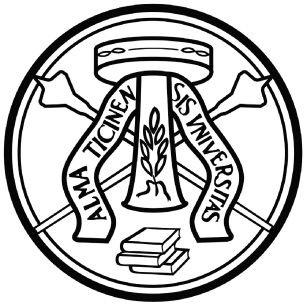


University of Pavia

Department of Electrical, Computer and Biomedical Engineering

Master's Degree in Industrial Automation Engineering

Multi-lane freeway traffic control with connected and automated vehicles

Supervisor:
Prof. Antonella Ferrara
Co-supervisor:
Dott. Nikolas Sacchi

Candidate:
Davide Raimondi

A.Y. 2024/2025

Abstract

Highway bottlenecks caused by traffic incidents frequently trigger a systemic breakdown of traffic flow, leading to the capacity drop phenomenon and severe congestion. To mitigate this inefficiency, this thesis proposes a Sequential Model Predictive Control (MPC) strategy that uses a fleet of Connected and Autonomous Vehicles (CAVs) to actively regulate upstream traffic flow. The physical plant is simulated using a macroscopic Cell Transmission Model (CTM) integrated with a customized incentive-based lateral flow logic. Operating over a finite prediction horizon, the MPC dynamically optimizes the speed of the CAV fleet to trying maintaining bottleneck density below the critical threshold, thereby preserving maximum discharge throughput.

To comprehensively evaluate the controller's robustness, a rigorous sensitivity analysis was conducted during a simulated 60-minute lane-closure incident. The experiments varied both the spatial control horizon, CAV activation distances from 5 to 20 km upstream, and the fleet sizes, ranging from 6 to 10 vehicles.

The simulation results demonstrate that the uncontrolled capacity drop imposes a systemic penalty of 363 vehicle-hours of delay. The optimal control configuration—deploying 10 CAVs 5 kilometers upstream of the incident—successfully recovered 94.7% of this delay, effectively neutralizing the capacity drop and forcing the network to operate almost at its theoretical physical limit. Furthermore, the analysis reveals a fundamental spatiotemporal trade-off in fleet deployment: activating control further upstream significantly increases the temporal footprint of each CAV, allowing smaller fleets to partially mitigate the incident. Conversely, proximity activation requires larger fleets to sustain temporal coverage but achieves higher absolute efficiency by minimizing initialization latency.

Finally, to validate the real-world feasibility of the proposed system, a probabilistic demand model and topological Monte Carlo simulation were conducted. The analysis proves that the Sequential MPC framework does not require near-total autonomous adoption; instead, it achieves over 99% spatial reliability for forming the moving bottleneck at a modest CAV market penetration rate of just 6.3%. These findings provide highly practical, deployment-ready strategies for future traffic management centers operating under realistic near-term conditions.

Sommario

I colli di bottiglia autostradali causati da incidenti stradali innescano frequentemente un collasso sistemico del flusso di traffico, portando al fenomeno del capacity drop e a gravi congestioni. Per mitigare questa inefficienza, la tesi propone una strategia di controllo predittivo basato sull'uso di un MPC sequenziale che utilizza una flotta di Veicoli Connessi e Autonomi per regolare attivamente il flusso di traffico a monte. L'impianto fisico è simulato utilizzando il Cell Transmission Model integrato con una logica di flusso laterale personalizzata basata su incentivi. Operando su un orizzonte di predizione finito, l'MPC ottimizza dinamicamente la velocità della flotta di CAV per cercare di mantenere la densità del collo di bottiglia al di sotto della soglia critica, preservando così la massima capacità di deflusso.

Per valutare in modo esaustivo la robustezza del controllore, è stata condotta un'analisi di sensibilità durante un incidente simulato di 60 minuti con chiusura di corsia. Gli esperimenti hanno variato sia l'orizzonte di controllo spaziale, distanze di attivazione dei CAV da 5 a 20 km dall'incidente, sia le dimensioni della flotta, da 6 a 10 formazioni.

I risultati della simulazione dimostrano che il capacity drop non controllato impone una penalità sistemica di 363 veicoli/ora di ritardo. La configurazione di controllo ottimale, dispiegando 10 CAV a 5 chilometri a monte dell'incidente, ha recuperato con successo il 94.7% di questo ritardo, neutralizzando di fatto il capacity drop e forzando la rete a operare quasi al suo limite fisico teorico. Inoltre, l'analisi rivela un compromesso spazio-temporale fondamentale nel dispiegamento della flotta: attivare il controllo più a monte aumenta significativamente l'impronta temporale di ciascun CAV, consentendo a flotte più piccole di mitigare parzialmente l'incidente. Al contrario, l'attivazione in prossimità richiede flotte più ampie per sostenere la copertura temporale, ma raggiunge un'efficienza assoluta maggiore minimizzando la latenza di inizializzazione.

Infine, per validare la fattibilità del sistema proposto, sono stati sviluppati un modello di domanda probabilistica e una simulazione topologica Monte Carlo. L'analisi dimostra che il framework raggiunge un'affidabilità spaziale superiore al 99% per la formazione di blocchi di controllo su più corsie con un modesto tasso di penetrazione di mercato dei CAV del 6,3%. Queste scoperte forniscono strategie dinamiche e pronte per l'implementazione, altamente pratiche per i futuri centri di gestione del traffico che opereranno in condizioni realistiche a breve termine.

Acknowledgments

Eccoci, la parte più difficile da scrivere. Partiamo dalle cose scontate.

Grazie all'università di Pavia, che mi ha permesso di continuare il mio percorso di studi nonostante la mia media in triennale (vero poli?).

Grazie alla prof. Ferrara, che ha accettato di essere la mia relatrice e mi ha fornito il tema per la tesi.

Grazie a Nikolas, che nonostante sia stato probabilmente il tesista meno coinvolto e presente della storia accademica, mi ha comunque supportato con materiale, consigli, critiche e correzioni.

Grazie a Gianluca Guidotti, che non so chi tu sia, ma solo dio sa perchè le tue credenziali eduroam funzionino, mentre le mie no. Sei l'eroe che noi del polo non meritavano ma di cui avevamo bisogno.

Grazie al mio ex datore prof. Stefano Secci e in generale le CNAM per avermi permesso di vivere un'esperienza indimenticabile come primo lavoro.

Grazie ad Alessandro, Mario, Noemi, Julie e Kamal per essere stati la mia compagnia in quel di Parigi, non vedo l'ora di venire a trovarvi, farmi autografare i volumi di Scarlet Queen e fare un altro karaoke con voi post ramen.

Grazie a Brahim, Adam e Azad per essere stati degli ottimi coinquilini, dal capodanno cinese, al costringervi a guardare Bakemonogatari assieme, dall'insultare Ruben (l'altro coinquilino voncione) all'osservare le macchine in fiamme sotto casa durante le rivolte. Bei ricordi.

Ok adesso inizia il bello. Non avendo fatto una tesi in triennale, fate che vale per entrambe. (ricordatevene quando direte che ho un ghiacciolo al posto del cuore)

Grazie a mio fratello Lore, che da un decennio mi tira in mezzo in situazioni mega ambigue rendendo la mia vita infinitamente più interessante e contagiandomi

con le sue passioni, prima le moto poi la montagna. Sei probabilmente la persona con cui più ho discusso in vita mia, a volte mi trovo a pensare che osserviamo il mondo con lenti diametralmente opposte, ma nonostante questo eccoci ancora qua, consapevoli che ci siamo sempre stati e sempre saremo l'uno per l'altro. Grazie per il sostegno che tu e la tua famiglia mi avete fornito durante il periodo covid, ospitandomi ogni giorno per tutti i mesi di lockdown in assenza di una connessione a casa mia. Questa laurea è anche tua.

Grazie a mio fratello Fra, che con la sua dannata bravura ha alimentato il mio spirito di competizione in ambito accademico. Purtroppo non sono riuscito a starti dietro ma, se non ti avessi preso come modello da battere, probabilmente mi sarei adagiato al livello del tizio qui sopra e non starei scrivendo questi ringraziamenti. Ogni tanto sento la tua mancanza, da qualche anno ci vediamo sempre meno e mi deprimo pensando che il nostro rapporto non è più lo stesso, poi quando ci rincontriamo realizzo che era tutta una mia fisma mentale, che hai ssemplicemente trovato il tuo equilibrio tra noi e Giulia e che comunque vada resteremo migliori amici. Questa laurea è anche tua.

Grazie ai miei genitori, per avermi permesso di studiare. So di essere tra i peggiori figli che si possano avere, non parlo mai, praticamente è come non avere un figlio. Non so bene come siamo arrivati a questo punto, forse siamo semplicemente persone incompatibili, tuttavia non vi siete mai tirati indietro dai vostri doveri e avete sempre cercato di supportarmi. Grazie.

Grazie alle mie sorelle, ancora una volta so di non essere il migliore dei fratelli. Magari complice la differenza di età, ma non riesco mai a sentirmi a mio agio ed essere me stesso con dei parenti affianco. Nonostante questo voglio che sappiate che per voi ci sarò sempre, specialmente adesso che ho scoperto di diventare zio. Vi auguro di avere un figlio migliore di quanto non sia stato io. Grazie.

A mio nonno, che ha sempre voluto avere un figlio o un nipote laureato. Non ce l'hai fatta a vedere questo sogno realizzato, e sebbene io non creda al paradiso, per una volta vorrei pensare che tu mi stia guardando da lassù e sia orgoglioso di me.

Grazie al Gruppo Bunker, anche se non mi faccio vedere se non quando ci sono feste o ragazze. In particolare:

Grazie a Pol, che anche tu non scherzi nel trascinarmi in situazioni ultra mega ambigue; grazie di cuore per tutte le vacanze che mi hai permesso di vivere nelle mille case che hai sparse in giro per il mondo e grazie per ospitarmi nelle serate a Parma, anche se poi ti ho spinto a farti quella tipa stramba. Sto ancora aspettando l'esperimento con il cuore di vacca da fare con il calabrone (bella Fra).

Grazie a Pollo, che anche se il nostro rapporto è sempre stato amore e odio, mi hai comunque aiutato quando ero a Parigi senza un tetto e senza un soldo. Sono sicuro che non riusciremo mai veramente ad andare d'accordo, ma sappi che se mai avrai bisogno di una mano io ci sarò.

Grazie a Marco, che da quando abbiamo iniziato a vederci più spesso sei diventato il mio compagno di confronti ingegneristici. Quando non capivo qualcosa ero certo che parlandone con te qualcosa ne sarebbe venuto fuori, anche se per te è significato sorbirti, tra le altre cose, il funzionamento del motore trifase all'una di notte. p.s. russa di meno, sia per me che per Chiara.

Grazie alle girls: Lu, Sofi e Giada. Un'amicizia iniziata con entrambi i piedi sbagliati, ma che poi si è rivelata tra le più preziose. Conoscervi mi ha fatto capire tantissimo di me e so di non essere molto sincero nell'esternare i miei sentimenti, ma spero, nel corso di questi anni, di avervi trasmesso l'importanza che avete per me. Grazie per essermi state così vicine durante il mio soggiorno in Francia, ricorderò con affetto tutte le nostre chiamate (scusa Lu per la bolletta di quella volta). Giada qui un po' meno, ma ti voglio bene lo stesso.

Grazie al Gruppo Ego Ideam, che mi ha accolto al momento del mio rientro in patria, quando la mia precedente compagnia si era sfaldata. in particolare:

Grazie a Ricky, ci conosciamo bene solo da un paio di anni ma ormai mi sembra di conoscerti da una vita. Hai una forza e una voglia di fare che poche volte ho trovato in altre persone ed è solo grazie a te se tutti noi abbiamo potuto unirci sotto l'associazione e fare nuove amicizie. Ricorderò sempre lo stress delle tue domande di automatica mentre cercavo di studiare.

Grazie a Tia, il mio fidato compagno di discussioni sui massimi sistemi della fisica. Ci bastano sei-sette drink, a te anche quindici-venti, ed eccoci a parlare di cose incomprensibili per ore. Adesso che ci siamo laureati entrambi possiamo finalmente progettare quel dannato essiccatore a freddo per Ezio.

Grazie al Gruppo Polo, che mi ha permesso di vivere la bellezza di avere un

gruppo di studio. Per noi pendolari l'università fa schifo. Sempre di fretta per non perdere le coincidenze, non abbiamo il tempo di socializzare con i nostri compagni di corso, perdendo una parte vitale del percorso di laurea, anche se ad ingegneria c'è poco da socializzare con gli strambi che la frequentano. Grazie a voi ho potuto vivere cosa si prova a trovarsi assieme per affrontare gli esami. In particolare:

Grazie a Blanca, che mi hai fatto scoprire le gioie di una carne cotta come dio comanda. A volte, di notte, mi capita di sognare il sapore della tua picanha, spero che le grigliate tutti insieme possano continuare per sempre.

Grazie a Elo, l'unico vero amico che è rimasto costante dalle superiori. Adesso hai trovato la forza di inseguire la vita che hai sempre desiderato in Olanda e spero di cuore che alla fine tu possa sentirti realizzato. Sappi che sarò sempre pronto a supportarti come tu hai fatto con me all'epoca. Buona fortuna.

Spero di non aver dimenticato nessuno, voglio scrivere tutto in una volta sola nel flusso di coscienza per buttare fuori l'interità dei miei sentimenti, senza più ricontrollare cosa ho scritto. Nel caso vi prego di scusarmi.

Dulcis in Fundo, per usare il latino che tanto disprezzo, una persona speciale. Come tutti saprete, io non sono proprio il ragazzo standard. In questi ventisei anni ho conosciuto decine se non centinaia di ragazze, tuttavia non ho mai provato quello che vedevo provare dai miei coetanei. Fino all'anno scorso.

Non ti ho nemmeno vista arrivare e come una goccia d'acqua hai scavato e scavato fino a trovare dei sentimenti che ormai pensavo di non avere. Mentre scrivo questa dedica non dovresti ancora essere consapevole di cosa tu mi abbia fatto provare in questi mesi, e la cosa divertente è che il pomeriggio stesso in cui sto scrivendo ho intenzione di rivelartelo, non riesco a pensare ad altro da settimane.

Comunque vada vorrei ringraziarti di esistere, perchè il solo fatto di averti conosciuta mi ha dato la speranza che al mondo anche io possa trovare una ragazza da amare.

Grazie.

Contents

List of Figures

Acronyms

Acronyms in alphabetical order

AV	Autonomous Vehicle
CACC	Cooperative Adaptive Cruise Control
CAV	Connected and Autonomous Vehicle
CFL	Courant-Friedrichs-Lewy
CTM	Cell Transmission Model
CV	Connected Vehicle
HDV	Human-Driven Vehicle
ICE	Internal Combustion Engine
ITS	Intelligent Transportation Systems
LWR	Lighthill-Whitham-Richards
MPC	Model Predictive Control
MTFC	Mainstream Traffic Flow Control
PDE	Partial Differential Equation
PM	Particulate Matter
PR	Penetration Rate
RM	Ramp Metering
SAE	Society of Automotive Engineers
TTS	Total Time Spent

TTT	Total Travel Time
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything
VSL	Variable Speed Limits

Chapter 1

Introduction

"There is a stubbornness about me that never can bear to be frightened at the will of others. My courage always rises at every attempt to intimidate me."

"Pride & Prejudice", Jane Austen

1.1 The General Context: From Mobility to Congestion

Mobility has long been understood as a base component of civilized society as well as a key driver of socio-economic change. With the past century as evidence, effective transport of people as well as goods has marked the success of a nation in a global environment. Moreover, the widespread use of the private car in the 20th century marked a staggering triumph for **individual liberty** in a way that granted individuals access to a means of transportation from any point to any other point in a manner that has never occurred in history.

However, it also led to a **problem**. The rapid growth of the world's population in urban areas [6] and the resultant growth in the number of vehicles on the road [7] have caused an unanticipated pressure on the road network. Even as cities have expanded in size and global supply chain patterns have increased in complexity to

reap the benefits of global trade, the demand on the transport network has increased in ways that physical infrastructure cannot meet. The **demand** in the transport network, represented by the desire to travel, is dynamic and unbounded, while the **supply**, represented by the road network, becomes static.

The clear consequence of such an inequality is **traffic congestion**. Congestion is created when,in a particular area, the **vehicle density** exceeds the number beyond which the speed reduces, leading to instability in the flow. It is an issue that signifies not only an annoyance to people commuting to work but also an inefficiency in the economy. Time, an invaluable asset, gets wasted in congestion, leading to losses in productivity.

Moreover, the congestion problem not only encompasses economic inefficiency [8]; it also encompasses a significant aspect of the critical **environmental crisis**. The transportation sector is one of the major consumers of fossil fuels [9] [10]. The **stop-and-go** conditions created by congested roads result in a substantial increase in fuel consumption as compared to conditions on free-flow roads. Further, with increased density on roads, safety conditions are affected as there are increased opportunities for vehicle interaction.

In this respect, the conventional method of merely building our way out of congestion by adding more road capacity has become increasingly antiquated and unsustainable. The geographical and environmental impact of road capacity has now reached its point of diminishing returns in most urban areas [11]. Consequently, the **paradigm shift** now necessary in the domain of twenty-first-century traffic engineering is the gradual transition away from physical capacity expansion towards the intelligent utilization of existing infrastructure. Initially, this objective drove the development of conventional macroscopic control methodologies, including Ramp Metering to regulate highway inflows, Variable Speed Limits (VSL) to harmonize mainline traffic pacing, and dynamic Route Guidance to optimize network-wide demand distribution. Today, this technological evolution has reached a new frontier with the inclusion of **Connected and Autonomous Vehicles (CAVs)** in the traffic flow, transitioning the field from fixed-point infrastructure control to mobile, predictive traffic regulation.

1.2 The Multidimensional Impact of Traffic Congestion

As established in the previous section, the imbalance between travel demand and the supply of infrastructure leads to congestion. However, for a complete under-

standing of the urgency in implementing advanced control strategies—like the ones presented in this thesis—it is important to analyze the consequences of this congestion. In transport economics, these impacts are termed **negative externalities**: costs generated by the user which are not borne by the user himself but imposed on society. The effect of these externalities is widespread, affecting economic productivity, public health, environmental sustainability, and the psychological condition of the population [8].

1.2.1 Economic Impact: Time Loss and Fuel Consumption

The most immediate and quantifiable consequence of traffic congestion is the systemic loss of time. In macroscopic traffic engineering, this inefficiency is formally evaluated using aggregate metrics such as Total Travel Time (TTT) or Total Time Spent (TTS). TTT represents the cumulative sum of the travel times experienced by all vehicles traversing a specific road network during a given period. Mathematically, it is defined as the integral of the number of active vehicles in the network, $N(t)$, over the entire observation horizon T:

$$TTT = \int_0^T N(t)dt$$

This metric is universally expressed in vehicle-hours (veh·h). TTS is a closely related formulation that often explicitly incorporates the additional delay accumulated by vehicles waiting in upstream virtual queues (e.g., at on-ramps) before physically entering the main network. Because time is an irreproducible economic resource, any increase in these metrics signifies that occupants are trapped in a non-productive state. Therefore, minimizing TTT and TTS is the primary objective of modern traffic control, as gridlock directly translates into substantial socioeconomic and environmental costs. For commuters, this means less leisure or fewer hours put into work.

The impact is even more severe for the logistics sector. Modern supply chains depend on Just-in-Time delivery models that rely on precise timing. However, reduced capacity during peak times magnifies the impact of non-recurrent incidents, introducing severe travel time variability and uncertainty into the network [12]. Waste also arises when a delivery truck stuck in traffic incurs a cost other than that of the driver's wages; it might delay an entire line of production or make the company forfeit some of its contractual deadlines.

Closely linked with time loss is the severe degradation of fuel efficiency [13]. Internal Combustion Engine (ICE) vehicles are designed to operate most efficiently

at steady, moderate speeds—usually from 60 to 90 km/h [1]. Under congestion, however, the regime of traffic flow switches to what is called *stop-and-go* waves [14]. Within this regime, the vehicles go through a series of rapid accelerations and sudden braking. This driving pattern is thermodynamically inefficient. The kinetic energy built up is lost as heat when the brakes are applied in deceleration, instead of being used for displacement [15]. Consequently, fuel consumption per kilometer escalates in congested scenarios. It is a direct economic loss to the drivers themselves and a gross national squandering of imported fossil fuels.

1.2.2 Environmental Impact: The Ecological Footprint

The environmental externalities of traffic are perhaps the most critical in the long term. The transport sector is the largest contributor to the degradation of the atmosphere, and congestion greatly enhances this contribution. The emissions produced by road traffic can be divided into two distinct groups, both of which are exacerbated by unstable flow conditions:

1. **Global Pollutants (Greenhouse Gases):** The major byproduct of burning fossil fuels is Carbon Dioxide (CO_2). While not toxic to humans directly, CO_2 represents the most dominant agent of anthropogenic climate change. As stated earlier, the increased fuel use inherent in stop-and-go driving leads to a proportional increase in CO_2 emissions. Therefore, reducing traffic volume is not just a mobility goal, but a decarbonization strategy.

2. **Local Pollutants:** These include Nitrogen Oxides (NO_x), Carbon Monoxide (CO), and Particulate Matter (PM_{10} and $PM_{2.5}$). Unlike CO_2, these substances have immediate adverse impacts on air quality in urban areas. The physics of combustion engines dictate that the emission rates for these pollutants are highest during transient engine loads—particularly during hard acceleration. In a congested highway situation, where a driver might accelerate hundreds of times over a short distance to close the gap with the vehicle in front [16], the emission of toxic pollutants is exponentially higher than in free-flow conditions. This creates hotspots of pollution along major arterial roads, posing severe health risks (respiratory and cardiovascular diseases) to communities living nearby [17] [18].

