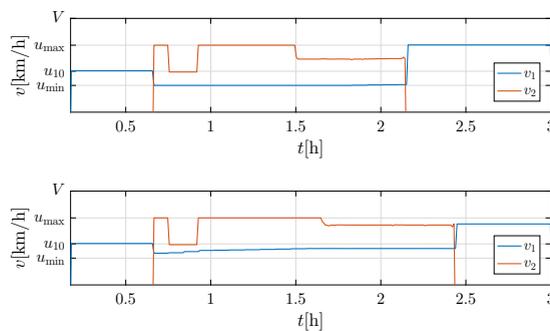
Figure 5.11: Traffic situation (subfigures (a) and (b)) and speed profiles (c) of one simulation run with no disturbances. Attempting to form a platoon using control law (2) led to an increase in energy cost by 0.71636%, whereas the control law (3) achieved energy savings of 0.50955%.



(a) Control law (2)



(b) Control law (3)



(c) Speed profiles under control law (2) (top) and (3) (bottom)

Figure 5.12: Traffic situation (subfigures (a) and (b)) and speed profiles (c) of one simulation run with follower speed disturbance. Attempting to form a platoon using control law (2) led to an increase in energy cost by 0.48012%, whereas the control law (3) achieved energy savings of 0.32183%.

### 5.3.3 Simulation results

Finally, we test the derived control laws in simulations. The metric we will be using is the percentage of energy saved, according to (5.12), compared to the case the vehicles would drive at a constant speed and arrive at  $x_f$  at time  $t_f$ . In total a 100 simulations were executed for each control law and traffic density range.

The simulation scenario in question is as follows:

1.  $t < t_{10}$ : The traffic is in free flow, with heterogeneous traffic density. The leader vehicle enters the road segment at  $t = t_{10}$ .
2.  $t_{10} \leq t < t_{20}$ : The leader vehicle travels at speed  $v_{01}$ , at which it would reach  $x_f$  at time  $t_f$ . The follower vehicle enters the road segment at  $t = t_{20}$ .
3.  $t_{20} \leq t < t_m$ : The leader and the follower adjust their speeds according to the specified control law, until they merge into a platoon.
4.  $t \geq t_m$ : The newly merged platoon proceeds and adjusts its speed so that it reaches  $X_f$  at time  $t_f$ .

If the platoon merging fails for any of the control laws, or if the vehicles violate the timing constraint (not arrive at  $x_f$  by  $t_f$ ), the vehicles proceed at their maximum speed until the end of the segment, and that simulation run is not counted in average cost calculations.

The initial background traffic conditions  $\rho_i(0)$  and inflow into the first cell  $q_0(t)$  are randomly generated heterogeneous free flow. We used three scenarios with different traffic density ranges, light traffic  $[\sigma/5, \sigma]$ , medium traffic  $[\sigma/3, \sigma]$  and heavy traffic  $[\sigma/2, \sigma]$ , resulting in average traffic density of  $0.6\sigma$ ,  $0.66\sigma$  and  $0.75\sigma$  respectively.

We are comparing three different control laws:

- **Control law (1):** The optimal reference speeds are calculated by solving (5.14) once at  $t = t_{20}$ , ignoring traffic conditions.
- **Control law (2):** The optimal reference speeds are calculated by solving (5.14) periodically during the merging phase, ignoring traffic conditions,
- **Control law (3):** The optimal reference speeds are calculated by solving (5.17) periodically during the merging phase, taking traffic conditions into account.

We also considered the case when some disturbance is acting on the vehicles. Namely, at a random time instant between 5 and 10 minutes after the merging has begun, we decrease the speed of the follower vehicle by 20km/h for 10 minutes. This delays the platoon merge and would result in lower energy savings, so recalculating optimal speeds is required. Examples of simulation runs without and with follower speed disturbance are shown in Figure 5.11 and Figure 5.12. Vehicles forming a

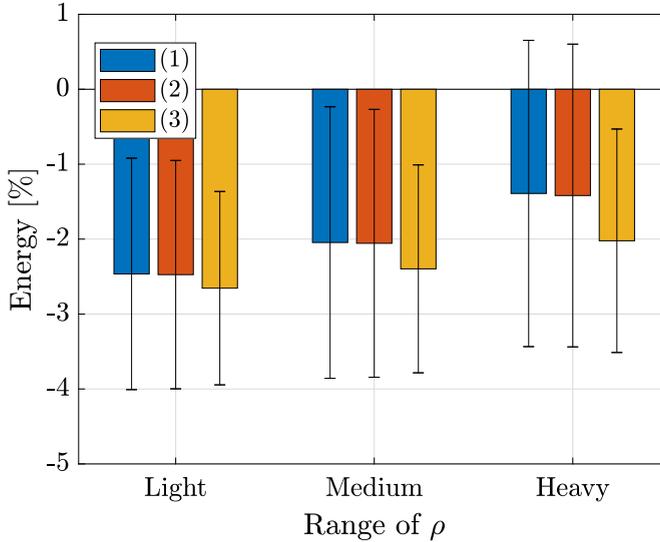


Figure 5.13: Relative energy consumption change with no disturbances.

$\rho$	(1)	(2)	(3)
Light	-2.4642%	-2.4741%	-2.6547%
Medium	-2.0452%	-2.0559%	-2.3969%
Heavy	-1.3909%	-1.4183%	-2.0220%

platoon are shown in traffic plots by red lines. Dashed black lines represent merging vehicles' trajectories if they would not be slowed down by the traffic, and dashed green lines in traffic plots for control law (2) represent the trajectories that would be followed if control law (3) was used instead. Crosses mark the position and time at which the platoon merge occurred.

The average energy savings are shown in Figure 5.13 and Figure 5.14. We can see that using control laws that take traffic conditions into consideration improves average energy savings, especially in heavy traffic.

In addition to calculating optimal merging speeds, this approach can also be used to predict when attempting to form a platoon is not beneficial. In Table 5.2 we see the number of “bad platooning attempts”, i.e. in how many simulation runs the vehicles failed to reach the goal position in time, or had higher overall energy cost. If the calculated optimal merging speeds are such that platoon merge is predicted to occur very close to the end of the common road segment  $x_f$ , we know that attempting to form a platoon will not yield fuel savings, and may abandon the attempt at the start, instead continuing driving at vehicles' own optimal speeds. We

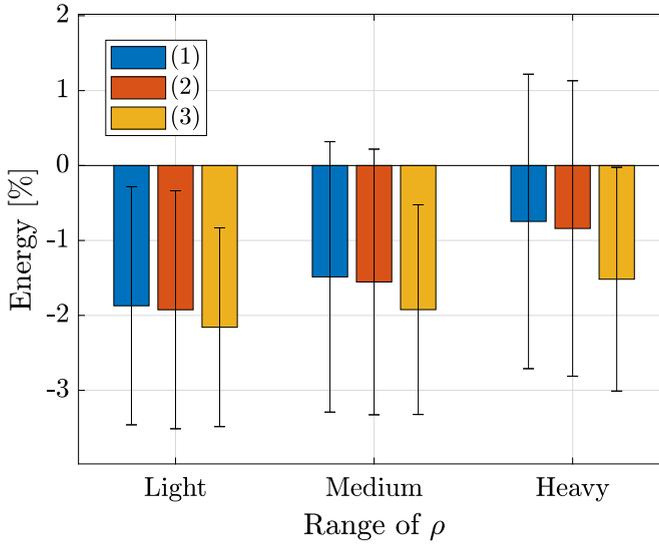


Figure 5.14: Relative energy consumption change with a disturbance.

$\rho$	(1)	(2)	(3)
Light	-1.8712%	-1.9245%	-2.1577%
Medium	-1.4860%	-1.5536%	-1.9228%
Heavy	-0.7458%	-0.8396%	-1.5171%

also see the number of times the algorithm correctly predicted this outcome (true positive), which could be used to pre-emptively abort the platooning attempt, as well as the number of times the bad outcome was not predicted (false negative) and falsely predicted (false positive).

Table 5.2: Bad platooning attempt prediction.

	Light	Medium	Heavy
Total bad attempts	5	15	23
Correctly predicted	5	14	19
Not predicted	0	1	4
Falsely predicted	6	3	1

## 5.4 Summary

Optimal platoon formation coordination is a complex problem, even when considering only a single pair of vehicles attempting to merge and form a platoon. In this case, the difficulties come from a large number of potential disturbances from the environment that can disrupt the process.

Selecting correct merging speeds for the vehicles is important because failing to do so might result in higher fuel consumptions. Even deciding whether the vehicles should attempt to platoon at all should be influenced by what kind of disturbances from the environment we might experience.

It is therefore important to have a good prediction model, for the merging vehicle trajectories in the merging process. Using this model, we tested different merging speed pairs and select the optimal one. Having predicted vehicle trajectories also enabled us to detect when the vehicles are deviating from the plan, and take appropriate actions to correct them.

It is clear that traffic conditions can play a significant role in the platoon merging phase, and that they cannot be ignored if we want to make a good prediction. By using the moving bottleneck framework, we were able to anticipate the effect the congestion formed upstream of the leader vehicle will have on the follower vehicle. This allowed us to calculate energy-optimal merging speeds, as well as to decide when platooning should be attempted.

## Chapter 6

# Bottleneck decongestion

**B**OTTLENECKS, such as lane drops, work zones, accident sites, or on-ramps, are often the main reason for congestion build-up and traffic flow disruption. For this reason, they are one of the main areas where traffic control is applied. Once the traffic flow arriving at the bottleneck exceeds what the bottleneck can handle in free flow, traffic breakdown ensues and congestion starts building up. This is mainly due to the arriving traffic flow being higher than the traffic flow discharging from the bottleneck. The situation is exacerbated by the capacity drop phenomenon, which manifests as a reduction in the traffic flow through the bottleneck once traffic breakdown happens, further reducing the discharging flow from the queue and representing a loss of efficiency for the traffic system. An additional difficulty in dealing with this phenomenon is in the fact that the traffic breakdown is known to happen stochastically. It may arise for an arbitrary value of traffic flow in some range, depending on the microscopic interactions of the vehicles changing lanes at their approach to the bottleneck.

Bottleneck decongestion has long been tackled by classical traffic control measures, such as ramp metering [60], and variable speed limits [62]. However, both of these control methods require additional fixed equipment to be installed upstream of the bottleneck. This limits their flexibility, especially when it comes to handling non-recurrent bottlenecks, such as work zones, accident sites etc. as it is not reasonable to assume the required equipment would be available wherever a bottleneck of that type arises. In this chapter, we propose an alternative approach to the problem, using platoons acting as controlled moving bottlenecks to restrict the traffic flow arriving at the bottleneck. By postponing and shaping the inflow to the bottleneck, we are able to keep it in free flow. This helps us avoid traffic breakdown and capacity drop, leading to significant reduction of total time spent of all vehicles.

The chapter starts with a more detailed analysis of the bottleneck decongestion problem in Section 6.1, which was outlined in Section 1.2.2. We then first tackle the case when the traffic breakdowns happen stochastically in Section 6.2. This

is achieved by creating controlled moving bottlenecks at desired points along the road, and then controlling them to achieve bottleneck decongestion and improve the throughput. Next, in Section 6.3 we design a more complex control law for the case when the traffic breakdowns are deterministic. This control law is based on the tandem queueing model with moving bottlenecks, coordinating the randomly arriving platoons to achieve a similar goal. Finally, the content of the chapter is summarized in Section 6.4.

## 6.1 Bottleneck decongestion problem

As outlined in Subsection 1.2.2, the congestion reduction problem, i.e., minimizing the total time spent, is equivalent to maximizing the early outflow from the road segment under consideration. The cost function to be maximized is

$$J_{\text{cr}} = \int_{t_0}^{t_{\text{end}}} (t - t_{\text{end}}) q^{\text{out}}(t) dt. \quad (6.1)$$

The outflow  $q^{\text{out}}(t)$  is determined by the traffic dynamics on the road, as well as by potential applied control actions, which will be designed to influence the traffic state so that 6.1 is maximized. Typically, the flow of the whole road is constrained by the most severe bottleneck, in which case we may focus on the stretch of road upstream of it, and call this problem bottleneck decongestion.

The simplest case of bottleneck decongestion considers a single bottleneck at the downstream end of the road, when the outflow from the road is equal to the outflow from the bottleneck,  $q^{\text{out}}(t) = q_b^{\text{out}}(t)$ . The queue length at the bottleneck is denoted  $n_b(t)$ , and it evolves based on the difference between the inflow to the bottleneck  $q_b^{\text{in}}(t)$ , and outflow from it  $q_b^{\text{out}}(t)$ ,

$$\begin{aligned} \dot{n}_b(t) &= q_b^{\text{in}}(t) - q_b^{\text{out}}(t), \\ q_b^{\text{out}}(t) &= \begin{cases} \min\{q_b^{\text{in}}(t), q_b^{\text{max}}\}, & m_b(t) = \text{FF}, \\ q_b^{\text{dis}}, & m_b(t) = \text{CD}, \end{cases} \end{aligned} \quad (6.2)$$

where  $m_b$  represents the current state of the bottleneck, with  $m_b = \text{FF}$  corresponding to the bottleneck in free flow, and  $m_b = \text{CD}$  to the congested bottleneck. We denote the maximum capacity of the bottleneck  $q_b^{\text{max}}$ , and the discharging flow from the bottleneck if it is congested  $q_b^{\text{dis}}$ . The outflow from the road therefore depends directly on the state of the bottleneck. Due to capacity drop, the discharging flow will be lower than the maximum capacity,  $q_b^{\text{dis}} < q_b^{\text{max}}$ . Note that such segment can be used as a building block for a more complex road network, e.g. with multiple stationary bottlenecks in cascade.

Assuming that the average inflow is below the bottleneck discharging flow  $q_{b,\text{avg}}^{\text{in}} < q_b^{\text{dis}}$ , the bottleneck queue will be stable. However, the potential variability

of the inflow can lead to traffic breakdowns and bottleneck becoming congested, delaying the outflow from the road. Let there be  $K_{\text{CD}}$  traffic breakdowns during the considered time frame  $t_0 \leq t \leq t_{\text{end}}$ , and let  $k$ -th traffic breakdown happen at time  $t = t_k^{\text{b}} \in (t_0, t_{\text{end}})$ , with  $m_{\text{b}}(t_k^{\text{b}}) = \text{CD}$  and  $q_{\text{b}}^{\text{in}}(t_k^{\text{b}}) > q_{\text{b}}^{\text{dis}}$ . After each traffic breakdown, the bottleneck returns to free flow at time  $t_k^{\text{c}}$  for which it holds that

$$\int_{t_k^{\text{b}}}^t q_{\text{b}}^{\text{in}}(t) dt > (t - t_k^{\text{b}}) q_{\text{b}}^{\text{dis}}, t \in (t_k^{\text{b}}, t_k^{\text{c}}), \quad \int_{t_k^{\text{b}}}^{t_k^{\text{c}}} q_{\text{b}}^{\text{in}}(t) dt = (t_k^{\text{c}} - t_k^{\text{b}}) q_{\text{b}}^{\text{dis}}. \quad (6.3)$$

Assuming the bottleneck is in free flow mode at  $t = t_{\text{end}}$ ,  $m_{\text{b}}(t_{\text{end}}) = \text{FF}$ , the congestion reduction cost function (6.1), can be written

$$J_{\text{cr}} = \int_{t_0}^{t_k^{\text{b}}} (t - t_{\text{end}}) q_{\text{b}}^{\text{in}}(t) dt + \sum_{k=1}^{K_{\text{CD}}} \left( \left( \frac{t^2}{2} - t_{\text{end}} t \right) \Big|_{t_k^{\text{b}}}^{t_k^{\text{c}}} q_{\text{b}}^{\text{dis}} + \int_{t_k^{\text{c}}}^{t_{k+1}^{\text{b}}} (t - t_{\text{end}}) q_{\text{b}}^{\text{in}}(t) dt \right),$$

where we set  $t_{K_{\text{CD}}+1}^{\text{b}} = t_{\text{end}}$ . For each traffic breakdown at the bottleneck, the TTS is increased by

$$\int_{t_k^{\text{b}}}^{t_k^{\text{c}}} (t - t_{\text{end}}) (q_{\text{b}}^{\text{dis}} - q_{\text{b}}^{\text{in}}(t)) dt > 0,$$

since  $t - t_{\text{end}} < 0$  and  $\int_{t_k^{\text{b}}}^{t_k^{\text{c}}} (q_{\text{b}}^{\text{dis}} - q_{\text{b}}^{\text{in}}(t)) dt < 0$ ,  $t \in (t_k^{\text{b}}, t_k^{\text{c}})$  due to (6.3), compared to the ideal case when the bottleneck is never congested.

Therefore, if we can by some means control the inflow to the bottleneck  $q_{\text{b}}^{\text{in}}(t)$ , the general control strategy for minimizing the TTS is the following:

- if the bottleneck is congested – return it to free flow as soon as possible, and
- if the bottleneck is in free flow – keep the flow as high as possible without causing a traffic breakdown.

This is consistent with the so-called “breakdown minimization principle” [220]. For this, we need to be able to regulate the traffic flow that reaches the bottleneck within some range  $[q^{\text{lo}}, q^{\text{hi}}]$ . Additionally, in order to be able to decongest the bottleneck, we need  $q^{\text{lo}} < q_{\text{b}}^{\text{dis}}$ , and in order to be able to improve the throughput without causing traffic breakdown, we need  $q_{\text{b}}^{\text{dis}} < q^{\text{hi}} \leq q_{\text{b}}^{\text{max}}$ .

More specifically, we study bottleneck decongestion using controlled moving bottlenecks as actuators, with their speed and severity as a control input. We assume the traffic state can be measured and observed using stationary sensors or connected vehicles. These moving bottlenecks are formed out of clusters of connected vehicles, which we will also refer to as *platoons*. By communicating with platoon  $\xi$ , we may change its reference speed  $u_{\xi}$  and formation, which in turn

affects the surrounding traffic and limits the traffic flow that can overtake it  $q_\xi^{\text{cap}}(t)$ . In general, reducing  $u_\xi$  and setting the platoon formation to take multiple lanes makes the platoon act as a more severe moving bottleneck. We assume that we can control the overtaking flow in some range  $[q^{\text{lo}}, q^{\text{hi}}]$ , e.g. the overtaking flow will be reduced if the platoon splits and drives side by side, taking multiple lanes, instead of taking only one lane. The maximum controllable overtaking flow  $q^{\text{hi}}$  corresponds to having the platoon take only one lane, leading to minimum moving bottleneck severity, and the minimum,  $q^{\text{lo}}$ , to having the platoon take the maximum number of lanes with maximum moving bottleneck severity. Therefore, we may use the platoons as controlled moving bottlenecks to control the inflow to the stationary bottleneck in order to decongest it and keep it in free flow, which increases the throughput.

While the behaviour of the bottleneck in either state is generally understood, there are conflicting paradigms about how the transition between them happens. In the remainder of this chapter, we will discuss the two main cases. We first focus on stochastic capacity and traffic breakdown, designing a control law that reacts to the changes in the bottleneck state. Then, we use a deterministic model to pro-actively coordinate platoons so that traffic breakdowns are avoided.

## 6.2 Bottleneck decongestion with stochastic traffic breakdown

In the vast majority of classic deterministic traffic models, it follows from the model structure that the capacity of a bottleneck  $q_b^{\text{max}}$  is constant for a given road geometry. Once traffic flow higher than the bottleneck capacity reaches the bottleneck, there is a traffic breakdown, and the congestion that is formed begins discharging from the bottleneck at a lower rate  $q_b^{\text{dis}} < q_b^{\text{max}}$ , due to the capacity drop phenomenon [37, 38, 39]. However, a plethora of empirical results points to traffic breakdown happening at the same bottleneck for different traffic conditions [40, 41, 42], interpreting the bottleneck capacity as a stochastic variable.

In this section, we tackle the bottleneck decongestion problem assuming that traffic breakdown happens stochastically. The proposed control law reacts to traffic breakdown, creating controlled moving bottlenecks that regulate the inflow to the stationary bottleneck in order to decongest it and keep it in free flow. The proposed reactive control law is then tested in simulation, using FTSM as the simulation model.

### 6.2.1 Stochastic traffic breakdown

A stationary bottleneck will be uncongested for low levels of incoming flow,  $q_b^{\text{in}}(t) < q_b^{\text{dis}}$ . As the flow increases, the probability of traffic breakdown will start increasing. The stochastic capacity of a bottleneck is given by specifying the probability of traffic breakdown within some time interval  $T$ , given the demand level  $q$ ,

as in [40]. The probability of traffic breakdown is taken to be Weibull-distributed,

$$F_B(T, q) = 1 - e^{-\frac{T}{T_0} \left(\frac{q}{k_0}\right)^{-m_0}},$$

where parameters  $T_0$ ,  $m_0$  and  $k_0$  are positive design parameters obtained from estimating the stochastic capacity of the bottleneck. Conversely, the time to breakdown is an exponentially distributed random variable parametrised by  $q$ ,  $\Theta_q \sim \exp(T_B(q)^{-1})$ . We denote by  $T_B(q)$  the mean time to breakdown, given as a function of the current traffic demand at the bottleneck,

$$T_B(q) = T_0 \left(\frac{q}{k_0}\right)^{-m_0}. \quad (6.4)$$

Therefore, depending on whether the stationary bottleneck is in free flow, or is congested, the dynamics of its queue length will follow

$$\begin{aligned} \dot{n}_b(t) &= q_b^{\text{in}}(t) - q_b^{\text{out}}(t), \\ q_b^{\text{out}}(t) &= \begin{cases} q_b^{\text{in}}(t), & m = \text{FF}, \\ q_b^{\text{dis}}, & m = \text{CD}, \end{cases} \end{aligned}$$

initialized at some time  $t_0$  with

$$\begin{aligned} n_b(t_0) &= 0, \\ \theta(t_0) &= \Theta_{q_b^{\text{in}}}, \\ q_b^{\text{old}}(t_0) &= q_b^{\text{in}}(t_0), \end{aligned}$$

where the time to breakdown  $\Theta_{q_b^{\text{in}}}$  is either zero, if  $q_b^{\text{in}} > q_b^{\text{max}}$ , or exponentially distributed otherwise,  $\Theta_{q_b^{\text{in}}} \sim \text{Exp}(T_B^{-1}(q_b^{\text{in}}(t_0)))$ , with  $T_B(q)$  given by (6.4). The mode of the bottleneck is determined by  $m \in \{\text{FF}, \text{CD}\}$ .

If the bottleneck is in free flow mode  $m = \text{FF}$ , and demand at its position stays  $q_b^{\text{in}}(t) = q_b^{\text{old}}$  longer than the time to breakdown  $\Theta_q$ , we say that there has been a traffic breakdown at the bottleneck, and its mode changes to capacity drop  $m = \text{CD}$ . If the demand at the bottleneck changes to  $q_b^{\text{in}}(t) \neq q_b^{\text{old}}$  before the time to breakdown has elapsed, we generate a new time to breakdown parametrised by the new demand  $\Theta_{q_b^{\text{in}}(t)}$ , update  $q_b^{\text{old}} = q_b^{\text{in}}(t)$ , and repeat the process. Once the bottleneck is active and  $m = \text{CD}$ , we consider it to have been decongested,  $m = \text{FF}$ , once its queue has been discharged and its inflow drops below  $q_b^{\text{in}}(t) < q_b^{\text{dis}}$ . Note that while technically a breakdown can happen for any  $q_b^{\text{in}}(t)$ , it is only necessary to consider the case when  $q_b^{\text{in}}(t) > q_b^{\text{dis}}$ , since otherwise the breakdown would immediately be resolved.

More formally, we may represent such queueing system as an open hybrid automaton with random reset map,

$$BN = (\mathcal{M}, \mathcal{X}, q_b^{\text{in}}, f, \mathcal{X}_0, \mathcal{D}, \mathcal{E}, \mathcal{G}, \mathcal{R}).$$

The automaton has two modes,  $\mathcal{M} = \{m\}$ , and  $m \in \mathbf{M} = \{\text{FF}, \text{CD}\}$ , and three continuous states,  $\mathcal{X} = \{n_b, \theta, q_b^{\text{old}}\}$ ,  $\mathcal{X} \in \mathbf{X} = \mathbb{R}^3$ . The input to the system  $q_b^{\text{in}}(t)$  is defined as a piecewise constant function of time with non-zero piece lengths, and the continuous dynamics are represented by

$$f(\text{FF}, \mathcal{X}, q_b^{\text{in}}) = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \quad f(\text{CD}, \mathcal{X}, q_b^{\text{in}}) = \begin{bmatrix} q_b^{\text{in}} - q_b^{\text{dis}} \\ -1 \\ 0 \end{bmatrix},$$

with domains over which each of these vector fields is valid given by

$$\begin{aligned} \mathcal{D}(\text{FF}, q_b^{\text{in}}) &= \{\mathcal{X} \in \mathbf{X} | n_b = 0, \theta \geq 0, q_b^{\text{old}} = q_b^{\text{in}}\}, \\ \mathcal{D}(\text{CD}, q_b^{\text{in}}) &= \{\mathcal{X} \in \mathbf{X} | n_b \geq 0\}. \end{aligned}$$

We initialize the bottleneck in free flow,  $\mathcal{X}_0 = [0 \quad \Theta_{q_b^{\text{in}}} \quad q_b^{\text{in}}]^\top$ , with  $\Theta_{q_b^{\text{in}}} = 0$  if  $q_b^{\text{in}} \geq q_b^{\text{max}}$ , or  $\Theta_{q_b^{\text{in}}} \sim \text{Exp}(T_B^{-1}(q_b^{\text{in}}))$  if  $q_b^{\text{in}} < q_b^{\text{max}}$ . There are three possible discrete transitions,  $\mathcal{E} = \{(\text{FF}, \text{FF}), (\text{FF}, \text{CD}), (\text{CD}, \text{FF})\}$ , which can occur when  $\mathcal{X}$  belongs to the guards

$$\begin{aligned} \mathcal{G}(\text{FF}, \text{FF}, q_b^{\text{in}}) &= \{\mathcal{X} \in \mathbb{R}^3 | n_b = 0, \theta \geq 0, q_b^{\text{old}} \neq q_b^{\text{in}}\}, \\ \mathcal{G}(\text{FF}, \text{CD}, q_b^{\text{in}}) &= \{\mathcal{X} \in \mathbb{R}^3 | n_b = 0, \theta \leq 0\}, \\ \mathcal{G}(\text{CD}, \text{FF}, q_b^{\text{in}}) &= \{\mathcal{X} \in \mathbb{R}^3 | n_b = 0, q_b^{\text{in}} < q_b^{\text{dis}}\}, \end{aligned}$$

and are characterised by reset maps

$$\begin{aligned} \mathcal{R}(\text{FF}, \text{FF}, \mathcal{X}, q_b^{\text{in}}) &= [n_b \quad \Theta_{q_b^{\text{in}}} \quad q_b^{\text{in}}]^\top, \\ \mathcal{R}(\text{FF}, \text{CD}, \mathcal{X}, q_b^{\text{in}}) &= [n_b \quad \theta \quad q_b^{\text{old}}]^\top, \\ \mathcal{R}(\text{CD}, \text{FF}, \mathcal{X}, q_b^{\text{in}}) &= [n_b \quad \Theta_{q_b^{\text{in}}} \quad q_b^{\text{in}}]^\top. \end{aligned}$$

Since  $\Theta_{q_b^{\text{in}}}$  is a random variable (6.4), these reset maps are stochastic.

### 6.2.2 Reactive bottleneck decongestion control

Assuming we know the traffic state along the road  $\rho(x, t)$ , and the state of the bottleneck  $m_b(t) \in \{\text{FF}, \text{CD}\}$ , we may design a control law for bottleneck decongestion and throughput improvement. As actuators, we will use controlled moving bottlenecks. We assume that we are able to create controlled moving bottlenecks at arbitrary positions on the road  $x_\varepsilon$ , and control how many lanes they take; in real application, this will depend on availability of suitable infrastructure-controllable vehicles. Considering a three lane road, we say that a moving bottleneck can either take two lanes, in which case its effect on the rest of the traffic is described by flux function  $Q^{\text{lo}}$ , or one lane, corresponding to flux function  $Q^{\text{hi}}$ , thus limiting the

overtaking flow at its position to  $q^{\text{lo}}$  and  $q^{\text{hi}}$ , respectively. This way, we are able to regulate the traffic flow and restrict it when and where it is required.

Since traffic breakdown at the bottleneck happens stochastically, as discussed in the previous subsection, we propose a control law that reacts to this event. When a traffic breakdown is detected at some time  $t = t_0$ , the controller can react by activating some vehicles on the road to act as controlled moving bottlenecks and help decongest the bottleneck. The control law needs to determine:

- the number of moving bottlenecks to be created  $\Xi(t_0)$ ,
- positions where the moving bottleneck will be created  $x_\xi(t_0)$ ,  $\xi = 1, \dots, \Xi(t_0)$ , with  $X_b > x_1(t_0) > \dots > x_{\Xi(t_0)}(t_0)$ ,
- velocities of the moving bottlenecks  $u_\xi(t)$ ,  $t \geq t_0$ , until  $x_\xi(t) = X_b$ ,  $\xi = 1, \dots, \Xi(t_0)$ , and
- overtaking flow limits of the moving bottlenecks  $q_\xi^{\text{cap}}(t)$ ,  $t \geq t_0$ , until  $x_\xi(t) = X_b$ ,  $\xi = 1, \dots, \Xi(t_0)$ .

If some number controlled moving bottlenecks  $\Xi(t_0-)$  were already active on the road at  $t = t_0$ , with positions  $x_\xi(t_0)$ ,  $\xi = 1, \dots, \Xi(t_0-)$ , the control law instead recalculates their velocities  $u_\xi(t)$  and overtaking flow limits  $q_\xi^{\text{cap}}(t)$ , and determines the number of new moving bottlenecks to be created  $\Xi(t_0) - \Xi(t_0-)$  and their positions  $x_\xi(t_0)$ ,  $\xi = \Xi(t_0-) + 1, \dots, \Xi(t_0)$ .

To simplify the control design, we adopt some assumptions and constraints. First, we assume that all free flow traffic moves at the same constant free flow speed  $V$ . Then, we assume that all moving bottlenecks move at the same speed  $u_\xi < V$ , and that we can only create controlled moving bottlenecks some minimum distance  $d^{\text{min}}$  away from other moving bottlenecks,  $x_\xi(t_0) - x_{\xi+1}(t_0) \geq d^{\text{min}}$ , and the stationary bottleneck  $X_b - x_1(t_0) \geq d^{\text{min}}$ , unless that moving bottleneck was already present on the road before time  $t_0$ . Finally, we assume that the overtaking flow limits of the moving bottlenecks can take one of two values,  $q_\xi^{\text{cap}}(t) \in \{q^{\text{lo}}, q^{\text{hi}}\}$ , with  $q^{\text{lo}} < q_b^{\text{dis}} < q^{\text{hi}}$ , i.e. the minimum overtaking flow limit is lower than the discharging flow from the congested bottleneck, which is lower than the maximum overtaking flow limit. Furthermore, this control input can be characterised by

$$q_\xi^{\text{cap}}(t) = \begin{cases} q^{\text{lo}}, & t_0 \leq t \leq t_\xi^c(t_0), \\ q^{\text{hi}}, & t > t_\xi^c(t_0), \end{cases}$$

and is uniquely determined by the switching time  $t_\xi^c(t_0)$ . Therefore, the set of control inputs that the simplified control law needs to determine reduces to:

- Number and positions of new moving bottlenecks to be created,  $\Xi(t_0) - \Xi(t_0-)$ , and  $x_\xi(t_0)$ ,  $\xi = \Xi(t_0-) + 1, \dots, \Xi(t_0)$ , and
- Switching times for the overtaking flow limits of all moving bottlenecks  $t_\xi^c(t)$ ,  $\xi = 1, \dots, \Xi(t_0)$ .

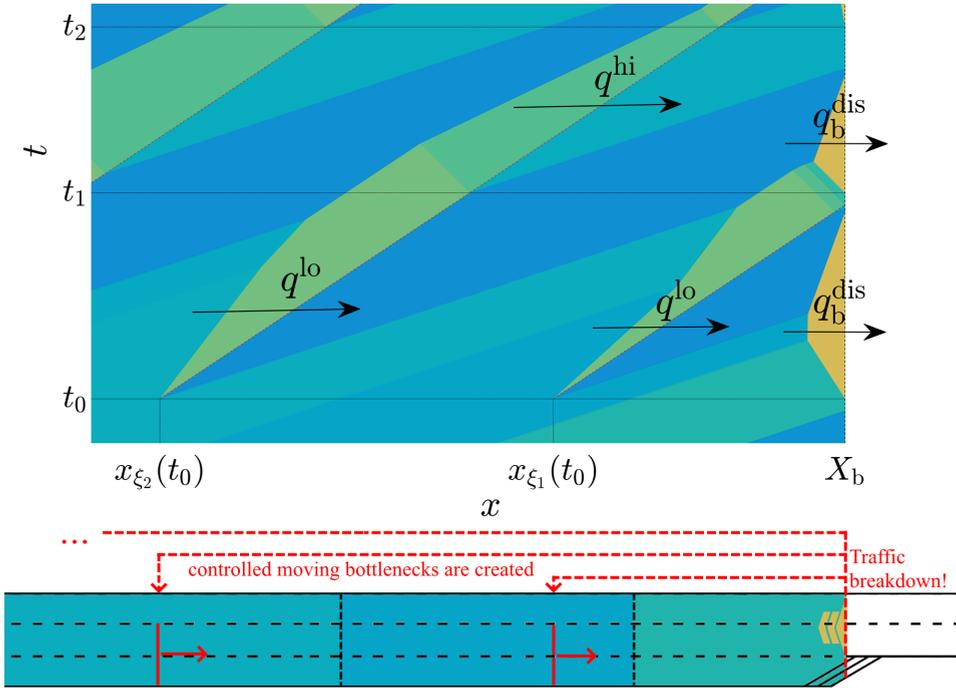


Figure 6.1: An illustration of how bottleneck decongestion control is calculated. The trajectories of the bottlenecks are indicated by dashed lines and the text above the arrows indicates the intensity of traffic flow past moving and static bottleneck. Brighter colour indicates higher traffic density. Once a traffic breakdown is detected, controlled moving bottlenecks are created at desired positions.

An example of control calculation is shown in Figure 6.1.

### Controlled moving bottleneck $\xi = 1$

Since the behaviour of the bottleneck is deterministic while it is in state  $m = \text{CD}$ , we may calculate  $t_1^c(t_0)$  that ensures that the bottleneck is decongested, thus maximizing the outflow from the road segment, exactly. The evolution of the bottleneck queue length in this state is given by

$$\dot{n}_b(t) = \begin{cases} V\rho(X_b - V(t - t_0), t_0) - q_b^{\text{dis}}, & t < t_1^c(t_0), \\ q^{\text{lo}} - q_b^{\text{dis}}, & t_1^c(t_0) \leq t < t_1^d(t_0), \end{cases}$$

where  $n_b(t_0) = 0$ ,  $t_1^V(t_0)$  is given by (4.33),  $t_1^V(t_0) = t_0 + (X_b - x_1(t_0))/V$ , and we denote

$$t_\xi^d(t_0) = t_\xi^V(t_\xi^c(t_0)).$$

This holds from time  $t = t_0$  until  $t = t_1^d(t_0)$  or  $n_b(t)$  becomes 0. If for some  $t \in (t_0, t_1^V(t_0)]$  we have  $n_b(t) \leq 0$ ,

$$V \int_{t_0}^t \rho(X_b - V(\tau - t_0), t_0) d\tau = \int_{X_b - V(t-t_0)}^{X_b} \rho(x, t_0) dx \leq q_b^{\text{dis}}(t - t_0),$$

the queue at the bottleneck will have dissipated before the overtaking flow from  $x_1(t_0)$  reaches it, and we may set  $t_1^c(t_0) = t_0$ ,  $q_1^{\text{cap}}(t) = q^{\text{hi}}$ . Otherwise, the queue at the bottleneck is dissipated at time  $t_b^c(t_0)$ ,

$$t_b^c(t_0) = \frac{\int_{x_1(t_0)}^{X_b} \rho(x, t_0) dx - q^{\text{lo}} t_1^V(t_0)}{q_b^{\text{dis}} - q^{\text{lo}}}. \quad (6.5)$$

If this happens for  $t_b^c(t_0) \leq t_1^u$ , where  $t_1^u = (X_b - x_1(t_0))/u_1$  according to (4.34), the queue at the bottleneck can be dissipated before moving bottleneck 1 reaches it. We may then set  $t_1^c(t_0)$  according to (6.5), with

$$\begin{aligned} t_b^c(t_0) &\leq t_1^d(t_0), \\ t_b^c(t_0) &\leq t_1^c(t_0) + \frac{X_b - x_1(t_0) - u_1 t_1^c(t_0) + u_1 t_0}{V}, \\ t_1^c(t_0) &\geq \frac{V t_b^c - X_b + x_1(t_0) - u_1 t_0}{V - u_1}. \end{aligned} \quad (6.6)$$

Otherwise, we set  $t_1^c(t_0) = t_1^u$ .

If there are no controlled moving bottlenecks present on the road at time  $t_0$ , and

$$\int_{X_b - d^{\text{min}}}^{X_b} \rho(x, t_0) dx \geq q_b^{\text{dis}} \frac{d^{\text{min}}}{V},$$

then creating controlled moving bottlenecks will make the dissipation of the queue at the bottleneck faster, and the control law creates at least one moving bottleneck. Otherwise, no moving bottlenecks will be created at time  $t_0$ . The first moving bottleneck can be created at position  $x_1(t_0)$  if it satisfies

$$X_b - x_1(t_0) \geq \max \left\{ d^{\text{min}}, \frac{u_1 \int_{x_1(t_0)}^{X_b} \rho(x, t_0) dx}{q^{\text{lo}} - q_b^{\text{dis}}} \right\}. \quad (6.7)$$

Therefore, we can achieve the fastest bottleneck decongestion if we choose  $t_1^c(t_0)$  and  $x_1(t_0)$  according to (6.6) and (6.7) with equality.

