

ABSTRACT

Title of Dissertation: A COORDINATED SIGNAL CONTROL SYSTEM FOR
COMMUTING ARTERIALS WITH HIGHLY
ASYMMETRIC DIRECTIONAL FLOWS

Yi-Ting Lin, Doctor of Philosophy, 2025

Dissertation directed by: Gang-Len Chang, Professor,
Department of Civil & Environmental Engineering

Urban commuting arterials often encounter unique operational challenges during peak periods due to highly asymmetric directional traffic patterns and the frequent queue spillbacks and mutual blockages. These traffic disruptions—particularly at intersections with short turning bays and limited link storage—can severely degrade the progression capacity and cause excessive delays, especially for high-volume turning and through movements. While considerable progress has been made by the traffic community over the years in developing various signal control systems, most existing strategies—both in traditional time-of-day coordination and advanced real-time adaptive control—have not fully addressed the complexities of mid-block congestion dynamics, cross-lane interference, and proactive mitigation of queue spillbacks as well as mutual blockages.

This dissertation introduces a modular-based and scalable arterial signal control system specifically designed for congested commuting arterials with directional demand imbalances and short turning-bay lengths. The developed system comprises three integrated modules: an offline time-of-day control module, a responsive real-time control module, and a proactive real-time control module. The time-of-day module is designed to produce the optimized phase sequences and signal offsets by integrating delay minimization for the high-volume commuting direction with

the progression operations for the opposite direction, while accounting for the roadway geometric constraints such as turning-bay and short link lengths. The responsive module functions to advance the pre-timed signal plans to real-time adaptation through a rule-based logic structure that detects and classifies congestion patterns, and then applies the selected signal adjustments to targeted intersections so as to mitigate the traffic impacts of queue spillbacks and lane blockages. Advancing the core control notion from responsive to proactive, this study further develops a proactive signal control module, which is incorporated with a multi-branch, multi-head Long Short-Term Memory (LSTM) architecture to predict queue states, based on the complex spatial-temporal dependencies of traffic states and queue patterns between the target intersection and its multiple upstream intersections. This embedded predictive function allows the system to execute preemptive control actions before the onset of critical congestion patterns. To ensure the operational reliability and stability, a real-time monitoring mechanism for evaluating the accuracy of prediction, triggering the fallback to responsive control when necessary, has also been included in this proactive signal module.

The developed system has been evaluated through traffic simulation experiments using a real-world arterial testbed on MD 355 Rockville Pike in Bethesda, Maryland, under evening peak-hour conditions. Experimental results with respect to common measures of effectiveness have confirmed the promising properties of the developed system, where 1) its offline time-of-day module appears to outperform state-of-the-art signal coordination methods such as MULTIBAND and TRANSYT-7F under congested peak-hour traffic conditions; 2) the responsive real-time module has further improved the time-of-day control's performance on dynamically mitigating queue spillbacks and mutual blockages; and 3) the system's advanced proactive module can achieve the highest overall performance with its unique functions to predict the traffic state and

congestion patterns in the predefined time horizon and then execute the optimal preemptive signal control actions in advance.

Overall, the experimental results from extensive traffic scenarios have collectively confirmed the flexibility and effectiveness of the developed system under various information availability, control environments, and geometric constraints, as well as asymmetric volume patterns over congested commuting arterials. With the support of state-of-the-art traffic sensors, the developed system offers an effective alternative for use in practice to contend with ever-increasing traffic congestion in urban commuting corridors.

**A COORDINATED SIGNAL CONTROL SYSTEM FOR COMMUTING ARTERIALS
WITH HIGHLY ASYMMETRIC DIRECTIONAL FLOWS**

By

Yi-Ting Lin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

2025

Advisory Committee:

Professor Gang-Len Chang, Chair

Professor Ali Haghani

Professor Martin Dresner

Associate Professor Xianfeng (Terry) Yang

Dr. Woon Kim at the Federal Highway Administration

© Copyright by
Yi-Ting Lin
Doctor of Philosophy, 2025

Dedication

**To my father, Tien-Jui Lin, and my mother, Li-Chuan Tsai
for their love and support**

Acknowledgments

I would like to express my sincere gratitude to my advisor, Dr. Gang-Len Chang, whose strict and uncompromising guidance has shaped me from a fresh university graduate with no research experience into a PhD scholar. Though never one to sugarcoat his words—often challenging us bluntly—his high standards and relentless drive pushed me to toughen up, think critically, work harder, and face challenges head-on. This tough instruction has made me braver, more disciplined, and better prepared to tackle complex problems, both in academia and beyond. I am deeply thankful for the opportunities he provided and for the trust he placed in my ability to grow through every challenge.

I am also profoundly grateful to my dissertation committee members—Dr. Ali Haghani, Dr. Martin Dresner, Dr. Xienfang Yang, and Dr. Woon Kim—for their time, valuable insights, critical questions, and constructive feedback. Their diverse expertise and perspectives have strengthened this work in ways I could not have achieved alone.

The Traffic Operation and Safety Lab has been more than a workplace—it has been a community. I am thankful to my colleagues and friends there for the countless brainstorming sessions, technical debates, and everyday conversations that have made this journey more collaborative, stimulating, and enjoyable. A special thank you goes to Dr. Yao Cheng. Your mentorship, patience, and constant encouragement have been invaluable to both my academic and personal growth. I am truly fortunate to have had your guidance throughout this journey.

Above all, I am forever grateful to my family. To my parents—thank you for your unwavering love and support, even from Taiwan: always just one phone call away with encouragement and reassurance, for flying across the ocean to share my important moments, and for the enormous suitcases of Taiwanese food you bring to keep me going. To my brother—thank

you for handling countless errands and holding down the fort back home so I could focus completely on my studies. To my sister—my personal PhD tutor—thank you for staying up late to answer my academic questions and guiding me with patience and humor. And to my husband—thank you for catching me in my lowest moments, lifting me up, and believing in me when I doubted myself; your love and encouragement have been my greatest source of strength.

Finally, I wish to extend my heartfelt thanks to everyone—friends, colleagues, and mentors—who, in ways big and small, have supported me along this journey. Whether through guidance, encouragement, or simply believing in me, your presence has been an essential part of reaching this milestone, and I will always be deeply grateful.

Table of Contents

Dedication.....	i
Acknowledgments	ii
Table of Contents	iv
List of Tables.....	vii
List of Figures.....	viii
Chapter 1 Introduction.....	1
1.1 Research Background.....	1
1.2 Research Objective.....	4
1.3 Dissertation Organization.....	5
Chapter 2 Literature Review	9
2.1 Introduction	9
2.2 Offline Arterial Signal Coordination Optimization.....	9
2.2.1 Control Objective: Progression Maximization	9
2.2.2 Control Objective: Delay Minimization	12
2.2.3 Multi-Objective Control.....	13
2.3 Commuting Arterial Traffic Controls.....	14
2.4 Adaptive Signal Control.....	17
2.4.1 Responsive Signal Control.....	17
2.4.2 Proactive Signal Control	20
2.5 Discussion	26
Chapter 3 System Structure and Key Components	28
3.1 System Structure Overview.....	28
3.2 Key Modules	35
3.2.1 Module I: Offline Signal Control for Commuting Arterials with Highly Asymmetric Directional Flows	35
3.2.2 Module II: Responsive Real-Time Arterial Signal Control for Turning-bay Overflows and Mutual Queue Blockages	36
3.2.3 Module III: Proactive Real-Time Arterial Signal Control for Commuting Arterials with Highly Asymmetric Directional Flows	39
3.3 Closure	42
Chapter 4 A Time-of-day Signal Control Module for Commuting Arterials with Highly Asymmetric Directional Flows.....	43
4.1 Introduction	43
4.2 Time-of-Day Arterial Signal Control Module.....	46

4.3	Model Formulations	49
4.3.1	Queue Formation and Evolution Process.....	49
4.3.2	Delay Estimation Based on the Time-Varying Queue Length	57
4.3.3	Traffic Dynamics Under Mutual Queue Blockage	62
4.3.4	Progression Design for the Low-Traffic-Volume Direction	74
4.3.5	Relations Between the Optimal Phase Sequences and the Starting Time of Each Phase	74
4.3.6	Summary.....	75
4.4	Case Study.....	76
4.4.1	Performance Analysis	78
4.4.2	Performance Evaluation.....	84
4.4.3	Case Study Summary	88
Chapter 5 A Responsive Real-Time Arterial Signal Control Module for Turning-bay Overflows and Mutual Queue Blockages.....		90
5.1	Introduction	90
5.2	Overview of the Responsive Real-Time Arterial Signal Control Module	92
5.3	Module components	96
5.3.1	Model 1: Queue Length Computation	97
5.3.2	Model 2: Identification of the Congestion Level.....	99
5.3.3	Model 3: Dynamic Responsive Signal Control	102
5.4	Numerical Experiments.....	107
5.4.1	Performance Evaluation for the Detection and Timely Response to Mutual Blockages.....	110
5.4.2	Performance Evaluation at the Critical Intersection	111
5.4.3	Module Performance Evaluation	116
5.5	Conclusions	119
Chapter 6 A Proactive Adaptive Signal Control Module with Queue Dynamics Prediction for Congested Arterials.....		121
6.1	Introduction	121
6.2	Overview of the Proactive Real-Time Arterial Signal Control Module.....	122
6.3	Functions and Formulations of the Key Embedded Models	125
6.3.1	Model 1: Queue Length State Prediction.....	125
6.3.2	Model 2: Prediction Monitoring and Mode Switching.....	155
6.3.3	Model 3: Congestion Level Assessment.....	158
6.3.4	Model 4: Proactive Strategy Generator.....	160

6.4	Case Study.....	161
6.4.1	Effectiveness in Preventing Mutual Blockage Events through Proactive Control Module 162	
6.4.2	Overall Module Performance Evaluation	164
6.5	Conclusions	167
Chapter 7	Conclusions.....	168
7.1	Research Summary.....	168
7.2	Future Research.....	171
7.2.1	System Enhancements	171
7.2.2	Potential Applications and Integration with ATMS Systems.....	173
Bibliography	175

List of Tables

Table 4.1 List of key notations.....	47
Table 4.2 Performance Comparison between the developed module and MAXBAND with the selected MOEs	82
Table 4.3 Performance Comparisons of the signal plans generated by three models.....	86
Table 5.1 Mathematical expression of mutual blockage types	98
Table 5.2 Detection of mutual blockages at each intersection.....	111
Table 5.3 Simulation results for performance comparison	118
Table 6.1 The results of the Granger causality test	143
Table 6.2 Optimized Hyperparameters for the Multi-Branch Multi-Head LSTM Model	143
Table 6.3 Summary of Spillover/Blockage Prediction Performance	144
Table 6.4 Effect of spatial input scope on prediction accuracy: three-, four-, and five-branch LSTM comparison	151
Table 6.5 Rule-Based Classification of Predicted Congestion Levels.....	160
Table 6.6 Proactive Control Actions Based on Predicted Congestion Level	161
Table 6.7 Blockage Type Observations under Responsive vs. Proactive Control Strategies	164
Table 6.8 Simulation results for performance comparison	166

List of Figures

Figure 1.1 Dissertation organization.....	8
Figure 3.1 Common undesirable traffic conditions on a commuting arterial with asymmetric traffic volume.....	30
Figure 3.2 Graphical illustration of queue evolution under non-uniform (left) and uniform (right) upstream arrival rates and the associated impacts on delays and blockage time.....	31
Figure 3.3 Graphical illustration of the adverse impacts caused by the mutual blockage.....	34
Figure 3.4 The developed system's overall structure.....	35
Figure 3.5 Flowchart of the developed responsive signal control for arterials with highly asymmetric directional traffic flows	39
Figure 4.1 The adverse impacts of mutual blockages (a) wasted green duration; (b) decreased capacity	45
Figure 4.2 Graphical illustration of the impact of queue formation patterns on a lane-blockage's starting time and the resulting delay	46
Figure 4.3 Structure of the developed time-of-day arterial signal control module.....	47
Figure 4.4 Graphical illustration of through queue formation with different arrival rates	50
Figure 4.5 Graphical illustration of left-turn queue formation under different arrival rates	51
Figure 4.6 Graphical illustration of the computation for the Through Queue Group- ($Q_{i,1,1}$)	54
Figure 4.7 Relationships between the start/end of the green phase and the queue formation duration	57
Figure 4.8 Graphical illustration of delay decomposition	58
Figure 4.9 Graphical illustration of the cumulative queue length and cumulation duration for Through Queue Group- $Q_{i,3,1}$	60
Figure 4.10 Graphical illustration for delay estimation	62
Figure 4.11 Graphical illustration of the blockage impact duration	65
Figure 4.12 Delay Computation accounted for the through-lane's capacity drop	70
Figure 4.13 Graphical illustration of the left-turn residual queue	73
Figure 4.14 (a) Geometric features and (b) peak-hour volume distributions over the candidate arterial segment.....	77
Figure 4.15 Volume scenario at the study site designed for module verification	78
Figure 4.16 Outputs of the developed module.....	79
Figure 4.17 Time-space diagram and queue evolution under (a) the developed module and (b) the MAXBAND model.....	81
Figure 4.18 Performance comparison with the selected measurements of effectiveness: (a) relation between CLV and total delay; (b) relation between the left-turn bay length and the onset time of blockages	83
Figure 4.19 Time-dependent queue length evolution at Intersection 2.....	86
Figure 5.1 Adverse impacts of blockages: A) wasted green time; B) increased queue formation speed; C) Through-lane capacity reduction	91
Figure 5.2 Framework of the developed module	96
Figure 5.3 Flowchart of the structure of Response Plan 2.....	105
Figure 5.4 (a) Geometric Features, and (b) distribution of peak hour volumes over the candidate arterial segment.....	108
Figure 5.5 Comparison of queue delay on the through and left-turn lanes.....	113
Figure 5.6 Queue delay comparison of (a) through vehicles; (b) left-turn vehicles	115

Figure 5.7 Through queue length comparison at Intersection 2	115
Figure 6.1 Structure of the Proactive Real-Time Arterial Signal Control Module	124
Figure 6.2 Spatiotemporal relations between upstream flows and resulting traffic states at downstream intersections over a congested commuting arterial	128
Figure 6.3 Operational structure of the developed multi-branch multi-head LSTM-based prediction model	130
Figure 6.4 Structure of the standard LSTM unit.....	134
Figure 6.5 Architecture of the temporally-aligned multi-branch LSTM model for downstream queue length prediction.....	138
Figure 6.6 (a) Geometric features of the selected site; (b) Flow rate distribution over the 3-hour period at Intersection 1.....	141
Figure 6.7 Training Loss Curve for the Multi-Branch Multi-Head LSTM Model	143
Figure 6.8 Comparison of predicted queue spillover/blockage events: (a) through movement prediction; (b) left-turn movement prediction; (c) binary spillover/blockage detection (1: blockage, 0: no blockage)	147
Figure 6.9 Comparison of Predicted and Actual Queue Blockage Events across Models (RNN, EKF, and the developed model).....	149
Figure 6.10 Comparison of predicted queue spillover/blockage events using three-, four-, and five-branch LSTM models: (a) through movement prediction; (b) left-turn movement prediction; (c) binary spillover/blockage detection (1: blockage, 0: no blockage).....	153
Figure 6.11 Predicted spillover/blockage duration for each event across three-, four-, and five-branch LSTM models	154
Figure 6.12 Flowchart for prediction monitoring and mode switching logic	158

Chapter 1 Introduction

1.1 Research Background

Commuting arterials connecting the central business districts (CBDs) and residential areas often experience highly asymmetric directional traffic flows, characterized by high traffic volumes toward the CBDs during morning peak hours and a reversal of this pattern in the evening. Aside from the impact of insufficient capacity at critical intersections due to geometric constraints, these congestion patterns are often exacerbated by signal control systems that fail to accommodate distinctly different traffic conditions in each direction. Consequently, drivers along such a commuting arterial are likely to experience frequent stops and excessive delays caused by the mingling of through and turning vehicles, queue spillbacks from turning bays, and underuse of the left-turn phase due to the bay blockage by excessive through-traffic queues. Moreover, failing to address such often-incurred saturated flows and the accompanying mutual queue blockages may pressure more through drivers to execute undesirable lane changes and thus reduce the available capacity of both the through and left-turn lanes.

To tackle all such critical issues in congested commuting arterials, the traffic community has proposed reversed lane operations as a viable solution in practice. However, due to safety concerns, high costs, and difficulties in signal coordination, implementing coordinated signal control has emerged as a more appealing alternative, particularly for those arterials not yet fully utilizing their available capacities (Bede and Péter, 2014). Early signal control methodologies for such needs, like MAXBAND (Little et al., 1981) and MULTIBAND (Gartner et al., 1991), adapted weighted progression bandwidths to accommodate the unbalanced flows for commuting arterials not reaching saturated traffic states. Along the same line, Peng and Wang (2023) proposed a three-

step coordinated control model to optimize cycle lengths, minimize delays, and maximize bandwidth under both oversaturated and undersaturated conditions.

Despite these advancements, several critical traffic congestion issues caused by geometric constraints such as insufficient bay length or short arterial links, compounded by highly unbalanced directional volumes, remain to be tackled. Moreover, many existing arterial control models are developed under the assumption of uniform arrival rates, which simplifies the design process but fails to reflect that traffic conditions are highly fluctuating and directional in typical commuting corridors. This assumption inherently limits most such models to the pre-timed control and fixed phase sequence, reducing their effectiveness in timely response to demand surge and/or rapid queue buildup. Consequently, the resulting delays are often underestimated, and the roadway's available capacity may not be fully utilized during the formation of excessive queues and mutual lane blockages.

While the pretimed optimized offsets and phase sequences, which properly reflect dynamic mutual dependence between intersection flows, have the potential to alleviate congestion—provided that all key contributors to queue formation are fully accounted for and traffic patterns remain approximately stable—such conditions are rarely sustained on congested commuting arterials during peak-hour operations. Under near-saturation conditions, even minor traffic disruptions can result in rapid propagation of local queue spillback, triggering standing queue waves across multiple links and disrupting pre-coordinated progression. These impacts inevitably compromise the efficiency of time-of-day operations, underscoring the need for real-time adaptive control strategies capable of dynamically responding to evolving traffic patterns and emerging localized congestion.

With advancements in high-resolution traffic data collection and predictive modeling techniques, there exists a growing opportunity to move beyond time-of-day control approaches. Adaptive signal control, either reactive or proactive in nature—holds the potential to prevent queue spillbacks and mutual blockages prior to their formation, thereby enhancing the arterial performance efficiency and reliability. Nevertheless, despite the growing recognition of such a dynamic control function, most existing studies on arterial coordinated control have yet to explicitly incorporate the complex impacts of mutual impedance between through and turning traffic queues in design of either time-of-day or real-time adaptive signal plans.

Hence, given the significant gaps that remain in the current body of research, this research, focused on developing a coordinated signal control system for congested commuting arterials with highly asymmetric directional flows, intends to address the following critical issues:

- How can signal control strategies be designed to accommodate highly directional volume imbalances, and to ensure efficient operations for the congested direction without degrading the performance in the non-congested direction?
- How can the adverse effects of mutual blockage between through and left-turn queues be reliably captured in time and effectively mitigated?
- How can mutual queue blockages and the resulting congestion impacts be reliably detected in advance to support the activation of customized proactive signal control plans? and
- How can signal control plans, designed for minimizing the impacts of various queue blockages, be correctly selected and executed in time to prevent further queue formation and the emergence of traffic blockage?

1.2 Research Objective

To address the critical issues identified in the research background, this study aims to develop a real-time coordinated signal control system customized for congested commuting arterials with highly unbalanced directional flows. More specifically, the proposed system for the research objective has been designed with the following essential functions:

- tackle the distinct traffic congestion patterns over commuting arterials with highly unbalanced directional flows during peak hours with an integrated objective function of minimizing delays in the high-volume commuting direction while maintaining the maximum progression in the opposing direction;
- concurrently optimize offsets and phase sequences to achieve the control objective of minimizing the system's total delay without compromising the traffic progression quality;
- account for the impacts of diverse upstream traffic streams and their complex interactions on their arrival rates in optimization of the target intersection's phase sequence for reducing the queue and overflow durations;
- compute different types and levels of traffic delays over each arterial link and intersection under various possible traffic demands and queue formation patterns;
- account for the impacts of various types of mutual blockages between through and left-turn vehicles in design of all arterial intersections' optimal signal offsets and phase sequences to minimize the associated adverse impacts and best use the available roadway capacity;
- classify the arterial's various congestion patterns into multiple levels in real time, and implement responsive signal control strategies to tackle the evolving traffic state and minimize the congestion impacts; and

- perform short-term queue length prediction, using spatial-temporal information and geometric characteristics, to project traffic queue states in advance, enabling early detection of potential overflows and subsequent execution of proactive control plans to minimize the likelihood of incurring queue spillbacks and mutual blockages.

1.3 Dissertation Organization

In alignment with the defined research objectives, this study is structured into seven distinct chapters, each dedicated to addressing specific primary tasks essential for the development of the proposed commuting arterial signal control system. Figure 1.1 illustrates the interconnectedness of these chapters. A concise description of the core substance in each chapter is presented below:

- **Chapter 2:** This chapter presents an extensive literature review, focusing on offline and real-time optimal control models for arterial signal coordination. Most key findings from the reviewed studies, along with remaining critical issues, are also reported in this chapter.
- **Chapter 3:** The structure and operational flows of the proposed coordinated control system for commuting arterials are presented in this chapter. Also included are detailed descriptions of the customized features and functions designed to tackle identified congestion issues, along with the required modules, inputs, and outputs.
- **Chapter 4:** This chapter introduces the base module for the system's time-of-day operations. It begins by introducing the objective function to accommodate the highly unbalanced directional flows in commuting arterials. This is followed by an in-depth discussion of the time-of-day module's core logic and associated sets of constraints, including varying arrival rates and the intricate interactions between through and left-turn vehicles. Additionally, this chapter delves into the general formulations of delay patterns, providing the flexibility for the module to effectively select the optimal set of phase

sequences and offsets. The results of extensive numerical analyses with a real-world commuting arterial for performance evaluation are also reported in this chapter.

- **Chapter 5:** This chapter introduces the responsive dynamic signal control module, the core of the proposed real-time adaptive signal control operation, which is built upon the time-of-day module developed in the previous chapter. The chapter begins with a detailed description of the methodology to monitor spatial-temporal queue distributions, followed by a discussion of rule-based methods for congestion pattern classification. The operational logic for selecting and activating customized signal plans in response to the detected and classified traffic conditions has also been reported in this chapter. Results of numerical experiments conducted on a congested arterial in Bethesda, Maryland, for performance assessment of the proposed adaptive system constitute the last part of this chapter.
- **Chapter 6:** This chapter presents the proactive dynamic signal control module, the second alternative within the developed real-time adaptive signal control framework. Same as the responsive module, it builds upon the time-of-day control foundation. The proactive module features its adoption of a prediction model for short-term traffic queue state, which, designed for early identification of potential queue overflows and blockages, enables the set of most effective signal control strategies to implement ahead of the likely formation of congestion and queue blockage. The rationale for adopting an LSTM-based approach for developing such a prediction model over conventional time-series models is discussed in detail, followed by the development of a modeling structure that enables event-based detection of predicted queue blockages. Numerical experiments for comparing the prediction performance of the proposed model against baseline methods are also reported in this chapter. The concluding section is dedicated to a case study demonstrating how the

prediction model is applied within the proactive control configuration to evaluate its effectiveness in mitigating congestion under real-world traffic conditions.

- **Chapter 7:** This chapter concludes the dissertation by summarizing its key contributions, including the essential features of the time-of-day base module and the demonstrated benefits of advancing to a real-time adaptive signal control operation. It has also highlighted potential directions for future research, including enhancing the developed system's computing algorithm, critical issues that may be addressed with emerging technologies, and major challenges associated with implementing such a system on an area-wide network level.

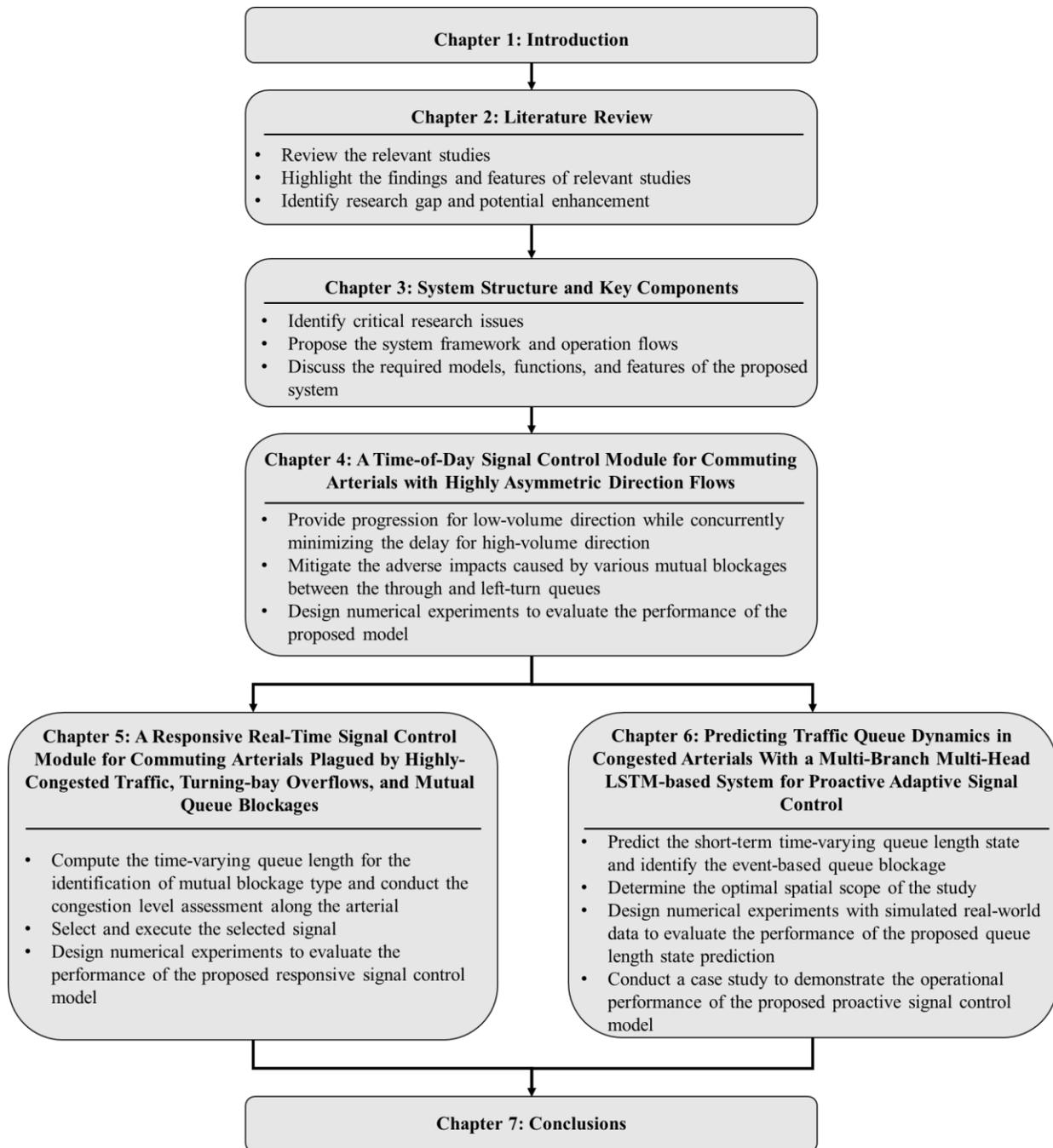


Figure 1.1 Dissertation organization

Chapter 2 Literature Review

2.1 Introduction

As mentioned in Chapter 1, the primary objective of this research is to devise an efficient signal control system tailored for commuting arterials characterized by highly asymmetric directional flows. The review presented in this chapter will concentrate exclusively on studies pertaining to offline arterial signal coordination optimization, commuting arterial controls, and adaptive real-time signal controls. Each subsequent section will explore key findings within these areas, describe their methodologies, including control objectives, underlying assumptions, limitations, and reported performance. Moreover, significant findings with potential relevance to this research and any critical issues identified in the literature, yet to be addressed, will be thoroughly discussed in the concluding section.

2.2 Offline Arterial Signal Coordination Optimization

In the domain of arterial signal coordination optimization, signal control algorithms are typically designed with two primary objective functions: 1) Progression maximization and 2) Delay minimization. Each objective function has its distinct features, with either progression maximization primarily applied in undersaturated conditions or delay minimization commonly used in saturated scenarios where traffic demands approach the roadway's capacity.

2.2.1 Control Objective: Progression Maximization

Existing studies with the control objective of progression maximization focus on enabling vehicles within these bands to traverse multiple intersections or even the entire arterial network seamlessly. The pioneering work on maximizing two-way progression bandwidths was initially proposed by Morgan and Little (1964). Little (1966) later extended this study using mixed-integer linear programming to transform the formulations into an operational optimization model. This work led

to the development of MAXBAND (Little et al., 1981), which incorporates initial intersection queues and left-turn phase sequences in the design of progression band maximization.

To adapt to fluctuations in link flows, arising from disparities in turning-in and turning-out volumes between intersections, and to address the diverse impacts of link volumes on progression band design, researchers have employed different methodologies. Examples of such studies include MULTIBAND (Gartner et al., 1991) and PASSER-IV (Chaudhary and Messer, 1993), both of which introduce variable bandwidths for arterial links based on their respective traffic volumes. Similarly, Messer et al (1973) have endeavored to generate optimized offsets coupled with multiphase signals, tailored to the volume observed across different links. Recent advancements in this area include: 1) Zhang et al. (2015) relaxing the requirement for symmetrical progression bands in MULTIBAND to allow the generation of asymmetrical bands with varying widths for different arterial segments; 2) Kim et al. (2016) utilizing closed-loop signal data to replicate actual dynamic bandwidth durations; and 3) Li et al. (2022) further considering factors such as phase pattern selection and queue clearing time for improved refinement of the model.

Grounded in the aforementioned landmarking models, which produce the optimal performance when an arterial's through traffic is the dominant flow (Zhang et al., 2015), several studies have explored arterials with various traffic patterns or geometry features. Yang et al. (2015) introduced a multi-path progression model tailored for arterials experiencing significant path-flows, a common occurrence, particularly in commuter arterials bridging freeway and surface-street traffic. In a similar vein, Yen et al. (2019) developed a three-staged signal optimization model aimed at ensuring smooth progression for both through and turning movements along arterials grappling with heavy turning flows within the constraints of bay lengths. Building upon this work, Yen et al. (2021) further refined their model to facilitate multi-path flow progression using sole

intersection turning counts. Subsequently, Lu et al. (2023) advanced the concept of path progression by extending it to the regional scale.

The manifold benefits of progression, including increased travel speed along arterials, reduced traffic emissions due to fewer stops, decreased probabilities of rear-end collisions owing to uniform vehicle speeds, and fully-utilized capacity due to smoother traffic flows, have been substantiated by previous studies. Nevertheless, the efficacy of progression inevitably diminishes as traffic conditions approach the capacity of any arterial link—either saturated or oversaturated—due to disruptions caused by excessive queues (Shepherd, 1994; Sun et al., 2015). Such states of traffic queues often disrupt traffic flow progression because: 1) spillback to upstream intersections would impede the progress of both arterial and cross-street flows toward downstream intersections; 2) under high demand, only limited bandwidth remains available for progression, as most green time is consumed in discharging initial queues. In such scenarios, standard signal plans designed to facilitate the smooth forward progression of vehicle platoons along arterials deteriorate rapidly, necessitating alternative control objectives (Quinn, 1992). Moreover, this issue can escalate to a more complex level, rendering conventional progression designs ineffective, especially when arterials must accommodate high volumes of turning-out-and-in traffic. Chen et al. (2019) addressed the impact of such turning bay overflows on through progression bands and proposed a model to mitigate this impedance caused by left-turn spillback. However, the traffic community has yet to adequately address various scenarios involving mutual queue blockage, the adverse impacts of resulting intermingling of through and left-turn vehicles, and the nature of high-demand distribution during peak hours, all with the overarching control objective of maximizing progression.

2.2.2 *Control Objective: Delay Minimization*

Within the realm of prioritizing delay minimization as the primary control objective, the family of TRANSYT models (Robertson, 1969) stands out as a leading-edge tool for arterial signal design, owing to its unique macroscopic-based simulation-optimization methodology (Wong, 1996; Wong et al., 2002). This model incorporates platoon dispersion between adjacent intersections into the computation of signal offsets and timing plans. Over time, it has evolved into TRANSYT-17, distinguished by several new features. These include a congested platoon dispersion model tailored for spillback scenarios, the ability to generate timings for prioritizing specific traffic types such as buses and trams, the integration of new computing algorithms such as genetic algorithms, and the simulation model with the newly added queue simulation technique for scenarios such as uneven lane usage/lane starvation.

TRANSYT-7F (McTrans, 2008) serves as the U.S. version of its family members. Other notable simulation models along similar lines include those developed by Park et al. (2000), Yun (2006), Zheng and Li (2023), and Zheng et al. (2023).

In addition to the simulation-based optimization models, numerous prominent models share the common objective of minimizing delay for signal design with various constraints, often delving into theoretically formulating delay estimations, exemplified by calculating the triangle area surrounded by the cumulative arrival and departure curve (Rouphail et al., 1999). Similarly, Chang and Lin (2000) developed a discrete dynamic optimization model centered around an arrival-departure curve targeting minimizing total delay during oversaturated periods while optimizing cycle length and green time under pretimed control. Other notable studies addressing delay minimization include Chang and Sun (2004), Yun and Park (2006), Stevanovic and

Stevanovic (2007), Liu and Chang (2011), Keyvan-Ekbatani et al. (2012 and 2013), Sun et al. (2015), and Yu et al. (2018).

Navigating the nonlinear nature of mathematically formulated delay functions presents a perennial challenge for researchers, who strive to strike a balance between computational efficiency and convenience for use in practice. As a result, heuristic approaches such as Genetic Algorithms (GA) are often embraced (Wong et al., 2002; Stevanovic et al., 2007; Liu and Chang, 2001; Arshad et al., 2020). However, in pursuit of simplicity for application, researchers frequently opt for uniform arrival rates rather than accounting for varying rates associated with upstream streams, which will constrain the study scope with a fixed phase sequence (Hillier and Rothery, 1967). While such a design may suffice for many cases when coupled with flexible offsets and green splits, it falls short during peak hours when fluctuations in turning-out volume are commonly observed. In such scenarios, achieving the desired level of performance necessitates the implementation of flexible phase sequencing. Thanks to the advancements in technology, it enables the efficient resolution of nonlinear challenges in modeling signal delay and control plans, fostering ongoing research into the development and application of optimizing phase sequences techniques (Achterberg, 2023).

2.2.3 Multi-Objective Control

Initially highlighted by Gartner et al. (1991), many traffic agencies recognize the benefits of concurrently having both progression maximization and delay minimization as control objectives. Consequently, some studies have proposed different models to integrate these strategies seamlessly. Two primary approaches have emerged: 1) modifying delay-based programs to maximize bandwidth, and 2) adjusting bandwidth-based signal timing plans to minimize delay.

Examples of studies applying the first approach include those by Cohen (1982), Skabardonis and May (1985), Cohen and Liu (1986), and Liu (1988). In Liu's (1988) study, a bandwidth-constrained delay minimization model was proposed, leveraging the bandwidth solution as the initial value for TRANSYT-7F (Wallace et al., 1984). This method specifies starting and ending time points at each intersection of the two-way band in an additional input stream, ensuring that the red time does not encroach on the through band, thereby preserving the progression effectiveness while optimizing delay reduction through optimized offsets and green splits. Conversely, for the second approach, Wallace and Courage (1982) introduced a new measure of progression quality, known as the number of progression opportunities (PROS), aimed at achieving the dual objectives of decreased stops and delay.

Recent studies in the past decade have proposed models with multi-objective frameworks to address various issues for different traffic conditions. For instance, Lertworawanich et al. (2011) introduced a new methodology for optimizing signal timing controls in oversaturated networks based on the cell transmission model, employing genetic algorithms for solution optimization. Zhou et al. (2017) developed optimization models enabling uneven double cycling (UDC) in arterial signal coordination to mitigate excessive delay at minor intersections, addressing disparities in traffic volumes. Meanwhile, Chen et al. (2022) enhanced progression maximization models for arterials with high turning volumes, offering flexibility in adjusting progression bandwidths to target delay reduction, particularly for turning traffic.