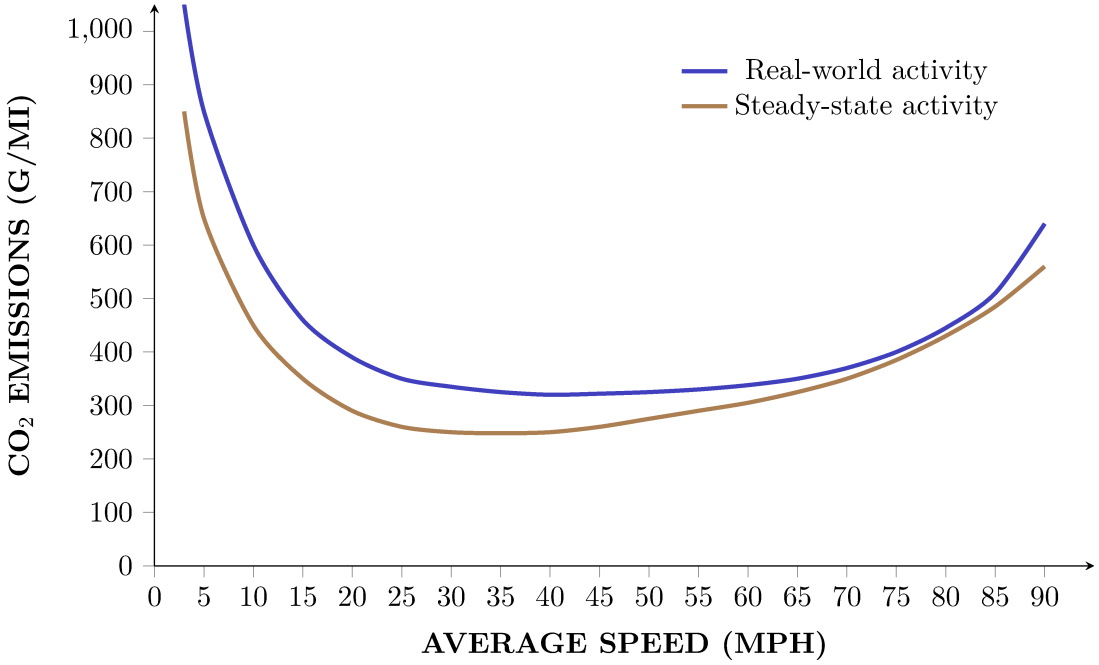


Figure 1.1: The relationship between vehicle speed and CO2 emissions [1]

.

1.2.3 Social and Human Impact: Safety and Stress

Beyond the economic and environmental metrics, traffic congestion also imposes a heavy toll on human capital, manifesting in safety risks and psychological stress.

Another important aspect is safety: there is a complex correlation between traffic density and accident rates. While fatal accidents at high speeds are less common in gridlock, congestion greatly increases the frequency of vehicle interactions. The stop-and-go phenomenon creates shockwaves which propagate upstream against the flow of traffic. Human drivers often fail to anticipate the sudden deceleration of the downstream platoon due to their inherent perception-reaction latency. During the backward propagation of a traffic shockwave, this delayed response frequently results in low-severity rear-end collisions—commonly referred to as “fender-benders”. While these incidents typically involve minimal structural damage, the vehicles involved remain stranded on the carriageway, effectively acting as secondary bottlenecks that instantly trigger further capacity drops and exacerbate network congestion. Research has shown that these traffic oscillations are a primary predictor of freeway crash occurrence [19] [20]. These fender-benders, while usually minor, can block lanes and thereby cause secondary accidents, further crippling the network capacity. This sets up a vicious cycle wherein congestion breeds accidents, and accidents breed

further congestion—a phenomenon often categorized as *non-recurrent congestion*.

The psychological burden of all this on drivers should not go unnoticed. The modern driving environment is a major cause of stress. This is a result not just of the duration of the trip, but of its unpredictability. The impedance of the commute—the effort required to overcome the distance—builds cognitive load and frustration. This process, often referred to as "commuter stress," has been linked to higher blood pressure, anxiety, and decreased overall life satisfaction [21]. In extreme cases, it manifests as aggressive driving behavior ("road rage"), which further compromises safety. By smoothing the driving task and reducing travel time variance, traffic control systems aim to mitigate not only the physical queues on the road but also the cognitive load on the driver.

1.3 Evolution of Traffic Control Strategies

The strategy for addressing traffic congestion has transformed greatly over the past few decades, moving from physical interventions to digital and algorithmic solutions. This development can be divided into three distinct phases: Infrastructure expansion, Road-Based traffic control, and the new paradigm of Vehicle-Based control [22].

1.3.1 Historical Approach: Infrastructure Expansion

Intuitively, we can image that increasing road capacity can solve the rising demand for travel. This included the construction of new roads as well as the addition of new lanes to existing ones and the design of complex interchanges.

Figure 1.2: The Almondsbury Interchange ("The Stack") in the UK. **Source:** [2].

Even though this approach gave temporary relief, it proved unsustainable. As demonstrated mathematically by Braess's Paradox [23], adding capacity to a network can counter-intuitively degrade overall performance rather than improve it, as drivers selfishly optimize their individual routes to reach a User Equilibrium rather than a System Optimum.

To illustrate this phenomenon, consider a classic four-node transportation network where 4000 vehicles must travel from a start node (S) to a destination node (D). The network initially contains two intermediate nodes, A and B, creating two possible symmetric routes: $S \to A \to D$ and $S \to B \to D$.

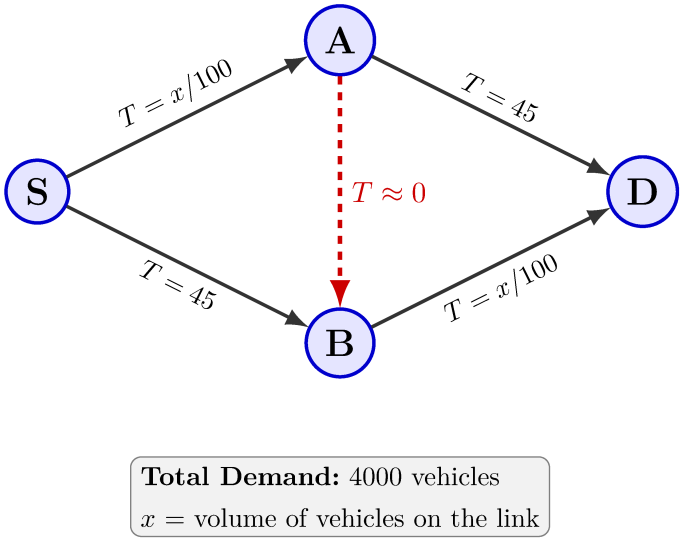


Figure 1.3: A classic four-node network demonstrating Braess's Paradox. The addition of the high-speed, zero-cost link ($A \to B$) causes all drivers to selfishly choose the $S \to A \to B \to D$ route, paradoxically increasing the total travel time for everyone from 65 minutes to 80 minutes.

The travel time on each link depends on the volume of traffic, x. The highway links ($S \to B$ and $A \to D$) have a massive capacity, resulting in a fixed travel time of 45 minutes regardless of congestion. The local bridge links ($S \to A$ and $B \to D$) are highly sensitive to congestion, with a travel time defined as $x/100$ minutes.

- **Before Expansion:** Vehicles distribute evenly to minimize their individual travel times. With 2000 vehicles on each route, the travel time for every driver is identical and optimal:

$$T = \frac{2000}{100} + 45 = 65 \text{ minutes}$$

- **After Expansion (The Paradox):** Suppose the transportation authority builds a highly efficient, high-speed shortcut connecting node A directly to node B, with a negligible travel time ($T_{A\to B} \approx 0$). Selfishly optimizing their routes, every driver will evaluate the network and realize the path $S \to A \to B \to D$ is strictly dominant. At node S, the path to A never exceeds 40 minutes (even if all 4000 cars take it), which always beats the fixed 45-minute path to B. At node A, taking the free shortcut to B and proceeding to D never exceeds 40 minutes, which again beats the 45-minute path directly to D. Consequently, the entire flow of 4000 vehicles shifts to this new central route. The new travel time for every driver becomes:

$$T = \frac{4000}{100} + 0 + \frac{4000}{100} = 80 \text{ minutes}$$

Despite the physical addition of a new, high-speed infrastructure link, the selfish routing behavior of individual drivers increases the total travel time for everyone in the network from 65 to 80 minutes. This mathematically proves the inefficiency of uncoordinated capacity expansion and highlights the necessity for active traffic control strategies.

In addition, the physical extension of infrastructure is limited by high costs and the lack of available space in metropolitan areas, along with serious implications for the environment.

1.3.2 Current State: Road-Based Traffic Control

Realizing the limitations of Civil Engineering solutions, attention drifted to Traffic Engineering, specifically **Traffic Management and ITS** (Intelligent Transportation Systems). The aim shifted from building more roads to using roads more efficiently [24] [25] [26]. This era introduced strategies known as "Road-Based Control," where control is applied to the aggregate traffic flow. Some of the most successful applications are:

- **Ramp Metering (RM):** A traffic light installed at the on-ramp regulates the rate at which vehicles enter the freeway [27].

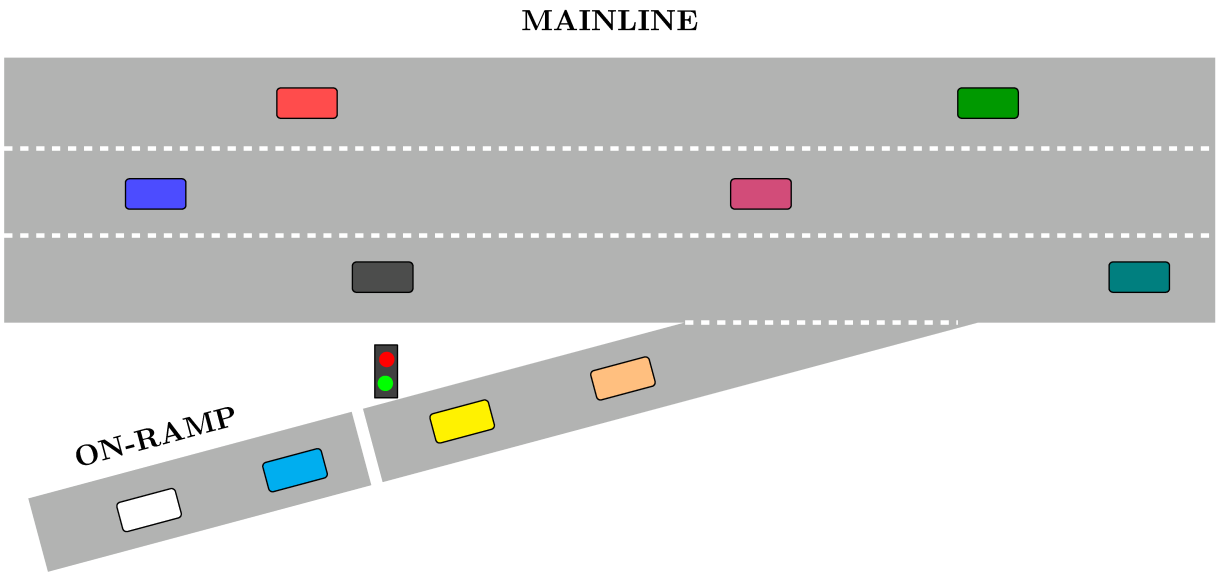


Figure 1.4: Schematic representation of a typical ramp metering system architecture and logic flow.

Ramp Metering (RM) has been widely demonstrated to effectively disrupt dense platoons of entering vehicles, thereby preventing the capacity drop that typically occurs when high-density merges destabilize the mainstream traffic flow [28]. However, the efficacy of this approach is fundamentally limited by its high infrastructure costs, spatial constraints, and the physical storage capacity of the on-ramps. Furthermore, if the accumulated queue threatens to spill back into the adjacent urban road network, the metering system must typically be overridden or deactivated to prevent gridlock on local streets. Consequently, RM often becomes structurally ineffective precisely during the most severe congestion peaks when continuous, mainline control is needed the most.

- **Variable Speed Limits (VSL):** Dynamic speed limits shown on overhead Gantries (Variable Message Signs) [26].

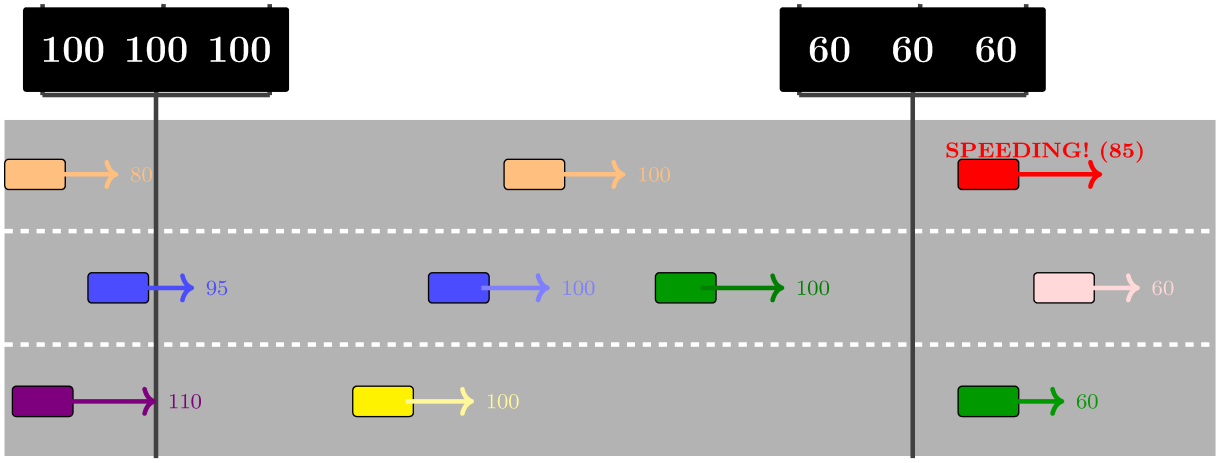


Figure 1.5: Schematic representation of a standard Variable Speed Limit (VSL) system architecture and control mechanism. The diagram illustrates how a uniform speed broadcast via gantries improves traffic homogenization and suppresses backward-propagating waves, while also explicitly highlighting the fundamental limitation of unpredictable human driver compliance.

Variable Speed Limits (VSL) are designed to improve the homogenization of traffic flow by significantly reducing speed variance among vehicles [29]. This macroscopic smoothing process is highly beneficial, as it mitigates the probability of rear-end collisions and actively suppresses the formation of backward-propagating stop-and-go waves. However, the operational success of VSL is fundamentally dependent on the unpredictable compliance of human drivers. Furthermore, traditional VSL systems act strictly as an open-loop, macroscopic intervention, broadcasting a uniform speed instruction via fixed gantry signs to all approaching traffic, and therefore lack the precision required to directly dictate individual vehicle dynamics or actively pace specific platoons.

Although these infrastructure-centric approaches improve network performance, they are hindered by fundamental limitations. Ramp Metering is rigidly localized to on-ramps and requires expensive fixed installations, while Variable Speed Limits suffer from unpredictable human compliance and cannot enforce specific vehicle trajectories. Ultimately, because both strategies are anchored to fixed physical locations, they lack the mobility required to dynamically track and mitigate traffic shockwaves across continuous, open highway segments.

1.3.3 The Future: Vehicle-Based Control

We are currently witnessing a paradigm shift towards **Vehicle-Based Control**. Unlike traditional methods that treat traffic as a fluid to be managed externally, this approach views the vehicle itself as an active *actuator* in the control loop. To fully understand this potential, it is necessary to distinguish between two converging technologies: Automation (SAE J3016 levels [3]) and Connectivity (V2X).

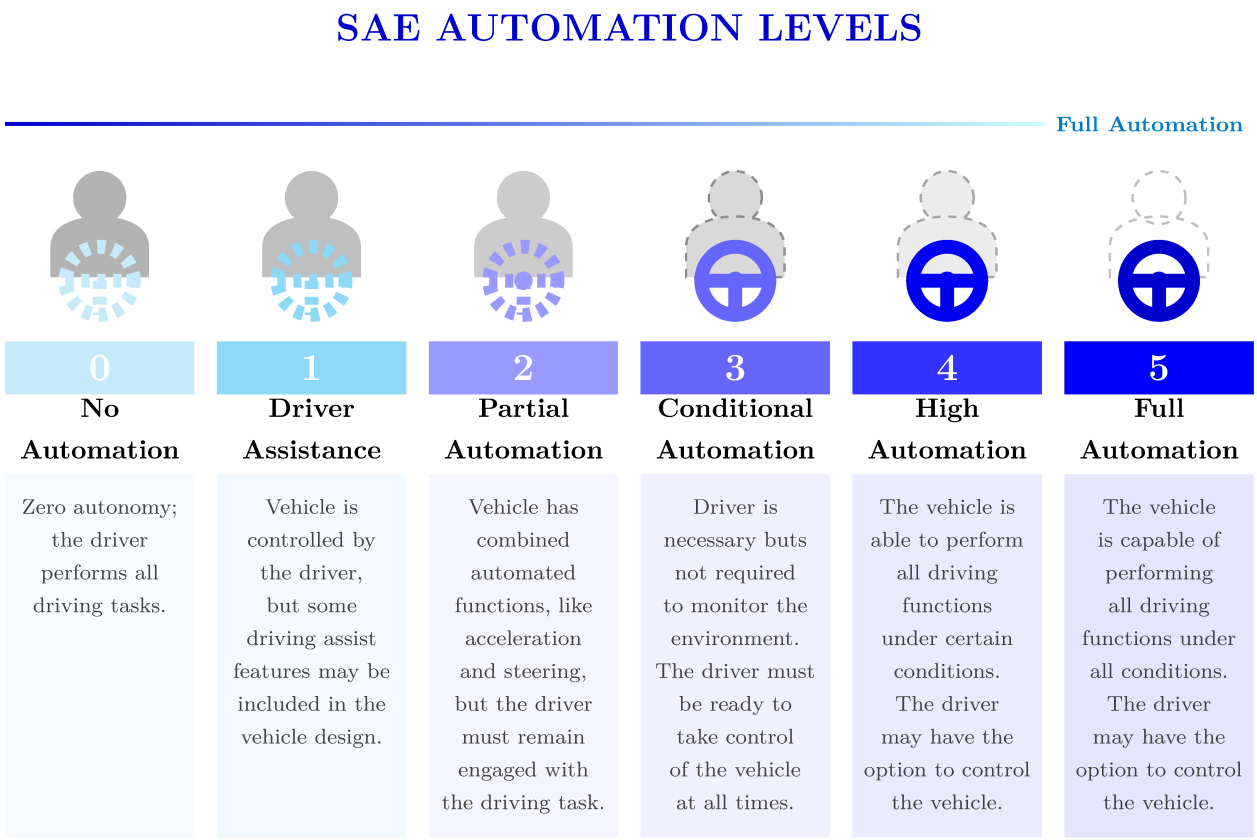


Figure 1.6: The six levels of driving automation as defined by the Society of Automotive Engineers (SAE) [3]

While often used interchangeably, these terms refer to distinct capabilities:

- **Autonomous Vehicles (AVs):** Refers to the capability of a vehicle to sense its environment and operate without human involvement. According to the **SAE J3016** standard, automation is classified into six levels, ranging from Level 0 (No Automation) to Level 5 (Full Automation). In the context of traffic control, we generally refer to vehicles at Level 3 or higher, capable of handling the dynamic driving task.

- **Connected Vehicles (CVs):** Refers to the ability to exchange data with other agents via V2X (Vehicle-to-Everything) communication. This includes V2V (Vehicle-to-Vehicle) and V2I (Vehicle-to-Infrastructure).

The synergy of these technologies results in the Connected and Autonomous Vehicle (CAV). In this thesis, a *vanguard* of CAVs, defined here as a synchronized, lateral formation spanning multiple lanes to act as a moving bottleneck, is utilized as a mobile actuator to enforce a specific control strategy known as **Mainstream Traffic Flow Control (MTFC) via Moving Bottlenecks**.

The control logic relies on two synergistic mechanisms:

1. **Cooperative Vanguard (The Actuator):** Three CAVs utilize V2V communication and CACC (Cooperative Adaptive Cruise Control) to travel in a tight formation across parallel lanes [30]. Unlike a single vehicle, this vanguard forms a cohesive physical block—essentially a "moving bottleneck"—that prevents human-driven vehicles from overtaking or cutting in, thereby granting the system full control over the speed of the traffic stream behind it.

2. **Moving Bottleneck Control (The Strategy):** This vanguard acts as a virtual, mobile traffic light. By intentionally slowing down upstream of a critical infrastructure bottleneck (such as a lane closure, work zone, or capacity drop), the formation creates a *moving bottleneck*. This action regulates the flow arriving at the real bottleneck, preventing it from oversaturating. This effectively replicates the benefits of Ramp Metering but is applied directly on the mainstream highway, preventing the breakdown of flow and the associated capacity drop.

1.4 Thesis Structure

The remainder of this thesis is organized as follows:

Chapter 2 establishes the theoretical foundations of macroscopic traffic flow. It introduces the fundamental hydrodynamic variables and their relationships across both free-flow and congested regimes using the fundamental diagram. The chapter further details the kinematics of traffic breakdown, specifically analyzing shockwave propagation, hysteresis, and the physical mechanisms behind the capacity drop phenomenon. Finally, it reviews the evolution of macroscopic traffic modeling, culminating in the foundational Cell Transmission Model (CTM).

Chapter 3 expands the standard CTM into a comprehensive multi-lane freeway environment to serve as the simulation's physical plant. It details the extended conservation law alongside a custom, density-based lane-changing model governed by cooperative pressure and dynamic permeability limits. Furthermore, this chapter formulates the mathematical representation of the capacity drop using a density-dependent capacity function, and strictly defines the temporal domain, traffic demand, and bottleneck event for the simulation scenarios.