**Controlled moving bottlenecks  $\xi > 1$** 

Once the control action for the first moving bottleneck has been determined, we may proceed to iteratively determine the control actions for each potential subsequent one. Since each moving bottleneck is created so that the stationary bottleneck is decongested before the moving bottleneck reaches it, the behaviour of the traffic downstream of moving bottleneck  $\xi > 1$  is not deterministic. However, if  $t_\xi^V(t_0) \geq t_{\xi-1}^d$ , and for some  $t \in (t_{\xi-1}^d(t_0), \min\{t_{\xi-1}^u, t_\xi^V(t_0)\}]$  we have that

$$\int_{x_{\xi-1}(t_0)-V(t-t_0)}^{x_{\xi-1}(t_0)} \rho(x, t_0) dx - q^{\text{lo}}(t_{\xi-1}^d(t_0) - t) - t_{\xi-1}^V(t_0) - q^{\text{hi}}(t - t_{\xi-1}^d(t_0)) \leq 0,$$

or if  $t_{\xi-1}^u < t_\xi^V(t_0)$  and for some  $t \in (t_{\xi-1}^u, t_\xi^V(t_0)]$  we have that

$$\int_{x_{\xi-1}(t_0)-V(t-t_0)}^{x_{\xi-1}(t_0)} \rho(x, t_0) dx - q^{\text{lo}}(t_{\xi-1}^d(t_0) - t) - t_{\xi-1}^V(t_0) - q^{\text{hi}}(t_{\xi-1}^u - t_{\xi-1}^d(t_0)) - q_b^{\text{dis}}(t - t_{\xi-1}^u) \leq 0,$$

the congestion downstream of moving bottleneck  $\xi$  will have cleared by  $t_\xi^V(t_0)$ , and we may set  $t_\xi^c(t_0) = t_0$ ,  $q_\xi^{\text{cap}}(t) = q^{\text{hi}}$ .

Next, if for some  $t \in (t_{\xi-1}^d(t_0), t_{\xi-1}^u]$  we have that

$$n_{\xi-1}(t_{\xi-1}^d(t_0)) - q^{\text{hi}}(t - t_{\xi-1}^d(t_0)) + q^{\text{lo}}(t - t_{\xi-1}^V(t_0)) \leq 0, \quad (6.8)$$

where  $n_{\xi-1}(t_{\xi-1}^d(t_0))$  denotes

$$n_{\xi-1}(t_{\xi-1}^d(t_0)) = \int_{x_\xi(t_0)}^{x_{\xi-1}(t_0)} \rho(x, t_0) dx - q^{\text{lo}}(t_{\xi-1}^d(t_0) - t_{\xi-1}^V(t_0)),$$

setting  $t_\xi^c(t_0)$  to

$$t_\xi^c(t_0) = \frac{Vt_\xi^d(t_0) - X_b + x_\xi(t_0) - u_\xi t_0}{V - u_\xi}, \quad (6.9)$$

where  $t_\xi^d(t_0)$  is given by

$$t_\xi^d(t_0) \geq \frac{n_{\xi-1}(t_{\xi-1}^d(t_0)) + q^{\text{hi}}t_{\xi-1}^d(t_0) - q^{\text{lo}}t_\xi^V(t_0)}{q^{\text{hi}} - q^{\text{lo}}},$$

ensures that the congestion downstream of moving bottleneck  $\xi$  is dissipated. If condition (6.8) does not hold for any  $t \in (t_{\xi-1}^d(t_0), t_{\xi-1}^u]$ , the congestion built up in the wake of moving bottleneck  $\xi - 1$  reaches the stationary bottleneck. This arrival of dense traffic will cause prompt traffic breakdown with high probability.

Therefore, in this case we conservatively assume that the traffic breakdown happens immediately at time  $t_{\xi-1}^u$ , and if for some  $t \in (t_{\xi-1}^u, t_{\xi}^u]$  we have that

$$n_{\xi-1}(t_{\xi-1}^d(t_0)) - q^{\text{hi}}(t_{\xi-1}^u - t_{\xi-1}^d(t_0)) + q^{\text{lo}}(t - t_{\xi}^V(t_0)) - q_{\text{b}}^{\text{dis}}(t - t_{\xi-1}^u) \leq 0,$$

setting  $t_{\xi}^c(t_0)$  according to (6.9), where  $t_{\xi}^d(t_0)$  is given by

$$t_{\xi}^d(t_0) \geq \frac{n_{\xi-1}(t_{\xi-1}^d(t_0)) - q^{\text{hi}}(t_{\xi-1}^u - t_{\xi-1}^d(t_0)) - q_{\text{b}}^{\text{dis}}t_{\xi-1}^u + q^{\text{lo}}t_{\xi}^V(t_0)}{q_{\text{b}}^{\text{dis}} - q^{\text{lo}}}, \quad (6.10)$$

ensures that the queue at the bottleneck can be dissipated before moving bottleneck  $\xi$  reaches it. Otherwise, we set  $t_{\xi}^c(t_0) = t_{\xi}^u$ .

Finally, we create a new moving bottleneck  $\xi$  if

$$\int_{X_{\text{b}} - Vt_{\xi-1}^u}^{X_{\text{b}}} \rho(x, t_0) dx - q^{\text{lo}}(t_{\xi-1}^d(t_0) - t_{\xi-1}^V(t_0)) - q^{\text{hi}}(t_{\xi-1}^u - t_{\xi-1}^d(t_0)) \geq 0,$$

at position  $x_{\xi}(t_0)$ ,

$$x_{\xi}(t_0) = \min\{x_{\xi-1}(t_0) - d^{\text{min}}, X_{\text{b}} - u_{\xi}t_{\xi}^d(t_0)\},$$

where  $t_{\xi}^d(t_0)$  is determined by (6.10) with equality.

### 6.2.3 Simulation setup

The proposed control law is tested on a simulated stretch of road upstream of a stationary bottleneck at position  $X_{\text{b}}$ , where we assume that flux function of the road is triangular. Let  $Q^{\Delta}(\sigma^s; \rho)$  be the triangular flux function, scaled so that  $\sigma^s$  is its critical density,

$$Q^{\Delta}(\sigma^s; \rho) = \begin{cases} V^{\text{FF}} \rho, & 0 \leq \rho \leq \sigma^s, \\ V^{\text{FF}} \frac{\sigma^0}{P^0 - \sigma^0} (P^0 \frac{\sigma^s}{\sigma^0} - \rho), & \sigma^s < \rho \leq P^0 \frac{\sigma^s}{\sigma^0}, \end{cases}$$

where  $V^{\text{FF}}$  is the free flow speed, and  $\sigma^0$  and  $P^0$  are the critical and jam density of the road upstream of the bottleneck, respectively. We use the Front Tracking Transition System Model (FTTSM), described in 4.5, to capture the evolution of traffic, the dynamics of the bottleneck, and the influence of controlled moving bottlenecks.

The model is initialized with

$$\mathcal{X}^0 = (N^0, t^0, z^0, \rho^0, Q^0, W^0, \mathcal{R}^0) = (N^0, 0, z^0, \rho^0, Q^0, (-V^{\text{FF}}, V^{\text{FF}}), \emptyset).$$

The initial wavefront positions are given iteratively by

$$z_{N^0}^0 = z_{N^0-1}^0 = X_{\text{b}},$$

$$z_i^0 = z_{i=1}^0 - \Delta z_i^0, \quad i = N^0 - 2, N^0 - 3, \dots, 1,$$

$$\Delta z_i^0 \sim \mathcal{U}\left(\frac{L}{2}, \frac{3L}{2}\right),$$

i.e. the initial length of the zones between two wavefronts upstream of the bottleneck is uniformly randomly distributed, with  $L$  as the average initial condition density length. The initial traffic densities are given by

$$\rho_i^0 \sim \mathcal{U}(2\sigma^0 - \bar{\rho}, \sigma^0), \quad i = 1, \dots, N^0 - 2,$$

$$\rho_i^0 = \frac{\sigma^0}{2}, \quad i = N^0 - 1, N^0, N^0 + 1,$$

where  $\bar{\rho}$  is the average of traffic density  $\rho_i^0$ ,  $i = 1, \dots, N^0 - 2$ , and we assume that  $V^{\text{FF}}\sigma^0/2 < q_b^{\text{dis}}$ . The flux functions are given by

$$Q^{0,i} = (V^{0,i}, \Sigma^{0,i}, \Lambda^{\pm,0}, \xi^{0,i}, F^0) = \left( \left[ V^{\text{FF}} \quad \frac{\sigma^0}{P^0 - \sigma^0} V^{\text{FF}} \right]^\top, [\sigma^0 \quad P^0]^\top, (0, 0), \xi^{0,i}, \emptyset \right),$$

with the identifier  $\xi^{0,i} = 0$  for  $i = 1, \dots, N^0 - 1$  and  $i = N^0 + 1$ , and  $\xi^{0,N^0} = \Xi_b$ , where  $\Xi_b$  can be taken as an arbitrary number larger than all other  $\xi$  that will be used in the course of the simulation. This flux function will model the state of the stationary bottleneck at position  $X_b$ .

Since the bottleneck is in free flow at time  $t = 0$ , it does not limit the traffic flow and its flux function can thus be taken to be the same as the flux function of the road upstream of it. However, once traffic breakdown happens, and capacity drop starts being active, the flux function of the bottleneck is changed to

$$Q_{i_{\Xi_b}}^{\text{CD}} = \left( \left[ V^{\text{FF}} \quad \frac{\sigma^0}{P^0 - \sigma^0} V^{\text{FF}} \right]^\top, [\sigma^{\text{d}} \quad P^{\text{d}}]^\top, (0, 0), \Xi_b, \emptyset \right),$$

where  $i_{\Xi_b}$  is given by  $\xi_i = \Xi_b$ , limiting the discharge rate to  $q_b^{\text{dis}} = V^{\text{FF}}\sigma^{\text{d}}$ . Once the congestion at the stationary bottleneck is dissipated and we have  $\rho_{\Xi_b} < \sigma^{\text{d}}$ , the flux function of the bottleneck is reverted to

$$Q_{i_{\Xi_b}}^{\text{FF}} = \left( \left[ V^{\text{FF}} \quad \frac{\sigma^0}{P^0 - \sigma^0} V^{\text{FF}} \right]^\top, [\sigma^0 \quad P^0]^\top, (0, 0), \Xi_b, \emptyset \right).$$

We may model the creation of a new moving bottleneck with unique identifier  $\xi < \Xi_b$  at time  $t_0$  and position  $x_\xi(t_0)$ , moving at speed  $u_\xi(t_0)$  and limiting the traffic flow to  $q_\xi^{\text{cap}}(t_0) = V^{\text{FF}}\sigma^\xi(t_0)$ , using one state insertion and one flux function transition.

1. Taking a transition  $+(\sigma^\xi(t_0), x_\xi(t_0))_{i_-}$ , a zone of density  $\sigma^\xi(t_0)$  is added downstream of front  $i_-$ ,  $z_{i_-} \leq x_\xi(t_0) \leq x_{i_-+1}$ . Since the newly created zone is infinitely narrow, it does not violate the conservation of traffic density.

2. Taking a transition  $\mathcal{Q}(Q^\xi(t_0), i_- + 1, i_- + 2)$ , where

$$Q^\xi(t_0) = \left( \left[ V^{\text{FF}} \quad \frac{\sigma^0}{P^0 - \sigma^0} V^{\text{FF}} \right]^\top, \left[ \sigma^\xi(t_0) \quad P^0 \frac{\sigma^\xi(t_0)}{\sigma^0} \right]^\top, (u_\xi(t_0), u_\xi(t_0)), \xi, \emptyset \right),$$

the flux function at the position of the moving bottleneck is scaled down so that its capacity is  $q_\xi^{\text{cap}}(t_0) = V\sigma^\xi(t_0)$ , and both the upstream and downstream ends of the bottleneck will move at speed  $u_\xi(t_0)$ .

A similar procedure can be applied to create a moving bottleneck of nonzero length. Note that it is required to ensure that  $u_\xi(t_0)$  is always such that

$$u_\xi(t) \left( \rho_+ - P^0 \frac{\sigma^\xi(t)}{\sigma^0} \right) \geq Q_+(\rho_+),$$

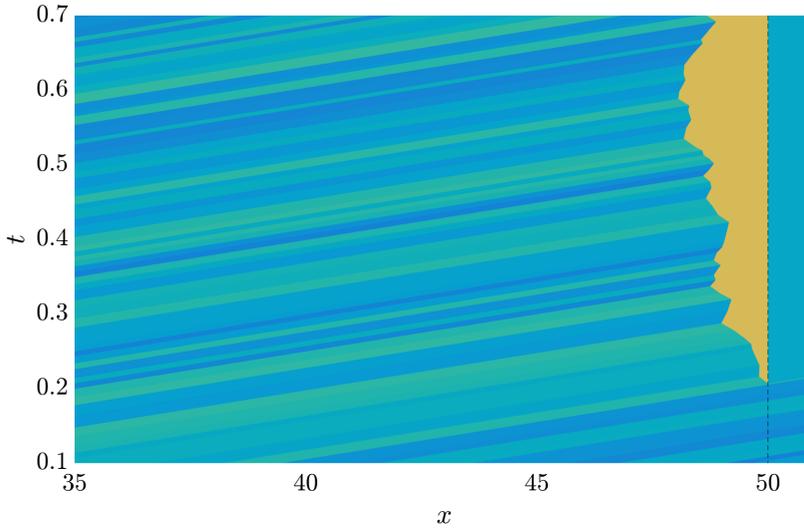
where  $\rho_+$  and  $Q_+$  are the traffic density and flux function immediately downstream of the moving bottleneck. The speed or capacity of the moving bottleneck  $\xi$  can be changed by changing  $\Lambda_{i_\xi}^\pm$  and  $\Sigma_{i_\xi}$ .

#### 6.2.4 Simulation results

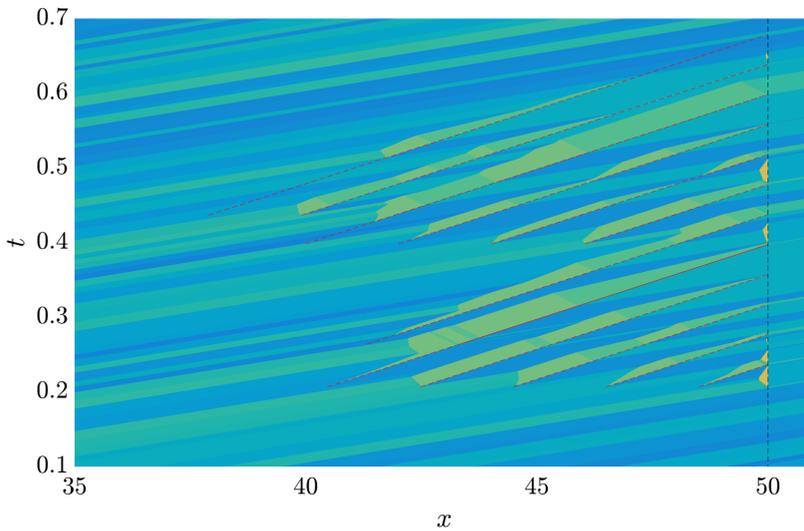
The simulation results of an example are shown in Figure 6.2, comparing the case where we apply no control and let the traffic evolve freely, and the case where we apply bottleneck decongestion control. We consider a stretch of highway, with no on- and off-ramps and a bottleneck at position  $X_B$ . The initial density  $\rho(0, x)$  is piecewise constant and randomly generated, with average value  $\bar{\rho}$ , resulting in a varying traffic flow at the position of the bottleneck. The time the first traffic breakdown happens (in this case at  $t_0 \approx 0.21$ ) is taken to be the same in both cases, and the simulations run independently starting with that point.

As can be seen from Figure 6.2b, by delaying the arrival of a part of the traffic, we are able to maintain free flow at the bottleneck. Control action is recalculated in order to react to changes at the bottleneck, and new controlled moving bottlenecks are added when needed. Since in this case the average initial traffic density is larger than the density at which the traffic flows out of the congested bottleneck,  $\bar{\rho} > \sigma^d$ , once a traffic breakdown happens, it is likely that congestion will persist and grow, since the average inflow to the queue will be larger than its outflow.

The flow at the position of the bottleneck is shown in Figure 6.3. We can see that the traffic flow follows the demand until  $t_0$ , when a traffic breakdown happens. In the controlled case, we manage to return to the unperturbed state around  $t = 0.65$ , whereas in the uncontrolled case the congestion at the bottleneck keeps accumulating. A total of  $n_c = 2106$  vehicles was served from  $t = 0.1$  to  $t = 0.7$  in the controlled case, compared to  $n_{\text{nc}} = 1990$  vehicles in the uncontrolled case, corresponding to a queue of  $n_b = 116$  vehicles at  $t = 0.7$ .



(a) No control.



(b) With control.

Figure 6.2: A simulation example comparing the evolution of traffic without and with control. Brighter colour indicates higher traffic density.

### 6.3 Platoon coordination for deterministic bottleneck decongestion

A common simplification is to assume that traffic breakdown happens deterministically once the traffic flow at the bottleneck exceeds some fixed value, although it is

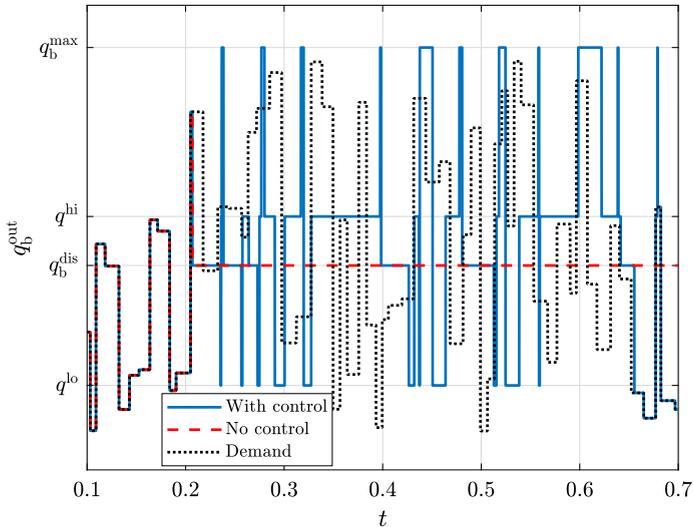


Figure 6.3: Traffic flow at  $X_B$  with and without control, compared with the demand.

broadly accepted, and empirically shown, that a breakdown happens stochastically, with different breakdown probabilities in range of traffic flow levels. Notwithstanding, events such as arrival of a platoon of heavy-duty vehicles at the bottleneck can make traffic breakdown certain, if the background traffic density is high enough. The traffic models that are able to capture the capacity drop phenomenon are also able to represent the traffic system with deterministic traffic breakdown. Namely,

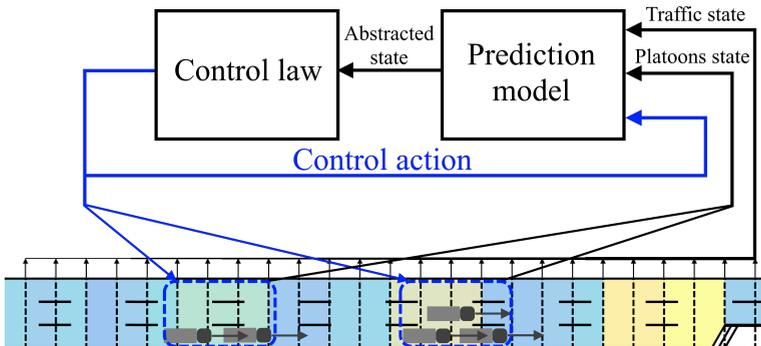


Figure 6.4: Schematic representation of the control system. We use the information about the current traffic state and state of controllable platoons to calculate the control actions in order to improve the traffic situation.

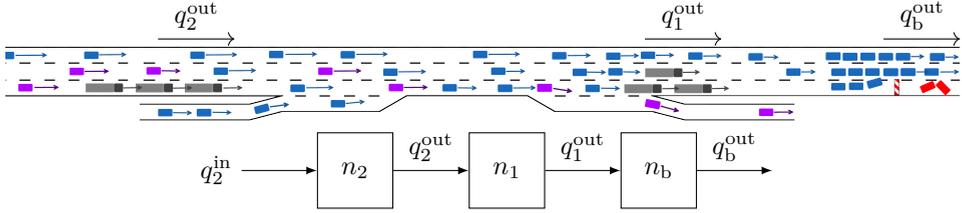


Figure 6.5: An illustration of the road segment under consideration. The stationary bottleneck corresponds to  $n_b$ , the downstream platoon to  $n_1$  and the upstream platoon to  $n_2$ , and the overtaking flows at the platoons are regulated accordingly. Both the inflow from the on-ramp and the outflow to the off-ramp will factor in the inflow to the downstream platoon queue.

in this section, we will use two such models:

- multi-class CTM, as a richer and more complex simulation model (described in Section 4.3), and
- tandem queueing model with moving bottlenecks, as a simplified representation of the model, used for prediction-based control design, analysis and implementation (described in Section 4.4).

The control action is calculated according to the simulated traffic state, and then applied on the more complex simulation model. An outline of the control loop is shown in Figure 6.4.

We use the proposed prediction-based control to restrict just enough inflow to the stationary bottleneck so that we first help dissipate the bottleneck queue, by restricting the flow as much as possible, and then keeping its flow as close as possible to its capacity. Depending on how directly we can see the influence of individual control actions on the relevant system states, we may be able to calculate optimal control in one shot, directly from the state prediction, or we may find the optimal control iteratively. For ease of demonstration, in this part we focus again on a single stationary bottleneck, with one on-ramp and one off-ramp upstream of its location acting as disturbances to the traffic flow, as shown in Figure 6.5.

### 6.3.1 Control laws

Owing to the simple deterministic model of traffic breakdown, it is now straightforward to calculate the control action for bottleneck decongestion. As discussed in Section 6.1, we can improve the TTS by ensuring

$$q_b^{\text{in}}(t) = \begin{cases} q^{\text{lo}}, & m_b(t) = \text{CD}, \\ q^{\text{hi}}, & m_b(t) = \text{FF}, \end{cases}$$

which we intend to do using platoons acting as controlled moving bottlenecks. The model of the bottleneck (6.2) is completed with

$$m_b(t) = \begin{cases} \text{FF}, & n_b(t) = 0, \\ \text{CD}, & n_b(t) \geq 0, \end{cases}$$

determining the bottleneck mode.

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  $V$  (which we assume is constant and could be given as an averaged value), minimum enforceable platoon speed  $U_{\min}$ , stationary bottleneck capacity  $q_b^{\text{cap}}$ , stationary bottleneck discharging flow  $q_b^{\text{dis}}$ , minimum enforceable overtaking flow at the moving bottlenecks  $q^{\text{lo}}$ , maximum enforceable overtaking flow at the moving bottleneck that is lower than the stationary bottleneck capacity  $q^{\text{hi}}$ , and average splitting ratio at the off-ramps  $R_k$ . The general road geometry, including the stationary bottleneck position  $X_b$  and the position of the on- and off-ramps  $X_k^{\text{r}}$ , is also assumed to be known.

We are looking to maximize the outflow from the bottleneck, which in case there are no off-ramps corresponds to minimizing the TTS. 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  $u_\xi(t)$ , which we control by changing the reference speed of the platooned vehicles, and the moving bottleneck formation, which is controlled by changing how many lanes the platoon takes and directly related to  $q_\xi^{\text{cap}}(t)$ . By doing this, we are 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 rely on the prediction model, and will be described in the remainder of the section.

### Ideal actuation

Before discussing platoon actuated control, in order to have a baseline for comparing the performance of the proposed control laws, we consider the ideal case, assuming we can fully control all traffic, and that we can control every class of traffic independently. This corresponds to having a 100% penetration rate of connected, communicating, and controlled vehicles, and knowing each vehicle's destination. Assuming the inflow of off-ramp-bound vehicles is lower than the capacity of the off-ramp, we only need to minimally delay the mainstream-bound background traffic so that the flow at the bottleneck never exceeds its capacity  $q_b^{\text{in}}(t^T) \leq q_b^{\text{cap}} = q_b^{\text{max}}$ .

In the multi-class CTM, which is used as the simulation model here, this is equivalent to ensuring that the traffic density immediately upstream of the bottleneck  $\rho_{i_b}(t^T) \leq \sigma^+$  for all  $t$ , where  $\sigma^+$  is the critical density of the stationary bottleneck, with  $U_i^k(t^T)$  as close as possible to  $V$ . Assuming the platooned vehicles belong to class  $a$ , and the background traffic bound for the bottleneck belongs to class  $b$ , this can be achieved by setting  $U_i^b(t^T)$  recursively,

$$\begin{aligned}
 U_{i_b}^b(t^T) &= V, \\
 U_i^b(t^T) &= V \min \left\{ 1, \max \left\{ \frac{U_{\min}^b}{V}, \frac{\rho_i^{b*}(t^T) - \frac{V - U_{i+1}^b(t^T)}{V} \rho_{i+1}^b(t^T)}{\rho_i^b(t^T)} \right\} \right\}, \quad i = 1, \dots, i_b - 1 \\
 \rho_i^{b*}(t^T) &= \begin{cases} \sigma_+ - \rho_\xi^*(t^T), & 0 < \frac{X_b - (i-1)L}{V} - \frac{X_b - x_\xi(t^T)}{u_\xi(t^T)} < \frac{l_\xi(t^T)}{u_\xi(t^T)} + \frac{L}{V}, \quad \xi = 1, \dots, \Xi, \\ \sigma_+, & \text{otherwise,} \end{cases}
 \end{aligned} \tag{6.11}$$

where  $i_b$  is the cell with the stationary bottleneck. This way, the mainstream-bound background traffic is regulated so that the total demand at the bottleneck, including the arriving platoons, is kept as close to its capacity as possible without exceeding it. The mainstream-bound background traffic is delayed minimally, while the platoons and the off-ramp-bound background traffic experience no delay, travelling at their respective maximum speeds.

### Platoon-actuation not aware of on- or off-ramps

The control objective, maximizing the throughput, i.e., the outflow  $q_b^{\text{out}}(t)$ , can be achieved by keeping  $n_b(t) = 0$  and  $q_b^{\text{in}}(t)$  as close as possible to  $q_b^{\text{cap}}$ . Additionally, we require that the queue at the platoon is already discharged when the platoon reaches the bottleneck,  $n_\xi(t_\xi^u(t_0)) = 0$ . Therefore we employ control law using tandem queueing model with moving bottlenecks as the prediction model,

$$\tilde{q}_\xi^{\text{cap}}(t) = \begin{cases} q^{\text{ref}}(t), & n_b(t) = 0 \wedge t \geq t_{\xi-1}^u(t_0), \\ \tilde{q}_{\xi-1}^{\text{cap}}(t), & \tilde{n}_{\xi-1}(t) = 0 \wedge t < t_{\xi-1}^u(t_0), \\ q^{\text{lo}}, & \text{otherwise,} \end{cases} \tag{6.12}$$

where we calculate the control action at time  $t_0$ , and the reference flow  $q^{\text{ref}}(t)$  can be externally determined. The coordinate shifted times  $t_\xi^V(t_0)$  and  $t_\xi^u(t_0)$  are given by (4.33) and (4.34), respectively. For maximizing the throughput, we set

$$q^{\text{ref}}(t) = q^{\text{hi}} - q_b^u(t),$$

taking the largest admissible  $q^{\text{hi}} \leq q_b^{\text{cap}}$ . In order to compute the current  $q_\xi^{\text{cap}}(t_0) = \tilde{q}_\xi^{\text{cap}}(t_\xi^V(t_0))$  for all platoons, we need to predict  $n_b(t)$  until  $t_\Xi^V(t_0)$ , which requires calculating  $q_\Xi^{\text{cap}}(t_0)$  and  $q_\xi^{\text{cap}}(t)$  for  $0 \leq t \leq \min \{ t_\xi^u(t_0), t_\Xi^V(t_0) \}$ .

Assuming this control law is applied, we set the speed of each platoon so that  $n_\xi(t_\xi^u(t_0)) = 0$  and  $n_b(t) = 0$ ,  $t_\xi^c(t_0) \leq t \leq t_\xi^u(t_0)$ , with minimum  $t_\xi^c(t_0)$ , where

$$t_\xi^c(t_0) \geq \max \left\{ t_\xi^V(t_0), t_{\xi-1}^u(t_0) + \frac{l_{\xi-1}}{V} \right\}.$$

This is achieved when

$$\tilde{n}_\xi(t_\xi^u(t_0)) = \tilde{n}_\xi(t_\xi^c(t_0)) + \int_{t_\xi^c(t_0)}^{t_1^u(t_0)} \tilde{q}_1^{\text{in}}(t) dt - q^{\text{hi}}(t_1^u(t_0) - t_1^c(t_0)) = 0. \quad (6.13)$$

For  $\xi = 1$ , in case it is known that  $t_2^V(t_0) < t_1^u(t_0)$ , (6.13) simplifies to

$$\begin{aligned} \tilde{n}_1(t_1^u(t_0)) &= \tilde{n}_1(t_2^V(t_0)) + q^{\text{lo}}(t_1^u(t_0) - t_1^c(t_0)) - q^{\text{hi}}(t_1^u(t_0) - t_1^c(t_0)) = 0, \\ u_1(t_0) &= \frac{(q^{\text{hi}} - q^{\text{lo}})(X_b - x_1(t_0))}{\tilde{n}_1(t_2^V(t_0)) + (q^{\text{hi}} - q^{\text{lo}})t_1^c(t_0)}, \end{aligned}$$

since we can explicitly calculate

$$\tilde{n}_1(t_2^V(t_0)) = \int_{t_1^V(t_0)}^{t_2^V(t_0)} V\rho(X_b - Vt, 0) dt - q^{\text{lo}}(t_2^V(t_0) - t_1^V(t_0)).$$

Otherwise,  $u_\xi(t_0)$  is calculated by solving (6.13) numerically, and can be obtained as a by-product of iterating the prediction steps for  $n_b(t)$  and  $\tilde{n}_\xi(t)$ . The simplest way of calculating  $u_\xi(t_0)$  is to initialize it to

$$u_\xi^{(0)}(t_0) = \min \left\{ U^{\text{max}}, u_{\xi-1}(t_0) \frac{X_b - x_\xi(t_0)}{X_b - x_{\xi-1}(t_0) + l_{\xi-1}} \right\},$$

and then decrease it until either  $u_\xi(t_0) = U^{\text{min}}$  or (6.13) is satisfied. This also ensures that  $u_\xi(t_0)$  is constrained to be within the range

$$U^{\text{min}} \leq u_\xi(t_0) \leq \min \left\{ U^{\text{max}}, u_{\xi-1}(t_0) \frac{X_b - x_\xi(t_0)}{X_b - x_{\xi-1}(t_0) + l_{\xi-1}} \right\},$$

which is required for the limitations to be met if there is no platoon merging.

### Platoon-actuation aware of on- or off-ramps

Consider now the case when there are on- or off-ramps. In order to predict the evolution of queues, which is needed for computing the control inputs, we need to know the ramp flows  $\tilde{q}_k^r(t)$  in advance. This information can be hard to obtain, since it will depend on the routing decisions of individual drivers constituting the background traffic. Therefore, we use the predicted ramp flows.

If ramp  $k$  is an on-ramp, we can replace the actual ramp flow with its average  $\hat{q}_k^r(t) = \bar{q}_k^r(t)$ , which in reality can be determined statistically. If ramp  $k$  is an off-ramp, we can employ the standard assumption that some constant ratio of vehicles  $R_k$  leave the road via the off-ramp. We can then write

$$\begin{aligned} \hat{q}_k^r(t) &= -R_k \left( \hat{q}_k^{\text{in},r}(t) + \sum_{l \in K_o^{r,k}(t)} \hat{q}_l^r(t) \right), \\ \hat{q}_k^{\text{in},r}(t) &= \begin{cases} q_b^V(t), & x_1^u(t) < x_k^r < X_b \\ \hat{q}_{\xi+1}^{\text{out}}(t), & x_{\xi+1}^u(t) < x_k^r < x_\xi^u(t) \end{cases} \\ K_o^{r,k}(t) &= \begin{cases} \{l | x_1^u(t) < x_l^r < x_k^r\}, & t > t_1^V(t_0), x_k^r < x_{\xi-1}^u(t) \\ \{l | x_\xi^u(t) < x_l^r < x_k^r\}, & x_k^r < x_{\xi-1}^u(t), \xi > 1 \\ \{l | X_b - Vt < x_l^r < x_k^r\}, & \text{otherwise} \end{cases} \\ x_\xi^u(t) &= x_\xi(t_0) + u_\xi(t_0)(t - t_0), \quad t \geq t_0, \end{aligned}$$

depending on the origin of the flow to off-ramp  $k$  at time  $t$ , and use  $\hat{q}_k^r(t)$  in place of  $q_k^r(t)$  in (4.35).

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

$$\tilde{n}_\xi(t_\xi^{r,k}(t_0)+) = (1 - R_k)\tilde{n}_\xi(t_\xi^{r,k}(t_0)), \quad x_\xi^u(t_\xi^{r,k}(t_0)) = X_k^r,$$

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 (6.12) 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  $\tilde{q}_\xi^{\text{cap}}(t) = q^{\text{hi}}$  if there are other platoons downstream that are regulating the inflow to the bottleneck. The updated control law is

$$\tilde{q}_\xi^{\text{cap}}(t) = \begin{cases} q^{\text{ref}}(t), & n_b(t) = 0 \wedge t \geq t_{\xi-1}^u(t_0), \\ q^{\text{hi}}, & K_\xi^{p-1*}(t) \neq \emptyset \wedge t < t_{\xi-1}^u(t_0), \\ \tilde{q}_{\xi-1}^{\text{cap}}(t), & K_\xi^{p-1*}(t) = \emptyset \wedge \tilde{n}_{\xi-1}(t) = 0 \wedge t < t_{\xi-1}^u(t_0), \\ q^{\text{lo}}, & \text{otherwise,} \end{cases} \quad (6.14)$$

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

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

### 6.3.2 Closed-loop system analysis

In order to understand the effects and limitations this control law will have in realistic situations, we first study it under simplified conditions, in an idealised setting.

Whereas in simulations the inflow of background traffic will vary in time, taking random values belonging to some range, we first assume constant background traffic inflow  $q^{\text{in}}(t) = q^{\text{in}}$ . Similarly, in simulations the platoons arrive with exponentially distributed gaps, but here we first assume periodic platoon arrivals, with period  $\tau_\pi$ , and each platoon consisting of  $n_\pi$  passenger car equivalents. Later, we will allow the inflow and gaps between two platoons vary within some range. In this subsection, 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.
3. An estimate of throughput given varying inflow and gap between platoons, i.e., the average inflow for which we are able to dissipate the congestion at the bottleneck with a predefined probability.

The bottleneck will be considered to have capacity  $q_b^{\text{cap}}$ , which is reduced to  $q_b^{\text{dis}}$  in case there is capacity drop,  $q_b^{\text{dis}} < q_b^{\text{cap}}$ . We study the case when the bottleneck is already congested at initial time. If the platoon arrives at a congested bottleneck, its vehicles are added to the bottleneck queue. Otherwise, if there is no queue at the platoon and it arrives at a bottleneck in free flow, the platoon passes through the bottleneck without causing traffic breakdown.

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

$$\dot{n}_b(t) = q_b^{\text{in}}(t) - q_b^{\text{out}}(t), \quad (6.15)$$

$$n_b(t_1^V(t_0)) = \mu_0, \quad (6.16)$$

$$q_b^{\text{in}}(t) = \begin{cases} \tilde{q}_\xi^{\text{out}}, & \max\{t_\xi^V(t_0), t_{\xi-1}^u(t_0)\} \leq t \leq t_\xi^u(t_0), \\ q^{\text{in}}(t), & \text{otherwise,} \end{cases} \quad (6.17)$$

$$q_b^{\text{out}}(t) = \begin{cases} q_b^{\text{in}}(t), & q_b^{\text{in}}(t) \leq q_b^{\text{cap}} \wedge n_b(t) = 0, \\ q_b^{\text{dis}}, & q_b^{\text{in}}(t) > q_b^{\text{cap}} \vee n_b(t) > 0, \end{cases} \quad (6.18)$$

$$\dot{\tilde{n}}_\xi(t) = \tilde{q}_\xi^{\text{in}}(t) - \tilde{q}_\xi^{\text{out}}(t), \quad t_\xi^V(t_0) < t < t_\xi^u(t_0), \quad (6.19)$$

$$\tilde{n}_\xi(t_\xi^V(t_0)) = 0, \quad (6.20)$$

$$\tilde{q}_\xi^{\text{in}}(t) = \begin{cases} \tilde{q}_{\xi+1}^{\text{out}}, & t_{\xi+1}^V(t_0) < t < t_{\xi+1}^u(t_0), \\ q^{\text{in}}(t), & t \leq t_{\xi+1}^V(t_0), \end{cases} \quad (6.21)$$

$$\tilde{q}_\xi^{\text{out}}(t) = \begin{cases} \tilde{q}_\xi^{\text{in}}(t), & \tilde{q}_\xi^{\text{in}}(t) \leq \tilde{q}_\xi^{\text{cap}}(t) \wedge \tilde{n}_\xi(t) = 0, \\ \tilde{q}_\xi^{\text{dis}}, & \tilde{q}_\xi^{\text{in}}(t) > \tilde{q}_\xi^{\text{cap}}(t) \vee \tilde{n}_\xi(t) > 0, \end{cases} \quad (6.22)$$

$$n_b(t_\xi^u(t_0)+) = \begin{cases} n_b(t_\xi^u(t_0)) + \tilde{n}_\xi(t_\xi^u(t_0)) + n_\pi, & n_b(t_\xi^u(t_0)) + \tilde{n}_\xi(t_\xi^u(t_0)) > 0, \\ 0, & n_b(t_\xi^u(t_0)) + \tilde{n}_\xi(t_\xi^u(t_0)) = 0, \end{cases} \quad (6.23)$$

$$\xi = 1, \dots, \Xi,$$

where  $\tilde{q}_\xi^{\text{cap}}(t)$  is governed by control law (6.12).