2.3 Commuting Arterial Traffic Controls

Commuting arterials linking central business districts (CBDs) and residential areas often exhibit unique traffic patterns, particularly during peak hours, with highly asymmetric directional flows. For instance, morning peak hours typically exhibit heavier traffic volume towards CBDs, while

the reverse occurs during evening peak hours. Addressing these asymmetries in traffic patterns is crucial for the design of an effective control plan. Reversible lane control has long been favored as a solution for such scenarios. Wolshon and Lambert (2006) provided an overview of reversible roadway applications and their efficacy in tackling traffic flow issues. Li et al. (2013a and 2013b) proposed a dynamic lane-use assignment model at signalized intersections to determine optimal phase sequences, lane-use assignments, and the need for reversible lanes. Similarly, Yu and Tian (2014) verified the effectiveness of reversible lanes and developed a bi-level programming model to optimize lane allocation and predict driver reactions. Advancing this concept, Yao et al. (2018) integrated left-turn bays and various vehicle categories into reversible lane control systems. To accommodate fluctuating traffic conditions, several studies proposed various frameworks for dynamic lane reversal, enabling rapid and automatic adjustments to lane directionality in response to real-time traffic conditions (Hausknecht et al., 2011; Mao et al., 2020; Chen et al., 2021; Zhou et al., 2023).

Despite its widespread use, reversible lane control still lacks standardized practices for implementation, leading to concerns about uniformity (Wolshon and Lambert, 2006). Bede and Péter (2014) highlighted safety concerns and practical challenges associated with altering system structures, especially the transition area. Additionally, strict operational requirements, high overhead costs, and the lack of intersection coordination pose further limitations on the application of reversible lanes (Peng and Wang, 2023). Acknowledging these limitations, some studies have turned to coordinated signal control as an alternative to address the issue of unbalanced directional flows on commuting arterials.

The earliest attempt to account for the uneven distribution included studies like MAXBAND (Little et al., 1981) and MULTIBAND (Gartner et al., 1991) family, which introduced

weighting factors to adjust the bandwidth according to unbalanced demands. However, traditional MAXBAND and MULTIBAND approaches faced some limitations, because the maximum bandwidth solution may not always be achievable without properly specified weighting factors. Lu et al. (2012) addressed these issues by proposing a novel two-way bandwidth maximization model that not only enhances total bandwidth but also incorporates a bandwidth proration impact factor, allowing users to control the importance of meeting target bandwidth demand ratios. Additionally, leveraging the unique design of overlapping phases, Ji and Cheng (2022) optimized signal phase sequences for arterials with unbalanced traffic flows. However, as noted earlier, forward progression maximization is typically applicable only when traffic demands are undersaturated, a condition rarely met during peak commuting hours when traffic approaches the intersection's capacity (Peng and Wang, 2023). Thus, they proposed a three-step algorithm to tackle such challenges. Firstly, they generated optimal cycle lengths and green splits to maximize throughput in both directions, followed by delay minimization using the Lighthill-Whitham-Richards theory for the oversaturated direction to obtain optimal offsets. Lastly, they identified phase schemes to ensure variable bandwidth maximization in unsaturated directions.

Despite the acknowledged importance of effective control measures for commuting arterials, there remains a noticeable lack of research dedicated to this area, leaving many issues unresolved. One particularly significant concern is the oversight of geometry constraints, notably turning bay capacity. While some studies have attempted to address queue overflows and their operational impacts, they typically focus on isolated intersections, limiting their applicability to broader, more complex scenarios involving multiple intersections in congested arterials (Yao et al., 2023; Zhao et al., 2024). The frequent overflow of queues beyond turning bay capacities often leads to mutual blockage between through and left-turn queues, highlighting the need for further

investigation into the adverse impacts of the intricate interactions between through and left-turn vehicles.

2.4 Adaptive Signal Control

Offline coordinated control is to optimize signal plans using historical traffic data rather than real-time information, thus limiting its effectiveness when traffic patterns deviate from historical norms. To address this limitation, a vast array of adaptive signal control models and systems have leveraged the advent of new detection technologies and emerged over the past two decades to offer dynamic responses to the fast-evolving traffic fluctuations and surges in real-time. Based on the approach to handling traffic conditions, adaptive signal control can be divided into two main types: responsive and proactive signal control. While the former reacts to current traffic conditions in real time, adjusting signal timing dynamically based on the feedback from sensors, the latter first forecasts future traffic conditions and then optimizes signal timing preemptively using real-time collected and stored data along with the embedded predictive models.

2.4.1 Responsive Signal Control

In the realm of responsive signal control systems, notable advancements have been made with deployments such as SCATS (Sims and Dobinson, 1980) and SCOOT (Hunt et al., 1982), which have demonstrated promising performance in field operations. SCATS, Sydney Coordinated Adaptive Traffic System, stands as a pinnacle of adaptive traffic management solutions. Originating from Australia, SCATS has garnered adoption in over 200 cities spanning 30 countries worldwide (NSW, 2025). Its operational logic revolves around the intricate decomposition of vast traffic networks into manageable subnetworks, each overseen by intersections operating on a synchronized cycle length. Within the subnetworks, SCATS utilizes a two-tiered hierarchy consisting of roadside control cabinets and remote computers. Roadside cabinets are responsible

for managing standard responsive signal control and processing traffic information, while remote computers are used to oversee area-based traffic control (Lowrie, 1982). The responsive control process relies on synthesized flow data derived from measured degrees of saturation and maximum flow rates from strategic detectors. This dataset serves as the foundation for determining critical signal parameters (McCann, 2014). Inter-subnetwork communication is orchestrated through a sophisticated “married/divorced” mechanism, which regulates one- or two-way progression based on directional traffic volumes (Luk and Sims, 1982). This mechanism relies on weighted local progression bandwidths, facilitating uninterrupted vehicle flow across subnetwork boundaries. Notably, SCATS' scalable architecture ensures that the addition of intersections does not compromise the system's performance, making it a resilient solution for evolving urban traffic management needs.

Originating from the Transport Research Laboratory (TRL) in the 1970s, SCOOT, known as the Split Cycle and Offset Optimization Technique, is often regarded as the traffic-responsive counterpart to TRANSYT. As SCOOT has evolved to version 7, it has continuously refined its capabilities by integrating advancements in technology, notably through the incorporation of various newly emerged detectors. This enhancement has significantly improved SCOOT's ability to capture and process quality real-time traffic data. By leveraging the platoon dispersion model in conjunction with this high-quality real-time data, SCOOT can dynamically adapt splits, offsets, and cycle lengths to minimize delays and optimize traffic flow efficiently (TRL, 2025).

While SCATS and SCOOT are widely adopted for their efficiency in urban traffic management, their performance becomes less reliable under saturated traffic conditions, as noted by Lee et al. (2002), raising concerns about their applicability in highly congested commuting corridors. Notably, neither system sufficiently accounts for the adverse impacts of queue spillovers

and mutual lane blockages in their control objectives or delay estimation methods. SCATS, with detectors placed at stop lines or slightly upstream, can only detect spillovers from turning bays after they have already interfered with adjacent traffic, limiting its responsiveness. SCOOT, on the other hand, typically places detectors at the roadway link's upstream to pre-estimate the downstream turning ratios, which restricts its ability to capture the dynamic interactions between through and turning movements. Moreover, both systems tend to rely on extending cycle lengths to improve intersection efficiency—a strategy that may be ineffective or even counterproductive when dealing with congestion patterns caused by turning bay overflows or residual through-traffic queues. These limitations underscore the need for control systems specifically designed to handle the complex conditions commonly observed in commuting arterials, such as highly unbalanced directional flows, intricate vehicle interactions, and frequent mutual blockages.

Other adaptive systems deployed in the field, such as OPAC, PRODYN, and ACS-Lite, have also made their unique contributions to traffic management under moderate or congested but not near-saturated traffic conditions. OPAC (Gartner, 1983) is a dynamic strategy for a demand-responsive and decentralized signal control system, developed and tested in the United States, and primarily designed for isolated intersections. PRODYN, developed by Henry et al. (1983), employs dynamic programming to minimize delays, supplemented by an upper-level supervisor for signal coordination between intersections. In contrast, ACS-Lite (Luyanda et al., 2003) focuses on cost-effective solutions by incorporating a regional “field master” to implement the simplified ACS logic compatible with existing signal controllers. This framework is designed to update traffic demand data at the time-of-day level, accommodating mid-term fluctuations from typical patterns.

Recent studies in responsive signal control, albeit largely theoretical exploration in nature, have been extensive in the traffic literature (Cai et al., 2009; Xie et al., 2012; Smith et al., 2013;

Zheng and Recker, 2013; Varaiya, 2013; Christofa et al., 2016; Ghasempour and Heydecker, 2019; Noaen et al., 2021). It is worth noting that most of such models in the literature primarily prioritize the benefits of through movements or focus solely on isolated intersections. Addressing heavy turning flows, both inbound and outbound, to ensure equitable benefits amid highly unbalanced directional flows in real-time control environments still remains a significant challenge to the traffic community.

2.4.2 Proactive Signal Control

Unlike the responsive signal control to react to the detected congestion, the proactive controls feature taking a preemptive approach to prevent traffic congestion from reaching critical levels. With the predicted congestion patterns and the advanced adjustments in signal plans, proactive systems can prevent or postpone the formation of congestion, thereby minimizing delays and maintaining traffic conditions. This proactive approach is especially crucial in heavily congested commuting arterials, where traffic queues and delays can quickly escalate, leading to oversaturated traffic conditions. Therefore, the implementation of proactive signal control strategies is essential for effectively managing traffic conditions in congested traffic environments.

One of the well-known proactive signal control models is RHODES (Mirchandani and Head, 2001), a real-time traffic-adaptive system designed to optimize traffic flow through a hierarchical control architecture. RHODES leverages real-time detector data to predict future traffic conditions at multiple resolution levels—ranging from individual vehicles to vehicle platoons—enabling it to alleviate conflicting demands across the network. At the intersection level, it incorporates the REALBAND module developed by Dell and Mirchandani in 1995, which facilitates acyclic signal phase transitions and minimizes delays for individual vehicles. Following RHODES, subsequent studies have proposed other proactive signal control models. For instance,

Haijema and Hendrix (2014) formulated a dynamic adaptive traffic signal control model based on a Markov decision process. Similarly, Yang and Jayakrishnan (2015) developed a real-time, network-wide traffic control scheme, structured around a two-level hierarchical architecture and designed to optimize both long-term and short-term signal operations for minimizing total delay.

Having a reliable traffic flow prediction, as highlighted by Mirchandani and Head (2001), is essential for the success of proactive signal control. If the predicted traffic flow deviates significantly from the actual conditions, the effectiveness of proactive signal control may be compromised, leading to a worsening of traffic conditions.

The realm of short-term traffic flow prediction has witnessed a surge in the traffic research community, particularly with the emergence of Intelligent Transportation Systems (ITSs) (Lv et al., 2015). However, traffic flow prediction is still a difficult problem due to the intrinsically highly nonlinear and stochastic characteristics of complex interactions between traffic streams (Kang et al., 2017). Studies in traffic prediction can generally be categorized into two main groups: parametric models and nonparametric models (Ma et al., 2015). Parametric models typically employ time-series analysis methods to address traffic prediction challenges. One notable example is the Kalman Filter method, which formulates the regression problem within a state space framework and updates the algorithm's parameters with actual measurements to minimize variance. (Xie et al., 2007; Ojeda et al., 2013; Guo et al., 2014; Zhou et al., 2019; Cai et al., 2019). Another widely utilized parametric model is the autoregressive integrated moving average (ARIMA) model, renowned for its linear structure. Kumar and Vanajakshi (2015) introduced a Seasonal ARIMA model specifically designed for short-term traffic flow prediction, demonstrating its capability to produce accurate forecasts with limited input data. Other notable studies exploring similar

approaches include those by Williams et al. (1998), Williams (2001), Cetin and Comert (2006), Tan et al. (2009), Zhang et al. (2014), and Alghamdi et al. (2019).

In addition to the Kalman Filter and ARIMA, several other well-established parametric methods have been explored in the literature. Zhou et al. (2019) applied historical average methods within a deep learning-based multi-model integration framework, demonstrating improved accuracy in handling large variations and uncertainties in traffic flows. Zeng et al. (2018) proposed a multi-variable Grey forecasting model featuring a dynamically optimized background-value coefficient, which significantly enhanced the prediction performance over traditional fixed-coefficient models. Spectral analysis approaches have also shown promise; for instance, Tchrakian et al. (2011) developed a real-time forecasting algorithm using modal functions derived from spectral analysis, enabling adaptive short-term traffic predictions with continuous online updates. Ghosh et al. (2009) contributed a multivariate structural time-series model capable of separately modeling trend, seasonality, and calendar effects, proving effective for real-time traffic forecasting across multiple urban intersections.

However, despite their widespread use, parametric models, including Kalman Filters and ARIMA, are constrained by inherent limitations. These models rely heavily on linear assumptions and fail to effectively capture the nonlinear dynamics and spatiotemporal dependencies that characterize real-world traffic systems. Kalman Filters, while effective for linear systems with Gaussian noise, struggle with the nonlinear, complex dynamics of traffic flow, particularly during rapid transitions from undersaturated to saturated conditions. Similarly, ARIMA assumes stationarity in traffic data, an assumption that rarely holds in practice, and its linear structure limits its ability to model interactions between variables, such as signal timings and existing queue length.

Both models also face challenges with outliers and missing data, making them less robust for real-time traffic monitoring and prediction.

Recognizing the limitations of traditional linear models, recent studies have adopted the Extended Kalman Filter (EKF) for traffic prediction tasks (Van et al., 2011; Chang et al., 2023; Shih et al., 2024). EKF addresses non-linearity by linearizing the system at each time step using Jacobian matrices, providing local approximations of system dynamics. However, this approach is only effective for small deviations around the linearization point. In real-world traffic systems, where non-linearities such as queue spillback, mutual blockage, and variable signal timings are common, EKF often fails to capture the system's complexity. Moreover, a major limitation of EKF lies in its scalability. Traffic prediction involves high-dimensional, multi-variable inputs (e.g., vehicle count, occupancy, speed, and signal plans) with non-linear interactions across intersections. EKF is not inherently suited for modeling such complexity, and extending it to multi-input, multi-output systems significantly increases computational cost due to the need for Jacobian calculations and matrix inversions at every step. Furthermore, EKF is designed for short-term state estimation and lacks the capability to model long-term dependencies, which are critical in predicting cumulative effects such as queue buildup and delay propagation. The method also assumes a Gaussian distribution and independent noise, which are inconsistent with field traffic data, where measurements are often noisy, correlated, and heteroscedastic—especially during peak hours or under spillover conditions. Additionally, EKF is sensitive to measurement inconsistencies, such as instances where flow is recorded as zero despite long queues due to sensor limitations. Such mismatches can cause the filter to diverge or yield unreliable estimates, particularly in large-scale, dynamic environments. Taken together, these limitations restrict the EKF's effectiveness for robust, multi-step traffic prediction in complex urban networks.

For non-parametric methods, neural network algorithms stand out prominently for their remarkable flexibility and robust learning, generalization, and prediction capabilities (Karlaftis and Vlahogianni, 2001). With the emergence of deep learning, models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have gained significant attention in traffic flow prediction, as these models are particularly effective at capturing the complex temporal dependencies and nonlinearities in traffic data. The Long Short-Term Memory (LSTM) network, a specialized variant of RNNs, has become especially popular due to its ability to capture long-range dependencies without suffering from the vanishing gradient problem, combined with its architecture of memory cells and gates, enables LSTMs to model intricate, nonlinear relationships through nonlinear activation functions. This makes LSTMs especially effective for short-term forecasting. Tian et al. (2015) introduced a Long Short-Term Memory Recurrent Neural Network (LSTM RNN) model, leveraging three multiplicative units within the memory block to dynamically determine the optimal time lags without a predefined time duration for input data.

Kang et al. (2017) explored the impact of various input settings on LSTM prediction performance, noting that incorporating occupancy/speed information can enhance overall model performance. Other studies utilizing LSTM include Shao and Soong (2016), Qiao et al. (2017), Liu et al. (2017), Mackenzie et al. (2018), Tian et al. (2018), Xiao et al. (2019), Mou et al. (2019), Rahman and Hasan (2021), Shu and Xiong (2021), and Redhu and Kumar (2023).

While extensive research has been devoted to short-term traffic flow prediction, much of it remains confined to individual intersections, often overlooking the spatial dependencies across a network of intersections (Narmadha and Vijayakumar, 2023) and underrepresenting the broader spatial dynamics of urban traffic systems (Yang et al., 2019). To address this limitation, recent

studies have increasingly emphasized the importance of capturing traffic pattern propagation by integrating both spatial and temporal dimensions into the structure of the employed predictive models. One particularly effective approach involves the use of Convolutional Neural Networks (CNNs), which have demonstrated exceptional capabilities in learning spatial features in fields such as computer vision and image processing.

Motivated by CNN's success in these domains, researchers have applied similar architectures to traffic prediction with promising outcomes. For example, Yu et al. (2017) proposed spatiotemporal recurrent convolutional networks that combine CNNs to extract spatial traffic patterns and LSTMs to model temporal congestion dynamics, achieving notable predictive accuracy. Subsequent studies have further refined these frameworks, integrating spatial-temporal learning for enhanced performance (Feng et al., 2018; Duan et al., 2018; Zhang et al., 2019; Chen et al., 2022), demonstrating the value of jointly modeling spatial dependencies and temporal evolution in traffic flow prediction.

Despite these advancements, current spatiotemporal models still face notable limitations. Many focus primarily on a single upstream intersection, which constrains their capacity to capture the extensive and complex propagation impacts of traffic flows. Even in models that incorporate multiple upstream intersections, the dynamic travel time between upstream and downstream intersections is often oversimplified, frequently neglecting key influencing factors such as signal timings, link lengths, and queue dynamics. As a result, these models may fall short in accurately reflecting the complex and evolving interactions that govern the efficiency of traffic flow movements across multiple intersections.

2.5 Discussion

This literature review reveals several critical gaps and challenges in the current body of research on coordinated signal control systems. While prior studies have demonstrated the potential of coordinated control under target traffic conditions, key limitations remain unaddressed—particularly in the context of commuting arterials characterized by highly directional peak-period flows. The following research gaps highlight the need for a more tailored and responsive control strategy:

- Tailoring coordinated signal control systems to accommodate the unique traffic patterns observed in commuting arterials. These areas often experience persistent congestion and long queues in one direction nearing capacity, while the roadway capacity for the opposing direction is underused.
- Optimizing the offsets and phase sequences concurrently under varying arrival patterns to effectively alleviate the congestion and minimize the delays along the arterial.
- Addressing the intricate interactions between through and left-turn vehicles, particularly concerning the adverse impacts of mutual blockage when queues spill over the turning lanes.
- Advancing offline coordinated signal control systems to operate in real-time responsive environments, allowing for timely adaptation to the evolving congestion patterns during peak commuting periods and the resulting complex interactions between through and left-turn vehicles.
- Developing a short-term traffic flow prediction model that integrates temporal, spatial, geometric, and dynamic traffic factors to enable accurate estimation of traffic queue state, thereby supporting the exercise of proactive signal control strategies.

To address these vital issues, this dissertation aims to develop a coordinated signal control system tailored specifically for commuting arterials with highly asymmetric directional flows.

Chapter 3 System Structure and Key Components

3.1 System Structure Overview

Most congested commuting corridors exhibit a unique traffic pattern where the predominant flows move towards the central business districts (CBDs) and/or office parks in the morning peak period, but the traffic pattern will reverse the distribution and travel direction during the evening peak hours. In the primary travel direction, most intersections, due to the surge of arriving vehicles from both upstream links and side streets, often experience excessive delays, overflows from turning bays, and queue spillback to upstream intersections. The resulting congested traffic state may incur various types of lane blockages and contribute to the formation as well as propagation of the so-called “congestion tree” over the entire commuting arterial. Depending on the spatial coverage of the congestion trees, such queue patterns may fail the entire signal plan for either progression maximization or delay minimization, despite that the roadway capacity for the low-volume directional flows is often underutilized.

An example of such a commuting arterial is MD 355, shown in Figure 3.1, which links the District of Columbia with the northern residential area in Bethesda, Maryland. During the evening peak hours, it often exhibits a highly asymmetric traffic distribution between the two directions, where the northbound volumes are significantly higher than those on the southbound lanes. Such highly unbalanced volume distribution over the congested commuting arterial will inevitably evolve its overall traffic conditions to the following undesirable states:

The steady high volume in the commuting direction may surpass the capacity of some intersections, thus resulting in residual queues from cycle to cycle;

Most intersection approaches in the high-volume direction often encounter excessive queues and delays, due to the upstream arriving flows and constantly turning-in traffic streams from the side streets, as shown in Point A in Figure 3.1;

Through queues in the high-volume direction are likely to spill over the left-turn bay, and consequently block left-turn vehicles, which may in turn prevent the through flows from progression or even cause mutual blockages and a rapid growth of congestion trees over the entire arterial, as shown in Point B in Figure 3.1;

- Left-turn queues may exceed the capacity of the turning bay and spill onto the neighboring through lane, thereby obstructing the effective progression band for through vehicles and reducing the arterial link's available capacity, as shown in Point C in Figure 3.1;
- Mutual blockage between through and turning vehicles, if incurred, over short arterial links may result in signal failure at critical intersections and hinder traffic streams from all approaches to moving onto their downstream links, as depicted in Point D in Figure 3.1;
- The rapid formation of traffic queues in the arterial's congested direction may also prevent side-street traffic at major intersections from turning into their target direction and lane, thus further spreading the traffic queues and undesirable delays to side streets and the entire corridor, as shown in Point E in Figure 3.1; and
- Traffic flows in the low-volume direction often cannot efficiently progress through consecutive intersections and suffer undue delays, because most such signal plans tend to favor the high-volume direction, as shown in Point F in Figure 3.1.

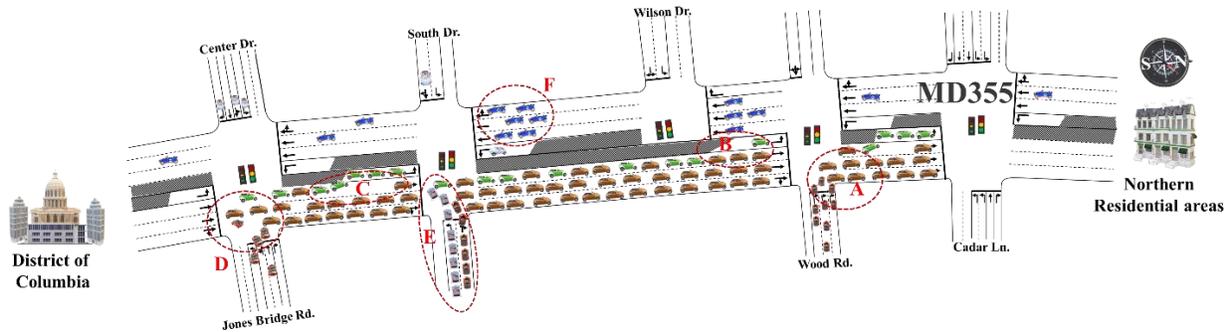


Figure 3.1 Common undesirable traffic conditions on a commuting arterial with asymmetric traffic volume

As discussed in Chapter 2, reversible lanes have been explored as a strategy to manage highly directional and congested traffic patterns on commuting arterials (Wolshon and Lamber, 2006; Li, 2013a, 2013b; Yu and Tian, 2014; Yao et al., 2018). While effective in certain contexts, reversible lane control faces significant limitations, including safety concerns, high implementation and operational costs, and difficulty in coordinating signals (Bede and Péter, 2014). Given these challenges, coordinated signal control has emerged as a more practical alternative—especially when many intersections along the arterial are underutilized (Peng and Wang, 2023).

Early efforts, such as MAXBAND (Little et al., 1981) and MULTIBAND (Gartner et al., 1991), introduced a weighted progression design to accommodate unbalanced demand. Although progression-based coordination can reduce delays, emissions, and safety risks under undersaturated conditions, its effectiveness diminishes as queues increase during peak periods (Shepherd, 1994; Sun et al., 2015), as elaborated in Chapter 2. Recognizing this, Peng and Wang (2023) proposed a three-step control structure for arterials with asymmetric flows: generating optimal signal timings to maximize throughput, minimizing delays using kinematic wave theory in oversaturated directions, and optimizing phase sequences to maintain progression in the less congested direction.

Despite the significant advancement by Peng and Wang (2023) in proposing a coordinated control model with integrated objectives for commuting arterials, two critical issues remain largely overlooked. Firstly, geometric constraints, notably turning bay length, pose significant challenges in design of signal progression for traffic streams entering an arterial's link from various upstream intersection approaches. The frequent overflows from a turning bay often lead to mutual blockages between the through and left-turn queues, as discussed in Figure 3.1. Secondly, existing studies often assume the vehicle arrival rates to follow a uniform distribution, thus neglecting the complex traffic queue dynamics and blockage patterns attributed to the highly fluctuating volume from different arriving traffic streams. Failing to tackle such complex traffic dynamics may result in difficulties in 1) accurately identifying the onset time of mutual blockage, 2) reliably estimating the resulting delays, as depicted in Figure 3.2; and 3) adopting advanced controls other than a fixed phase sequence (Hillier and Rothery, 1967).

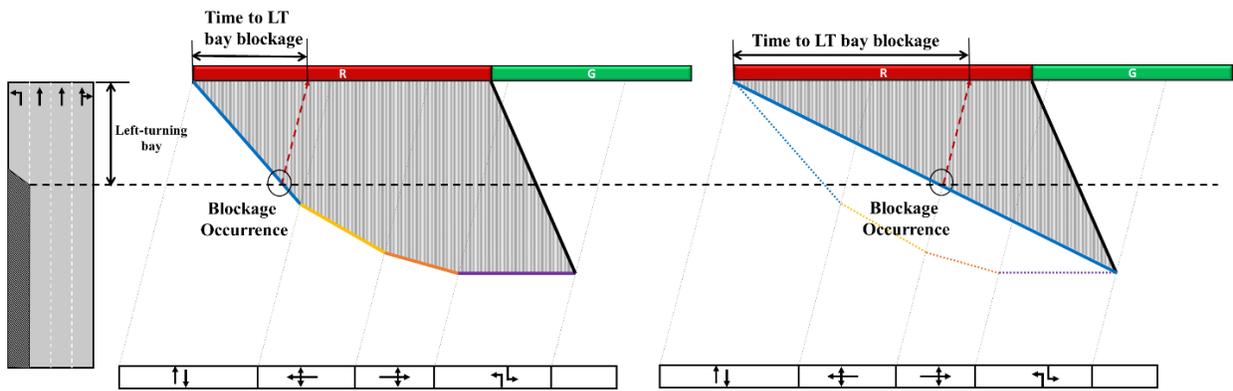


Figure 3.2 Graphical illustration of queue evolution under non-uniform (left) and uniform (right) upstream arrival rates and the associated impacts on delays and blockage time

In view of the above-mentioned critical issues as discussed in Chapter 2, this study has developed a dynamic coordinated signal optimization system tailored specifically for congested commuting arterials. This system is designed to tackle the complex challenges posed by highly

asymmetric flows during congested peak hours with either time-of-day or real-time controls. Central to its efficacy are the proposed system's multifaceted features, allowing users to compute the optimized offsets and phase sequences customized to tackle the unique traffic dynamics of the target commuting arterial. These features include:

- Integrating the objective of delay minimization for the congested direction with the control of maximizing progression for the opposing low-volume direction to optimal use of the arterial's available capacity. This integration methodology intends to address the unique traffic dynamics observed in commuting arterials, where traffic flows can vary significantly between different directions (as shown in Figure 3.1). Typically, during peak hours, one direction often experiences congestion, leading to delays and excessive queues, while the opposite direction may have lower traffic volume and excessive capacity. The integration of two different control objectives for two opposing directional arterial flows accommodating substantially different volumes can better ensure both efficiency and equity for the roadway users.
- Employing optimized offsets and phase sequences concurrently to minimize the total delay of the traffic flows in the high-volume commuting direction. This is one of the essential and imperative system features because delays, depending on the queue evolution process, are determined by the implemented offsets and phase sequences.
- Incorporating the minimization of potential blockages in design of the optimal signal plan, based on the estimated queue formation pace. Blockages can lead to left-turn vehicles spilling over to the through queues, causing additional queues and delays, as depicted in Figure 3.3. This feature is designed to minimize delays in response to the projected queue formation patterns and the potential blockage scenarios.

- Addressing time-varying arrival rates from diverse upstream traffic streams to ensure the effectiveness of the optimized phase sequences, especially in reducing the queue length or overflow duration, as shown in Figure 3.2.
- Accounting for the aggregate delay, resulting from different possible queue formation and evolution patterns contributed by all upstream arriving flows of different rates, in design of the dynamic phase sequence. This feature is imperative as it enables the control system to dynamically generate a new set of optimized phase sequences to minimize the total delays under various possible traffic demands and queue formation patterns.
- Accounting for the frequency and formation sequence of all possible mutual blockages between through and left-turn queues in design of the optimal signal timings and phase sequences. This system feature can also be exercised to estimate unused left-turn green time, reductions in through capacity, and the duration of mutual blockage between left-turn and through vehicles.
- Customizing the optimal control strategy to respond to time-varying traffic as well as congestion patterns in real time. This is proposed for the system to dynamically adjust the control plan during the transition of traffic conditions from undersaturated to saturated states over the arterial. For instance, when congestion levels and queue formation speeds are moderate, then the system will not re-optimize the signal plan but to execute dynamic adjustments of key control parameters such as green splits.
- Integrating with a short-term traffic flow prediction model to estimate the traffic queue state by capturing both arrival rates from multiple upstream streams and their

spatiotemporal evolution, enabling the execution of real-time proactive signal control.

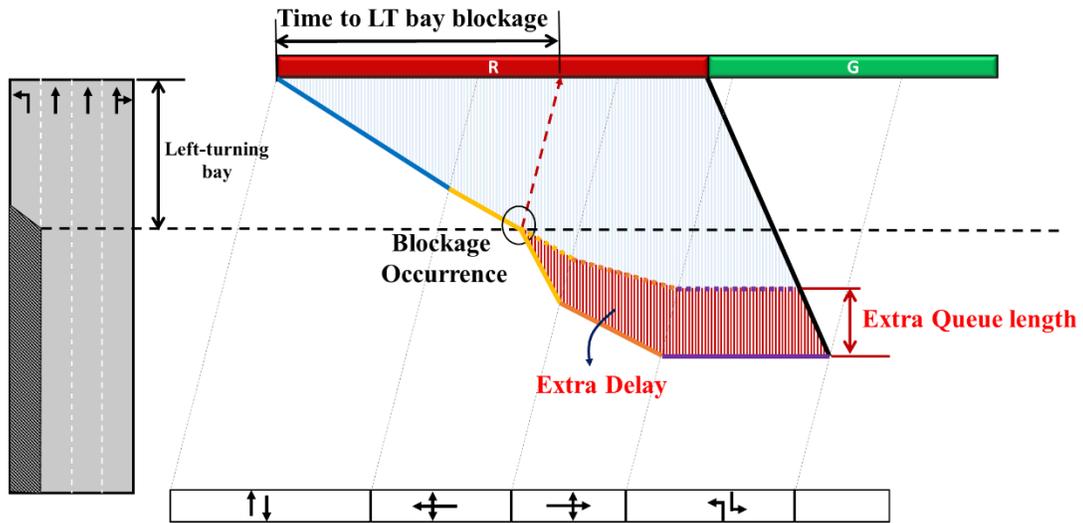


Figure 3.3 Graphical illustration of the adverse impacts caused by the mutual blockage

Figure 3.4 outlines the structure of such a system and its key modules. Module I serves as the base module for addressing the issue of asymmetric directional flows in commuting arterials under time-of-day pretimed control. Module II and Module III function to further enhance the base module to provide a timely response to the detected traffic congestion pattern.

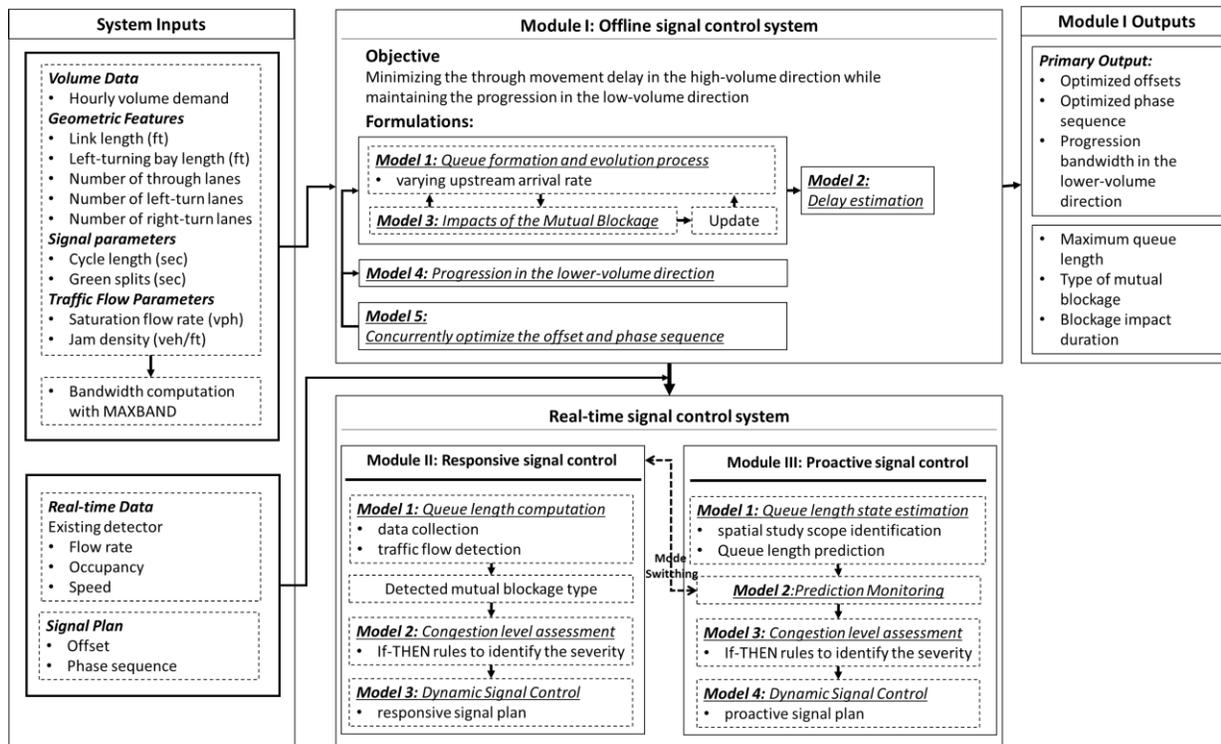


Figure 3.4 The developed system's overall structure

The remaining sections will first provide an overview of the primary functions embedded in each model, including the necessary data, operational logic, and resulting control measures. The main tasks involved in developing the proposed system are highlighted in the last section.

3.2 Key Modules

3.2.1 Module I: Offline Signal Control for Commuting Arterials with Highly Asymmetric Directional Flows

As shown in Figure 3.4, the control objective of this module is to minimize delay in the arterial's high-volume direction while maintaining progression in the arterial's low-volume direction. Hence, it is designed with the following functions:

- **A mechanism to estimate the queue length and duration attributed to each traffic stream.** This function is proposed to precisely assess the formation of queue length

originating from upstream traffic streams with varying arrival rates. Additionally, it serves as the base dataset for computing the resulting delay.

- **A flexible computation process for estimating the nature and duration of delays along the arterial.** This function enables the system to flexibly quantify the total delay under diverse queue evolution patterns and also produce the necessary input data to support the objective of optimizing signal control.
- **A mutual-blockage estimation function to compute the impacts of various types of lane blockage.** This function is designed to ensure that the control objective can reliably target the traffic dynamics of intermingled through and left-turn queues in most commuting arterials during peak hours. Failing to tackle such frequent mutual blockages between different traffic streams will cause the arterial to experience excessive queue lengths, increased delays, and worsening congestion levels.

The primary output of this module is the set of optimized time-of-day offsets and phase sequences for each intersection within the study area. Equipped with these functions, Module I provides a robust foundation for designing pre-timed signal control systems tailored to commuting arterials characterized by highly asymmetric directional traffic flows. Additionally, it also serves as the base module for its extensions, including Module II and Module III, designed for real-time signal control operations.

3.2.2 Module II: Responsive Real-Time Arterial Signal Control for Turning-bay Overflows and Mutual Queue Blockages

As illustrated in Figure 3.4, Module II builds upon Module I by incorporating real-time data and detection to enhance timely responsiveness to dynamic traffic conditions. This advanced feature

enables the system to effectively respond to varying levels of congestion, including different types of mutual blockages that commonly occur along commuting arterials. To fulfill this objective, Module II comprises a series of models, with each executing its uniquely designed function in a predefined sequence. The sequential operating structure of these models is illustrated in Figure 3.5, highlighting the coordinated process essential for responsive dynamic control.