Chapter 4 formulates the proposed traffic control strategy, utilizing CAVs as Lagrangian actuators to regulate dynamic flow. It defines the Sequential Model Predictive Control (MPC) framework and its receding horizon principle. A primary focus is placed on the normalized multi-objective cost function, which aims to minimize TTT by physically trading spatial headway for increased bottleneck capacity through a discrete action space.

Chapter 5 presents the primary simulation results and performance analyses. It first establishes theoretical performance bounds by comparing an ideal, capacity-drop-free incident against an uncontrolled baseline. The chapter then evaluates the control strategy through an extensive simulation matrix, providing a deep dive into the spatiotemporal dynamics of the system. It concludes by analyzing the critical trade-offs between CAV activation distances, temporal saturation, and initialization latency to determine the system's overall efficacy.

Chapter 6 investigates the geometric scalability and robustness of the control strategy, testing the framework's adaptability across reduced-capacity scenarios, including two-lane environments and pure longitudinal control on single-lane corridors (ancora da fare).

Chapter 7 summarizes the key findings of this research, discusses the practical implications for future traffic management, and outlines necessary avenues for future work.

Chapter 2

Theoretical Background on Traffic Modelling

Before describing the specific control strategies adopted in this thesis, it is necessary to introduce the mathematical foundations of traffic flow theory. This chapter defines the fundamental variables used to represent traffic dynamics at a macroscopic level and illustrates the physical phenomena characterizing freeway operations, such as congestion, bottlenecks, and capacity drop.

2.1 Fundamental Traffic Variables and the Hydrodynamic Relation

To analyze the complex dynamics of freeway traffic and design effective control strategies, it is necessary to adopt a mathematical framework that describes the system's behavior. In the context of this thesis, the traffic stream is modeled according to the **macroscopic approach** [26]. Unlike microscopic models, which track the trajectory of every single vehicle, macroscopic theory treats traffic as a continuous compressible fluid [31]. This continuum approximation allows the state of the traffic flow to be completely described by three aggregate variables: density, flow, and speed.

2.1.1 Density (ρ)

Traffic density, denoted by ρ, is the spatial concentration of vehicles. It is defined as the number of vehicles occupying a specific length of road at a given instant in time. The standard unit of measurement is vehicles per kilometer (veh/km). Density is a

critical parameter because it directly reflects the crowding of the road; it ranges from zero (empty road) to a maximum value known as *jam density* (ρ_{jam}), which depends on road structure, where vehicles are completely stopped in a bumper-to-bumper formation.

2.1.2 Flow (q)

Traffic flow, denoted by q, represents the temporal rate of the traffic stream. It is defined as the number of vehicles passing a specific cross-section of the roadway during a unit of time. It is typically expressed in vehicles per hour (veh/h). While density is a "snapshot" of the road at a fixed time, flow is a measure of throughput over a duration.

It is important to note that flow is zero in two opposing conditions: when there are no cars on the road, and when traffic is completely stopped (jam density).

2.1.3 Speed (v)

Speed is the distance traveled per unit of time, expressed in kilometers per hour (km/h). However, in macroscopic traffic theory, the definition of "average speed" requires careful distinction [32]. There are two ways to calculate the mean speed of a group of vehicles:

- **Time Mean Speed (v_t):** The arithmetic average of spot speeds measured at a specific point (e.g., by a radar gun).
- **Space Mean Speed (v_s):** The harmonic mean of speeds of vehicles traversing a road segment.

For the fundamental equations of traffic flow to hold valid, the **Space Mean Speed** must always be used. This variable captures the true average velocity of the traffic stream over a spatial segment.

2.1.4 The Fundamental Equation

These three variables are not independent; they are intrinsically linked by the fundamental relationship of traffic flow theory, which mirrors the continuity equation in fluid dynamics:

$$q = \rho \cdot v \tag{2.1}$$

This equation states that the flow passing a point is equal to the product of the density and the space mean speed. This relationship holds true for any steady-state traffic condition. It implies that for a constant flow, a reduction in speed (due to congestion) must be mathematically accompanied by a proportional increase in density. This equation serves as the cornerstone for the Fundamental Diagram, which graphically describes the equilibrium states of the traffic system.

2.2 The Fundamental Diagram of Traffic Flow

The hydrodynamic relationship 2.1 introduced in the previous section is necessary but not sufficient to describe traffic behavior, as it contains three variables. To close the system of equations, a functional relationship between flow and density is required. This relationship is graphically represented by the **Fundamental Diagram**, which describes the steady-state equilibrium of the traffic stream.

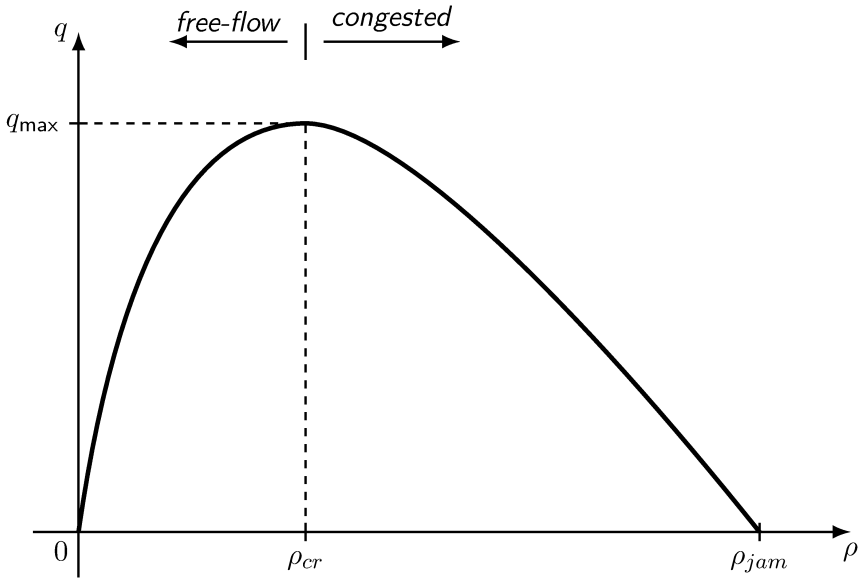


Figure 2.1: The macroscopic fundamental diagram (flow-density relationship), illustrating the free-flow branch and the congested branch.

The Fundamental Diagram, denoted as $q(\rho)$, plots traffic flow on the vertical axis against traffic density on the horizontal axis. While empirical data often shows scattering due to human heterogeneity, the theoretical curve is generally concave, starting at the origin $(0, 0)$ and ending at jam density $(\rho_{jam}, 0)$. This curve is divided into two distinct regimes by a crucial inflection point: the **Free-Flow Regime** and the **Congested Regime**.

2.2.1 Free-Flow Regime

The left side of the diagram in figure 2.1, where density ranges from zero to a critical value (ρ_{cr}), represents the Free-Flow regime. In this state, the density of vehicles is low enough that interactions between drivers are negligible. Vehicles can travel at their desired speed, limited only by the speed limit or road geometry. Mathematically, this section of the curve is often approximated as linear. The slope of this line represents the **Free-Flow Speed** (v_f).

$$q = v_f \cdot \rho \quad \text{for} \quad 0 \leq \rho \leq \rho_{cr} \tag{2.2}$$

In this regime, an increase in density results in a proportional increase in flow. This is the stable operating domain of the highway, where traffic control is generally passive.

2.2.2 Critical Density and Capacity

The peak of the Fundamental Diagram represents the maximum possible flow that the road section can sustain. This maximum value is the **Capacity** of the road (q_{max}). The density value corresponding to this peak is termed the **Critical Density** (ρ_{cr}). Critical density acts as the tipping point for traffic stability. It represents the optimal utilization of the infrastructure. Ideally, traffic control strategies aim to maintain density exactly at this critical value to maximize throughput. Once density exceeds ρ_{cr}, the system undergoes a phase transition.

2.2.3 Congested Regime

The right side of the diagram 2.1 ($\rho > \rho_{cr}$) represents the Congested regime (often referred to as unstable or forced flow). Here, the density is so high that vehicles obstruct one another. As density increases further towards jam density, the space mean speed drops significantly. Counter-intuitively, in this regime, adding more vehicles reduces the flow. This creates a dangerous inefficiency: the road is full of cars, but fewer cars are actually passing through the bottleneck per hour. This instability implies that once congestion sets in, it tends to persist or worsen unless the demand drops significantly. The curve eventually intersects the density axis at **Jam Density** (ρ_{jam}), where vehicles are bumper-to-bumper, speed is zero, and consequently, flow is zero.

2.2.4 The Speed-Density Relationship

While the Fundamental Diagram 2.1 ($q - \rho$) describes the system's throughput, the driver's perspective is best represented by the relationship between **Speed (v)** and **Density (ρ)**.

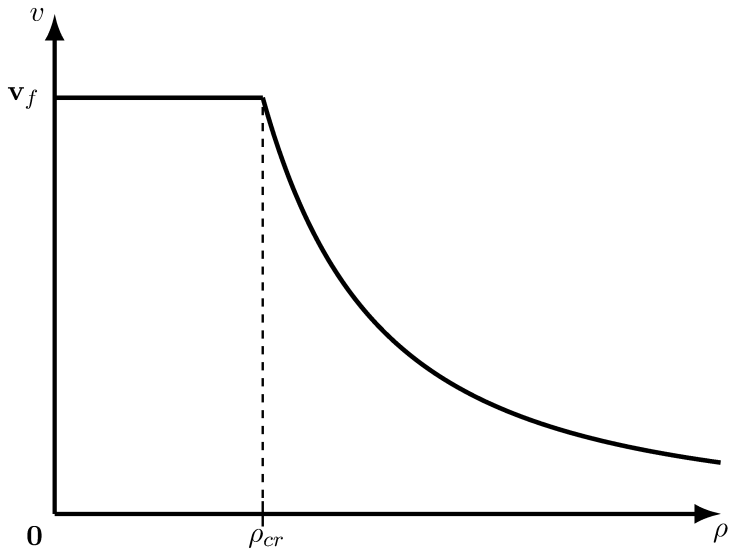


Figure 2.2: Speed-density ($v - \rho$) relationship, featuring a piecewise-continuous constant free-flow speed (v_f) followed by a nonlinear speed degradation starting at the critical density (ρ_{cr}).

As illustrated in Figure 2.2, this relationship is monotonically decreasing.

- **Free-Flow Condition:** At low densities ($\rho \approx 0$), interactions between vehicles are negligible. Drivers travel at their desired speed, limited only by the physical road geometry or speed limits. This constant value is the **Free-Flow Speed** (v_f).

- **Deceleration:** As density increases, the spacing between vehicles (headway) decreases. To maintain a safe stopping distance, drivers are forced to reduce their speed.

- **Jam Condition:** As density approaches the maximum value (ρ_{jam}), the speed asymptotically approaches zero.

This behavior is critical because the reduction in speed is exactly what causes the flow to drop once density exceeds the critical threshold.

2.3 Bottlenecks, Capacity Drop, and Shockwave Propagation

The Fundamental Diagram described in the previous section represents the equilibrium states of a homogeneous road section. However, in real-world freeway networks, homogeneity is the exception. The operational limits of the infrastructure are defined by specific constraints known as **bottlenecks**.

2.3.1 Bottlenecks and Traffic Breakdown

A bottleneck is defined as any location on the roadway where the downstream capacity is lower than the upstream capacity (e.g., lane drops, tunnels, or tight curves) or where the arriving demand exceeds the available capacity (e.g., high on-ramp flows).

When the incoming flow exceeds the bottleneck capacity, the system undergoes a phenomenon known as **Traffic Breakdown**. In terms of the Fundamental Diagram, the traffic state jumps from the stable "free-flow" branch to the unstable "congested" branch. This transition is not merely a reduction in speed; it represents a fundamental change in the physics of the traffic stream.

2.3.2 Shockwave Phenomena

The transition between free-flow and congestion is not instantaneous over space; it propagates through the traffic stream as a **Shockwave**. In macroscopic traffic theory, a shockwave represents the boundary discontinuity between two different density states, this is observed when brake lights propagate upstream.

Mathematically, if the density downstream of a bottleneck is high and the density upstream is low, the interface between them moves backward. This **backward forming shockwave** is the mechanism by which queues grow in space. The speed of this wave depends on the difference in flow and density between the two states. Control strategies, such as the one proposed in this thesis, aim to intercept these waves and dissipate them before they become severe.

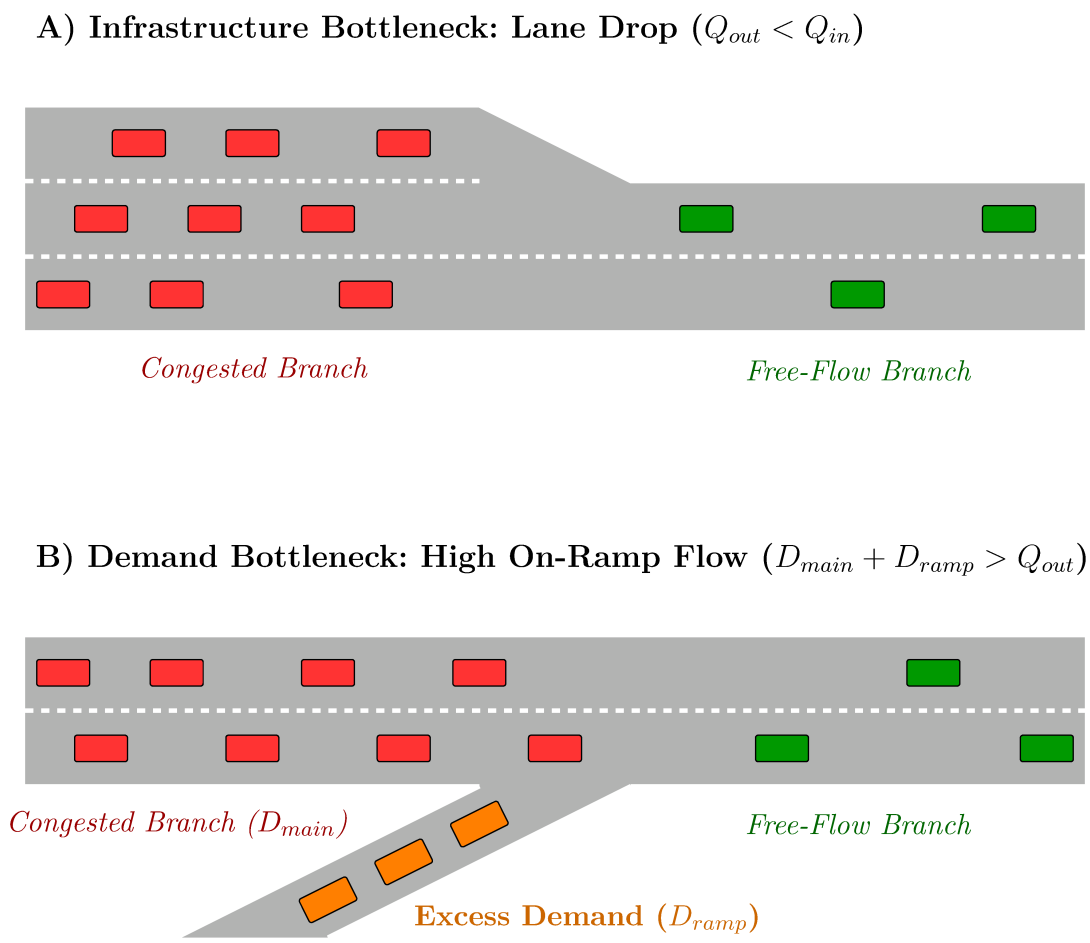


Figure 2.3: Visual representation of the two primary mechanisms of traffic breakdown: (A) an infrastructure bottleneck caused by a physical reduction in capacity, such as a lane drop, and (B) a demand bottleneck caused by overwhelming incoming flow from an on-ramp merging into the mainline. Both scenarios trigger a transition from the free-flow regime to the congested regime, creating a backward-propagating shockwave.

2.3.3 The Capacity Drop

The most critical consequence of traffic breakdown is the **Capacity Drop** phenomenon [33] [34] [28] [26]. Classical theory might suggest that a bottleneck, once congested, continues to discharge vehicles at its maximum capacity (q_{max}). However, empirical studies and second-order models reveal that this is not the case. The discharge flow rate from a congested queue is significantly lower than the road's free-flow capacity.

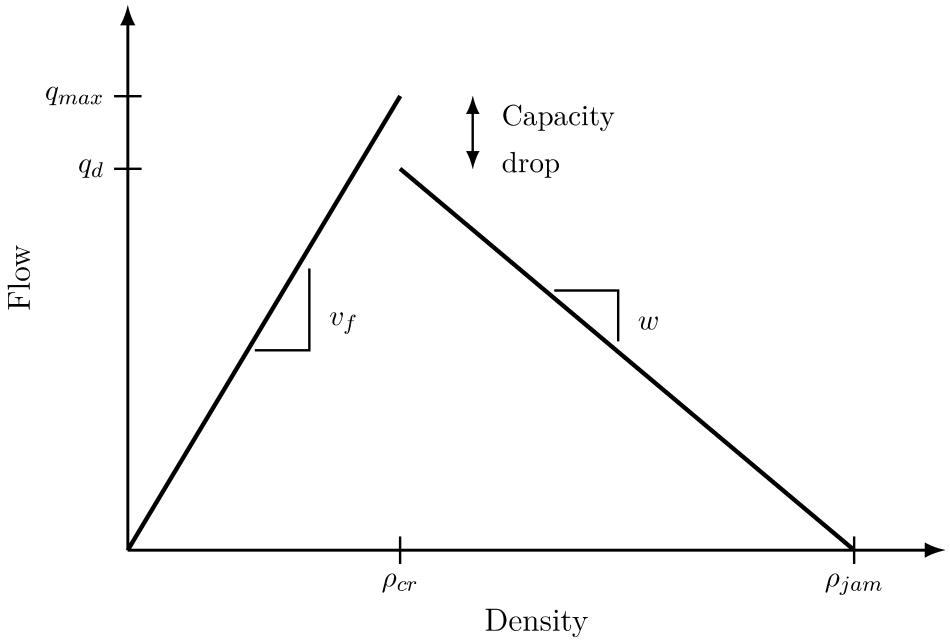


Figure 2.4: A triangular macroscopic fundamental diagram illustrating the capacity drop phenomenon. The graph distinguishes between the theoretical free-flow capacity (q_{max}) and the reduced queue discharge flow (q_d), alongside their respective characteristic wave speeds (v_f and w).

Typically, this drop ranges from 5% to 30% of the nominal capacity [35]. The physical mechanism underlying this phenomenon is rooted in microscopic driver kinematics during the transition from congested to free-flow states. When vehicles attempt to accelerate out of a stop-and-go wave to discharge through the bottleneck, human drivers inherently exhibit sluggish acceleration. Due to perception-reaction latency and the psychological need to establish larger safety margins at higher speeds, drivers maintain significantly greater time headways than they would under steady-state cruising conditions. This delayed vehicle-following behavior generates an acceleration void at the head of the queue, systematically restricting the bottleneck's actual discharge rate to a value substantially below its theoretical maximum capacity.

2.3.4 Hysteresis and Control Implications

The Capacity Drop creates a **hysteresis** loop in the traffic state:

- **Entering Congestion:** The transition from a stable free-flow regime to a congested state occurs rapidly when transient demand spikes exceed the physical

capacity of the bottleneck, triggering the immediate formation of a backward-propagating shockwave.

- **Exiting Congestion:** Conversely, the network exhibits a structurally delayed recovery when attempting to return to free-flow conditions. This asymmetry is governed by the *queue discharge rate* ($q_{discharge}$), formally defined as the maximum flow at which vehicles can accelerate and depart from an established upstream queue. Because q_d is strictly lower than the pre-congestion capacity, the accumulated queue dissipates at a significantly slower pace than it formed. Consequently, even if upstream demand recedes to the original capacity level, the bottleneck remains actively congested because the effective throughput of the system has fundamentally shrunk.

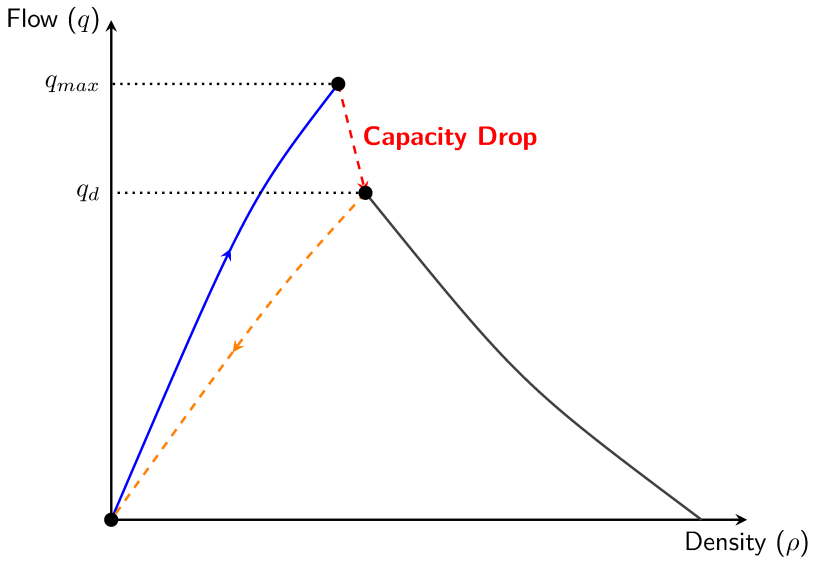


Figure 2.5: Fundamental diagram illustrating the traffic hysteresis loop. The capacity drop forces the system into a congested state, where at the same density the flux is reduced and recovery is strictly governed by the lower queue discharge rate ($q_d < q_{max}$).