We control the platoons by setting their speeds and formation. The platoon speed  $u_\xi(t_0)$  can range from some set minimum speed  $U^{\text{min}}$  to the free flow speed of all traffic  $V$ , and it will determine the time at which the platoon will reach the bottleneck  $t_\xi^u(t_0)$ . By setting the platoon formation (i.e. number of lanes a platoon occupies), we control the maximum overtaking flow at its position  $\tilde{q}_\xi^{\text{cap}}(t)$ , with overtaking flow of  $q^{\text{lo}}$  corresponding to maximum number of lanes taken, and  $q^{\text{hi}}$  corresponding to one lane taken. To simplify the presentation, we omit writing  $(t_0)$ .

### 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.  $u_\xi = V$ ,  $t_\xi^u = t_\xi^V$ , and  $\tilde{q}_\xi^{\text{cap}}(t) = q^{\text{hi}}$ ), the system under consideration simplifies to (6.15)–(6.18) and (6.23) with  $\tilde{n}_\xi(t_\xi^u) = 0$ . The system is stable if

$$q^{\text{in}} + \frac{n_\pi}{\tau_\pi} < q_b^{\text{dis}},$$

i.e., if the average total inflow is less than the dissipating flow of the bottleneck, its queue length will go to zero regardless of its initial value.

If the platoons can be controlled, we are able to extend the range of  $q^{\text{in}}$  for which the system is stable. In this case, it is of interest to study what is the maximum initial queue length  $\mu_0$  for which the system is stable for a given  $q^{\text{in}}$ . The length of the considered road segment is  $X_b - X_0$  and a platoon moving at speed  $u_k$  traverses it and reaches the bottleneck after  $\tau_k^u = \frac{X_b - X_0}{u_k}$ . Assuming the first platoon enter the road at time  $t = 0$  and  $t_1^V = (X_b - X_0)/V$ , we define the initial queue length  $\mu_0 = n_b(t_1^V)$  as the queue length at the bottleneck at the time when the overtaking flow from the platoon reaches it. This is equivalent to saying that there is  $\mu_0$  excess congestion to be dissipated, signifying how many vehicles need to be temporarily removed from the inflow 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  $\mu_k$  will be distributed between the bottleneck and downstream platoons that previously entered the road.

The case we are considering assumes that the flow values are arranged as

$$q^{\text{lo}} < q_b^{\text{dis}} < q^{\text{in}} < q^{\text{in}} + \frac{n_\pi}{\tau_\pi} < q^{\text{hi}} \leq q_b^{\text{cap}}, \quad (6.24)$$

and the uncontrolled system is unstable.

The system is stable if  $\mu_{k+1} < \mu_k$  until  $\mu_k = 0$  for some  $k$ . This means that every subsequent platoon has less excess congestion to dissipate until the system returns to the unperturbed state. Since we are looking for maximum  $\mu_k$  for which this holds, we study the situation when maximum control action is applied, i.e. the platoon speed is  $u_k = U^{\min}$  and maximum overtaking flow is  $q^{\text{lo}}$  until the queue at the bottleneck is dissipated, which happens at  $\tau_k^{\text{dis}} = \mu_k / (q_b^{\text{dis}} - q^{\text{lo}})$ . Moving at minimum speed, a platoon will reach the bottleneck after  $\tau^{\text{max}} = (X_b - X_0) / U^{\min}$ , so a necessary condition to be able to begin dissipating the congestion is  $\tau^{\text{max}} > \tau_k^{\text{dis}}$ , which yields

$$\mu_0 < (q_b^{\text{dis}} - q^{\text{lo}}) \tau^{\text{max}}.$$

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 when they reach the bottleneck. In the recovery phase, each subsequent platoon will have a higher speed, until the traffic returns to the unperturbed state. The recovery phase begins with platoon  $k$  that is able to dissipate all excess congestion and reach the bottleneck with no queue, which happens when  $\mu_k < \mu_{q^{\text{in}}}^{\text{sat}}$ , where

$$\mu_{q^{\text{in}}}^{\text{sat}} = \frac{1}{a} \left( (q^{\text{hi}} - q^{\text{lo}}) \tau^{\text{max}} - (q^{\text{in}} - q^{\text{lo}}) \tau_\pi \right). \quad (6.25)$$

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

$$\mu_{k+1} = a\mu_k + b, \quad (6.26)$$

$$a = \frac{q^{\text{hi}} - q^{\text{lo}}}{q_b^{\text{dis}} - q^{\text{lo}}} > 1, \quad (6.27)$$

$$b = \tau_\pi (q^{\text{in}} - q_b^{\text{dis}}) + n_\pi - \tau^{\text{max}} (q^{\text{hi}} - q_b^{\text{dis}}) < 0. \quad (6.28)$$

Therefore, the excess congestion (6.26) will decrease if

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

Starting with  $\mu_0$ , we can calculate  $\mu_k$  by recursing (6.26),

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

Note that since the discharging flow  $q_b^{\text{dis}}$  by definition lies between  $q^{\text{lo}}$  and  $q^{\text{hi}}$ , we may represent it as a weighted average between these two values,

$$q_b^{\text{dis}} = \frac{a-1}{a} q^{\text{lo}} + \frac{1}{a} q^{\text{hi}},$$

where  $a$  is given by (6.27), 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  $\mu_0$ , the transition into the second phase of congestion dissipation happens after  $k^{\text{sat}}$  platoons, where  $k^{\text{sat}}$  is the lowest integer such that

$$a^{k^{\text{sat}}} \mu_0 + \sum_{i=1}^{k^{\text{sat}}-1} a^i b \leq \mu_{q^{\text{in}}}^{\text{sat}}.$$

The recovery phase is characterized by the lack of congestion at the stationary bottleneck,  $n_b(t) = 0$ , i.e. all the congestion is in the queues at the platoons.

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

$$\tau_{k^{\text{sat}}}^u = \frac{\mu_{k^{\text{sat}}}}{q_b^{\text{dis}} - q^{\text{lo}}} + \tau_\pi \frac{q^{\text{in}} - q^{\text{lo}}}{q^{\text{hi}} - q^{\text{lo}}},$$

travelling at speed  $u_{k^{\text{sat}}} = \frac{X_b - X_0}{\tau_{k^{\text{sat}}}^u}$ . Since the stationary bottleneck will be in free flow, starting with  $k = k^{\text{sat}}$ , the excess congestion will follow

$$\mu_{k+1} = \mu_k - \tau_\pi (q^{\text{hi}} - q^{\text{in}}),$$

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

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

The dynamics of  $\mu_k$  through both phases of congestion dissipation can be jointly described as

$$\mu_{k+1} = \begin{cases} a\mu_k + b, & \mu_k \geq \mu_{q^{\text{in}}}^{\text{sat}}, \\ \mu_k - \tau_\pi (q^{\text{hi}} - q^{\text{in}}), & \tau_\pi (q^{\text{hi}} - q^{\text{in}}) \leq \mu_k < \mu_{q^{\text{in}}}^{\text{sat}}, \\ 0, & \mu_k < \tau_\pi (q^{\text{hi}} - q^{\text{in}}), \end{cases}$$

with  $a$  and  $b$  given by (6.27) and (6.28).

We summarize the analysis in this proposition:

**Proposition 1.** *Assuming constant inflow  $q^{\text{in}}(t) = q^{\text{in}}$ , periodic arrival of platoons with period  $\tau_\pi$  and ordering of flow values (6.24), the queue length  $n_b(t)$  of system (6.15)–(6.23) controlled by control law (6.12) is stable and will remain 0 after some time  $t$ , if the initial queue length satisfies*

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

where  $a$  and  $b$  are given by (6.27) and (6.28), respectively. Furthermore, if this condition is satisfied, the system returns to the unperturbed state with  $n_b(t) = 0$  and  $\tilde{n}_\xi(t) = 0$  after platoon  $k^{\text{rec}}$  reaches the bottleneck.

Conversely, substituting  $b = (1 - a)\mu_0$  into (6.28), we may derive the maximum  $q^{\text{in}}$  for which the system will be stable for a given  $\mu_0$ ,

$$q_{\mu_0}^{\text{in}} = q_b^{\text{dis}} - \frac{n_\pi}{\tau_\pi} + \frac{a-1}{a}(q^{\text{hi}} - q^{\text{lo}})\frac{\tau^{\text{max}}}{\tau_\pi} - (a-1)\frac{\mu_0}{\tau_\pi}. \quad (6.29)$$

### 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 (6.26), which then becomes equal to:

$$\mu_{k+1} = a\mu_k + b + \varepsilon_k, \quad (6.30)$$

where  $\varepsilon_k = \varepsilon_k^{\tau_\pi}(q^{\text{in}} - q_b^{\text{dis}}) + \tau_\pi\varepsilon_k^{q^{\text{in}}} + \varepsilon_k^{\tau_\pi}\varepsilon_k^{q^{\text{in}}}$  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  $\tau_\pi$ , and  $\varepsilon_k^{q^{\text{in}}}$  is the difference of the average inflow from  $q^{\text{in}}$  during that time. We may also write

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

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

$$\mu_k < \frac{b+E}{1-a}, \quad (6.31)$$

*with  $a$  given by (6.27) and  $b$  by (6.28), then system (6.30) is stable, so we are able to dissipate the congestion at the bottleneck. Conversely, if for any  $k$  we have*

$$\mu_k > \frac{b-E}{1-a}, \quad (6.32)$$

*then system (6.30) is unstable, so in that case the queue at the bottleneck will grow unbounded.*

Consequently the conclusions about stability can be extended to system (6.15)–(6.23) if a suitable bound on uncertainty  $E$  can be derived.

For the initial excess congestion between these two values,  $\frac{b+E}{1-a} < \mu_0 < \frac{b-E}{1-a}$ ,  $\mu_k$  will almost surely satisfy either condition (6.31) or (6.32) for some  $k$ , after which the queue stability does not depend on  $\varepsilon_k$ . Assuming uniformly distributed  $\varepsilon_k$ , with  $\mathbb{E}\{\varepsilon_k\} = 0$ ,  $\text{Var}\{\varepsilon_k\} = \frac{E^2}{3}$  (e.g., if  $\varepsilon_k \sim \mathcal{U}[-E, E]$ ), the probability of  $\mu_k$  satisfying (6.32) (i.e., failing to decongest the bottleneck) closely follows the logistic curve depending on  $\mu_0$ ,

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

and the probability of  $\mu_k$  satisfying (6.31) for some  $k$  (i.e., successfully decongesting the bottleneck) is  $\mathcal{P}_{\text{sta}}(\mu_0) = 1 - \mathcal{P}_{\text{uns}}(\mu_0)$ .

Finally, given an appropriately chosen  $\mu_0$ , we may define the estimate of throughput of the controlled system as the maximum  $q^{\text{in}}$  for which the control algorithm is able to decongest the bottleneck with probability  $\mathcal{P}^*$ . This yields the bound on  $q^{\text{in}}$ ,

$$\begin{aligned} q_{\mu_0, \Delta}^{\text{in}} &= q_{\text{b}}^{\text{dis}} - \frac{n_{\pi}}{\tau_{\pi}} + \frac{q^{\text{hi}} - q_{\text{b}}^{\text{dis}}}{\tau_{\pi}} \left( \tau^{\text{max}} - \frac{\mu_0 + \Delta}{q_{\text{b}}^{\text{dis}} - q^{\text{lo}}} \right), \\ q_{\mu_0, \Delta}^{\text{in}} &= q_{\text{b}}^{\text{dis}} - \frac{n_{\pi}}{\tau_{\pi}} + \frac{a-1}{a} (q^{\text{hi}} - q^{\text{lo}}) \frac{\tau^{\text{max}}}{\tau_{\pi}} - (a-1) \frac{\mu_0 + \Delta}{\tau_{\pi}}, \end{aligned} \quad (6.33)$$

where

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

is the measure of combined uncertainty in the system. Moreover, if we substitute  $\Delta = 0$  into (6.33), we recover (6.29). This means that the deterioration of  $q_{\mu_0, \Delta}^{\text{in}}$  due to the introduction of varying inflow and platoon arrivals is

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

Note that this bound is only valid for  $\mu_0 > \mu_{q_{\mu_0, \Delta}^{\text{in}}}^{\text{sat}}$ , for which the dissipation process starts in the saturation phase. Substituting (6.33) into (6.25), we find the minimum  $\mu_0$  for which this holds,

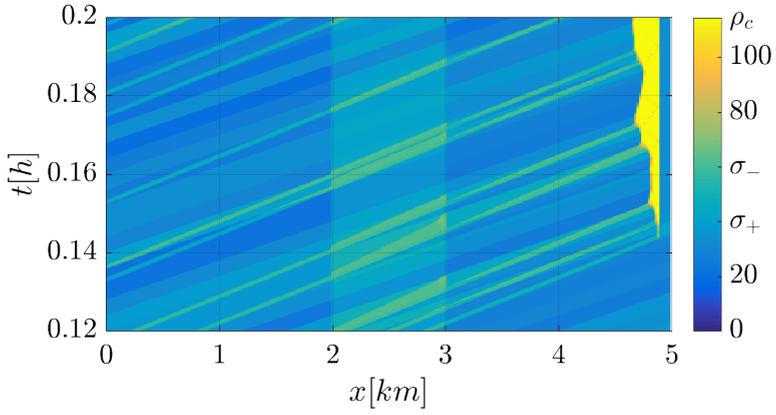
$$\mu_0^{\text{sat}} = \frac{(q^{\text{hi}} - q^{\text{lo}}) (\tau^{\text{max}} - \tau_{\pi})}{a} + n_{\pi} + (a-1)\Delta,$$

in which case we have

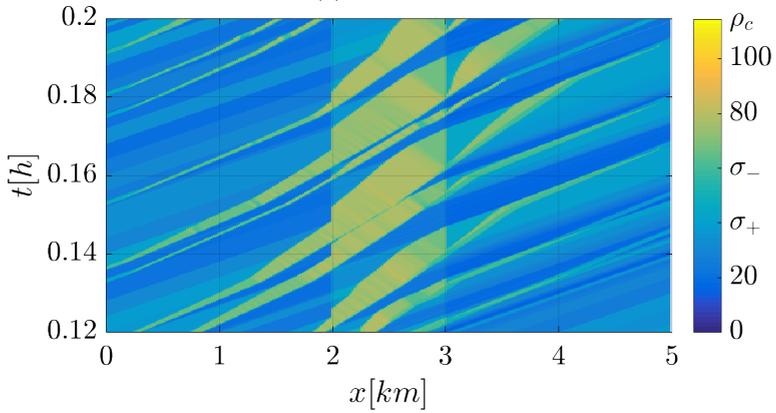
$$q_{\mu_0^{\text{sat}}, \Delta}^{\text{in}} = q^{\text{hi}} - a \frac{n_{\pi}}{\tau_{\pi}} - a(a-1) \frac{\Delta}{\tau_{\pi}}. \quad (6.34)$$

### 6.3.3 Simulation setup

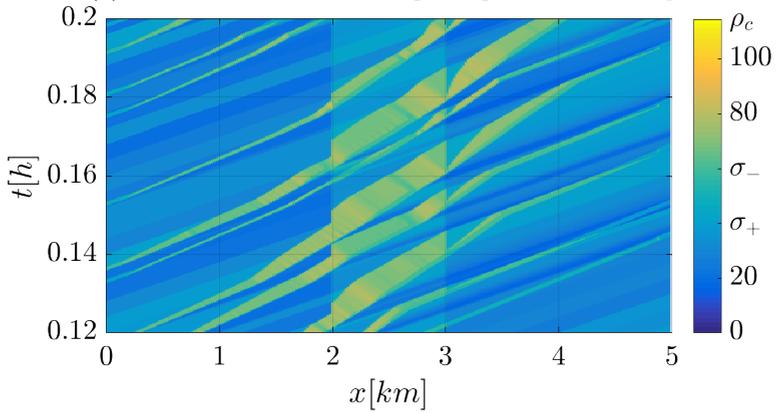
The proposed control laws were tested in simulations, executed on a 5 km long stretch of highway, illustrated by Figure 6.5, with an on-ramp around the 2 km mark, and an off-ramp around the 3 km mark. Most of the highway stretch has three lanes, corresponding to a critical density of  $\sigma^- = 60$  veh/km and capacity of  $q_-^{\text{max}} = 6000$  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  $q_+^{\text{max}} = 4000$  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 for this road configuration.



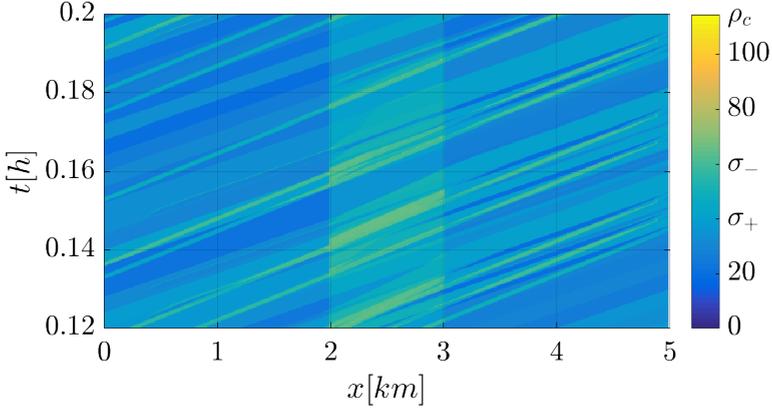
(a) No control



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



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



(d) Ideally actuated control

Figure 6.6: An example comparing the outcome of the four simulation cases. Traffic density is color-coded, with warmer color representing higher density.

The simulation model we used was the multi-class CTM, given in Section 4.3, 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 Poisson arrival rate of  $\lambda = 81$  platoon/h,  $\tau_\pi = 0.0123$  h. We assume that each platoon consists of 2 passenger car equivalents, although in reality, due to having shorter inter-vehicular gaps, these platoons might be up to about five passenger cars or about three trucks long. 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 demand of mainstream-bound background traffic takes values in  $\phi_1^b(t) \sim \mathcal{U}(1000, 2000)$  veh/h, and the demand of off-ramp bound traffic is  $\phi_1^c(t) \sim \mathcal{U}(750, 1250)$  veh/h. Since the on-ramp and off-ramp are close, we assume that none of the vehicles entering the highway via the on-ramp will exit it via the off-ramp,  $\phi_{i_{\text{on}}}^c(t) = 0$  veh/h. The inflow of mainstream-bound traffic at the on-ramp is modelled as  $\phi_{i_{\text{on}}}^b(t) \sim \mathcal{U}(900, 1500)$  veh/h.

With the parameters specified in the previous paragraph, we may calculate an estimate of the throughput that we may achieve by applying the presented control law. Using (6.34) with  $E = \tau_\pi (\max(\phi_1^b + \phi_{i_{\text{on}}}^b) - \mathbb{E}\{\phi_1^b + \phi_{i_{\text{on}}}^b\})$  and  $\mathcal{P}^* = 0.9$ , yielding  $\Delta \approx 5.4$ , we estimate that the throughput would be improved from  $q_{\text{unc}}^{\text{in}} + \frac{n_\pi}{\tau_\pi} = 3273$  veh/h to  $q_{\mu_0, \Delta}^{\text{in}} + \frac{n_\pi}{\tau_\pi} = 3513.2$  veh/h. Note that in deriving (6.34) 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 and

ensure fair comparison between different control laws. Simulations are done with four cases of control:

- (n.c.) No control.
- (i.r.) Platoon-actuated control ignoring on- and off-ramps, with  $\tilde{q}_\xi^{\text{cap}}(t)$  given by (6.12).
- (c.r.) Platoon-actuated control taking on- and off-ramps into consideration, with  $\tilde{q}_\xi^{\text{cap}}(t)$  given by (6.14).
- (i.a.) Ideally actuated control, with  $U_i^b(t)$  given by (6.11).

In order to demonstrate the effect applying these control laws has on the traffic, a part of one simulation run is shown in Figure 6.6.

#### 6.3.4 Simulation results

Consider the uncontrolled case shown in Figure 6.6a. Around time  $t = 0.144$  h, the aggregate density of the platooned vehicles and background traffic arriving at the bottleneck is too high, and the aggregate demand exceeds bottleneck capacity. This causes a traffic breakdown, and after a brief transient, congestion is formed and bottleneck capacity is reduced. Because of this, even though the incoming traffic density is lower after  $t = 0.154$  h, and would 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 stay congested until the inflow to the highway segment is reduced close to the end of the simulation run.

In contrast to this, in the ideally actuated case shown in Figure 6.6d, the mainstream-bound background traffic is controlled directly. A part of this traffic is delayed so that when there is a platoon arriving at a stationary bottleneck, the density of the remaining traffic is low enough so as not to cause traffic breakdown and capacity drop. In this way, free flow is maintained and throughput is close to its theoretical maximum.

As shown in Figure 6.6b and Figure 6.6c, the performance of the two proposed control laws is similar. However, in case the influence of on- and off-ramps is ignored while predicting the evolution of the system, the applied control action is more severe than required. This leads to more congestion upstream of the off-ramp and overall lower efficiency. The control law that takes the on- and off-ramps into account comes close to emulating the ideal actuation case. However, it achieves somewhat worse performance because it is unable to selectively affect only one class of background traffic, since it only has access to the average splitting ratio for the off-ramp, and requires delaying the platoons.

We executed 50 Monte Carlo simulations, with the same platoon arrival times and background traffic inflow profiles for each control case. The resulting average

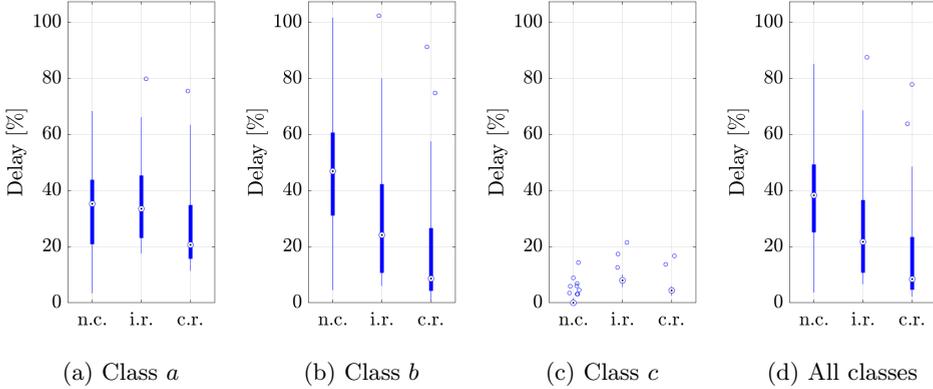


Figure 6.7: Box plots showing the increase in TTS compared to the ideal actuation case. The different cases of control are: no control (n.c.), ignoring on- and off-ramps (i.r.) and considering on- and off-ramp (c.r.).

Table 6.1: Average and median TTS for each vehicle class and all vehicles.

TTS [veh $h$ ]	Class $a$		Class $b$		Class $c$		Total	
	average	median	average	median	average	median	average	median
(n.c.)	22.62	22.94	369.84	374.13	56.62	56.04	449.08	453.94
(i.r.)	23.25	23.03	329.91	315.75	60.62	60.37	413.78	398.18
(c.r.)	21.77	21.34	304.90	278.63	58.60	58.41	385.27	357.58
(i.a.)	17.00	16.91	255.00	254.09	55.92	55.93	327.92	326.42

Table 6.2: Average and median delay for individual vehicle classes and all vehicles.

Delay [%]	Class $a$		Class $b$		Class $c$		Total	
	average	median	average	median	average	median	average	median
(n.c.)	33.1	35.3	45.0	46.9	1.2	0.0	36.9	38.3
(i.r.)	36.8	33.5	29.4	24.1	8.4	8.0	26.2	21.7
(c.r.)	28.1	20.7	19.6	8.6	4.8	4.4	17.5	8.4

and median TTS are shown in Table 6.1. We show the TTS of each vehicle class, and for all vehicles combined. Apart from comparing the TTS, we also considered the delay, defined as the difference in TTS compared to the ideal actuation case, which is taken as a benchmark for minimum achievable TTS of each simulation run. The delay is shown as percentage of minimum travel time, and it is shown as a box plot in Figure 6.7, and given in Table 6.2. For example, if a vehicle would traverse the road segment in 3 minutes if it travelled at free flow speed, and actually traverses it in 4.5 minutes, we say that it had a 50% delay.

We can see that even by applying control that ignores the existence of on- and off-ramps, as described in Section 6.3.1, 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, when the control from Section 6.3.1 was used, the TTS of both class *a* (the platooned vehicles) 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 action, their TTS will be lower, since they will avoid waiting in congestion upstream of 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 not to optimize the overall traffic performance, but to minimize only their own travel time. The TTS of class *c* vehicles is still increased compared to the uncontrolled case, but less so 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%.

It is notable that while the proposed control laws achieve significant reduction of both average and median TTS, there is a number of outliers corresponding to particularly unfavourable simulation runs. Since the arrival of platoons is modelled as a Poisson process, we can expect to occasionally have long gaps between two platoons. If this occurrence coincides with a higher demand of mainstream-bound background traffic, we will not be able to prevent the traffic breakdown, since there would be no platoons available to actuate the control action, resulting in a build-up of congestion and higher TTS.

## 6.4 Summary

Decongesting the traffic bottlenecks, especially those of non-recurrent nature, is a challenging traffic control task. The equipment necessary for most classical traffic control methods might not be available at the location where a non-recurrent bottleneck arises. For this reason, an alternative, Lagrangian approach is preferred. In this chapter, we used controlled moving bottleneck to decongest stationary bottlenecks and keep them in free flow.

First, a control law that reacts to stochastic traffic breakdown is presented and tested in simulation. While it effectively decongests the stationary bottleneck, the described control law is derived for worst case and does not consider the stochastic nature of bottleneck dynamics, leading to some inefficiencies. Furthermore, the proposed control law only reacts to traffic breakdowns that have already happened, and will not attempt to pre-empt it by spreading the inflow to the bottleneck more evenly over time, thus making it more likely that the traffic will stay in free flow.

Next, we study the deterministic case in more detail, proposing a prediction-based control law for platoon coordination so that the stationary bottleneck is decongested and kept in free flow. We employ a tandem queueing model with moving bottlenecks as the prediction model, and the control law uses platoon speed and formation as control inputs. The performance of these control laws is tested in multi-class CTM simulations, on a 5 km long stretch of highway upstream of a bottleneck at an accident site, going from three lanes to two. The considered highway segment also includes an on-ramp and an off-ramp. The achieved TTS using these control laws is compared to the case when no control is used, as well as with the case when we have ideal actuation, and can fully control all individual vehicles. It has been 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.

## Chapter 7

# Single stop-and-go wave dissipation

STOP-AND-GO waves are a peculiar traffic phenomenon which is well-known empirically, but somewhat problematic to describe in traffic models. They typically arise due to a disturbance in the traffic flow, and get amplified as they propagate upstream due to the hysteresis effect of acceleration and deceleration in human driving behaviour, until they are dissipated. Such irregularities in the traffic flow are not only bad for the throughput of the road, they also cause increased fuel consumption and are a potential safety hazard. For all these reasons, stop-and-go wave dissipation is a control problem worth investigating. Classical traffic control approaches can have difficulties with stop-and-go wave dissipation, since this moving phenomenon can arise at locations where traffic control equipment is not available. And while control approaches using variable speed limits have been proposed, with one notable example being SPECIALIST [67], stop-and-go wave dissipation remains an ideal control problem for Lagrangian traffic control paradigm.

In this chapter, we study the simplest case of stop-and-go wave dissipation using controlled moving bottlenecks, using a single controlled moving bottleneck to dissipate a single stop-and-go wave. In Section 7.1, the stop-and-go wave problem is discussed in more detail, first using a general traffic flow control framework, and then focusing on using controlled moving bottlenecks. We approach the problem from two angles. First, in Section 7.2, we assume that the controlled moving bottleneck, e.g. a heavy-duty vehicle, is already present on the road, and control its speed so that the stop-and-go wave is dissipated without creating excessive additional disturbance to the traffic. This approach is suitable for use on highways with few lanes, due to limitations on the control action that a single vehicle can actuate. Then, in Section 7.3.1, we generalize this approach by adding an additional phase to the control, in which we gather multiple connected vehicles into a more severe controlled moving bottleneck, in order to be able to apply stronger control. Once enough connected vehicles are gathered, we may proceed to use them to dissipate the stop-and-go wave in a similar way.

## 7.1 Single stop-and-go wave dissipation problem

Stop-and-go wave dissipation problem is a variant of the congestion reduction problem, outlined in Subsection 1.2.2, where the outflow from the considered road segment is limited by the discharging flow from some stop-and-go wave. The objective of the control is to minimize the total time spent (TTS) of all vehicles, which is achieved by maximizing the early outflow from the road,

$$J_{\text{cr}} = \int_{t_0}^{t_{\text{end}}} (t - t_{\text{end}}) q^{\text{out}}(t) dt.$$

As discussed in Section 4.1, the presence of stop-and-go waves will reduce the throughput of the road, so we endeavour to accelerate their dissipation using whatever means of traffic control we have available.

Consider a road segment with constant free flow speed  $V$ , and a single stop-and-go wave at some position  $z_{\psi}(t_0)$ , propagating upstream with some constant velocity  $\lambda_{\psi} < 0$ ,

$$z_{\psi}(t) = z_{\psi}(t_0) + \lambda_{\psi}(t - t_0), \quad t \leq t_{\psi}^{\text{c}},$$

where  $t_{\psi}^{\text{c}}$  is the time when the stop-and-go wave is cleared. We write the accumulation of congestion at the stop-and-go wave  $n_{\psi}(t)$ , and

$$\dot{n}_{\psi}(t) = \frac{V - \lambda_{\psi}}{V} (q_{\psi}^{\text{in}}(t) - q_{\psi}^{\text{dis}}), \quad t \leq t_{\psi}^{\text{c}},$$

where at some initial time  $t_0$ ,  $n_{\psi}(t_0) = 0$ ,  $q_{\psi}^{\text{in}}(t_0) > q_{\psi}^{\text{dis}}$ , and with  $n_{\psi}(t_{\psi}^{\text{c}}) = 0$ ,  $q_{\psi}^{\text{in}}(t_{\psi}^{\text{c}}) < q_{\psi}^{\text{dis}}$ , i.e. the stop-and-go wave dissipates at time  $t_{\psi}^{\text{c}} > t_0$ , when

$$n_{\psi}(t_{\psi}^{\text{c}}) = \int_{t_0}^{t_{\psi}^{\text{c}}} \frac{V - \lambda_{\psi}}{V} (q_{\psi}^{\text{in}}(t) - q_{\psi}^{\text{dis}}) dt = 0, \quad n_{\psi}(t) > 0, t \in (t_0, t_{\psi}^{\text{c}}). \quad (7.1)$$

If the road is in free flow downstream of the stop-and-go wave, the traffic flow that discharges from the stop-and-go wave at time  $t$  reaches the downstream boundary of the road  $X_{\text{b}}$  at time  $t_{\psi}^{\text{V}}(t)$ , given by

$$t_{\psi}^{\text{V}}(t) = t + \frac{X_{\text{b}} - z_{\psi}(t)}{V} = \frac{X_{\text{b}} - z_{\psi}(t_0) + (V - \lambda_{\psi})t + \lambda_{\psi}t_0}{V},$$

and we have

$$q^{\text{out}}(t) = q_{\psi}^{\text{dis}}, \quad t \leq t_{\psi}^{\text{d}},$$

where  $t_{\psi}^{\text{d}} = t_{\psi}^{\text{V}}(t_{\psi}^{\text{c}})$ . Conversely, we denote by  $t_{\psi}^{-\text{V}}(t)$  the inverse of  $t_{\psi}^{\text{V}}(t)$ ,

$$t_{\psi}^{-\text{V}}(t) = \frac{Vt - X_{\text{b}} + z_{\psi}(t_0) - \lambda_{\psi}t_0}{V - \lambda_{\psi}}.$$

Assume that we can influence the traffic flow on the road upstream of the stop-and-go wave, so that we are able to control  $q_\psi^{\text{in}}(t)$  after some time  $t_u$ . We may then use this control to expedite the dissipation of the stop-and-go wave, which can improve the throughput and reduce the TTS. Let  $q_\psi^{\text{in}}(t) = q_u^{\text{out}}(t)$  for  $t \geq t_u$ , where  $q_u^{\text{out}}(t)$  denotes the outflow from some regulated accumulation of vehicles,

$$\begin{aligned} \dot{n}_u(t) &= \frac{V - \lambda_\psi}{V} (q_u^{\text{in}}(t) - q_u^{\text{out}}(t)), \quad t \in [t_u, t_\psi^c], \\ q_u^{\text{out}}(t) &= \begin{cases} \min\{q_u^{\text{in}}(t), q_u^{\text{cap}}(t)\}, & n_u(t) = 0, \\ q_u^{\text{cap}}(t), & n_u(t) > 0, \end{cases} \end{aligned}$$

with  $n_u(t_u) = 0$ , and assume we can control  $q_u^{\text{cap}}(t)$  within some range  $[q^{\text{lo}}, q^{\text{hi}}]$ . If  $q^{\text{lo}} < q_\psi^{\text{dis}}$ , we may dissipate the stop-and-go wave by time  $t_\psi^c$  by setting  $q_u^{\text{cap}}(t) = q^{\text{lo}}$  for  $t \leq t_\psi^c$ , in which case

$$t_\psi^c = t_\psi^{-V} (t_\psi^{\text{d}}), \quad t_\psi^{\text{d}} = t_\psi^V (t_u) + \frac{n_\psi(t_u)}{q_\psi^{\text{dis}} - q^{\text{lo}}},$$

assuming  $\int_{t_u}^t q_u^{\text{in}}(\tau) d\tau > (t - t_u)q^{\text{lo}}$  for  $t \in (t_u, t_\psi^{\text{d}})$ . After the stop-and-go wave is dissipated, it ceases to obstruct the traffic flow, and we will be able to control

$$q^{\text{out}}(t) = q_u^{\text{out}} \left( t - \frac{X_b - X_u}{V} \right), \quad t > t_\psi^{\text{d}},$$

assuming we control the traffic flow at position  $X_u = z_\psi(t_\psi^c)$  for  $t \geq t_\psi^c$ . We may then set  $q_u^{\text{cap}}(t) = q^{\text{hi}}$  for  $t \in (t_\psi^c, t_u^c]$  in order to discharge the accumulated traffic  $n_u(t)$ ,

$$\dot{n}_u(t) = q_u^{\text{in}}(t) - q_u^{\text{out}}(t), \quad t \in (t_\psi^c, t_u^c],$$

by time  $t_u^c$ , given by

$$n_u(t_\psi^c) + \int_{t_\psi^c}^{t_u^c} q_u^{\text{in}}(t) dt = q^{\text{cap}}(t_u^c - t_\psi^c),$$

and keep the road in free flow afterwards. The traffic flow discharging from  $n_u(t)$  at time  $t \in (t_\psi^c, t_u^c)$  reaches the downstream boundary of the road at time  $t + (X_b - X_u)/V$ , and we denote  $t_u^{\text{d}} = t_u^c + (X_b - X_u)/V$ .