- ***Model 1 – Queue Length Computation:*** This model estimates the time-varying evolution of both through and left-turn queue lengths with real-time processed detector data. By continuously capturing traffic conditions, it quantifies the temporal progression of queue formation and dissipation at each intersection. Based on these computed profiles, the module identifies the timing, duration, and types of mutual blockages that have occurred along the arterial. These blockage classifications, along with the corresponding queue assessment, form the output of Model 1 and serve as the critical input for subsequent identification of the congestion level in Model 2.
- ***Model 2 – Congestion Level Assessment:*** Building upon the outputs of Model 1, this model serves to evaluate the overall congestion level along the arterial by applying a structured set of predefined rule-based criteria. These *if-then* rules are designed to capture a comprehensive picture of traffic conditions, which are categorized into spatial, temporal, and spatiotemporal dimensions. Spatial rules are used to examine the simultaneous occurrence of blockages at multiple intersections to detect localized or arterial-wide congestion. Temporal rules are incorporated with the information of cycle-to-cycle queue evolution trends and feedback mechanisms to identify if the congestion patterns are escalating or persisting steadily over time. Spatiotemporal rules function to integrate the data from both dimensions to detect complex queue propagation dynamics across space

and time. Based on these assessments, the model will classify congestion severity into distinct levels—ranging from localized short-term disturbances to arterial-wide spillovers. The results from such classification serve as the critical input for Model 3, ensuring that the subsequent control strategy is best adopted to contend with the prevailing traffic conditions.

- ***Model 3 – Responsive Strategy Generator***: Informed by the congestion levels identified in Model 2 and the real-time queue length profiles from Model 1, this model is responsible for selecting and deploying appropriate responsive control strategies. Each pre-developed signal control strategy is designed to address a specific level of congestion patterns and blockage types observed along the arterial. Upon classification of the current traffic condition, the system will select the most effective signal plan and update signal timing parameters—such as offsets, phase sequences—for each intersection within the impact area. If the severity of congestion indicates the need for a more comprehensive recalibration, this model can trigger the core optimization module established in Module I to perform the arterial-wide re-optimization. This ensures that even under dynamically changing conditions, the system can maintain an effective and coordinated response to minimize the excessive queuing of vehicles and delays with optimal control of both the arriving and discharging traffic streams across the whole arterial.

The unique design features of Module II ensure that both localized and moderately congested conditions can be effectively managed at the intersection level, without activating an arterial-wide re-optimization of the control plan. With such real-time signal plan adjustments, focusing only on the targeted congested queue segment/intersection, the module can enhance the traffic efficiency while preserving computational and operational resources.

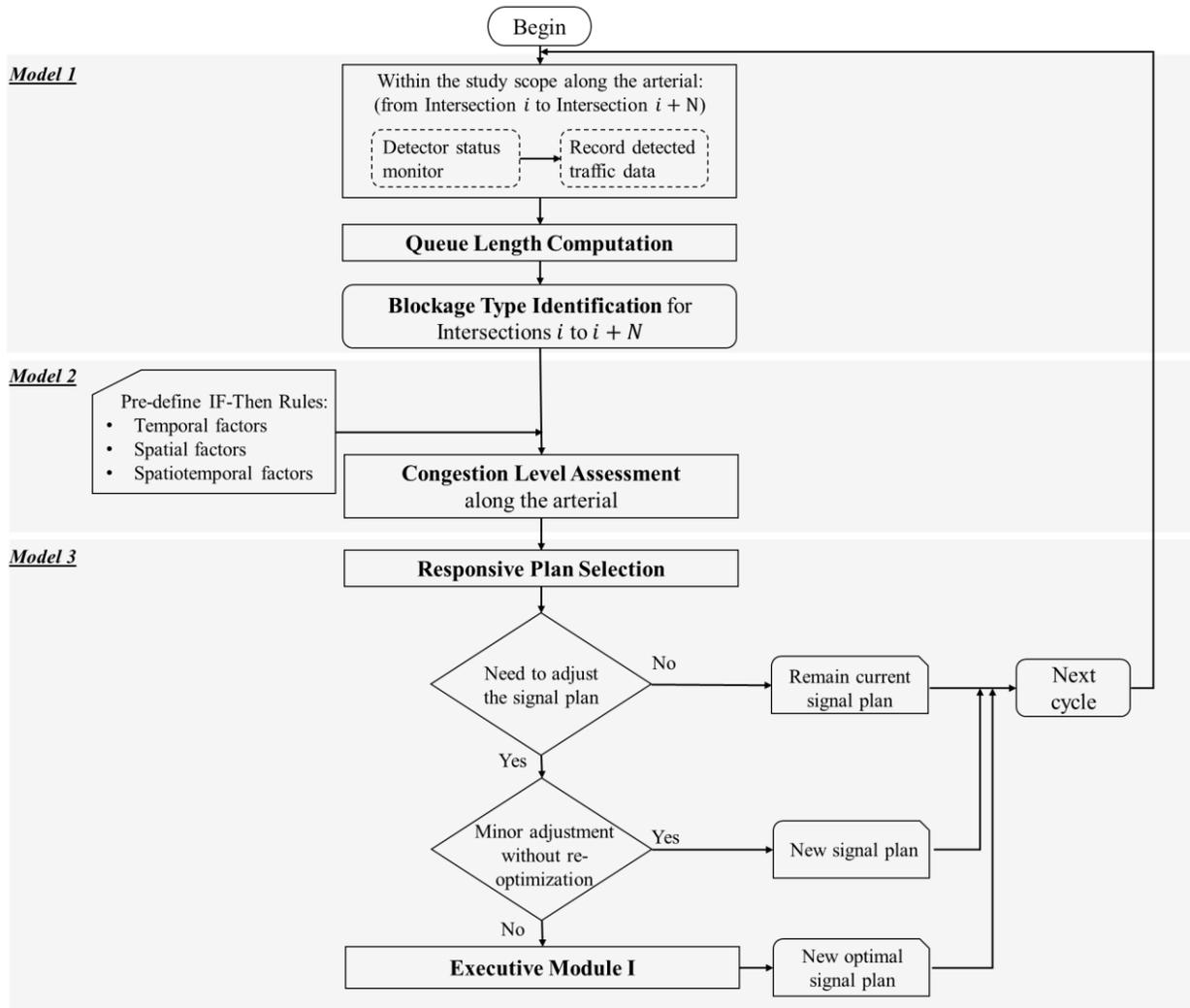


Figure 3.5 Flowchart of the developed responsive signal control for arterials with highly asymmetric directional traffic flows

3.2.3 Module III: Proactive Real-Time Arterial Signal Control for Commuting Arterials with Highly Asymmetric Directional Flows

As illustrated in Figure 3.4, Module III, building upon the same time-of-day control logic in Module I, is specially designed with a predictive function for proactive signal control. In contrast to the reactive nature of Module II, which adjusts signal timings in response to observed congestion, Module III is expected to predict and prevent the excessive queues and mutual blockages prior to

their formation and propagation. This proactive control operation is particularly essential for arterials with highly asymmetric directional flows, where queue spillbacks and multi-staged mutual blockages can develop rapidly and propagate to all intersections. While maintaining the modular structure consistent with that of Module II, Module III introduces an additional model to monitor the reliability of prediction results and to enable dynamic switching between proactive and responsive models based on real-time traffic conditions. As shown in Figure 3.4, Module III consists of the following four sequentially operating models:

- ***Model 1 – Queue Length State Prediction:*** This model forms the core of Module III, which allows the system to execute a forward-looking signal control through the prediction of queue state, focusing on projecting whether or not the queue spillbacks or mutual blockages will occur at each intersection, along with their timing and duration if present. These predictions provide the necessary lead time for proactive adjustments of the signal control plan. The model utilizes a spatiotemporal deep learning structure that captures both temporal trends from recent traffic patterns—and spatial interactions—with real-time data from multiple upstream intersections. The model also accounts for the dynamics of travel times influenced by link length, signal timings, existing queues, and potential blockages. To enhance the module’s responsiveness, the prediction model can output multiple indicators to characterize the predicted traffic state, including turning ratios and queue lengths. Additionally, the spatial scope of the prediction is optimized by selecting the most relevant upstream intersections to balance accuracy with computational efficiency.
- ***Model 2 – Prediction Monitoring and Mode Switching:*** This model evaluates the accuracy of predicted queue states by comparing them against real-time traffic data. If the discrepancy between predicted and observed queue conditions exceeds a predefined

threshold, the system will deactivate the proactive strategy and revert to the responsive signal control operations defined in Module II. This monitoring function ensures the reliability of proactive decision-making and prevents potential misjudgments due to inaccurate predictions.

- ***Model 3 – Congestion Level Assessment:*** This model mirrors the structure and logic of its counterpart in Module II, employing a comprehensive set of pre-defined *if-then* rules to classify arterial congestion levels based on spatial, temporal, and spatiotemporal criteria. The key distinction lies in its input, that is, instead of relying on the observed queue length states, this model will evaluate the predicted queue conditions generated by Model 1. With such forward-looking traffic state information, the module can have an early identification of potential congestion patterns and queue distribution. Such a function can allow the deployed signal control system to have sufficient time in advance to respond preemptively to the evolving congestion dynamics, with available control strategies embedded in Model 3.
- ***Model 4 – Proactive Strategy Generator:*** This model functions to generate predictive signal control strategies based on anticipated congestion patterns. Using the predicted queue length states from Model 1 and the congestion level assessed in Model 2, it is responsible for identifying the timing and the signals for advanced adjustments to prevent excessive queue formation and spillbacks. While structurally similar to the responsive strategy generator in Module II, the key distinction lies in its predictive capability that enables the signal system to execute preemptive interventions before the presence of any mutual lane blockage. Upon determining the best response plan, the system will recalculate the signal settings at those identified critical intersections. If re-optimization is required,

the base signal plan established by Module I will be re-engaged to ensure globally coordinated controls across the arterial.

This integrated design equips Module III with the capability to predict and mitigate congestion prior to its formation, while simultaneously preserving a fallback mechanism to revert to responsive control when the reliability of predictions is insufficient. Together, all such features provide a flexible and intelligent control framework that is well-suited to address dynamic and complex traffic conditions over commuting arterials.

3.3 Closure

In summary, this chapter has presented a comprehensive multi-level signal control system, tailored for urban commuting arterials characterized by significant directional traffic imbalances and potential mutual blockages between through and left-turn flows. The key research concerns, essential system functionalities, core inputs, and expected outputs for such a system to operate effectively and efficiently under both pre-timed and real-time control scenarios constitute the core of this chapter. The structure and interrelations of the three core models—along with their constraints and operational logic—have been thoroughly described. The next chapter will first present the core design logic and associated key formulations as well as their interrelations to constitute the time-of-day pretimed signal control system.

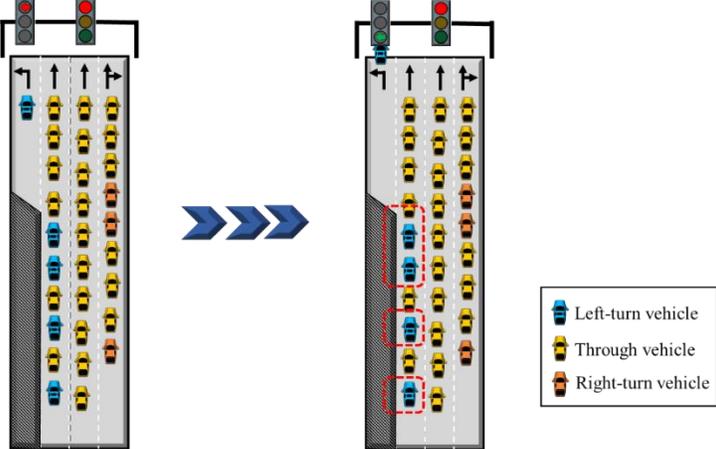
Chapter 4 A Time-of-day Signal Control Module for Commuting Arterials with Highly Asymmetric Directional Flows

4.1 Introduction

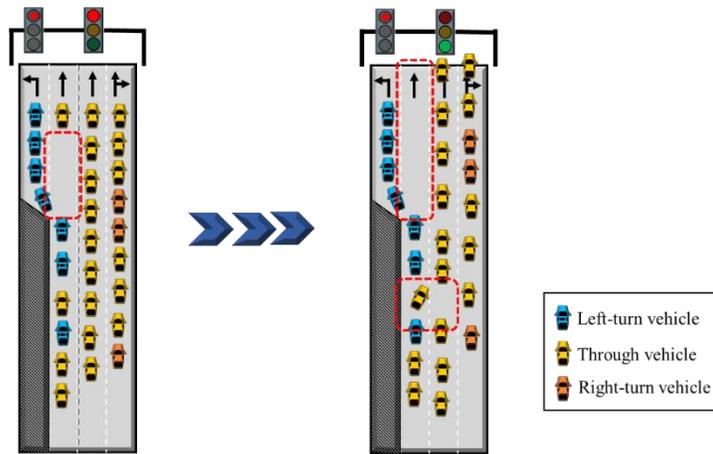
The design of signal control plans for major arterials in commuting corridors poses a significant challenge to traffic professionals due to their highly congested yet asymmetric directional flows, particularly during peak periods. In the high-volume commuting direction, intersections along such arterials often encounter a myriad of issues, including excessive delays, prolonged queues, low progression efficiency, and mutual blockages between through and left-turn queues. In contrast, in the low-volume directions, the available capacity at most intersections is often underutilized, leading to undesirably high traffic speeds and safety concerns. Developing an effective coordinated signal control system for such arterials is inherently difficult due to the presence of both undersaturated and near-saturated traffic characteristics in their two opposing directional flows.

Traditionally, most studies on coordinated signal control tend to focus either on undersaturated or oversaturated traffic conditions, due to their distinct traffic patterns and discrepancies in the adopted control objective. However, the unique characteristics of commuting arterials necessitate an integrated approach to address the challenges posed by unbalanced directional flows (Peng and Wang, 2022). In addition, what has often posed challenges to traffic management for commuting arterials is the frequent occurrence of mutual blockages between through and turning flows. For instance, when the excessive through queues block left-turn vehicles from entering the turning bay in time, the allocated green time for the left-turn phase, as illustrated in Figure 4.1(a), may not be fully used. Similarly, when the left-turn queues spill over from the turning bay to a neighboring through lane, it will naturally reduce the capacity of the through lanes, as depicted in Figure 4.1(b). Additionally, during the through-green phase, the lane-

changing maneuvers by those impeded through vehicles may contribute to slowdowns of traffic flows in adjacent lanes. Failing to address such mutual blockages can result in a significant underestimate of the total traffic delays and compromise the effectiveness of any optimized signal plan for the arterial.



(a)



(b)

Figure 4.1 The adverse impacts of mutual blockages (a) wasted green duration; (b) decreased capacity

To address these challenges, this chapter has presented a time-of-day control module specifically designed to tackle the critical queue-formation issues prevalent in commuting arterials.

Key operational features customized for such arterials include:

- Integrating the following two objectives to optimally use the arterial's available capacity and minimize the likelihood of incurring turning-bay overflows: 1) delay minimization for the congested direction; and 2) maximizing progression for the opposing low-volume direction;
- Employing the optimized offsets and phase sequences as the two key control variables to minimize the total delay of the through traffic in the high-volume commuting direction;
- Estimating the through and left-turn queue formation pace and incorporating the information in design of the optimal signal plan;
- Accounting for the total delay, resulting from different queue formation and evolution patterns, contributed by different arriving rates and sequences of all upstream coming flows.

Notably, the last two features are introduced with the recognition that the total delay in an intersection's congested approach varies not only on the maximum queue per cycle, but also on the queue patterns and their arriving sequence. For instance, as depicted in Figure 4.2, the formation process of two queues of equal length can result in varying total delays under different phase sequences: 1) the resulting signal delays, indicated by the shaded area, may vary markedly when the traffic streams of varying rates arrive in different sequences; 2) the formation pace of through queues in most scenarios may accelerate significantly after surpassing the turning bay's length, consequently increase the resulting total delay-justifying one of the developed module's features that is to reliably estimate the potential onset time for different types of bay blockages at each major intersections along the commuting arterial.

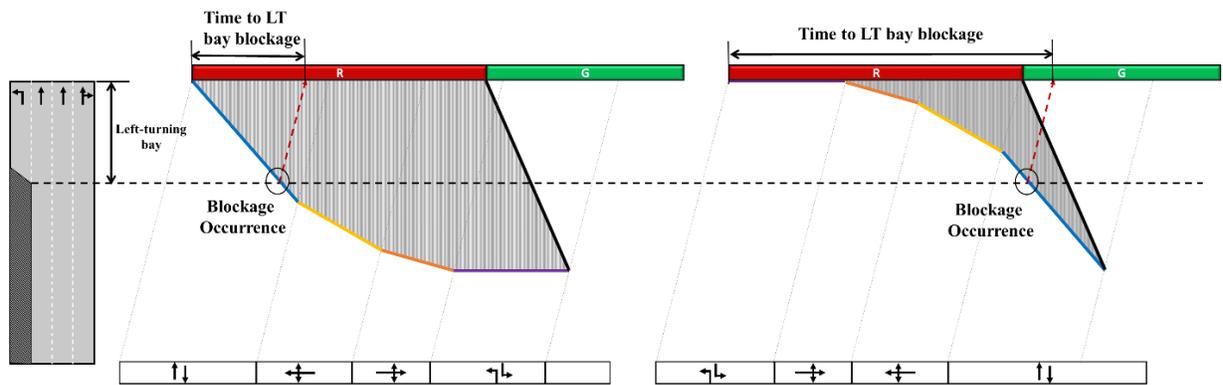


Figure 4.2 Graphical illustration of the impact of queue formation patterns on a lane-blockage's starting time and the resulting delay

4.2 Time-of-Day Arterial Signal Control Module

Figure 4.3 illustrates the structure of the developed module, highlighting the interrelationships between its five primary models along with the associated formulations. The remaining chapter is organized as follows: Section 4.3 details the formulations of the models, including all constraints

and objective functions; Section 4.4 presents case studies to evaluate module performance. Key variables and parameters used in the module are listed in Table 4.1.

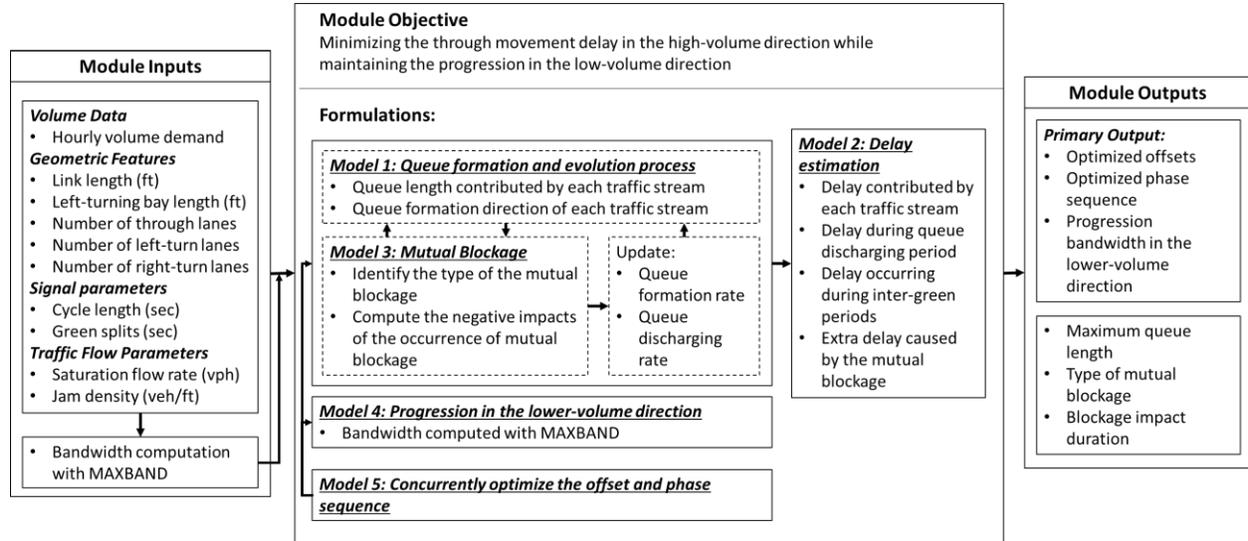


Figure 4.3 Structure of the developed time-of-day arterial signal control module

Table 4.1 List of key notations

Indices, Sets, and Operators	
n	Turning movement. $n \in \{T, L\}$ (T: arterial through; L: arterial left-turn)
j	Queue Group indicator. $j \in \{1,2,3,4\}$
p	Cycle indicator. $p \in \{1,2\}$
Parameters	
π_i	Travel time to traverse through the link between upstream intersection $i + 1$ and downstream intersection i (sec)
C	Common cycle length (sec)
v	Speed (ft/sec)
k	Jam density (veh/ft)
h	Saturation headway (sec)
L_i	Link length between intersections i and $i + 1$ (ft)
L_i^*	Left-turn bay length at the high-volume direction between intersections i and $i + 1$ (ft)
$g_i^n (\bar{g}_i^n)$	The effective green duration for movement n departing (arriving) intersection i (sec)
r_i^n	The red-light duration for movement n departing intersection i (sec)
$q_{i,j}^n$	Arrival rate of the movement n 's queue Group- j at intersection i (veh/sec)
I	Inter-green duration (sec)
$d_{i,j}$	Demand from traffic stream contributing to Queue Group- j at intersection i (vph)
ρ_i^n	Turning ratio of movement n at intersection i
N_i	Number of through lanes in the high-volume direction at intersection i
\bar{B}'	Pre-calculated MAXBAND bandwidth for the low-traffic-volume direction (sec)
Variables	

$x_i(\bar{x}_i)$	Binary variable indicating the phase sequence, which equals 1 if the green phase for the main street's through (left-turn) movement in the high-volume direction is ahead of the main street's left-turn (through) movement in the low-volume direction
\bar{B}	The bandwidth for the low-volume direction (sec)
$\bar{\omega}_i$	The time difference between a through phase in the low-volume direction and the starting time of the progression band at intersection i (sec) in the same cycle
$K_i^n(\bar{K}_i^n)$	Integer variables to synchronize movement n 's green phase in the high-volume (low-volume) direction at intersection i with its corresponding signal cycle at the upstream intersection
θ_i	The offset of the intersection i (sec)
$\theta_i^n(\bar{\theta}_i^n)$	The starting time of the green phase for movement n in the high-volume (low-volume) direction at intersection i (sec)
ε_i^n	The starting time of the n movement's queue formation at intersection i (sec)
$u_{i,j,p}$	The duration for Through Queue Group- j - p to contribute to the queue formation at intersection i (sec)
$Q_{i,j,p}$	Number of vehicles joining the through queue from Queue Group- j - p at intersection i (veh)
Q_i^n	Maximum queue length of the n movement at intersection i (veh)
$S'_{i,j,p}$	The starting time of the green phase for the associated Through Queue Group- j - p at intersection i (sec)
$E'_{i,j,p}$	The ending time of the green phase for the associated Through Queue Group- j - p at intersection i (sec)
$S_{i,j,p}$	The starting time of the duration for the Through Queue Group- j - p contributing to the queue formation at intersection i (sec)
$E_{i,j,p}$	The ending time of the duration for the Through Queue Group- j - p contributing to the queue formation at intersection i (sec)
$A_{i,j,p}$	Number of cumulative vehicles after the Through Queue Group- j - p joining the queue at intersection i (veh)
$Z_{i,j,p}$	Cumulative duration after the Through Queue Group- j - p joining the queue at intersection i (sec)
F'_i	Number of blocked left-turn vehicles constituting the residual queue at intersection i
F_i	Total number of blocked left-turn vehicles at intersection i
$U_{i,j,p}$	Height of the delay area contributed by the Through Queue Group- j - p at intersection i (sec)
\hat{D}_i	Delay area contributed by queue groups at intersection i . (veh·sec)
D_i^*	Delay area generated by queue groups during inter-green duration at intersection i (veh·sec)
D'_i	Delay area caused by the blocked left-turn vehicles that stay as residual queues at intersection i (veh·sec)
D_i	Total delay at intersection i (veh·sec)
$\delta_{i,j,p}^n$	Binary variables indicating if the n movement in queue Group- j - p at intersection i initiates the bay blockage
$\alpha_{i,j,p}$	Binary variables indicating if Queue Group- j - p at intersection i joins the queue after the onset of the left-turn bay blockage
$\beta_{i,j,p}$	Binary variables indicating if the cumulative queue length after the Through Queue Group- j - p at intersection i should be included in the delay area computation
φ_i	Binary variables indicating if a through-lane's capacity drop occurs at intersection i
γ_i	Binary variables indicating if any blocked left-turn vehicles contribute to the residual queues at intersection i
Y_i^n	The time point when the last queuing vehicle of movement n reached the stop line at intersection i (sec)
\tilde{t}_i^n	The starting time of the blockage at intersection i due to the n movement's queue, without considering the interaction between initial and subsequent blockages.

t_i^n	The starting time of the blockage at intersection i due to the n movement's queue, considering the interaction between initial and subsequent blockages.
W_i^n	The end time of the blockage impact at intersection i due to the n movement's queue, considering the interaction between initial and subsequent blockages.
H_i^n	The time duration for the queue from movement n to exceed the turning bay's length at intersection i

4.3 Model Formulations

This section will present the following sets of core equations for each of the models in the developed module:

- Queue formation and evolution process
- Delay estimation based on the time-varying queue length
- Traffic dynamics under the state of mutual queue blockage
- The progression design for the low-traffic-volume direction
- Relations between the optimal phase sequences and the starting time of each phase

4.3.1 Queue Formation and Evolution Process

Figure 4.2 illustrates how different queue evolution patterns lead to different types of queuing delays. Further insights into the through queue formation process and its interdependent relations with upstream traffic waves are provided in Figure 4.4. To facilitate the presentation and analysis, the following provides a detailed breakdown of the different vehicle groups comprising the queue structure, where:

$Q_{i,j,p}$ - denotes the through queues at intersection i , formed by traffic stream j from the upstream intersection's cycle p . Where $j = 1,2,3,4$ and $p = 1,2$.

- When $j = 1$, it indicates the through flows from the main street;
- When $j = 2$, it indicates the left-turn flows from the main street;

- When $j = 3$, it indicates the left-turning-in flows from the side street; and
- When $j = 4$, it indicates the right-turning-in flows from the side street.

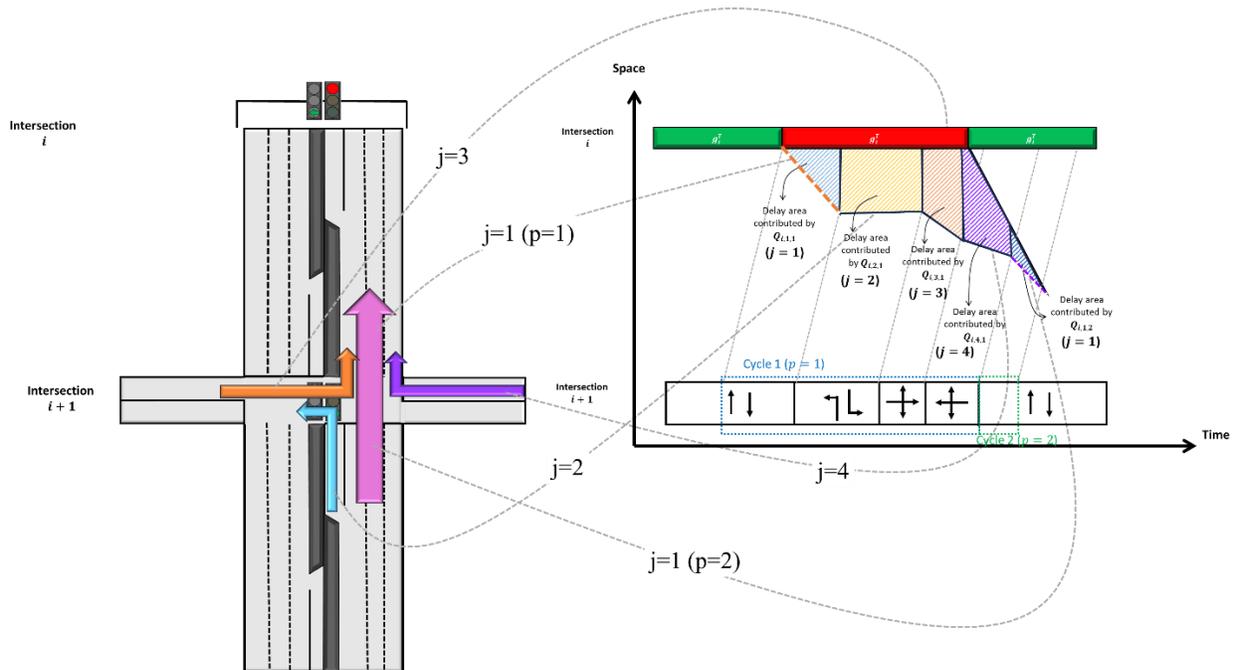


Figure 4.4 Graphical illustration of through queue formation with different arrival rates

Note that these definitions are not indicative of their order of sequence within the queue formation process but rather reflect different movements of the arriving traffic streams. Depending on the arterial's signal coordination design, traffic streams, contributing to an intersection's through queue, may come from its upstream intersection during different cycles. Hence, queue groups contributed by the same traffic streams from an upstream intersection but during different cycles are numbered accordingly.

For instance, Figure 4.4 illustrates an example queuing pattern that consists of five queue groups from two cycles. The queue groups indicated by the orange dashed line and the purple dashed line both originate from the upstream intersection's through flows, but from different cycles.

As such, the through queue group represented by the orange dashed line is denoted as $Q_{i,1,1}$ and the group represented by the purple dashed line is denoted as $Q_{i,1,2}$.

Therefore, the developed formulations, aimed at estimating all types of incurred delays, must comprehensively capture the entire queue formation process, which includes vehicles contributing to different groups of the total queues from various cycles. Furthermore, a similar logic can be applied to the formation of left-turn queues, as illustrated in Figure 4.5.

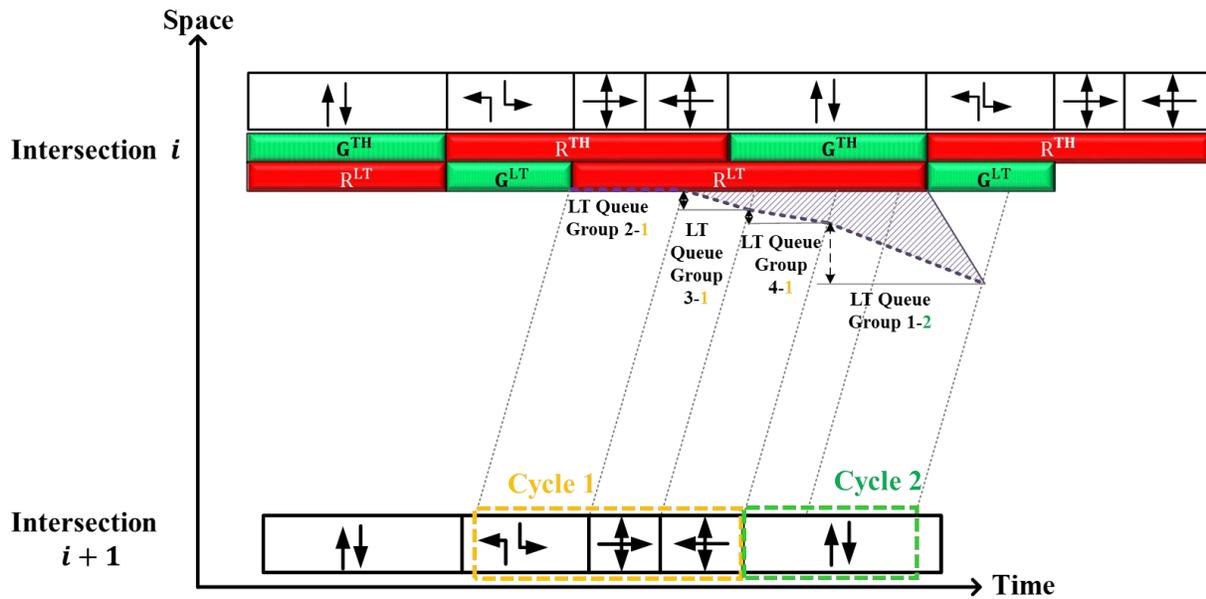


Figure 4.5 Graphical illustration of left-turn queue formation under different arrival rates

To achieve this, the developed module must be capable of performing the following tasks:

- Compute the length of each queue group within the total queue
- Compute the queue duration from formation to dissipation for each queue group

Note that the same logic is also applicable for computing the properties of left-turn queues.

Therefore, the following presentation will focus only on the through queues.

4.3.1.1 Computing the length of each vehicle group within the total queue

To explicitly formulate the queue evolution process with varying arrival rates from different traffic streams, it is necessary to calculate the length contributed by each queue group, denoted as $Q_{i,j,p}$. This computation involves multiplying the duration, $u_{i,j,p}$, which represents the time period contributing to the queue formation, by the associated arrival rates. The superscription p indicates whether the queue group is formed from the same or the subsequent cycle of the upstream intersection. Mathematically, this relationship is expressed as follows:

$$Q_{i,j,p} = u_{i,j,p} \cdot q_{i,j}^T \quad \forall j = 1,2,3,4; \forall p = 1,2 \quad (4-1)$$

The vehicle arrival rate for both the through queue and left-turn queue, denoted as $q_{i,j}^n$ in Equation 4-1, can be derived from the following equations:

$$q_{i,j}^T = \frac{\frac{d_{i+1,j}}{g_{i+1}^n} \times \frac{C}{3600} \times (\rho_i^T + \rho_i^R)}{N_i} \quad \forall j = 1,2,3,4 \quad (4-2a)$$

$$q_{i,j}^L = \frac{d_{i+1}^j}{g_{i+1}^n} \times \frac{C}{3600} \times \rho_i^L \quad \forall j = 1,2,3,4 \quad (4-2b)$$

Where d_{i+1}^j , g_{i+1}^n , ρ_i^T , ρ_i^R , and ρ_i^L are defined in Table 4-1. Note that the associated upstream green phase and demand are determined based on the queue group being analyzed. For instance, to compute the arrival rate for Through Queue Group- ($Q_{i,1,1}$), the associated upstream green phase would be g_{i+1}^T and the demand would be $d_{i+1,1}$. Note that the duration for an upstream traffic stream to contribute to the downstream queue formation, $u_{i,j,p}$, may be shorter than its allocated green phase duration since a part of the green time may be in a progression band and does not incur queues.

Given the estimated size for each group of the queue, the maximum queue length at intersection i , denoted as Q_i^T , can be determined by summing the lengths of all groups of the queue and expressed as follows:

$$Q_i^T = \sum_{p=1}^2 \sum_{j=1}^4 Q_{i,j,p} \quad (4-3)$$

To streamline the discussion, the subsequent analysis will concentrate on the formulations for Through Queue Group- ($Q_{i,1,1}$). The same process and logic can be applied to other groups of both the through and left-turn queues.

4.3.1.2 Queue formation duration of the Through Queue Group- ($Q_{i,1,1}$)

Note that the key variable in Equation 4-1, $u_{i,j,p}$, varies with the arriving sequence of all upstream traffic streams. As illustrated in Figure 4.6, vehicles traveling within the progression band can traverse the downstream intersection without stopping, delineated by the shaded area labeled “TH progression”, thereby resulting in a shorter formation duration for Through Queue Group- ($Q_{i,1,1}$), compared to the full available green phase. Hence, to estimate the resulting delay, one shall first identify each group of the queue’s onset and ending times (e.g., points A and B for Through Queue Group - $Q_{i,1,1}$ in Figure 4.6) with the following steps:

- **Step 1:** Define the range of queue formation by identifying the initiating and ending instances of the queue evolution, denoted by ε_i^T and Y_i^T (i.e., points A and C in Figure 4.6, respectively)
- **Step 2:** Compute the starting and ending of the green phase corresponding to a group of the queue, as represented by points D and B in Figure 4.6.

- **Step 3:** Compute the duration of Through Queue Group- ($Q_{i,1,1}$), contributing to the total through queue length, as represented by the segment between points A and B in Figure 4.6, by accounting for the relationships between initiating/ending instances of queue evolution and the start/end of the green phase for Through Queue Group- ($Q_{i,1,1}$).