This phenomenon provides the theoretical justification for the **Moving Bottleneck** strategy proposed in this thesis. The objective of using CAVs to regulate flow *upstream* is to prevent the critical bottleneck from activating. By metering the flow before it reaches the constraint, the system ensures the bottleneck operates at its higher free-flow capacity (q_{max}) rather than degrading to the dropped discharge rate (q_d).

2.4 Macroscopic Traffic Modeling

Having defined the fundamental traffic variables and their equilibrium relationships, it is necessary to introduce a mathematical framework capable of describing their evolution over time and space. While the Fundamental Diagram provides a static map of possible traffic states, it does not explain how the system transitions from one state to another (e.g., how a queue forms and propagates upstream).

To address this, Traffic Control Engineering relies on dynamic macroscopic models [26]. These models are analogous to fluid dynamics, treating traffic flow as a continuous medium. They are classified based on the number of differential equations used:

- **First-Order Models:** Based on the conservation of mass (vehicles) and a static speed-density relationship. They are computationally efficient and robust for simulating shockwaves.
- **Second-Order Models:** Introduce a second equation for speed dynamics (momentum), allowing for acceleration effects and capacity drop phenomena.

For the control strategy proposed in this thesis, a first-order approach is sufficient to capture the essential dynamics of congestion propagation.

2.4.1 The LWR Model

The cornerstone of macroscopic traffic theory is the **Lighthill-Whitham-Richards (LWR)** model. It was developed independently by Lighthill and Whitham [31] and Richards [36] in the mid-1950s. The model is built upon a single conservation law, assuming that vehicles are neither created nor destroyed along a road section (excluding on-ramps and off-ramps).

Consider a one-dimensional road segment of length dx over a time interval dt. The principle of *Conservation of Vehicles* states that the change in the number of vehicles within the segment must equal the difference between the flow entering and the flow exiting. Mathematically, this balance is expressed by the continuity equation:

$$\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \tag{2.3}$$

Where:

- $\rho(x,t)$ is the density at location x and time t.

- $q(x,t)$ is the flow at location x and time t.

Equation 2.3 alone contains two unknowns (ρ and q). To close the system, the LWR model assumes that traffic is always in equilibrium, meaning the speed adapts instantaneously to the local density. This allows us to substitute the fundamental relationship $q = Q(\rho)$ directly into the continuity equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = 0 \tag{2.4}$$

By applying the chain rule ($\frac{\partial Q}{\partial x} = \frac{dQ}{d\rho} \cdot \frac{\partial \rho}{\partial x}$), we obtain the quasi-linear hyperbolic partial differential equation (PDE):

$$\frac{\partial \rho}{\partial t} + Q'(\rho) \cdot \frac{\partial \rho}{\partial x} = 0 \tag{2.5}$$

The term $Q'(\rho) = \frac{dQ}{d\rho}$ represents the slope of the tangent to the Fundamental Diagram. Physically, this value corresponds to the **wave speed**—the velocity at which small disturbances (or information) travel through the traffic stream.

The LWR model is particularly powerful because it admits discontinuous solutions known as **shockwaves**. This mathematical feature allows it to accurately replicate the sudden formation of queues (e.g., when fast-moving traffic hits a bottleneck), making it the standard choice for designing freeway traffic controllers.

2.4.2 The Cell Transmission Model (CTM)

While the LWR model provides an elegant analytical framework, the partial differential equation (PDE) presented in Eq. 2.5 is difficult to solve analytically for complex boundary conditions and variable road geometries. To simulate traffic on a computer, the continuous time and space of the LWR model must be discretized.

The **Cell Transmission Model (CTM)**, developed by Daganzo [37], is the most widely adopted numerical scheme for this purpose. Mathematically, it is a Godunov-type finite difference approximation of the LWR PDE. It assumes a triangular Fundamental Diagram and tracks the number of vehicles in discrete road segments (cells) at discrete time steps. Due to its computational efficiency and ability to accurately reproduce shockwaves, CTM serves as the simulation engine for the control strategies developed in this thesis.

Discretization and Definitions

The road section is divided into M homogeneous segments, referred to as **cells**, indexed by $i = 1, \ldots, M$. Time is discretized into steps of duration T, indexed by

$k = 0, 1, 2, \ldots$. The state of the system is defined by the variable $\rho_i(k)$, representing the density of vehicles (occupancy) in cell i at time step k.

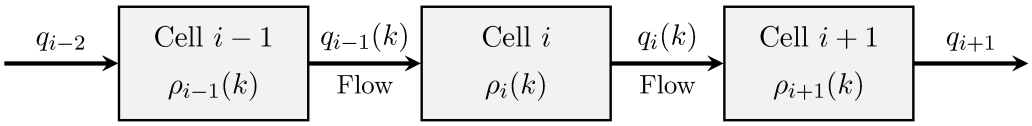


Figure 2.6: Schematic representation of the Cell Transmission Model

Each cell i is characterized by fixed physical parameters:

- L_i: The length of the cell (km).
- v_f: The free-flow speed (km/h).
- w: The backward congestion wave speed (km/h).
- q_i: The maximum flow capacity (veh/h) normalized to the time step ($Q_i \cdot T$).
- ρ_{jam}: The jam density (storage capacity).

The Transmission Logic: Demand and Supply

The model assumes a **Triangular Fundamental Diagram**, where the equilibrium flow is determined by:

$$q(\rho) = \min\left(v_f \cdot \rho,\ w \cdot (\rho_{jam} - \rho),\ q_{max}\right) \tag{2.6}$$

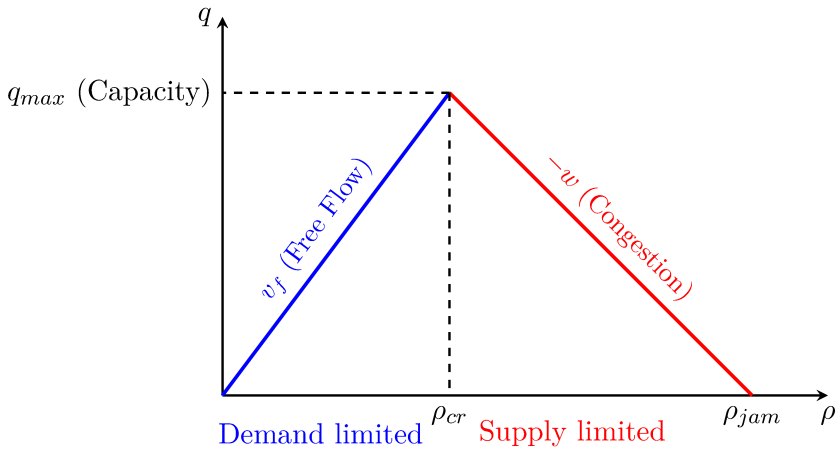


Figure 2.7: The Triangular Fundamental Diagram assumed by the CTM..

This assumption allows the CTM to decouple the flow calculation into two distinct functions: **Demand** (the upstream ability to send flow, corresponding to the

left side of the triangle) and **Supply** (the downstream ability to receive flow, corresponding to the right side).

For every time step k, we calculate two quantities at the boundary between cell i and cell $i+1$:

1. **Sending Function (Demand) $D_i(k)$:** This represents the maximum number of vehicles that *want* to leave cell i. It is limited by the number of cars currently in the cell and the capacity of the cell's outlet.

$$D_i(k) = \min\left(v_f \cdot \rho_i(k),\ q_i\right) \tag{2.7}$$

2. **Receiving Function (Supply) $S_{i+1}(k)$:** This represents the maximum number of vehicles that cell $i+1$ *can accept*. It is limited by the available empty space (storage) in the cell and its inlet capacity.

$$S_{i+1}(k) = \min\left(q_{i+1},\ w \cdot (\rho_{jam,i+1} - \rho_{i+1}(k))\right) \tag{2.8}$$

The term $(\rho_{jam,i+1} - \rho_{i+1})$ represents the "holes" or empty space in the downstream cell. The ratio w/v_f accounts for the speed at which these holes propagate backward into the traffic stream.

Flow Computation and State Update

The actual flow $q_i(k)$ moving from cell i to $i+1$ is determined by the minimum of what can be sent and what can be received. This is analogous to a fluid constriction: flow is limited by either the upstream pump or the downstream valve.

$$q_i(k) = \min\left(D_i(k),\ S_{i+1}(k)\right) \tag{2.9}$$

Once the boundary flows are calculated for all cells, the state of the system is updated using the conservation equation (mass balance):

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i}\left(q_{i-1}(k) - q_i(k)\right) \tag{2.10}$$

This recursive process allows the simulation to evolve over time, naturally capturing complex phenomena:

- **Free-flow:** When $D_i < S_{i+1}$, flow is determined by the upstream density.
- **Congestion:** When $D_i > S_{i+1}$, flow is throttled by the downstream bottleneck, causing n_i to accumulate. This accumulation is the CTM equivalent of a queue forming.

Numerical Stability (CFL Condition)

For the discrete CTM to accurately approximate the continuous physics of traffic, the simulation time step T and cell length L_i cannot be chosen arbitrarily. They must satisfy the **Courant-Friedrichs-Lewy (CFL)** condition:

$$v_f \cdot T \leq L_i \tag{2.11}$$

Physically, this ensures that a vehicle traveling at free-flow speed cannot "skip" an entire cell in a single time step. It guarantees that traffic propagates cell-by-cell without numerical dispersion.

Baseline CTM Simulation: Single and Multi-Lane Scenarios

To evaluate the fundamental behavior of the model, a base scenario is first presented. This simulation considers a single-lane highway stretch of 60 km over a 3-hour horizon. The upstream demand follows a rectangular profile with a peak of 2200 veh/h, matching the nominal capacity of the cells. A bottleneck is induced at cell 40 by reducing its capacity to 1700 veh/h.

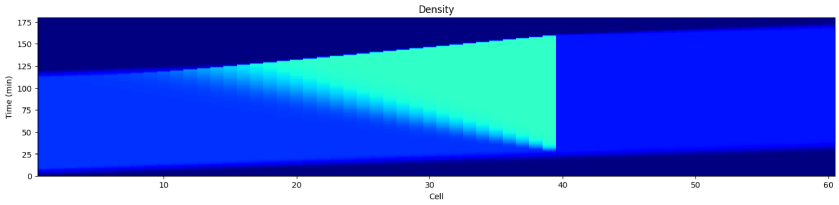


Figure 2.8: Spatiotemporal density evolution in a single-lane CTM. The capacity reduction in cell 40 triggers a clear backward-propagating shockwave as demand exceeds the bottleneck throughput.

As shown in Figure 2.8, the CTM correctly identifies the transition from free-flow to congestion, with the queue tail moving upstream. Building upon this, the model is tested on a three-lane highway where the left lane is subjected to a complete blockage between minutes 35 and 95.

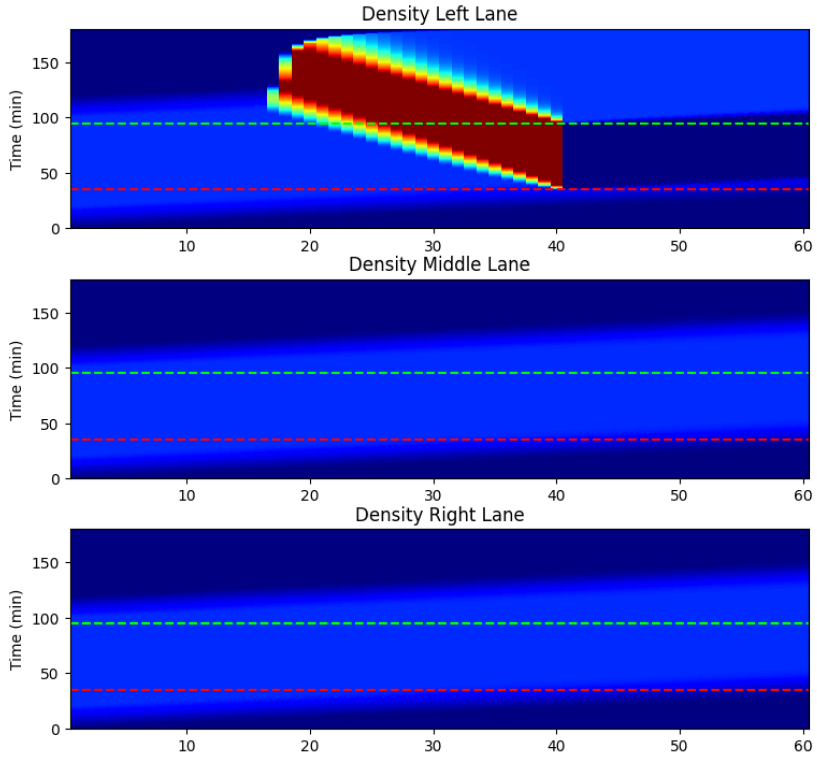


Figure 2.9: Spatiotemporal density heatmap of the classic CTM applied to a three-lane highway. The left lane experiences a complete blockage, while the middle and right lanes remain completely unaffected due to the lack of lateral interaction logic.

As illustrated in Figure 2.9, while the original formulation successfully captures the shockwave in the affected lane, a major limitation emerges: the classic CTM processes each lane as an isolated pipe. It lacks the ability to represent lateral interactions, such as lane-changing maneuvers to avoid the blockage, or complex phenomena like the capacity drop. Consequently, to realistically simulate multilane traffic dynamics, the baseline model requires the advanced extensions developed in the following sections.

Chapter 3

Modeling the Multi-Lane Freeway Environment

The theoretical framework established in Chapter 2 provides the fundamental laws governing traffic flow. However, the standard CTM assumes a homogeneous, single-lane link, which prevents the analysis of complex lateral interactions such as lane-changing friction and asymmetric congestion.

This chapter details the formulation of the **Multi-Lane Cell Transmission Model** [38] developed for this thesis. This model serves as the plant, the virtual reality representing the physical highway system, upon which the control logic described in Chapter 4 will operate. The proposed environment extends the classical theory by explicitly modeling the parallel geometry of a multi-lane freeway, introducing dynamic logic for lateral mass exchange, and incorporating a density-dependent capacity drop mechanism.

3.1 Multi-Lane Network Topology

The physical domain considered in this study is a unidirectional freeway segment of length L_{total}. Unlike single-lane models that aggregate capacity, this environment simulates the highway as a set of $N_{lanes} = 3$ parallel, interacting links, indexed by $l \in \{A, B, C\}$. This topology allows the system to capture distinct traffic states in each lane, enabling the simulation of scenarios where specific lanes are blocked or congested while others remain free-flowing.

The spatial domain is discretized into N homogeneous segments, referred to as **cells**, each with a length of L_i. The temporal evolution of the system is computed at discrete time steps of duration T, chosen to satisfy the Courant-Friedrichs-Lewy

(CFL) condition.

The specific configuration of the lanes is defined as follows:

- **Lane C:** The leftmost lane, primarily used for overtaking and high-speed travel.
- **Lane B:** The middle lane, subject to interactions from both sides.
- **Lane A:** The rightmost lane, typically utilized by slower traffic and vehicles entering or exiting the highway.

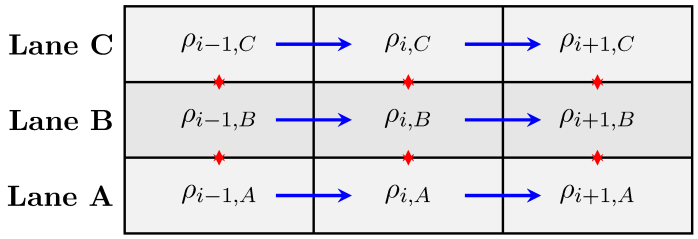


Figure 3.1: Schematic representation of the Multi-Lane CTM topology. Blue arrows represent longitudinal flow Φ_{long}, and red arrows represent lateral interactions Φ_{lat}.

3.1.1 The Extended Conservation Law

The fundamental difference between the standard CTM and the multi-lane formulation lies in the conservation of mass. In the standard model, the vehicle count in a cell changes only due to longitudinal flow (upstream inflow minus downstream outflow). In a multi-lane environment, vehicles are also conserved laterally.

For any given cell i in lane l, the change in the number of vehicles is the net result of longitudinal transport along the lane and lateral transport from adjacent neighbors. The state update equation is thus generalized as:

$$\rho_{i,l}(k+1) = \rho_{i,l}(k) + \frac{T}{L_i}\left[\Phi_{long}(i,l) + \Phi_{lat}(i,l)\right] \tag{3.1}$$

Where:

- $\Phi_{long}(i,l) = q^{in}_{i-1,l}(k) - q^{out}_{i,l}(k)$ represents the net longitudinal flow. This is calculated using the standard demand-supply logic described in Chapter 2, applied independently to each lane.
- $\Phi_{lat}(i,l)$ represents the net lateral flow. It is the sum of vehicles entering lane l from its neighbors minus the vehicles leaving lane l to its neighbors.

By decoupling these two transport mechanisms, the model ensures that mass is strictly conserved: a vehicle leaving Lane A to enter Lane B is instantaneously subtracted from $\rho_{i,A}$ and added to $\rho_{i,B}$, preventing the creation or destruction of vehicles during maneuvers.

3.2 The Lane Changing Model

In a multi-lane environment, the coupling between parallel cells is modeled through a dynamic lateral flow mechanism. Unlike microscopic models that track individual gaps, this macroscopic approach determines the net flow of vehicles $q^{LC}_{i,l\to m}$ migrating from a source lane l to a target lane m.

This process is governed by three distinct factors: the local density incentives, the cooperative pressure from blocked lanes, and the dynamic physical limits of the maneuver.

3.2.1 Density-Based Incentives

The primary driver for lane changing is the density gradient. Vehicles are motivated to migrate from a highly congested lane to a less congested neighbor to maximize their speed. The baseline incentive for moving from lane l to m in cell i, denoted as $I_{l\to m}(i)$, is proportional to the normalized density difference:

$$I_{l\to m}(i) = K_{inc} \cdot \frac{\rho_{i,l}(k) - \rho_{i,m}(k)}{\rho_{jam}} \tag{3.2}$$

Where K_{inc} is a calibration parameter representing the sensitivity of drivers to density variations. A positive value indicates a net desire to move towards lane m.

3.2.2 Cooperative Pressure and Blockage Logic

A specific feature of this model is the handling of bottlenecks (e.g., incidents) through a Look-Ahead and Cooperative logic.

Mandatory Change

If a downstream cell $i+1$ in lane l is blocked ($q_{max} \approx 0$), the incentive logic is overridden to force an exit. The incentive $I_{l\to m}(i)$ is set to the maximal value (K), while entry into the blocked lane is inhibited ($I_{m\to l}(i+1) = -1$).

Cooperative Pressure propagation

In high-density scenarios, a simple incentive based on density is insufficient to clear a blocked lane. To address this, the model implements a **Cooperative Pressure** mechanism. If vehicles in the leftmost lane (Lane C) are forced to merge into the center lane (Lane B) due to a blockage, this pressure is mathematically propagated to Lane B drivers. Physically, this simulates drivers in the center lane proactively moving to the right to create space for vehicles merging from the left, facilitating the clearing of the bottleneck.

The incentive for Lane B drivers to move to Lane A is augmented by the merging desire of Lane C:

$$I_{B\to A}^{coop}(i) = I_{B\to A}(i) + K_{coop} \cdot \max(0, I_{C\to B}(i)) \tag{3.3}$$

Where K_{coop} is a gain factor.

3.2.3 Dynamic Permeability Limits

Lane changing becomes increasingly difficult as traffic density rises due to the lack of available gaps. To replicate this friction, we define a **Dynamic Permeability Function** $Q_{LC}^{max}(\rho)$ that scales linearly with the density of the *target* lane.

The maximum allowable lateral flow drops from a nominal value (Q_{LC}^{nom}) in free-flow to a minimal residual value (Q_{LC}^{min}) in jam conditions:

$$Q_{LC}^{max}(\rho) = \begin{cases} Q_{LC}^{nom} & \text{if } \rho \le \rho_{cr} \\ Q_{LC}^{nom} - (Q_{LC}^{nom} - Q_{LC}^{min}) \cdot \frac{\rho - \rho_{cr}}{\rho_{jam} - \rho_{cr}} & \text{if } \rho > \rho_{cr} \end{cases} \tag{3.4}$$

This ensures that lateral flow effectively ceases when the target lane is gridlocked, preventing unrealistic vehicle teleportation.

3.2.4 Lateral Flow Computation and State Update

The actual lateral flow $q_{i,l\to m}^{LC}(k)$ is computed by bounding the desired flow with the physical limits and the target supply:

$$q_{i,l\to m}^{LC} = \min\left(\underbrace{v_f \rho_{i,l} \cdot \max(0, I_{l\to m})}_{\text{Desired}}, \quad \underbrace{Q_{LC}^{max}(\rho_{i,m})}_{\text{Limit}}, \quad \underbrace{S_{i,m}}_{\text{Supply}} \right) \tag{3.5}$$

Mass Conservation and Numerical Stability: The conservation of mass is primarily enforced by the definition of the Demand, eq. 2.7, and Supply, eq.

2.8, functions, which mathematically constrain the flows to the available number of vehicles (upstream) and the available physical space (downstream).

Furthermore, the simulation time step T is chosen to strictly satisfy the **Courant-Friedrichs-Lewy (CFL)** condition, eq. 2.11. This ensures that the total displacement of vehicles in a single time step generally remains within the physical bounds of the cell.

However, to guarantee that the state variables remain strictly within the physical domain $\mathcal{D} = [0, \rho_{jam}]$ despite finite-precision floating-point arithmetic, a final saturation step is applied to the updated densities:

$$\rho_{i,l}(k+1) = \max\left(0,\ \min\left(\rho_{i,l}^{calc},\ \rho_{jam}\right)\right) \tag{3.6}$$

Where $\rho_{i,l}^{calc}$ is computed by applying the conservation law defined in Section 3.1 using the computed longitudinal and lateral fluxes. This step eliminates negligible numerical noise without affecting the macroscopic conservation properties of the model.