In a simplified case when the stop-and-go wave is at the downstream boundary of the road segment at time  $t_0$ ,  $z_\psi(t_0) = X_b$ , the inflow  $q_u^{\text{in}}(t) = q^{\text{in}}$  is constant for  $t \geq t_u$ , and the congestion at the stop-and-go wave at time  $t_u$  is given, the congestion reduction cost function (1.2) simplifies to

$$J_{\text{cr}} = \left( \frac{t^2}{2} - t_{\text{end}} t \right) \Big|_{t_0}^{t_\psi^{\text{d}}} q_\psi^{\text{dis}} + \left( \frac{t^2}{2} - t_{\text{end}} t \right) \Big|_{t_\psi^{\text{d}}}^{t_u^{\text{d}}} q^{\text{hi}} + \left( \frac{t^2}{2} - t_{\text{end}} t \right) \Big|_{t_u^{\text{d}}}^{t_{\text{end}}} q^{\text{in}},$$

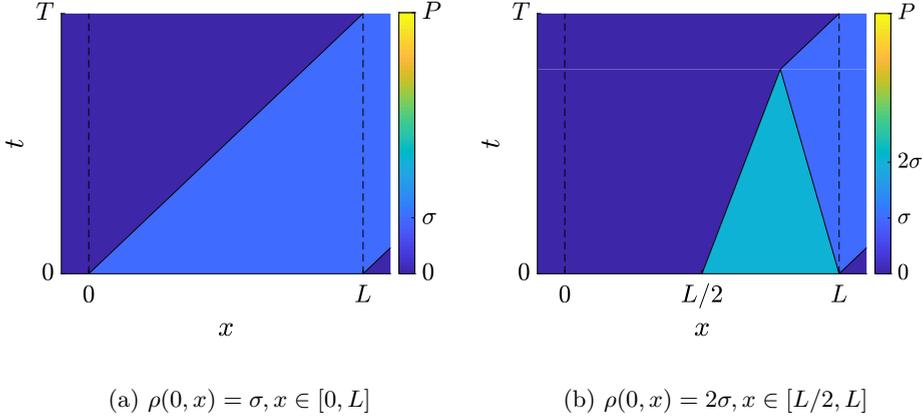


Figure 7.1: Example of two situations with the same TTS, but different T.V.

with  $t_\psi^d$  and  $t_u^d$  given by

$$t_\psi^d = t_\psi^V(t_u) + \frac{n_\psi(t_u)}{q_\psi^{\text{dis}} - q^{\text{lo}}}, \quad t_u^d = t_\psi^d + \frac{q^{\text{in}} - q^{\text{lo}}}{q^{\text{hi}} - q^{\text{in}}} (t_\psi^d - t_\psi^V(t_u)).$$

In comparison, if no control is applied, the cost function is

$$J_{\text{cr}}^{\text{nc}} = \left( \frac{t^2}{2} - t_{\text{end}} t \right) \Big|_{t_0}^{t_\psi^{\text{d,nc}}} q_\psi^{\text{dis}} + \left( \frac{t^2}{2} - t_{\text{end}} t \right) \Big|_{t_\psi^{\text{d,nc}}}^{t_{\text{end}}} q^{\text{in}},$$

$$t_\psi^{\text{d,nc}} = t_\psi^d + \frac{n_\psi(t_u)(q^{\text{in}} - q^{\text{lo}})}{(q_\psi^{\text{dis}} - q^{\text{in}})(q_\psi^{\text{dis}} - q^{\text{lo}})},$$

and the improvement in TTS achieved by applying control is

$$J_{\text{cr}}^{\text{nc}} - J_{\text{cr}} = \left( \frac{t^2}{2} - t_{\text{end}} t \right) \Big|_{t_\psi^d}^{t_u^d} (q_\psi^{\text{dis}} - q^{\text{hi}}) + \left( \frac{t^2}{2} - t_{\text{end}} t \right) \Big|_{t_u^d}^{t_\psi^{\text{d,nc}}} (q_\psi^{\text{dis}} - q^{\text{in}}).$$

Note that the cumulative outflow for the two cases is equal for  $t > t_\psi^{\text{d,nc}}$ , and

$$(t_u^d - t_\psi^d) (q_\psi^{\text{dis}} - q^{\text{hi}}) = (t_\psi^{\text{d,nc}} - t_u^d) (q_\psi^{\text{dis}} - q^{\text{in}})$$

so in essence, the control has shifted a part of the outflow to an earlier time, leading to a reduction in TTS.

Unsurprisingly, in case there is no capacity drop at the stop-and-go wave, the achieved TTS in the controlled and uncontrolled case will be equal, since in that case  $q_\psi^{\text{dis}} = q^{\text{cap}} = q^{\text{hi}}$ , and  $t_u^d = t_\psi^{\text{d,nc}}$ . This is due to the fact that TTS does not directly

depend on the traffic conditions on the road, only on the outflow, which might be identical for very different traffic conditions, as shown in Figure 7.1. Therefore, there is a need for an additional cost function that indicates how favourable the current traffic profile is. One such cost function is the Average Total Variation (ATV) of the smoothed traffic density profile  $\bar{\rho}(x, t)$ , i.e. the Total Variation (T.V.),

$$\text{T.V.}(\bar{\rho}(x, t)) = \int_0^{X_b} \left| \frac{\partial \bar{\rho}(x, t)}{\partial x} \right| dx,$$

averaged over the considered time horizon, from  $t_0$  to  $t_{\text{end}}$ . Since the traffic density might be discontinuous due to shocks, we need to apply spatial smoothing before differentiation, which we do here using a spatial moving average filter with window length  $L$ . We may then write the ATV as

$$\text{ATV} = \int_{t_0}^{t_{\text{end}}} \int_L^{X_b} \frac{|\rho(x, t) - \rho(x - L, t)|}{(t_{\text{end}} - t_0)L} dx dt.$$

We may use the ATV as a measure of traffic homogeneity, where lower ATV (higher homogeneity) is preferable to higher ATV (lower homogeneity). Essentially, the existence of a stop-and-go wave corresponds to high total variation of traffic density, so if the stop-and-go wave is dissipated quicker, we can expect the ATV to be lower. Note that if the traffic density  $\rho(x, t)$  is given as an average value inside  $N$  cells of length  $L$ ,  $X_b = NL$ , during time periods of length  $T$ , with  $t_{\text{end}} = t_{\text{end}}^T T$ ,  $t_{\text{end}}^T \in \mathbb{N}$  and  $t_0 = t_0^T T$ ,  $t_0^T \in \mathbb{N}$ ,

$$\rho(x, t) = \rho_i(t^T), x \in [(i-1)L, iL),$$

where  $t^T = \lfloor t/T \rfloor$ , the ATV simplifies further to

$$\text{ATV} = \sum_{t^T=t_0^T}^{t_{\text{end}}^T-1} \sum_{i=2}^N \frac{|\rho_i(t^T) - \rho_{i-1}(t^T)|}{t_{\text{end}}^T - t_0^T}.$$

## 7.2 Moving bottleneck control

While there are many traffic control schemes that can regulate the inflow of traffic to a stop-and-go wave, e.g. using variable speed limits [67], here in particular we are interested in using controlled moving bottlenecks as Lagrangian actuators. In this section we adapt the general stop-and-go wave dissipation control to this case. The designed control law will be tested in simulations, and used as a component of the overall control strategy throughout the remainder of this thesis.

### 7.2.1 Control law

Consider a road segment with traffic density  $\rho(x, t)$ , with downstream boundary in  $X_b$ , where the traffic travels at a constant speed  $V$  in free flow. Let there be a single CAV at some position  $x_\xi(t)$ , whose reference speed  $u_\xi(t)$  we can directly control within some range,  $u_\xi(t) \in [u_\xi^{\min}, u_\xi^{\max}]$ , with  $u_\xi^{\min} \geq 0$  and  $u_\xi^{\max} \leq V$ . If the road downstream of the CAV is in free flow, the traffic flow overtaking it at time  $t$  reaches the downstream boundary of the road at time

$$t_\xi^V(t) = t + \frac{X_b - x_\xi(t)}{V},$$

and we write the inverse of this transformation  $t_\xi^{-V}(t), t_\xi^{-V}(t_\xi^V(t)) = t$ . In case there is a stop-and-go wave downstream of the CAV,  $z_\psi(t) > x_\xi(t)$ , with the downstream front position of the stop-and-go wave denoted  $z_\psi(t)$ , we can attempt to control  $u_\xi(t)$  in order to have the CAV act as a moving bottleneck and help dissipate the congestion. The overtaking flow from the moving bottleneck at  $x_\xi(t)$ , which we denote  $q_\xi^{\text{out}}(t)$ , reaches the stop-and-go wave at time  $t_\psi^{-V}(t_\xi^V(t))$ , and we may write

$$q_\psi^{\text{in}}(t_\psi^V(t)) = q_\xi^{\text{out}}(t_\xi^V(t)), \quad q_\psi^{\text{in}}(t) = q_\xi^{\text{out}}(t_\xi^{-V}(t_\psi^V(t))),$$

until the stop-and-go wave is dissipated at time  $t_\psi^c$ , or the CAV itself reaches its downstream front at some time  $t_\xi^\psi$ ,  $x_\xi(t_\xi^\psi) = z_\psi(t_\xi^\psi)$ . Here,  $q_\xi^{\text{out}}(t)$  is given as the outflow from the queue at the moving bottleneck,

$$\begin{aligned} \dot{n}_\xi &= \frac{V - u_\xi(t)}{V} (q_\xi^{\text{in}} - q_\xi^{\text{out}}(t)), \\ q_\xi^{\text{out}}(t) &= \begin{cases} \min\{q_\xi^{\text{in}}(t), q_\xi^{\text{cap}}(t)\}, & n_\xi(t) = 0, \\ q_\xi^{\text{cap}}(t), & n_\xi(t) > 0, \end{cases} \end{aligned}$$

and we have  $q_\xi^{\text{out}}(t) \leq q_\xi^{\text{cap}}(t)$ .

We can therefore attempt to emulate the control action described in the previous section using the moving bottleneck, with  $t_u = t_\psi^{-V}(t_\xi^V(t_0))$ , if the (possibly speed-dependent) capacity limitation of the moving bottleneck,  $q_\xi^{\text{cap}}(t) = q_\xi^{\text{lo}}(u_\xi(t))$ , restricts the flow more severely than the stop-and-go wave,  $q_\xi^{\text{lo}}(u_\xi(t)) < q_\psi^{\text{dis}}$ . Assuming the inflow to the moving bottleneck  $q_\xi^{\text{in}}(t)$  is higher than  $q_\xi^{\text{lo}}(u_\xi(t))$ , the evolution of the accumulated traffic at the stop-and-go wave is then given by

$$n_\psi(t) = n_\psi(t_u) - (t_\psi^V(t) - t_\psi^V(t_u))q_\psi^{\text{dis}} + \int_{t_\psi^V(t_u)}^{t_\psi^V(t)} q_\xi^{\text{lo}}(u_\xi(t_\xi^{-V}(t_\psi^V(t)))) dt > 0, \quad t \in (t_u, t_\psi^c),$$

and  $n_\psi(t_\psi^c) = 0$ .

Since the moving bottleneck only affects the inflow to the stop-and-go wave while it is upstream of it,  $x_\xi(t) < z_\psi(t)$ , in order to dissipate the stop-and-go wave,

we need to have  $t_{\psi}^c \leq t_{\xi}^{\psi}$ . Conversely, in order avoid excessively restricting the traffic flow after the stop-and-go wave has dissipated, we need to ensure that the CAV stops acting as a moving bottleneck after time  $t_{\xi}^c$ ,

$$t_{\xi}^c = t_{\xi}^{-V}(t_{\psi}^d).$$

Assuming  $u_{\xi}^{\max} = V$ , we may do this by setting  $u_{\xi}(t) = V$  for  $t \geq t_{\xi}^c$ . Therefore, the conditions for optimal stop-and-go wave dissipation using a controlled moving bottleneck are

$$\begin{aligned} t_{\xi}^{\psi} &\leq t_{\psi}^c, \\ u_{\xi}(t) &= V, t > t_{\xi}^c. \end{aligned}$$

The simplest choice of  $u_{\xi}(t)$  that satisfies these conditions is

$$u_{\xi}(t) = \begin{cases} u_{\xi}^*, & t \leq t_{\xi}^c, \\ u_{\xi}^{\max}, & t > t_{\xi}^c, \end{cases}$$

where  $u_{\xi}^*$  is given by

$$n_{\psi}(t_u) - (q_{\psi}^{\text{dis}} - q_{\xi}^{\text{lo}}(u_{\xi}^*)) \left( t_{\psi}^V \left( t_0 + \frac{z_{\psi}(t_0) - x_{\xi}(t_0)}{u_{\xi}^* - \lambda_{\psi}} \right) - t_{\psi}^V(t_u) \right) = 0, \quad (7.2)$$

i.e. the moving bottleneck moves with constant speed  $u_{\xi}^*$ , reaches the stop-and-go wave just as it dissipates, and afterwards continues on at free flow speed. Note that in this case we do not need to assume that  $u_{\xi}^{\max} = V$ . If (7.2) yields  $u_{\xi}^* < u_{\xi}^{\min}$ , the moving bottleneck is unable to dissipate the stop-and-go wave.

Assume the traffic dynamics are modelled by the CTM with moving bottlenecks, as discussed in Section 4.2.2. The road is split into  $N$  cells of length  $L$ , with the traffic conditions described by average traffic densities in the cells,  $\rho_i(t^T)$ ,  $i = 1, \dots, N$ , at time instant  $t^T = t_0^T, \dots, t_{\text{end}}^T - 1$ . We write the position of a CAV  $x_{\xi}(t)$ , and the traffic density directly upstream of it is denoted  $\rho_{\xi-}(t^T)$ . The CAV is thus located in cell  $i_{\xi}(t^T) = \lfloor (x_{\xi}(t)/L) \rfloor + 1$ ,  $t = t^T T$ . Along with  $\rho_i(t^T)$ ,  $x_{\xi}(t)$  and  $\rho_{\xi-}(t^T)$  constitute the system state and evolve according to the CTM with moving bottlenecks.

If the cell traffic densities are known, we may use the controlled vehicle to act on the rest of the traffic as a moving bottleneck, restricting the overtaking flow to some value and helping dissipate any potential stop-and-go wave. Here, a stop-and-go wave is defined as a region downstream of  $x_{\xi}(t)$  where the traffic speed is lower than the maximum CAV speed  $u_{\xi}^{\max}$ ,  $\mathcal{V}(\rho_i(t^T)) < u_{\xi}^{\max}$ , and denote the last cell where that is the case  $i_{\psi}(t^T)$ . The position of the downstream front of the stop-and-go wave is thus given by  $z_{\psi}(t^T T) = i_{\psi}(t^T)L$ .

We use the reference speed of the CAV  $u_{\xi}(t^T) \in [u_{\xi}^{\min}, u_{\xi}^{\max}]$  as a control input and endeavour to minimize the TTS. Every time instant  $t^T$ , we calculate

$$u_{\xi}(t^T) = \min(u_{\xi}^{\max}, \max(u_{\xi}^{\min}, u_{\xi}^*(t^T))),$$

where  $u_\xi^*$  is given by (7.2). The parameters  $t_0$ ,  $t_u$ ,  $n_\psi(t_u)$ ,  $q_\psi^{\text{dis}}$ ,  $q_\xi^{\text{lo}}(u)$ , and  $\lambda_\psi$  are given as states and parameters of the traffic model that is used. Time  $t_0 = t^T T$  is the initial time for calculating the control. Since in basic CTM there is no capacity drop, the discharging flow from the stop-and-go wave will be equal to the road capacity,  $q_\psi^{\text{dis}} = V\sigma$ , and the stop-and-go wave propagates upstream with speed  $\lambda_\psi = -W$ . Therefore, we have

$$t_u = t_\psi^{-V}(t_\xi^V(t_0)) = \frac{z_\psi(t_0) - x_\xi(t_0)}{V + W},$$

at which time the accumulated congestion at the stop-and-go wave will be

$$n_\psi(t_u) = (z_\psi(t_0) - x_\xi(t_0))\rho_{x_\xi}^{z_\psi}(t^T) - (t_\psi^V(t_u) - t_\psi^V(t_0))q_\psi^{\text{dis}},$$

where  $\rho_{x_\xi}^{z_\psi}(t^T)$  denotes the average traffic density between the CAV and the stop-and-go wave,

$$\rho_{x_\xi}^{z_\psi}(t^T) = \frac{\sum_{i=i_\xi(t^T)+1}^{i_\psi(t^T)} \rho_i(t^T)}{i_\psi(t^T) - i_\xi(t^T)},$$

since by that time, all the vehicles currently between the CAV and the stop-and-go wave will have entered its congestion, and some will have left it. Finally, the speed-dependent overtaking flow limitation imposed by the moving bottleneck is given by

$$\begin{aligned} q_\xi^{\text{lo}}(u) &= Q(\rho_\xi^{\text{d}}(u)), \\ Q(\rho_\xi^{\text{d}}(u)) - Q_\xi(\rho_\xi^{\text{r}}(u)) &= u(\rho_\xi^{\text{r}}(u) - \rho_\xi^{\text{d}}(u)), \quad \rho_\xi^{\text{d}}(u) \in [0, \sigma_Q^{\text{max}}], \\ \rho_\xi^{\text{r}}(u) &= \arg \max_\rho Q_\xi(\rho) - u\rho, \quad \rho \geq 0. \end{aligned}$$

If the flux function in the zone of the moving bottleneck is a scaled-down version of the flux function elsewhere on the road,

$$Q_\xi^\beta(\rho) = (1 - \beta)Q\left(\frac{\rho}{1 - \beta}\right),$$

the overtaking flow limit  $q_\xi^{\text{lo}}(u_\xi^*)$  no longer depends on the moving bottleneck speed,  $q_\xi^{\text{lo}} = V(1 - \beta)\sigma$ . In this case,  $u_\xi^*(t^T)$  can be expressed explicitly,

$$u_\xi^*(t^T) = \frac{V - \lambda_\psi}{V} \frac{\beta\sigma}{\rho_{x_\xi}^{z_\psi}(t^T) - (1 - \beta)\sigma} + \lambda_\psi. \quad (7.3)$$

We can see that the reference moving bottleneck speed does not depend on the initial distance between the moving bottleneck and the stop-and-go wave, although the average traffic density between them will change as their distance decreases.

We recalculate the controlled moving bottleneck reference speed  $u_\xi(t^T)$  every time instant using the current traffic density profile. Once all congestion downstream of the CAV acting as a moving bottleneck has dissipated, its reference speed is set to maximum  $u_\xi(t) = u_\xi^{\max}$ , and it continues traversing the remainder of the road. The effect of the control action  $u_\xi(t^T)$  will be included in the updated  $\rho_{x_\xi}^{z_\psi}(t^T + 1)$ , thus providing feedback in the control loop.

Note that the CAV will only be able to dissipate the stop-and-go wave if  $u_\xi^*(t_0^T) \geq u_\xi^{\min}$ , in which case, it will avoid going through the congestion with minimum delay. Otherwise, the CAV is not able to improve the TTS, since the congestion that builds up in its wake will get returned to the stop-and-go wave once the controlled vehicle reaches it. Therefore, if it is certain that a CAV will be unsuccessful in stop-and-go wave dissipation, it should continue driving along with the rest of traffic, with reference speed equal to its maximum.

### 7.2.2 Simulation Results

We tested the control law in simulations using the CTM with moving bottlenecks described in Section 4.2.2 as the simulation model. The simulated road consists of  $N = 110$  cells with  $L = 0.5$  km, and  $t_{\text{end}} = 1.5$  h. The traffic flow is modelled with a triangular fundamental diagram, with free flow speed  $V = 110$  km/h, critical density  $\sigma = 40$  veh/km and jam density  $P = 186.66$  veh/km, yielding  $W = 30$  km/h. We assume that a moving bottleneck halves the capacity of the road,  $\beta = 0.5$ ,  $(1 - \beta)\sigma = 20$  veh/km, and  $V_\xi = 110$  km/h.

The simulation scenario in question is as follows:

1.  $t < t_0 = 5$  minutes: The traffic is in free flow, with heterogeneous traffic density. The controlled automated vehicle is moving at speed  $u_{\max}$ .
2.  $t_0 \leq t < t_1 = 20$  minutes: A traffic jam is caused by blocking the road at position  $z_\psi(t_0)$ . The CAV is acting as a moving bottleneck, and its speed is controlled so that the congestion is cleared as soon as possible.
3.  $t \geq t_1, x_\xi(t) \leq z_\psi(t)$ : The blockage is removed and the stop-and-go wave starts propagating upstream. The automated vehicle's speed  $u_\xi$  is controlled so that it helps dissipate and avoids the congestion with minimum delay.

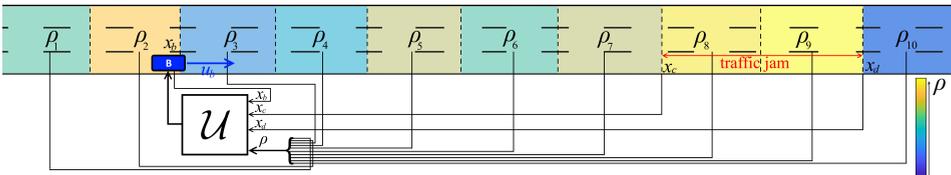
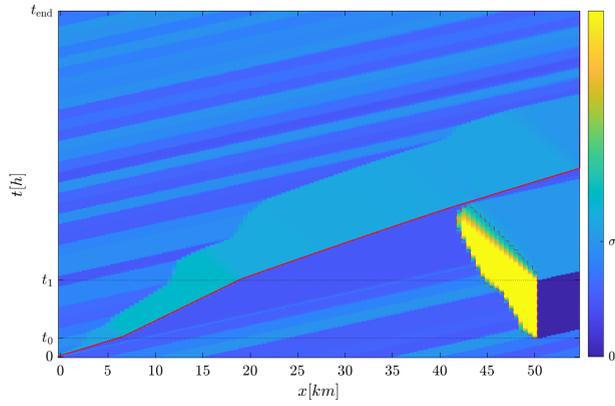
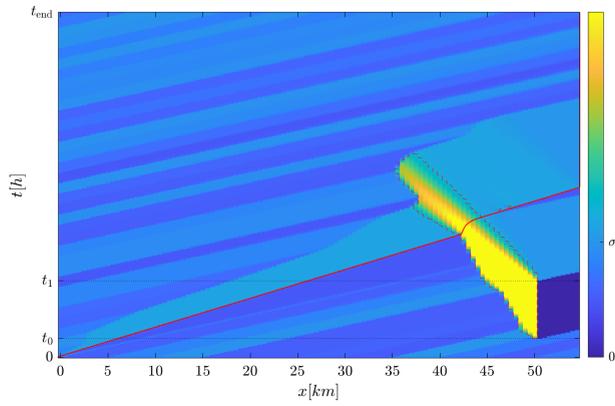


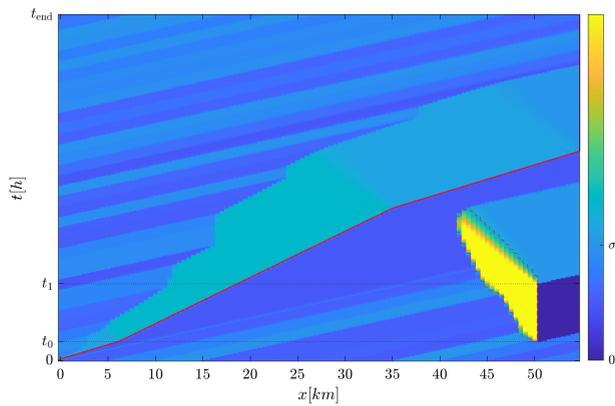
Figure 7.2: Control loop example. Cell traffic densities  $\rho_i$  are color-coded (warmer is higher density).



(a) Controlled



(b) Fast



(c) Slow

Figure 7.3: Traffic densities and moving bottleneck trajectories for the three cases.

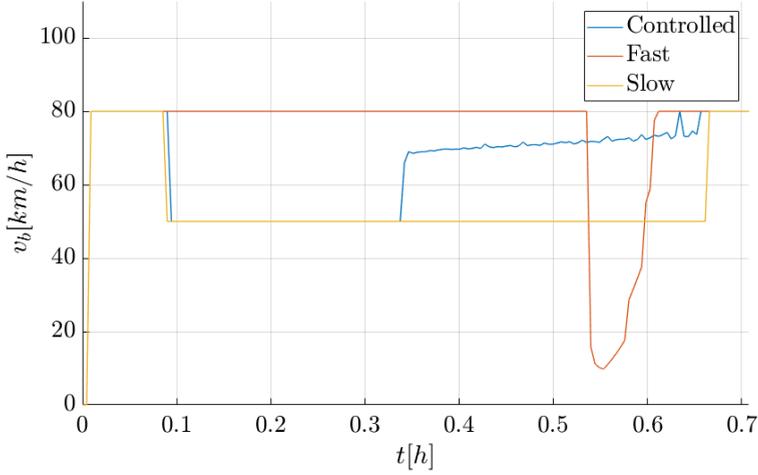


Figure 7.4: Controlled vehicle speeds for the three cases.

4.  $x_\xi(t) > z_\psi(t)$  or the stop-and-go wave has been dissipated: The vehicle continues at speed  $u_\xi^{\max}$ .

The simulation results for  $u_\xi^{\max} = 80$  km/h are shown on Figure 7.3. Warmer colors represent higher traffic density and the congestion, where  $\mathcal{V}(\rho) \leq u_{\max}$ , is outlined in dashed red line. The trajectory of the controlled CAV is represented by the solid red line. The blockage happens at  $t_1 = 5$  minutes and lasts for 15 minutes. During this time, the road is blocked at  $z_\psi(t_0) = 50$  km. The minimum and maximum speeds of the controlled vehicle is taken to be  $u_{\min} = 50$  km/h. We compare three cases:

1. Case 1 (Controlled moving bottleneck): The vehicle is controlled according to the control law (7.3) until there is no more congestion downstream of it.
2. Case 2 (Fast moving bottleneck): The vehicle does not reduce its speed, and continues at  $u_\xi^{\max}$  until it is forced to slow down as it enters congestion.
3. Case 3 (Slow moving bottleneck): The vehicle reduces its speed to  $u_\xi^{\min}$  until there is no more congestion downstream of it.

We can see that by implementing this control strategy, the controlled vehicle avoids the traffic jam with little delay, while also helping dissipate it faster. In the second case, the controlled vehicle does traverse the road segment the fastest out of the three cases, but it does not help clear the traffic jam, and is forced to sharply reduce its speed while inside the congestion, as shown on Figure 7.4. In

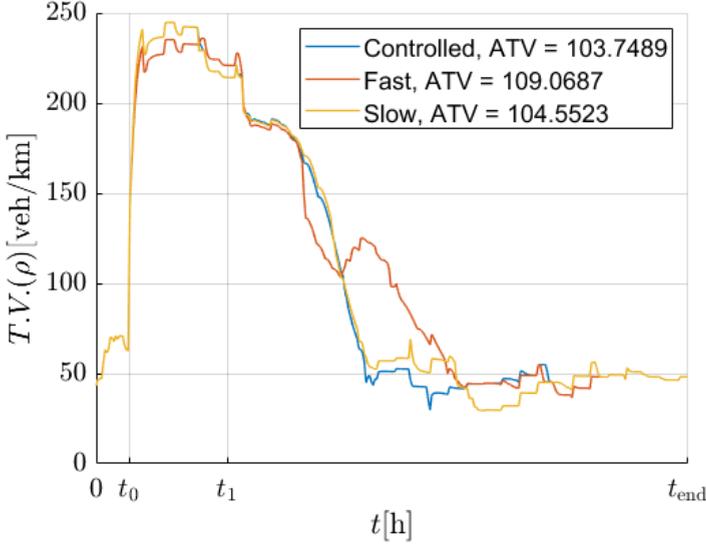


Figure 7.5: Total variation of traffic density for the three cases.

case the vehicle reduces its speed to  $u_{\xi}^{\min}$ , it helps resolve the traffic jam, but it is unnecessarily delayed, and has held up more traffic than necessary.

The achieved Total Time Spent for the three scenarios are  $TTS = 3212.3$  h for case 1,  $TTS = 3207.9$  h for case 2, and  $TTS = 3288.6$  h for case 3. We can see that applying the control law from Section 7.2.1 results in almost no increase in TTS compared to the fast moving bottleneck case, and we see an increase in TTS in the slow moving bottleneck case due to it causing unnecessary additional congestion. Total variations of traffic density for the three cases are shown in Figure 7.5. Although it caused a very slight increase in TTS, we see that the controlled moving bottleneck is able to decrease the ATV of traffic density, thus having a calming effect on the overall traffic without impeding the throughput.

We examined the average influence this control law has on the surrounding traffic, through 100 simulation runs for randomly generated background traffic in the range  $[\sigma/2, \sigma]$  and three different values for  $u_{\xi}^{\max}$ . As performance metric, we considered the increase in TTS compared to the ideal case where the road is always in free flow (results are shown on Figure 7.6 and in Table 7.1) and the ATV results are shown on Figure 7.11 and in Table 7.2.

We can see that employing the described control law leads to an improvement in traffic conditions, in addition to ensuring more desirable conditions for the controlled vehicle, since it avoids going through congestion. The TTS for this case is only very slightly higher than in the case of a fast moving bottleneck, while the ATV is lower. In the slow moving bottleneck case, although we avoid entering

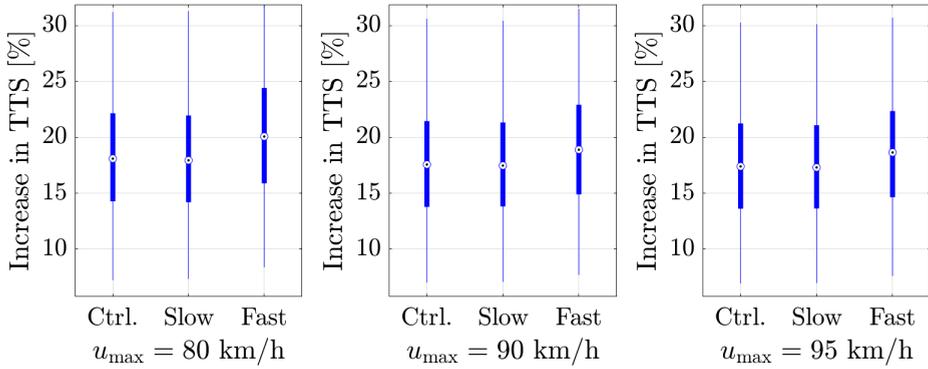


Figure 7.6: TTS increase comparison.

Table 7.1: TTS increase compared to minimum TTS without congestion, when the average travel time is 0.5 h.

$u_{\max}$ [km/h]	Controlled	Fast	Slow
80 km/h	19.33 %	19.19 %	21.09 %
90 km/h	18.73 %	18.63 %	19.98 %
95 km/h	18.50 %	18.40 %	19.56 %

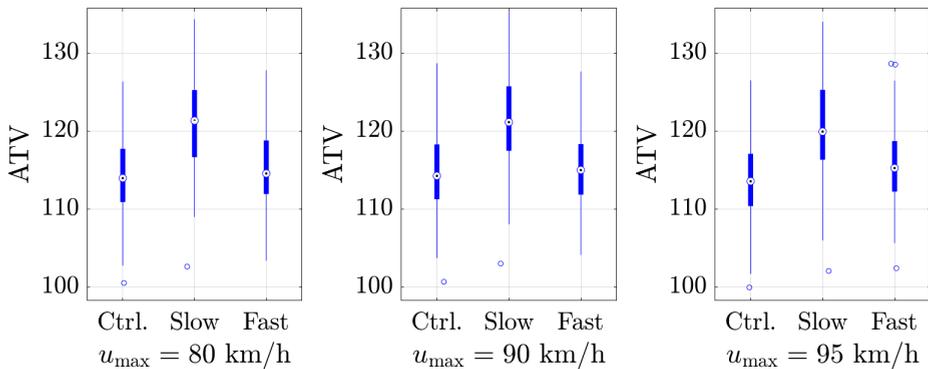


Figure 7.7: Average total variations of traffic density comparison.

Table 7.2: Average total variations of traffic density.

$u_{\max}$	Controlled	Fast	Slow
80 km/h	101.94	108.11	102.67
90 km/h	101.86	108.32	102.61
95 km/h	101.88	108.15	102.59

the traffic jam with the controlled vehicle, the TTS is increased due to a drop in throughput, and the ATV is comparable to than in the controlled case.

### 7.3 Accumulated moving bottlenecks control

It follows from the analysis done in the previous section that in order for the condition (7.1) to be satisfied, and a stop-and-go wave dissipated, we need to have  $q_\psi^{\text{in}}(t) < q_\psi^{\text{dis}}$  for at least a part of the time, which in case we are using moving bottleneck control is equivalent to  $\min\{q_\xi^{\text{in}}(t), q_\xi^{\text{cap}}(t)\} < q_\psi^{\text{dis}}$ . Therefore, in order to be able to guarantee that this condition holds independently of the traffic conditions upstream of the moving bottleneck, we need to ensure  $q_\xi^{\text{cap}}(u_\xi(t)) < q_\psi^{\text{dis}}$ , i.e. we need to be able to restrict the traffic flow using the moving bottleneck more than the capacity drop at stop-and-go wave restricts it. In the previous section, since a two-lane road was considered, a single CAV acting as a bottleneck had a strong effect on the surrounding traffic. However, the more lanes the considered road has, the less of an effect a single vehicle will have on the rest of the traffic. Coupled with the limitations on the minimum reference speed  $u_\xi^{\text{min}}$  a controller may set for a CAV, this severely limits the impact a control law using only a single vehicle can have. Therefore, in this section we explore using multiple CAVs acting together in a formation, in order to apply control action on the rest of the traffic.

#### 7.3.1 Moving bottleneck accumulation

Let there be two classes of vehicles, class  $a$  of CAVs that can be controlled from the infrastructure and class  $b$  of human-driven background traffic that cannot be directly controlled. Therefore,  $U_i^a(t)$  is the control input that we can change in some range  $0 \leq U^{\text{min}} \leq U_i^a(t) \leq U^{\text{max}} \leq V$ , and we set  $U_i^b(t) = V$  except where a different free flow speed is needed to properly model stop-and-go waves. In case the control region is in free flow,  $\rho_i^{\text{K}}(t) < \sigma$ , we will have  $q_i^{\text{K}}(t) = U_i^{\text{K}}(t)\rho_i^{\text{K}}(t)$ , and it is easy to show that  $\rho_i^{\text{b}}$  is not controllable by  $U_i^a$ . If class  $a$  vehicles only represent a very small portion of the traffic, the effect of solely controlling these vehicles in free flow will be very small; instead, we need to also indirectly control the background traffic by creating a controlled congestion. Therefore, the control we propose will consist of three phases:

1. selecting the initial point where we start accumulating controllable vehicles
2. collecting enough controllable vehicles so that they can affect the rest of traffic, and
3. using the collected controllable vehicles as a controlled moving bottleneck.

The overall control structure is outlined in Figure 7.8.

The simplest way of creating an accumulation of vehicles is to enforce a moving free flow speed gradient on the directly controllable vehicle class  $a$ . Let  $\rho^a(x, t)$  be

the CAV traffic density profile,

$$\rho^a(x, t) = \rho_i^a(t^T), \quad X_i \leq x < X_{i+1}, t^T T \leq t < (t^T + 1)T.$$

We can attempt to create a controlled moving bottleneck at position  $x_\xi(t)$ , evolving in time according to

$$\dot{x}_\xi(t) = u_\xi(t),$$

which will then be used to for stop-and-go wave dissipation control. Starting with  $t = t_0$  and setting  $u_\xi(t) = u_\xi = U^{\max} = V$ , if the position-dependent free flow speed of class  $a$  vehicles is set to

$$U^a(x, t) = \begin{cases} U^{\max}, & x \leq x_\xi(t), \\ U^{\min}, & x > x_\xi(t), \end{cases}$$

we can expect them to accumulate close to  $x_\xi(t)$ ,

$$x_\xi(t) = x_\xi(t_0) + (t - t_0)u_\xi.$$

In the ideal situation, the number of accumulated vehicles around  $x_\xi(t)$  is given by

$$n_\xi^a(t) = \int_{x_\xi(t_0)}^{x_\xi^a(t)} \rho^a(x, t_0) dx, \tag{7.4}$$

$$x_\xi^a(t) = x_\xi(t_0) + (t - t_0)(U^{\max} - U^{\min}),$$

i.e. all vehicles that at time  $t_0$  are in  $[x_\xi(t_0), x_\xi^a(t)]$  will have accumulated around  $x_\xi(t)$  by time  $t$ . Denote by  $n_\xi^*(t_0)$  the goal accumulation of controllable vehicles that is required to limit the traffic flow to some value  $q_\xi^{\text{lo}}$ . According to (7.4), this accumulation  $n_\xi^a(t) = n_\xi^*(t_0)$  will have been achieved at time

$$t_\xi^* = t_0 + \frac{\mathcal{R}^{a-1}(\mathcal{R}^a(x_\xi(t_0)) + n_\xi^*(t_0)) - x_\xi(t_0)}{U^{\max} - U^{\min}}, \tag{7.5}$$

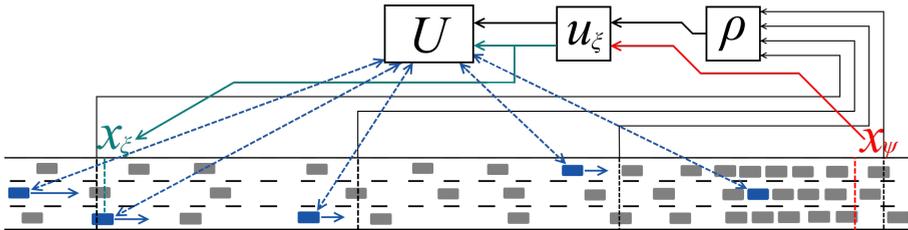


Figure 7.8: Control loop sketch. We use the traffic density data and information about stop-and-go wave boundaries to calculate reference speeds for controlled vehicles. Vehicles upstream of  $x_\xi(t)$  drive faster than those downstream, causing them to eventually accumulate at that point.

where we denote by  $\mathcal{R}^a(x)$  the integral of  $\rho^a(x, t_0)$ ,  $\frac{d\mathcal{R}^a(x)}{dx} = \rho^a(x, t_0)$ .

In the multi-class CTM framework, we create a free flow speed gradient akin to (4.20), setting

$$U_i^a(t) = \begin{cases} \min \left\{ U^{\max, \max} \left\{ U^{\min, \frac{\rho_\xi^*(t) - \frac{V-U_i^a}{V+1} \rho_{i+1}^a(t)}{\rho_i^a(t)}} \right\} \right\}, & i_\xi^-(t) \leq i < i_\xi(t), \\ \min \left\{ U^{\max, \max} \left\{ U^{\min, \frac{\rho_\xi^*(t) \frac{x_\xi(t+1) - X_{i_\xi(t+1)}}{L}}{\rho_i^a(t)}} \right\} \right\}, & i = i_\xi(t), \\ U^{\min}, & i_\xi(t) < i \leq i_\xi^+(t), \end{cases}$$

where  $X_{i_\xi(t^T)} < x_\xi(t^T) \leq X_{i_\xi(t^T)+1}$  will cause class  $a$  vehicles to accumulate in cell  $i_\xi(t)$  until they achieve the reference density  $\rho_\xi^*(t_0)$ . The process of accumulation will be somewhat different due to diffusion. Assuming  $U^{\max} = V$  and  $L = VT$ , we have

$$n_\xi^a(t^T) = \sum_{k=0}^{t^T-t_0^T} A_k^{t^T-t_0^T} \rho_{i_\xi(t_0)}^a(t_0^T),$$

where  $A_k^K$  is given iteratively,

$$A_k^K = \begin{cases} 1, & k = 0, \\ \frac{V-U^{\min}}{V} A_{k-1}^{K-1} + \frac{U^{\min}}{V} A_k^{K-1}, & 0 < k \leq K, \\ 0, & k > K, \end{cases}$$

which is only approximately equivalent to (7.4). In order to be able to restrict the traffic flow to  $q^{\text{lo}} = V(\sigma - \rho_\xi^*(t_0))$  in this framework, we need to accumulate  $n_\xi^*(t_0) = 2L\rho_\xi^*(t_0)$  CAVs, which is achieved at time  $t_\xi^{*T}$ ,  $n_\xi^a(t_\xi^{*T} - 1) < n_\xi^*(t_0) \leq n_\xi^a(t_\xi^{*T})$ . In practice, (7.5) can be used as a good approximation for  $t_\xi^{*T}$ , although it underestimates the time required to achieve the desired accumulation. Such accumulation will then act as a moving bottleneck moving at speed  $u_\xi(t)$ , and limiting the traffic flow to  $q_\xi^{\text{cap}}(t) = q_\xi^{\text{lo}}(t) = V(\sigma - \rho_\xi^*(t_0))$ , as given by (4.24).