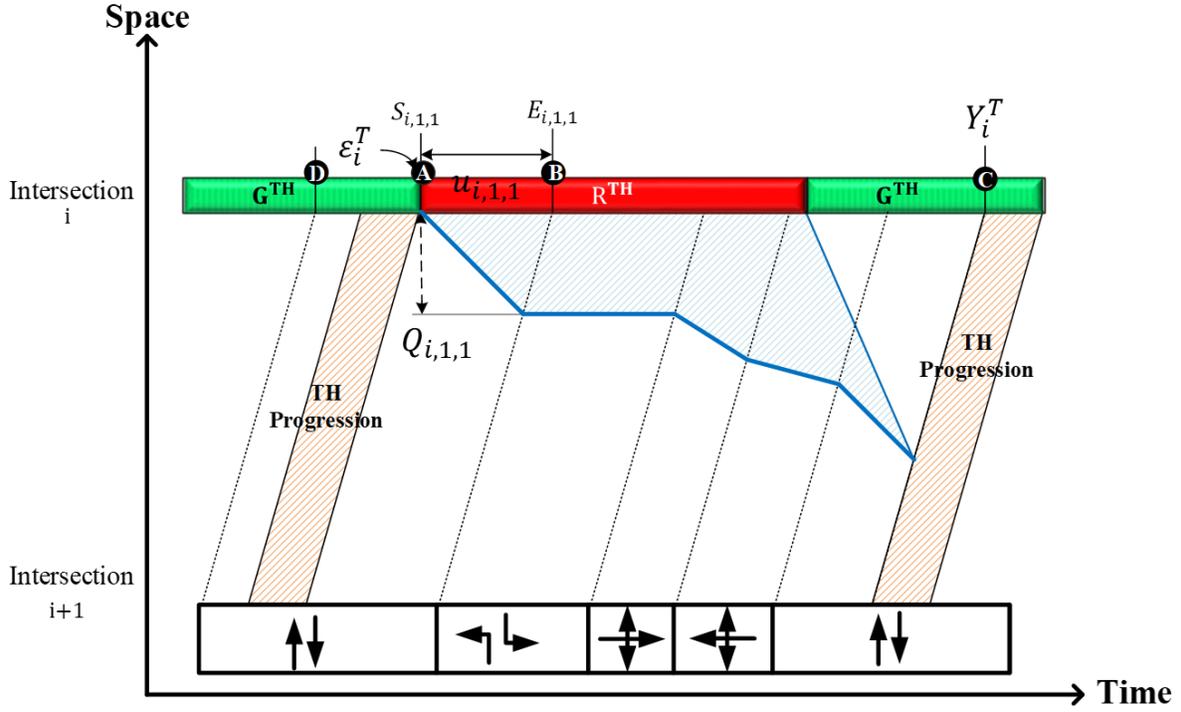


Figure 4.6 Graphical illustration of the computation for the Through Queue Group- ($Q_{i,1,1}$)

The formulations associated with each step are enumerated as follows:

Step 1:

$$\varepsilon_i^T = \theta_i^T + (g_i^T + I) + K_i^T \cdot C \quad (4-4)$$

$$\theta_{i+1} + \pi_i \leq \varepsilon_i^T \leq \theta_{i+1} + \pi_i + C \quad (4-5)$$

$$Y_i^T = \varepsilon_i^T + r_i^T + Q_i^T \cdot h + \frac{Q_i^T}{k \cdot v} \quad (4-6)$$

Equation 4-4 illustrates that ε_i^T , serving as a reference point, is specified to define when the through queue accumulation would begin at intersection i , which in this case, is at the end of the green phase. To ensure the synchronization of ε_i^T with the passage of vehicles from upstream intersection $i + 1$, Equation 4-5 establishes its boundary conditions within the temporal scope encompassing the arrival of the first and last vehicles from the upstream intersection $i + 1$. Furthermore, Equation 4-6 introduces Y_i^T as the temporal marker denoting the arrival of the final through-queueing vehicle at intersection i .

Step 2:

$$S'_{i,1,p} = \theta_{i+1}^T + \pi_i + (p - 1) \cdot C \quad \forall p = 1,2 \quad (4-7)$$

$$0 \leq S'_{i,1,p} \quad \forall p = 1,2 \quad (4-8)$$

$$E'_{i,1,p} = (\theta_{i+1}^T + g_{i+1}^T) + \pi_i + (p - 1) \cdot C \quad \forall p = 1,2 \quad (4-9)$$

$$0 \leq E'_{i,1,p} \quad \forall p = 1,2 \quad (4-10)$$

The right-hand side of Equation 4-7 consists of the onset of the through phase at intersection $i + 1$, the travel time from upstream intersection $i + 1$ to downstream intersection i , and the last term, $(p - 1) \cdot C$, is to ensure that the applicability to any upstream queue cycle. Similarly, the end of the green phase, denoted as $E'_{i,1,p}$, follows a similar rationale, augmented by an additional term representing the duration of the through phase, as expressed in Equation 4-9.

Step 3:

$$S_{i,1,p} = \begin{cases} \varepsilon_i^T & S'_{i,1,p} < \varepsilon_i^T \\ S'_{i,1,p} & \varepsilon_i^T \leq S'_{i,1,p} \leq Y_i^T \\ Y_i^T & Y_i^T < S'_{i,1,p} \end{cases} \quad \forall p = 1,2 \quad (4-11)$$

$$E_{i,1,p} = \begin{cases} \varepsilon_i^T & E'_{i,1,p} < \varepsilon_i^T \\ E'_{i,1,p}, & \varepsilon_i^T \leq E'_{i,1,p} \leq Y_i^T \\ Y_i^T & Y_i^T < E'_{i,1,p}, \end{cases} \quad \forall p = 1,2 \quad (4-12)$$

$$u_{i,1,p} = E_{i,1,p} - S_{i,1,p} \quad \forall p = 1,2 \quad (4-13)$$

$$0 \leq u_{i,1,p} \leq g_{i+1}^T \quad \forall p = 1,2 \quad (4-14)$$

To compute the duration of $Q_{i,1,1}$, given the start/end of the upstream through green phase, one shall first identify their relationships with the initiating/ending time of the queue formation (i.e., ε_i^T and Y_i^T), which have four possible scenarios, as shown in Figure 4-7. If either the start or end of the green phase does not fall within the temporal bounds of ε_i^T and Y_i^T , the duration contributing to the queue formation would be shorter than the green phase itself (refer to Figure 4-7(a) and (b)). In the scenario where both the start and end of the green phase lie beyond this range - indicating no contribution to the queue accumulation during this phase - the duration is considered zero (see Figure 4-7(c)). Hence, the start/end of the queue formation duration should be adjusted accordingly, as shown in Equation 4-11 and Equation 4-12. The resulting duration of $Q_{i,1,1}$ can then be computed with its start and end times, as shown in Equation 4-13.

Note that such a general formulation process can be applied to all four possible scenarios depicted in Figure 4.7, under any sequence of traffic streams contributing to this group of the queue.

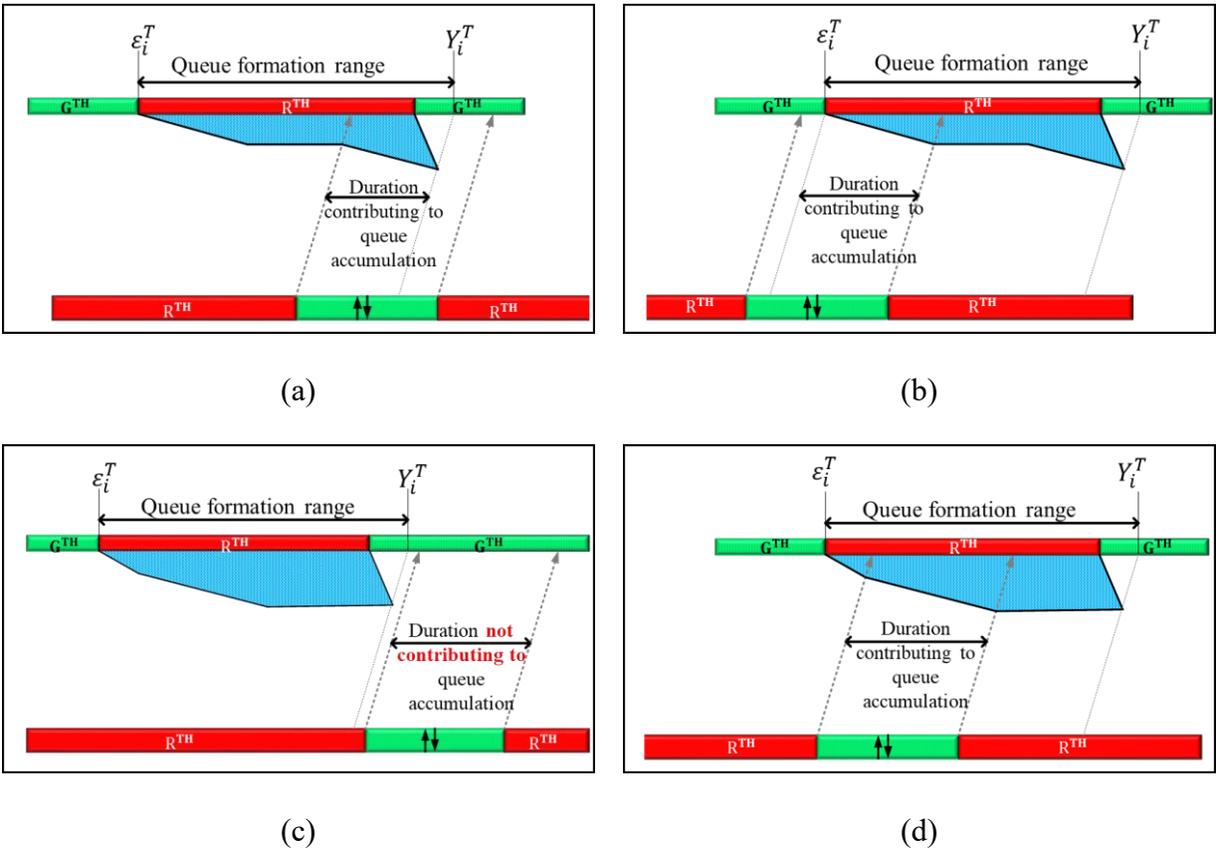


Figure 4.7 Relationships between the start/end of the green phase and the queue formation duration

4.3.2 Delay Estimation Based on the Time-Varying Queue Length

Primary computation works for the delay estimation include:

- dividing the total delay areas into properly defined sub-areas
- formulating the key variables: cumulative queue length and cumulative queue duration
- quantifying each sub-delay area

4.3.2.1 Dividing the total delay areas into properly defined sub-areas

With the explicit formulations for time-varying queue length, one can compute the delay based on the area between the queue accumulation and dissipation waves, as shown in Figure 4.8. This area can be further delineated into three distinct sub-areas:

- Area A: Represents delays attributed to the initial formation of the queue, visualized by the red triangular region in Figure 4.8.
- Area B: Denotes delays contributed by additional groups of the queues that progressively join the queue length, illustrated by the blue area in Figure 4.8.
- Area C: Signifies delays occurring during inter-green periods, as indicated with the grey rectangular regions in Figure 4.8.

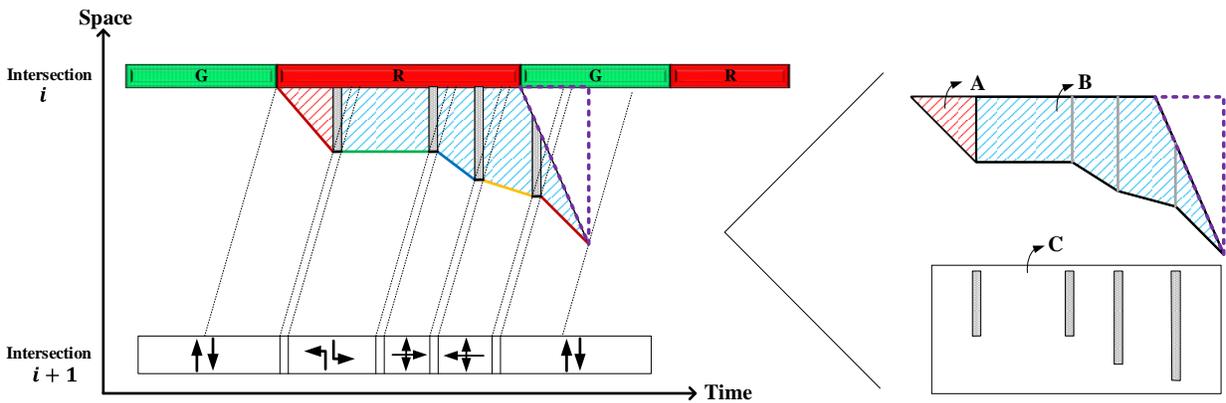


Figure 4.8 Graphical illustration of delay decomposition

4.3.2.2 Formulating the key variables: cumulative queue length and duration

To quantify the delay area, it is imperative to compute two key variables: the time-varying cumulative queue length, denoted as $A_{i,j,p}$, and the cumulative queue duration, represented by $Z_{i,j,p}$ as illustrated in Figure 4.9). These variables are essential for computing the total time-varying delay area contributed by various types of queues.

Cumulative queue length

$$A_{i,1,p} = A_{i,4,p-1} + Q_{i,1,p} + (1 - x_{i+1}) \cdot Q_{i,2,p} \quad \forall p = 1,2 \quad (4-15)$$

$$A_{i,2,p} = A_{i,4,p-1} + Q_{i,2,p} + x_{i+1} \cdot Q_{i,1,p} \quad \forall p = 1,2 \quad (4-16)$$

$$A_{i,3,p} = A_{i,4,p-1} + Q_{i,1,p} + Q_{i,2,p} + Q_{i,3,p} \quad \forall p = 1,2 \quad (4-17)$$

$$A_{i,4,p} = A_{i,4,p-1} + Q_{i,1,p} + Q_{i,2,p} + Q_{i,3,p} + Q_{i,4,p} \quad \forall p = 1,2 \quad (4-18)$$

Equations 4-15 to 4-18 outline the computation of cumulative queue lengths for each of the four queue groups. Considering Queue Group- $(Q_{i,3,p})$ for illustration, Equation 4-17 demonstrates that the cumulative queue length after Through Queue Group- $(Q_{i,3,p})$ merges with the existing queue, including the cumulative queue length of the preceding cycle $(A_{i,4,p-1})$, and the current cycle p 's Queue Group-1,2, and 3 (i.e., $Q_{i,1,p}$, $Q_{i,2,p}$, $Q_{i,3,p}$).

When $p = 1$, $A_{i,4,p-1} = 0$. It should be noted that in this study, vehicle demands from both directions of the side street are presumed to be at the same level. Consequently, the phase sequence between side street directions is predetermined with Queue Group-3 $(Q_{i,3,p})$ always preceding Queue Group-4 $(Q_{i,4,p})$. However, for Queue Group-1 $(Q_{i,1,p})$ and Queue Group-2 $(Q_{i,2,p})$, the sequence is adaptable. Hence, the cumulative queue length after Queue Group-1 $(Q_{i,1,p})$ joins the existing queue may incorporate Queue Group-2 $(Q_{i,2,p})$, contingent on the phase sequence at the upstream intersection, indicated by x_{i+1} . Similar logic applies to $A_{i,2,p}$.

Cumulative queue duration

$$Z_{i,1,p} = Z_{i,4,p-1} + u_{i,1,p} + (1 - x_{i+1}) \cdot u_{i,2,p} \quad \forall p = 1,2 \quad (4-19)$$

$$Z_{i,2,p} = Z_{i,4,p-1} + u_{i,2,p} + x_{i+1} \cdot u_{i,1,p} \quad \forall p = 1,2 \quad (4-20)$$

$$Z_{i,3,p} = Z_{i,4,p-1} + u_{i,1,p} + u_{i,2,p} + u_{i,3,p} \quad \forall p = 1,2 \quad (4-21)$$

$$Z_{i,4,p} = Z_{i,4,p-1} + u_{i,1,p} + u_{i,2,p} + u_{i,3,p} + u_{i,4,p} \quad \forall p = 1,2 \quad (4-22)$$

The cumulative duration following the joining of Through Queue Group- j - p ($Q_{i,j,p}$) at intersection i , denoted as $Z_{i,j,p}$, can be formulated using Equations 4-19 to 4-22, following the same principles applied in the calculation of the cumulative queue length. Note that when $p = 1$, $Z_{i,4,p-1} = 0$.

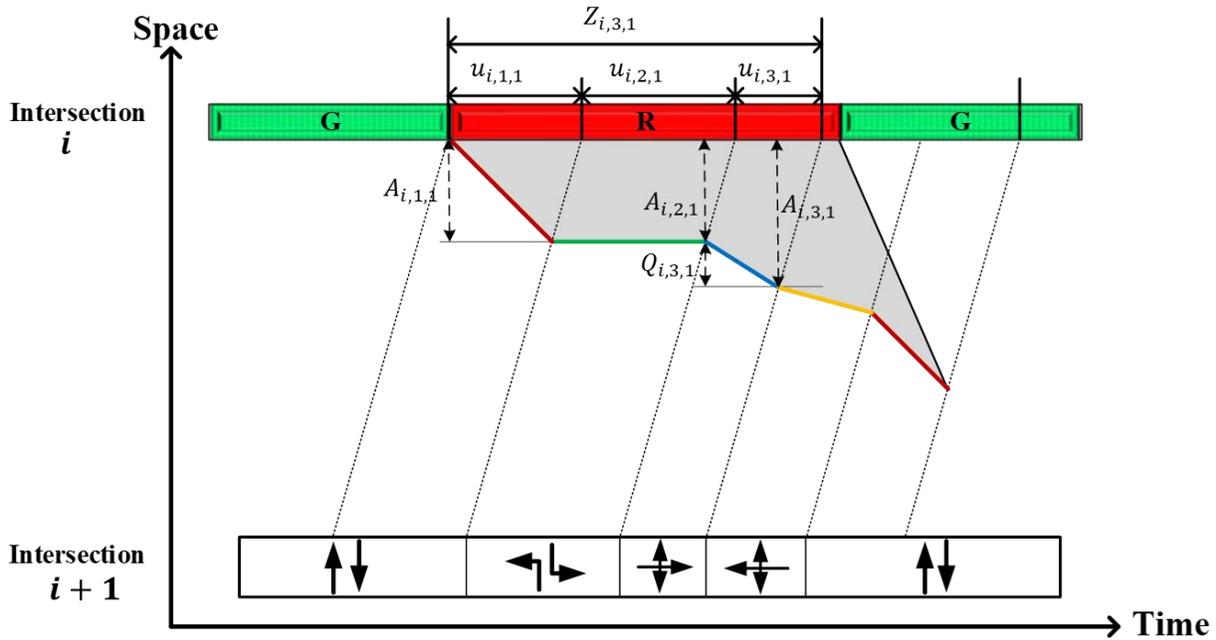


Figure 4.9 Graphical illustration of the cumulative queue length and cumulation duration for Through Queue Group- $Q_{i,3,1}$.

4.3.2.3 Quantifying each sub-delay area

Following the establishment of these key variables, namely, $A_{i,j,p}$ and $Z_{i,j,p}$, the formulations associated with each delay area are listed as follows:

Areas A and B:

$$U_{i,j,p} = u_{i,j,p} - \frac{Q_{i,j,p}}{k \cdot v} \quad \forall j = 1,2,3,4; \forall p = 1,2 \quad (4-23)$$

$$\widehat{D}_i = \left(\sum_{p=1}^2 \sum_{j=1}^4 \frac{(2 \cdot A_{i,j,p} - Q_{i,j,p}) \cdot U_{i,j,p}}{2} \right) - \frac{Q_i^T \cdot h \cdot Q_i^T}{2} \quad (4-24)$$

Equation 4-24 calculates Areas A and B, \widehat{D}_i , by summing all triangular and trapezoidal regions formed by the cumulative queue length over time. Each region's area is calculated using the standard formula for a trapezoid: the sum of the lengths of the parallel sides multiplied by the height, divided by two. Note that the last term in Equation 4-24, $\frac{Q_i^T \cdot h \cdot Q_i^T}{2}$, denotes the area of the purple dashed triangle in Figure 4.8, which should be subtracted from the total delay computation. In the case of Area A, one of the parallel sides will be equal to 0. The height of each region, denoted as $U_{i,j,p}$, and illustrated in Figure 4.10 as well as formulated in Equation 4-23, is obtained by subtracting the travel time of the last queuing vehicle from the duration of the corresponding queue group. Here, the travel time refers to the duration it takes for the last vehicle of a given queue group to reach the end of the preceding queue.

Area C:

$$\beta_{i,j,p} = \begin{cases} 1 & \text{when } \varepsilon_i^n \leq E'_{i,1,p} \leq Y_i^T \\ 0 & \text{otherwise} \end{cases} \quad \forall j = 1,2,3,4; \forall p = 1,2 \quad (4-25)$$

$$D_i^* = I \cdot \left(\sum_{p=1}^2 \sum_{j=1}^4 \beta_{i,j,p} \cdot A_{i,j,p} \right) \quad (4-26)$$

To achieve a more accurate delay estimation, it is essential to include the inter-green delay area, denoted as D_i^* (Area C). The computation of this area, as depicted in Equation 4-25 (see Figure 4.10), involves a process of determining which queue group's cumulative queue length should be taken into account.

The total delay at the intersection i can be obtained as shown in the following equation:

$$D_i = \widehat{D}_i + D_i^* \quad (4-27)$$

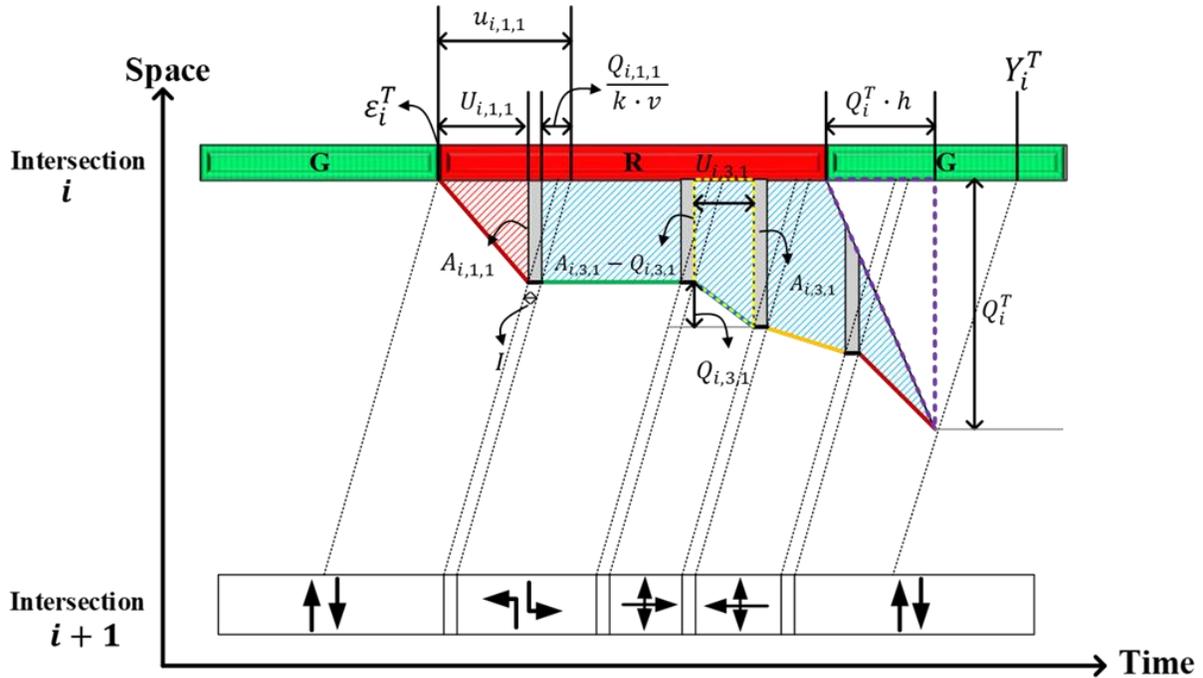


Figure 4.10 Graphical illustration for delay estimation

4.3.3 Traffic Dynamics Under Mutual Queue Blockage

As discussed previously, the intricate interplay between vehicles from the through queue and left-turn queue significantly influences an intersection's operational efficiency. The complexity of mutual blockage manifests in a multitude of negative effects, including wasted left-turn green time, reduced through-lane capacity, and excessive residual queues resulting from obstructed vehicles. Four distinct types of mutual blockage can be classified, depending on the sequence and frequency of blockages within a signal cycle and the initial signal timing plan. The types of mutual blockage and their corresponding impacts on intersection efficiency are detailed below:

- **Blockage Type 1:** The through queue initially blocks left-turn vehicles, preventing some from entering the turning bay in time. This leads to wasted green time for the left-turn movement and the accumulation of residual queues across subsequent cycles.
- **Blockage Type 2:** Left-turn queues spill over the turning bay, extending into the through lane and forcing through vehicles to merge into the adjacent lane, thereby causing a capacity reduction for the through-lanes
- **Blockage Type 3:** The through queue initially blocks left-turn vehicles, resulting in wasted green time and a residual left-turn queue. As the through queue discharges, the left-turn queue builds up and subsequently blocks the through queue, causing an additional capacity drop for through lanes.
- **Blockage Type 4:** Left-turn queues spill over the turning bays and block the through traffic. During the discharging period for left-turn queues, the through queue continues to accumulate and eventually obstructs the left-turn vehicles, leading to both underutilization of the green time for left-turn vehicles and a capacity loss for through lanes.

The primary works for addressing potential mutual queue blockage include:

- Blockage type identification
- Blockage duration computation
- Negative impact of the mutual blockage formulation

4.3.3.1 Blockage type identification

To determine the blockage type, it is crucial to consider both the sequence and frequency of queue blockages. This can be achieved through the following steps:

- **Step 1:** Identify the queue group that incurs the spillover or blockages in both the through and left-turn traffic streams;

- **Step 2:** Calculate the onset time for each turning bay spillover or blockage; and
- **Step 3:** Compute the duration between the onset of spillover or blockage and the beginning of the queue accumulation.

Taking the blockage incurred by the through traffic stream as an illustrative example, the corresponding formulations for each computational step are presented as follows:

Step 1:

$$\delta_{i,j,p}^T = \begin{cases} 1 & \text{if } (A_{i,j,p} - Q_{i,j,p}) \leq L_i^* \cdot k \leq A_{i,j,p} \\ 0 & \text{otherwise} \end{cases} \quad \forall j = 1,2,3,4; \forall p = 1,2; \quad (4-28)$$

Equation 4-28 introduces a set of binary variables, denoted as $\delta_{i,j,p}^T$, indicating the through queue group responsible for initiating the blockage in the left-turn bay. Specifically, $\delta_{i,j,p}^T$ equals 1 if Through Queue Group- ($Q_{i,j,p}$) at intersection i is the queue group that starts to impede left-turn vehicles; $\delta_{i,j,p}^T = 0$ if this queue group does not cause the blockage. In this equation, $L_i^* \cdot k$ represents the length of the left-turning bay measured by vehicle units, while $A_{i,j,p} - Q_{i,j,p}$ denotes the cumulative queue length prior to the formation of Through Queue Group- ($Q_{i,j,p}$).

Step 2:

$$\tilde{t}_i^T = \varepsilon_i^T + \sum_{p=1}^2 \sum_{j=1}^4 \delta_{i,j,p}^T \cdot \left((Z_{i,j,p} - u_{i,j,p}) + \frac{(L_i^* \cdot k) - (A_{i,j,p} - Q_{i,j,p})}{q_{i,j}^T} \right) \quad (4-29)$$

Equation 4-29 computes the exact onset time of the blockage, marking the point at which the original queue formation pace begins to change, as illustrated by the red-dashed arrow in Figure 4.11. The term $(L_i^* \cdot k) - (A_{i,j,p} - Q_{i,j,p})$ calculates the number of vehicles required for Through Queue Group- $Q_{i,j,p}$ to extend to the length of the left-turning bay at the intersection i .

Step 3:

$$H_i^T = \tilde{t}_i^T - \varepsilon_i^T \quad (4-30)$$

Equation 4-30 computes the time it takes for the through queue to reach the left-turning bay, denoted as H_i^T . This is determined by subtracting the start of the queue accumulation, computed in Equation 4-4, from the time point when the blockage occurs.

The same logic applies to computing the left-turn spillover from the left-turn bay. Steps 1 through 3 are repeated, with $\delta_{i,j,p}^L$ indicating the left-turn queue group responsible for initiating the spillover, \tilde{t}_i^L represents the onset time of the spillover event, and H_i^L represents the time it takes for the left-turn queue to extend and reach the entrance of the left-turning bay.

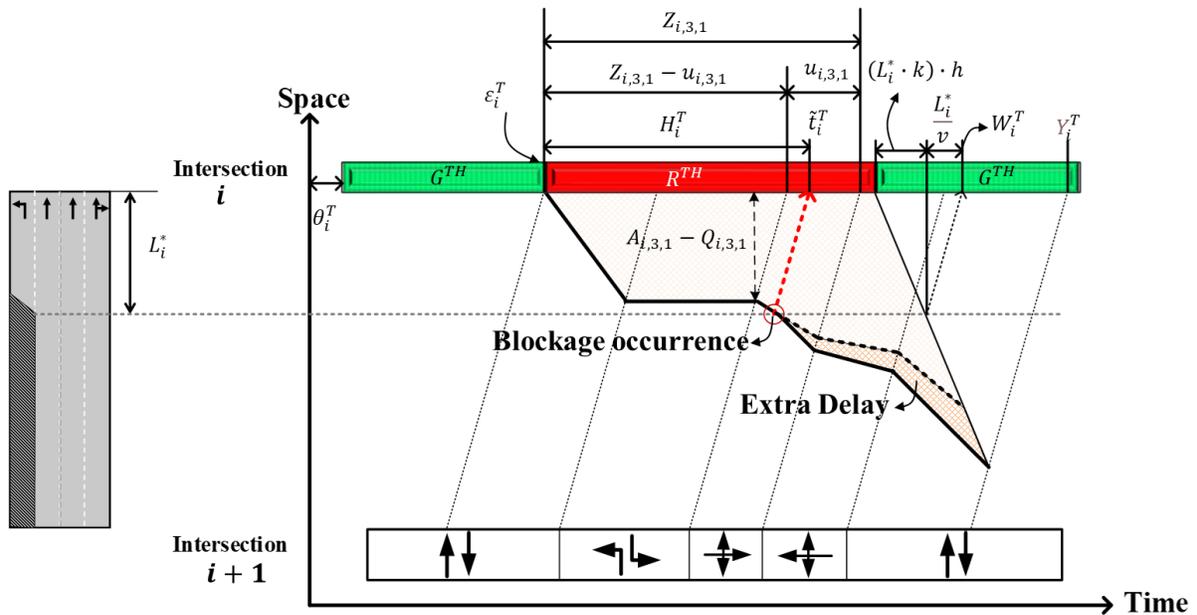


Figure 4.11 Graphical illustration of the blockage impact duration

The criteria for identifying each blockage type, based on the previously established variables, are shown as follows:

- **Blockage type 1: Blockage in the through queue only**

$$H_i^T > 0 \text{ and } H_i^L = 0$$

In this case, the through queue reaches the left-turning bay, causing a blockage, while the left-turn queue does not. This condition is characterized by a positive H_i^T and a zero H_i^L .

- **Blockage type 2: Blockage in the left-turn queue only**

$$H_i^T = 0 \text{ and } H_i^L > 0$$

Here, the left-turn queue reaches the left-turning bay, resulting in a blockage, while the through queue does not. This type is identified by a positive H_i^L and a zero H_i^T .

- **Blockage type 3: Mutual blockage caused initially by the through queue**

$$H_i^T > 0 \text{ and } H_i^L > 0 \quad \tilde{t}_i^T < \tilde{t}_i^L$$

This type occurs when both the through and left-turn queue lengths extend beyond the left-turning bay, with the through queue causing the initial blockage. It is identified by positive values for both H_i^T and H_i^L , with the onset time of the through blockage (\tilde{t}_i^T) preceding that of the left-turn queue (\tilde{t}_i^L).

- **Blockage type 4: Mutual blockage initiated by the left-turn queue**

$$H_i^T > 0 \text{ and } H_i^L > 0 \text{ and } \tilde{t}_i^T > \tilde{t}_i^L$$

This type occurs when both the left-turn and through queue lengths exceed the left-turning bay, with the left-turn queue causing the initial mutual blockage. It is identified by positive values for both H_i^T and H_i^L , and the onset time of the left-turn queue blockage (\tilde{t}_i^L) being earlier than that of the through queue (\tilde{t}_i^T).

4.3.3.2 Blockage duration computation

The duration of blockage begins with the onset of the spillover or blockage and concludes when the queue in front has been discharged, allowing the blocked vehicles to resume moving forward. For Blockage Types 3 and 4, where multiple blockages occur within one cycle, the onset time of subsequent blockage must be recalculated based on the dissolving time of the initial blockage. The ending time of the blockage, denoted as W_i^n , is calculated by adding the time required to discharge the bay-length queuing vehicles to the starting time of the green phase. This calculation is illustrated in Figure 4.11 and is expressed by the following equation, applicable to both the first and subsequent blockages:

$$W_i^n = \theta_i^n + C + (L_i^* \cdot k) \cdot h + \frac{L_i^*}{v} \quad \forall n = (T, L) \quad (4-31)$$

Note that the ending time of the first blockage marks the moment when the previously blocked vehicles resume moving forward and join the queue.

To address blockage types involving multiple blockages within a single cycle (i.e., Type 3 and Type 4), the onset of the blockage in Equation 4-29, originally denoted as \tilde{t}_i^n , should be adjusted to account for the duration of the initial blockage. The adjusted onset time is denoted as t_i^n . Specifically, if the initial blockage is caused by the left-turn queue exceeding the bay capacity, the adjustment incorporates the term $(W_i^L - t_i^L)$, as shown in Equation 4-23a. Alternatively, if the through queue initially causes the blockage by surpassing the bay length, the adjustment includes the term $(W_i^T - t_i^T)$, as presented in Equation 4-32b.

$$t_i^T = \begin{cases} \tilde{t}_i^T & \text{if } H_i^T \times H_i^L = 0 \\ t_i^{T'} + (W_i^L - t_i^L) & \text{otherwise} \end{cases} \quad (4-32a)$$

$$t_i^L = \begin{cases} \tilde{t}_i^L & \text{if } H_i^T \times H_i^L = 0 \\ \tilde{t}_i^L + (W_i^T - t_i^T) & \text{otherwise} \end{cases} \quad (4-32b)$$

4.3.3.3 Formulating the negative impacts of mutual queue blockage

To fully address the probable impacts arising from various types of mutual blockages, it is necessary to augment the existing delay formulations to encompass the effects associated with each blockage event. Those impacts can be classified into three primary categories, based on the blockage characteristics, their duration, and signal timing:

- Increase in the speed of the through queue formation
- Reduced capacity on through lanes
- Number of blocked left-turn vehicles and residual queues

Increase in the speed of through queue formation.

If the through queue exceeds the bay length before it starts to dissipate, the arrival rate for the through-queue formation during the remaining period should include left-turn flows. This impact can be incorporated by updating the vehicle arrival rate in Equation 4-1 with the following equations. These adjustments reflect the effects of the blockage and the computation of queue length and delay.

$$\alpha_{i,j,p} = \begin{cases} 1 & \text{if } W_i^T > S'_{i,j,p} > t_i^T \\ 0 & \text{otherwise} \end{cases} \quad \forall j = 1,2,3,4; \forall p = 1,2 \quad (4-33)$$

$$Q_{i,j,p} = u_{i,j,p} \cdot (q_{i,j}^T + \alpha_{i,j,p} \cdot q_{i,j}^L) + \delta_{i,j,p}^T \cdot [(E_{i,j,p} - t_i^T) \cdot q_{i,j}^L] \quad \begin{matrix} \forall j = 1,2,3,4; \\ \forall p = 1,2 \end{matrix} \quad (4-34)$$

With such a critical variable, $\alpha_{i,j,p}$, being identified, Equation 4-33 is introduced to determine whether vehicles from the Through Queue Group- ($Q_{i,j,p}$) at intersection i have joined the queues after the onset of blockage. Specifically, $\alpha_{i,j,p}$ is set to 1 if vehicles from the group have joined the queue, and 0 otherwise. Given that vehicle arrival rates following the onset of blockage must account for the presence of left-turn vehicles, Equation 4-1 is accordingly modified and replaced with Equation 4-34. In this adjustment, the last term, $\delta_{i,j,p}^T \cdot [(E_{i,j,p} - t_i^T) \cdot q_{i,j}^L]$, signifies the number of left-turn vehicles joining the through queue from the queue group that initially triggers the blockage.

Reduced the capacity on the through lanes

The overflow of left-turn vehicles into the through lanes reduces the effective through-lane capacity, necessitating corresponding adjustments in the delay computation to accurately reflect the diminished discharge rate. As shown in Figure 4.12(a), at time point A, the left-turn queue begins to intertwine with the through queue, evidenced by a noticeable reduction in the dissipation wave propagation speed.

Consequently, the delay computations must be modified in Area B, where the dissipation wave profile transitions from a straight line to a segmented (polyline) curve. Specifically, the last term in Equation 4-24, $\frac{Q_i^n \cdot h \cdot Q_i^n}{2}$, representing the area of the dashed purple triangle in Figure 4.10, is no longer applicable when a drop in through-lane capacity is observed. Instead, the triangle is replaced with a quadrilateral, as shown in Figure 4.12. This quadrilateral can be further subdivided into Area B1 and Area B2, as shown in Figure 4.12(b), with the corresponding formulations for these subareas provided below:

$$U_i^* = \frac{t_i^L - (\theta_i + C)}{h + \frac{1}{k \cdot v}} \quad (4-35)$$

Area B1:

$$\hat{D}_i^1 = U_i^{*2} \cdot h \cdot \frac{1}{2} \quad (4-36)$$

Area B2:

$$\hat{D}_i^2 = \frac{(U_i^* + Q_i^T) \cdot (U_i^* \cdot \frac{1}{k \cdot v})}{2} \quad (4-37)$$

Equation 4-35 calculates the base of Area B1 and the short parallel side of Area B2, U_i^* , which is essential for determining the areas of B1 and B2. U_i^* can be derived given the onset of the blockage. This calculation is based on the time it takes for the last queuing vehicle to reach the stop line before the blockage occurs, represented as $t_i^L - (\theta_i + C)$, as shown in Figure 4.12(b).