3.3 Modeling the Capacity Drop

A fundamental limitation of the standard first-order CTM is its inability to capture the *Capacity Drop* phenomenon. As detailed in 3.3, the empirical observation that the discharge flow from a bottleneck decreases significantly once traffic breaks down and a queue forms.

In the standard CTM framework described in 2.4.2, the capacity of a cell is treated as a constant parameter q_{max}. Even during severe congestion ($\rho > \rho_{cr}$), the model assumes that the bottleneck can continue to discharge vehicles at the maximum rate q_{max}, provided there is sufficient downstream space. This assumption contradicts real-world observations, where the "discharge capacity" of an active bottleneck is typically 5% to 30% lower than its free-flow capacity [35]. This reduction creates a hysteresis loop: once a traffic jam forms, recovering to free-flow conditions requires the demand to drop significantly below the original capacity level.

To accurately simulate this inertia, which is the primary target for the control strategies developed in this thesis, we adopt the **Density-Dependent Capacity** formulation proposed by Piacentini et al. [4].

3.3.1 The Density-Dependent Capacity Formulation

The proposed model modifies the Fundamental Diagram by making the maximum flow capacity $Q_{i,l}^{max}$ a function of the instantaneous density state $\rho_{i,l}(k)$.

The capacity function is defined piecewise:

- **Free-Flow Regime ($\rho \leq \rho_{cr}$):** The cell operates at its nominal design capacity $C_{nominal}$.
- **Congested Regime ($\rho > \rho_{cr}$):** The capacity is no longer constant. Instead, it decays linearly as the density increases towards the jam density.

This behavior is governed by the drop parameter $\alpha \in [0, 1]$. The dynamic capacity $Q_{i,l}^{max}(k)$ is computed at each simulation step k as follows:

$$Q_{i,l}^{max}(k) = \begin{cases} C_{nominal} & \text{if } \rho_{i,l}(k) \leq \rho_{cr} \\ C_{nominal} \cdot \left[1 + (\alpha - 1)\frac{\rho_{i,l}(k) - \rho_{cr}}{\rho_{jam} - \rho_{cr}}\right] & \text{if } \rho_{i,l}(k) > \rho_{cr} \end{cases} \tag{3.7}$$

Where:

- $C_{nominal}$ is the physical capacity of the lane (set to 2200 veh/h in this study).
- ρ_{cr} and ρ_{jam} are the critical and jam densities, respectively.
- The term $\frac{\rho - \rho_{cr}}{\rho_{jam} - \rho_{cr}}$ represents the "congestion fraction," ranging from 0 (at breakdown) to 1 (at standstill).

3.3.2 Physical Interpretation of α

The parameter α explicitly defines the severity of the capacity drop. It represents the ratio of the discharge capacity at jam density relative to the nominal capacity.

- If $\alpha = 1$, the model reverts to the standard CTM (constant capacity C).
- If $\alpha < 1$, the capacity decreases in congestion.

In this implementation, we define $\alpha = 0.7$. This implies that as the traffic density approaches ρ_{jam}, the effective capacity of the road drops to 70% of its nominal value.

Mathematically, this modification transforms the demand function of the CTM. Instead of the standard trapezoidal shape, the demand function slopes downward in the congested region, as illustrated in Figure 3.2. This ensures that a congested cell sends fewer vehicles downstream than a free-flowing cell, effectively modeling the friction and inefficiency of stop-and-go waves.

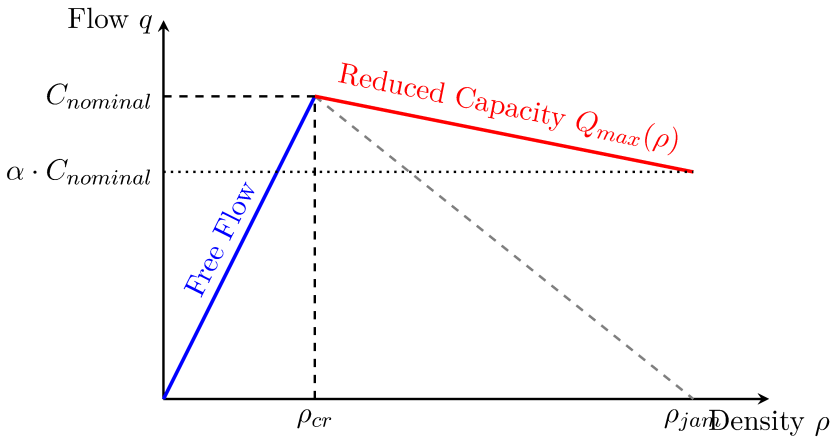


Figure 3.2: The modified Fundamental Diagram incorporating Capacity Drop. Adapted from [4].

3.4 Simulation Scenario

To validate the multi-lane dynamics and capacity drop, a stress test scenario was designed: a three-lane highway operating under heavy demand and subjected to a severe, temporary bottleneck.

Each lane is characterized by identical physical parameters derived from standard freeway characteristics. The state of the system is defined by the density matrix $\boldsymbol{\rho}(k) \in \mathbb{R}^{N \times N_{lanes}}$, where $\rho_{i,l}(k)$ represents the density in cell i of lane l. The fundamental parameters used in the simulation are listed in Table 3.1.

Parameter	**Symbol**	**Value**	**Unit**
Free-Flow Speed	v_f	120	km/h
Congestion Wave Speed	w	17.15	km/h
Nominal Capacity (per lane)	Q_{max}	2200	veh/h
Jam Density (per lane)	ρ_{jam}	140	veh/km
Critical Density	ρ_{cr}	17.5	veh/km
Cell Length	L_i	1.0	km
Simulation Time Step	T	1	s
Number of Cells	N	60	-
Lane Changing Incentive	K_{inc}	100	-
Cooperative Incentive	K_{coop}	0.0085	-

Table 3.1: Physical parameters of the simulated freeway environment.

This setup serves as the *Open Loop* baseline (uncontrolled case) against which the control strategies developed in Chapter 4 will later be evaluated.

3.4.1 Temporal Domain and Traffic Demand

The simulation covers a time horizon of $T_{sim} = 3$ hours. The traffic demand enters the highway at the upstream boundary (Cell 1) and follows a **Trapezoidal Profile**. This shape is chosen to evaluate the system's performance across three distinct regimes: the onset of congestion, the sustained peak, and the recovery phase.

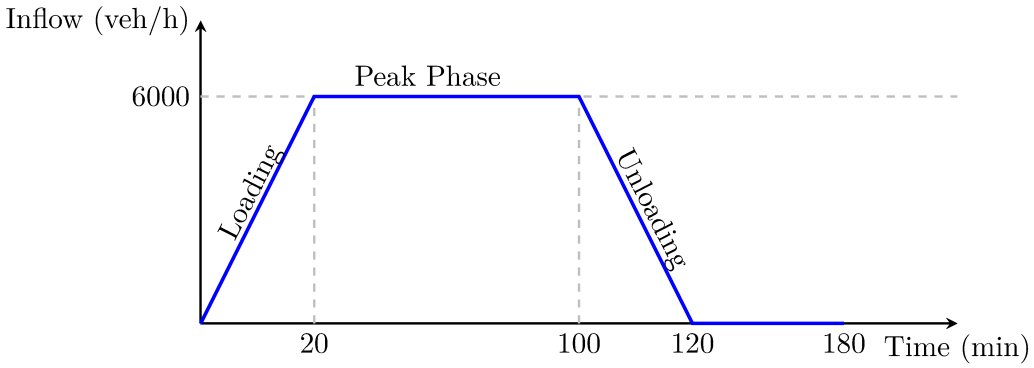


Figure 3.3: The trapezoidal inflow demand profile applied to the upstream boundary.

The inflow profile $q_{in}(t)$ is identical for all three lanes and evolves as follows:

1. **Loading Phase** ($0 \rightarrow 20$ **min):** Demand increases linearly from 0 to a peak value of 2000 veh/h/lane.

2. **Peak Phase** ($20 \rightarrow 100$ **min):** Demand remains constant at 2000 veh/h/lane. With 3 lanes, this yields a total inflow of 6000 veh/h, which loads the network to approximately 90% of its nominal capacity ($3 \times 2200 = 6600$ veh/h).

3. **Unloading Phase** ($100 \rightarrow 120$ **min):** Demand decreases linearly back to 0.

4. **Dissipation Phase** ($120 \rightarrow 180$ **min):** Zero inflow, allowing the network to clear any remaining congestion.

3.4.2 The Bottleneck Event

To trigger complex lateral interactions and force the activation of the capacity drop mechanism, a temporary lane closure is simulated.

- **Location:** The incident occurs in **Lane C** (leftmost lane) at cells $i = 40$ and $i = 41$, corresponding to the segment between km 40 and km 42.
- **Severity:** The capacity of these cells is reduced to 0 veh/h, representing a complete blockage (e.g., a severe accident or stalled vehicle).
- **Duration:** The incident is active from $t = 35$ min to $t = 95$ min.

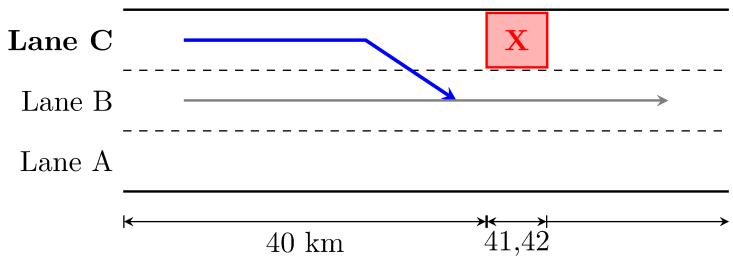


Figure 3.4: Schematic of the simulation scenario. A complete blockage in Lane C at km 41 forces traffic to merge into Lane B, creating a bottleneck.

Expected System Dynamics

When the incident activates at $t = 35$ min, vehicles in Lane C encounter a hard block. This forces a massive migration of traffic from Lane C to Lane B via the mandatory lane-changing logic (Section 3.2).

Consequently, the density in Lane B rises rapidly above its critical threshold ρ_{cr}. This triggers the **Capacity Drop** mechanism (Section 3.3) in Lane B, reducing its effective throughput. If the density in Lane B becomes critical, congestion will further spill over into Lane A.

3.5 Performance Bounds and Model Comparison

To evaluate the impact of the newly introduced lateral interactions and the capacity drop phenomenon, it is necessary to establish the absolute performance boundaries of the simulated multi-lane highway network.

The incident scenario (a complete left-lane closure between minutes 35 and 95) was simulated under two distinct multi-lane conditions. These baselines define the theoretical lower bound and the realistic upper bound, effectively isolating the specific delay caused by the capacity drop. Finally, these results are directly contrasted with the classic CTM behavior demonstrated in the previous sections.

3.5.1 Ideal Lower Bound

The first scenario represents a theoretical ideal. In this simulation, the incident physically blocks the leftmost lane, forcing vehicles to change lanes, but the capacity drop mechanism is artificially disabled (setting the drop parameter $\alpha = 1.0$).

In this state, the bottleneck discharges vehicles at the maximum theoretical rate allowed by the remaining two open lanes. The queue that forms is purely a geometric consequence of the reduced road space, rather than a systemic breakdown of traffic flow.

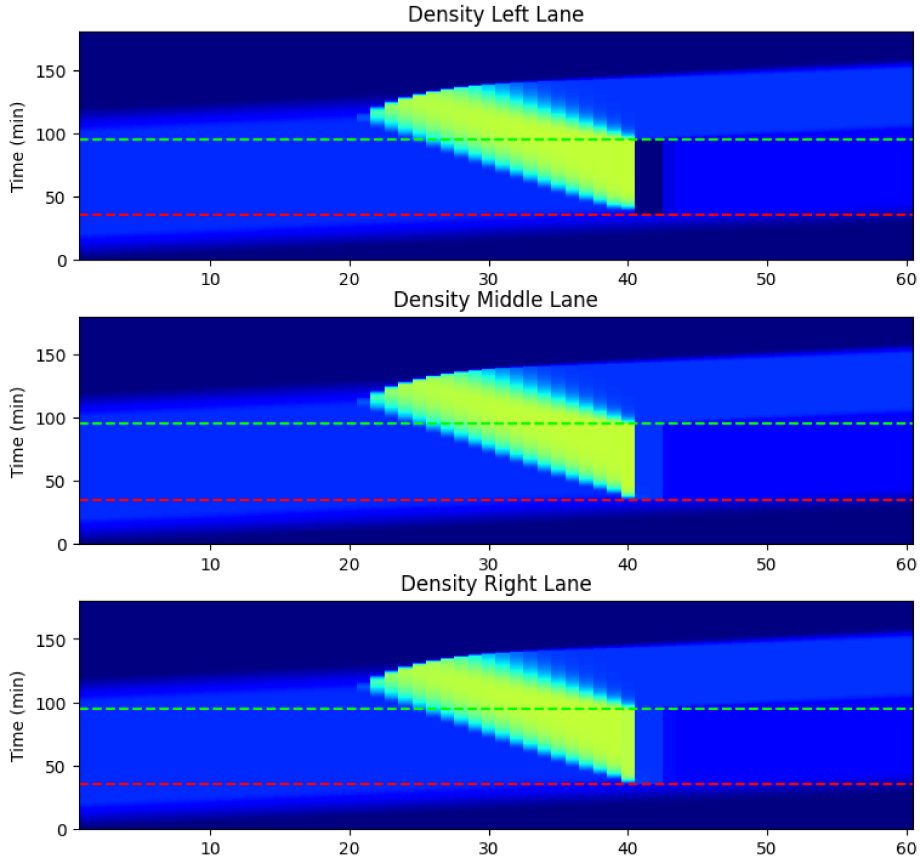


Figure 3.5: Spatiotemporal evolution of traffic density for the theoretical ideal scenario, where the capacity drop phenomenon is disabled ($\alpha = 1.0$).

- **Total Travel Time (TTT):** 6935 veh·h

In this ideal scenario, the absence of the capacity drop mechanism allows the bottleneck to discharge vehicles at the absolute maximum rate sustainable by the two remaining open lanes. Consequently, this value represents the **absolute minimum TTT** achievable by the system for this specific incident profile. Furthermore,

this scenario clearly illustrates the fundamental shift from the classic CTM formulation presented in figure 2.9. Instead of vehicles being artificially trapped in the blocked lane while adjacent lanes remain empty, the implemented lateral logic enables dynamic lane-changing. Vehicles actively diverge from the obstructed left lane and distribute themselves into the middle and right lanes. While this maneuver maximizes the utilization of the remaining road capacity and significantly reduces the overall delay compared to an isolated-lane model, it inevitably propagates the congestion laterally.

3.5.2 Realistic Baseline

The second scenario represents the realistic behavior of human-driven traffic within the advanced multi-lane framework. Here, the capacity drop mechanism is active.

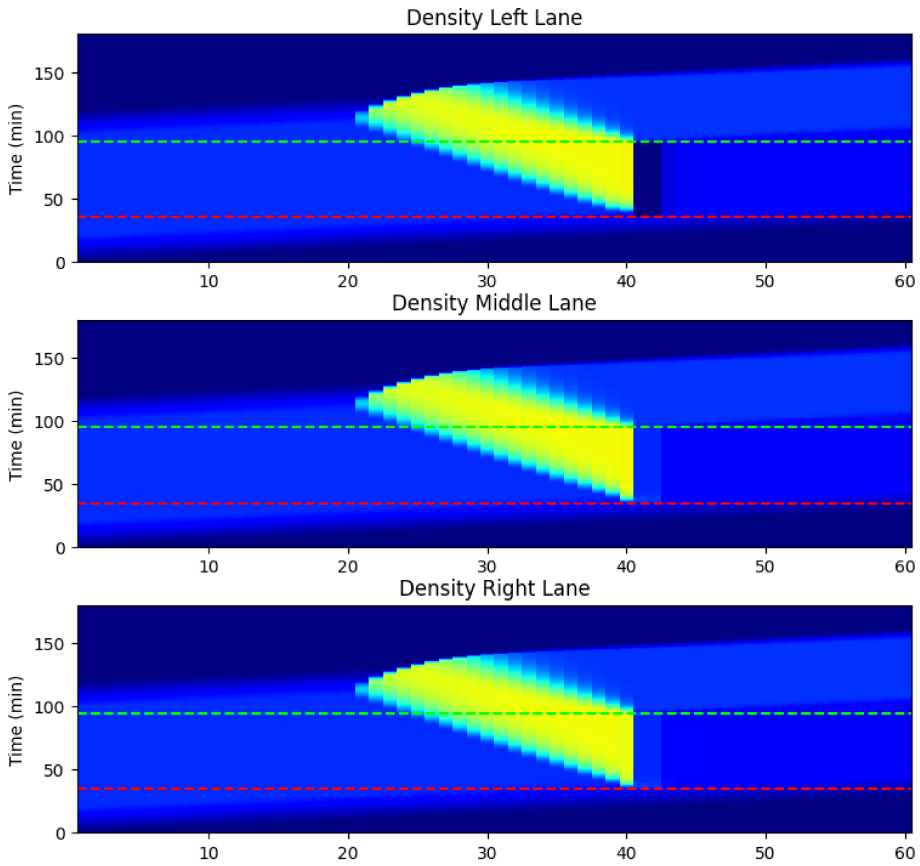


Figure 3.6: Spatiotemporal density showing stronger congestion due to the active capacity drop and lane-changing friction.

- **Total Travel Time (TTT):** 7298 veh·h

When the incident occurs, the traffic flow breaks down and the discharge capacity of the bottleneck plummets compared to its nominal value. Visually, this degradation is evident in the spatiotemporal heatmap: the brighter yellow regions, when compared to the ideal scenario without capacity drop, indicate a significantly higher concentration of vehicle density. This localized intensification is exclusively driven by the introduction of the capacity drop mechanism, which restricts the throughput of bottleneck cells. Consequently, this starvation of the downstream section forces a stronger, backward-propagating shockwave that takes significantly longer to dissipate even after the incident is cleared.

3.5.3 The Control Objective

Comparing these two baselines reveals that the systemic inefficiency imposes a penalty of **363 veh·h** onto the network.

The primary objective of the deployment of CAVs is to recover this specific margin. By proactively regulating the upstream traffic flow, these control action aim to prevent the local density at the bottleneck from reaching the critical breakdown state that triggers the capacity drop.

Chapter 4

Traffic Control Strategy

Considering the model introduced in Chapter 3, this chapter introduces the control framework designed to mitigate congestion. While traditional traffic management relies on static infrastructure to regulate flow, this thesis proposes a dynamic approach using CAVs as mobile actuators.

This chapter details the conceptual shift from fixed to mobile actuation, defines the specific control objective regarding the TTT, and formulates the sequential Model Predictive Control (MPC) algorithm used to coordinate the fleet.

The transition from human-driven traffic to mixed traffic flows introduces new degrees of freedom for control engineering [39] [40]. In this study, we do not treat CAVs merely as vehicles that drive themselves more efficiently; rather, we treat them as active, controllable agents capable of influencing the macroscopic behavior of the entire traffic stream. By adjusting their longitudinal speed, these vehicles can act as Moving Bottlenecks [41] regulating the flow of human-driven vehicles (HDVs) surrounding them.

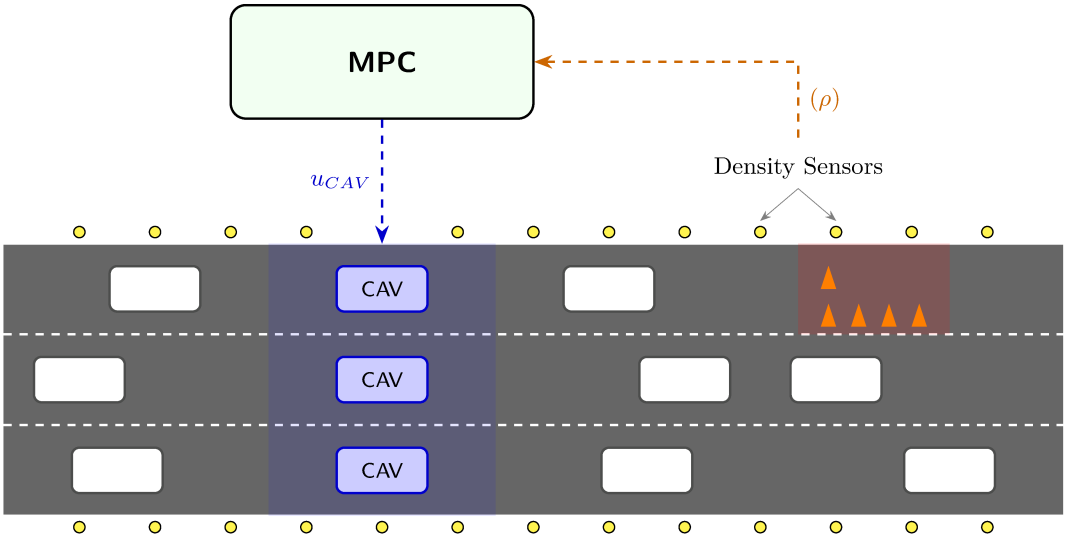


Figure 4.1: Physical representation of the closed-loop control architecture. A single-lane incident blocks the leftmost lane downstream. A network of distributed roadside sensors continuously collects real-time traffic density data (ρ) and transmits it to the MPC. The controller computes the optimal target speed (u_{CAV}) and broadcasts it to the Vanguard CAV formation. By spanning all three lanes, the CAVs create a synchronized moving bottleneck that regulates the total inflow, proactively preventing the traffic breakdown at the downstream incident site.