### 7.3.2 Control law

As opposed to the control law discussed in Section 7.2.1, where the initial position of the moving bottleneck was given, we are now free to choose any point  $x_\xi(t_0)$  where we begin accumulating vehicles, as a control input. Once a large enough accumulation of controlled vehicles is achieved, the controller proceeds with stop-and-go wave dissipation according to the control law (7.3), with  $\beta\sigma = \rho_\xi^*(t_0)$  as described in Section 7.2.1, where  $\rho_\xi^*(t_0)$  is another decision variable. Therefore, when a stop-and-go wave is detected, we need to select suitable  $x_\xi(t_0)$  and  $\rho_\xi^*(t_0)$ , and then proceed with the next step, with  $u_\xi^*$  given by

$$n_\psi(t_u) - (q_\psi^{\text{dis}} - q_\xi^{\text{lo}}) \left( t_\psi^V \left( t_\xi^* + \frac{z_\psi(t_\xi^*) - x_\xi(t_\xi^*)}{u_\xi^* - \lambda_\psi} \right) - t_\psi^V(t_u) \right) = 0, \quad (7.6)$$

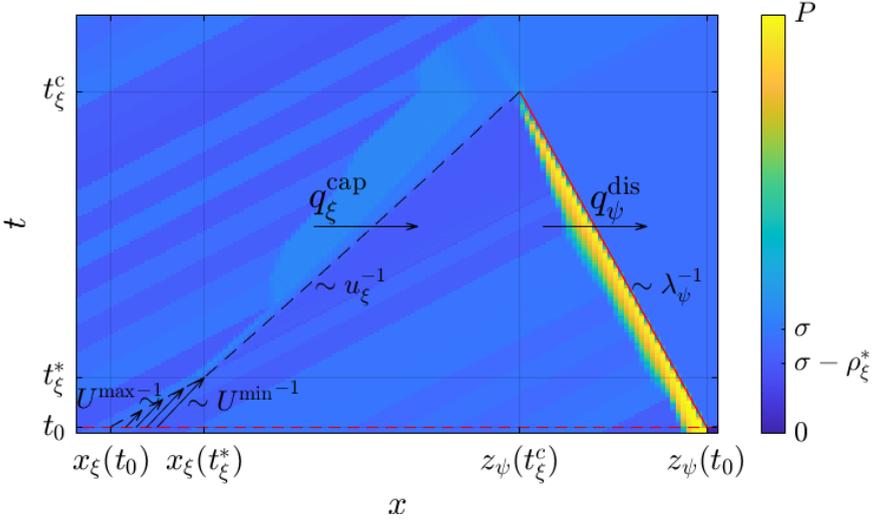


Figure 7.9: Calculating  $x_\xi(t_0)$  given  $u_\xi$ . We calculate  $x_\xi(t_0)$  so that we accumulate enough vehicles after  $t_\xi^* - t_0$ , continue moving the controlled bottleneck at speed  $u_\xi$  and reach the stop-and-go wave at time  $t_\xi^c$ , exactly as it dissipates.

where  $n_\psi(t_u)$  and  $t_\psi^V(t_u)$  depend on  $x_\xi(t_0)$ ,  $t_\xi^*$  and  $x_\xi(t_\xi^*)$  depend on both  $x_\xi(t_0)$  and  $\rho_\xi^*(t_0)$ , and  $q_\xi^{\text{lo}}$  depends only on  $\rho_\xi^*(t_0)$ . The decision variables  $x_\xi(t_0)$  and  $\rho_\xi^*(t_0)$  may be found by solving the congestion reduction problem with cost function (1.1) with additional constraint (7.6). A simpler option is to first choose  $u_\xi^*$  and  $\rho_\xi^*(t_0)$ , and then solve (7.6) to find  $x_\xi(t_0)$ . The full process of calculating the control actions is outlined in Figure 7.9.

A good choice for  $u_\xi^*$  is  $u_\xi^* = U^{\min} - \Delta u$ , with small  $\Delta u > 0$ , leading to quick stop-and-go wave dissipation, while still being robust for delayed accumulation of CAVs due to diffusion or some other disturbances. The choice of  $\rho_\xi^*(t_0)$  will typically be constrained to physically achievable cases, with CAVs occupying some number of lanes,  $\rho_\xi^*(t_0) = \beta\sigma$ , where  $\beta$  is the ratio of the number of occupied lanes to total number of lanes.

Calculating  $x_\xi(t_0)$ ,  $\rho_\xi^*(t_0)$ , and  $u_\xi^*(t)$ , requires us to have a very detailed knowledge about the system state. If some part of that state is not directly measurable, or is uncertain, it would be estimated from the data that is available. In particular, when calculating  $u_\xi^*(t)$  in (7.3), calculating  $\rho_{x_\xi}^{z_\psi}(t^T)$  might be challenging, since it requires information about traffic density in each cell. We may instead use the estimated average traffic density between  $x_\xi(t)$  and  $z_\psi(t)$ ,  $\hat{\rho}_{x_\xi}^{z_\psi}(t^T)$ , which is initialized at some time, and then follows

$$\hat{\rho}_{x_\xi}^{z_\psi}(t^T + 1) = \hat{\rho}_{x_\xi}^{z_\psi}(t^T) - \frac{V - \lambda_\psi q_\psi^{\text{dis}} - (V - u_\xi(t^T))(\sigma - \rho_\xi^*)}{z_\psi(t^T) - x_\xi(t^T)} T. \quad (7.7)$$

Note that since we are not taking in new measurements of traffic density, nor of the stop-and-go wave, this variant essentially corresponds to feedforward control. Since we assume we can communicate with class  $a$  vehicles, and know their positions, we may use exact information about their density  $\rho_i^a(t^T)$ .

### 7.3.3 Simulation results

The parameters of the multi-class CTM model that were used are  $V = 110$  km/h,  $L = 0.5$  km,  $T = L/V$ ,  $\sigma = 60$  veh/km,  $P = 240$  veh/km,  $W = V\sigma/(P - \sigma)$ , and we assume a capacity drop of severity  $\alpha = 0.1$ . The initial traffic density  $\rho_i(0)$  is randomly generated, with every 5 adjacent cells taking a uniformly distributed value from  $[0.8\sigma, \sigma]$ . Similarly, the inflow into the first cell  $q_0(t)$  is randomly generated in the same way, with every 5 samples taking a uniformly distributed value from  $[0.8V\sigma, V\sigma]$ . The ratio of CAVs in the initial traffic density,  $r_i^0$ , and in inflow,  $r^a(t)$ , also takes an uniformly distributed random value from  $[0, 2\bar{r}]$ , where  $\bar{r}$  is the average ratio. For example, the initial density of class  $a$  vehicles in cell  $i$  is  $\rho_i^a(0) = r_i^0 \rho_i(0)$ , while the initial density of class  $b$  is  $\rho_i^b(0) = (1 - r_i^0) \rho_i(0)$ , etc. We are considering 50 km of highway with no on- or off-ramps. At the beginning of each simulation, a stop-and-go wave is induced by fully closing the road for 2 minutes at a point close to the end of the considered stretch, after which we proceed with phase 1 of the described control law. Since the average inflow is equal to the discharge rate of the stop-and-go wave, it is likely that, unless some control action is applied, the wave will remain until the end of the simulation run.

We choose  $\rho_\xi^*(t_0) = \frac{\sigma}{3}$ , i.e. the moving bottleneck will cover one third of the lanes, and use  $u_\xi^* = 60$  km/h,  $U_{\min} = 50$  km/h. Two different versions of the control law described in Subsection 7.3.2 will be compared. In the first one, we use exact information about the current traffic density when calculating (7.3), while in the second one we calculate  $u_\xi^*(t^T)$  approximately, using estimated average density (7.7).

We evaluate the two versions of the proposed control law in 100 simulation runs, with different average ratios of class  $a$  vehicles  $\bar{r}^k$ . The box plots of relative change of TTS and ATV, compared to the base case when no stop-and-go wave is induced,  $\Delta \text{Index}_s^k = \frac{\text{Index}_s^k - \text{Index}_s^0}{\text{Index}_s^0}$ , with  $s$  denoting the number of the simulation run, are shown in Figure 7.10 and Figure 7.11, respectively. The mean and median relative change of these indices are also given in Table 7.3 and Table 7.4.

We can see that applying either version of the control law leads to improvements in both performance indices, even with penetration rates as low as 3%, with higher penetration rates leading to larger improvements, especially in case we use exact traffic density data. Approximate feedforward control is also more likely to fail to dissipate the stop-and-go wave, since it can underestimate the amount of vehicles between the controlled moving bottleneck and the downstream end of the stop-and-go wave, as witnessed by a higher spread or number of outliers in Figures 7.10 and 7.11.

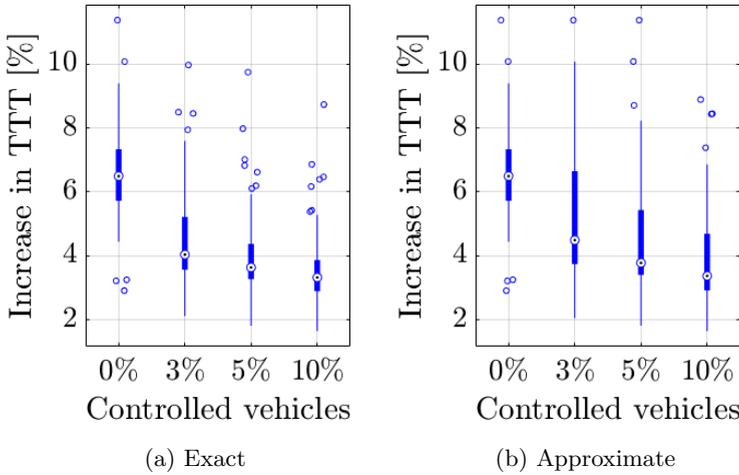


Figure 7.10: Box plots of TTS change compared to the base case.

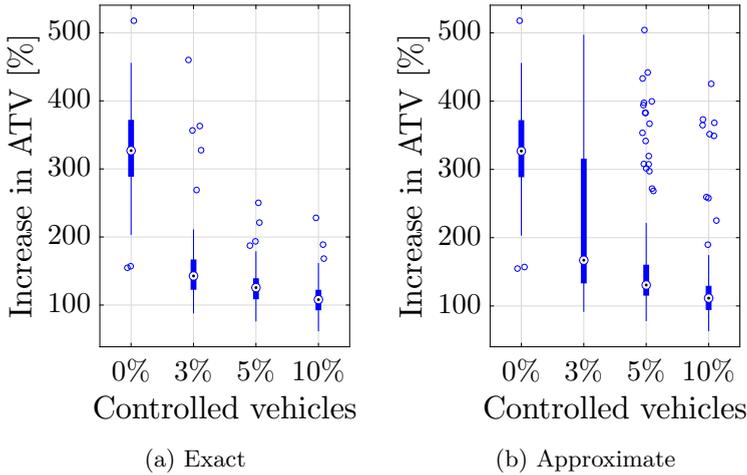
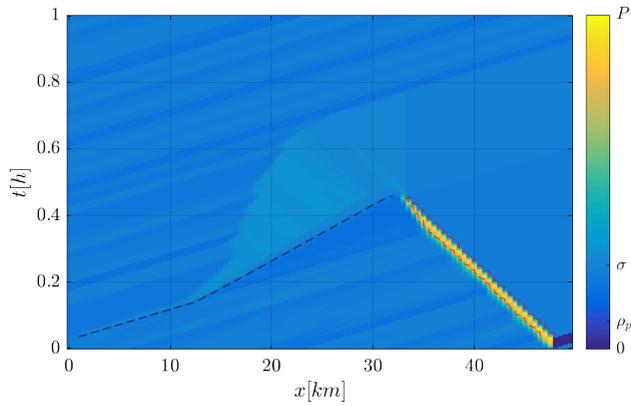
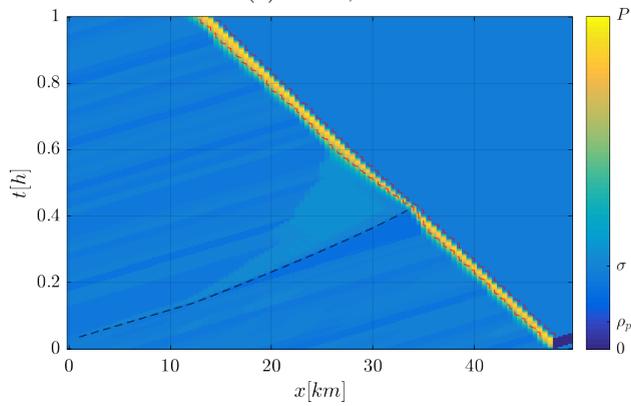


Figure 7.11: Box plots of ATV change compared to the base case.

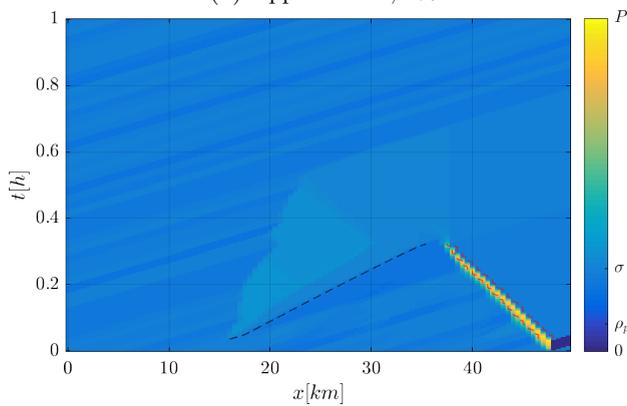
To further illustrate the influence the ratio of controlled vehicles has, as well as highlight the difference in performance of the two control versions, in Figure 7.12 we compare the execution of one simulation run for different penetration rates and control law versions. In Figure 7.12a and 7.12b we compare the exact and the approximate feedforward control laws under penetration rate of  $\bar{r} = 5\%$ . In the latter case, the speed of the controlled moving bottleneck is too high due to underestimating the traffic volume, causing it to arrive at the stop-and-go too wave early and fail to dissipate it.



(a) Exact, 5%



(b) Approximate, 5%



(c) Exact, 10%

Figure 7.12: One example simulation run for different ratios of controlled vehicles and version of the control algorithm.

Table 7.3: Mean and median change in TTT compared to the base case with no stop-and-go wave formed.

$\bar{r}^k$ [%]	exact		approximate	
	mean	median	mean	median
0%	6.6%	6.49%	6.6%	6.49%
3%	4.51%	4.04%	5.22%	4.49%
5%	3.99%	3.64%	4.49%	3.78%
10%	3.58%	3.32%	3.82%	3.37%

Table 7.4: Mean and median change in ATV compared to the base case with no stop-and-go wave formed.

$\bar{r}^k$ [%]	exact		approximate	
	mean	median	mean	median
0%	331.13%	326.87%	331.13%	326.87%
3%	153.46%	142.76%	216.99%	167.01%
5%	127.73%	125.63%	169.84%	130.67%
10%	109.92%	107.99%	129.88%	111.45%

Additionally, by comparing Figure 7.12a and Figure 7.12c, we see the benefit of having a higher penetration rate of connected automated vehicles. In case  $\bar{r} = 10\%$ , we are both able to start collecting the vehicles closer to the stop-and-go wave, and finish collecting enough of them much quicker than in case  $\bar{r} = 5\%$ . This allows us to dissipate the stop-and-go wave quicker, leading to improvement in performance.

## 7.4 Summary

In this chapter we considered the simplest case of the stop-and-go wave dissipation problem, dissipating a single stop-and-go wave using a single controlled moving bottleneck. The proposed control laws were tested in simulation and shown to achieve good results, successfully dissipating the stop-and-go wave. In case there was no capacity drop, dissipating the stop-and-go wave did not improve the TTS, as expected, but it did improve the ATV, making the traffic density profile smoother and more homogeneous. In case the discharging flow of the stop-and-go wave was lower than the capacity of the road, dissipating it reduced the TTS. By accumulating enough controlled vehicles, we are able to implement stop-and-go dissipating control on roads with an arbitrary number of lanes, provided that there is enough time and space to execute the control actions.

The simplest case of stop-and-go wave dissipation presented here is expanded upon in the following chapter, and generalized to the case with multiple stop-and-go waves, using multiple controlled moving bottlenecks as actuators.



## Chapter 8

# Reconstruction-based multiple stop-and-go wave dissipation

USING connected vehicles as actuators for traffic control has been shown to achieve good results in the previous chapters. In this chapter, we consider using these vehicles to also fill other roles in the control loop. So far, the full traffic state was assumed to be known, and this information was used for control action calculation. If we assume multiple connected vehicles are present on the road, we may use their trajectories and potential local traffic state measurements to estimate and reconstruct the traffic state, as well as to learn the traffic model governing the behaviour of the traffic flow. This traffic state reconstruction can then be used in place of actual traffic state in case the full state information is not available.

Apart from considering reconstruction-based control instead of full-information control, in Section 8.1 we also continue the work presented in the previous chapter by generalizing the stop-and-go wave problem to the case when we dissipate multiple randomly arising stop-and-go waves using multiple randomly arriving connected vehicles used as controlled moving bottlenecks. Stop-and-go waves originating from congestion in the wake of moving bottlenecks are also considered. First, in Section 8.2, a simple traffic state reconstruction scheme is considered, assuming the underlying traffic model is known and presenting heuristic rules for selecting a subset of available CAVs whose local measurements will be used. Then, in Section 8.3, we drop the assumption that the traffic model is known, and use the FTSM to control a more complex traffic model, simultaneously reconstructing the traffic state, learning the traffic model, and calculating the control actions, all based on the CAVs acting as both sensors and actuators.

### 8.1 Multiple stop-and-go wave dissipation

The stop-and-go wave dissipation problem described in Section 7.1 can easily be generalized to the case when there are multiple stop-and-go waves  $\psi$ , and multiple

controlled moving bottlenecks  $\xi$ . At some time  $t_0$ , let the positions of the stop-and-go waves and moving bottlenecks be  $z_\psi(t_0)$  and  $x_\xi(t_0)$ , where  $\psi \in \Psi(t_0)$ ,  $\xi \in \Xi(t_0)$ ,  $\Psi(t_0)$  is the set of all stop-and-go waves present on the considered road segment at time  $t_0$ , and  $\Xi(t_0)$  is the set of all moving bottlenecks present at time  $t_0$ . We denote by  $\vec{\psi}$  and  $\overleftarrow{\psi}$  the first stop-and-go wave downstream and upstream of stop-and-go wave  $\psi$ , respectively,

$$\begin{aligned}\vec{\psi} &= \arg \min_{k \in \mathbf{K}} z_k(t_0), \quad \mathbf{K} = \{m \in \Psi(t_0) | z_m(t_0) > z_\psi(t_0)\}, \\ \overleftarrow{\psi} &= \arg \max_{k \in \mathbf{K}} z_k(t_0), \quad \mathbf{K} = \{m \in \Psi(t_0) | z_m(t_0) < z_\psi(t_0)\}.\end{aligned}$$

Similarly,  $\vec{\xi}$  and  $\overleftarrow{\xi}$  denote the first moving bottleneck downstream and upstream of moving bottleneck  $\xi$ , respectively,

$$\begin{aligned}\vec{\xi} &= \arg \min_{k \in \mathbf{K}} x_k(t_0), \quad \mathbf{K} = \{m \in \Xi(t_0) | x_m(t_0) > x_\xi(t_0)\}, \\ \overleftarrow{\xi} &= \arg \max_{k \in \mathbf{K}} x_k(t_0), \quad \mathbf{K} = \{m \in \Xi(t_0) | x_m(t_0) < x_\xi(t_0)\}.\end{aligned}$$

We may also order the indices of stop-and-go waves and moving bottlenecks so that  $z_\psi(t_0) > z_{\psi+1}(t_0)$ ,  $\psi \in \Psi(t_0)$ , and  $x_\xi(t_0) > x_{\xi+1}(t_0)$ ,  $\xi \in \Xi$ . The multiple stop-and-go wave dissipation problem can be decomposed into  $|\Xi|$  single stop-and-go wave problems, where each of the controlled moving bottlenecks  $\xi$  tries to dissipate the most severe stop-and-go wave downstream of it, also taking into account the new stop-and-go waves that may arise in the wake of the moving bottleneck  $\vec{\xi}$ .

Let  $\Psi_\xi(t_0)$  be the set of stop-and-go waves that are downstream of the moving bottleneck  $\xi$  at time  $t_0$ ,

$$\Psi_\xi(t_0) = \{\psi \in \Psi | z_\psi(t_0) > x_\xi(t_0)\}.$$

Firstly, we denote the downstream-most controlled moving bottleneck  $\xi = 1$ . For this moving bottleneck, at time  $t = t_0$ , we may calculate the optimal reference speeds for dissipating each stop-and-go wave  $\psi \in \Psi_1(t_0)$  according to (7.2). We denote these reference speeds  $u_1^{\psi*}(t_0)$ , calculated for stop-and-go wave  $\psi$ , assuming all other stop-and-go waves were not impeding the traffic flow. The overall reference speed is then given as the minimum of all reference speeds for single stop-and-go waves,

$$u_1^*(t_0) = \min_{\psi \in \Psi_1(t_0)} u_1^{\psi*}(t_0).$$

If  $u_\xi^*(t_0) > u_\xi^{\min}$ , the moving bottleneck  $\xi$  will be able to fully dissipate all stop-and-go waves downstream of it. In this case, the most severe stop-and-go wave  $\psi_\xi$  is the one for which the minimum reference speed was calculated,

$$\psi_\xi = \arg \min_{\psi \in \Psi_\xi(t_0)} u_\xi^{\psi*}(t_0).$$

Otherwise, if for any  $\psi \in \Psi_\xi(t_0)$  we get  $u_\xi^{\psi^*}(t_0) < u_\xi^{\min}$ , the moving bottleneck  $\xi$  will not be able to fully dissipate at least one stop-and-go wave. We denote the set of these remaining stop-and-go waves  $\Psi_\xi^f(t_0)$ ,

$$\Psi_\xi^f(t_0) = \left\{ \psi \in \Psi_\xi(t_0) \mid u_\xi^{\psi^*}(t_0) < u_\xi^{\min} \right\}.$$

Each subsequent controlled moving bottleneck  $\xi$  tries to dissipate the stop-and-go waves that originate from the area between itself and moving bottleneck  $\vec{\xi}$ ,  $\Psi_\xi(t_0) \setminus \Psi_{\vec{\xi}}(t_0)$ , and the stop-and-go waves left from moving bottleneck  $\vec{\xi}$ . If the previous moving bottleneck fails to dissipate some stop-and-go waves,  $\Psi_{\vec{\xi}}^f(t_0) \neq \emptyset$ , moving bottleneck  $\xi$  attempts to dissipate them as well. However, if moving bottleneck  $\vec{\xi}$  successfully dissipated all stop-and-go waves from  $\Psi_{\vec{\xi}}(t_0)$ , the congestion left in its wake can form a new, less severe stop-and-go wave, that moving bottleneck  $\xi$  will have to dissipate. We denote this newly created stop-and-go wave  $\psi'_{\vec{\xi}}$ , it is created at time  $t_{\psi'_{\vec{\xi}}}^c$  at position  $z_{\psi'_{\vec{\xi}}}^c(t_{\psi'_{\vec{\xi}}}^c) = x_{\vec{\xi}}(t_{\psi'_{\vec{\xi}}}^c)$ , and we have  $q_{\psi'_{\vec{\xi}}}^{\text{dis}} > q_{\psi_{\vec{\xi}}}^{\text{dis}}$ .

Finally, the reference speed of controlled moving bottleneck  $\xi$  is given again by

$$u_\xi^*(t_0) = \min_{\psi \in \Psi_\xi(t_0)} u_\xi^{\psi^*}(t_0). \quad (8.1)$$

where  $u_\xi^{\psi^*}(t_0)$  is calculated differently depending on  $\psi$ :

- For the stop-and-go waves that moving bottleneck  $\vec{\xi}$  was not able to fully dissipate,  $\psi \in \Psi_\xi(t_0) \setminus \Psi_{\vec{\xi}}(t_0) \cup \Psi_{\vec{\xi}}^f(t_0)$ ,  $u_\xi^{\psi^*}(t_0)$  is given by (7.2), with

$$t_u = t_\psi^{-V}(t_\xi^V(t_0)).$$

- For the stop-and-go wave left in the wake of moving bottleneck  $\vec{\xi}$ ,  $\psi = \psi'_{\vec{\xi}}$ , if  $\Psi_{\vec{\xi}}^f(t_0) = \emptyset$ , we calculate  $u_\xi^{\psi^*}(t_0)$  according to

$$n_{\psi'}(t'_u) - \left( q_{\psi'}^{\text{dis}} - q_\xi^{\text{lo}}(u_\xi^{\psi^*}(t_0)) \right) \left( t_\psi^V \left( t_0 + \frac{z_\psi(t_0) - x_\xi(t_0)}{u_\xi^{\psi^*}(t_0) - \lambda_\psi} \right) - t_\psi^V(t'_u) \right) = 0,$$

where  $t'_u = \min\{t_\psi^{-V}(t_\xi^V(t_0)), t_\psi^c\}$ , and

$$n_{\psi'}(t'_u) = \int_{x_\xi(t_0)}^{x_{\vec{\xi}}(t_0)} \rho(x, t_0) dx + \max\{t_\psi^d - t_\xi^V(t_0), 0\} \left( q_\xi^{\text{lo}}(u_\xi^{\psi^*}(t_0)) - q_{\vec{\xi}}^{\text{lo}}(u_{\vec{\xi}}^*(t_0)) \right).$$

- For all other  $\psi$ , which are dissipated by moving bottleneck  $\vec{\xi}$  or dissipate on their own, we set  $u_\xi^{\psi^*}(t_0) = u_\xi^{\max}$ .

As described in Section 4.3.5, the discharging flow of a stop-and-go wave  $\psi$  depends on the maximum traffic density  $\rho_\psi^c(t_0)$  in it,

$$q_\psi^{\text{dis}} = Q(\rho_\psi^d(t_0)),$$

where  $\rho_\psi^d(t_0)$  is given as the minimum solution of

$$Q(\rho_\psi^d(t_0)) = Q(\rho_\psi^c(t_0)) + \lambda_\psi(\rho_\psi^d(t_0) - \rho_\psi^c(t_0)), \quad \rho \geq 0$$

and  $Q$  is the flux function of the overall traffic. The overtaking flow of moving bottleneck  $\xi$  dissipating stop-and-go wave  $\psi$  is given by

$$q_\xi^{\text{lo}}(u_\xi^{\psi*}(t_0)) = Q(\rho_\xi^d(t_0)),$$

where  $\rho_\xi^d(t)$  is the minimum  $\rho > 0$  for which

$$Q(\rho) = Q_\xi(\rho^\tau) + u_\xi^{\psi*}(t)(\rho - \rho^\tau) \tag{8.2}$$

holds, and  $\rho^\tau$  is the traffic density at which a line with slope  $u_\xi^{\psi*}(t)$  is tangent to the moving bottleneck flux function  $Q_\xi(\rho)$ ,

$$\rho^\tau = \arg \max_{\rho} Q_\xi(\rho) - u_\xi^{\psi*}(t)\rho.$$

Conversely, we denote the maximum solution of (8.2) as  $\rho_\xi^c(t)$ , which is the traffic density of the congestion that accumulates in the wake of the moving bottleneck, assuming the incoming traffic flow is higher than  $q_\xi^{\text{lo}}(u_\xi^{\psi*}(t_0))$ . After the stop-and-go wave  $\psi$  is dissipated by moving bottleneck  $\xi$ , the congestion in its wake creates a new stop-and-go wave with

$$q_{\psi'}^{\text{dis}} = Q(\rho_{\psi'}^d(t_0)),$$

and  $\rho_{\psi'}^d(t_0)$  calculated based on

$$\rho_{\psi'}^c(t_0) = \rho_\xi^c(t_0).$$

Therefore, it is easy to show that  $q_{\psi'}^{\text{dis}} > q_\psi^{\text{dis}}$ , and even though dissipating a stop-and-go wave using a moving bottleneck causes another stop-and-go wave to be created, the new stop-and-go wave is less severe than the original one and the overall traffic throughput will be increased.

## 8.2 Basic traffic state reconstruction and probe vehicle selection

Since we are now using multiple CAVs to dissipate multiple stop-and-go waves, possibility arises to rely solely on the CAVs to act both as the only sensors and as the only actuators on the road, with measurements communicated among them in order to estimate the traffic state and calculate the control action. We may differentiate between three types of CAVs:

1. Inactive CAVs – not currently acting as sensors or actuators,
2. Probe CAVs – acting only as sensors, and
3. Actuator CAVs – acting as both sensors and actuators,

based on the level of capabilities and compliance of the vehicles. Here, actuator CAVs represent the highest level of CAV automation and compliance, that are able to measure the local traffic conditions, communicate them with the infrastructure, and follow the commands that is sent to them from the infrastructure. Probe CAVs are either unable or unwilling to receive commands from the infrastructure, but they share their local traffic measurements, and inactive CAVs are those that are able to act as probe vehicles, but are either currently refusing to share their measurements with the infrastructure, or have been (potentially temporarily) commanded not to send their measurements e.g. to reduce the communication burden. Furthermore, collecting local measurements from probe vehicles might incur some costs, so deactivating those probe vehicles whose measurements are not needed might be beneficial.

### 8.2.1 Traffic state reconstruction in CTM

Consider a stretch of road divided into  $N$  cells of length  $L$ , with cell  $i$  stretching from  $X_{i-1} = (i - 1)L$  to  $X_i = iL$ . We assume that the dynamics of traffic are given by the CTM (3.16), as defined in Section 3.3, with time step  $T = L/V$ . Let  $\Xi^s(t^T)$  be the set of all CAVs acting as sensors at discrete time  $t^T$ ,  $t = t^T T$ , whose local traffic measurements we can access. The position of probe CAV  $\xi \in \Xi^s(t^T)$  is denoted  $x_\xi(t^T)$ , the cell it is currently in is  $i_\xi(t^T)$ ,  $X_{i_\xi(t^T)-1} < x_\xi(t^T) \leq X_{i_\xi(t^T)}$ . We assume that probe vehicles can measure local traffic densities in cells  $i \in \mathcal{I}^s(t^T)$  adjacent to the cell they are in,

$$\mathcal{I}^s(t^T) = \{i \in \{1, \dots, N\} \mid |i - i_\xi(t^T)| \leq 1, \xi \in \Xi^s(t^T)\},$$

assuming that the cell length  $L$  is chosen so that the sensors on CAVs can indeed cover this range. This set will typically change every time step, since the CAVs move along the road, leave the road segment at its downstream end, and new ones will arrive at its upstream end. The position of the CAVs is updated as

$$x_\xi(t^T + 1) = x_\xi(t^T) + T v_\xi(t^T),$$

where the speed of the CAV  $v_\xi(t^T)$  depends on its reference speed and the speed of the traffic immediately downstream of it,

$$v_\xi(t^T) = \min \left\{ u_\xi(t^T), \frac{\check{q}_{i_\xi(t^T)}(t^T)}{\check{\rho}_{i_\xi(t^T)}(t^T)} \right\}.$$

Using these measurements, we can attempt to approximately reconstruct the traffic density,

$$\rho_i(t^T) = \begin{cases} \check{\rho}_i(t^T), & i \in \mathcal{I}^s(t^T), \\ \rho_i(t^T - 1) + \frac{T}{L} (q_{i-1}(t^T - 1) - q_i(t^T - 1)), & i \notin \mathcal{I}^s(t^T). \end{cases}$$

Here we denote by  $\check{\rho}$  and  $\check{q}$  the true traffic density and flow, and by  $\rho$  and  $q$  the traffic density and flow estimates, reconstructed using the available traffic data. The traffic flows  $q_i(t^T)$  are defined using the reconstructed traffic density  $\rho_i(t^T)$  instead of the real traffic density  $\check{\rho}_i(t^T)$ .

Since we assume that the traffic flow model is known, there are only three sources of uncertainty in the traffic density estimate: the initial condition  $\check{\rho}_i(0)$ , the inflow  $\check{q}_0(t^T)$ , and the conditions at the downstream end of the road segment, which in this case will consist of stop-and-go waves arriving from downstream. We assume that at least the average inflow  $\bar{q}_0$  is known, which in practice could be learned from historical data, or could be acquired from some stationary sensor upstream of the road segment under consideration. Then, we may use this value as the estimated inflow,  $q_0(t^T) = \bar{q}_0$ , if no other information is available, as well as for initializing the traffic density estimates,  $\rho_i(0) = \bar{q}_0/V$ . However, unless they can be measured in some other way, changes in the traffic conditions downstream of the road segment will only be detected once a probe vehicle reaches their spillback.

Note that in the proposed simple reconstruction algorithm it is required that all probe vehicles communicate their measurements at each time step, potentially straining the communication resources. However, since many of the vehicles will be in free flow, moving at the same speed as the vehicles around them, the measurements that they would communicate are often redundant. Therefore, it can be beneficial to develop an algorithm that will only activate the potential probe vehicles when their measurements are needed.

### 8.2.2 Probe vehicle selection

It is clear that the quality of traffic state reconstruction can only increase if we gain access to more information, i.e. use more probe vehicles. However, in a situation where the communication channel bandwidth is limited, it might be useful to reconstruct the density with fewer sensors, eliminating the redundant information. Information about the congestion and stop-and-go waves is particularly important and will significantly improve the control performance, whereas the information about the rest of the road that is in free flow is less impactful.

There are numerous ways of selecting which CAVs are used as probe vehicles, and the selection will depend on the intended purpose. Here, we propose a simple adaptive probe vehicle selection scheme. Denote by  $\Xi$  the set of all CAVs. First, we use a subset  $\Xi_0^s$  of CAVs that are always activated. The set of actuator vehicles is denoted  $\Xi^a$ ,  $\Xi^a \subset \Xi_0^s$ , and  $\Xi^s(t^T)$  is the set of probe vehicles activated at time  $t$ . When the vehicles in  $\Xi_0^s$  detect any congestion, we activate all vehicles that are less

Case	Sensors	Actuators
No control	$\emptyset$	$\emptyset$
Predefined subset of CAVs	$\Xi_0^s \subset \Xi$	$\Xi^a \subset \Xi_0^s$
Adaptive subset of CAVs	$\Xi^s(t^T) \subset \Xi$	$\Xi^a \subset \Xi_0^s$
All CAVs	$\Xi$	$\Xi^a \subset \Xi$
Full-information	all the road	$\Xi^a \subset \Xi$

Table 8.1: Summary of different state reconstruction and control scenarios using CAVs.

than some predefined distance  $\ell$  away from any cell where  $\rho_i(t^T) > \sigma$ , i.e. where we expect to encounter congestion, leading to the following definition:

$$\Xi^s(t^T) = \Xi_0^s \cup \left\{ \xi \in \Xi \mid \rho_i(t^T) > \sigma, i - i_\xi(t^T) \in \left[ 0, \left\lfloor \frac{\ell}{L} \right\rfloor \right] \right\}.$$

After these vehicles leave the congestion, they will be deactivated. With this selection scheme, we aim to improve the quality of traffic state reconstruction that is most impactful with regards to the control performance.

In the following subsection, we will compare this probe vehicle selection scheme to using only the probe vehicles from  $\Xi_0^s$ , as well as with using all CAVs as probe vehicles.

### 8.2.3 Simulation results

The efficiency of the proposed traffic density reconstruction scheme and control law is studied in simulations with different parameters. The initial traffic densities, flow into the road segment and stop-and-go waves entering the road from downstream are randomized and the same values are taken for all parameter combinations. The arrival of CAVs is modelled as Poisson arrival process with average gap of  $G$ , and the newly arrived vehicle is a probe vehicle with probability  $p_s$ , an actuator vehicle with probability  $p_a$  (in which case it also acts as a probe vehicle), and an inactive CAV if it is neither a probe nor an actuator vehicle.

We executed 100 simulations for every different combination of parameter  $G$  and  $p_s$  values, with  $G \in \{0.5\text{km}, 1\text{km}, 1.5\text{km}, 2.5\text{km}\}$ ,  $p_s \in \{0.1, 0.3, 0.5, 0.7\}$ , and  $p_a = 0.3$ . For each simulation run, five control cases were used, as summarized in

$\bar{q}_0 = 3200 \text{ veh/h}$	$\sigma = 40 \text{ veh/km}$	$\alpha = 0.25$
$V = 100 \text{ km/h}$	$W = 50 \text{ km/h}$	$u^{\min} = 30 \text{ km/h}$

Table 8.2: Simulation parameters.