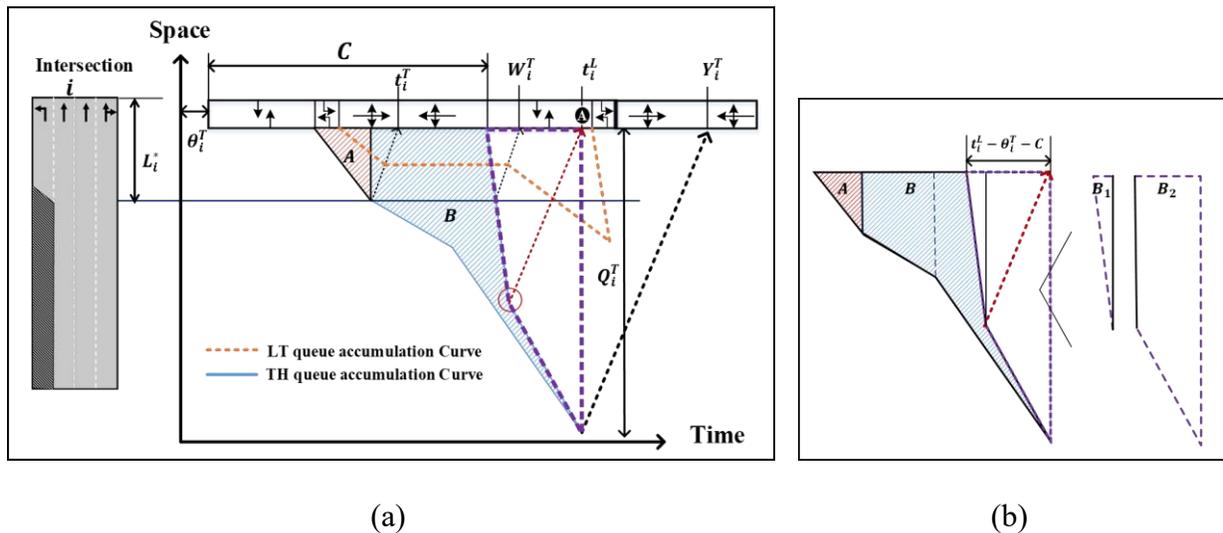


Figure 4.12 Delay Computation accounted for the through-lane's capacity drop

The delay formulation of Equation 4-24 can therefore be generalized with the following equations:

$$\varphi_i = \begin{cases} 1 & \text{if } H_i^L > 0 \text{ and } W_i^L > \theta_i^T + C \\ 0 & \text{otherwise} \end{cases} \quad (4-38)$$

$$\widehat{D}_i = \left(\sum_{p=1}^2 \sum_{j=1}^4 \frac{(2 \cdot A_{i,j,p} - Q_{i,j,p}) \cdot U_{i,j,p}}{2} \right) - (1 - \varphi_i) \cdot \frac{Q_i^T \cdot h \cdot Q_i^T}{2} - \varphi_i \cdot (\widehat{D}_i^1 + \widehat{D}_i^2) \quad (4-39)$$

Equation 4-38 introduces a new set of binary variables, φ_i . When the impact of the left-turn blockage extends to the onset of the through green phase, a through-lane capacity drop is inevitable, thus $\varphi_i = 1$. In Equation 4-39, when $\varphi_i = 1$, the term equals $(1 - \varphi_i) \cdot \frac{Q_i^T \cdot h \cdot Q_i^T}{2}$ zero. Conversely, when $\varphi_i = 0$, it indicates that the blockage does not affect the through queue discharge, the term $\varphi_i \cdot (\widehat{D}_i^1 + \widehat{D}_i^2)$ equals zero.

Number of blocked left-turn vehicles and residual queues

When the through-queue exceeds the length of the left-turn bay, left-turn vehicles are blocked and must wait until the through queue begins to discharge. If these blocked left-turn vehicles cannot enter the turning bay in time to pass through the intersection during the left-turn green phase, they will form a residual queue and experience an additional cycle of delay. To properly account for such excessive delays, the objective function is augmented to include the minimization of extra delay experienced by these vehicles. Specifically, only the left-turn vehicles that fail to access the turning bay promptly and subsequently become part of the residual queue are considered in this adjustment.

Based on the relationship between the end of the through blockage impact and the left-turn green phase, the number of left-turn vehicles unable to enter the turning bay in time due to the blockage by the through queue, denoted as F'_i , can be quantified using the following equation:

$$\gamma_i = \begin{cases} 1 & \text{if } H_i^T > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4-40)$$

$$F_i = \sum_{p=1}^2 \sum_{j=1}^4 u_{i,j,p} \cdot (\alpha_{i,j,p} \cdot q_{i,j}^L) + \delta_{i,j,p}^T \cdot [(E_{i,j,p} - t_i^T) \cdot q_{i,j}^L] \quad (4-41)$$

$$F'_i = \begin{cases} \gamma_i \cdot (F_i) & W_i^T > \theta_i^L + C + g_i^L \\ 0 & W_i^T < \theta_i^L + C \\ \left(Y_i^T - (\theta_i^L + C + g_i^L) \right) \times \frac{k \cdot v}{h \cdot k \cdot v + 1} \times \rho_i^L & \text{otherwise} \end{cases} \quad (4-42)$$

Equation 4-40 introduces a set of binary variables, γ_i , which equals 1 if the through queue blocks the left-turn vehicles, and 0 otherwise. In Equation 4-41, the total number of blocked left-turn vehicles, denoted as F_i , is computed based on Equation 4-34. The blocked left-turn vehicles that become the residual queue for the next cycle are calculated using Equation 4-42. If the impact of the through blockage persists beyond the end of the left-turn green phase, none of the affected left-turn vehicles will be able to pass through the intersection within the current cycle. Consequently, all such blocked left-turn vehicles will remain in the residual queue and must wait until the subsequent signal cycle to proceed. If the through blockage impact ends before the activation of the left-turn green phase, all blocked left-turn vehicles can enter the bay in time, and not form any residual queue. In the third case, where the impact ends during the left-turn green phase, some blocked left-turn vehicles will become residual queues. This calculation involves estimating the total number of vehicles (both through and left-turn) remaining undischarged

between the end of the left-turn green phase and the arrival of the last through-queue vehicle at the intersection stop bar, multiplied by the left-turn turning ratio (see Figure 4.13).

The additional delay experienced by the obstructed left-turn vehicles, represented by D'_i , is shown in Equation 4-43. These vehicles must wait for an extra cycle before they can be discharged, resulting in additional delay.

$$D'_i = F'_i \times C \tag{4-43}$$

The original total delay at intersection i , as defined in Equation 4-27, can thus be updated to incorporate this additional residual queue delay. Specifically, the total delay can be modified as:

$$D_i = \widehat{D}_i + D_i^* + D'_i \tag{4-44}$$

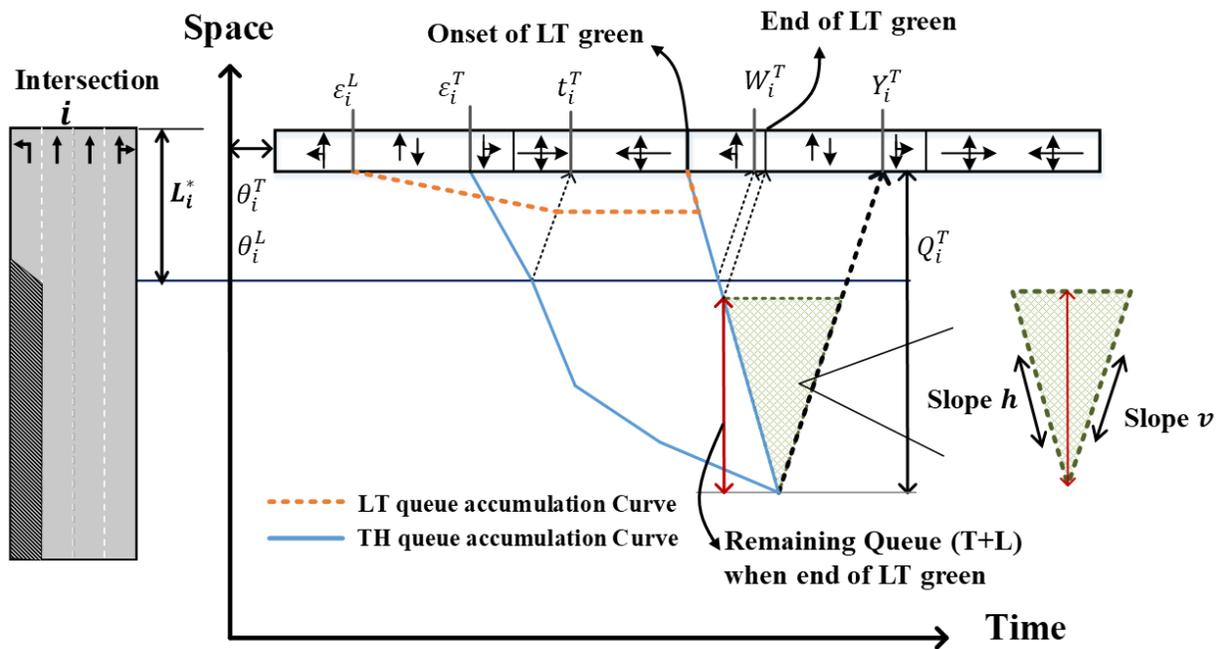


Figure 4.13 Graphical illustration of the left-turn residual queue

4.3.4 Progression Design for the Low-Traffic-Volume Direction

As previously stated, the developed module emphasizes the preservation of progression in the direction experiencing low traffic volume. In order to ensure the desired progression, one can formulate the constraints tailored for the low-traffic-volume direction, adopting a similar notion as to MAXBAND (Little et al., 1981) as follows:

$$\bar{\theta}_i^T + \bar{\omega}_i + \pi_i + \bar{K}_i^T \cdot C = \bar{\theta}_{i+1}^T + \bar{\omega}_{i+1} \quad (4-45)$$

$$\bar{B} + \bar{\omega}_i \leq \bar{g}_i^T \quad (4-46)$$

$$\bar{B} \geq \bar{B}' \quad (4-47)$$

Where \bar{B}' refers to the pre-calculated bandwidth of the low-traffic-volume direction computed by MAXBAND. Equation 4-47 ensures the low-traffic-volume direction has the minimum bandwidth, as obtained from MAXBAND.

4.3.5 Relations Between the Optimal Phase Sequences and the Starting Time of Each Phase

To enhance the model's depiction of queue evolution and delay estimation, the following expressions outline the configurations for left-turn phase patterns. These patterns are crucial in developing a signal control plan for arterial roads, with a specific emphasis on optimizing phase sequences and offsets in this study. The starting times for the green phase of each movement should be calculated concurrently with the optimization of the intersection's offsets and phase sequence. The constraints for this process, as applied in MAXBAND (Little et al., 1981) and MULTIBAND (Gartner et al., 1991), are specified as follows:

$$\theta_i^T = \theta_i + (1 - x_i) \cdot (I + \bar{g}_i^L) \quad (4-48)$$

$$\theta_i^L = \theta_i + (1 - \bar{x}_i) \cdot (I + \bar{g}_i^T) \quad (4-49)$$

Where θ_i represents the offset of intersection i ; θ_i^L indicates the starting time of a green phase for left-turn movements at the same intersection; when the binary variable \bar{x}_i equals 1, it signifies the leading left-turn phase of the low-volume direction at intersection i ; I is the inter-green duration.

4.3.6 Summary

In brief, to minimize the through delay in the high-traffic-volume direction while maintaining the progression bandwidth in the low-traffic-volume direction, the objective function of the developed module can be specified as follows:

$$\text{Min} \sum_i D_i \quad \forall i \quad (4-50)$$

The formulations for traffic conditions associated with the asymmetric directional volumes can be summarized as follows:

- Queue length computation: Equations 4-2 to 4-3
- Queue evolution pattern: Equations 4-4 to 4-14
- Delay estimation: Equations 4-15 to 4-27
- Constraint for the progression in the low-traffic volume direction: Equations 4-45 to 4-47
- Varying phase sequences and offsets: Equations 4-48 and Equation 4-49
- Mutual blockage Identification: Equations 4-28 to 4-32
 - Impact of mutual blockage- Type 1: Equations 4-33 to 4-34 and 4-40 to 4-44
 - Impact of mutual blockage- Type 2: Equations 4-35 to 4-39
 - Impact of mutual blockage- Type 3: Equations 4-33 to 4-44
 - Impact of mutual blockage- Type 4: Equations 4-33 to 4-44

Note that in scenarios where no blockage occurs (e.g., no exclusive left-turn lane, or when left-turns are not allowed), one can simplify the developed module by removing the set of constraints related to mutual blockage (i.e., Equations 4-28 to 4-44).

In summary, by accommodating varying phase sequences, offsets, and flexible formulations for queue length computation—capable of handling any number and sequence of queue groups—the module can accurately compute the delay associated with any queue evolution pattern. This comprehensive framework enables the identification of an optimal pattern that not only minimizes queuing delays but also reduces the duration of mutual blockages by prioritizing streams with low traffic flow. The optimal pattern is achieved by setting the objective to minimize overall through delays.

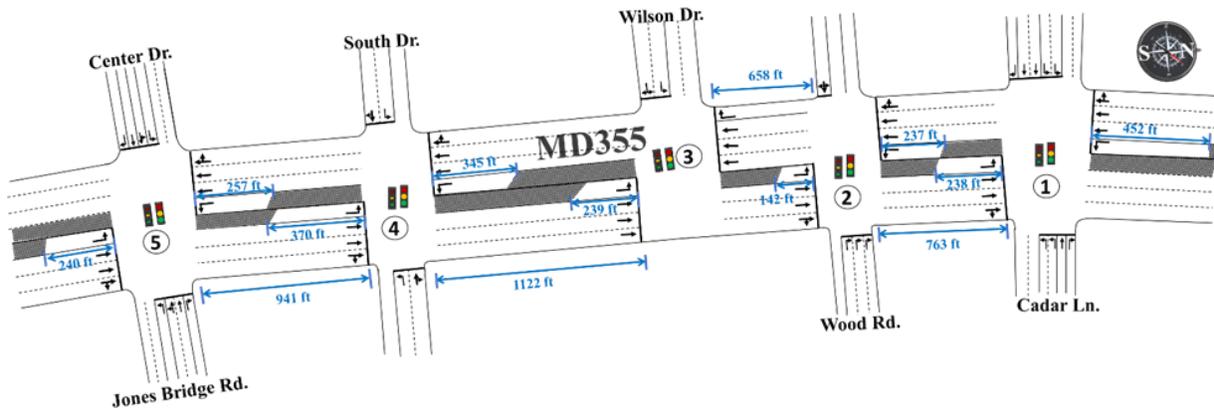
4.4 Case Study

The case study, containing both numerical experiments and simulations, is designed for the following two tasks:

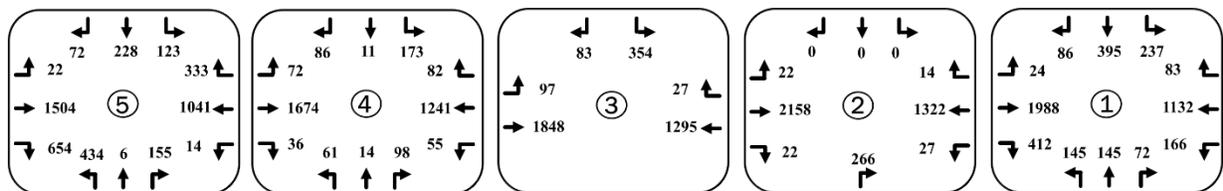
- Comparing the signal plans generated by the developed module against those from benchmark models, focusing on the resulting queue evolution processes and associated delays.
- Conducting simulation experiments to observe the delays and number of stops under both the developed and benchmark models to verify the developed module's contributions with respect to mitigating traffic congestion and queue-incurred blockages in major commuting arterials accommodating highly asymmetric traffic volumes.

The selected site for experimental analyses is a segment of MD 355 Rockville Pike in Bethesda, Maryland, encompassing five signalized intersections. Figure 4.14 presents detailed information about this site, including its geometric layout (Figure 4.14(a)) and evening peak-hour

traffic volume distribution (Figure 4.14(b)). This arterial segment plays a critical role as a connector between residential areas and the central business district near the District of Columbia. During the evening peak period, a pronounced directional imbalance emerges, with northbound traffic volumes significantly exceeding those in the southbound direction. Such characteristics make this site an ideal setting for evaluating the developed module's ability to manage congestion and queue spillovers along arterials experiencing highly asymmetric traffic flows. Note that the evening peak-hour volumes are sourced from the Maryland State Highway Administration's Internet Traffic Monitoring System (ITMS).



(a)



(b)

Figure 4.14 (a) Geometric features and (b) peak-hour volume distributions over the candidate arterial segment

4.4.1 Performance Analysis

The analyses conducted at this stage are mainly for verifying:

- Whether all developed formulations can reliably capture complex interrelations between various queue streams for the module to produce the signal control plan with the minimum delay in the high-volume direction;
- The necessity and benefits of the developed module that fully accounts for major complex queue formations between the arterial’s intersections;
- The consistency between the traffic volumes and the resulting delays, as well as the queue length distribution among intersections.

Note that the real-world volumes shown in Figure 4.14(b) have been revised to highlight the target commuting arterial’s congested and asymmetric traffic distribution in two opposite directions, as shown in Figure 4.15. The pedestrian volume is moderate; therefore, assumed to be manageable within the red phase duration for conflicting movement. The cycle length is set as 150 seconds, and the green splits are computed with the critical lane volumes (CLV). The developed module was solved by Gurobi 9.0 on a workstation with 16GB RAM in 30 seconds.

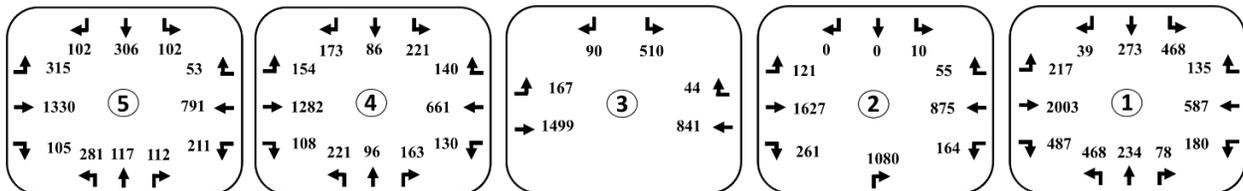


Figure 4.15 Volume scenario at the study site designed for module verification

Figure 4.16 shows the outputs of the developed module. Figure 4.17 further visualizes the output results with time-space diagrams showing the offsets and phase sequences produced by the developed module and the MAXBAND (Little et al., 1981) model, along with the queue evolution

process and the bidirectional progression bands. Table 4.2 summarizes selected-intersection specific features and measures of effectiveness (MOEs) derived from the developed module and MAXBAND.

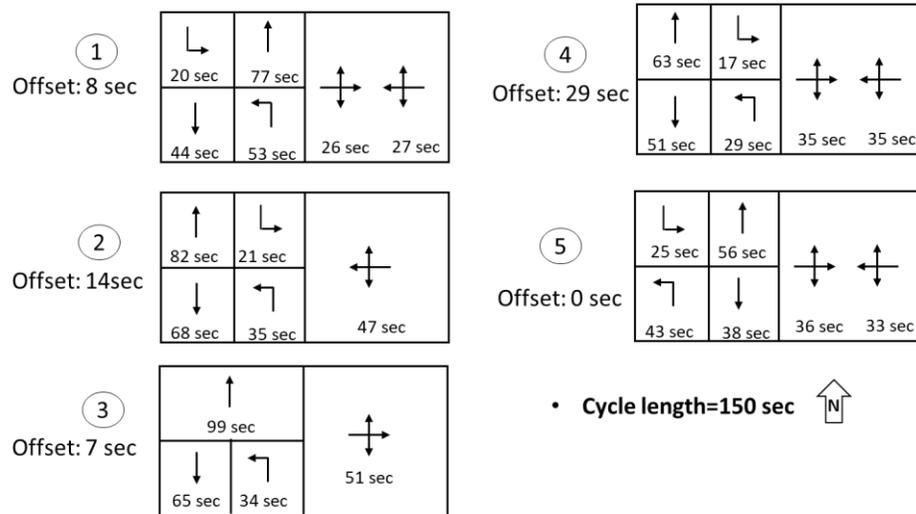
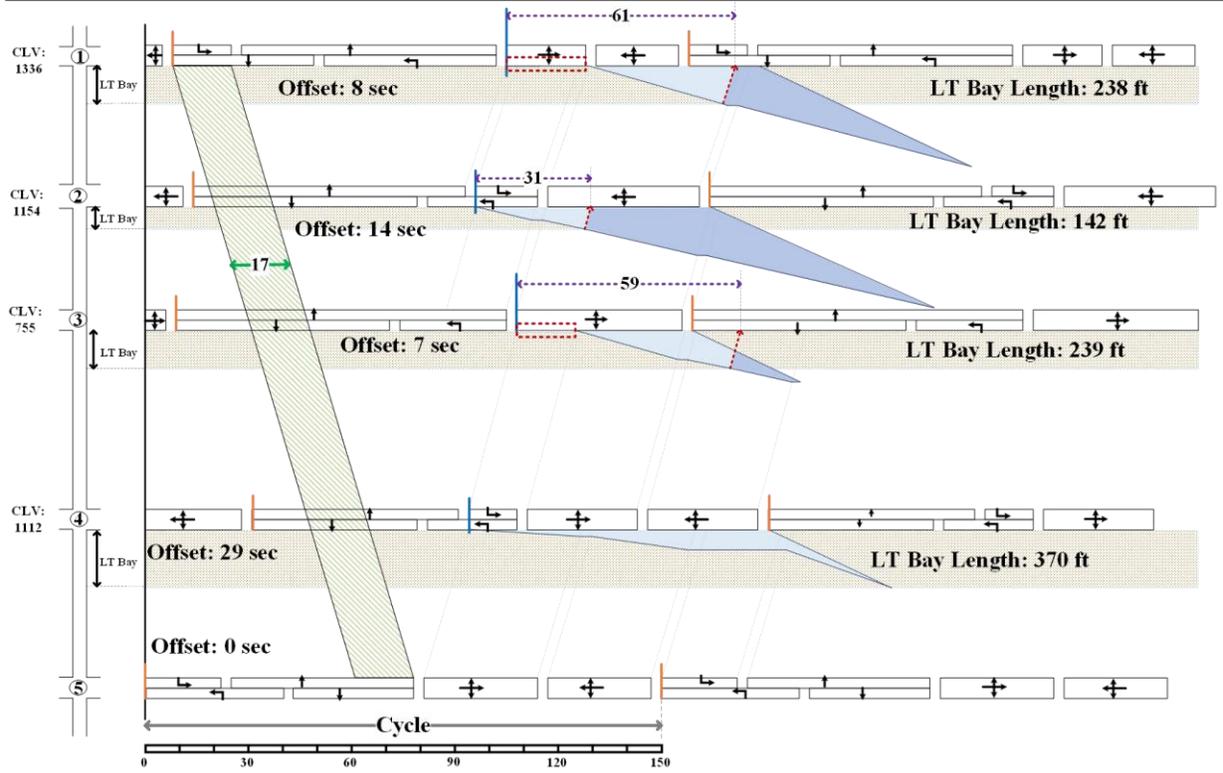
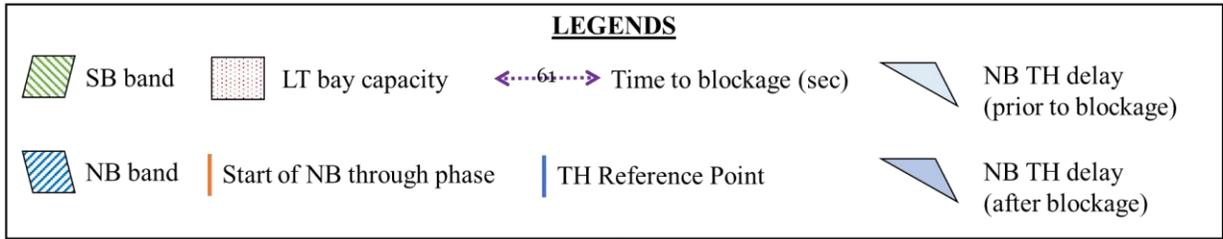
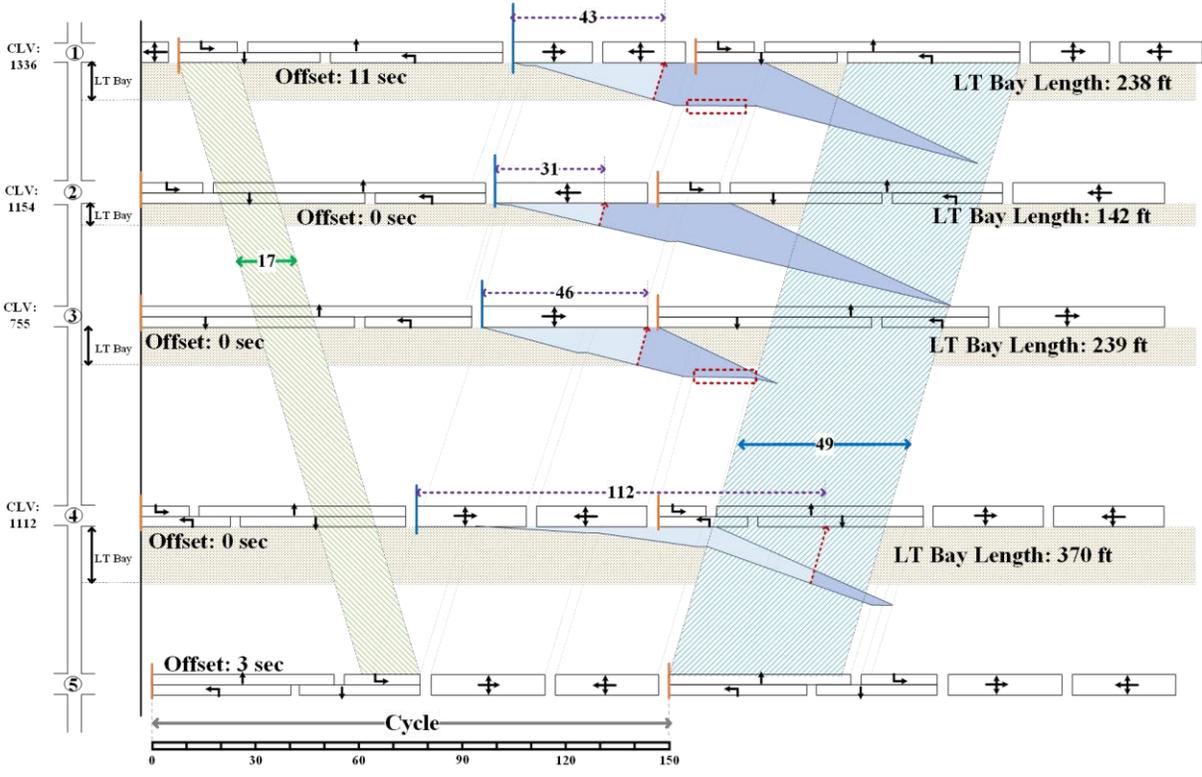
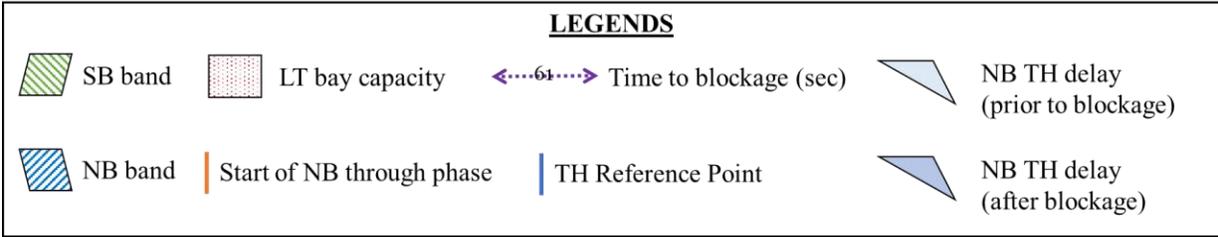


Figure 4.16 Outputs of the developed module



(a)



(b)

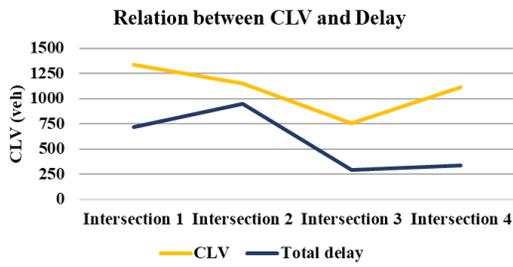
Figure 4.17 Time-space diagram and queue evolution under (a) the developed module and (b) the MAXBAND model

Table 4.2 Performance Comparison between the developed module and MAXBAND with the selected MOEs

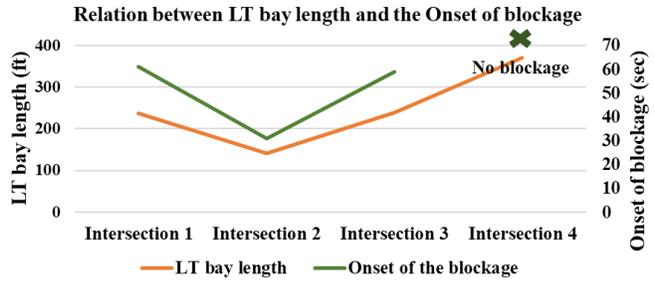
Inter-section	CLV (veh)	LT bay length (ft)		Duration to the onset of blockage(sec)	Maximum queue length (veh)	Total delay (veh.sec /ln.cyc)	Average delay (sec/veh.ln.cyc)
1	1336	238	Developed module	61	29	721	25.75
			MAXBAND	43	29	1000	35.71
2	1154	142	Developed module	31	29	951	36.25
			MAXBAND	31	29	936	35.68
3	755	239	Developed module	59	15	294	14.11
			MAXBAND	46	17	309	14.83
4	1112	370	Developed module	(No blockage)	14	338	17.51
			MAXBAND	112	23	410	21.24
Total			Developed module			2304	24.41
			MAXBAND			2655	28.13

The major findings are summarized below:

The results in Figure 4.17(a) and (b) and Table 4.2 demonstrate that the module’s embedded formulations are capable of explicitly reflecting the complex interaction between different streams of queue flows along the arterial, as evidenced by the consistency between the distribution of traffic volumes and their resulting delays as well as queue among intersections (see Figure 4.18(a)). Specifically, as shown in Table 4.2 and Figure 4.18(a), Intersection 3 with the lowest CLV has the lowest total and average delays for Northbound through vehicles among all intersections. The higher delay observed at Intersections 1 and 2 resulted from not only their high CLVs but also the short left-turn bays, which may incur early queue blockage in each cycle. For instance, the blockage takes place as soon as 31 seconds after the onset of the queue formation at Intersection 2, as shown in Table 4.2. At Intersection 4, with a longer left-turn bay length, the turning bay blockage is not expected to occur.



(a)



(b)

Figure 4.18 Performance comparison with the selected measurements of effectiveness: (a) relation between CLV and total delay; (b) relation between the left-turn bay length and the onset time of blockages

Figure 4.17(a) further shows that, as expected, the signal timing plan produced by the developed module ensures that vehicles during the phase experiencing low speeds of the queue wave can join the queues ahead of those traffic streams of higher arrival rates. Specifically, at Intersection 1 and Intersection 3, traffic streams not contributing to the Northbound through queue (see red boxes in Figure 4.17) will receive the right-of-way ahead of other movements in each cycle. At Intersection 4, the traffic stream with the highest flow rates (i.e., the steepest queue accumulation curve) joins the queues after all other streams of vehicles have arrived. With such phase sequences, the target arterial can not only have the minimum total delay of all vehicles but also slow the queue formation pace and hold up the onset time of overflowing the left-turn bay if it is inevitable.

The comparison results between Figure 4.17(a) and (b) and Table 4.2 assert the benefits of incorporating the delay-minimization feature for the high-volume direction in the design of two-way progression for commuting arterials accommodating highly asymmetric flows. Noticeably,

the developed module can enhance the optimal signal progression plan by MAXBAND (Little et al., 1981) on the following aspects:

1. Preventing the potential left-turn queue blockage at Intersection 4;
2. Delaying the increase in the queue formation speed, even if the left-turn bay blockage due to the long through queue is highly likely;
3. Optimizing the phase sequences to allow traffic from low-volume streams (in the red dashed squares) to arrive at intersections ahead of other movements so as to reduce the time-varying queue formation pace and the resulting total delay.

As such, although the resulting maximum queue lengths were comparable between the two models, the developed module can yield a lower total and average delay at almost all intersections, as shown in Table 4.2.

4.4.2 Performance Evaluation

For performance assessment, this study has adopted the traffic simulator with VISSIM (PTV, 2018) to compare the developed module under field traffic scenarios with the results from three well-recognized benchmarks, Synchro (Cubic, 2019), TRANSYT-7F (Wallace et al., 1984) and MULTIBAND (Gartner et al., 1991) which understandably were developed mainly for more generalized traffic patterns in commuting corridors. All simulated real-world traffic demands used for performance evaluation with simulation are shown in Figure 4.14(b). Table 4.3 summarizes the selected MOEs obtained from the results of simulation analyses. Note that fifteen random simulations have been conducted for each model's simulation analysis, each with a 300-second warm-up period and 3600-second simulation time. Additionally, to demonstrate the benefit of the developed module in optimizing arriving phase sequences, Figure 4.19 illustrates the time-

dependent queue length evolution over three cycles from a selected simulation run, with each time interval representing 5 seconds. This Figure uses Intersection 2 as an example, which is a critical intersection in the study due to its short link and short left-turn bay.

The comparison results in Figure 4.19 show that the developed module consistently achieves a shorter queue length across all cycles when compared to the other benchmark models. This is particularly important in the context of traffic management, as it indicates a more efficient handling of vehicle accumulation at the intersection, thereby reducing the likelihood of spillback into upstream lanes. Given the longer queue, the resulting increase in discharge time directly translates into additional delays. For instance, during the first cycle (i.e., from Time Interval 1 to Time Interval 30), the total queue time (from the onset of queue accumulation to complete dissipation) spans only 3 time intervals under the developed module. In contrast, the queueing time persists for over 5 intervals under Synchro, 6 intervals under MULTIBAND, and 8 intervals under TRANSYT-7F. The noticeably shorter queue duration achieved by the developed module proves its ability to better minimize the total delay.

Furthermore, the queue evolution pace under the developed module is relatively slower than under the other counterparts. The resulting slow pace is one of the module's main advantages as it can prevent or at least delay the onset of queue spillback up to the left-turn bay, a common issue that can severely disrupt traffic flow and result in excessive delays. For example, the left-turn bay at Intersection 2, about 142 feet long, is most likely to experience blockages when queues grow quickly. The developed module, as expected, can effectively prevent turning-bay blockage or reduce its duration, thereby preventing potential subsequent delay impacts on adjacent lanes.