4.0.1 Eulerian vs. Lagrangian Control

Historically, the distinction between Eulerian (fixed in space) and Lagrangian (moving with the flow) reference frames in traffic engineering literature was primarily applied to sensing and data acquisition [42]. Traditional Eulerian sensors, such as inductive loop detectors, measure traffic variables at fixed infrastructural locations, whereas Lagrangian sensors, such as GPS-equipped probe vehicles, collect data continuously as they navigate through the network.

However, in recent years, with the advent of Connected and Automated Vehicles (CAVs) and advanced algorithms, this nomenclature has been formally extended to classify traffic control strategies as well [43]. Consequently, modern control frameworks are categorized into two distinct paradigms:

- **Eulerian Control:** Relying on stationary roadside infrastructure (e.g., Ramp Metering or overhead Variable Speed Limits) to regulate the traffic flow as vehicles pass a specific geographic cross-section.
- **Lagrangian Control:** Utilizing CAVs as mobile actuators that operate di-

rectly within the traffic stream. These vehicles dynamically adjust their own longitudinal behavior (speed and spacing) to proactively shape the surrounding traffic state and mitigate congestion phenomena from within the flow itself.

Eulerian Control: The state-of-the-art solution for highway bottleneck management is the VSL. As introduced in 1.5 they consist of fixed gantries located at specific coordinates that display a speed limit to all passing vehicles. While effective in theory, VSL systems suffer from two major limitations:

1. **Spatial Discretization:** Control can only be applied at the precise location of the gantry. If a shockwave forms between two gantries, the system may be too slow or spatially inaccurate to intercept it effectively.

2. **Compliance and Enforcement:** VSL relies on human compliance. Drivers often ignore speed limits displayed on overhead signs, especially if the reason for the reduction (e.g., downstream congestion) is not immediately visible [44]. This compliance rate acts as a disturbance in the control loop, often requiring the controller to be overly aggressive to achieve the desired flow reduction.

Lagrangian Control: In contrast, the strategy employed in this thesis utilizes a Lagrangian approach. The actuator (e.g., CAV) moves with the traffic stream. This fundamental difference transforms the control problem from regulating a specific *point* on the road to regulating a specific *cluster* of vehicles.

This approach offers distinct advantages for the bottleneck avoidance problem:

- **Spatial Continuity:** The control action is not bound to a gantry. A CAV can start functioning at any coordinate x to intercept a backward-propagating wave, providing infinite spatial resolution.

- **Physics-Based Compliance:** Unlike VSL, which relies on legal enforcement, CAV-based control relies on physical constraints. By forming a moving bottleneck and reducing speed across all lanes, the CAV physically constrains the velocity of the following vehicles. A human driver cannot drive through them; they are forced to adapt to the new speed. This effectively creates a compliance rate of 100% for the vehicles trapped immediately upstream of the actuator.

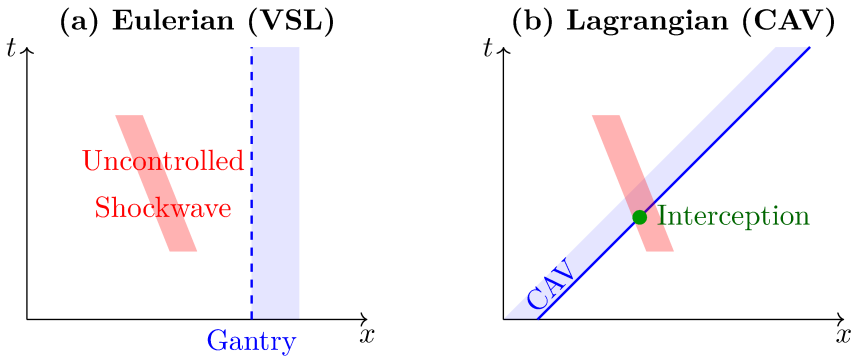


Figure 4.2: Comparison of control reference frames. (a) Eulerian control is limited to fixed gantry locations, potentially missing local phenomena. (b) Lagrangian control moves with the traffic stream, enabling continuous spatial coverage.

4.0.2 Mechanism of Control: Dynamic Flow Regulation

The fundamental mechanism employed in this study is the localized modification of the traffic state via speed regulation. In a macroscopic framework, a vanguard of CAVs traveling at a controlled speed u_{CAV} effectively imposes a constraint on the local transport speed of the traffic stream.

In the numerical implementation (as detailed in the CTM algorithm), the presence of a CAV in cell i at time k overrides the nominal free-flow speed v_f of that cell. The maximum speed of the cell becomes:

$$v_i(k) = \min(v_{free}, u_{CAV}(k)) \tag{4.1}$$

This modification has a direct impact on the **Sending Function** (Demand) of the cell. By reducing the speed, the controller limits the maximum flux that can leave the cell, effectively metering the flow passed to downstream segments. [4]

$$q_{i \to i+1} = \min(\rho_i \cdot u_{CAV},\ Q_{max},\ S_{i+1})$$

This creates a controllable "gate" moving at the speed of the traffic. By manipulating u_{CAV}, the controller can precisely modulate the inflow into the bottleneck area (km 40). The strategy effectively redistributes density: it retains vehicles in a high-capacity upstream section to prevent them from overwhelming the reduced-capacity section at the incident site. This "upstream storage" strategy prevents the bottleneck density from exceeding by far ρ_{cr}, thereby avoiding or at least reducing the capacity drop.

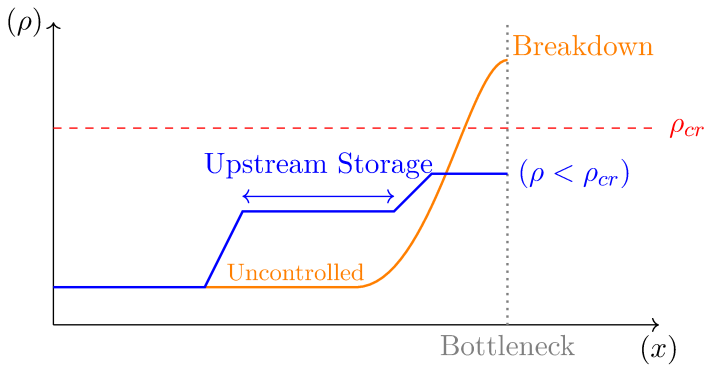


Figure 4.3: The uncontrolled scenario (orange) allows density to spike above ρ_{cr} at the bottleneck, causing capacity drop. The controlled strategy (blue) deliberately increases density upstream to better spread the density across the space.

4.0.3 Sequential Fleet Deployment Architecture

A single actuator (one vanguard of CAVs) has a limited effective window; once it traverses the critical segment and passes the bottleneck, it loses its ability to influence the upstream flow. To maintain control authority throughout the entire duration of the incident (35 to 95 minutes), a **Sequential Fleet Deployment** strategy is adopted.

The control system utilizes a fleet of N_{fleet} distinct CAV vanguards, injected into the simulation at fixed temporal intervals Δt_{inject}. This staggered deployment creates a relay effect:

1. **Continuous Coverage:** As the n-th vanguard approaches the bottleneck and its control horizon expires, the $(n+1)$-th vanguard is already positioned upstream, ready to assume the flow regulation task.

2. **Distributed Actuation:** This architecture transforms the single massive control problem into a sequence of smaller, overlapping control windows.

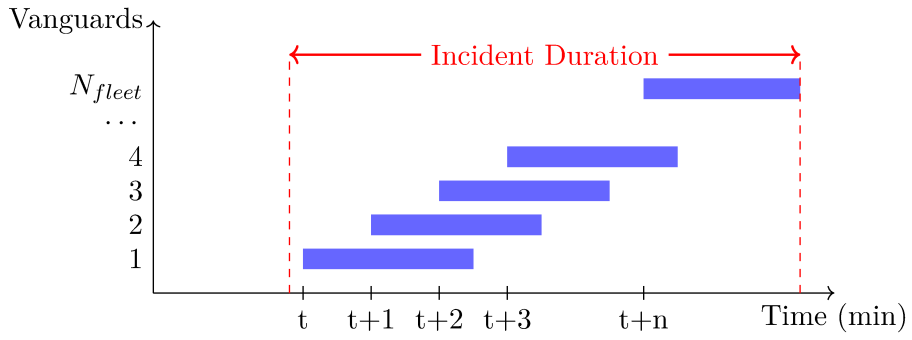


Figure 4.4: Sequential Fleet Deployment strategy. The overlap between the active windows of consecutive vanguard ensures continuous control authority over the bottleneck throughout the incident duration.

This design allows the control algorithm to operate sequentially, optimizing the trajectory of one active vanguard at a time while ensuring that the aggregate effect successfully mitigates the bottleneck congestion.

4.1 Control Objective and Cost Function Formulation

The primary objective of the proposed control strategy is to maximize the operational efficiency of the highway network during severe bottleneck events. In traffic engineering, efficiency is inversely proportional to the aggregate time all vehicles spend in the system. Therefore, the global control goal is the minimization of the **TTT**.

However, minimizing TTT in a system prone to capacity drop requires a nuanced approach. A greedy controller that maximizes instantaneous speed often leads to faster breakdown and, paradoxically, higher total delay. To address this, the control problem is formulated with a composite cost function that balances efficiency, throughput maximization, and congestion prevention.

4.1.1 TTT as a Performance Metric

The Total Travel Time represents the sum of the time spent by all users in the network over the simulation horizon T_{sim}. In the discrete-time, discrete-space CTM framework, TTT is calculated as the integral of the vehicle accumulation over space and time:

$$TTT = \sum_{k=0}^{K} \sum_{i=1}^{N_{cells}} \sum_{l=1}^{N_{lanes}} \rho_{i,l}(k) \cdot L_i \cdot T \tag{4.2}$$

Where:

- $\rho_{i,l}(k)$ is the density in cell i, lane l at step k [veh/km].
- L_i is the length of cell i [km].
- T is the simulation time step [h].

Minimizing this quantity is equivalent to maximizing the average network speed. Under uncongested conditions, the TTT is minimal when all vehicles travel at free-flow speed v_f. However, under saturated demand, the TTT is minimized by maintaining the bottleneck flow at its maximum possible value.

4.1.2 The Physics of Recovery: Trading Space for Capacity

The fundamental justification for applying upstream flow control lies in the **Capacity Drop** phenomenon described in Chapter 3.3.

When a bottleneck activates (e.g., the incident at km 40), the discharge capacity of the road depends on the traffic state upstream of the blockage:

1. **Uncontrolled State (Breakdown):** If traffic is allowed to rush into the bottleneck uncontrolled, the density rises above ρ_{cr}. The discharge capacity drops to the congested value $Q_{cong} \approx \alpha \cdot C_{nom}$.
2. **Controlled State (Recovery):** If the controller holds traffic back using CAV vanguards, maintaining the density at the bottleneck slightly below ρ_{cr}, the discharge capacity remains at C_{nom}.

The control strategy, therefore, trades space for capacity. The controller deliberately induces a delay upstream to create a storage queue. While this locally increases travel time for the vehicles in the formation, it prevents the capacity drop at the bottleneck. The massive time savings gained by maintaining high bottleneck throughput (C_{nom}) far outweigh the minor delays invested upstream.

4.1.3 The Normalized Multi-Objective Cost Function (J)

To realize this strategy within the MPC framework, we define a composite cost function J that guides the optimization of the CAV velocity profiles.

A major challenge in multi-objective optimization is the disparity in physical units (e.g., TTT in vehicle-hours vs. density in veh/km). To ensure numerical stability and intuitive tuning, **each term is normalized** against a reference physical value representing the critical state of the system.

The optimizer minimizes the cost J over the prediction horizon N_p:

$$J = \sum_{k=1}^{N_p} \left[\beta_1 \cdot \hat{J}_{TTT}(k) - \beta_2 \cdot \hat{J}_{Flow}(k) + \beta_3 \cdot \hat{J}_{Density}(k) \right] \tag{4.3}$$

Each normalized term ($\hat{J}$) targets a specific aspect of the control logic:

1. Spatially-Weighted Efficiency Term ($\hat{J}_{TTT}$)

To improve the controller's responsiveness to queue formation, the total travel time evaluation incorporates a spatial weighting mechanism. Rather than treating all cells equally, the controller assigns a spatial penalty weight w_i to each cell.

The normalized cost term $\hat{J}_{TTT}(k)$ evaluates the "weighted equivalent vehicles" currently in the network, divided by the ideal number of weighted vehicles if the entire road were operating perfectly at the critical density ρ_{cr}:

$$\hat{J}_{TTT}(k) = \frac{\sum_{i=1}^{N_{cells}} \sum_{l \in \{A,B,C\}} w_i \cdot \rho_{i,l}(k)}{\rho_{cr} \cdot N_{lanes} \cdot \sum_{i=1}^{N_{cells}} w_i} \tag{4.4}$$

Higher weights ($w_i = 10.0$) are applied to the upstream sections to heavily penalize queues propagating backward, while the segments downstream of the control zone are assigned a lower baseline weight ($w_i = 1.0$). A transition function bridges these zones across the controlled segment.

This robust normalization ensures that the term approaches 1.0 when the network is at maximum free-flow utilization and strictly exceeds 1.0 when congestion waves form.

2. Throughput Maximization Term ($\hat{J}_{Flow}$)

This term incentivizes maximizing the outflow at the bottleneck location. It is normalized by the nominal capacity of the available lanes at the bottleneck ($N_{Flow}^{ref} \approx$ 4400 veh/h for 2 lanes).

$$\hat{J}_{Flow}(k) = \frac{q_{out}^{bottleneck}(k)}{N_{Flow}^{ref}} \tag{4.5}$$

Note: The negative sign in the global equation ($-\beta_2$) indicates maximization.

3. Congestion Queue Volume Penalty ($\hat{J}_{Density}$)

This term acts as a soft constraint that penalizes the physical mass of the congestion queue forming at the bottleneck. Rather than evaluating a simple relative error, the

updated formulation sums the strictly positive excess density across all lanes within the critical bottleneck area.

$$\hat{J}_{Density}(k) = \frac{\sum_{l\in\{A,B,C\}} \sum_{i\in\mathcal{B}} \max(0, \rho_{i,l}(k) - \rho_{cr})}{D_{norm}} \tag{4.6}$$

Where $\mathcal{B} = \{38, 39, 40\}$ represents the subset of cells immediately upstream of and at the incident site, while D_{norm} is the result of multiplying the critical density by the number of cells evaluated in the bottleneck zone and the number of lanes.

By applying the $\max(0, \dots)$ function, the controller ensures that only the density strictly exceeding the critical threshold ρ_{cr} contributes to the penalty. This actively forces the optimizer to measure and suppress the formation of the bottleneck queue without unnecessarily rewarding excessively empty cells.

4.1.4 Weighting Coefficients

The behavior of the controller is determined by tuning the weights β. Thanks to the normalization step, these weights now represent the direct relative importance of each objective, independent of units:

- β_1 (TTT Weight) = 1.
- β_2 (Outflow Weight) = 5.
- β_3 (Density Weight) = 1.3.

Outflow is the dominant factor as keeping the maximum flux is the top priority. TTT and Density have almost the same weight, slightly more favorable towards the latter in order to push the vehicles closer to the bottleneck (TTT) but not too close (Density).

4.2 The Sequential Model Predictive Control Algorithm

To solve the optimal control problem in real-time, a standard global optimization approach (optimizing all N_{fleet} vehicles simultaneously) is computationally prohibitive due to the curse of dimensionality. To address this, we propose a **Sequential Optimization Strategy** coupled with a Receding Horizon framework [45]..

4.2.1 Receding Horizon Principle

The controller operates in a closed loop. At every control step t_c, the system solves an open-loop optimization problem over a finite prediction horizon N_p:

1. **Measure:** The current state of the network (densities ρ and CAV positions) is observed.
2. **Optimize:** The optimal sequence of speed commands $\mathbf{u}^* = [u(t), u(t+1), \ldots, u(t+N_p)]$ is computed to minimize the cost function J defined in Section 4.2.
3. **Apply:** Only the first control action $u(t)$ is applied to the actuators.
4. **Recede:** The time window shifts forward by one step, and the process repeats.

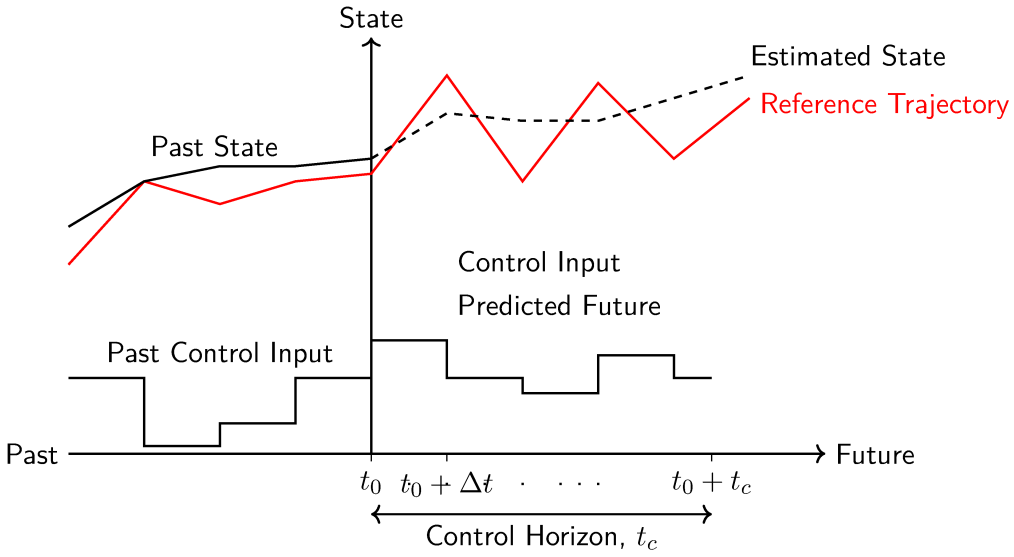


Figure 4.5: Principle of Receding Horizon Control. At each step, an optimal control sequence is computed over a finite horizon, but only the first action is applied (Source: [5]).

4.2.2 From Global to Sequential Optimization

Before introducing the chosen control architecture, it is essential to define the baseline **Global Optimization** logic. In a purely centralized MPC scheme, a single coordinator possesses full knowledge of the traffic network and computes the optimal control inputs for the entire CAV vanguards simultaneously.

Mathematically, this involves finding a joint optimal control matrix $\mathbf{U}^*$ over the prediction horizon H_p, which contains the future actions for all n agents:

$$\mathbf{U}^* = \arg\min_{\mathbf{U}} J(\mathbf{X}, \mathbf{U})$$

where $\mathbf{U} = [\mathbf{u}_{CAV_1}, \ldots, \mathbf{u}_{CAV_n}]^T$, $\mathbf{X}$ represents the predicted traffic states, and J is the global cost function.

While this centralized approach theoretically guarantees the absolute mathematical optimum, it is fundamentally hindered by the computational complexity. As the number of automated vehicles or the prediction horizon increases, the search space expands exponentially. Given the highly non-linear and discontinuous nature of the multi-lane CTM—particularly due to the capacity drop phenomenon and lane-changing friction—solving this massive Non-Linear Programming (NLP) problem simultaneously becomes computationally demanding for real-time applications.

To overcome this computational bottleneck while maintaining high traffic efficiency, a **Sequential Optimization Logic** is adopted in this study. Instead of searching for a joint optimal vector $\mathbf{U}$, which would require exploring a massive state space, the algorithm optimizes one agent at a time in a "greedy" fashion.

The optimization process at each time step k proceeds as follows:

1. **Freeze Neighbors:** Assume all other CAVs maintain their previously computed trajectories or current speeds.
2. **Optimize Leader:** Solve for the optimal speed profile of the first active vanguard (CAV 1) that minimizes the global network cost J.
3. **Lock & Proceed:** Fix the trajectory of CAV 1. Then, solve for CAV 2, treating CAV 1 as a known disturbance.
4. **Iterate:** Repeat for all active CAVs formations ($i = 1, 2, \dots$).

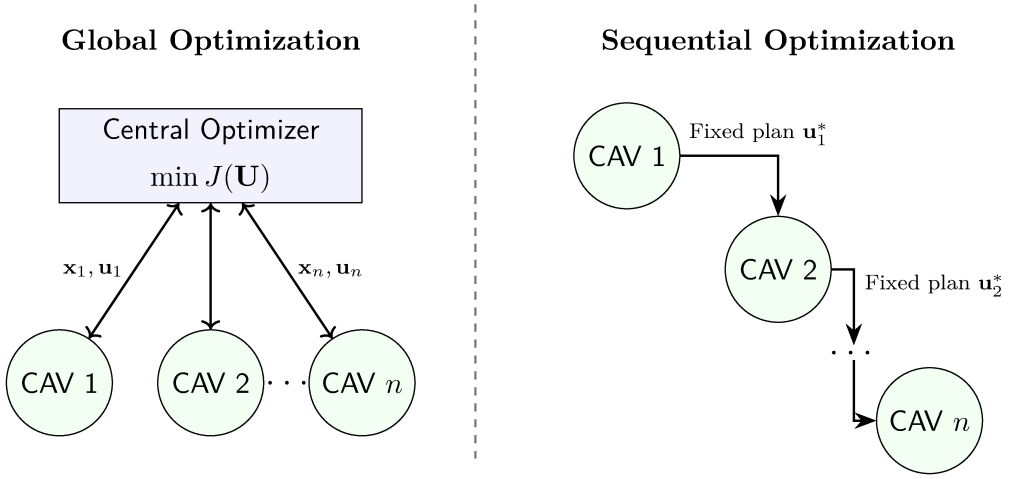


Figure 4.6: Conceptual comparison between Global Optimization (left), where a central controller solves the joint optimal problem for all vehicles simultaneously, and Sequential Optimization (right), where each CAV optimizes its own trajectory iteratively and shares its planned actions with the subsequent agents.