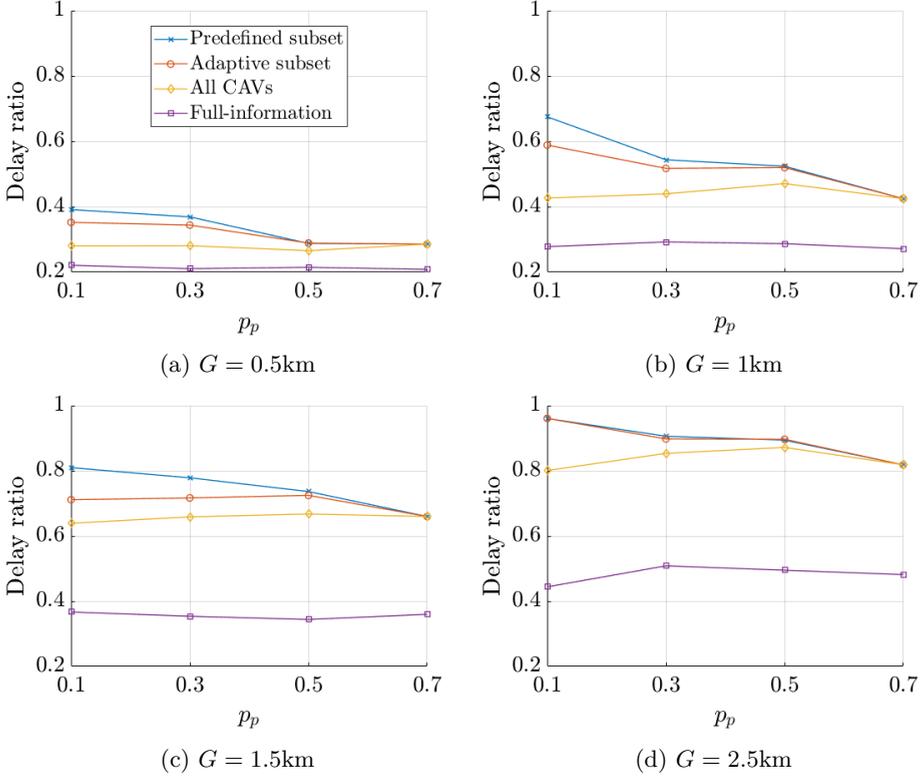


Figure 8.1: Median delay compared to the no control case shown for different  $p_s$  and  $G$  for a given  $p_a = 0.3$ .

Table 8.1. The performance metric that was used is the median of the delay ratio,

$$\frac{TTS - TTS^{\min}}{TTS^{\text{unc}} - TTS^{\min}},$$

i.e. the ratio between the increases of  $TTS$  in the controlled and uncontrolled case. The increase is calculated compared to the minimum  $TTS^{\min} = \frac{\bar{q}_0}{V} NL t_{\text{sim}}$ , where  $NL = 5$  km is the length of the simulated road segment and  $t_{\text{sim}} = 1$  h the duration of each simulation run.

The simulation results are shown in Figure 8.1. We can see how increasing  $p_s$  affects the delay ratio for constant  $G$  and  $p_a$ . Unsurprisingly, we can see that control performance deteriorates as we use less and less information. When  $G = 0.5$  km, the full-information control achieves the best performance, eliminating close to 80% of delay, whereas using all CAVs as sensors eliminates around 72% of delay for the same  $G$ . In case we are using a subset of CAVs as sensors, the performance

will improve as  $p_s$  increases, starting from eliminating around 60% and 65% of the delay, using a predefined and adaptive subset of CAVs as sensors respectively, and approaching the performance of the case where we use all CAVs as sensors as  $p_s + p_a$  goes to 1, when the same subsets of CAVs are used. We can also see that the main factor determining the control performance is the average gap between two CAVs. When  $G$  is very low, all control schemes achieve good results, and probe vehicle-based control approaches the full-information control.

To further illustrate the control and state reconstruction algorithms, in Figure 8.2 we show a detail from one of the simulation runs. The figures show the density profile along the road, with brighter colours representing denser traffic.

The baseline case, with no control, is shown in Figure 8.2a, and the full-information control case is shown in Figure 8.2b. A stop-and-go wave originating from downstream enters the road around  $t = 0.52$  h, and propagates upstream unless dissipated by applying some control action.

Figures 8.2c and 8.2e show the attempt to dissipate the congestion using only the predefined set of CAVs (shown in green and red) as sensors, with the real traffic situation  $\check{\rho}_i(t^T)$  shown in Figure 8.2c, and the reconstructed estimation of the traffic state  $\rho_i(t^T)$  shown in Figure 8.2e. Around  $t = 0.536$  h, an actuator vehicle runs into the stop-and-go wave, detecting it as it goes through it. The actuator vehicle upstream reacts by slowing down and restricting the flow. Four inactive CAVs reach the stop-and-go wave before the actuator vehicle, but since they transmit no information, the control law underestimates the width of the wave and the CAV fails to completely dissipate it. However, in case we use the proposed adaptive probe vehicle activation, once these four inactive CAVs get close to the congestion, they are temporarily activated, as shown in Figures 8.2d and 8.2f (shown in dashed green). The additional information corrects the underestimation, and the stop-and-go wave is successfully dissipated.

### 8.3 Reconstruction-based control with model learning

In the previous section, the use of local traffic measurements provided by probe vehicles to reconstruct the traffic state and calculate the control action was demonstrated, assuming that only the average traffic density was known. However, we also assumed that the traffic model was known exactly, and used the same model for both simulation, and state reconstruction and control design. In this section, we explore an even more general problem, dropping the assumption that the traffic model is known. We use the FTSM framework for traffic state reconstruction, model learning, and finally control design and implementation, whereas the simulation model on which the proposed algorithms are tested is of different form and, moreover, noisy. While the previous section focused more on the process of selecting which CAVs would be used as probe vehicles, here we assume that all available CAVs can act as actuator vehicles, and focus on the remainder of the control loop. The overall structure of the closed-loop system is shown in Figure 8.3.

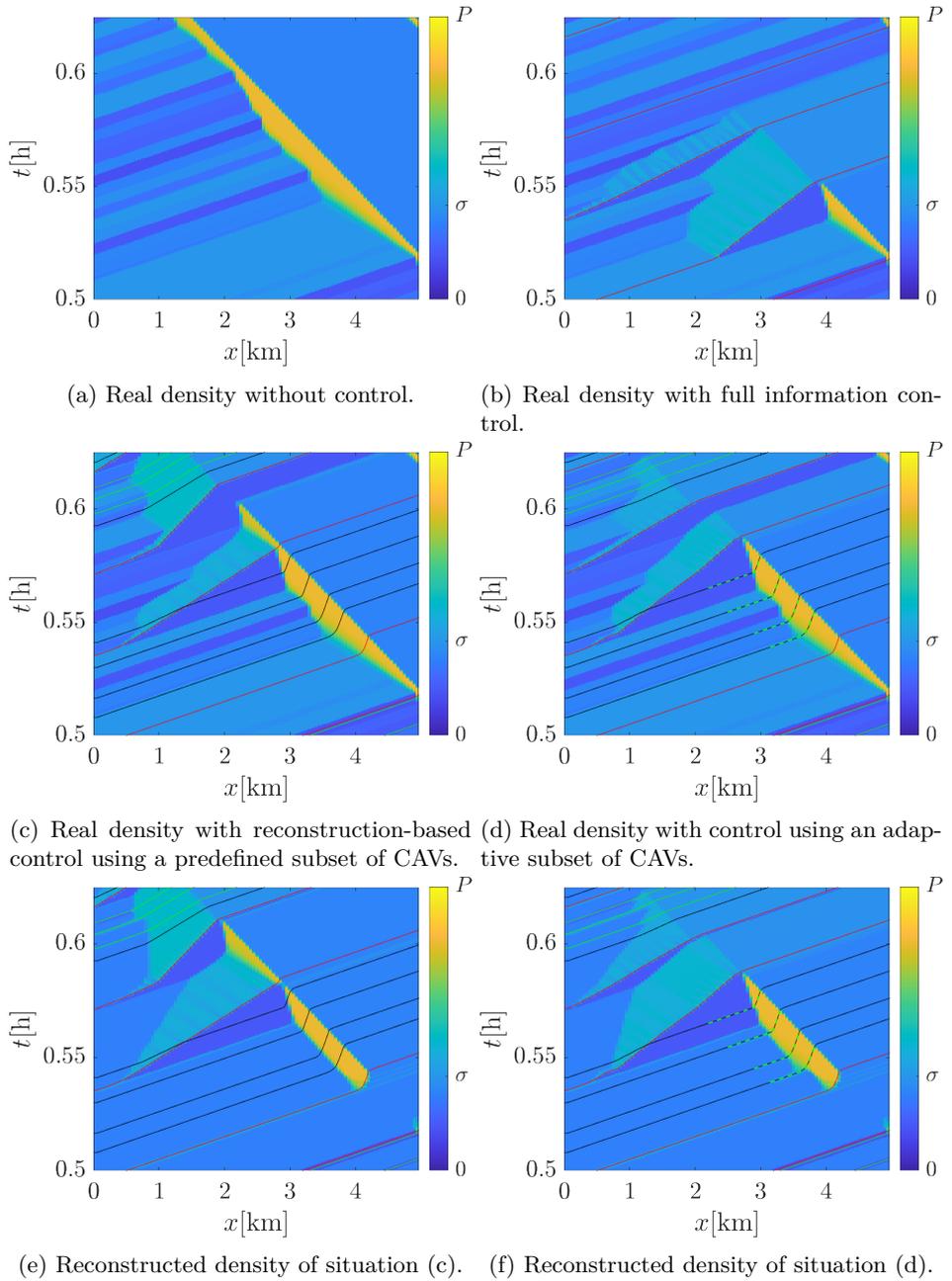


Figure 8.2: Traffic density obtained from a numerical simulation in different cases. The trajectories of inactive CAVs are in black, probe vehicles in green and actuator vehicles in red.

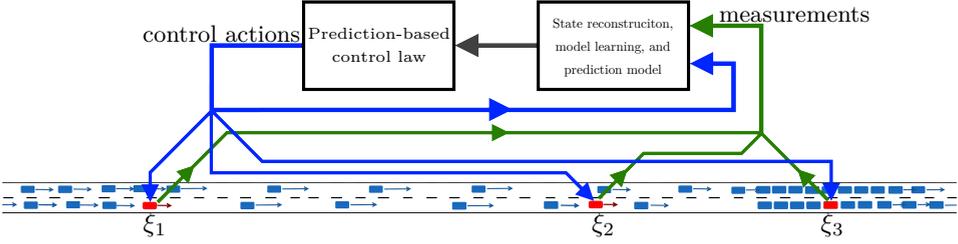


Figure 8.3: Closed-loop system structure. Local traffic measurements are collected by the CAVs and sent to the infrastructure, where traffic state reconstruction and model learning is implemented. Based on the current state estimate, control actions for the CAVs are calculated using the learned predictions model.

### 8.3.1 Traffic state reconstruction and model learning

As described in Section 4.5, we assume that CAVs can measure the local traffic density  $\check{\rho}_\xi$  and traffic speed  $\check{v}_\xi$ , from which we can get the local traffic flow  $\check{q}_\xi = \check{\rho}_\xi \check{v}_\xi$ . In case the CAV  $\xi$  is moving slower than the surrounding traffic, we also assume that it can measure the overtaking flow  $\check{\omega}_\xi$ . These measurements acquired from CAVs will be used to reconstruct the traffic state, as well as to identify the flux functions and other parameters that describe the traffic system.

We first describe how local traffic density measurements can be used to attempt to reconstruct the broader traffic state, which will later be used to calculate control actions. While reconstructing the traffic state, we assume that all flux functions and other parameters of the traffic model are known. We then propose algorithms that can be used to identify each of these aspects of the traffic model. The proposed algorithms exploit the continuous piecewise linear structure of the flux functions to simplify calculations and avoid making assumptions about their shape. Indeed the only assumption made is that the dynamics of the traffic can, at least approximately, be described as a first order conservation law, ignoring the dynamics of the traffic speed.

#### State reconstruction

Firstly, we use the local measurements of traffic density  $\check{\rho}_\xi$  to attempt to reconstruct the traffic state, and other measurements are used only in model learning. This information can be included into the FTSM by forcing the traffic density to be equal to the measured one at the position of the CAV  $\xi$ , by setting  $F_\xi = \check{\rho}_\xi$ . If the flux functions of all road segments are known, the influence of these measurements will spread to the rest of the space.

Assuming we have a homogeneous stretch of road that can be described by a

single flux function, we can initialize the traffic model at some start time  $t^0 = 0$  to

$$\begin{aligned}\mathcal{X}^0 &= (N^0, t^0, z^0, \rho^0, Q^0, W^0, \mathcal{R}^0) = (0, 0, \emptyset, \rho^{\text{avg}}, Q^0, W^0, \emptyset), \\ Q^0 &= (V^0, \Sigma^0, \Lambda^{\pm, 0}, \xi^0, F^0) = ([0], [\rho^{\text{max}}], (0, 0), 0, \emptyset),\end{aligned}$$

with constant initial traffic density  $\rho^0$  using the overall average density known from historical data, wavespeed bounds  $W^0$  such that all wavespeeds are allowed, and  $\rho^{\text{max}}$  chosen large enough so that the traffic density will never exceed it. The flux functions and other parameters will be updated as new measurements become available, according to the algorithms described in the remainder of this section.

We update the traffic state in an event-based manner at some time  $t'$  when

1. a new CAV enters the road,
2. the difference between the current traffic density measurement of some CAV  $\check{\rho}'_{\xi^*}$ , and the previous forced traffic density  $F_{\xi^*}$ , exceeds some bound,  $|\check{\rho}'_{\xi^*} - F_{\xi^*}| > B_\rho$ , or
3. before prediction of the traffic state evolution.

If the state  $\mathcal{X}$  was last updated at time  $t$ , we may calculate the current state  $\mathcal{X}'$  at time  $t'$  as described in Section 4.5, chaining transitions that make the current state admissible, with the passage of time transition  $\tau(t')$  until the current time becomes  $t'$ . During this update, the CAVs are described by their previous forced densities  $F_\xi$ , and boundary speeds  $\lambda_\xi^\pm = \frac{x_\xi(t') - x_\xi(t)}{t' - t}$ , which ensure that the positions of the CAVs in  $\mathcal{X}'$  correspond to their real positions on the road  $x_\xi(t')$ . If the state update was caused by the entry of a new CAV, we assign a new unique identifier  $\xi^* > \max \Xi$  to it, where  $\Xi$  is the set of identifiers of all CAVs that were on the road before the new arrival. The newly arrived CAV  $\xi^*$  is then added to the set  $\Xi' = \Xi \cup \{\xi^*\}$ , and added to the model at position  $x_{\xi^*}(t')$ , with forced traffic density set to  $F'_{\xi^*} = \check{\rho}'_{\xi^*}$ . Otherwise, if the state update was caused by a large deviation of the currently measured traffic density  $\check{\rho}'_{\xi^*}$  from  $F_{\xi^*}$ , we update its forced traffic density to  $F'_{\xi^*} = \check{\rho}'_{\xi^*}$  once the state reaches  $\mathcal{X}'$ . The process then repeats at the next update time  $t''$ . We can influence the frequency of updates by choosing a higher  $B_\rho$ , which also decreases the sensitivity to measurement noise, but may yield a higher state reconstruction error.

### Learning the background traffic flux function $Q$

Consider a CAV  $\xi$  travelling through a road segment described by flux function  $Q$ , and let its reference speed be  $u_\xi \geq v_Q^{\text{max}}$ . Then the actual speed of the CAV will depend only on the speed of the traffic immediately downstream of it, and it will not influence the surrounding traffic, so  $\lambda_\xi^\pm = \check{v}_\xi$ , and since the CAV is following the flow of the traffic, we have  $\check{\omega}_\xi = 0$ . The measurements should then be consistent with the flux function,  $Q(\check{\rho}_\xi) \approx \check{q}_\xi$ .

Denote by  $(\check{\rho}^Q, \check{q}^Q)$  the set of all the collected measurement using all the CAVs inside the considered road segment. We may fit the data to a piecewise function with some number of breakpoints  $m$  satisfying the conditions imposed on flux functions (4.47), minimizing a choice of error metric using some batch processing algorithm. However, this batch processing approach might lead to sluggish update of  $Q$ , as well as unnecessary computational complexity, and might thus be unsuitable for real-time control.

Instead, here we propose a stream processing event-triggered method for on-line flux function update, summarized in Algorithm 1. Operator  $\text{avg}$  denotes finding the average value, and operator  $|\check{\rho}_\xi^{Q, B_\sigma}|$  is the number of elements in set  $\check{\rho}_\xi^{Q, B_\sigma}$ . Let  $\Xi_Q$  be the set of CAVs collecting measurements about the flux function  $Q$ . If, for any CAV  $\xi$ , the deviation of the measured traffic flow  $\check{q}_\xi$  from the modelled traffic flow based on the measured traffic density  $Q(\check{\rho}_\xi)$  is greater than some margin  $B_q$ , we update the flux function based on the measurement. Choosing higher  $B_q$  reduces the update frequency and sensitivity to measurement noise, at the cost of higher model error. Instead of using all the collected data, the set of measurements which we use for flux function fitting  $(\check{\rho}^Q, \check{q}^Q)$  will now consist only of the measurements that triggered a flux function update.

First, we add the  $(\check{\rho}_\xi, \check{q}_\xi)$  for which  $|Q(\check{\rho}_\xi) - \check{q}_\xi| > B_q$  to  $(\check{\rho}^Q, \check{q}^Q)$ . Next, if the flux function already has any breakpoints close to the measurement,  $|\sigma_i - \check{\rho}_\xi| < B_\sigma$ ,

---

**Algorithm 1:** Updating the background traffic flux function  $Q$  using measurements from CAVs travelling with the flow

---

**Input:**  $\check{\rho}_\xi, \check{q}_\xi, \xi \in \Xi_Q, Q$

**Output:**  $Q'$

**for**  $\xi \in \Xi_Q$  **do**

**if**  $|Q(\check{\rho}_\xi) - \check{q}_\xi| > B_q$  **then**

        Add  $(\check{\rho}_\xi, \check{q}_\xi)$  to  $(\check{\rho}^Q, \check{q}^Q)$ ;

        Remove from  $\Sigma_Q$  all  $\{\sigma \in \Sigma_Q : |\sigma - \check{\rho}_\xi| < B_\sigma\}$ ;

        Find  $(\check{\rho}_\xi^{Q, B_\sigma}, \check{q}_\xi^{Q, B_\sigma}) = \{(\rho, q) \in (\check{\rho}^Q, \check{q}^Q) : |\rho - \check{\rho}_\xi| < B_\sigma\}$ ;

        Add  $\sigma' = \text{avg} \check{\rho}_\xi^{Q, B_\sigma}$  to  $\Sigma_Q$ ;

        Recalculate  $V_Q$ , so that  $Q'(\sigma') = \text{avg} \check{q}_\xi^{Q, B_\sigma}$ ;

**if**  $|\check{\rho}_\xi^{Q, B_\sigma}| > B_n$  **then**

        Remove from  $(\check{\rho}^Q, \check{q}^Q)$  and  $(\check{\rho}_\xi^{Q, B_\sigma}, \check{q}_\xi^{Q, B_\sigma})$

$(\rho, q) = \arg \max_{(\rho, q) \in (\check{\rho}_\xi^{Q, B_\sigma}, \check{q}_\xi^{Q, B_\sigma})} \left( \frac{\rho}{\check{\rho}_\xi} - 1 \right)^2 + \left( \frac{q}{\check{q}_\xi} - 1 \right)^2$

**end**

**end**

**end**

---

they will be replaced by a new breakpoint; otherwise a new breakpoint is added without moving any of the other breakpoints. Choosing higher  $B_\sigma$  decreases the number of breakpoints in flux function  $Q$ , at the cost of potentially increasing the frequency of updates and higher model error. The new breakpoint  $\sigma'$  and its flux  $Q(\sigma')$  are taken as the average values of all measurements from  $(\check{\rho}^Q, \check{q}^Q)$  that are close to the new measurement,  $(\check{\rho}_\xi^{Q, B_\sigma}, \check{q}_\xi^{Q, B_\sigma}) = \{(\rho, q) \in (\check{\rho}^Q, \check{q}^Q) : |\rho - \check{\rho}_\xi| < B_\sigma\}$ . We therefore add the new breakpoint  $\sigma'$  to  $\Sigma_Q$  and recalculate  $V_Q$  so that we have  $Q'(\sigma') = q'$ , where  $\sigma'$  is the average of  $\check{\rho}_\xi^{Q, B_\sigma}$  and  $q'$  the average of  $\check{q}_\xi^{Q, B_\sigma}$ . The updated  $\Sigma_{Q'}$  and  $V_{Q'}$  are given by

$$\begin{aligned} \Sigma_{Q'} &= [ \sigma_1 \quad \dots \quad \sigma_{i_{\xi^-}} \quad | \quad \sigma' \quad | \quad \sigma_{i_{\xi^-}} \quad \dots \quad \sigma_m ]^\top, \\ V_{Q'} &= [ V_1 \quad \dots \quad V_{i_{\xi^-}} \quad | \quad V'_- \quad V'_+ \quad | \quad V_{i_{\xi^-}+1} \quad \dots \quad V_m ]^\top, \\ i_{\xi^-} &= \begin{cases} \max i \text{ s.t. } \sigma_i \in \Sigma_Q, \sigma_i < \check{\rho}_\xi - B_\sigma, & \sigma_1 \geq \check{\rho}_\xi - B_\sigma, \\ 0, & \sigma_1 < \check{\rho}_\xi - B_\sigma, \end{cases} \\ i_{\xi^+} &= \min i \text{ s.t. } \sigma_i \in \Sigma_Q, \sigma_i > \check{\rho}_\xi + B_\sigma, \\ V'_- &= \frac{q' - Q(\sigma_{i_{\xi^-}})}{\sigma' - \sigma_{i_{\xi^-}}}, \quad V'_+ = \frac{Q(\sigma_{i_{\xi^+}}) - q'}{\sigma_{i_{\xi^+}} - \sigma'} \end{aligned}$$

where the first blocks of  $\Sigma_{Q'}$  and  $V_{Q'}$  are empty if  $i_{\xi^-} = 0$ , and the third block of  $V_{Q'}$  is empty if  $i_{\xi^+} = m$ . We also need to ensure by suitable selection of initial  $Q$  that  $\sigma_m > \check{\rho}^{\max} - B_\sigma$ , i.e. the final breakpoint of  $Q$  is never moved.

Finally, in order to accelerate the adaptation of  $Q$  to potentially new condition on the road, we may choose to always calculate the breakpoint densities and flows using only the  $B_{\check{n}}$  most relevant measurements. Then, if after adding the new measurement to  $(\check{\rho}^Q, \check{q}^Q)$  we have  $|\check{\rho}_\xi^{Q, B_\sigma}| > B_{\check{n}}$ , i.e. the number of measurements close to  $\check{\rho}_\xi$  is higher than  $B_{\check{n}}$ , we may remove the measurement that differs from the measure  $(\check{\rho}_\xi, \check{q}_\xi)$  the most.

### Learning the bottleneck diagram $Q_\xi$

Consider again a CAV  $\xi$  travelling through a road segment described by flux function  $Q$ , but with reference speed that is lower than the speed of the surrounding traffic,  $u_\xi < \check{v}_\xi$ . Then the CAV will follow its reference speed, the overtaking flow will be  $\check{\omega}_\xi > 0$ , and the presence of the slow moving vehicle might be affecting the behaviour of the surrounding traffic. We aim to model this influence by using a different flux function  $Q_\xi$  at the position of the CAV, with  $\lambda_\xi^\pm = u_\xi$ . Then we need the measurement to satisfy

$$\begin{aligned} (\forall \rho \geq 0) \quad Q_\xi(\rho) &\leq \check{\omega}_\xi + \rho u_\xi, \\ (\exists \rho \geq 0) \quad Q_\xi(\rho) &= \check{\omega}_\xi + \rho u_\xi, \end{aligned}$$

i.e. the line  $\check{\omega}_\xi + u_\xi \rho$  should be tangent to  $Q_\xi(\rho)$ .

---

**Algorithm 2:** Updating the bottleneck diagram flux function  $Q_\xi$  using measurements of overtaking flow

---

**Input:**  $\check{\omega}_\xi, u_\xi, Q_\xi$

**Output:**  $Q'_\xi$

Find  $\omega_\xi^{\max} = \max_{\sigma \in \Sigma_{Q_\xi}} Q_\xi(\sigma) - u_\xi \sigma = Q_\xi(\sigma_{i_\xi^\omega}) - u_\xi \sigma_{i_\xi^\omega}$ ;

**if**  $\omega_\xi^{\max} > \check{\omega}_\xi + B_\omega$  **then**

    Find  $\sigma_- = \min \rho \geq 0$  s.t.  $Q_\xi(\rho) = u_\xi \rho + \check{\omega}_\xi$ , and

$\sigma_+ = \max \rho \geq 0$  s.t.  $Q_\xi(\rho) = u_\xi \rho + \check{\omega}_\xi$ ;

    Remove from  $\Sigma_{Q_\xi}$  all  $\{\sigma \in \Sigma_{Q_\xi} : \sigma_- \leq \sigma \leq \sigma_+\}$ ;

    Add  $\{\sigma_-, \sigma_+\}$  to  $\Sigma_{Q_\xi}$ ;

    Recalculate  $V_{Q_\xi}$ , so that  $Q'_\xi(\rho) = \check{\omega}_\xi + u_\xi \rho, \sigma_- \leq \rho \leq \sigma_+$ ;

**else if**  $\omega_\xi^{\max} < \check{\omega}_\xi - B_\omega$  **then**

    Recalculate  $V_{Q_\xi}$ , so that  $Q'_\xi(\sigma_{i_\xi^\omega}) = \check{\omega}_\xi + u_\xi \sigma_{i_\xi^\omega}$ ;

**end**

---

If the conditions (8.3) are violated by more than some margin  $B_\omega$ , we update flux function  $Q_\xi$  according to the proposed Algorithm 2 so that (8.3) will hold. Choosing a higher  $B_\omega$  decreases the frequency of updates of  $Q_\xi$  and sensitivity to measurement noise, but increases the model error. We first calculate the maximum overtaking flow  $\omega_\xi^{\max}$  that  $Q_\xi$  allows in case the boundary speed is  $u_\xi$ , which is achieved for traffic density  $\sigma_{i_\xi^\omega}$ ,  $\omega_\xi^{\max} = Q_\xi(\sigma_{i_\xi^\omega}) - u_\xi \sigma_{i_\xi^\omega}$ . Since  $Q_\xi$  is piecewise linear, we only need to search for  $\sigma_{i_\xi^\omega}^\omega$  within the set of its breakpoints  $\Sigma_{Q_\xi}$ . Then, if  $\omega_\xi^{\max} > \check{\omega}_\xi + B_\omega$ , the current flux function  $Q_\xi$  admits a higher overtaking flow than the measured one, so we proceed to update it with

$$\begin{aligned} \Sigma_{Q'_\xi} &= [ \sigma_1 \quad \dots \quad \sigma_{i_{\xi^-}} \mid \sigma_- \quad \sigma_+ \mid \sigma_{i_{\xi^-}} \quad \dots \quad \sigma_m ]^\top, \\ V_{Q'_\xi} &= [ V_1 \quad \dots \quad V_{i_{\xi^-}+1} \mid u_\xi \quad \mid V_{i_{\xi^+}} \quad \dots \quad V_m ]^\top, \\ \sigma_- &= \min \rho \geq 0 \text{ s.t. } Q_\xi(\rho) = u_\xi \rho + \check{\omega}_\xi \\ \sigma_+ &= \max \rho \geq 0 \text{ s.t. } Q_\xi(\rho) = u_\xi \rho + \check{\omega}_\xi \\ i_{\xi^-} &= \begin{cases} \max i \text{ s.t. } \sigma_i \in \Sigma_{Q_\xi}, \sigma_i < \sigma_-, \sigma_1 \geq \sigma_-, \\ 0, & \sigma_1 < \sigma_-, \end{cases} \\ i_{\xi^+} &= \min i \text{ s.t. } \sigma_i \in \Sigma_{Q_\xi}, \sigma_i > \sigma_+, \end{aligned}$$

where the first block of  $\Sigma_{Q'_\xi}$  is empty if  $i_{\xi^-} = 0$ . Otherwise, if  $\omega_\xi^{\max} < \check{\omega}_\xi - B_\omega$ , the current flux function  $Q_\xi$  admits a lower overtaking flow than the measured one, so we proceed to update it with

$$V_{Q'_\xi} = [ V_1 \quad \dots \quad V_{i_{\xi^-}-1} \mid V'_- \quad V'_+ \mid V_{i_{\xi^+}+2} \quad \dots \quad V_m ]^\top,$$

---

**Algorithm 3:** Updating the wavefront speed bounds  $W$  using traffic density and flow measurements

---

**Input:**  $\check{\rho}_\xi, \check{q}_\xi, \check{v}_\xi, u_\xi, \xi \in \Xi, Q, W$   
**Output:**  $W'$

```

for  $\xi \in \Xi$  do
  if  $|\check{\rho}'_\xi - \check{\rho}_\xi| > B_\rho$  then
    if  $u_\xi > \check{v}_\xi, u_\xi > \check{v}'_\xi, \check{\rho}_\xi \tilde{\Sigma}_{Q'}^{\check{\rho}'_\xi} \neq [\check{\rho}_\xi \check{\rho}'_\xi]^\top$  then
      if  $\check{\rho}_\xi > \check{\rho}'_\xi$  then
        if  $\check{w}_\xi > \check{\rho}_\xi \tilde{V}_{Q,1}^{\check{\rho}'_\xi}$  then
          Add  $\check{w}_\xi$  to  $\check{w}^{\check{W}'_-}$ ;
        else if  $\check{w}_\xi < \check{\rho}_\xi \tilde{V}_{Q,1}^{\check{\rho}'_\xi}$  then
          Add  $\check{w}_\xi$  to  $\check{w}^{\check{W}'_+}$ ;
        end
      else
        if  $\check{w}_\xi > \check{\rho}_\xi \tilde{V}_{Q,1}^{\check{\rho}'_\xi}$  then
          Add  $\check{w}_\xi$  to  $\check{w}^{\check{W}'_-}$ ;
        else if  $\check{w}_\xi < \check{\rho}_\xi \tilde{V}_{Q,1}^{\check{\rho}'_\xi}$  then
          Add  $\check{w}_\xi$  to  $\check{w}^{\check{W}'_+}$ ;
        end
      end
      Recalculate all  $W$  for which  $\check{w}^{\check{W}'_\pm} \neq \emptyset$  so that  $\check{W}'_\pm = \text{avg} \check{w}^{\check{W}'_\pm}$ ;
    end
  end
end
end

```

---

$$V'_- = \frac{\check{w}_\xi + u_\xi \sigma_{i_\xi^\omega} - Q_\xi(\sigma_{i_\xi^\omega - 1})}{\sigma_{i_\xi^\omega} - \sigma_{i_\xi^\omega - 1}}, \quad V'_+ = \frac{Q_\xi(\sigma_{i_\xi^\omega + 1}) - \check{w}_\xi - u_\xi \sigma_{i_\xi^\omega}}{\sigma_{i_\xi^\omega + 1} - \sigma_{i_\xi^\omega}},$$

where if  $i_\xi^\omega = 1$ , we take  $\sigma_{i_\xi^\omega - 1} = 0$ , and the first block of  $V_{Q'_\xi}$  is empty.

### Learning the wavefront speed bounds $W$

Finally, we may also use the traffic density and speed measurements of the CAVs to learn the wavefront speed bounds  $W$ . We may detect the influence of wavefront speed bounds upon a change in the measurements of some CAV  $\xi$  as it enters a zone of different traffic density, when  $|\check{\rho}'_\xi - \check{\rho}_\xi| > B_\rho$ , where  $\check{\rho}_\xi$  is the old traffic density measurement, and  $\check{\rho}'_\xi$  the new one. If the CAV  $\xi$  is travelling with the traffic flow without affecting it ( $u_\xi > \check{v}_\xi$  and  $u_\xi > \check{v}'_\xi$ ), then as discussed in Section 4.5.1, the solution to the Riemann problem between these two traffic densities should

include zones of traffic density  $\check{\rho}_\xi \tilde{\Sigma}_Q^{\check{\rho}'_\xi}$ , according to the flux function that describes the traffic flow  $Q$ . Essentially, the transition from  $\check{\rho}_\xi$  to  $\check{\rho}'_\xi$  should happen along an upper concave envelope of  $Q$  if  $\check{\rho}_\xi > \check{\rho}'_\xi$ , or over a lower convex envelope of  $Q$  if  $\check{\rho}_\xi < \check{\rho}'_\xi$ . Therefore, if  $\check{\rho}_\xi \tilde{\Sigma}_Q^{\check{\rho}'_\xi} \neq [\check{\rho}_\xi \check{\rho}'_\xi]^\top$ , we can suspect that the reason for this discrepancy is the limit on wavefront speeds, which we will use to estimate  $W = (\check{W}_-, \check{W}_+, \hat{W}_-, \hat{W}_+)$ , as outlined in Algorithm 3.

We denote the wavefront speed based on these measurements  $\check{w}_\xi = \frac{\check{q}'_\xi - \check{q}_\xi}{\check{\rho}'_\xi - \check{\rho}_\xi}$ . If  $\check{\rho}_\xi \tilde{\Sigma}_Q^{\check{\rho}'_\xi} \neq [\check{\rho}_\xi \check{\rho}'_\xi]^\top$ , this wavefront speed offers information about one wavefront speed bound:  $\check{W}_\pm$  or  $\hat{W}_\pm$  depending on whether  $\check{\rho}_\xi < \check{\rho}'_\xi$  or  $\check{\rho}_\xi > \check{\rho}'_\xi$ , respectively, and  $\check{W}_-$  or  $\check{W}_+$  depending on whether  $\check{w}_\xi > \check{\rho}_\xi \tilde{V}_{Q,1}^{\check{\rho}'_\xi}$  or  $\check{w}_\xi < \check{\rho}_\xi \tilde{V}_{Q,1}^{\check{\rho}'_\xi}$ , respectively. Finally, we calculate those wavefront speed bounds  $\check{W}_\pm$ , for which we have at least one such measurement,  $\check{w}^{\check{W}_\pm} \neq \emptyset$ , as the average over the whole set,  $\check{W}_\pm = \text{avg} \check{w}^{\check{W}_\pm}$ .

### 8.3.2 Control laws

Four cases of control are considered: control with full access to the information about both the traffic state and traffic model, using the actual traffic state but without considering the variation in the traffic model, using the actual traffic model and reconstructed traffic density, and prediction-based control using the FTTSM. The control action of CAV  $\xi$  is its reference speed,  $u_\xi(t) \in [u^{\min}, u^{\max}]$ . The only restriction on the reference speed is in practice the minimum speed  $u_\xi(t) \geq u^{\min}$ , as the maximum speed of the CAV will be dictated by the speed of the surrounding traffic. The first three control laws differ only in what information they use, i.e. whether they use the reconstructed or actual traffic state, and exact or estimated model, and are discussed in more detail in Section 8.1. Here we briefly present the simplified approximate control laws, and outline based on which information the control action is calculated.

**Full-information control:** When calculating the full-information control at time  $t_0$ , we assume that the full traffic state  $\rho(x, t_0)$ , the flux function of the road  $Q(\rho)$ , and the stop-and-go wave speed  $\lambda_\psi = \check{W}_-$  are all known. We may simplify the calculation of  $u_\xi^{\psi^*}(t)$  by approximating  $q_\xi^{\text{lo}}(u_\xi^{\psi^*}(t_0)) \approx q_\xi^{\text{lo}}(u^{\min})$ , and ignoring the stop-and-go waves that are created in the wake of the preceding CAVs. In fact, since  $q_\xi^{\text{lo}}(u_\xi^{\psi^*}(t_0)) \leq q_\xi^{\text{lo}}(u^{\min})$  if  $u_\xi^{\psi^*}(t_0) \geq u^{\min}$ , the traffic flow overtaking moving bottleneck  $\xi$  is overestimated if the approximate  $q_\xi^{\text{lo}}$  is used, leading to more robust control. Denoting the traffic density of the traffic flow overtaking the moving bottleneck moving at minimum speed as  $\rho_\xi^{\text{d,max}}$ ,  $Q(\rho_\xi^{\text{d,max}}) = q_\xi^{\text{lo}}(u^{\min})$ , the traffic density of the discharging traffic flow from the stop-and-go wave as  $\rho_\psi^{\text{d}}$ ,  $q_\psi^{\text{dis}} = Q(\rho_\psi^{\text{d}})$ , and the average traffic density from  $x_\xi(t_0)$  to  $z_\psi(t_0)$  as  $\rho_{x_\xi}^{z_\psi}(t_0)$ , we

can express the reference speed  $u_\xi^{\psi*}(t)$  in closed form,

$$u_\xi^{\psi*}(t) = \frac{q_\xi^{\text{lo}}(u^{\text{min}}) - q_\psi^{\text{dis}} + \hat{W}_- \left( \rho_\psi^d - \rho_{x_\xi}^{z_\psi}(t_0) \right)}{\rho_\xi^{\text{d,max}} - \rho_{x_\xi}^{z_\psi}(t_0)}. \quad (8.4)$$

The reference speed is thus given by (8.1), where the set  $\Psi_\xi(t_0)$  contains the stop-and-go waves downstream of  $x_\xi(t_0)$  that the preceding CAVs are not able to dissipate.