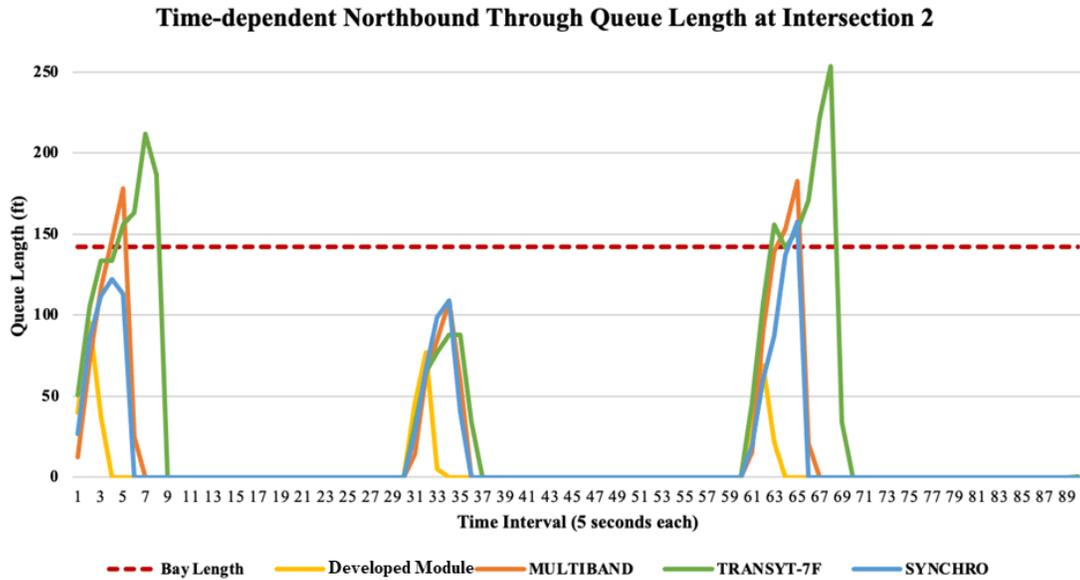


Figure 4.19 Time-dependent queue length evolution at Intersection 2

Table 4.3 Performance Comparisons of the signal plans generated by three models

Direction	Link No.	Synchro	TRANSYT-7F	MULTIBAND	Developed Module	% change compared to Synchrho ¹	% change compared to TRANSYT-7F ¹	% change compared to MULTIBAND ¹
Through delay along the arterial (sec/veh)								
Northbound ³	⑤→①	66.5	74.33	71.25	65.27	-1.85%*	-12.19%*	-8.39%*
Southbound ³	①→⑤	38.99	35.81	35.21	35.24	-9.62%*	-1.59%*	0.09%
Average delay of through vehicles from all upstream streams² on individual links (sec/veh)								
	②→①	2.08	2.87	2.26	1.92	-7.69%*	-33.10%*	-15.04%*
Northbound ³	③→②	0.95	1.43	1.31	0.86	-9.47%*	-39.86%*	-34.35%*
	④→③	1.34	3.34	2.67	1.79	33.58%	-46.41%*	-32.96%*
	⑤→④	23.26	37.82	33.42	21	-9.72%*	-44.47%*	-37.16%*
Average number of stops of through vehicles along the arterial								
Northbound ³	⑤→①	1.54	1.54	1.59	1.35	-12.34%*	-12.34%*	-15.09%*
Southbound ³	①→⑤	0.84	0.8	0.77	0.78	-7.14%*	-2.50%*	1.30%
Network average delay (sec/veh)								
Both directions		126.63	137.36	126.69	121.83	-3.79%*	-11.31%*	-3.84%*
Network average number of stops								
Both directions		1.8	2.17	1.79	1.78	-1.11%*	-17.97%*	-0.56%

Note:

1. The Pair-T test results of percent changes are denoted with an asterisk: * = significance level of at least 0.05
2. All vehicles from the upstream intersections, including through and side-street turning-in vehicles heading in the northbound direction, will be accounted for at the downstream intersection
3. The northbound direction is the high-volume direction; the southbound direction is the low-volume direction

Major findings from the performance evaluation results, shown in Table 3, are summarized below:

- The developed module has proved its effectiveness in reducing through delays along such arterials of highly asymmetric volumes in two opposing directions when compared with the results with Synchro, TRANSYT-7F, and MULTIBAND in the case study. Specifically, the northbound through delay along the arterial was reduced by 1.85% compared to Synchro, 12.19% compared to TRANSYT-7F, and 8.39% compared to MULTIBAND. These reductions highlight the module's effectiveness in tackling high-volume traffic efficiently, especially where some intersections (e.g., Intersections 1-3) are likely to experience turning-bay overflows and lane blockage.
- When evaluating the average delay of through vehicles across all upstream traffic streams on individual links in the northbound direction, the developed module can consistently outperform the benchmark state-of-the-practice models. For example, the average delay has been reduced by 7.69%, 33.10%, and 15.04% compared to those under Synchro, TRANSYT-7F, and MULTIBAND, respectively, on the segment from Intersection 2 to Intersection 1. Similar improvements can also be observed on other links, further supporting the module's effectiveness in minimizing delays for all traffic streams over commuting arterials with such geometric features and volume distribution patterns.
- As an indicator of the progression efficiency in both directions, the average number of stops for the northbound through movement with the developed module is 12.34%, 12.34%, and 15.09% lower compared to those with Synchro, TRANSYT-7F, and MULTIBAND, respectively. In the southbound direction of lower traffic volumes, the

developed module still achieves notable reductions of 7.14% and 2.50%, compared to those with Synchro and TRANSYT-7F, but only a slight increase of 1.30% when compared to MULTIBAND. These results seem to support that the developed module's minimization of the total delay for the high-volume direction is not accomplished at the cost of progression efficiency for the lower-volume direction.

- At the network-wide level, the developed module has effectively reduced both the average delay and the number of stops across the entire network. The networkwide average delay has been reduced by 3.79%, compared to those with Synchro, 11.31% with TRANSYT-7F, and 3.84% with MULTIBAND. Similarly, the network-wide average number of stops exhibits reductions of 1.11%, 17.97%, and 0.56% when compared to Synchro, TRANSYT-7F, and MULTIBAND, respectively.

4.4.3 Case Study Summary

In summary, the results of case studies show that the developed formulations can reliably reflect the complex interactions between various time-varying queues and signal control plans, thus allowing the developed module to effectively reduce the traffic delay in the high-volume direction while maintaining the same maximized progression bandwidth for vehicles in the low-volume direction. More specifically, the developed formulations for the intersection through-movement delays have accounted for most interdependent time-varying factors contributing to the formation and evolution of queue patterns over the arterial links, as well as the impact of turning bays. Hence, with the specified objective function, the developed module can produce the signal plans that enable traffic in the high-volume direction to minimize the total delay by optimally sequencing the traffic streams of high flow rate to join the intersection queue formation after those of low volume. Such a unique property along with the adoption of two different control objectives for the arterial's

two highly asymmetric traffic directions allow the developed module to also yield expected efficiency with respect to the networkwide delay reduction, decreased number of vehicle stops, a lower likelihood of queue spillbacks at tuning bays, and a shorter left-turn bay blockage duration if not preventable.

Chapter 5 A Responsive Real-Time Arterial Signal Control Module for Turning-bay Overflows and Mutual Queue Blockages

5.1 Introduction

As demonstrated in Chapter 4, the time-of-day signal control module designed for congested commuting arterials with asymmetric directional flows effectively optimizes signal offsets along the arterial and coordinates phase sequences at critical intersections. This pre-timed module can minimize excessive queues caused by lane blockages and mitigate the resulting congestion, provided that all contributing factors to queue formation are accounted for in the design of signal plans. However, traffic patterns suited to a pre-timed signal plan are unlikely to remain consistent during peak hours, especially on arterials with highly asymmetric traffic volumes and near-saturated conditions at critical intersections. In such a case, even a minor queue blockage at one intersection can rapidly trigger standing queue waves upstream and across neighboring links, disrupting the time-varying arrival rates at downstream intersections within the impacted boundaries. The adverse impacts caused by these mutual lane blockages—including unused left-turn green time, accelerated through-queue formation, and reduced capacity of through lanes—are illustrated in Figure 5.1.

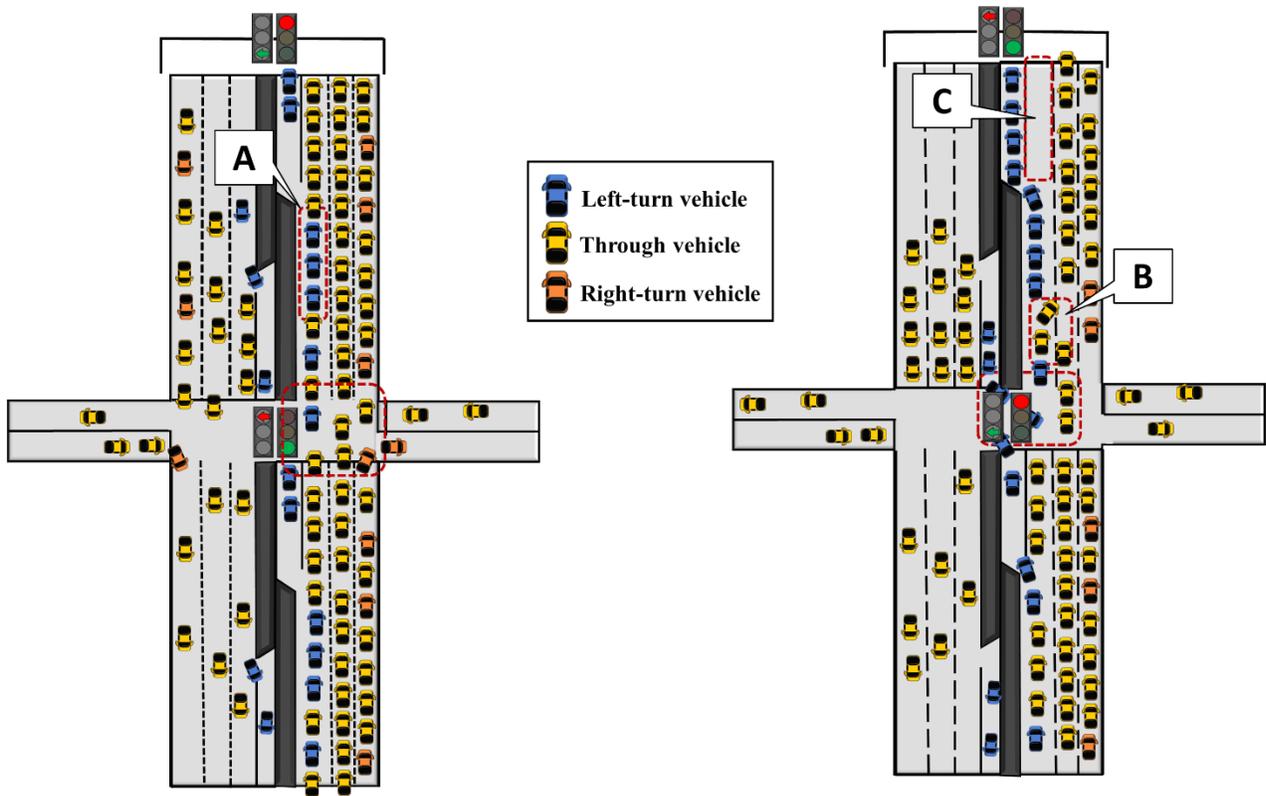


Figure 5.1 Adverse impacts of blockages: A) wasted green time; B) increased queue formation speed; C) Through-lane capacity reduction

To address these challenges, this chapter introduces a responsive signal control module that leverages real-time data collected from detectors deployed along the arterial. Serving as a critical component of an adaptive arterial signal control system, this module advances the traditional time-of-day control by offering a timely response to unstable traffic states and emerging congestion patterns resulting from turning bay overflows and/or mutual queue blockages.

As such, the responsive control module is designed with three key functions:

- **Detecting traffic states in real-time and estimating their ensuing evolution patterns:**

With the sensor data from each lane, the function will be used to robustly estimate the time-varying queue evolution of both through and left-turn movements, identifying any queue

formation over all intersections within the control boundaries and the likelihood of incurring mutual queue blockages.

- **Classifying detected queue types and blockages into pre-characterized congestion levels:** A customized heuristic-based classification method maps observed queue distribution/traffic states across intersections to pre-defined levels of congestion.
- **Selecting and executing control strategies tailored to each congestion level:** Based on the identified congestion level, this control function is responsible for selecting and executing the set of corresponding signal plans customized to alleviate current bottlenecks and prevent further congestion propagation.

5.2 Overview of the Responsive Real-Time Arterial Signal Control Module

The responsive module operates as one of the signal plans for the arterial-wide real-time control, focusing on managing traffic flows across multiple congested intersections along an arterial. The module's primary operations include executing sequential control tasks—from identifying the traffic states, and characterizing congestion levels, to selecting as well as activating the best responsive control strategy—to ensure its real-time adaptability to evolving traffic conditions. The operation flows of the developed module are organized into the following steps:

- **Step 1: Data Collection and Time-Varying Queue Length Estimation**

At each time step t , the module will first collect real-time traffic volume data from all detectors across the intersections, including vehicle counts for both through and left-turn movements. Using these counts, the module will then compute the time-varying queue length for each movement and then perform the update in real-time as traffic conditions evolve throughout the cycle. At the end of each signal cycle $C_{i,m}$, where $C_{i,m}$ refers to m^{th} cycle at intersection i (i.e.,

the current cycle) within the control period, and the subsequent cycle is denoted as $C_{i,m+1}$, the module will classify each intersection's blockage type as follows:

- a. **Blockage Type-N:** No discernible queue blockage.
- b. **Blockage Type-1:** One blockage per cycle, where the through queue blocked left-turn vehicles.
- c. **Blockage Type-2:** One blockage per cycle, where the left-turn queue blocked through vehicles.
- d. **Blockage Type-3:** Two blockages per cycle, where the through queue first blocked left-turn vehicles, and then incoming through vehicles in turn are impeded by those blocked left-turn queues.
- e. **Blockage Type-4:** Two blockages per cycle, where the left-turn queue exceeding the bay first blocked the through vehicles, causing the formation of through queues which subsequently blocked the arriving left-turn vehicles.

- **Step 2: Congestion Level Assessment**

When the blockage is detected, the module will apply pre-defined sets of IF-THEN rules to evaluate the severity of the queues, considering their temporal and spatial distribution along the arterial. At the end of the cycle at each intersection, the system will classify the evaluation results into one of the following four levels:

- a. **Congestion Level-N:** No discernible queue blockage at any intersection.
- b. **Congestion Level-1:** Localized congestion, with blockages occurring independently at some intersections.

- c. **Congestion Level-2:** Blockages occur at most intersections or propagate interdependently over consecutive intersections during the same cycle, or a critical intersection suffers sustained mutual blockages over consecutive cycles.
- d. **Congestion Level-3:** The most severe congestion level, where Level-2 congestion persists consistently across multiple intersections and over multiple cycles.

- **Step 3: Responsive Plan Selection**

Based on the results of the congestion assessment, the module will select the optimal control strategy to respond to the detected congestion level at the start of the subsequent cycle $C_{i,m+1}$.

The four responsive signal plans, each customized to contend with the four classified congestion levels, are defined as follows:

- a. **Response Plan-N:** Remain the current signal plan.
- b. **Response Plan-1:** Green extension for either through or left-turn movements, or both.
- c. **Response Plan-2:** Re-optimize phase sequences and offsets.
- d. **Response Plan-3:** Adjust the cycle lengths and re-optimize phase sequences and offsets based on the updated green splits.

- **Step 4: Execute the Selected Signal Plan and Update the Resulting Traffic Conditions**

Upon selecting a responsive signal plan, the module will compute the new signal settings for each intersection during the subsequent cycle ($C_{i,m+1}$). These updated signal settings are then implemented in the following cycle ($C_{i,m+2}$). It is important to note that for Response Plan-2 and Response Plan-3, which involve arterial-wide adjustments, the activation of the updated signal plan at each intersection must account for their respective offsets to ensure proper coordination across the system.

Figure 5.2 illustrates the framework of the developed module, which consists of three principal models. Note that each model is designed to perform its designated role in the process, where traffic monitoring and queue length computation are performed in Model 1; congestion level identification is to be executed in Model 2; and dynamic signal control strategies are to be selected and activated in Model 3. These models operate sequentially and continuously to update the arterial's signal plan in response to the real-time detected and classified traffic conditions. The following sections detail the formulations and underlying logic of each model, concluding the chapter with comprehensive numerical experiments to demonstrate the module's performance and effectiveness.

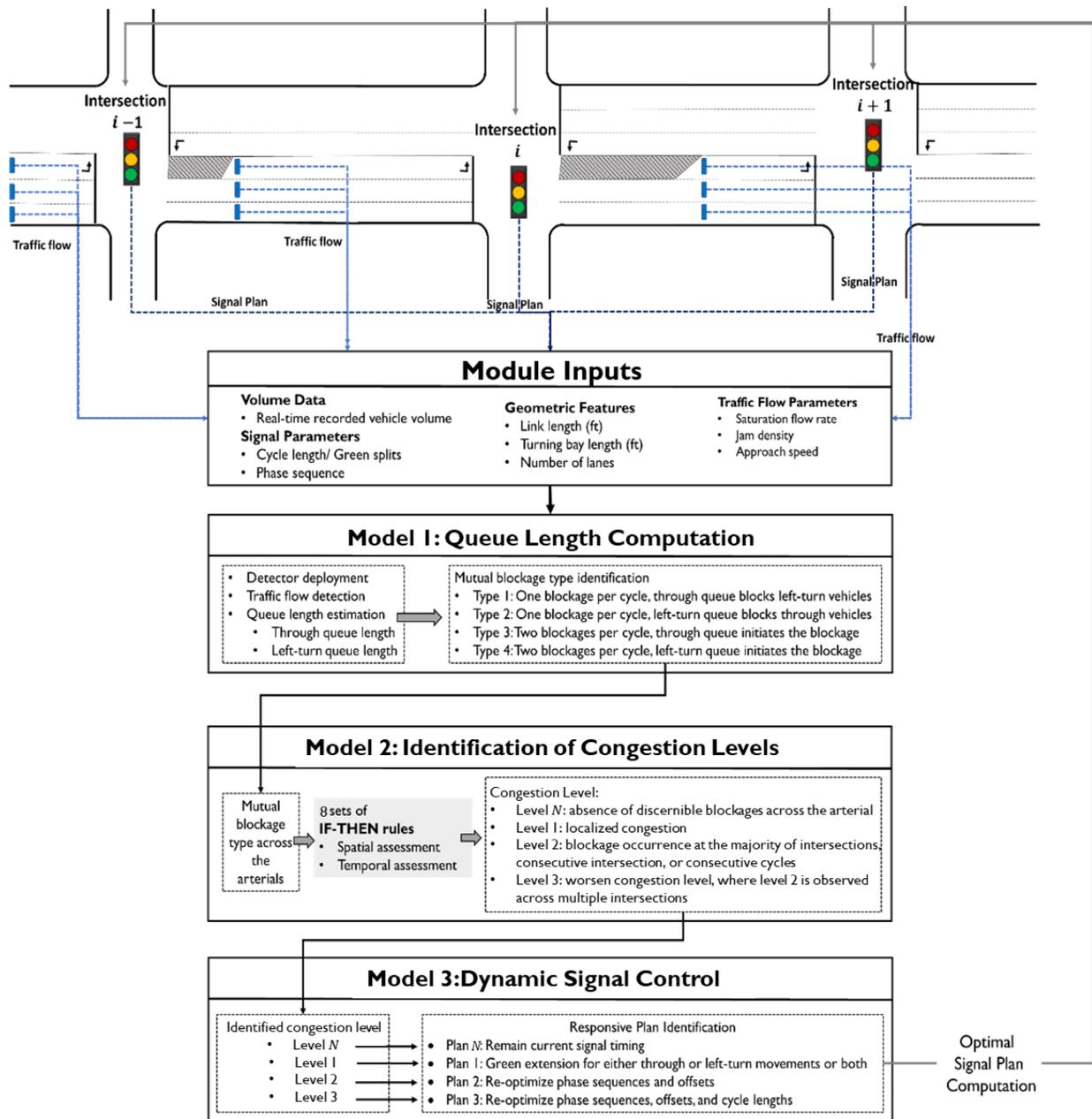


Figure 5.2 Framework of the developed module

5.3 Module components

This section provides a detailed description of the three core models that constitute the responsive control module. Each model is presented in a dedicated subsection, with emphasis on its unique

functions, mathematical formulations, and the logic underpinning its integration within the overall control structure. The core function of each model is as follows:

- **Model 1: Queue length computation**
- **Model 2: Identification of the congestion levels**
- **Model 3: Dynamic signal control**

5.3.1 Model 1: Queue Length Computation

This model aims to identify the type of queue blockage using the detected inflow rate from movement n at time step t , denoted as \hat{x}_t^n . The location of detectors is installed at the end of turning bays at every intersection approach to ensure the timely capture of the blockage. It is important to note that a zero-inflow rate can result from either a lack of demand (i.e., no vehicles arriving) or a queue that has extended to the detector, preventing further vehicle movement. In either case, this information feeds into the estimation of time-varying queue lengths for both through and left-turn movements. The resulting queue length estimates are then used to determine the type of blockage that has occurred, as described by the following equation:

$$Q_i^n(t) = Q_i^n(t-1) + \hat{x}_t^n - \left(\tau_i^n(t) \times \Delta t \times \frac{1}{h} \right) \quad n \in \{T, L\} \quad (5-1)$$

where $Q_i^T(t)$ and $Q_i^L(t)$ denote the queue length in the through and left-turn movements at time step t , respectively; $\tau_i^n(t)$ is a binary variable to identify whether movement n has a green phase at a given time step t ; Δt is the length of each time step; h is the saturation headway. It is important to note that the queue length refers only to vehicles that have come to a complete stop, and does not include vehicles that are slowing down to join the queue.

With Equation 5-1, one can characterize possible queue blockages into the four most likely types as shown in Table 5.1. Each cycle is divided into M time slices in the queue computation,

and the queue size in the turning bay at intersection i is represented with $L_i^* \cdot k$, where L_i^* denotes the length of the left-turn bay in feet and k is the jam density. Note that the comparison of left-turn and through queue lengths pertains specifically to the adjacent lanes.

Table 5.1 Mathematical expression of mutual blockage types

Blockage type-N	$Q_i^T(t) < L_i^* \cdot k, \forall t$	(5-2)
	$Q_i^L(t) < L_i^* \cdot k, \forall t$	(5-3)
	$0 \leq t \leq M$	(5-4)
Blockage type-1	$Q_i^T(m_i^T) \geq L_i^* \cdot k$	(5-5)
	$Q_i^L(t) < L_i^* \cdot k, \forall t$	(5-6)
	$0 < m_i^T < M$	(5-7)
	$0 \leq t \leq M$	(5-8)
Blockage type-2	$Q_i^T(t) < L_i^* \cdot k, \forall t$	(5-9)
	$Q_i^L(m_i^L) \geq L_i^* \cdot k$	(5-10)
	$0 < m_i^L < M$	(5-11)
	$0 \leq t \leq M$	(5-12)
Blockage type-3	$Q_i^T(m_i^T) \geq L_i^* \cdot k$	(5-13)
	$Q_i^L(m_i^L) \geq L_i^* \cdot k$	(5-14)
	$0 < m_i^T < m_i^L < M$	(5-15)
Blockage type-4	$Q_i^T(m_i^T) \geq L_i^* \cdot k$	(5-16)
	$Q_i^L(m_i^L) \geq L_i^* \cdot k$	(5-17)
	$0 < m_i^L < m_i^T < M$	(5-18)

Equations 5-2 and 5-3 illustrate a scenario in which neither the through queue nor the left-turn queue exceeds the length of the corresponding turning bay throughout the entire cycle. This condition indicates the absence of any queue-induced blockage and is thus classified as Blockage type-N. In contrast, Blockage type-1, as shown in Equations 5-5 through 5-8, is to reflect the queue pattern in which the through-queue at time m_i^T has exceeded the length of the turning bay, but no spillover has occurred in the left-turn queue within the cycle. By the same token, Blockage type-2 is defined as the traffic state where the left-turn queue, rather than the through-queue, has spilled over the turning bay, as shown in Equations 5-9 through 5-12.

Note that Blockage type-3 and Blockage type-4 are to characterize mutual blockages caused by different sequences of queue formation within a single cycle, where the former indicates that the mutual blockage is caused first by the through queues, then followed by the turning vehicle queues (described in Equations 5-13 through 5-15). For the latter type, Blockage type-4, captures the reverse sequence, where the left-turn queue first exceeds the bay and initiates a blockage that is later compounded by the through queue, as formulated in Equations 5-16 through 5-18.

5.3.2 Model 2: Identification of the Congestion Level

This model is designed to evaluate the congestion level along the arterial using a customized heuristic that accounts for the estimated temporal and spatial distribution of queue blockages. It builds on the output from Model 1 by utilizing the detected blockage types to construct an augmented matrix, denoted as Λ , which indicates the presence or absence of queue blockages at each intersection over time. Specifically, the matrix Λ is constructed by concatenating the monitoring results from a total of X signal cycles, including the current cycle (m^{th} cycle) and the preceding $X - 1$ cycles across I intersections. This yields a historical record of congestion conditions ranging from $C_{i,m-X+1}$ to $C_{i,m}$, where $i = 1, \dots, I$, and where Intersection I corresponds to the most upstream location. Note that the value of X should be selected in accordance with the total number of intersections: longer arterials with a greater number of intersections may require a larger X to sufficiently capture the spatial evolution and propagation of congestion patterns. The structure of matrix Λ is formally expressed as follows:

$$\Lambda = \begin{bmatrix} S_{1,m-X+1} & \dots & S_{I,m-X+1} \\ \vdots & & \vdots \\ S_{1,c} & \ddots & S_{I,c} \\ \vdots & & \vdots \\ S_{1,m} & \dots & S_{I,m} \end{bmatrix}$$

where $s_{i,c} \in \{0, 1\}$ denotes the presence of a queue blockage at intersection i during the c^{th} cycle within the control period: $s_{i,c} = 1$ if a blockage was present, and $s_{i,c} = 0$ if there is no blockage.

By analyzing the augmented matrix Λ , one can characterize the congestion levels with a series of rules involving spatial, temporal, and spatiotemporal pattern assessments, as detailed below:

- **Rule 1:** If no intersection along the arterial experienced blockages during the evaluation period, as expressed in Equation 5-19, it indicates the absence of discernible congestion in the current traffic state, then classified as *Congestion Level-N*.

$$\sum_{i=1}^I \sum_{c=m-X+1}^m s_{i,c} = 0 \quad (5-19)$$

- **Rule 2:** If, in any cycle within the evaluation period, fewer than half of the intersections have encountered mutual blockages, as formulated in Equation 5-20, it is indicative of localized congestion, which should be classified as *Congestion Level-1*.

$$0 < \sum_{i=1}^I s_{i,c} < \frac{1}{2}I \quad \forall c \in [m - X + 1, m] \quad (5-20)$$

- **Rule 3:** If, in any cycle within the evaluation period, blockages are detected at any of two consecutive intersections along the arterial, as formulated in Equation 5-21, it is indicative of a high congestion level, which should be classified as *Congestion Level-2*.

$$s_{i,c} + s_{i+1,c} = 2 \quad \begin{array}{l} \forall i \in [1, I - 1]; \\ \forall c \in [m - X + 1, m] \end{array} \quad (5-21)$$

- **Rule 4:** If, in any cycle within the evaluation period, more than half of the intersections experienced mutual blockages, as formulated in Equation 5-22, this also indicates a significant congestion pattern, warranting classification as *Congestion Level-2*.

$$\sum_{i=1}^I s_{i,c} \geq \frac{1}{2}I \quad \forall c \in [m - X + 1, m] \quad (5-22)$$

- **Rule 5:** If at least one intersection experiences the queue blockages over the previous consecutive cycles (for a given evaluation period of X cycles), it implies that the response plan for localized congestion (i.e., *Congestion Level-1*) is not effective, and a new signal plan is needed to cope with the sustained congestion. The traffic state is then elevated to *Congestion Level-2*. This rule is expressed mathematically as:

$$\sum_{c=m-X+1}^m s_{i,c} = X \quad \forall i \in [1, I] \quad (5-23)$$

- **Rule 6:** If a local queue blockage has been observed to propagate progressively to upstream intersections, it indicates the need to revise the current signal plan and to reclassify the new congestion pattern as *Congestion Level-2*. The rule can be expressed as follows:

$$s_{i,c} + s_{i+1,c+1} = 2 \quad \forall i \in [1, I - 1]; \quad (5-24)$$

$$\forall c \in [m - X + 1, m]$$

- **Rule 7:** If *Congestion Level-1* is observed consistently over prior consecutive X cycles, it indicates that the traffic state has degraded to *Congestion Level-2*.
- **Rule 8:** If *Congestion Level-2* was persistently present over prior consecutive X cycles, one can conclude that the traffic state has further deteriorated and should be classified as *Congestion Level-3*.

Note that the results of real-time monitoring and classification of congestion levels with the above rules offer the developed module the most timely and reliable information to revise the

signal plan so as to cope with the time-varying congestion patterns and the resulting traffic states in a timely manner.

5.3.3 Model 3: Dynamic Responsive Signal Control

This model is responsible for selecting and executing signal control strategies that are specifically tailored to mitigate the congestion level identified in Model 2 based on the spatial-temporal distribution of detected blockage types. For instance, if a localized congestion is detected, the control model will apply only minor adjustments to signal timings at the affected intersections. Conversely, in response to severe and widespread congestion, a new set of coordinated signal plans must be generated and implemented across the entire arterial to prevent further deterioration of traffic conditions and to alleviate gridlock.

Accordingly, in addition to Response Plan-N, which maintains the current signal timing plan, Model 3 includes three additional responsive control strategies, each designed to address a specific congestion level. These strategies incorporate varying levels of signal plan adjustments, ranging from localized refinements to full arterial-level updates. It is important to note that all three responsive control strategies build upon the base signal plans developed by the time-of-day control methodology introduced in Chapter 4. This foundation ensures consistency in the operational objective while enhancing flexibility and responsiveness under dynamic traffic conditions.

5.3.3.1 Response plan 1: green extension for the needed movement, through and/or left-turn traffic flows

Response Plan 1 is designed for local traffic congestion (i.e., Congestion Level-1) at critical intersections, where local traffic blockage due to queue formation needs to be responded to in time

to prevent it from further spreading to neighboring intersections. The implementation of Response Plan 1 comprises the following steps:

- **Step 1: Minimum Green Time Check**

Before extending green time, it is essential to ensure that the green time from the side-street phase has not been fully utilized. Otherwise, Response Plan 1 may not be the optimal choice, and Response Plan 2 ought to be the candidate plan.

- **Step 2: Blockage Time Computation**

The blockage impact commences when queues persist and spill over or propagate to upstream segments. Hence, it is essential to estimate the blockage time based on the difference between the end of the red phase and the onset of the blockage.

- **Step 3: Green Splits Adjustments**

With the computed blockage time as the preferred green extension and the minimum green time required for the side street, the developed module can adopt the revised green splits to the detected local congestion as follows:

$$G_i^* = g_i - \tilde{g}_i \quad (5-25)$$

$$\bar{G}_i^n = (\tilde{m}_i^n - m_i^n) \times \Delta t \quad (5-26)$$

$$G_i = \begin{cases} \min(G_i^*, \bar{G}_i^n), & G_i^* > 0 \\ \text{Computed with Responsive Plan 2,} & G_i^* = 0 \end{cases} \quad (5-27)$$

Where G_i^* , \bar{G}_i^n , and G_i denote the allowable green, target green, and finally adopted green extensions of the critical intersection i , respectively; g_i and \tilde{g}_i represent the side-street green duration and the minimum side-street green duration, respectively; g_i^n is the green

duration of movement n ; and \tilde{m}_i^n is the time step at the onset of the green light for the overflow stream n .

5.3.3.2 *Response plan 2: Re-optimize phase sequences and offsets for intersections within the congestion impact boundaries*

Figure 5.3 illustrates the activation and procedural framework of Response Plan 2, which is designed to re-optimize signal offsets and phase sequences in response to mutual queue blockages identified along the arterial. The process begins with identifying the movement group that initiates the blockage, followed by determining the onset time and the duration for the queue to propagate to the turning bay. If multiple blockage events occur within a single signal cycle, the model recalculates the onset time of the subsequent blockage to account for compounding congestion effects.

To support this re-optimization, the system quantifies the adverse impacts of mutual blockages—such as accelerated through-queue formation, reduced lane capacity, and the accumulation of blocked left-turn vehicles and residual queues. These negative impacts, previously introduced in Section 4.3.3, *Traffic Dynamics Under Mutual Blockage*, are incorporated into a restructured objective function that better reflects the real-time traffic inefficiencies induced by queue spillovers.

The actual optimization procedure for generating the revised offsets and phase sequences follows the methodology established in the time-of-day control module (also detailed in Chapter 4). Specifically, the control objective remains the core objective of minimizing delays in the major commuting direction while maintaining progression in the opposite direction. By embedding real-time blockage information into the optimization logic, Response Plan 2 provides a responsive and targeted adjustment to alleviate evolving congestion along the arterial.

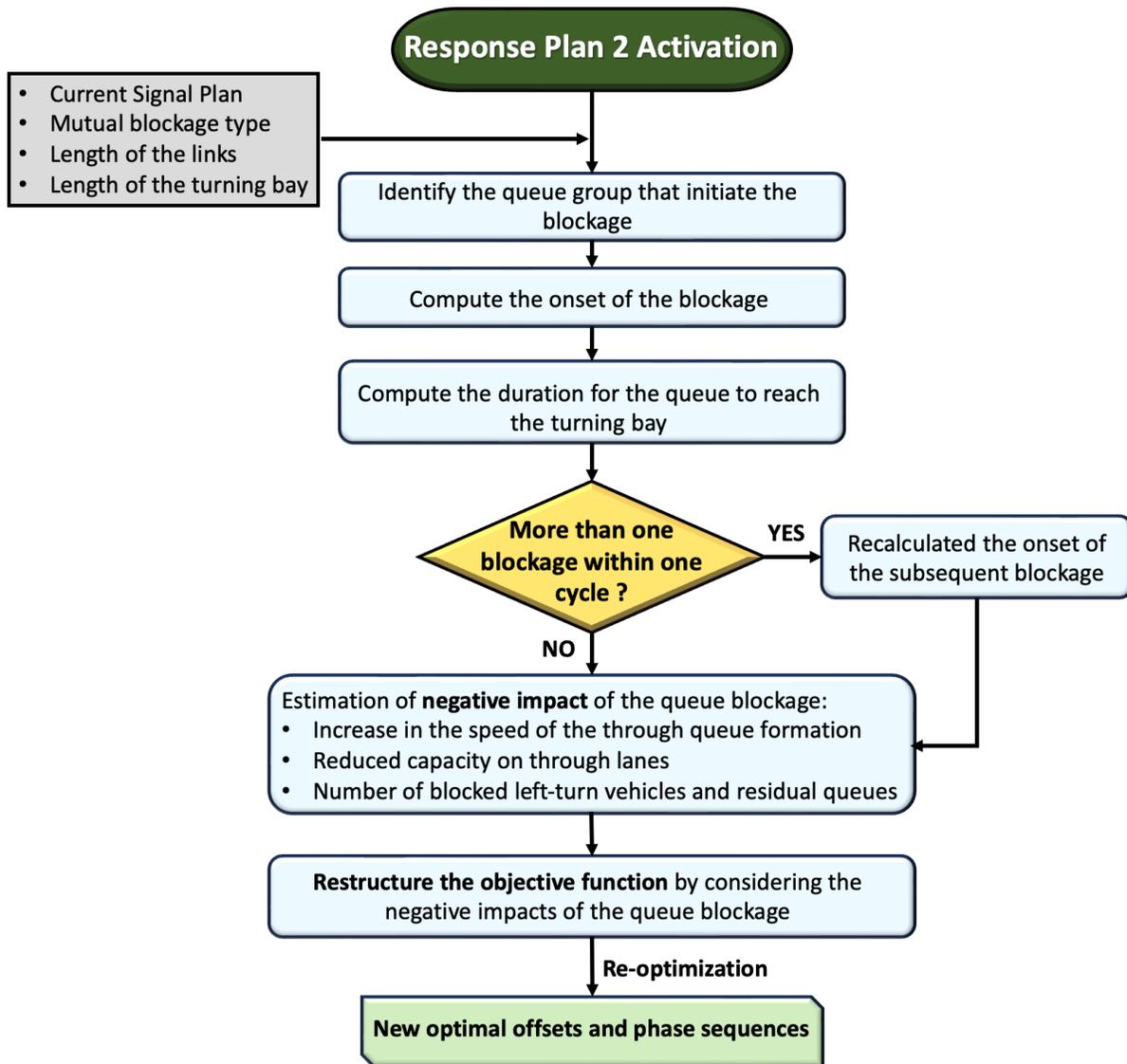


Figure 5.3 Flowchart of the structure of Response Plan 2

5.3.3.3 Response plan 3: Adjust cycle lengths and green splits, then re-optimize phase sequences and offsets within the congestion impact boundaries

Response Plan 3 is to be activated for the most severe congestion level (Level-3) over the arterial, where the entire signal plan (i.e., cycle length, green splits, offsets, and phase design) needs to be re-optimized and re-implemented in a timely manner. To effectively identify the optimal signal plan, a stepwise approach is developed as follows:

- **Step 1: Adjust the Cycle Length**

The first action is to decrease the cycle length progressively by a small interval (e.g., 5 seconds). Since the reduction is decremental, it is less likely to cause sudden disruptions to traffic flows. The rationale lies in that under severe multi-blockage congestion due to either a short turning bay or link length, reduced cycle length, followed by re-optimization of phase sequence and offsets, is one of the most effective ways to prevent excessive queue accumulation in real time. Theoretically, one could include the cycle length in the signal re-optimization process. However, practically, to do so will inevitably make the control formulations rather complex and cumbersome, and render the solution process intractable as well as inefficient for use in real-time responsive operations.

Note that the adjusted cycle length should not be shorter than the pre-defined threshold, as formulated in Equation 5-28, because a shorter cycle length inherently results in a higher proportion of lost time, and thus a lower intersection capacity utilization.

$$C_{min} = \frac{1}{\epsilon} \times L \quad (5-28)$$

Where ϵ denotes a pre-defined percentage, and L denotes the total lost time per cycle (sec/cycle), which can be calculated based on the start-up lost time and clearance lost time, as outlined in Roger et al., 2004.