This approach reduces the complexity from exponential (M^N) to linear ($M \times N$), where M is the number of action choices and N is the fleet size. While theoretically suboptimal compared to a global solver, this strategy is highly effective for traffic streams where interactions are primarily unidirectional (upstream vehicles affect downstream flow, but not vice versa).

4.2.3 Discrete Action Space

To further accelerate the solver, the control variable u_{CAV} is not treated as a continuous variable. Instead, the speed commands are selected from a quantized set of candidates:

$$\mathcal{U} = \{5, 10, 15, \ldots, 120\} \text{ km/h}$$

This discretization transforms the optimization into a finite search problem, allowing the controller to rapidly evaluate all feasible speed candidates and select the one that yields the lowest cost J without requiring complex gradient-based solvers.

Chapter 5

Simulation Results and Performance Analysis

This chapter presents the comprehensive simulation results of the proposed Sequential MPC strategy. The primary objective of these experiments is to quantify the ability of a CAV fleet to mitigate congestion and prevent the capacity drop phenomenon during a severe bottleneck event. To properly contextualize the performance of the controller, the chapter first establishes the theoretical and uncontrolled physical boundaries of the network. Subsequently, it details an extensive sensitivity analysis, evaluating the algorithm's effectiveness across a matrix of different spatial configurations (varying CAV injection locations) and temporal control authorities (varying fleet sizes).

5.1 Reference Baselines and Control Objective

As established during the preliminary analysis of the multi-lane CTM (see Section 3.5), rigorously evaluating the proposed Vanguard CAV strategy requires comparing its performance against the absolute physical boundaries of the highway network.

To contextualize the control results presented in this chapter, it is useful to recall the two benchmark scenarios defined for the incident profile:

- **Ideal Lower Bound (TTT = 6935 veh·h):** The theoretical best-case scenario where the capacity drop is artificially disabled ($\alpha = 1.0$). The bottleneck discharges vehicles at the maximum residual efficiency, representing the absolute minimum delay physically achievable.
- **Uncontrolled Baseline (TTT = 7298 veh·h):** The realistic behavior of

human-driven traffic, where the capacity drop is active ($\alpha = 0.7$). Traffic breakdown and lateral merging friction strictly limit the bottleneck's throughput, causing a severe shockwave.

The gap between these two boundaries isolates a systemic inefficiency penalty of **363 veh·h**, caused exclusively by the capacity drop phenomenon.

5.2 Simulation Matrix and Setup

To comprehensively evaluate the controller's robustness and efficiency, a sensitivity analysis was conducted by varying two primary operational parameters:

1. **Starting Position:** The location where the CAVs are activated and begin executing the MPC optimization. This dictates the spatial buffer available to store vehicles upstream of the bottleneck. Tested at km 20, 25, 30, and 35.

2. **Number of Vanguards:** The number of consecutive CAVs deployed to manage the bottleneck. Because vehicles are typically injected at a constant interval of 6 minutes, the number of vanguards directly dictates the total duration the controller can actively manipulate the traffic flow. Tested ranging from 6 to 10 vehicles.

5.2.1 Aggregate Performance Analysis

The TTT for all simulated configurations is summarized in Table 5.1. Every tested configuration successfully mitigated a significant portion of the congestion compared to the uncontrolled baseline (7298 veh·h).

Starting Position	6 CAV	7 CAV	8 CAV	9 CAV	10 CAV
20 km	7048	7033	7033	7034	7034
25 km	7050	7020	7004	7004	7004
30 km	**7036**	**7009**	**6977**	6964	6965
35 km	7043	7011	6982	**6956**	**6954**

Table 5.1: Total Travel Time (veh·h) for all the configurations. Bold results highlight the optimal starting position for each specific fleet size, illustrating the spatiotemporal trade-off.

The absolute best performance was achieved by the 10-CAV fleet starting at km 35, which yielded a TTT of 6954 veh·h. Recalling the control objective established in Section 3.5.3, the systemic penalty of the capacity drop was 363 veh·h. By reducing the TTT to 6954 veh·h, this optimal configuration recovered an astounding 94.7% of the preventable delay, effectively neutralizing the capacity drop and forcing the highway to operate near its theoretical physical limit (6935 veh·h).

5.2.2 Visualizing the Control Strategy

Before proceeding with the detailed performance analysis, it is essential to establish how to correctly interpret the spatiotemporal diagrams generated by the simulation.

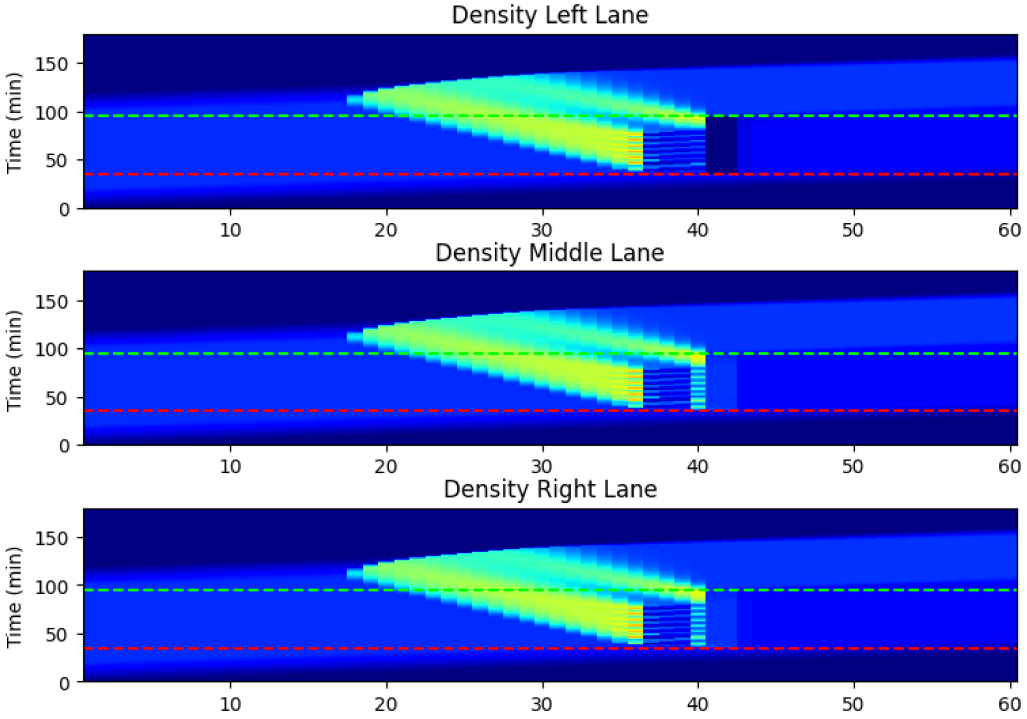
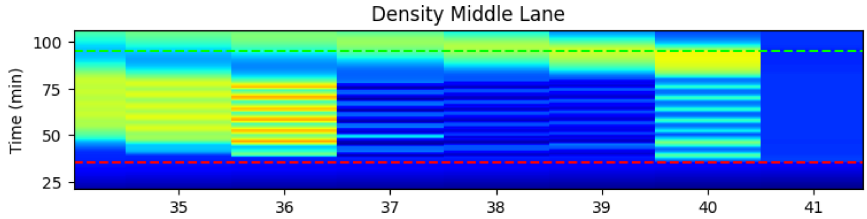


Figure 5.1: Complete spatiotemporal evolution of traffic density and velocity under sequential MPC control. The charts show the macroscopic shockwave mitigation and the trajectories of the CAV fleet.

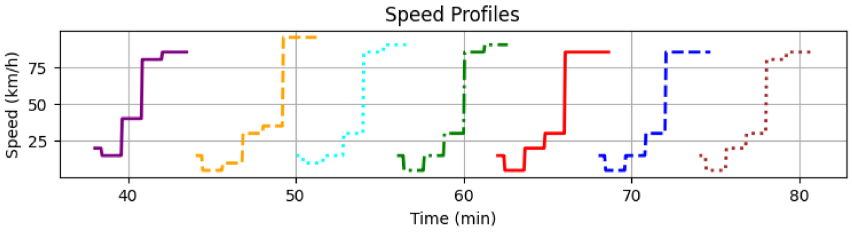
Figure 5.1 illustrates the macroscopic traffic dynamics across the three lanes. The distinct diagonal discontinuities visible within the heatmaps represent the trajectories of the CAV vanguards actively enforcing the speed control policy. To fully grasp this mechanism, a more detailed view is required.

While the aggregate heatmaps demonstrate the overall effectiveness of the strategy, a closer examination of the control zone reveals the specific behavior of the

CAVs. Figure 5.2 juxtaposes a magnified view of the traffic conditions within the control region with the actual speed trajectories dictated by the controller.



a) Control Zone Spatiotemporal Zoom



b) CAVs Optimal Speed Profiles

Figure 5.2: Detailed analysis of the controlled region: zommed version of the CAVs control action (top) and the corresponding optimal speed trajectories generated by the sequential MPC for the CAV (bottom).

In the initial phase, the MPC algorithm detects the critical density buildup at the downstream bottleneck and responds by imposing a strictly reduced velocity on the CAVs. This deliberate deceleration generates a controlled moving bottleneck, which locally accumulates vehicles immediately upstream of the CAVs—visible as the high-density red region in the magnified heatmap. Conversely, if we observe the same time instant downstream of the CAV formation, the plot displays a deep blue region indicating near-zero density. This deliberate "starvation" of flow temporarily prevents new vehicles from reaching the incident area, granting the primary bottleneck sufficient time to discharge before the congestion becomes unmanageable.

As the downstream density dissipates, the associated penalty in the MPC cost function becomes less dominant compared to the objectives of maintaining the outflow and minimizing the Total Travel Time (TTT). Driven by this formulation, the controller progressively commands the CAVs to accelerate back to free-flow conditions. Having established this interpretative framework, the quantitative analysis of the spatiotemporal dynamics can now be addressed.

5.2.3 Spatiotemporal Dynamics: The Activation Trade-off

The data reveals that the overall system efficiency is governed by a strict kinematic relationship between the CAV's starting position and its active control duration. The temporal authority provided by a single CAV is limited to the time it spends traveling from its activation point to the bottleneck at km 40.

Analyzing the optimal configurations (bolded in Table 5.1) reveals a distinct shift in strategy depending on the available vehicular resources. This behavior is driven by the fundamental trade-off between acting quickly upon the bottleneck versus acting for a sufficient duration.

Scarce Resources (6, 7, and 8 CAVs): When the fleet size is small, the primary risk is the premature expiration of the control action. If control is activated close to the incident (e.g., at km 35), the CAVs have a short distance to travel. They traverse the controlled segment and exit the network rapidly. Consequently, their individual temporal footprint is reduced, and the aggregate control horizon of the entire fleet expires well before the 60-minute incident is resolved. This leaves the final portion of the event uncontrolled, triggering a severe, late-stage capacity drop (yielding, for example, 7043 veh·h for 6 CAVs at 35 km).

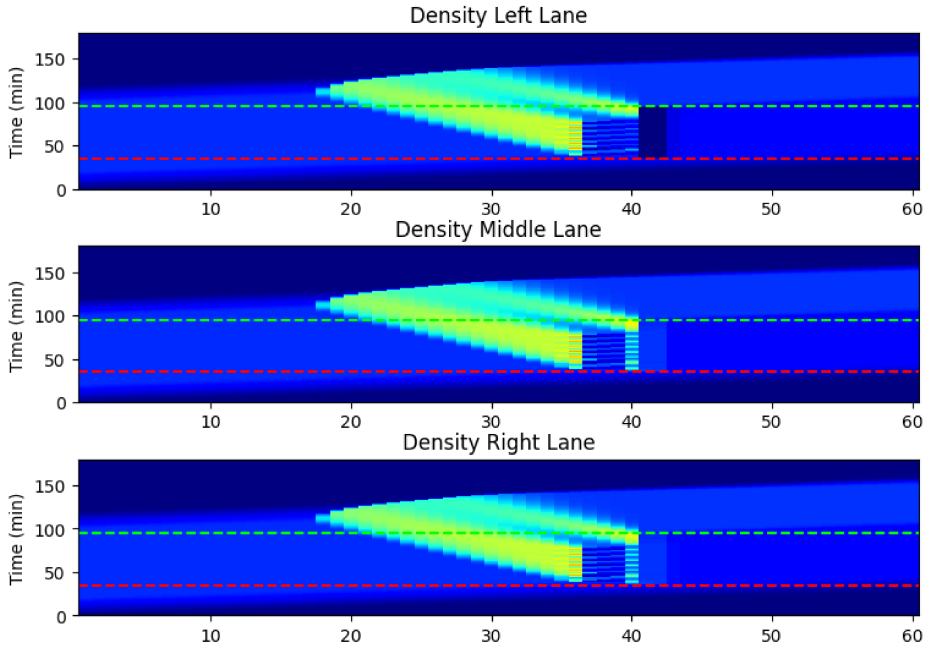


Figure 5.3: Spatiotemporal density profile for a scarce fleet (7 CAVs) activated at 35 km. This loss of temporal control authority leaves the final portion of the incident unmitigated, triggering a severe, delayed capacity drop and yielding suboptimal system efficiency.

Therefore, for smaller fleets (6, 7, and 8 CAVs), the optimal starting position shifts upstream to **km 30**. Activating the fleet further back means that the first CAV has a longer physical distance to travel before it can begin regulating the flow entering the bottleneck at km 40. This increased travel time lengthens the initialization latency, meaning that the bottleneck remains completely uncontrolled for a slightly longer period at the very beginning of the incident, which inevitably accumulates some initial delay (sacrificing some TTT upfront).

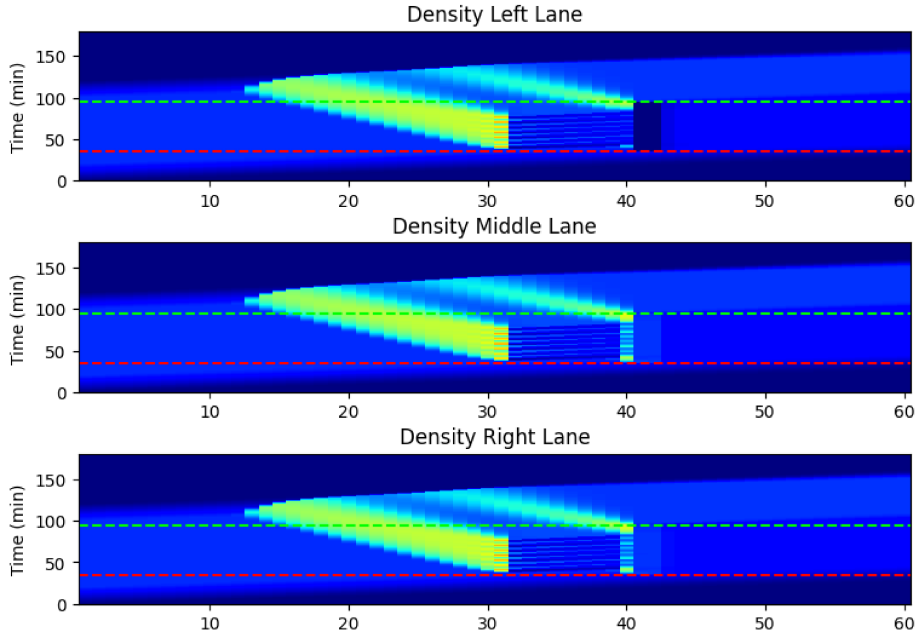


Figure 5.4: Spatiotemporal density profile for 7 CAVs activated at 30 km. By shifting the activation point further upstream, the physical travel time of the vehicles is increased, which significantly extends the fleet's aggregate temporal coverage, reducing the late-stage flow breakdown and yielding the optimal Total Travel Time (7009 veh·h) for this fleet size.

However, this early penalty is strictly outweighed by the macroscopic benefit: starting further upstream keeps each CAV in the network longer, significantly stretching the fleet's aggregate temporal coverage. This allows the scarce fleet to fully cover the latter half of the incident, effectively avoiding the massive late-stage capacity drop and ultimately improving the overall system efficiency.

Abundant Resources (9 and 10 CAVs): Conversely, when deploying a larger fleet of 9 or 10 CAVs, the total temporal coverage is already sufficient to span the entire 60-minute incident period. Because premature expiration is no longer a risk, the priority shifts entirely to minimizing the initialization delay.

For these larger fleets, the optimal starting position is **km 35**. Starting closer to the bottleneck means the control action reaches the congested area much sooner. The immediate, localized mitigation minimizes the early-stage congestion that forms immediately after the incident begins, yielding the absolute best performance (6954 veh·h for 10 CAVs).

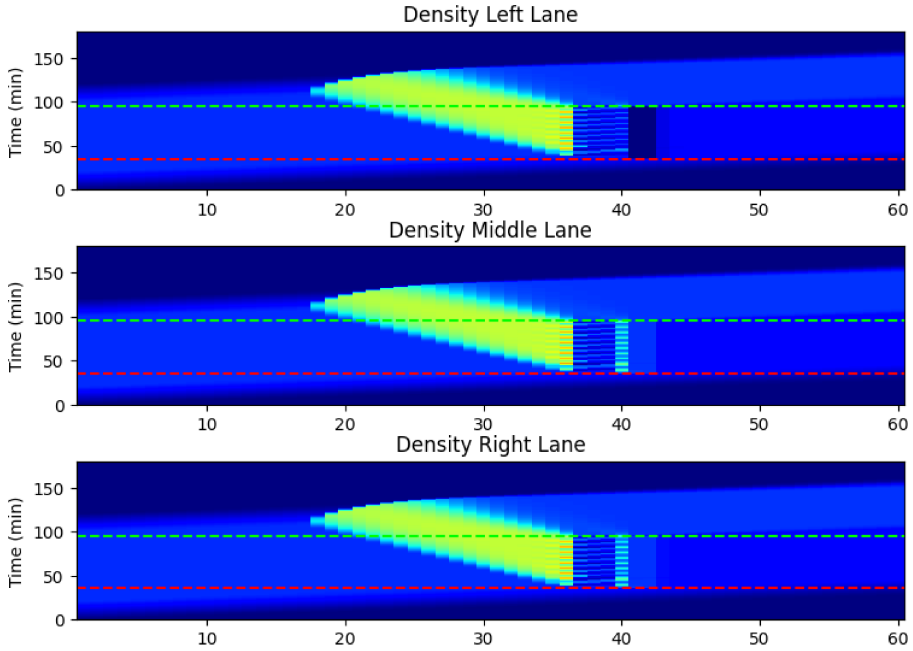


Figure 5.5: Spatiotemporal density profile for the optimal control configuration (10 CAVs activated at 35 km). The full temporal coverage across the entire 60-minute incidente, simultaneously with the proximity to the bottleneck, allows the system to achieve maximum throughput and the lowest Total Travel Time (6954 veh·h).

5.2.4 Temporal Saturation and Initialization Latency

This kinematic relationship also explains the deliberately highlighted temporal saturation points in the matrix (e.g., TTT stabilizing at 7033–7034 veh·h at 20 km, and 7004 veh·h at 25 km).

When a simulation reaches this plateau, it indicates that the active fleet has successfully stretched across the entire 60-minute incident horizon. Because CAVs starting at 20 km spend a vast amount of time on the road, a smaller fleet of just 7 or 8 CAVs is entirely sufficient to maintain optimal bottleneck throughput for the full hour. Adding a 9th or 10th CAV yields identical results because the temporal horizon is already fully mitigated. This confirms that the required fleet size to cover an event is inversely proportional to the activation distance.

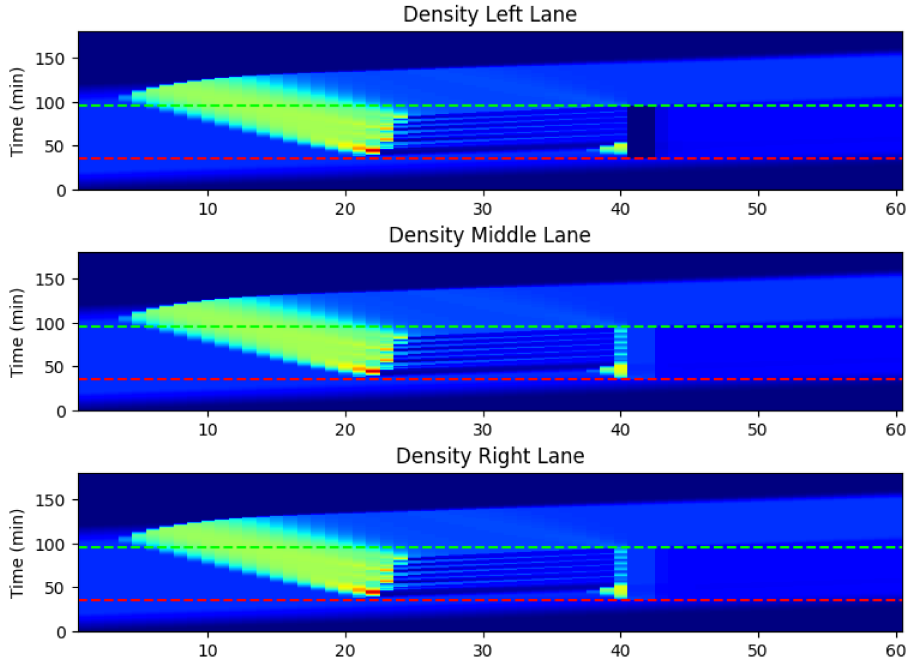


Figure 5.6: Temporal saturation achieved from a distant activation point (7 CAVs at 20 km). The extended travel time allows a smaller fleet to cover the entire incident duration, though it induces significant upstream delays.