**Estimated-model control:** In the full-information case, we assume that both the model and the current traffic density profile are fully known. In reality, the parameters of the traffic model might change due to varying weather or other conditions, and the traffic density profile will have to be reconstructed using available measurements. If the true model is not known, we may calculate the reference speeds using its best available estimate,

$$u_\xi^{\psi*}(t) = \frac{\hat{q}_\xi^{\text{lo}}(u^{\text{min}}) - \hat{q}_\psi^{\text{dis}} + \hat{W}_- \left( \hat{\rho}_\psi^d - \hat{\rho}_{x_\xi}^{z_\psi}(t_0) \right)}{\hat{\rho}_\xi^{\text{d,max}} - \hat{\rho}_{x_\xi}^{z_\psi}(t_0)}. \quad (8.5)$$

Note that the average density  $\rho_{x_\xi}^{z_\psi}(t_0)$  is assumed to be known exactly. If the actual model changes, or if it deviates from the estimated one, the calculated control action may be overly conservative, leading to superfluous traffic disruption, or overly optimistic, leading to failure to dissipate stop-and-go waves.

**Reconstructed-state control:** Conversely, if the traffic density profile is not known, we may use the estimated average density  $\hat{\rho}_{x_\xi}^{z_\psi}(t_0)$  instead,

$$u_\xi^{\psi*}(t) = \frac{q_\xi^{\text{lo}}(u^{\text{min}}) - \hat{q}_\psi^{\text{dis}} + \hat{W}_- \left( \hat{\rho}_\psi^d - \hat{\rho}_{x_\xi}^{z_\psi}(t_0) \right)}{\rho_\xi^{\text{d,max}} - \hat{\rho}_{x_\xi}^{z_\psi}(t_0)}. \quad (8.6)$$

The stop-and-go wave discharging density  $\hat{\rho}_\psi^d$  and flow  $\hat{q}_\psi^{\text{dis}}$  depend on the maximum traffic density of the stop-and-go wave, and will therefore have to be estimated based on the reconstructed traffic profile  $\hat{\rho}(x, t_0)$ . We may use the simple traffic state reconstruction algorithm described in Section 8.2.1. In this case, the full set of stop-and-go waves  $\Psi(t_0)$  may not be known, since some of the stop-and-go waves are possibly not detected, and we use  $\hat{\Psi}(t_0)$  based on  $\hat{\rho}(x, t_0)$  instead.

**FTTSM-based control:** Finally, we employ the FTTSM to calculate control actions for all CAVs. Employing the traffic state reconstruction and model learning algorithms described in Section 8.3, we acquire the current estimated FTTSM state  $\hat{\mathcal{X}}(t_0)$ , which includes both the traffic density and model information. The control actions  $u_\xi(t_0)$  will be based on the predicted state  $\hat{\mathcal{X}}(t_0 + \theta)$  with  $\theta$  ranging from 0 to the prediction horizon  $\Theta$ , taken long enough that all CAVs are guaranteed to either dissipate or run into stop-and-go waves. During the process of finding  $\hat{\mathcal{X}}(t + \theta)$ , all

traffic densities are unforced  $\hat{F}_\xi = \emptyset$ , and the boundary speeds of all CAVs are set to

$$\hat{\lambda}_\xi^\pm = \begin{cases} u^{\min}, & \hat{v}_{\xi+} > u^{\min} \wedge (\exists i \geq i_{\xi+}) \hat{\rho}_i \geq \hat{\sigma}^* \\ \hat{v}_{\xi+}, & \text{otherwise,} \end{cases}$$

where  $v_{\xi+}$  is given by (4.49),  $i_{\xi+}$  by (4.50), and  $\hat{\sigma}^*$  is the minimum traffic density considered as a stop-and-go wave,

$$\hat{\sigma}^* = \begin{cases} \sigma_{\hat{Q}}^{\max}, & (\nexists i) \hat{V}_i < \hat{W}_-, \\ \max \left\{ \sigma_{\hat{Q}}^{\max}, \sigma_{\hat{Q}}^{\hat{W}_-} \right\}, & (\exists i) \hat{V}_i < \hat{W}_-, \end{cases}$$

$$\sigma_{\hat{Q}}^{\hat{W}_-} = \hat{\sigma}_i, \quad i = \max j \text{ s.t. } \hat{V}_k > \hat{W}_-, k = 1, \dots, j.$$

Essentially, the CAVs are driven at minimum speed, applying maximum control action, until they run into congestion or there is no congestion downstream of their location.

Next, for each CAV  $\xi$ , we identify the position and prediction time  $(\hat{x}_\xi^*, \hat{\tau}_\xi^*)$  at which the final zone  $i$  where  $\hat{\rho}_i > \hat{\sigma}^*$  is dissipated, i.e. the congestion is dissipated and the CAV no longer needs to apply control action. The prediction time is given in relative coordinates, compared to the initial time  $t$  at which the prediction is calculated. This congestion will include the stop-and-go waves initially present in the system, as well as potential congestion created in the wake of CAVs downstream of the considered one. If the CAV  $\xi$  fails to dissipate all congestion, and instead enters congested traffic,  $\hat{v}_{\xi+} < u^{\min}$ , we instead set  $(\hat{x}_\xi^*, \hat{\tau}_\xi^*)$  to the point at which this happened, and if there was no congestion downstream of CAV  $\xi$  at the beginning of the prediction, we set  $(\hat{x}_\xi^*, \hat{\tau}_\xi^*) = (x_\xi(t), 0)$ . Then, the reference speed  $u_\xi(t)$  is set to

$$u_\xi(t) = \begin{cases} u^{\max}, & \hat{\tau}_\xi^* = 0, \\ \max \left\{ u^{\min}, \min \left\{ u^{\max}, \frac{\hat{x}_\xi^* + \hat{x}_{\vec{\xi}} - \hat{x}_\xi(t) + u^{\min} \hat{\tau}_\xi^*}{\hat{\tau}_\xi^*} \right\} \right\}, & \hat{\tau}_\xi^* > 0, \end{cases} \quad (8.7)$$

where  $\vec{\xi}$  denotes the first CAV downstream of CAV  $\xi$ . The control action is therefore calculated by first predicting the evolution of the system under maximum control effort, and then reducing the control effort where possible, based on the prediction.

### 8.3.3 Simulation setup

Apart from the FTTSM, which is used as the model for traffic state reconstruction and control calculation for the FTTSM-based control, we also use a variant of the CTM with process noise as the simulation model to test the proposed traffic state reconstruction, model learning and control algorithms. The model used consists of

(3.16), (3.17), with the demand and supply functions based on(3.20),

$$D_i(t^T) = \underline{\rho}_i(t^T) \min \left\{ U_i(t^T), \max \left\{ 0, \mathcal{V}(\underline{\rho}_i(t^T)) + \delta_i(t^T) \right\} \right\},$$

$$S_i(t^T) = \bar{\rho}_i(t^T) \min \left\{ \max \left\{ 0, \mathcal{V}(\bar{\rho}_i(t^T)) + \delta_{i-1}(t^T) \right\} \right\},$$

where  $\underline{\rho}_i(t^T) = \min\{\rho_i(t^T), \sigma_Q^{\max}\}$ ,  $\bar{\rho}_i(t^T) = \max\{\rho_i(t^T), \sigma_Q^{\max}\}$ , and  $\delta_i(t^T)$  is an added normally distributed noise term  $\delta_i(t) \sim \mathcal{N}(0, \Delta)$  that models the stochastic nature of human driving.

The flux function used is (3.11), yielding average speed of each cell is given by the speed-density relation

$$\mathcal{V}(\rho) = V_Q^{\max} \exp \left( -\frac{1}{m} \left( \frac{\rho}{\sigma_Q^{\max}} \right)^m \right). \quad (8.8)$$

The inflow to the first cell  $q_0(t^T)$  is given as an external input, and can be delayed if the conditions in cell 1 are such that they can not admit such flow, and the outflow from the final cell  $q_N(t^T)$  will be limited by extraneously setting  $S_{N+1}(t^T)$  in order to generate stop-and-go waves entering the road segment from downstream. We use reference maximum speed  $U_i(t^T)$ , to model the influence of moving bottlenecks with (4.27) and stop-and-go waves with (4.31), as described in Section 4.3.3 and 4.3.4.

The simulations were executed on a 10 km stretch of a two-lane road, which consists of  $N = 100$  cells of length  $L = 100$  m each. The simulation length is taken to be  $t^{\text{end}} = 3$  h, and the simulation time step is  $T = 3$  s. The inflow to the road segment randomly varies in time, changing every minute, and is uniformly distributed  $q_0(t) \sim \mathcal{U}(1450, 4350)$  veh/h. The CAVs arrive with random time gaps between them,  $g_\xi = \max\{\gamma_\xi, 30\}$  s,  $\gamma_\xi \sim \text{Exp}(\frac{1}{30})$ . We create the stop-and-go waves, arriving from downstream, by limiting the maximum outflow at the downstream end of the road segment to  $\mathcal{U}(200, 400)$  veh/h for 30 s, causing congestion to build up. Once this restriction is removed, the congestion will start dissipating, with capacity drop, propagating upstream as a stop-and-go wave. The time gap between two generated stop-and-go waves is uniformly distributed  $g_\psi \sim \mathcal{U}(360, 1080)$  s (6 to 18 minutes).

In order to demonstrate how control adapts to time-varying traffic model, e.g. due to a change in weather conditions when it starts raining [194], we change the flux function at time  $t = 0.5$  h. Both flux function are given by  $Q(\rho) = \rho\mathcal{V}(\rho)$ , (8.8), with parameters  $m = 2.34$ ,  $V_Q^{\max} = 120$  km/h,  $\sigma_Q^{\max} = 51.1$  veh/km, yielding capacity  $Q^{\max} = 4000$  veh/h, for  $t \leq 0.5$  h, and  $m = 3.4$ ,  $V_Q^{\max} = 75$  km/h,  $\sigma_Q^{\max} = 60.4$  veh/km, yielding capacity  $Q^{\max} = 3375$  veh/h, for  $t > 0.5$  h. The variance of the additive process noise of the velocity is in both cases  $\Delta = 16$  km<sup>2</sup>/h<sup>2</sup>.

The FTTSM state reconstruction and model learning algorithms described in this section used traffic flow bound  $B_q = 1000$ , flux function breakpoint bound  $B_\sigma = 20$ , bound on the maximum number of used data points  $B_n = 20$ , overtaking

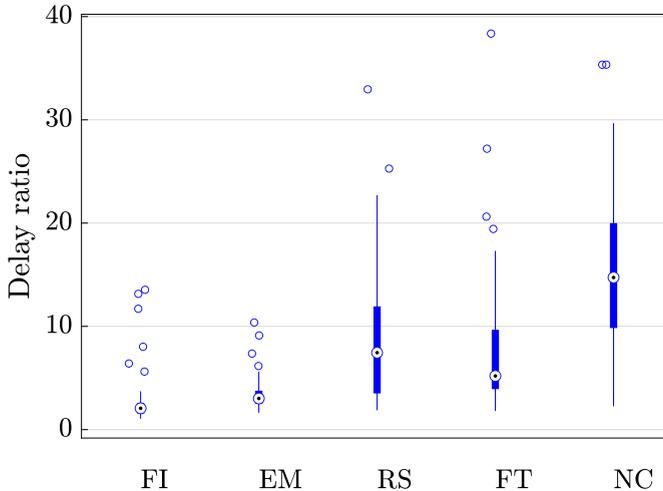


Figure 8.4: Box plots of delay ratios with different control laws. FI: Full-information (8.4), EM: Estimated-model (8.5), RS: Reconstructed-state (8.6), FT: FTSM-based (8.7), NC: No control.

Delay ratio [%]	FI	EM	RS	FT	NC
Mean	2.92	3.44	9.44	8.02	15.69
Median	2.06	3.02	7.45	5.20	14.73

Table 8.3: Achieved mean and median delay ratios of the different control laws.

flow bound  $B_\omega = 500$ , and traffic density bound  $B_\rho = 30$ . Furthermore, all of the measurements were perturbed with multiplicative noise  $\varepsilon \sim \mathcal{N}(1, 0.0025)$ .

### 8.3.4 Simulation results

Finally, we executed 50 simulation runs, comparing the performance of the four described control laws. For each simulation run, six simulations with the same realization of all random variables were executed, one using each of the control laws, one with no control, and a benchmark simulation with no stop-and-go waves. The performance metric used was the delay ratio, defined as the percentage increase of Total Time Spent (TTS) compared to the TTS of the benchmark simulation,  $DR = \frac{TTS_i^{\text{case}} - TTS_{\text{min}}^{\text{min}}}{TTS_{\text{min}}^{\text{min}}} 100\%$ . The results are shown in Figure 8.4 and Table 8.3.

We can see that all control cases achieve significant reduction of delay compared to the uncontrolled case. In particular, control laws that use full information about the traffic density profile (FI and EM) perform significantly better than those based

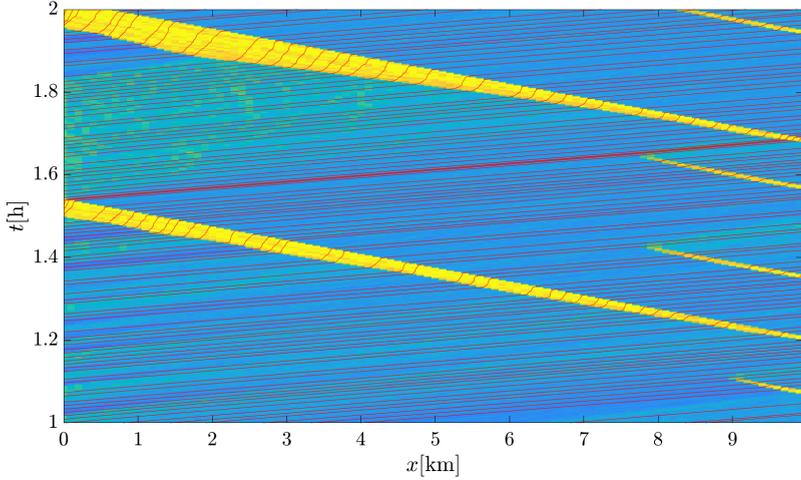


Figure 8.5: Detail from the simulations, no control

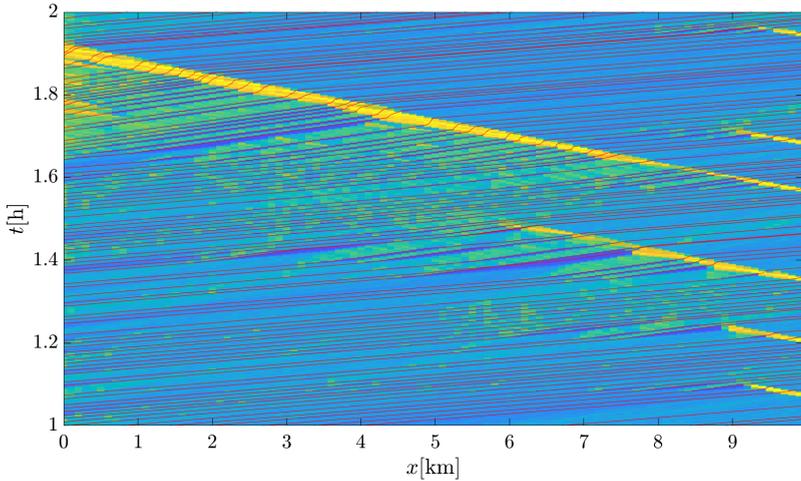


Figure 8.6: Detail from the simulations, reconstructed-state control

on state reconstruction (RS and FT). This is not surprising, since in this case, the control law is able to react as soon as a stop-and-go wave appears on the road, instead of waiting for one CAV to detect it first. In this case, not knowing the exact model only deteriorated the control performance slightly (in case of EM compared to FI), since the change in model parameters was not severe.

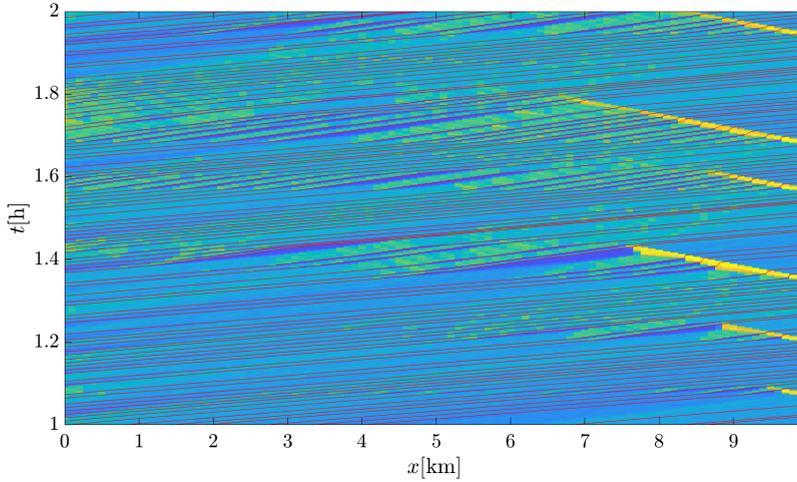


Figure 8.7: Detail from the simulations, full-information control

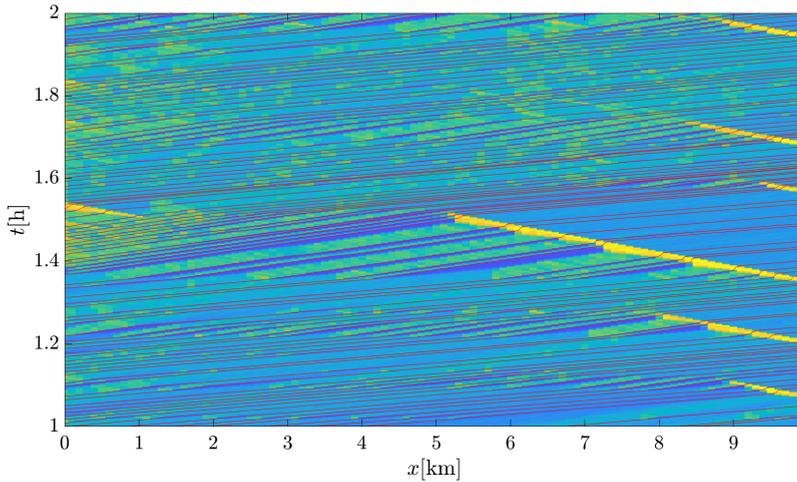


Figure 8.8: Detail from the simulations, FTSM-based control

Reconstructed-state control performs worse than the other control laws, probably due to the fact that it does not correctly capture the build-up of congestion in the wake of CAVs. Conversely, since the traffic state reconstruction is done model-based in case of the FTSM-based control, this control law achieves better performance, in spite of using less information. Overall, in spite of being fully

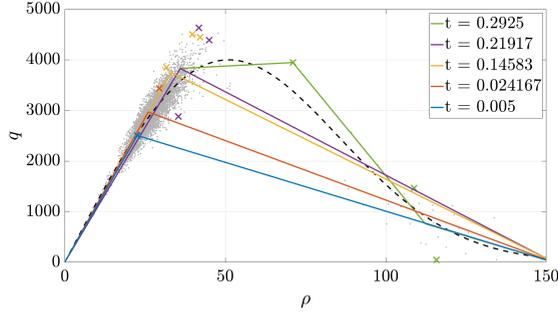


Figure 8.9: Overview of the changes to  $\hat{Q}$  through time in one simulation run before  $t = 0.5$ . Some updates when  $Q$  is changed are shown. Density-flow measurement pairs for  $t < 0.5$  are shown as grey dots and the current  $Q$  is shown in dashed black. The estimation of  $\hat{Q}$  is done based on measurements  $(\hat{\rho}^Q, \hat{q}^Q)$ , shown as  $\times$ -es, and newly added  $\times$ -es are shown in the same colour as the current  $\hat{Q}$ .

data-driven, this control law is able to dissipate stop-and-go waves and significantly reduce the overall Total Time Spent.

In Figure 8.8 we show a detail from one of the simulations. In case no control is applied, stop-and-go waves will grow and propagate upstream. In this simulation run, the reconstructed-state control failed to dissipate the stop-and-go wave, and the other control laws were successful, albeit with the FTSM-based control causing some more congestion further upstream.

The background traffic flux function  $Q$  is the most impactful component of the overall traffic prediction model, so the process of estimating it using Algorithm 1 is shown in Figures 8.9 and 8.10. It can be seen that the first estimates of  $Q$  have a single breakpoint and are triangular, because all available measurements are still tightly clustered at that time. Later, as measurements of congested traffic become available, more breakpoints are added and the flux function takes a somewhat more complex shape. Once the flux function is changed at  $t = 0.5$ , the algorithm gradually adapts the estimate to the new flux functions and either offsets or eliminates the measurements that are the worst outliers. Even though the traffic speed is modelled with significant process noise, and the measurements used are also noisy, the learning algorithms are able to quickly estimate the flux function using a low number of memorized measurements. Moreover, even though the piecewise-linear flux function deviates from the form of the underlying actual flux function, this deviation does not reflect particularly negatively on the control performance, while adopting such a simple form of the flux function greatly increases the numerical efficiency of the prediction model.

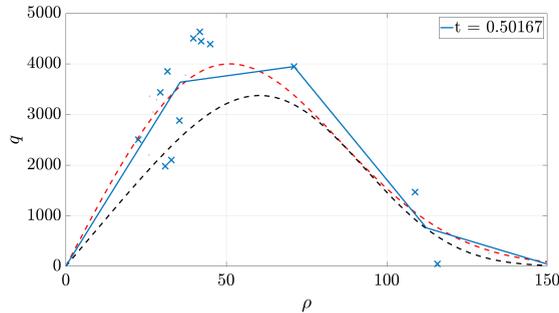
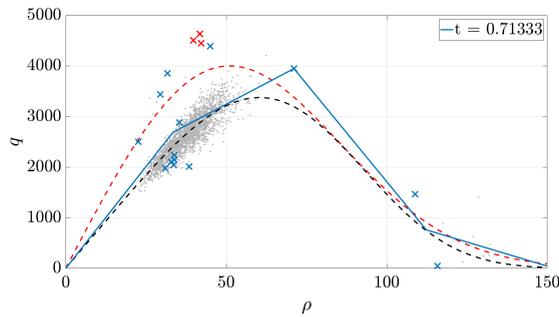
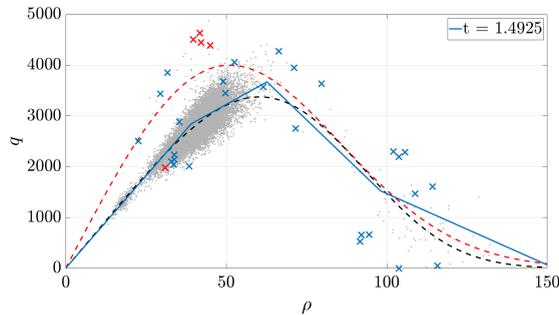
(a) At  $t = 0.50167$ (b) At  $t = 0.71333$ (c) At  $t = 1.4925$ 

Figure 8.10: Overview of the changes to  $\hat{Q}$  through time in one simulation run after  $t = 0.5$ . In 8.10a, 8.10b, and 8.10c,  $\hat{Q}$  is shown at three different times  $t > 0.5$ . Here, only the measurements from  $t = 0.5$  to the time when  $\hat{Q}$  was estimated are shown. The current  $Q$  is shown in dashed black and the old  $Q$  in dashed red. Blue  $\times$ -es are the measurements used to estimate  $\hat{Q}$ , and red  $\times$ -es are the “forgotten” measurements, that were used at some previous time, but have since been removed from  $(\check{\rho}^Q, \check{q}^Q)$ .

## 8.4 Summary

In this chapter we explored the effect of using reconstructed traffic state and learned traffic model for multiple stop-and-go waves dissipation control. Both the state reconstruction and actuation are executed using connected automated vehicles, and the only other information that we assume is known is the average traffic flow of the road. As the number and density of connected vehicles increases, the reconstruction-based control approaches its full-information counterpart. A simple traffic state reconstruction method was proposed, along with an empirical probe CAV activation scheme. Then, an improved traffic state reconstruction method using the FTTSM is presented, along with algorithms for learning different components of the traffic model from data. The proposed control laws were tested in simulations and shown to be able to increase the throughput of the road by dissipating stop-and-go waves entering from its downstream end, leading to a reduction in delay that the vehicles experience.

**Part IV**

**Conclusions**



## Chapter 9

# Summary and future research

FINALLY, in this chapter we conclude the thesis. In Section 9.1 we summarize and discuss the presented results, and in Section 9.2 outline some possible extensions and future work on the considered topic.

### 9.1 Summary

The central question of this thesis was how traffic control can be implemented using exclusively a small portion of connected vehicles, whose trajectories and local measurements we can use, and over which we have direct control. Focusing on the Lagrangian framework of traffic control allows us to exploit the largely untapped potential that the emerging technologies of Connected and Autonomous Vehicles can bring to the field of Intelligent Transportation Systems in the near future. The mobile nature of sensors and actuators is especially advantageous in case the road of interest is not instrumented with the required traffic control equipment, as well as if the traffic phenomenon is mobile itself.

In order to design and implement the Lagrangian traffic control laws, we need an appropriate traffic model. This model needs to be able to capture the mutual influence that the directly controlled vehicle and the rest of the traffic have on each other, and we model this interaction by considering the connected vehicles as controlled moving bottlenecks. Furthermore, other traffic phenomena that are relevant to the considered problems, such as stop-and-go waves and capacity drop, also need to be properly modelled. In this thesis we considered four related modelling frameworks that can represent all of the relevant traffic phenomena, and we studied three control problems: platoon merging coordination, congestion dissipation, and traffic state reconstruction.

First, we introduced moving bottlenecks to the cell transmission model by exploiting its correspondence with the Lighthill-Whitham-Richards model, where moving bottlenecks were modelled by introducing constraints on capacity, or a moving zone with a reduced flux function. Second, in the multi-class cell transmission

model, we used the reference speed of each class in different cells to force the density profile to reflect the influence of platoons acting as moving bottlenecks and stop-and-go waves. In both these cell-based traffic models, it was shown that capacity drop at a stationary bottleneck can be introduced in a straightforward way by modifying the demand or supply functions. Next, the tandem queueing model includes capacity drop in its basic form. Next, we extended the tandem queueing model to include moving bottlenecks and stop-and-go waves, by allowing the delay between two queues to vary in time, and allowing the structure of the model to change. Finally, we proposed the front tracking transition system model, in which moving bottlenecks are conveniently modelled in the same way as with the Lighthill-Whitham-Richards model. We modelled stop-and-go waves by introducing a new type of weak solution, with constraints on the rarefaction wave front speed. We have presented a variety of macroscopic traffic models, and their advantages and disadvantages.

The first traffic control problem we studied was coordination of a pair of trucks merging into a platoon on a highway. In order to improve the reliability and fuel economy of the platoon merging process, all external influences need to be considered, including the influence of the road grade and the surrounding traffic. We used experimental data to identify the influence of road grade on both vehicles, allowing us to better predict when and where they will merge into a platoon. By modelling the trucks as moving bottlenecks, we were able to predict how much the follower vehicle will be delayed by the congestion left in the wake of the leader vehicle, and formulate a control law for platoon merging that minimizes the fuel consumption. Our approach has the additional benefit of being able to indicate when attempting to form a platoon would not yield fuel savings, and thus prevent unnecessary deviations from vehicles' original plans.

The congestion dissipation problem was studied in three settings: decongestion of a single stationary bottleneck, dissipation of a single stop-and-go wave using a single controlled moving bottleneck, and dissipation of multiple stop-and-go waves using multiple controlled moving bottlenecks. We used platoons of vehicles to decongest stationary bottlenecks and keep them in free flow, by controlling the speed of the platoons, and how many lanes they occupy. This way, we were able to counteract the generally negative effect that arrivals of platoons have on the traffic situation around bottlenecks. By shaping the traffic flow that reaches the bottleneck, we were able to dissipate its queue and keep it in free flow, eliminating a large portion of the additional delay that would be incurred due to capacity drop. Connected vehicles acting as both sensors and actuators were shown to be a particularly good match for dissipating stop-and-go waves, since they can be employed wherever a stop-and-go wave arises. In the simplest case, if there are only a few lanes of traffic, a single controlled vehicle can be used to dissipate it. By slowing down the controlled vehicle and using it as a moving bottleneck, we were able to restrict the inflow of the traffic to the stop-and-go wave until it dissipates. We used appropriate traffic models to predict the evolution of the stop-and-go wave and calculate the optimal moving bottleneck speed, minimizing the total time spent

on the road for all vehicles. This approach was then extended to dissipation of multiple stop-and-go waves, using multiple controlled moving bottlenecks. In this case, each moving bottleneck attempts to dissipate both the stop-and-go waves that are currently present downstream of it, and the stop-and-go waves that will arise from the congestion left in the wake of other moving bottlenecks, further improving the traffic situation.

Finally, we considered the traffic state reconstruction problem in tandem with multiple stop-and-go wave dissipation, by dropping the assumption that the traffic state was known, and instead using an estimate based on the local measurements from the connected vehicles. The connected vehicles were used both as sensors, to detect the congestion and stop-and-go waves, and as actuators, to restrict the traffic flow and dissipate the stop-and-go waves. We proposed some simple traffic state reconstruction algorithms, together with algorithms for learning components of the traffic dynamics, such as the flux function and the wave speed bound which governs the propagation speed of the stop-and-go waves. In order to demonstrate the effectiveness of the proposed traffic state reconstruction and control under realistic settings, we included both measurement and process noise in the simulations.

## 9.2 Future research directions

There is a plethora of ways the work presented in this thesis can be continued and expanded. First and foremost, a more thorough validation of the proposed models and control laws is needed. We have conducted preliminary validation of the multi-class cell transmission model and tandem queueing model with moving bottlenecks upstream of a stationary bottleneck, using microscopic traffic simulations done in SUMO to represent real traffic. The model extensions were mostly derived based on theoretical properties of the base models, with few additional assumptions. However, a thorough validation of all aspects of the models and the control laws needs to be conducted using suitable microscopic traffic simulations. For example, the dynamics of the mechanism of mutual influence between the directly controlled vehicles and the rest of the traffic should be examined in depth. We employed different macroscopic models for simulation, and for prediction and control implementation. However, using a microscopic traffic simulator, and applying traffic control calculated using the macroscopic traffic models will serve to both validate, and reinforce the general applicability of all proposed control laws. Furthermore, additional implementation issues that stem from a microscopic simulation framework will surely need to be dealt with, and other phenomena that are not apparent from the macroscopic models might need to be explained.

Although some of the discussed models can be readily extended to more complex traffic network configurations consisting of multiple road links, these general network structures were not explicitly considered in any of the considered control problems. The focus of this thesis was mostly on phenomena that happen on a section of the road, possibly upstream of a stationary bottleneck. However, these

sections can be used as building blocks for more general traffic networks, and similar control approaches can be generalized to this case. Scalability is an issue that is bound to arise when transferring these control laws to more general settings, and it will have to be addressed, potentially by employing approximate and multi-scale adaptations of the control laws. Furthermore, while the goal of this thesis was to present the Lagrangian traffic control paradigm, focusing only on the sensing and actuation done by connected vehicles, it is indisputable that best performance would be achieved by a combination of Lagrangian and Eulerian approaches. Moving bottleneck control, which was the centrepiece of this thesis, should be combined with other available types of traffic control, such as variable speed limits, ramp metering and route guidance. The local traffic measurements from the probe vehicles should be combined with the available traffic measurements coming from fixed sensors, such as inductive loops.

There are also many theoretical aspects of the considered control problems that should be further analysed. Analysis of the closed-loop systems, particularly focusing on stability and performance bounds of the proposed control laws, is needed. Indications about the required level of availability of connected vehicles to be used for control should also be derived. Many aspects of the control, such as pro-active control in case of stochastic traffic breakdowns, or active control of probe vehicles in order to improve the traffic state reconstruction quality, have not yet been fully explored. Furthermore, treating the traffic from fluid dynamics perspective, described by deterministic PDE models, disregards the inherent stochasticity of human driving behaviour. Therefore, robustness of any proposed control law needs to be examined and ensured, in order to ensure they would remain effective in real-world applications.

Last but not least, the traffic control approaches proposed here should also be implemented and tested in real-world experiments. Indeed most of the required additional equipment on the connected vehicles could to some degree be emulated by a simple mobile app that would communicate reference speeds to the drivers. The drivers could then be compensated, monetarily or with services, for following the assigned reference speeds, as well as for sharing their trajectories and potential local traffic measurements. The Lagrangian traffic control paradigm has so far been underutilized, and has the potential to greatly broaden the extent of control the authorities have over the traffic, leading to more efficient utilization of the available infrastructure, and a better future with less traffic congestion.

# Bibliography

- [1] European Commission, “Roadmap to a Single European Transport Area—Towards a competitive and resource efficient transport system,” *White Paper, Publications Office of the European Union*, 2011, COM(2011) 144 final.
- [2] M. Barth and K. Boriboonsomsin, “Real-world carbon dioxide impacts of traffic congestion,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2058, pp. 163–171, 2008.
- [3] European Commission, “EU transport in figures: Statistical pocketbook,” *Publications Office of the European Union*, 2019.
- [4] United States Environmental Protection Agency, “Fast facts. u.s. transportation sector greenhouse gas emissions 1990-2018,” *Office of Transportation and Air Quality*, 2020.
- [5] European Commission, “Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. The European Green Deal,” *Publications Office of the European Union*, 2019, COM/2019/640 final.
- [6] D. J. Fagnant and K. Kockelman, “Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations,” *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 167–181, 2015.
- [7] J. Walker, “The self-driving car timeline – predictions from the top 11 global automakers,” *Tech Emergence*, 2018.
- [8] P. Bansal and K. M. Kockelman, “Forecasting Americans’ long-term adoption of connected and autonomous vehicle technologies,” *Transportation Research Part A: Policy and Practice*, vol. 95, pp. 49–63, 2017.
- [9] T. Litman, *Autonomous vehicle implementation predictions*. Victoria Transport Policy Institute, 2017.

- [10] K. Dresner and P. Stone, “Multiagent traffic management: A reservation-based intersection control mechanism,” in *Third International Joint Conference on Autonomous Agents and Multiagent Systems*, vol. 3. IEEE Computer Society, 2004, pp. 530–537.
- [11] S. Shladover, D. Su, and X.-Y. Lu, “Impacts of cooperative adaptive cruise control on freeway traffic flow,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2324, pp. 63–70, 2012.
- [12] A. Talebpour and H. S. Mahmassani, “Influence of connected and autonomous vehicles on traffic flow stability and throughput,” *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 143–163, 2016.
- [13] M. Alonso Raposo, B. Ciuffo, M. Makridis, and C. Thiel, “The r-evolution of driving: from connected vehicles to coordinated automated road transport (C-ART). Part I: Framework for a safe & efficient coordinated automated road transport (C-ART) system,” *Publications Office of the European Union*.
- [14] Z. Wadud, D. MacKenzie, and P. Leiby, “Help or hindrance? the travel, energy and carbon impacts of highly automated vehicles,” *Transportation Research Part A: Policy and Practice*, vol. 86, pp. 1–18, 2016.
- [15] F. Dressler, H. Hartenstein, O. Altintas, and O. Tonguz, “Inter-vehicle communication: Quo vadis,” *IEEE Communications Magazine*, vol. 52, no. 6, pp. 170–177, 2014.
- [16] J. C. Herrera, D. B. Work, R. Herring, X. J. Ban, Q. Jacobson, and A. M. Bayen, “Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment,” *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568–583, 2010.
- [17] J. Macfarlane, “When apps rule the road: The proliferation of navigation apps is causing traffic chaos. it’s time to restore order,” *IEEE Spectrum*, vol. 56, no. 10, pp. 22–27, 2019.
- [18] C. Bergenheim, S. Shladover, E. Coelingh, C. Englund, and S. Tsugawa, “Overview of platooning systems,” in *Proceedings of the 19th ITS World Congress*, Vienna, Austria, 2012.
- [19] A. Alam, B. Besselink, V. Turri, J. Martensson, and K. H. Johansson, “Heavy-duty vehicle platooning for sustainable freight transportation: A cooperative method to enhance safety and efficiency,” *IEEE Control Systems*, vol. 35, no. 6, pp. 34–56, 2015.
- [20] L. D. Baskar, B. De Schutter, and H. Hellendoorn, “Optimal routing for automated highway systems,” *Transportation Research Part C: Emerging Technologies*, vol. 30, pp. 1–22, 2013.

- [21] R. Nishi, A. Tomoeda, K. Shimura, and K. Nishinari, "Theory of jam-absorption driving," *Transportation Research Part B: Methodological*, vol. 50, pp. 116–129, 2013.
- [22] Z. He, L. Zheng, L. Song, and N. Zhu, "A jam-absorption driving strategy for mitigating traffic oscillations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 802–813, 2017.
- [23] S. Shellenbarger, "One driver can prevent a traffic jam," *Wall Street Journal*, 2016.
- [24] M. Čičić and K. H. Johansson, "Traffic regulation via individually controlled automated vehicles: a cell transmission model approach," in *21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, US, 2018.
- [25] M. Čičić and K. H. Johansson, "Energy-optimal platoon catch-up in traffic in moving bottleneck framework," in *18th European Control Conference (ECC)*, Napoli, Italy, 2019, pp. 3674–3679.
- [26] M. Čičić and K. H. Johansson, "Stop-and-go wave dissipation using accumulated controlled moving bottlenecks in multi-class CTM framework," in *IEEE 58th Conference on Decision and Control (CDC)*, Nice, France, 2019, pp. 3146–3151.
- [27] M. Čičić, X. Xiong, L. Jin, and K. H. Johansson, "Coordinating vehicle platoons for highway bottleneck decongestion and throughput improvement," *IEEE Transactions on Intelligent Transportation Systems*, 2021, accepted, also presented at Transportation Research Board 99th Annual Meeting.
- [28] M. Čičić, M. Barreau, and K. H. Johansson, "Numerical investigation of traffic state reconstruction and control using connected automated vehicles," in *IEEE 23rd Intelligent Transportation Systems Conference (ITSC)*, Rhodes, Greece, 2020, pp. 1–6.
- [29] M. Čičić and K. H. Johansson, "Front tracking transition system model for traffic state reconstruction and control," in preparation.
- [30] M. Čičić, K.-Y. Liang, and K. H. Johansson, "Platoon merging distance prediction using a neural network vehicle speed model," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 3720–3725, 2017.
- [31] M. Čičić, I. Mikolášek, and K. H. Johansson, "Front tracking transition system model with controlled moving bottlenecks and probabilistic traffic breakdowns," in *The 21st IFAC World Congress*, 2020.