- **Step 2: Adjust Green Splits**

Following the adjustment of the cycle length, the green splits must be adjusted accordingly to reflect the reduced total green time available for each phase. In this study, green splits are allocated based on entry flows to each intersection approach—i.e., the discharging flow rates from upstream intersections—which can more accurately reflect the time-varying traffic

demands under congested conditions. Note that the green splits are calculated using the established methodology outlined in Roger et al. (2004), with the effective green time for each phase computed using the formula:

$$g_i = g_{TOT} \cdot \left(\frac{V_{ci}}{V_c} \right) \quad (5-29)$$

where g_i is the effective green time for Phase i (sec); g_{TOT} is the total effective green time for the cycle (sec); V_{ci} is the critical lane volume for Phase i (veh/h); V_c is the sum of the critical-lane volumes (veh/h).

- **Step 3: Re-optimize Offsets and Phase Sequence**

Once the cycle length and green splits have been revised to reflect the updated congestion level, the final step involves re-optimizing the signal offsets and phase sequences. This re-optimization once again adopts the same methodology introduced in the time-of-day control module (Chapter 4), where the primary control objective is to minimize delay along the major commuting direction while maintaining progression for the opposing flow. The optimization framework is also adjusted to account for the adverse effects of mutual blockages, as detailed in Section 4.3.3, ensuring the resulting signal plans are both responsive to evolving congestion conditions and consistent with the underlying traffic dynamics of the arterial.

5.4 Numerical Experiments

The primary objectives of these numerical experiments are first to verify the developed module's effectiveness in detecting various types of traffic queue blockage in a timely manner, and then to assess its performance in real-time generation as well as execution of responsive signal plans to minimize the resulting queue propagation and delay.

Figure 5.4 shows the selected arterial segment of five intersections on the MD 355 Rockville Pike in Bethesda for model evaluation, including its geometric features, land configuration, and distributions of evening peak-hour volumes. Note that this segment connects the residential areas and the business district around the District of Columbia, thus exhibiting a dominating northbound traffic volume during the evening peak hours. To increase the congestion level during the evening peak hours, an increase of 2.55% has been added to each of the four 15-minute field volumes, from the Internet Traffic Monitoring System (ITMS), as shown in Figure 5.4 (b).

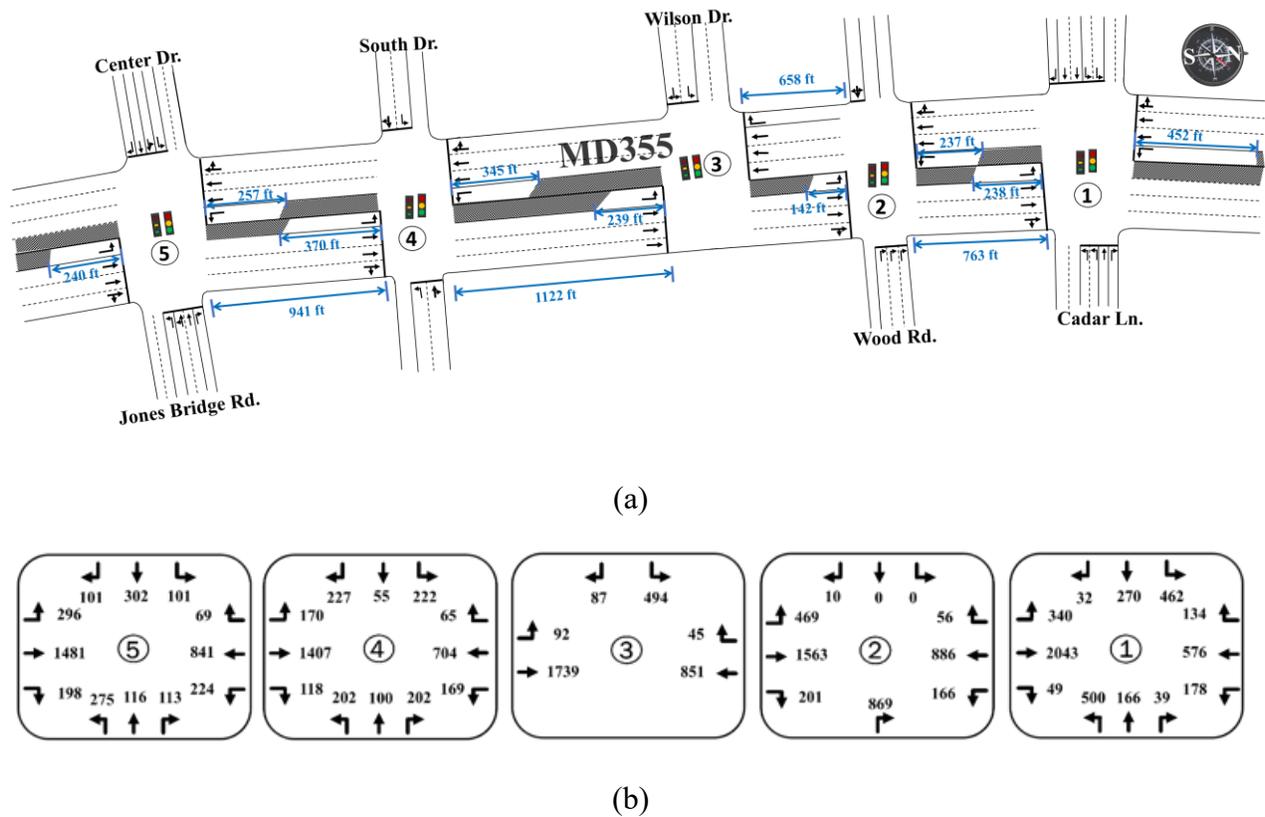


Figure 5.4 (a) Geometric Features, and (b) distribution of peak hour volumes over the candidate arterial segment

The performance of the developed responsive real-time signal control module was evaluated against the time-of-day signal control module introduced in Chapter 4. This module is selected as the baseline due to its demonstrated superiority over several established signal optimization tools, including MULTIBAND, TRANSYT-7F, and Synchro. A detailed comparison of performance metrics across these models is provided in Section 4.4.2 (Model Performance Evaluation), with summary results presented in Table 4.3. Notably, the developed time-of-day control module achieved a 12.19% reduction in northbound (i.e., commuting direction) arterial delay per vehicle per hour relative to TRANSYT-7F, along with an 8.39% improvement over MULTIBAND and a 1.85% improvement compared to Synchro. In terms of stop frequency, the average number of stops per northbound through vehicle per hour was 12.34% lower than under TRANSYT-7F and Synchro, and 15.09% lower than under MULTIBAND. These findings justify the selection of the time-of-day module reported in Chapter 4 as the benchmark for evaluating the benefits of transitioning from a fixed time-of-day operation to the real-time responsive signal control.

While a comparison with other real-time control models would provide additional insights, it was not feasible in this study due to two key reasons. First, the operational details of existing real-time control systems are proprietary and not publicly available, which limits the ability to fairly assess their effectiveness under comparable conditions. Secondly, the design logic of most real-time control systems is grounded in the assumption of contending with moderately congested traffic conditions without overflows or lane blockages. As a result, their performance under congested scenarios involving mutual blockages would likely be suboptimal. Comparing the developed responsive real-time signal control module to such models would therefore not provide a fair or meaningful evaluation. Instead, adopting an offline module with well-recognized

performance as the benchmark provides a more robust assessment of the benefits of incorporating real-time responsive features in signal controls for congested arterials.

5.4.1 Performance Evaluation for the Detection and Timely Response to Mutual Blockages

Table 5.2 presents the simulation results regarding the detection of mutual blockage scenarios at each intersection over 24 cycles. As shown in Table 5.2, the developed responsive real-time signal control module can, as expected, detect the onset of mutual blockage in time for the signal to take any needed responsive actions. Some examples of such detection exercises revealed from the simulation results are listed below:

- Due to the continuous increase in traffic volume, Blockage Type-1 was detected by the developed module to persist from Cycles 13 to 15 at Intersection 2, and the resulting traffic state was classified as Congestion Level-2. In light of this, Response Plan 2 was triggered, and the optimal signal plan was produced in Cycle 16 and then implemented in Cycle 17. In contrast, the blockages under the time-of-day control module without timely detection and response have persisted over more cycles, i.e., further persist to Cycle 17.
- During Cycle 18, more than half of the intersections (Intersections 1-3) experience mutual blockages and evolve to Congestion Level-2. This triggered the need to implement Response Plan 2. Hence, a new signal plan, generated in Cycle 19 and executed in Cycle 20 by the developed responsive module, has successfully eliminated blockages at all intersections (see blue cells in Table 5.2).
- The performance comparison between the developed responsive real-time control module and the time-of-day control module demonstrates significant improvements, particularly at Intersection 3 during Cycles 21-23. Under the offline signal control, consecutive blockages were unavoidable due to spillback from Intersection 2. In contrast, the responsive real-time

control module, with its timely intervention at Cycle 20, can effectively mitigate queue spillback from Intersection 2, allowing the queue to dissipate before reaching Intersection 3 (see red boxes in Table 5.2).

Table 5.2 Detection of mutual blockages at each intersection

Cycle	Intersection 1		Intersection 2		Intersection 3		Intersection 4	
	Responsive Control Module	Time-of-day Module						
1-12	-	-	-	-	-	-	-	-
13	-	-	S_1	S_1	-	-	-	-
14	S_1	S_1	S_1	S_1	-	-	-	-
15	-	-	S_1	S_1	S_1	S_1	-	-
16	-	-	S_1	S_1	-	-	-	-
17	-	-	-	S_3	S_1	S_1	-	-
18	S_1	S_1	S_1	S_1	-	-	S_1	S_1
19	S_1	S_1	S_1	S_1	S_1	S_1	S_3	S_3
20	-	S_2	-	S_1	-	-	-	S_1
21	-	S_3	-	S_1	-	S_1	-	S_1
22	-	-	S_1	S_3	-	S_1	-	S_1
23	-	S_1	-	S_1	-	S_1	S_1	S_1
24	-	S_1	S_1	S_1	S_1	S_2	-	S_1

Note: S_i denotes the observed Blockage type- i , the cell without any letter represents no mutual blockage observed

5.4.2 Performance Evaluation at the Critical Intersection

The performance evaluation of the developed responsive control module in tackling mutual blockages between through and left-turn vehicles is focused on the arterial's critical intersections. Using Intersection 2—characterized by its short link length—as a case study, the analysis has examined its susceptibility to mutual blockages under both the time-of-day control module and the developed responsive real-time control module. The evaluation is conducted using the following two measures of effectiveness (MOEs):

- **Throughput per cycle at each intersection:** The number of vehicles passing through the stop line at the downstream intersection during each signal cycle.

- **Queue delay:** Per vehicle's total time in the queue.

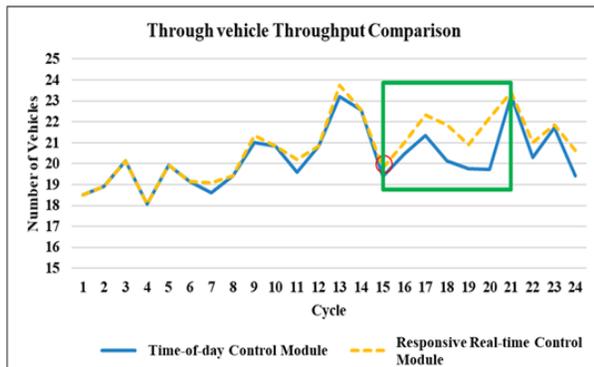
5.4.2.1 Comparison results of throughput over through and left-turn lanes

Throughput is selected as a key indicator in the ensuing comparisons to assess the impact of blockages, as mutual blockages—whether left-turn vehicles are obstructed by through queues or vice versa—will inevitably reduce the number of vehicles passing through the intersection. Figure 5.5 illustrates the throughput differences between the developed responsive control module and the time-of-day control module. The key findings from the graphically revealed results are summarized below:

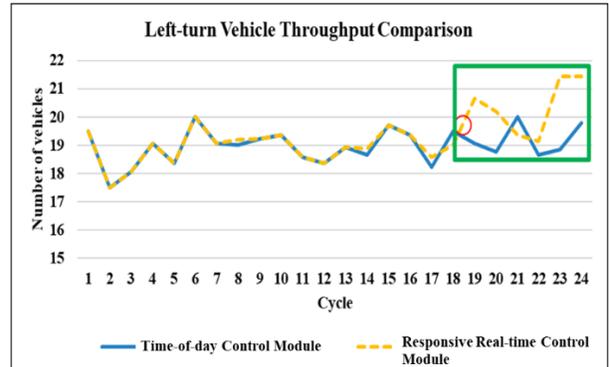
- As shown in Figure 5.5, at the outset of the low traffic volume period, both the developed responsive control module and the time-of-day operations exhibit comparable performance. However, as traffic volume increases, the emergence of mutual blockage between through and left-turn vehicles becomes apparent (see Cycle 13 in Table 5.2). It is at this juncture that the developed responsive control module's capability to manage such scenarios effectively becomes evident, as reflected in its higher throughput (see green box in Figure 5.5).
- Figure 5.5(a) illustrates that the developed responsive control module, responding to a lane blockage starting in Cycle 13, consistently achieves higher throughput in the through lanes from Cycle 15 onward compared to the time-of-day control module. This improvement stems from the responsive control module's ability to generate and implement a new signal plan within one cycle. Without timely adjustments, as demonstrated by the time-of-day control module, the blockage impact may persist from Cycle 13 to Cycle 21 and exacerbate the congestion level.

- Figure 5.5(b) shows that the developed responsive control module consistently achieves higher throughput for the left-turn lane starting from Cycle 19, in response to the blockage detected in Cycle 17. This result underscores that the developed module can effectively address the blockage in both the through and left-turn lanes, thereby minimizing the excessive delays for traffic flows of different movements.

Figure 5.5(a) and Figure 5.5(b) reveal that Intersection 2 experiences prolonged blockage impacts on the through lane beginning in Cycle 13, and on the left-turn lane starting from Cycle 17. The developed responsive control module has effectively mitigated these mutual blockages by dynamically adjusting the signal plan to accommodate the varying traffic conditions. Notably, during the periods of volume surge (from Cycle 13 to Cycle 24), the developed responsive module consistently achieves higher throughput compared to the time-of-day control module. This confirms the developed responsive control module’s ability to minimize spillover impacts and queue propagation, ensuring better traffic progression even under highly congested conditions.



(a)



(b)

Figure 5.5 Comparison of queue delay on the through and left-turn lanes

Queue delay is selected as a crucial indicator to assess the impact of mutual blockages. During the period of mutual queue blockages, the allocated green phases may not be fully utilized,

and some blocked vehicles may have to wait for the green time in subsequent cycles, and thus endure excessive queue delays.

Figure 5.6 and Figure 5.7 illustrate performance comparisons between the developed responsive real-time control module and the time-of-day control module with respect to the queue delay and queue length, respectively. Some key insights into their respective performances are summarized below:

- At the beginning of the simulated period (see Figure 5.6), both modules demonstrate comparable average queue delays. However, the differences in such MOE between the two modules in the later cycles become pronounced. More specifically, the responsive real-time control module begins to outperform its counterpart in reducing queue delays for the through movement from Cycle 15 and for the left-turn traffic stream from Cycle 19. This improvement can be attributed to the module's ability to effectively identify mutual queue blockages early (from Cycle 15 and Cycle 19) and respond promptly with a new signal plan, thus significantly shortening the blockage duration for high-volume traffic flows.
- As traffic volumes increase, the queue delays under the responsive real-time control module progressively become lower than those observed under the time-of-day control. For example, as highlighted in the purple box in Figure 5.6, from Cycle 17 to Cycle 24, the responsive module consistently achieves reduced queue delays. This result underscores the module's ability to mitigate the effects of mutual blockages, prevent the propagation of queues, and maintain smoother traffic progression through adaptively updated signal plans.
- As shown in Figure 5.7, both modules start with comparable queue lengths under lower traffic volumes (i.e., from Cycle 1 to Cycle 12). However, as the volume increases (i.e., from Cycle 13 to Cycle 24), the performance under the time-of-day control module

degrades progressively. This is evident from the sudden spike in the queue length for the benchmark module at Cycle 21, which has its queue spilled back to the upstream intersection (i.e., Intersection 3). In the worst case, the queue spillback extends to two upstream intersections (i.e., Intersection 4), as indicated by the red circles in Figure 5.6. The responsive real-time control module with timely queue detection and signal revision is able to prevent the formation of such severe congestion and ensure minimal delay for traffic flows during peak hours.

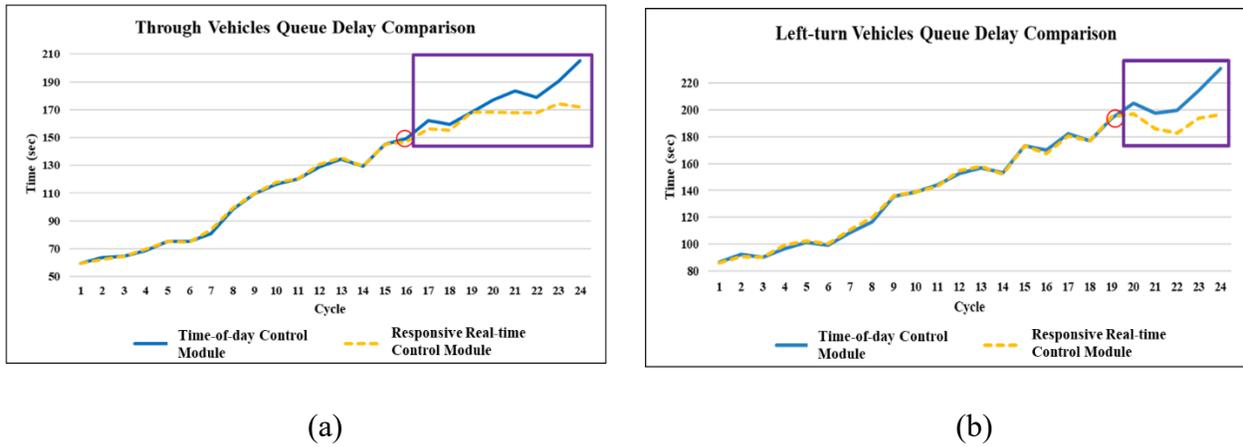


Figure 5.6 Queue delay comparison of (a) through vehicles; (b) left-turn vehicles

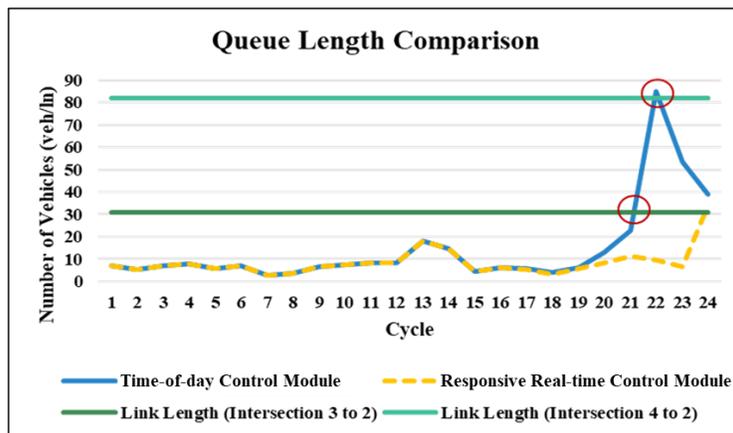


Figure 5.7 Through queue length comparison at Intersection 2

5.4.3 *Module Performance Evaluation*

Given its capability to detect mutual blockages and dynamically adjust signal plans, the developed real-time control module is evaluated against the time-of-day control module. As shown in Table 5.3, the comparison includes key performance metrics such as the average delay per vehicle per hour and number of stops per vehicle per hour for through vehicles over the arterial, as well as the average delay per vehicle per hour for through and left-turn movements from all upstream streams on individual links.

The simulation results in Table 5.3 show that the expected performance improvements of the responsive control module reflect the benefits of having a responsive function to tackle the traffic surge in a timely manner. Key findings from the performance evaluation are summarized below:

- The developed responsive real-time control module significantly reduced the average delay per vehicle per hour along the northbound arterial by 7.39%, decreasing from 234.32 seconds to 217.01 seconds. Notably, there was no significant difference in the average delay per vehicle per hour along the southbound arterial, demonstrating that the system can facilitate traffic flows in the main commuting direction of high traffic volume without impacting the non-commuting direction.
- The average number of stops per vehicle per hour serves as an indicator of progression quality and interference from queue spillovers and mutual blockages between the through and left-turn vehicles. The developed system, as expected, can significantly decrease the average number of stops per vehicle per hour for northbound through vehicles by 5.87% (from 6.61 to 6.22 stops), achieving statistical significance at the 0.05% level. A slight reduction of 0.09% was also observed for the southbound through vehicles, from 2.00 to

1.98 stops per vehicle per hour. This demonstrates that the developed responsive real-time control module can effectively mitigate the mutual blockage effect, particularly in the northbound direction (i.e., commuting direction), where mutual blockages are expected to take place more often.

- Significant delay reductions per vehicle per hour for the northbound through and turning-in vehicles with the developed responsive real-time control module were observed across all evaluated links, ranging from 9.05% to 12.57%. Specifically, the improvements in delay on the links from Intersections 2 to 1, 3 to 2, 4 to 3, and 5 to 4 were 9.05%, 11.61%, 12.57%, and 10.67%, respectively, all achieving statistical significance at least at the 0.05 level.
- The developed responsive real-time control module can significantly reduce the delays per vehicle per hour for left-turning vehicles, ranging from 6.02% to 12.26% across all evaluated links. This highlights the module's capability to effectively manage traffic overflows not only for through vehicles but also for those making left turns, particularly during commuting peak hours when the volume distribution between through and turning vehicles exhibits a high level of fluctuation. Notably, left-turn vehicles from Intersection 3 to Intersection 2 experienced the most delay reduction, because Intersection 2 has the shortest turning bay, which is most likely to be blocked by traffic queues.

Overall, the developed responsive real-time control module under those experimental scenarios can achieve a network-wide average delay reduction of 5.46%, from 131.16 seconds to 124.00 seconds per vehicle per hour, and reduce the average number of stops across the network by 3.38%, from 3.19 to 3.09 stops per vehicle per hour. These improvements demonstrate that while the responsive control primarily focuses on optimizing traffic flows along the main arterial,

the module can also effectively enhance the overall network's performance. Specifically, this is achieved without compromising the efficiency of traffic flows on the side streets.

In summary, the results of laboratory experiments and field data from MD 355 Rockville Pike in Bethesda, Maryland, have convincingly demonstrated the benefits of the developed responsive real-time control module's effectiveness in reducing queue blockage duration and minimizing the resulting impacts on traffic delay and increasing throughput for all traffic streams across the entire arterial.

Table 5.3 Simulation results for performance comparison

Direction		Time-of-day control module	Responsive real-time control module	% change	
Average delay of through vehicles along the arterial per vehicle per hour (sec)					
Northbound	⑤→①	234.32	217.01	-7.39%	*
Southbound	①→⑤	141.65	141.46	-0.13%	
Average number of stops of through vehicles along the arterial per vehicle per hour					
Northbound	⑤→①	6.61	6.22	-5.87%	*
Southbound	①→⑤	2.00	1.98	-0.90%	
Average delay of through vehicles from all upstream streams on individual links per vehicle per hour (sec)					
	②→①	26.60	24.19	-9.05%	*
Northbound	③→②	52.19	46.13	-11.61%	*
	④→③	48.02	41.98	-12.57%	*
	⑤→④	41.56	37.12	-10.67%	*
Average delay of left-turn vehicles from all upstream streams on individual links per vehicle per hour (sec)					
	②→①	41.29	38.80	-6.02%	*
Northbound	③→②	64.54	56.62	-12.26%	*
	④→③	82.97	75.20	-9.36%	*
	⑤→④	33.48	30.09	-10.13%	
Average delay per vehicle per hour (sec)					
Network		131.16	124.00	-5.46%	*
Average number of stops per vehicle per hour					
Network		3.19	3.09	-3.38%	*

Note: * indicates the significance level of at least 0.05

5.5 Conclusions

This chapter has introduced a responsive real-time arterial signal control module designed to mitigate the adverse effects of turning-bay overflows and mutual queue blockages, which are commonly observed on commuting arterials with high and asymmetric directional traffic volumes. Building upon the time-of-day control methodology developed in Chapter 4, this responsive module offers a dynamic enhancement by detecting evolving queue patterns, classifying the severity of congestion, and executing tailored control strategies in real time.

The developed module consists of three core components: (1) a traffic monitoring model for detecting time-varying queue blockages across all intersections; (2) a congestion level classification model that applies spatiotemporal heuristics to identify congestion severity; and (3) an event-based control model that selects and implements signal plans responsive to the identified traffic states. These integrated components allow the module to respond effectively to dynamic traffic patterns and provide timely signal adjustments to prevent the escalation of the congestion level.

Extensive numerical experiments using a simulated commuting arterial along MD 355 Rockville Pike in Bethesda, Maryland, have confirmed the promising performance of the proposed responsive control module. Compared with the time-of-day signal control module presented in Chapter 4, the responsive module has demonstrated substantial improvements in delay reduction, duration for queue dissipation, and throughput for both through and left-turning movements—particularly at intersections susceptible to turning-bay spillovers and short link blockages. On average, the module has achieved a reduction of 5.46% in total network delay and a reduction of 3.38% in the number of stops per vehicle per hour, while ensuring efficient traffic operations in both commuting and non-commuting directions.

Overall, the findings from experimental scenarios have confirmed that the responsive real-time control module offers a robust and adaptive signal plan for congested commuting arterials, especially under time-varying traffic congestion patterns that cannot be effectively addressed by pre-timed signal plans and most existing models.

Chapter 6 A Proactive Adaptive Signal Control Module with Queue Dynamics Prediction for Congested Arterials

6.1 Introduction

Chapter 5 has introduced a responsive arterial signal control module that detects queue formation in real time and selects the optimal set of signal timing strategies to mitigate the adverse impacts incurred by mutual lane blockages and turning-bay overflows. While such a responsive approach is effective in managing emerging congestion, it is inherently constrained by the timing of blockage detection—response remedies are only initiated after the congestion pattern has already emerged. Hence, if sufficiently reliable prediction information is available, a proactive control strategy that takes all necessary adjustments in signal plans ahead of predicted congested traffic patterns is expected to be the most effective.

As such, this chapter presents the development of a proactive real-time arterial signal control module, designed to forecast queue dynamics for the control center to take preventive actions before the emergence of queue spillovers or mutual blockages. This approach is particularly crucial for commuting arterials characterized by highly asymmetric directional traffic flows, where short-term traffic surges and localized interactions between turning and through movements can escalate rapidly and propagate across multiple intersections.

The developed proactive module, building upon the same structure of the responsive control module introduced in Chapter 5, has been designed with two essential advancements. First, it contains a short-term queue state prediction model—based on a multi-branch, multi-head Long Short-Term Memory (MBMH-LSTM) neural network—that forecasts queue lengths and the onset time of any queue blockage using upstream traffic flows' spatiotemporal patterns. Second, to ensure the reliability of prediction-driven control, a monitoring mechanism is introduced to assess

the consistency between predicted and observed queue lengths in real time. This feature is particularly important given the expected variability in traffic conditions such as incidents, which may compromise prediction accuracy. When significant discrepancies between the predicted and actual traffic states are detected, the system is designed to seamlessly revert from proactive to responsive operations, thereby maintaining robust and effective traffic controls under both stable and abnormal conditions.

The remainder of this chapter is organized as follows. Section 6.2 presents an overview of the proactive control module, highlighting its architecture and operational flows. Section 6.3 provides a detailed description of the module's core components, with a primary focus on the LSTM-based queue state prediction model and the associated monitoring functions. Following this, the prediction model's performance evaluation using multiple benchmark metrics is presented in the same section. Finally, Section 6.4 presents a case study that applies the developed proactive strategy in a real-world context and compares its operational effectiveness against the responsive real-time control module introduced in Chapter 5, highlighting the module's effectiveness in tackling the dynamically evolving congestion patterns with preventive control strategies.

6.2 Overview of the Proactive Real-Time Arterial Signal Control Module

Building upon the logic of time-of-day control developed in Chapter 4 and the responsive real-time control strategy introduced in Chapter 5, this chapter presents the third module—a proactive real-time arterial signal control module. This module is designed to predict emerging traffic conditions and implement signal control strategies in advance of queue spillbacks or mutual blockages. Such predictive capability is particularly valuable for commuting arterials characterized by highly unbalanced directional flows and rapidly evolving traffic states, where delaying control responses can result in substantial performance degradation.

The primary distinction between this proactive and the responsive control modules lies in the incorporation of a short-term queue state prediction function. Rather than relying solely on detected queue conditions, the proactive module leverages predictive insights to capture the traffic patterns and initiate preemptive signal adjustments. In addition, a monitoring mechanism is embedded within the module to evaluate the reliability of the predicted results. If the deviations between predicted and observed traffic states exceed a predefined threshold, the system will suspend the proactive operations and revert to the responsive controls outlined in Chapter 5. This adaptive fallback mechanism ensures both the robustness and reliability of the system's performance under various levels of traffic uncertainties.

Figure 6.1 illustrates the structure of the developed proactive module, which consists of four sequentially operated models. Specifically, Model 1 performs queue state prediction; Model 2 monitors prediction accuracy and determines whether a mode switch is necessary; Model 3 evaluates the predicted congestion level; and Model 4 generates the set of optimal proactive signal control strategies. Since Models 3 and 4 maintain the structural logic similar to that introduced in Chapter 5, the emphasis of this chapter is placed on Models 1 and 2—both are newly developed to distinguish the proactive control module from its responsive counterpart.

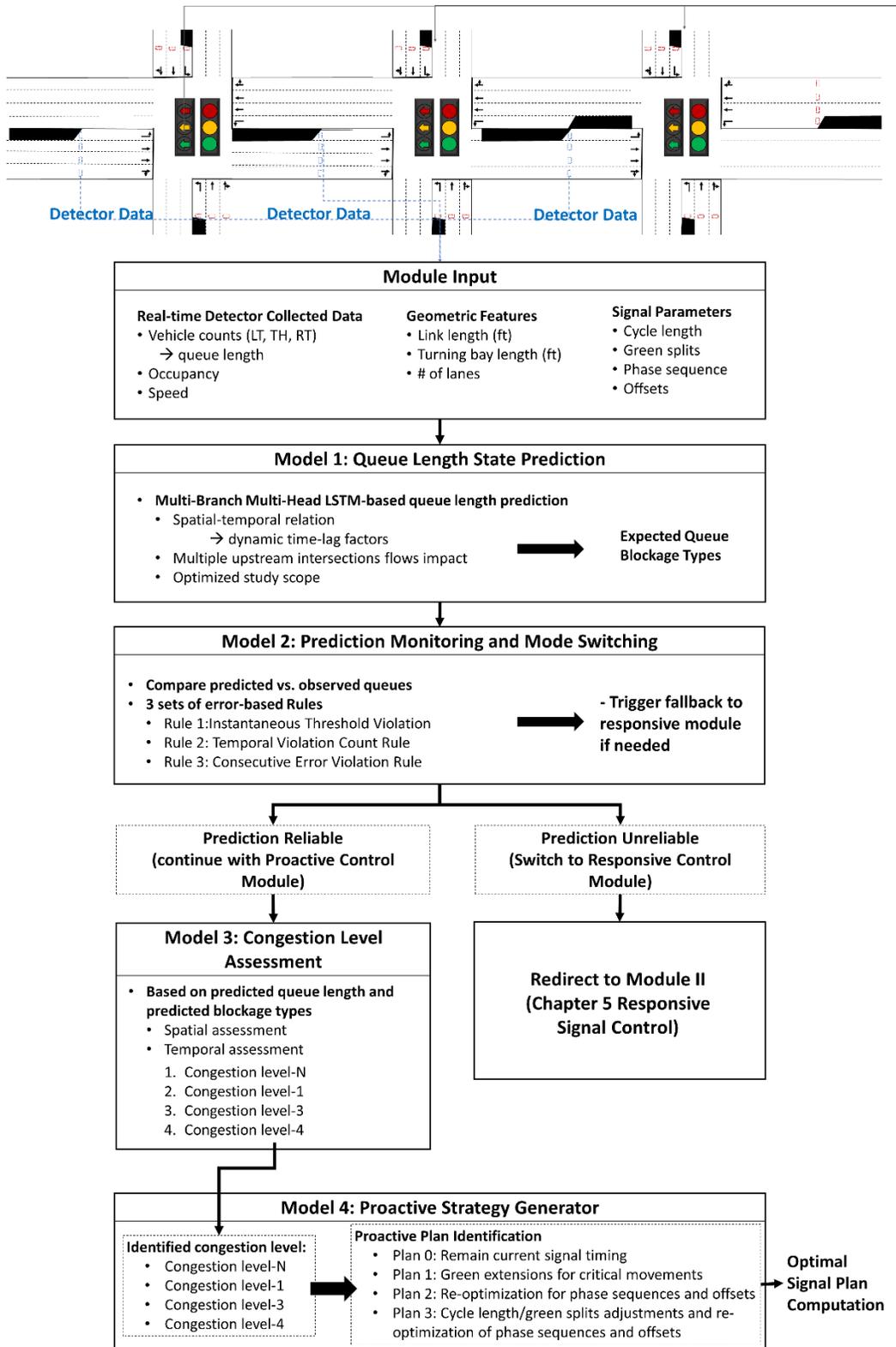


Figure 6.1 Structure of the Proactive Real-Time Arterial Signal Control Module

6.3 Functions and Formulations of the Key Embedded Models

This section presents the four core models that constitute the proactive real-time arterial signal control module. Each is designed to perform a specialized function, operating sequentially to enable forward-looking signal adjustments in anticipation of evolving congestion patterns. The overall design builds upon the modular structure of the responsive control module, but with two key enhancements: (1) replacement of real-time queue detection with a predictive queue state estimation model, and (2) introduction of a monitoring function and mode switching mechanism to continuously evaluate the prediction accuracy and ensure operational reliability.

6.3.1 Model 1: Queue Length State Prediction

The first core model of the proactive control module is for the queue length state prediction. Specifically, the model is designed to generate short-term predictions of time-varying queue lengths for through and left-turn movements at all monitored intersections. These prediction results provide essential information in advance for the module to take proactive signal control actions aimed at preventing potential spillovers and queue mutual blockages. The following subsections present the problem nature, modeling challenges, and the proposed solution—a multi-branch, multi-head Long Short-Term Memory (MBMH-LSTM) network tailored to the dynamic and spatially correlated nature of arterial traffic flows.

6.3.1.1 Problem nature and modeling challenges

As emphasized by Mirchandani and Head (2001), the effectiveness of any proactive signal control system depends fundamentally on the reliability of its underlying traffic prediction model. Accurate forecasts of key short-term traffic variables—such as approaching volumes, turning ratios, and queue lengths—are essential information for timely and effective control decisions to

proactively prevent or mitigate emerging congestion patterns. Among these variables, the state of the queue length—indicating whether or not a turning bay will overflow or a queue blockage will occur, its onset time, and the lasting duration—serves as one of the most critical indicators to trigger the activation of proactive control strategies in a timely manner.

However, having a short-term predictive function at the desirable level of accuracy for signal control systems presents significant challenges to the traffic control community. Because urban traffic systems are inherently stochastic, nonlinear, mutually-dependent, and time-varying by nature, the collective manifestation of such complex properties is especially pronounced during transitional periods between peak and off-peak hours (Kang et al., 2017). Moreover, the prediction's reliability is further complicated by the spatial-temporal interdependence between all involved traffic streams that contribute to the traffic dynamics and congestion patterns in commuting arterials.

The dynamic interactions between traffic congestion and the accompanied queue blockages that often plague such commuting corridors can be illustrated with the temporospatial relations between all contributors in Figure 6.2, where the formation of the queue length at a downstream target intersection (Intersection 1) depends not only by its immediate upstream input (i.e., Intersection 2), but also by a series of its upstream traffic streams from Intersections 5 through 2. Each pair of adjacent intersections' traffic states (e.g., 5→4, 4→3, etc.) is subject to disruptions by various types of congestion patterns and their resulting traffic queues, including:

- Queue overflows and mutual blockages, which impede vehicle progression (Figure 6.2, Area A);
- Residual queues from last cycles, which will delay upstream arrivals (Figure 6.2, Area B);
- Signal phasing changes, which periodically interrupt the traffic stream evolution and contribute to the queue formation (Figure 6.2, Area C).

The impacts from such disruptions do not take place in isolation; rather, they accumulate and propagate downstream, interacting dynamically across both spatial and temporal dimensions. For instance, a disruption to traffic at Intersection 5 may delay vehicles heading to Intersection 4, where their local queues and responsive signal timings consequently impact the vehicle arrival patterns at Intersection 3. Such sequential impacts will propagate downstream and ultimately reshape the queue formation dynamics at the target Intersection 1. Such collective impacts on a target intersection, characterized by nonlinear, time-lagged, and spatially distributed dependencies, demand a prediction model to be capable of inferring not only immediate upstream conditions, but also all indirect, delayed, and temporally dispersed contributors originating from multiple upstream intersections.