The Absolute Physical Limit: To further investigate the limits of this temporal saturation, an additional test was conducted at the 30 km activation point using 10 CAVs, but with a reduced injection interval of 5 minutes instead of 6 minutes. The resulting TTT remained almost identical to the standard 6-minute configuration (yielding roughly 6964 veh·h).

This proves that once a fleet provides full temporal coverage, increasing the control density (injecting CAVs more frequently) provides no further benefit. The residual TTT loss is not due to inadequate injection rates, but is caused strictly by **initialization latency**. This is the unavoidable, uncontrolled time window between the exact moment the incident begins (minute 35) and the moment the very first CAV physically arrives at the km 40 bottleneck to apply the first control action.

This universal physical principle explains why even the absolute best configuration (10 CAVs at km 35, yielding 6954 veh·h) cannot perfectly match the theoretical ideal bound (6935 veh·h). Even though the activation point is much closer, the first CAV must still travel 5 kilometers to reach the disruption. During that brief travel window, the capacity drop begins to manifest uncontrolled. Therefore, initialization latency defines the absolute physical limit of any reactive, distance-based control strategy, as the system can never fully recover the delay accumulated before the

controller physically arrives on site.

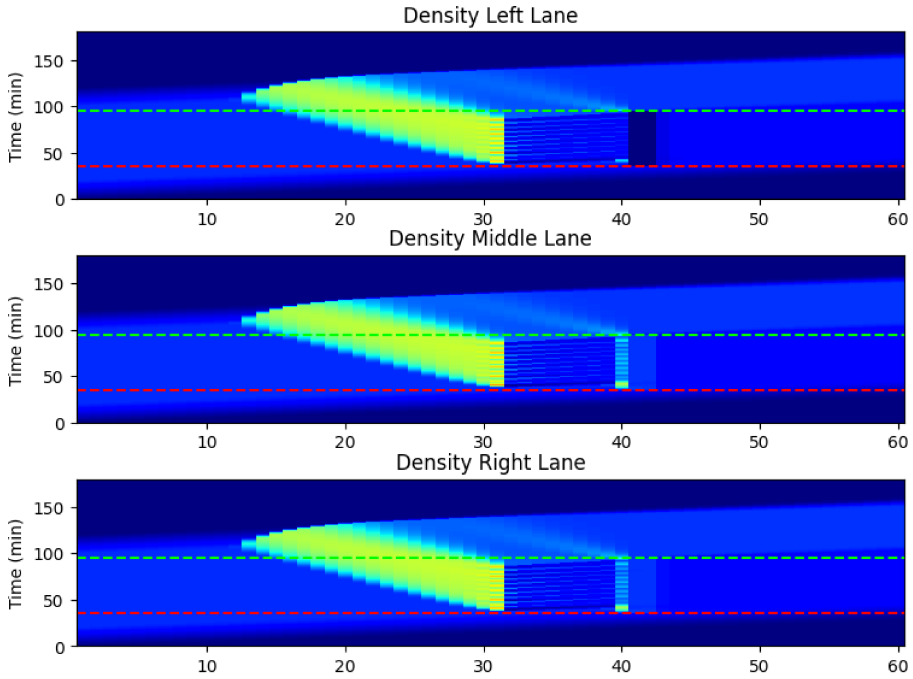


Figure 5.7: Spatiotemporal density profile for 10 CAVs activated at 30 km with a reduced 5-minute injection interval.

5.3 Feasibility Analysis and Market Penetration Requirements

The simulation results presented in the previous sections demonstrate that a fleet of 6 to 10 Connected and Autonomous Vehicles (CAVs) can effectively eliminate the capacity drop and restore optimal highway throughput. However, from a practical deployment perspective, this assumes the immediate availability of a sufficient number of autonomous vehicles precisely when and where the incident occurs. To validate the real-world feasibility of the Sequential MPC strategy, a probabilistic analysis was conducted to determine the minimum required Market Penetration Rate (PR) of CAVs within the general traffic stream.

5.3.1 Probabilistic Demand Modeling: The Poisson Approach

To enable the execution of the control strategy at any desired moment, the system must be capable of deploying a synchronized moving bottleneck. As a preliminary assumption, since this formation must span all three available lanes to effectively

regulate the upstream flow, a minimum arrival of 3 CAVs per minute is hypothesized as the baseline requirement to form a complete lateral cluster.

To translate this operational target into a required market penetration rate (PR), the overall traffic demand is modeled as a stochastic process. During the peak demand phase, the highway experiences a total inflow of 6000 veh/h, corresponding to a mean arrival rate of $\lambda_{total} = 100$ veh/min for the general vehicle population. Assuming that CAVs are distributed randomly within this traffic stream, their arrivals can be modeled using a Poisson distribution with an expected rate of $\lambda_{CAV} = PR \cdot \lambda_{total}$.

To ensure the robustness of the control strategy, we impose a strict reliability constraint: there must be at least a 95% probability that 3 or more CAVs arrive within any given minute. By evaluating this statistical threshold against the total vehicular inflow, the minimum penetration rate can be mathematically derived by solving the following cumulative probability function:

$$P(X \geq 3) = 1 - e^{-\lambda_{CAV}} \left(1 + \lambda_{CAV} + \frac{\lambda_{CAV}^2}{2}\right) \geq 0.95 \tag{5.1}$$

Solving this inequality for the mean arrival rate of autonomous vehicles (λ_{CAV}) yields a minimum required rate of $\lambda_{CAV} \approx 6.3$ CAVs/min. Given the total network flow of $\lambda_{total} = 100$ veh/min, this mathematically translates to a required PR of 6.3%.

5.3.2 Topological Validation via Monte Carlo Simulation

While the Poisson distribution dictates the temporal probability of arrival, it does not guarantee that the vehicles are spatially close enough to coordinate on the asphalt. To validate the spatial feasibility of the 6.3% PR, a Monte Carlo simulation was developed in Python. The simulation modeled a continuous highway stream slightly under the critical density conditions of the MPC scenario (50 veh/km). A realistic technological constraint was introduced, limiting the maximum Vehicle-to-Vehicle (V2V) communication gap to 500 meters. A cluster was deemed successfully formed if at least 3 CAVs linked together within this spatial tolerance.

The results of the Monte Carlo validation, run over 5000 randomized iterations, are presented in Figure 5.8.

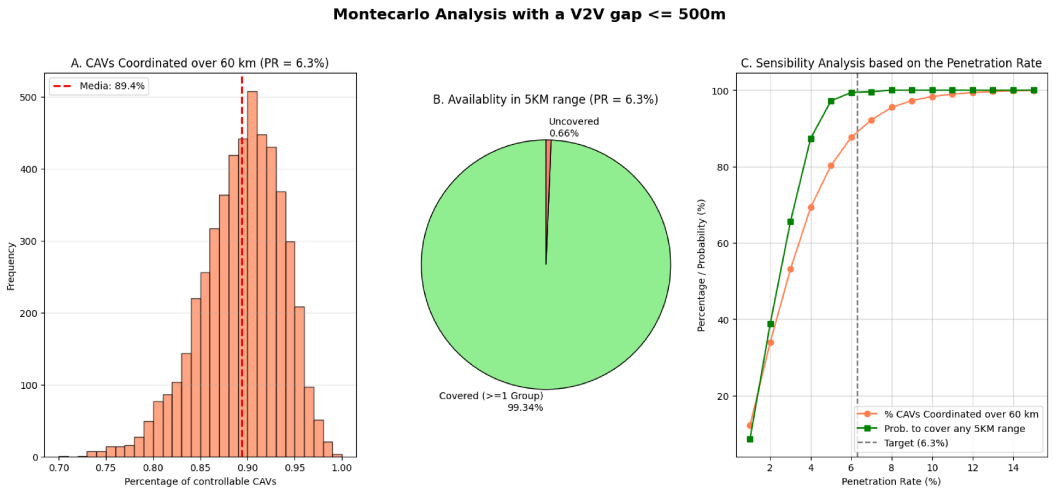


Figure 5.8: Topological validation of the vehicular network under a 500m V2V communication constraint. (A) Distribution of coordinated CAVs over a 60 km network at 6.3% PR. (B) Reliability of forming a control cluster within a 5 km critical activation zone. (C) Sensitivity analysis demonstrating the inflection point of system reliability across varying penetration rates.

The topological simulation strongly supports the analytical Poisson derivation. As shown in Figure 5.8A, at the target 6.3% PR, an average of 89.6% of all autonomous vehicles on a 60 km highway segment successfully establish V2V links to form usable clusters.

More importantly, evaluating the strictly localized "activation zone" required for the Sequential MPC strategy (a 5 km stretch upstream of the bottleneck), Figure 5.8B demonstrates a **99.34% probability** of successfully forming at least one valid control cluster. This proves that a 6.3% market penetration is overwhelmingly sufficient to guarantee the physical availability of the required control fleet.

Finally, a broader sensitivity analysis (Figure 5.8C) evaluated penetration rates ranging from 1% to 15%. The resulting probability curve reveals that the calculated 6.3% PR sits precisely at the critical inflection point of the system. At penetration rates below 5%, the probability of forming a cluster drops precipitously, rendering reactive control impossible. Conversely, penetration rates above 8% offer severely diminishing returns in spatial reliability.

Consequently, this analysis confirms that the proposed Sequential MPC framework does not require futuristic, near-total market adoption of autonomous vehicles. The control strategy is highly feasible and can be reliably deployed as soon as CAV market penetration reaches the modest threshold of 6.3%.

Chapter 6

Conclusions and Future Work

6.1 Summary of Contributions

Highway bottlenecks caused by incidents or lane closures frequently trigger a systemic breakdown of traffic flow. This breakdown, known as the capacity drop phenomenon, severely degrades the discharge rate of the bottleneck and causes backward-propagating shockwaves that significantly increase TTT.

To mitigate this inefficiency, this thesis proposed and evaluated a Sequential MPC strategy. The proposed system leverages a fleet of CAVs to actively regulate upstream traffic flow. By optimizing the speed and trajectories of the CAVs over a finite prediction horizon, the controller successfully prevents the density at the bottleneck from reaching the critical breakdown threshold, thereby maintaining maximum physical throughput. The physical plant was rigorously simulated using a macroscopic CTM integrated with a customized, incentive-based lateral flow logic to accurately replicate the dynamics of a degrading highway section.

Furthermore, this research extended beyond theoretical control formulation to rigorously assess the real-world feasibility of the system. Through probabilistic demand modeling using a Poisson distribution and subsequent topological validation via Monte Carlo simulations, the study mathematically established the minimum vehicular resources required for deployment. The analysis proved that the Sequential MPC strategy achieves over 99% spatial reliability at a modest Market PR of just 6.3%.

6.2 Key Findings

Through an extensive sensitivity analysis varying both the spatial activation point and the temporal fleet size during a 60-minute incident, this research uncovered several fundamental kinematic relationships governing reactive traffic control:

- **Near-Optimal Delay Recovery:** The uncontrolled capacity drop imposed a systemic penalty of 363 veh·h. The optimal CAV configuration (10 vehicles deployed 5 km upstream) successfully reduced the total delay to 6954 veh·h, recovering 94.7% of the preventable penalty and forcing the degraded network to operate near its theoretical physical limit.

- **The Spatiotemporal Trade-off:** The optimal activation distance is strictly dictated by the available vehicular resources, governed by a trade-off between initialization latency and temporal lifespan.

  - **Abundant Resources:** When a large fleet (e.g., 9–10 CAVs) is available, the temporal horizon is naturally covered. The optimal strategy is to activate control close to the incident (e.g., at 35 km) to minimize initialization latency and apply corrective action as quickly as possible.

  - **Scarce Resources:** When the fleet is limited (e.g., 6–8 CAVs), activating too close to the bottleneck causes the vehicles to exit the control zone prematurely, triggering a severe late-stage capacity drop. Moving the activation point further upstream (e.g., to 30 km) extends the travel time and aggregate lifespan of the fleet, successfully spanning the full duration of the incident and avoiding systemic breakdown.

- **The Absolute Physical Limit:** Tests utilizing increased injection frequencies proved that the residual delay cannot be eliminated by simply adding more vehicles. The ultimate efficiency of the system is strictly bounded by *initialization latency*—the unavoidable time required for the very first CAV to physically travel from its activation point to the site of the disruption.

- **Feasibility and Market Penetration:** Probabilistic demand modeling and topological Monte Carlo simulations demonstrated that the control strategy does not require a fully autonomous vehicle fleet. The system achieves over 99% spatial reliability for forming multi-lane control clusters at a modest CAV market penetration rate of just 6.3%.

6.3 Practical Implications

From a traffic engineering perspective, these findings provide actionable, dynamic deployment strategies for highway management centers. The Sequential MPC algorithm demonstrates high resilience, allowing operators to dynamically adjust the spatial activation point based on the real-time market penetration or availability of CAVs. Operators can strategically trade immediate, localized capacity for sustained temporal regulation, guaranteeing robust congestion mitigation even under varying resource constraints.

6.4 Limitations and Future Work

While the proposed control strategy demonstrates exceptional theoretical performance, several avenues remain for future research to bridge the gap toward real-world implementation:

- **Dynamic Incident Horizons:** The current simulation tested a fixed 60-minute incident. Future work should evaluate the controller's adaptability to incidents of unpredictable or fluctuating durations, potentially integrating real-time incident duration estimation to dynamically adjust the fleet's temporal lifespan.
- **Mixed-Traffic Dynamics and Human Compliance:** While the topological analysis confirms that a 6.3% penetration rate is sufficient to form a mobile bottleneck, the current macroscopic model inherently assumes perfect compliance from the human-driven vehicles trailing the control fleet. In reality, human drivers may exhibit aggressive lane-changing, tailgating, or delayed reaction times when paced at reduced speeds. Future research should integrate microscopic car-following and lane-changing models to evaluate the physical stability of the human-driven traffic stream behind the CAV wall and assess potential safety implications.
- **Infrastructure and Sensing Requirements:** The Sequential MPC relies heavily on accurate, real-time macroscopic traffic states along the entire prediction horizon to optimize the CAV trajectories. Implementing this logic in the real world necessitates significant infrastructure investments, such as ubiquitous density sensors (e.g., inductive loop detectors, radar, or camera systems) deployed at high frequencies along the highway, or advanced V2I

communication networks. Future studies could explore the robustness of the MPC algorithm under conditions of noisy, delayed, or missing sensor data.

Bibliography

[1] M. Barth and K. Boriboonsomsin, "Traffic congestion and greenhouse gases," *Access Magazine*, vol. 1, no. 35, pp. 2–9, 2009.

[2] Wikimedia Commons, "Stack interchange." `https://en.wikipedia.org/wiki/The_Stack`, 2024. Accessed: 2026-01-14.

[3] SAE International, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," Tech. Rep. J3016, SAE International, 2016.

[4] G. Piacentini, A. Ferrara, I. Papamichail, and M. Papageorgiou, "Highway traffic control with moving bottlenecks of connected and automated vehicles for travel time reduction," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 3140–3145, IEEE, 2019.

[5] H. Wei, W. Ross, S. Varisco, P. Krief, and S. Ferrari, "Modeling of human driver behavior via receding horizon and artificial neural network controllers," in *52nd IEEE Conference on Decision and Control (CDC)*, pp. 6778–6785, IEEE, 2013.

[6] U. N. D. of Economic and S. Affairs, *Demographic Yearbook, 75th issue, 2024*. United Nations, 2024. Accessed: 2026-03-01.

[7] Federal Highway Administration, "Traffic volume trends: Archive 2002–2020." Office of Highway Policy Information, U.S. Department of Transportation, 2020. Accessed: 2026-03-01.

[8] K. Button, *Transport Economics*. Edward Elgar Publishing, 3rd ed., 2010.

[9] European Environment Agency, "Greenhouse gas emissions from transport in europe," indicator assessment, European Environment Agency, 2024.

[10] Y. Huang, Y. Zhang, F. Deng, D. Zhao, and R. Wu, "Impacts of built-environment on carbon dioxide emissions from traffic: A systematic literature review," *International Journal of Environmental Research and Public Health*, vol. 19, no. 24, p. 16898, 2022.

[11] G. Duranton and M. A. Turner, "The fundamental law of road congestion: Evidence from us cities," *American Economic Review*, vol. 101, no. 6, pp. 2616–2652, 2011.

[12] R. B. Noland and J. W. Polak, "Travel time variability: A review of theoretical and empirical issues," *Transport Reviews*, vol. 22, no. 1, pp. 39–54, 2002.

[13] M. Barth and K. Boriboonsomsin, "Real-world carbon dioxide impacts of traffic congestion," *Transportation Research Record*, vol. 2058, no. 1, pp. 163–171, 2008.

[14] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E*, vol. 62, no. 2, pp. 1805–1824, 2000.

[15] T. D. Gillespie, *Fundamentals of Vehicle Dynamics*. Warrendale, PA: Society of Automotive Engineers (SAE), 1992.

[16] H. Yao, Q. Li, and X. Li, "A study of relationships in traffic oscillation features based on field experiments," *Transportation Research Part A: Policy and Practice*, vol. 141, pp. 339–355, 2020.

[17] World Health Organization, "Ambient (outdoor) air pollution," 2024. Accessed: 2026-01-13.

[18] K. Zhang and S. Batterman, "Air pollution and health risks due to vehicle traffic," *Science of The Total Environment*, vol. 450–451, pp. 307–316, 2013.

[19] Z. Zheng, S. Ahn, and C. M. Monsere, "Impact of traffic oscillations on freeway crash occurrence," *Accident Analysis & Prevention*, vol. 42, no. 2, pp. 626–636, 2010.

[20] C. Lee, F. Saccomanno, and B. Hellinga, "Analysis of crash precursors on instrumented freeways," *Transportation Research Record*, vol. 1784, no. 1, pp. 1–8, 2002.

[21] R. W. Novaco, D. Stokols, J. Campbell, and J. Stokols, "Transportation, stress, and community psychology," *American Journal of Community Psychology*, vol. 7, no. 4, pp. 361–380, 1979.

[22] S. Siri, C. Pasquale, S. Sacone, and A. Ferrara, "Freeway traffic control: A survey," *Automatica*, vol. 130, p. 109655, 2021.

[23] D. Braess, A. Nagurney, and T. Wakolbinger, "On a paradox of traffic planning," *Transportation Science*, vol. 39, no. 4, pp. 446–450, 2005.

[24] J. Sussman, *Introduction to Transportation Systems.* Boston, MA: Artech House, 2000.

[25] R. P. Roess, E. S. Prassas, and W. R. McShane, *Traffic Engineering.* Upper Saddle River, NJ: Pearson, 4th ed., 2011.

[26] A. Ferrara, S. Sacone, and S. Siri, *Freeway Traffic Modelling and Control.* Cham, Switzerland: Springer, 2018.

[27] N. Sacchi, M. Cucuzzella, and A. Ferrara, "Constrained ramp metering control based on sliding mode unknown input observers," *european Journal of Control*, vol. 85, p. 101329, 2025.

[28] M. J. Cassidy and J. Rudjanakanoknad, "Increasing the capacity of an isolated merge by metering its on-ramp," *Transportation Research Part B: Methodological*, vol. 39, no. 10, pp. 896–913, 2005.

[29] S. Li, T. Wang, H. Ren, B. Shi, and X. Kong, "Variable speed limit strategies based on the macro hierarchical control traffic flow model," *Journal of Advanced Transportation*, vol. 2021, p. Article ID 9910097, 2021.

[30] S. E. Shladover, D. Su, and X.-Y. Lu, "Impacts of cooperative adaptive cruise control on freeway traffic flow," *Transportation Research Record*, vol. 2324, no. 1, pp. 63–70, 2012.

[31] 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.

[32] J. G. Wardrop, "Some theoretical aspects of road traffic research," *Proceedings of the Institution of Civil Engineers*, vol. 1, no. 3, pp. 325–362, 1952.

[33] L. C. Edie, "Car-following and steady-state theory for noncongested traffic," *Operations Research*, vol. 9, no. 1, pp. 66–76, 1961.

[34] R. C. Carlson, I. Papamichail, M. Papageorgiou, and A. Messmer, "Optimal mainstream traffic flow control of large-scale motorway networks," *Transportation Research Part C: Emerging Technologies*, vol. 18, pp. 193–212, Apr 2010.

[35] Z. Wang, R. Carlson, I. Grassi, and M. Papageorgiou, "Freeway traffic control in presence of capacity drop," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1397–1409, 2021.

[36] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956.

[37] 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.

[38] S. L. Shvetsov and D. Helbing, "Extending the cell transmission model to multiple lanes and lane-changing," *International Journal of Civil Engineering*, vol. 15, pp. 507–535, Jul 2015.

[39] C. Roncoli, I. Papamichail, and M. Papageorgiou, "Traffic flow optimization in presence of vehicle automation and communication systems—part ii: Optimal control for multi-lane motorway," *Transportation Research Part C: Emerging Technologies*, vol. 57, pp. 260–275, 2015.

[40] R. E. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, H. Pohlmann, F. Wu, B. Piccoli, *et al.*, "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.

[41] D. C. Gazis and R. Herman, "The moving and "phantom" bottlenecks," *Transportation Science*, vol. 26, no. 3, pp. 223–229, 1992.

[42] Y. Yuan, J. W. C. Van Lint, R. E. Wilson, F. van Wageningen-Kessels, and S. P. Hoogendoorn, "Real-time lagrangian traffic state estimator for freeways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, pp. 59–70, Mar 2012.

[43] E. Vinitsky, K. Parvate, A. Kreidieh, C. Wu, and A. Bayen, "Lagrangian control through deep-rl: Applications to bottleneck decongestion," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 759–765, IEEE, Nov 2018.

[44] 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.

[45] E. F. Camacho and C. Bordons, *Model Predictive Control*. Springer London, 2nd ed., 2007.