- [32] L. Jin, M. Čičić, S. Amin, and K. H. Johansson, “Modeling the impact of vehicle platooning on highway congestion: A fluid queuing approach,” in *Proceedings of the 21st International Conference on Hybrid Systems: Computation and Control (part of CPS Week)*. ACM, 2018, pp. 237–246.
- [33] G. Piacentini, M. Čičić, A. Ferrara, and K. H. Johansson, “VACS equipped vehicles for congestion dissipation in multi-class CTM framework,” in *18th European Control Conference (ECC)*, Napoli, Italy, 2019, pp. 2203–2208.
- [34] A. Ibrahim, M. Čičić, D. Goswami, T. Basten, and K. H. Johansson, “Control of platooned vehicles in presence of traffic shock waves,” in *IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, 2019, pp. 1727–1734.
- [35] L. Jin, M. Čičić, K. H. Johansson, and S. Amin, “Analysis and design of vehicle platooning operations on mixed-traffic highways,” *IEEE Transactions on Automatic Control*, 2020.
- [36] J. Liu, M. Barreau, M. Čičić, and K. H. Johansson, “Learning-based traffic state reconstruction using probe vehicles,” in *16th IFAC Symposium on Control in Transportation Systems*, Lille, France, 2021, submitted.
- [37] A. Srivastava and N. Geroliminis, “Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model,” *Transportation Research Part C: Emerging Technologies*, vol. 30, pp. 161–177, 2013.
- [38] M. Kontorinaki, A. Spiliopoulou, C. Roncoli, and M. Papageorgiou, “First-order traffic flow models incorporating capacity drop: Overview and real-data validation,” *Transportation Research Part B: Methodological*, vol. 106, pp. 52–75, 2017.
- [39] M. Saberli and H. S. Mahmassani, “Hysteresis and capacity drop phenomena in freeway networks: Empirical characterization and interpretation,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2391, no. 1, pp. 44–55, 2013.
- [40] W. Brilon, J. Geistefeldt, and M. Regler, “Reliability of freeway traffic flow: a stochastic concept of capacity,” in *16th International symposium on transportation and traffic theory*. College Park, Maryland, 2005, pp. 125–143.
- [41] B. S. Kerner, “Three-phase traffic theory and highway capacity,” *Physica A: Statistical Mechanics and its Applications*, vol. 333, pp. 379–440, 2004.
- [42] F. L. Hall and K. Agyemang-Duah, “Freeway capacity drop and the definition of capacity,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 1320, 1991.

- [43] B. S. Kerner and H. Rehborn, "Experimental properties of phase transitions in traffic flow," *Physical Review Letters*, vol. 79, no. 20, p. 4030, 1997.
- [44] M. Schönhof and D. Helbing, "Empirical features of congested traffic states and their implications for traffic modeling," *Transportation Science*, vol. 41, no. 2, pp. 135–166, 2007.
- [45] B. S. Kerner, "Experimental features of the emergence of moving jams in free traffic flow," *Journal of Physics A: Mathematical and general*, vol. 33, no. 26, p. L221, 2000.
- [46] Y. Sugiyama, M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S.-i. Tadaki, and S. Yukawa, "Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam," *New journal of physics*, vol. 10, no. 3, p. 033001, 2008.
- [47] M. R. Flynn, A. R. Kasimov, J.-C. Nave, R. R. Rosales, and B. Seibold, "Self-sustained nonlinear waves in traffic flow," *Physical Review E*, vol. 79, no. 5, p. 056113, 2009.
- [48] K. Yuan, V. L. Knoop, L. Leclercq, and S. P. Hoogendoorn, "Capacity drop: a comparison between stop-and-go wave and standing queue at lane-drop bottleneck," *Transportmetrica B: transport dynamics*, vol. 5, no. 2, pp. 145–158, 2017.
- [49] A. de Palma and R. Lindsey, "Traffic congestion pricing methodologies and technologies," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1377–1399, 2011.
- [50] D. Ettema, J. Knockaert, and E. Verhoef, "Using incentives as traffic management tool: empirical results of the "peak avoidance" experiment," *Transportation Letters*, vol. 2, no. 1, pp. 39–51, 2010.
- [51] D. B. Lee Jr, L. A. Klein, and G. Camus, "Induced traffic and induced demand," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1659, no. 1, pp. 68–75, 1999.
- [52] T. Roughgarden, "On the severity of Braess's paradox: designing networks for selfish users is hard," *Journal of Computer and System Sciences*, vol. 72, no. 5, pp. 922–953, 2006.
- [53] J. Eliasson and L.-G. Mattsson, "Equity effects of congestion pricing: quantitative methodology and a case study for Stockholm," *Transportation Research Part A: Policy and Practice*, vol. 40, no. 7, pp. 602–620, 2006.
- [54] A. A. Kurzhanskiy and P. Varaiya, "Traffic management: An outlook," *Economics of Transportation*, vol. 4, no. 3, pp. 135 – 146, 2015.

- [55] A. Keimer, N. Laurent-Brouty, F. Farokhi, H. Signargout, V. Cvetković, A. M. Bayen, and K. H. Johansson, "Information patterns in the modeling and design of mobility management services," *Proceedings of the IEEE*, vol. 106, no. 4, pp. 554–576, 2018.
- [56] M. Papageorgiou, H. Hadj-Salem, and J.-M. Blosseville, "Alinea: A local feedback control law for on-ramp metering," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1320, no. 1, pp. 58–67, 1991.
- [57] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: An overview," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 271–281, 2002.
- [58] G. Gomes and R. Horowitz, "Optimal freeway ramp metering using the asymmetric cell transmission model," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 4, pp. 244–262, 2006.
- [59] L. D. Baskar, B. De Schutter, J. Hellendoorn, and Z. Papp, "Traffic control and intelligent vehicle highway systems: a survey," *IET Intelligent Transport Systems*, vol. 5, no. 1, pp. 38–52, 2011.
- [60] 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 Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2024–2039, 2014.
- [61] M. Papageorgiou, E. Kosmatopoulos, and I. Papamichail, "Effects of variable speed limits on motorway traffic flow," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2047, pp. 37–48, 2008.
- [62] M. Hadiuzzaman, T. Z. Qiu, and X.-Y. Lu, "Variable speed limit control design for relieving congestion caused by active bottlenecks," *Journal of Transportation Engineering*, vol. 139, no. 4, pp. 358–370, 2012.
- [63] 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," *Journal of Intelligent Transportation Systems*, vol. 20, no. 6, pp. 559–572, 2016.
- [64] Y. Zhang and P. A. Ioannou, "Combined variable speed limit and lane change control for highway traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1812–1823, 2017.
- [65] A. Talebpour, H. S. Mahmassani, and S. H. Hamdar, "Speed harmonization: Evaluation of effectiveness under congested conditions," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2391, no. 1, pp. 69–79, 2013.

- [66] J. Ma, X. Li, S. Shladover, H. A. Rakha, X.-Y. Lu, R. Jagannathan, and D. J. Dailey, "Freeway speed harmonization," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 78–89, 2016.
- [67] A. Hegyi, S. Hoogendoorn, M. Schreuder, H. Stoelhorst, and F. Viti, "SPECIALIST: A dynamic speed limit control algorithm based on shock wave theory," in *11th IEEE International Conference on Intelligent Transportation Systems*, 2008, pp. 827–832.
- [68] R. C. Carlson, I. Papamichail, M. Papageorgiou, and A. Messmer, "Optimal motorway traffic flow control involving variable speed limits and ramp metering," *Transportation Science*, vol. 44, no. 2, pp. 238–253, 2010.
- [69] A. Hegyi, B. De Schutter, and H. Hellendoorn, "Model predictive control for optimal coordination of ramp metering and variable speed limits," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 185–209, 2005.
- [70] X.-Y. Lu, P. Varaiya, R. Horowitz, D. Su, and S. Shladover, "Novel freeway traffic control with variable speed limit and coordinated ramp metering," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2229, pp. 55–65, 2011.
- [71] M. Ben-Akiva, J. Bottom, and M. S. Ramming, "Route guidance and information systems," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 215, no. 4, pp. 317–324, 2001.
- [72] E. Schmitt and H. Jula, "Vehicle route guidance systems: Classification and comparison," in *IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2006, pp. 242–247.
- [73] L.-E. Y. Mimbela, L. A. Klein *et al.*, "Summary of vehicle detection and surveillance technologies used in intelligent transportation systems," *Federal Highway Administration's (FHWA) Intelligent Transportation Systems Program Office*, 2007.
- [74] A. Festag, "Cooperative intelligent transport systems standards in europe," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 166–172, 2014.
- [75] K. Sjoberg, P. Andres, T. Buburuzan, and A. Brakemeier, "Cooperative intelligent transport systems in europe: current deployment status and outlook," *IEEE Vehicular Technology Magazine*, vol. 12, no. 2, pp. 89–97, 2017.
- [76] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended kalman filter: a general approach," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 141–167, 2005.

- [77] S. Blandin, A. Couque, A. Bayen, and D. Work, “On sequential data assimilation for scalar macroscopic traffic flow models,” *Physica D: Nonlinear Phenomena*, vol. 241, no. 17, pp. 1421–1440, 2012.
- [78] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, “Highway traffic state estimation with mixed connected and conventional vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3484–3497, 2016.
- [79] S. E. Shladover, “Connected and automated vehicle systems: Introduction and overview,” *Journal of Intelligent Transportation Systems*, vol. 22, no. 3, pp. 190–200, 2018.
- [80] A. Vinel, L. Lan, and N. Lyamin, “Vehicle-to-vehicle communication in C-ACC/platooning scenarios,” *IEEE Communications Magazine*, vol. 53, no. 8, pp. 192–197, 2015.
- [81] V. L. Knoop, P. B. van Erp, L. Leclercq, and S. P. Hoogendoorn, “Empirical MFDs using Google traffic data,” in *21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 3832–3839.
- [82] B. Sana, J. Castiglione, D. Cooper, and D. Tischler, “Using Google’s aggregated and anonymized trip data to support freeway corridor management planning in San Francisco, California,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2643, no. 1, pp. 65–73, 2017.
- [83] C. De Fabritiis, R. Ragona, and G. Valenti, “Traffic estimation and prediction based on real time floating car data,” in *11th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2008, pp. 197–203.
- [84] D. Work, S. Blandin, O.-P. Tossavainen, B. Piccoli, and A. Bayen, “A distributed highway velocity model for traffic state reconstruction,” *Applied Research Mathematics eXpress (ARMX)*, vol. 1, pp. 1–35, 2010.
- [85] E. Cristiani, C. de Fabritiis, and B. Piccoli, “A fluid dynamic approach for traffic forecast from mobile sensor data,” *Communications in Applied and Industrial Mathematics*, vol. 1, no. 1, pp. 54–71, 2010.
- [86] R. Arnott, A. De Palma, and R. Lindsey, “Does providing information to drivers reduce traffic congestion?” *Transportation Research Part A: General*, vol. 25, no. 5, pp. 309–318, 1991.
- [87] T. Cabannes, F. Shyu, E. Porter, S. Yao, Y. Wang, M. A. S. Vincentelli, S. Hinardi, M. Zhao, and A. M. Bayen, “Measuring regret in routing: assessing the impact of increased app usage,” in *21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2589–2594.

- [88] S. Chen, H. Yu, and M. Krstić, “Regulator design for a congested continuum traffic model with app-routing instability,” in *American Control Conference (ACC)*. IEEE, 2020, pp. 4515–4520.
- [89] V. Milanés, S. E. Shladover, J. Spring, C. Nowakowski, H. Kawazoe, and M. Nakamura, “Cooperative adaptive cruise control in real traffic situations.” *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 296–305, 2014.
- [90] T. Kim and K. Jerath, “Mitigation of self-organized traffic jams using cooperative adaptive cruise control,” in *International Conference on Connected Vehicles and Expo (ICCVE)*. IEEE, 2016, pp. 7–12.
- [91] A. Askari, D. A. Farias, A. A. Kurzhanskiy, and P. Varaiya, “Effect of adaptive and cooperative adaptive cruise control on throughput of signalized arterials,” in *Intelligent Vehicles Symposium (IV)*. IEEE, 2017, pp. 1287–1292.
- [92] R. E. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, R. Haulcy, H. Pohlmann, F. Wu, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work, “Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments,” *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 205 – 221, 2018.
- [93] S. Eilers, J. Mårtensson, H. Pettersson, M. Pillado, D. Gallegos, M. Tobar, K. H. Johansson, X. Ma, T. Friedrichs, S. S. Borojeni *et al.*, “Companion-towards co-operative platoon management of heavy-duty vehicles,” in *IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, 2015, pp. 1267–1273.
- [94] S. Tsugawa, S. Jeschke, and S. E. Shladover, “A review of truck platooning projects for energy savings,” *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 68–77, 2016.
- [95] C. Bonnet and H. Fritz, “Fuel consumption reduction in a platoon: Experimental results with two electronically coupled trucks at close spacing,” 2000.
- [96] H. L. Humphreys, J. Batterson, D. Bevely, and R. Schubert, “An evaluation of the fuel economy benefits of a driver assistive truck platooning prototype using simulation,” SAE Technical Paper, Tech. Rep., 2016.
- [97] A. Chottani, G. Hastings, J. Murnane, and F. Neuhaus, “Distraction or disruption? autonomous trucks gain ground in us logistics,” *McKinsey Smith Co.*, 2018.
- [98] J. Lioris, R. Pedarsani, F. Y. Tascikaraoglu, and P. Varaiya, “Platoons of connected vehicles can double throughput in urban roads,” *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 292–305, 2017.

- [99] S. Moridpour, E. Mazloumi, and M. Mesbah, “Impact of heavy vehicles on surrounding traffic characteristics,” *Journal of advanced transportation*, vol. 49, no. 4, pp. 535–552, 2015.
- [100] R. Horowitz and P. Varaiya, “Control design of an automated highway system,” *Proceedings of the IEEE*, vol. 88, no. 7, pp. 913–925, 2000.
- [101] M. E. Khatir and E. J. Davison, “Decentralized control of a large platoon of vehicles using non-identical controllers,” in *American Control Conference (ACC)*, vol. 3, Portland, US, 2004, pp. 2769–2776.
- [102] V. Turri, B. Besselink, J. Mårtensson, and K. H. Johansson, “Fuel-efficient heavy-duty vehicle platooning by look-ahead control,” in *53rd IEEE Conference on Decision and Control (CDC)*, Los Angeles, US, 2014, pp. 654–660.
- [103] J. Ploeg, D. P. Shukla, N. van de Wouw, and H. Nijmeijer, “Controller synthesis for string stability of vehicle platoons,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 854–865, 2014.
- [104] A. Alam, A. Gattami, and K. H. Johansson, “An experimental study on the fuel reduction potential of heavy duty vehicle platooning,” in *13th IEEE International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2010, pp. 306–311.
- [105] L. Aarts and G. Feddes, “European truck platooning challenge,” in *HVTT14: International Symposium on Heavy Vehicle Transport Technology*, Rotorua, New Zealand, 2016.
- [106] R. Bishop, D. Bevly, L. Humphreys, S. Boyd, and D. Murray, “Evaluation and testing of driver-assistive truck platooning: phase 2 final results,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2615, no. 1, pp. 11–18, 2017.
- [107] A. K. Bhoopalam, N. Agatz, and R. Zuidwijk, “Planning of truck platoons: A literature review and directions for future research,” *Transportation Research Part B: Methodological*, vol. 107, pp. 212–228, 2018.
- [108] M. Wang, S. van Maarseveen, R. Happee, O. Tool, and B. van Arem, “Benefits and risks of truck platooning on freeway operations near entrance ramp,” *Transportation Research Record: Journal of the Transportation Research Board*, 2019.
- [109] A. Duret, M. Wang, and A. Ladino, “A hierarchical approach for splitting truck platoons near network discontinuities,” *Transportation Research Part B: Methodological*, 2019.

- [110] K.-Y. Liang, J. Mårtensson, and K. H. Johansson, “Fuel-saving potentials of platooning evaluated through sparse heavy-duty vehicle position data,” in *Intelligent Vehicles Symposium (IV)*. IEEE, 2014, pp. 1061–1068.
- [111] S. van de Hoef, K. H. Johansson, and D. V. Dimarogonas, “Fuel-efficient en route formation of truck platoons,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 102–112, 2018.
- [112] K.-Y. Liang, J. Mårtensson, and K. H. Johansson, “When is it fuel efficient for a heavy duty vehicle to catch up with a platoon?” *IFAC Proceedings Volumes*, vol. 46, no. 21, pp. 738–743, 2013.
- [113] K.-Y. Liang, J. Mårtensson, and K. H. Johansson, “Experiments on platoon formation of heavy trucks in traffic,” in *19th IEEE International Conference on Intelligent Transportation Systems (ITSC)*. Los Angeles, US: IEEE, 2016, pp. 1813–1819.
- [114] M. Saeednia and M. Menendez, “Analysis of strategies for truck platooning: Hybrid strategy,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2547, pp. 41–48, 2016.
- [115] W. Zhang, E. Jenelius, and X. Ma, “Freight transport platoon coordination and departure time scheduling under travel time uncertainty,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 98, pp. 1–23, 2017.
- [116] K.-Y. Liang, Q. Deng, J. Mårtensson, X. Ma, and K. H. Johansson, “The influence of traffic on heavy-duty vehicle platoon formation,” in *Intelligent Vehicles Symposium (IV)*. Seoul, Korea: IEEE, 2015, pp. 150–155.
- [117] R. Larsen, J. Rich, and T. K. Rasmussen, “Hub-based truck platooning: Potentials and profitability,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 127, pp. 249–264, 2019.
- [118] A. Johansson, V. Turri, E. Nekouei, K. H. Johansson, and J. Mårtensson, “Truck platoon formation at hubs: An optimal release time rule,” in *The 21st IFAC World Congress*, 2020.
- [119] J. Axelsson, “Business models and roles for mediating services in a truck platooning system-of-systems.” in *14th Annual Conference System of Systems Engineering, SoSE*, 2019, pp. 113–118.
- [120] A. Johansson and J. Mårtensson, “Game theoretic models for profit-sharing in multi-fleet platoons,” in *IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 3019–3024.
- [121] J. Axelsson, T. Bergh, A. Johansson, B. Mårdberg, P. Svenson, and V. Åkesson, “Truck platooning business case analysis,” *RISE Report*.

- [122] A. Kotsialos and M. Papageorgiou, "The importance of traffic flow modeling for motorway traffic control," *Networks and Spatial Economics*, vol. 1, no. 1-2, pp. 179–203, 2001.
- [123] C. F. Daganzo, "Requiem for second-order fluid approximations of traffic flow," *Transportation Research Part B: Methodological*, vol. 29, no. 4, pp. 277–286, 1995.
- [124] A. Aw and M. Rascle, "Resurrection of "second order" models of traffic flow," *SIAM journal on applied mathematics*, vol. 60, no. 3, pp. 916–938, 2000.
- [125] M. Treiber and A. Kesting, *Traffic Flow Dynamics: Data, Models and Simulation*. Springer, 2013.
- [126] S. P. Hoogendoorn and P. H. Bovy, "State-of-the-art of vehicular traffic flow modelling," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 215, no. 4, pp. 283–303, 2001.
- [127] A. Ferrara, S. Sacone, and S. Siri, *Freeway Traffic Modelling and Control*. Springer, 2018.
- [128] F. van Wageningen-Kessels, H. Van Lint, K. Vuik, and S. Hoogendoorn, "Genealogy of traffic flow models," *EURO Journal on Transportation and Logistics*, vol. 4, no. 4, pp. 445–473, 2015.
- [129] S. C. Dafermos, "The traffic assignment problem for multiclass-user transportation networks," *Transportation science*, vol. 6, no. 1, pp. 73–87, 1972.
- [130] G. Wong and S. Wong, "A multi-class traffic flow model—an extension of lwr model with heterogeneous drivers," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 9, pp. 827–841, 2002.
- [131] M. Saifuzzaman and Z. Zheng, "Incorporating human-factors in car-following models: a review of recent developments and research needs," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 379–403, 2014.
- [132] D. C. Gazis, R. Herman, and R. W. Rothery, "Nonlinear follow-the-leader models of traffic flow," *Operations research*, vol. 9, no. 4, pp. 545–567, 1961.
- [133] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation Research Part B: Methodological*, vol. 15, no. 2, pp. 105–111, 1981.
- [134] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000.

- [135] A. Kesting, M. Treiber, and D. Helbing, “Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity,” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1928, pp. 4585–4605, 2010.
- [136] F. Wu and D. B. Work, “Connections between classical car following models and artificial neural networks,” in *21st IEEE International Conference on Intelligent Transportation Systems (ITSC)*, Maui, US, 2018.
- [137] J. Barceló and J. Casas, “Dynamic network simulation with AIMSUN,” in *Simulation approaches in transportation analysis*. Springer, pp. 57–98.
- [138] M. Fellendorf, “Vissim: A microscopic simulation tool to evaluate actuated signal control including bus priority,” in *64th Institute of Transportation Engineers Annual Meeting*. Springer, 1994, pp. 1–9.
- [139] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, “Sumo—simulation of urban mobility,” in *The Third International Conference on Advances in System Simulation (SIMUL), Barcelona, Spain*, vol. 42, 2011.
- [140] S. Ossen and S. P. Hoogendoorn, “Heterogeneity in car-following behavior: Theory and empirics,” *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 2, pp. 182–195, 2011.
- [141] S. Benzoni-Gavage and R. M. Colombo, “An  $n$ -populations model for traffic flow,” *European Journal of Applied Mathematics*, vol. 14, no. 5, pp. 587–612, 2003.
- [142] S. P. Hoogendoorn and P. H. Bovy, “Continuum modeling of multiclass traffic flow,” *Transportation Research Part B: Methodological*, vol. 34, no. 2, pp. 123–146, 2000.
- [143] M. W. Levin and S. D. Boyles, “A multiclass cell transmission model for shared human and autonomous vehicle roads,” *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 103–116, 2016.
- [144] S. Logghe and L. H. Immers, “Multi-class kinematic wave theory of traffic flow,” *Transportation Research Part B: Methodological*, vol. 42, no. 6, pp. 523–541, 2008.
- [145] J. Van Lint, S. Hoogendoorn, and M. Schreuder, “Fastlane: New multiclass first-order traffic flow model,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2088, pp. 177–187, 2008.
- [146] S. Fan and D. B. Work, “A heterogeneous multiclass traffic flow model with creeping,” *SIAM Journal on Applied Mathematics*, vol. 75, no. 2, pp. 813–835, 2015.

- [147] C. Caligaris, S. Sacone, and S. Siri, "Model predictive control for multiclass freeway traffic," in *European Control Conference (ECC)*. IEEE, 2009, pp. 1764–1769.
- [148] M. Burkhardt, H. Yu, and M. Krstić, "Stop-and-go suppression in two-class congested traffic," *Automatica*, vol. 125, 2021.
- [149] P. Bagnerini and M. Rascle, "A multiclass homogenized hyperbolic model of traffic flow," *SIAM journal on mathematical analysis*, vol. 35, no. 4, pp. 949–973, 2003.
- [150] M. Wright, G. Gomes, R. Horowitz, and A. A. Kurzhanskiy, "A new model for multi-commodity macroscopic modeling of complex traffic networks," *arXiv preprint arXiv:1509.04995*, 2015.
- [151] J.-P. Lebacque and M. M. Khoshyaran, "A variational formulation for higher order macroscopic traffic flow models of the gsom family," *Procedia-Social and Behavioral Sciences*, vol. 80, pp. 370–394, 2013.
- [152] S. Fan and B. Seibold, "Data-fitted first-order traffic models and their second-order generalizations: Comparison by trajectory and sensor data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2391, no. 1, pp. 32–43, 2013.
- [153] S. Fan, M. Herty, and B. Seibold, "Comparative model accuracy of a data-fitted generalized aw-rascle-zhang model," *Networks and Heterogeneous Media*, vol. 9, no. 2, pp. 239–268, 2014.
- [154] A. Khelifi, H. Haj-Salem, J.-P. Lebacque, and L. Nabli, "Lagrangian generic second order traffic flow models for node," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 5, no. 1, pp. 14–27, 2018.
- [155] R. Wang, Y. Li, and D. B. Work, "Comparing traffic state estimators for mixed human and automated traffic flows," *Transportation Research Part C: Emerging Technologies*, vol. 78, pp. 95–110, 2017.
- [156] J. C. Muñoz and C. F. Daganzo, "Moving bottlenecks: a theory grounded on experimental observation," *15th International Symposium on Transportation and Traffic Theory*, pp. 441–462, 2002.
- [157] K. Fadhloun, H. Rakha, and A. Loulizi, "Analysis of moving bottlenecks considering a triangular fundamental diagram," *International journal of transportation science and technology*, vol. 5, no. 3, pp. 186–199, 2016.
- [158] B. S. Kerner and S. L. Klenov, "A theory of traffic congestion at moving bottlenecks," *Journal of Physics A: Mathematical and Theoretical*, vol. 43, no. 42, p. 425101, 2010.

- [159] J.-P. Lebacque, J. Lesort, and F. Giorgi, “Introducing buses into first-order macroscopic traffic flow models,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 1644, pp. 70–79, 1998.
- [160] C. Lattanzio, A. Maurizi, and B. Piccoli, “Moving bottlenecks in car traffic flow: a pde-ode coupled model,” *SIAM Journal on Mathematical Analysis*, vol. 43, no. 1, pp. 50–67, 2011.
- [161] M. L. Delle Monache and P. Goatin, “A front tracking method for a strongly coupled PDE-ODE system with moving density constraints in traffic flow,” *Discrete and Continuous Dynamical Systems-Series S*, vol. 7, no. 3, pp. 435–447, 2014.
- [162] M. D. Simoni and C. G. Claudel, “A fast simulation algorithm for multiple moving bottlenecks and applications in urban freight traffic management,” *Transportation Research Part B: Methodological*, vol. 104, pp. 238–255, 2017.
- [163] G. Piacentini, P. Goatin, and A. Ferrara, “A macroscopic model for platooning in highway traffic,” *SIAM Journal on Applied Mathematics*, vol. 80, no. 1, pp. 639–656, 2020.
- [164] W. J. Schakel and B. Van Arem, “Improving traffic flow efficiency by in-car advice on lane, speed, and headway,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1597–1606, 2014.
- [165] P. Ioannou, Y. Wang, and H. Chang, “Integrated roadway/adaptive cruise control system: Safety, performance, environmental and near term deployment considerations,” *California PATH Program, Institute of Transportation Studies, University of California at Berkeley*, 2007.
- [166] B. Hellinga and M. Mandelzys, “Impact of driver compliance on the safety and operational impacts of freeway variable speed limit systems,” *Journal of Transportation Engineering*, vol. 137, no. 4, pp. 260–268, 2011.
- [167] I. Karafyllis, N. Bekiaris-Liberis, and M. Papageorgiou, “Analysis and control of a non-standard hyperbolic pde traffic flow model,” *arXiv preprint arXiv:1707.02209*, 2017.
- [168] L. Zhang, C. Prieur, and J. Qiao, “PI boundary control of linear hyperbolic balance laws with stabilization of ARZ traffic flow models,” *Systems & Control Letters*, vol. 123, pp. 85–91, 2019.
- [169] H. Yu and M. Krstic, “Traffic congestion control for Aw–Rascle–Zhang model,” *Automatica*, vol. 100, pp. 38–51, 2019.
- [170] B. Besselink and K. H. Johansson, “String stability and a delay-based spacing policy for vehicle platoons subject to disturbances,” *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4376–4391, 2017.

- [171] C. Wu, A. M. Bayen, and A. Mehta, “Stabilizing traffic with autonomous vehicles,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 1–7.
- [172] Y. Zhou, S. Ahn, M. Wang, and S. Hoogendoorn, “Stabilizing mixed vehicular platoons with connected automated vehicles: An h-infinity approach,” *Transportation Research Procedia*, vol. 38, pp. 441–461, 2019.
- [173] S. Cui, B. Seibold, R. Stern, and D. B. Work, “Stabilizing traffic flow via a single autonomous vehicle: Possibilities and limitations,” in *Intelligent Vehicles Symposium (IV)*. IEEE, 2017, pp. 1336–1341.
- [174] Y. Zheng, J. Wang, and K. Li, “Smoothing traffic flow via control of autonomous vehicles,” *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 3882–3896, 2020.
- [175] M. L. Delle Monache, T. Liard, A. Rat, R. Stern, R. Bhadani, B. Seibold, J. Sprinkle, D. B. Work, and B. Piccoli, “Feedback control algorithms for the dissipation of traffic waves with autonomous vehicles,” in *Computational Intelligence and Optimization Methods for Control Engineering*. Springer, 2019, pp. 275–299.
- [176] R. E. Stern, S. Cui, M. L. D. Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, R. Haulcy, H. Pohlmann, F. Wu, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work, “Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments,” *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 205 – 221, 2018.
- [177] E. Vinitzky, K. Parvatey, A. Kreidiehz, C. Wu, and A. Bayen, “Lagrangian control through deep-rl: Applications to bottleneck decongestion,” in *21st IEEE International Conference on Intelligent Transportation Systems (ITSC)*, Maui, US, 2018.
- [178] American Traffic Safety Services Association, “Guidelines on rolling roadblocks for work zone applications,” 2013.
- [179] R. A. Ramadan and B. Seibold, “Traffic flow control and fuel consumption reduction via moving bottlenecks,” in *Transportation Research Board 96th Annual Meeting*, 2017.
- [180] G. Piacentini, P. Goatin, and A. Ferrara, “Traffic control via moving bottleneck of coordinated vehicles,” in *15th IFAC symposium on control in transportation systems*, 2018.
- [181] G. Piacentini, C. Pasquale, S. Sacone, S. Siri, and A. Ferrara, “Multiple moving bottlenecks for traffic control in freeway systems,” in *18th European Control Conference (ECC)*. IEEE, 2019, pp. 3662–3667.

- [182] T. Liard, R. Stern, and M. L. Delle Monache, "Optimal driving strategies for traffic control with autonomous vehicles," in *The 21st IFAC World Congress*, 2020.
- [183] J. G. Wardrop and G. Charlesworth, "A method of estimating speed and flow of traffic from a moving vehicle." *Proceedings of the Institution of Civil Engineers*, vol. 3, no. 1, pp. 158–171, 1954.
- [184] C. Wright, "A theoretical analysis of the moving observer method," *Transportation Research*, vol. 7, no. 3, pp. 293–311, 1973.
- [185] R.-P. Schäfer, K.-U. Thiessenhusen, and P. Wagner, "A traffic information system by means of real-time floating-car data," in *ITS World Congress*, 2002.
- [186] A. D. Patire, M. Wright, B. Prodhomme, and A. M. Bayen, "How much gps data do we need?" *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 325–342, 2015.
- [187] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura, "Traffic state estimation on highway: A comprehensive survey," *Annual Reviews in Control*, vol. 43, pp. 128–151, 2017.
- [188] T. Seo and T. Kusakabe, "Probe vehicle-based traffic state estimation method with spacing information and conservation law," *Transportation Research Part C: Emerging Technologies*, vol. 59, pp. 391–403, 2015.
- [189] E. Lovisari, C. C. de Wit, and A. Y. Kibangou, "Density/flow reconstruction via heterogeneous sources and optimal sensor placement in road networks," *Transportation Research Part C: Emerging Technologies*, vol. 69, pp. 451–476, 2016.
- [190] M. Wright and R. Horowitz, "Fusing loop and gps probe measurements to estimate freeway density," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3577–3590, 2016.
- [191] E. S. Canepa and C. G. Claudel, "Networked traffic state estimation involving mixed fixed-mobile sensor data using hamilton-jacobi equations," *Transportation Research Part B: Methodological*, vol. 104, pp. 686–709, 2017.
- [192] M. L. Delle Monache, T. Liard, B. Piccoli, R. Stern, and D. Work, "Traffic reconstruction using autonomous vehicles," *SIAM Journal on Applied Mathematics*, vol. 79, no. 5, pp. 1748–1767, 2019.
- [193] M. Barreau, A. Selivanov, and K. H. Johansson, "Dynamic traffic reconstruction using probe vehicles," in *59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 233–238.

- [194] R. Billot, N.-E. El Faouzi, and F. De Vuyst, “Multilevel assessment of the impact of rain on drivers’ behavior: standardized methodology and empirical analysis,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2107, no. 1, pp. 134–142, 2009.
- [195] S. Moridpour, M. Sarvi, and G. Rose, “Lane changing models: a critical review,” *Transportation letters*, vol. 2, no. 3, pp. 157–173, 2010.
- [196] M. J. Lighthill and G. B. Whitham, “On kinematic waves II. a theory of traffic flow on long crowded roads,” *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 229, no. 1178, pp. 317–345, 1955.
- [197] P. I. Richards, “Shock waves on the highway,” *Operations research*, vol. 4, no. 1, pp. 42–51, 1956.
- [198] P.-E. Mazaré, A. H. Dehwah, C. G. Claudel, and A. M. Bayen, “Analytical and grid-free solutions to the Lighthill–Whitham–Richards traffic flow model,” *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1727–1748, 2011.
- [199] C. G. Claudel and A. M. Bayen, “Lax–hopf based incorporation of internal boundary conditions into hamilton–jacobi equation. part i: Theory,” *IEEE Transactions on Automatic Control*, vol. 55, no. 5, pp. 1142–1157, 2010.
- [200] M. Papageorgiou, J.-M. Blosseville, and H. Hadj-Salem, “Macroscopic modelling of traffic flow on the boulevard périphérique in paris,” *Transportation Research Part B: Methodological*, vol. 23, no. 1, pp. 29–47, 1989.
- [201] M. Carey and M. Bowers, “A review of properties of flow–density functions,” *Transport Reviews*, vol. 32, no. 1, pp. 49–73, 2012.
- [202] B. Greenshields, W. Channing, H. Miller *et al.*, “A study of traffic capacity,” in *Highway research board proceedings*, vol. 1935. National Research Council (USA), Highway Research Board, 1935.
- [203] C. F. Daganzo, “The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory,” *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.
- [204] G. F. Newell, “A simplified theory of kinematic waves in highway traffic, part i: General theory,” *Transportation Research Part B: Methodological*, vol. 27, no. 4, pp. 281–287, 1993.
- [205] P. D. Lax, “Nonlinear hyperbolic equations,” *Communications on Pure and Applied Mathematics*, vol. 6, no. 2, pp. 231–258, 1953.

- [206] H. Holden and N. H. Risebro, *Front tracking for hyperbolic conservation laws*. Springer, 2015, vol. 152.
- [207] M. Papageorgiou, J.-M. Blosseville, and H. Hadj-Salem, “Modelling and real-time control of traffic flow on the southern part of boulevard peripherique in paris: Part I: Modelling,” *Transportation Research Part A: General*, vol. 24, no. 5, pp. 345–359, 1990.
- [208] C. F. Daganzo, “The cell transmission model, part ii: network traffic,” *Transportation Research Part B: Methodological*, vol. 29, no. 2, pp. 79–93, 1995.
- [209] J.-P. Lebacque, “The Godunov scheme and what it means for first order traffic flow models,” *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pp. 647–677, 1996.
- [210] D. Heidemann and H. Wegmann, “Queueing at unsignalized intersections,” *Transportation Research Part B: Methodological*, vol. 31, no. 3, pp. 239–263, 1997.
- [211] T. Van Woensel and N. Vandaele, “Modeling traffic flows with queueing models: a review,” *Asia-Pacific Journal of Operational Research*, vol. 24, no. 04, pp. 435–461, 2007.
- [212] H. M. Zhang, “A non-equilibrium traffic model devoid of gas-like behavior,” *Transportation Research Part B: Methodological*, vol. 36, no. 3, pp. 275–290, 2002.
- [213] C. C. de Wit and A. Ferrara, “A variable-length cell road traffic model: application to ring road speed limit optimization,” in *55th IEEE Conference on Decision and Control (CDC)*, 2016.
- [214] Y. Han, Y. Yuan, A. Hegyi, and S. P. Hoogendoorn, “New extended discrete first-order model to reproduce propagation of jam waves,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2560, pp. 108–118, 2016.
- [215] Y. Lu, S. C. Wong, M. Zhang, and C.-W. Shu, “The entropy solutions for the lighthill-whitham-richards traffic flow model with a discontinuous flow-density relationship,” *Transportation Science*, vol. 43, no. 4, pp. 511–530, 2009.
- [216] P. Tabuada, *Verification and control of hybrid systems: a symbolic approach*. Springer Science & Business Media, 2009.
- [217] E. Hellström, M. Ivarsson, J. Åslund, and L. Nielsen, “Look-ahead control for heavy trucks to minimize trip time and fuel consumption,” *Control Engineering Practice*, vol. 17, no. 2, pp. 245–254, 2009.

- [218] V. Turri, B. Besselink, and K. H. Johansson, “Cooperative look-ahead control for fuel-efficient and safe heavy-duty vehicle platooning,” *IEEE Transactions on Control Systems Technology*, vol. 25, no. 1, pp. 12–28, 2017.
- [219] L. Guzzella, A. Sciarretta *et al.*, *Vehicle propulsion systems*. Springer, 2007, vol. 1.
- [220] B. S. Kerner, “Breakdown minimization principle versus wardrop’s equilibria for dynamic traffic assignment and control in traffic and transportation networks: A critical mini-review,” *Physica A: Statistical Mechanics and its Applications*, vol. 466, pp. 626–662, 2017.