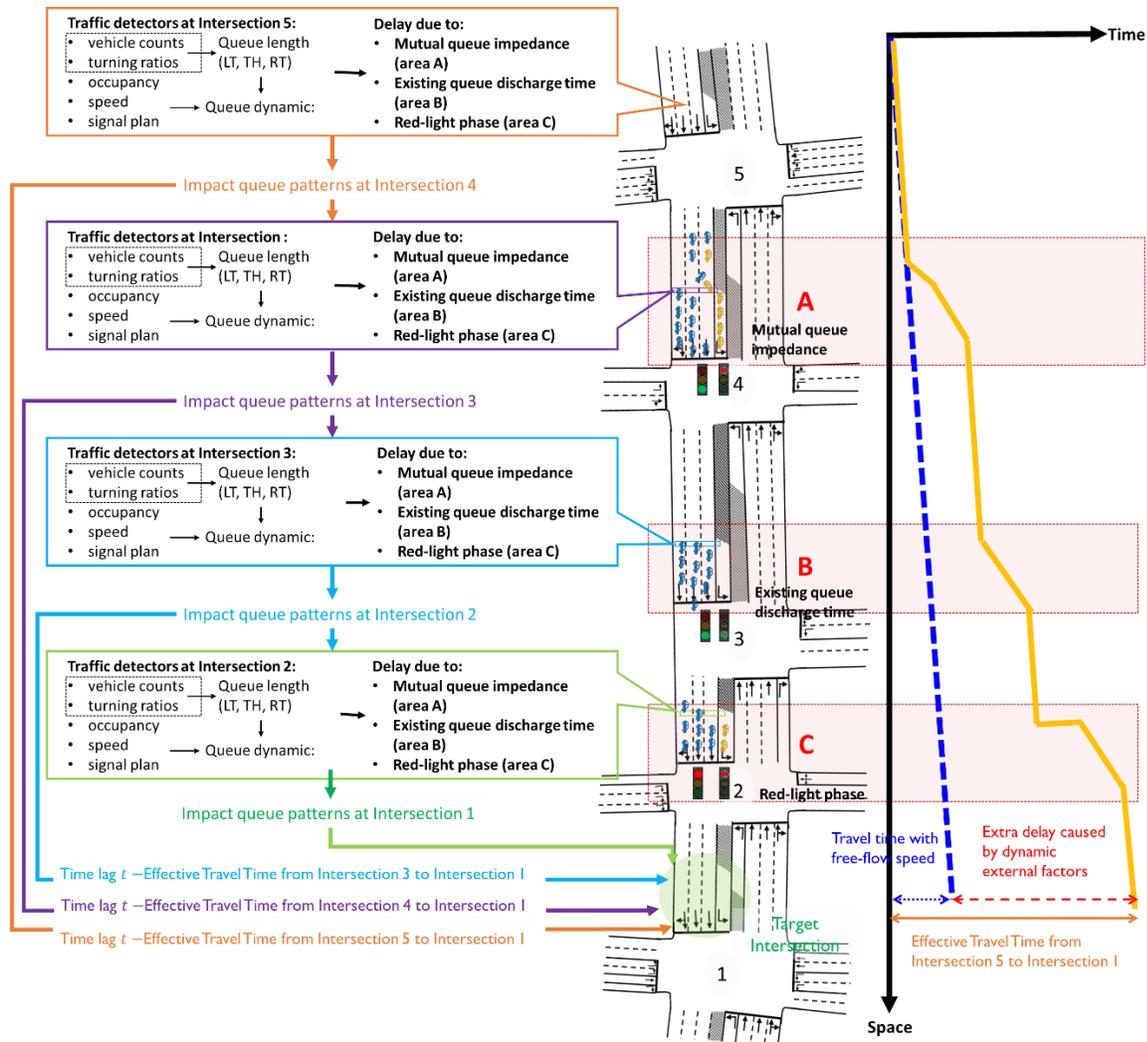


Figure 6.2 Spatiotemporal relations between upstream flows and resulting traffic states at downstream intersections over a congested commuting arterial

To tackle such challenges with a robust prediction model, it is imperative to sort out the spatial range of upstream traffic information to be included in the model. While incorporating real-time traffic data from more upstream intersections may improve the predictive accuracy, it is likely to demand undue computational overhead and introduce some more data noise that may degrade a model’s real-time prediction reliability. Furthermore, the temporal impacts of traffic states at each upstream intersection vary with its distance from the target intersection, the quality of signal

coordination, and the prevailing congestion levels. Such heterogeneous and intermingled traffic dynamics add another layer of modeling difficulty, especially when attempting to balance the tradeoff between model performance and computing efficiency for real-time implementation.

Taken together, these challenges underscore the need for a predictive modeling framework that can learn and infer temporospatial dependencies across multiple upstream intersections while maintaining operational tractability. The modeling approach presented in this chapter addresses these functional requirements through the following multi-branch, multi-head deep learning architecture:

6.3.1.2 Operational structure of the prediction model

In the developed architecture, each upstream intersection is represented with an independent LSTM branch, which allows the developed model to learn the temporal patterns unique to each location (see Figure 6.3). The developed model structure can enhance its flexibility by preserving local dynamics at the intersection level while also supporting the integration of diverse input variables—such as vehicle volumes, occupancies, signal coordination, and queue dynamics—specific to each intersection.

At a higher level of the architecture, the multi-head fusion mechanism aggregates the learned temporal features from each LSTM branch into a shared representation. This fusion layer plays a critical role in synthesizing information across space, enabling the model to accurately infer how traffic dynamics at multiple upstream intersections interact and cumulatively affect downstream queue formation. For instance, traffic conditions at Intersection 5 may influence those at Intersection 4, which subsequently impact Intersections 3 and 2, ultimately shaping the queue evolution at the target downstream intersection. This cumulative effect is particularly pronounced in congested commuting corridors characterized by tightly spaced intersections and unstable

directional flows. A schematic depiction of how the developed structure is used to reflect the states of traffic flows spatially and temporally across the arterial is provided in Figure 6.2.

By aligning the model architecture with the topology of the arterial network, the developed multi-branch LSTM offers both the desirable interpretability and scalability. It is well-positioned to support proactive signal control systems with accurate and reliable predictions of queue spillovers and blockage events under various dynamic and congested traffic conditions.

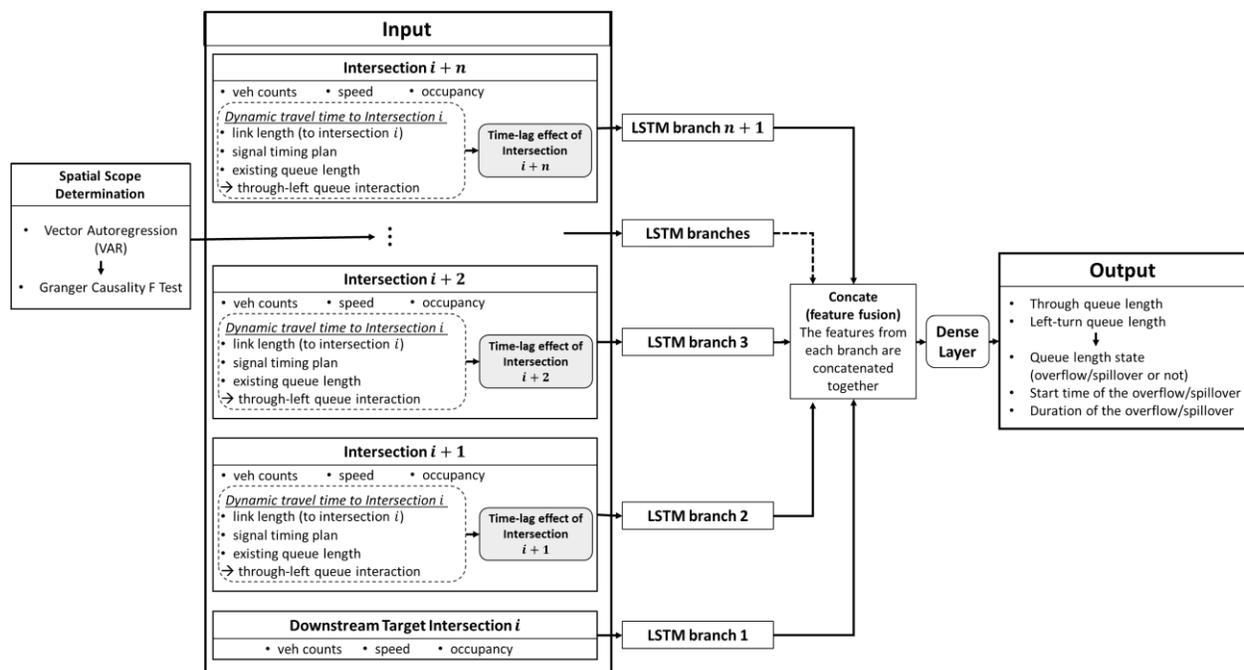


Figure 6.3 Operational structure of the developed multi-branch multi-head LSTM-based prediction model

6.3.1.3 Methodological design

The Multi-Branch Multi-Head LSTM-based model is developed through a structured and sequential process. Each step corresponds to a key modeling decision aimed at capturing the delayed, direction-specific, and temporally aligned interactions between multiple upstream intersections and the target downstream intersection. These steps are outlined as follows:

- **Step 1: Spatial Scope Determination**

Identify and select upstream intersections that pose significant impacts on the downstream intersection's queue dynamics, based on the roadway geometric features (e.g., link length, turning-bay length), signal plan, and observed traffic interactions.

- **Step 2: Multi-Branch LSTM Architecture Construction**

Construct a dedicated LSTM branch for each upstream intersection to independently learn its temporal traffic patterns. A separate branch is also created for the downstream intersection.

- **Step 3: Input Alignment with Adaptive Lag Time and Temporal Feature Learning**

Align the input sequences from upstream intersections by dynamically estimating time-varying propagation delays (lags), ensuring temporal coherence with the downstream intersection's queue state. This enables each LSTM branch to learn features based on the lag-adjusted historical input.

- **Step 4: Temporal-Spatial Feature Fusion and Multi-Head Queue Length Prediction**

Fuse those temporally aligned hidden representations from all upstream branches into the target downstream branch. A unified spatiotemporal state representation is then passed to multiple heads to predict left-turn and through queue lengths at a future decision point $t + \delta$, to support the proactive signal control.

- **Step 5: Event-Based Queue Blockage Identification**

Use the predicted queue lengths and known geometric constraints to detect potential queue blockages (e.g., spillback or bay overflow) and their durations, enabling the signal system to take an event-based control intervention in a timely manner.

6.3.1.3.1 The prediction model's spatial scope

To balance prediction accuracy and computational efficiency, a two-stage statistical screening process is conducted to identify and include only those upstream intersections that have a significant influence on the downstream queue state. This process involves:

- 1) Bivariate Vector Autoregression (VAR) model estimation, and
- 2) 2. Granger causality testing between the traffic features at each upstream intersection and the queue state at the downstream intersection of interest.

Let Q_t denote the downstream queue length at time t , and let $\mathbf{X}_{j,t}$ represent the vector of selected traffic features (e.g., traffic volumes, occupancy) from the upstream Intersection j . Specifically,

$$\mathbf{X}_{j,t} = \{q_{j,t}, o_{j,t}, u_{j,t}, s_{j,t}, U_{j,t}\} \quad (6-1)$$

where $q_{j,t}$ is the observed vehicle counts at Intersection j at time t ; $o_{j,t}$ and $u_{j,t}$ are the detected occupancy and speed at Intersection j at time t , respectively; $s_{j,t}$ is the signal phase indicator at Intersection j at time t ; $U_{j,t}$ is the estimated existing queue length at Intersection j at time t .

A bivariate VAR model of order p is constructed for each upstream intersection j as follows:

$$\begin{bmatrix} Q_t \\ \mathbf{X}_{j,t} \end{bmatrix} = \sum_{k=1}^p \begin{bmatrix} A_{11}^k & A_{12}^k \\ A_{21}^k & A_{22}^k \end{bmatrix} \begin{bmatrix} Q_{t-k} \\ \mathbf{X}_{j,t-k} \end{bmatrix} + \begin{bmatrix} \varepsilon Q_t \\ \varepsilon \mathbf{X}_{j,t} \end{bmatrix} \quad (6-2)$$

Here, A_{ab}^k denotes the coefficient matrices associated with the k -th lag, where the subscripts a and b refer to the dependent and explanatory variables, respectively. Specifically, A_{11}^k captures the autoregressive effect of past downstream queue lengths Q_{t-k} on the current Q_t ; A_{12}^k represents the influence of past upstream traffic features $\mathbf{X}_{j,t-k}$ on Q_t . The error terms εQ_t , $\varepsilon \mathbf{X}_{i,t}$ are assumed to be white noise processes. To assess whether upstream intersection i provides statistically significant predictive information, a Granger causality test is applied to the estimated VAR model.

The null hypothesis of the Granger causality test states that the lagged values of $X_{j,t-k}$ do not improve the prediction of Q_t beyond what is already explained by its own past values.

That is:

$$H_0: A_{12}^1 = A_{12}^2 = \dots = A_{12}^p = 0$$

Rejection of H_0 implies that the upstream traffic features Intersection j Granger-cause the downstream queue length, and thus this intersection is considered influential and should be selected for constituting the model input.

The model order p is determined based on information criteria such as AIC or BIC. The F-statistic for the Granger causality test is computed by comparing the residual sum of squares (RSS) from the unrestricted VAR model and a restricted version where the parameters A_{12}^k are set to zero.

$$F = \frac{(RSS_r - RSS_u)/m}{RSS_u/(T - 2p - 1)} \quad (6-3)$$

where RSS_r and RSS_u are the residual sum of squares of the restricted and unrestricted models, respectively; m is the number of restrictions (i.e., total coefficients in A_{12}^k); T is the number of observations.

Only upstream intersections that exhibit statistically significant Granger Causality (i.e., at a 95% confidence level) are retained for downstream queue prediction modeling.

6.3.1.3.2 Multi-branch LSTM architecture construction

Each selected upstream intersection is assigned with a dedicated LSTM branch within the developed multi-branch architecture. While all branches share the same internal configuration, they operate on the input sequences that represent traffic conditions specific to their respective locations over the arterial. This structure enables the prediction model to individually learn and

preserve the temporal dynamics unique to each upstream intersection.

Each input sequence consists of multi-variable traffic data—including vehicle counts, speed, occupancy, signal phase, and existing queue length—collected over a prespecified period at time intervals of fixed length. As illustrated in Figure 6.4, the uniform LSTM structure, adopted across all branches, incorporates a memory cell and three gating mechanisms: the input gate, forget gate, and output gate. These mechanisms collectively regulate the flows of information over time and update two internal states: the cell state, c_t , which serves as the long-term memory, and the hidden state, h_t , which captures the short-term context for use as the output at each time step. The gating mechanism selectively retains or discards information in these states, allowing the model to capture long-term dependencies while filtering out noise and random fluctuations.

The modularized design of this architecture not only facilitates scalable modeling across varying network sizes (i.e., different numbers of upstream intersections) but also supports parallelized learning for all data streams. This forms the backbone of the model’s capability to encode temporally rich patterns in a distributed yet coherent manner.

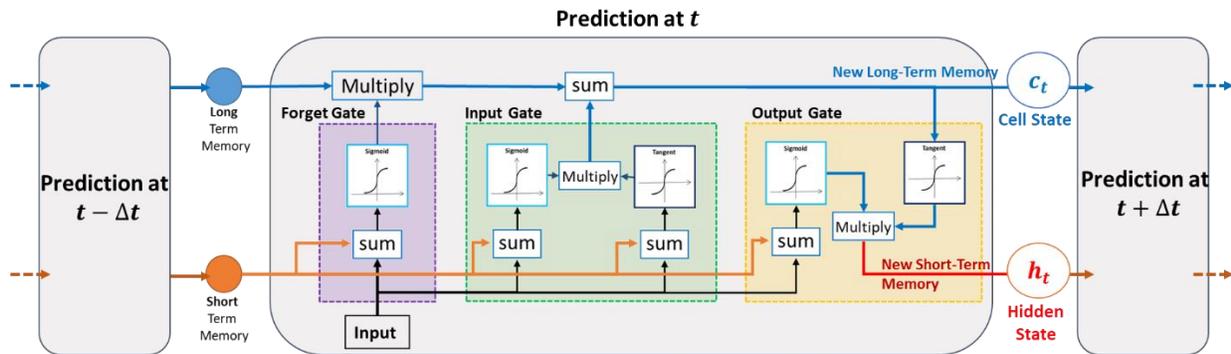


Figure 6.4 Structure of the standard LSTM unit

6.3.1.3.3 Incorporating the adaptive lag time and spatial-temporal information in the model input

As is well recognized, upstream intersections may influence downstream traffic conditions with varying levels of delay, shaped by physical distance, prevailing speed, and local signal operations. Accurate modeling of these delayed interactions requires a precise temporal alignment of upstream inputs with downstream outcomes. To account for these heterogeneous and time-varying delays, the developed model introduces an adaptive lag mechanism to ensure that the model's input from all upstream intersections is consistent with their temporal and spatial alignment.

For a downstream intersection i , the prediction of queue length at time t relies not only on its own recent traffic states but also on lag-adjusted historical information from each upstream Intersection $j \in U_i$ (the set of upstream intersections for downstream intersection i). The time delay (or lag), $\tau_{j,t}$, from upstream Intersection j to downstream i is dynamically estimated based on the estimated travel time $T_{j \rightarrow i,t}$, which integrates physical and operational traffic factors, as defined below:

$$T_{j \rightarrow i,t} = \frac{L_{ji}}{u_{j,t}} + Q_{j,t}^* + s_{j,t}^* \quad (6-4)$$

$$\tau_{j,t} = \left\lceil \frac{T_{j \rightarrow i,t}}{\Delta t} \right\rceil \quad (6-5)$$

where L_{ji} is the distance between upstream Intersection j and downstream Intersection i ; $Q_{j,t}^*$ denotes the queue-induced delay at Intersection j ; $s_{j,t}^*$ represents the signal-related delay at Intersection j ; and Δt is the time interval.

To construct the input for predicting downstream queue length state at time t , the developed model is designed to aggregate the following two types of temporal information:

- 1) local historical data from the downstream intersection i : $\{x_{i,t-k} \mid k = 1, \dots, H\}$

2) lag-adjusted historical data from each upstream intersection j : $\{x_{j,t-\tau_{j,t-k}} \mid k = 1, \dots, H\}$, where H denotes the fixed input sequence length.

This alignment ensures that for any historical downstream observation at time $t - k$, the model can incorporate all inputs from the corresponding upstream time $t - k - \tau_{j,t}$, that is, the moment when upstream traffic conditions are expected to begin their impacts on the downstream traffic states. By modeling these lag structures explicitly, the architecture functions to maintain the temporal coherence and to enhance the model's ability to learn from the information revealed in the delayed spatiotemporal dependencies.

In addition to the temporal alignment of inputs, the frequency at which predictions are made must account for the minimum time it takes a vehicle to traverse the shortest link between any two intersections. This ensures the model can respond at the level of efficiency sufficient to capture fast-evolving queue conditions. Accordingly, the prediction frequency is defined as (6):

$$\text{Prediction Frequency} = \min\left(\frac{l_i}{v}, \frac{l_{i+1}}{v}, \dots, \frac{l_{i+N}}{v}\right) \quad (6-6)$$

where l_i denotes the distance between detector locations at adjacent Intersections i and $i + 1$, and v is the prevailing traffic speed. This formulation ensures that predictions are generated at a frequency sufficient to reflect real-time traffic dynamics, especially on closely spaced or fast-moving arterial segments.

6.3.1.3.4 Temporal-spatial fusion feature for multi-head queue length prediction

Following the temporal-spatial alignment, the next step involves fusing the extracted features from multiple upstream intersections into the downstream queue length prediction process. The objective is to predict the queue length state at a fixed future time step, $t + \delta$, where the prediction horizon, δ , is so selected to allow sufficient lead time for downstream controllers to compute and

implement proactive signal plans based on the projected queue conditions.

Each upstream Intersection j is modeled with a dedicated LSTM branch that processes its lag-adjusted input sequence:

$$h_{j,t}^k = f_{LSTM} \left(X_{j,t-\tau_{j,t-k}}, h_{j,t}^{k-1}; \theta_j \right) \quad k = 1, 2, \dots, H \quad (6-7)$$

where θ_j represents the parameters of the LSTM branch for Intersection j . This process yields a final hidden state $h_{j,t}^H$, which summarizes upstream traffic conditions aligned to impact the downstream intersection at time t . Note that the model parameters θ_j and θ_i are learned via backpropagation using the training dataset.

At the downstream Intersection i , a separate LSTM processes its own local traffic history while integrating temporally aligned upstream states in the prediction process. The final downstream hidden state is computed using (8):

$$h_{i,t}^H = f_{LSTM} \left(X_{i,t-1}, \sum_{j \in U_i} h_{j,t}^H, h_{i,t}^{k-1}; \theta_i \right) \quad k = 1, 2, \dots, H \quad (6-8)$$

The resulting downstream hidden state $h_{i,t}^H$ represents a temporally fused encoding of both local and upstream traffic dynamics, based on their respective influence delays.

The fusion mechanism is illustrated in Figure 6.5, where each upstream intersection contributes a lag-adjusted hidden state that is integrated into the downstream LSTM's computation. These fused results are subsequently passed through separate prediction heads to estimate the left-turn and the through queue length at time step $t + \delta$:

$$\hat{Q}_{i,t+\delta}^n = W^n \cdot h_{i,t}^H + b^n \quad n \in \{LT, TH\} \quad (6-9)$$

Here, $\hat{Q}_{i,t+\delta}^{LT}$ and $\hat{Q}_{i,t+\delta}^{TH}$ denote the predicted left-turn and through queue lengths at time step $t +$

δ , respectively. This multi-head design enables the model to jointly learn different types of queue dynamics while leveraging the relationship of a shared temporal-spatial feature.

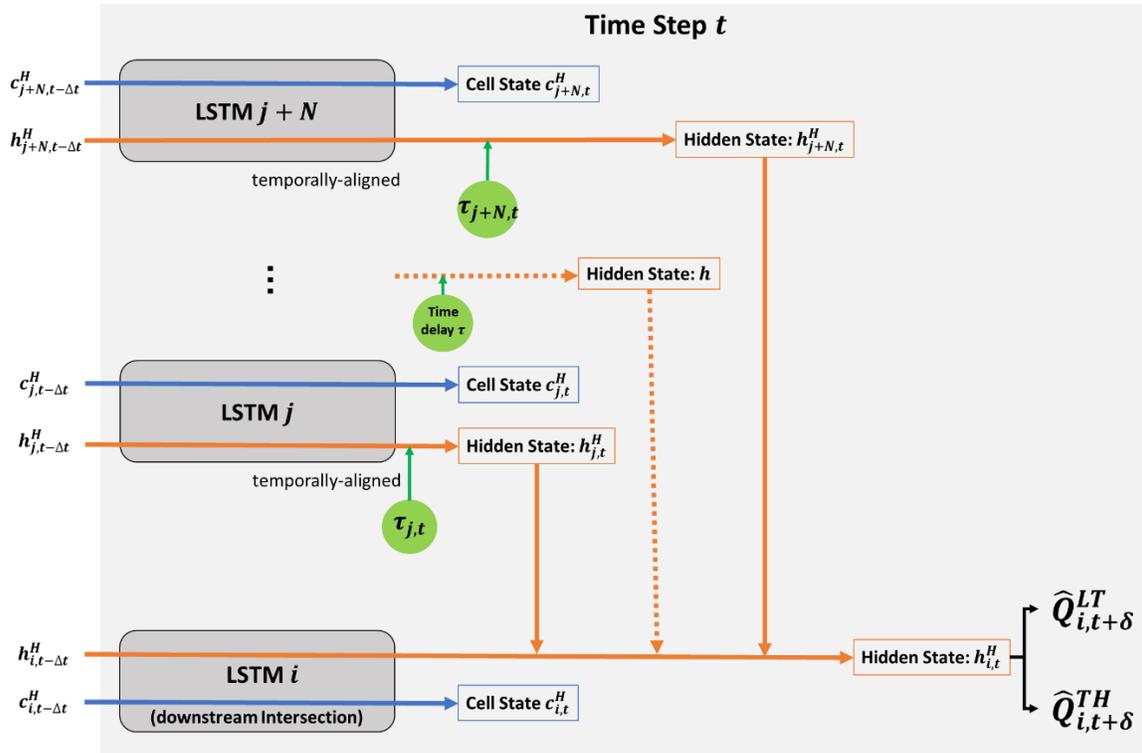


Figure 6.5 Architecture of the temporally-aligned multi-branch LSTM model for downstream queue length prediction

6.3.1.3.5 Event-based blockage identification

To support the proactive control decisions, the model's predicted queue lengths will be transformed into quantifiable traffic events—specifically, the occurrence and duration of blockage events caused by excessive left-turn or through queues.

A blockage event is defined as a period during which the predicted queue length for a given movement (i.e., left-turn or through) exceeds its physical storage capacity, resulting in spillovers

or blockage of movements on adjacent lanes. The following two types of blockages should be monitored:

- Left-turn spillover: occurs when the predicted left-turn queue exceeds the length of the left-turn bay.
- Through-queue blockage: occurs when an excessively long through queue prevents left-turning vehicles from entering the bay

Given a prediction horizon of δ seconds, and predicted queue lengths $\hat{Q}_{i,t+\delta}^n$ for each movement $n \in \{LT, TH\}$, the model will check, at each prediction time t , whether:

$$\hat{Q}_{i,t+\delta}^n > L_i^*$$

where L_i^* denotes the left-turn bay length at Intersection i .

If this condition is satisfied, a blockage flag will be raised, and the event will be tracked across consecutive prediction points to identify:

- Blockage start time
- Blockage end time
- Blockage duration

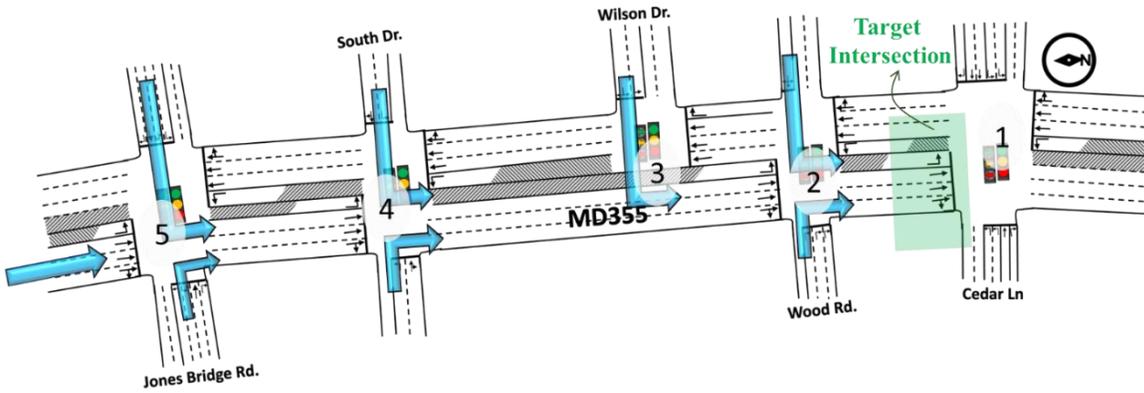
This event-based approach enables the model to predict any critical queue spillovers prior to their actual occurrence and prevent operational disruptions. The predicted results, if sufficiently reliable, can serve to guide the selection and execution of the real-time control strategies.

6.3.1.4 Performance evaluation

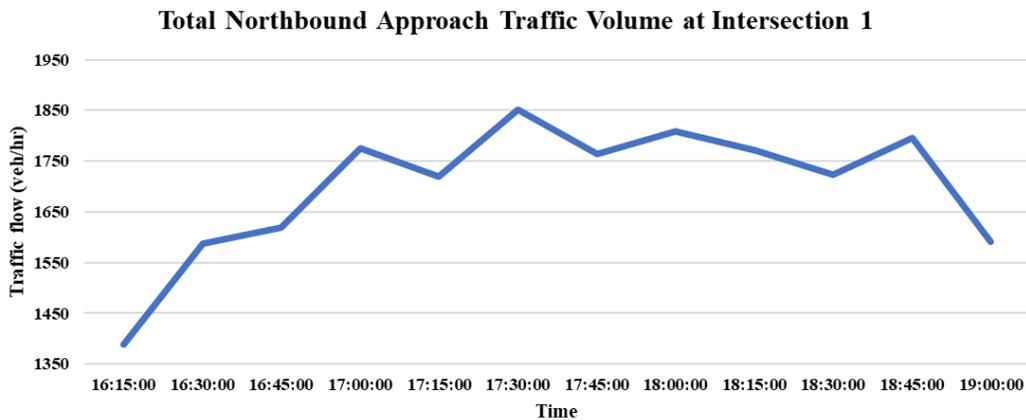
The performance of the developed model is evaluated under a series of extensive simulated traffic scenarios to (1) determine the optimal number of upstream intersection data sources to be incorporated into the prediction process, and (2) assess the model's suitability for real-time operations. The evaluation is based on the following Measures of Effectiveness (MOEs):

- **Event Detection Rate:** The proportion of actual spillover/blockage events (i.e., those generated in the simulated traffic scenario) that are correctly identified by the model.
- **False Alarm Rate:** The proportion of normal queue states incorrectly classified as spillover/blockage events.
- **Average Start Time Deviation:** The mean difference between the predicted and actual start times of all detected spillover events. A positive (negative) value indicates that the detected event is predicted to take place later (earlier) than its actual onset time.
- **Average Duration Deviation:** The mean difference between the predicted and actual durations of detected spillover events. A positive value implies an overestimation of the blockage duration, whereas a negative value indicates an underestimate

Figure 6.6(a) illustrates the selected study arterial: a five-intersection segment of MD 355 Rockville Pike in Bethesda, Maryland. Geometric configurations and lane arrangements of each intersection are shown in the figure. Input data were obtained from the Internet Traffic Monitoring System (I-TMS). As the northbound direction represents the primary commuting flows during the PM peak period, it was selected as the target direction for prediction. Vehicle counts from I-TMS were used directly for all northbound inflows, including through, left-turn, and right-turn movements at the most upstream intersections (Intersection 5), as well as for turning-in movements from side streets at all intersections (as indicated by the blue arrows in Figure 6.6 (a)). The prediction target is the northbound time-varying queue length at the most downstream intersection—MD 355 at Cedar Lane (Intersection 1). Figure 6.6 (b) displays the northbound traffic flow rates at Intersection 1 over a three-hour period, covering the hour preceding the PM peak, the peak hours, and the post-peak hours.



(a)



(b)

Figure 6.6 (a) Geometric features of the selected site; (b) Flow rate distribution over the 3-hour period at Intersection 1

6.3.1.4.1 Numerical experiment setup

To support the model training and evaluation, a dataset was generated using VISSIM to simulate a realistic urban traffic environment. A total of 20 independent simulation replications were conducted, each covering a three-hour operational period. The key variables recorded in each simulated scenario include vehicle counts, speed, occupancy, signal plan, and time-varying queue length.

To prepare the dataset for exercising the real-time prediction, a sliding window segmentation approach was employed. This technique maintains the temporal structure of the data

and allows the model to capture sequential dependencies effectively. At each time step t , the input sequence is defined as $X_t = \{x_t, x_{t+1}, \dots, x_{t+W-1}\}$, and the corresponding target output is $Y_t = x_{t+W}$, where W is the window size. To accommodate differences in travel times from various upstream intersections, a dynamic sliding window structure was used to ensure accurate temporal alignment between the input features and their respective outputs.

Predictions from the adopted models were generated every 30 seconds, matching both the data aggregation interval and the estimated minimum travel time from upstream detectors to the target intersection. Each candidate prediction model for evaluation was tasked with predicting the time-varying queue length on the northbound approach of the target intersection, with a prediction horizon of 300 seconds—equivalent to two complete signal cycles to ensure sufficient lead time for proactive signal control. These resulting datasets were partitioned using an 80/20 split, yielding per intersection of 5168 samples for model training and 1292 samples for performance evaluation.

6.3.1.4.2 Model construction

As previously discussed, determining the number of upstream intersections' traffic information to be included in the prediction system is the first step in model construction. As shown in Table 6.1, Intersections 2, 3, and 4 were found to meet the Granger-cause conditions with Intersection 1 (i.e., H_0 of no causality was rejected at the 5% significance level), whereas Intersection 5 did not exhibit a statistically significant causal relationship within the prediction's spatial-temporal horizon. Based on these results, all information and traffic states from the three upstream intersections (i.e., Intersections 2, 3, 4) were selected as the input sources. Accordingly, the developed model architecture consists of four principal branches: three corresponding to the selected upstream intersections and one representing the target intersection itself.

Table 6.1 The results of the Granger causality test

Upstream Intersection	P-value	Interpretation
Intersection 2	0.000	Reject H_o
Intersection 3	0.000	Reject H_o
Intersection 4	0.001	Reject H_o
Intersection 5	0.208	Fails to reject H_o

Each branch of the model processes its respective time series of input data independently before merging into the subsequent layers of the model. All model hyperparameters—including the number of LSTM units, learning rate, and batch size—were optimized through empirical tuning using the training dataset. The final values of these hyperparameters are summarized in Table 6.2. The training results, exhibiting the stable converging property, are shown in Figure 6.7, which represents the loss curve across training epochs.

Table 6.2 Optimized Hyperparameters for the Multi-Branch Multi-Head LSTM Model

Hyperparameters	Optimized Value
Units	64
Batch Size	32
Number of branches	4
Epochs	50
Learning rate	0.001

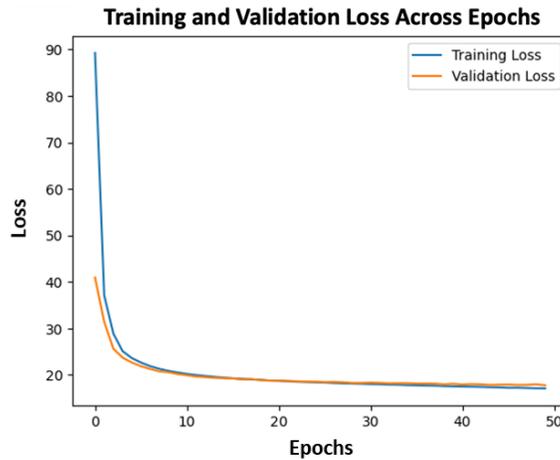


Figure 6.7 Training Loss Curve for the Multi-Branch Multi-Head LSTM Model

6.3.1.4.3 Model Performance Evaluation

To assess the performance of the developed multi-branch, multi-head LSTM model in predicting queue spillovers and blockage events, a comparative evaluation was conducted against two benchmark models: the Extended Kalman Filter (EKF) and a standard Recurrent Neural Network (RNN). All three models were trained and tested on identical input datasets, comprising traffic state features from three upstream intersections and the target intersection. While the developed model preserves the spatial structure by assigning each upstream intersection with its own input branch, both the RNN and EKF rely on concatenated inputs without spatial differentiation.

To focus on those time periods with surging traffic volumes and high blockage likelihood, the evaluation was conducted for traffic states during the PM peak hours. The time-varying traffic states, simulated with VISSIM using the actual volume data, are viewed as the “ground truth” for performance comparison needed in this study. Table 6.3 summarizes the performance of all three models using the selected MOEs, including detection rate, false alarm rate, and deviations in predicted start time and duration.

Table 6.3 Summary of Spillover/Blockage Prediction Performance

Metric	MBMH-LSTM	RNN	EKF
Event Detection Rate	100%	75%	25%
False alarm rate	0%	0%	0%
Average Start Time Deviation (sec/event)	0	87 ^a	40 ^a
Average Duration Deviation (sec/event)	0	-20 ^b	-30 ^b

Note:

a) A positive (negative) value indicates that the detected event is predicted to take place later (earlier) than its actual onset time.

b) A positive (negative) value implies an overestimation (underestimation) of the blockage duration.

Event detection and false alarm rates

Figure 6.8(a) and Figure 6.8 (b) present the predicted versus actual queue lengths for the through and left-turn movements, respectively, while Figure 6.8 (c) visualizes the binary classification of queue spillover/blockage states, with a value of 1 indicating the occurrence of blockage. Key findings from the event analysis are summarized below and further supported with Table 6.3:

- The developed model successfully detected all 12 true spillover/blockage events, achieving a perfect detection rate (100%) with no false positives. Notably, the model also exhibited zero average deviation in the prediction of the event start time, indicating precise detection of queue presence and its exact blockage onset time. These results reflect the model's strong temporal awareness and responsiveness, which are essential for timely traffic signal interventions.
- The RNN model correctly identified 9 out of 12 blockage events, resulting in a 75% detection rate with no false alarms. However, its accuracy in predicting the onset of these blockage events varied from case to case. While one prediction closely aligned with the actual start time (i.e., Event 2), the detection for blockage Events 3, 4, 6, and 7 showed minor delays behind their actual starting times (1 to 2 time intervals). In contrast, the remaining four events exhibited more pronounced delays, suffered a pronounced late detection of approximately one full signal cycle behind the actual onset time (equivalent to 4 to 5 time intervals). These discrepancies, highlighted in yellow in Figure 6.8(c), led to an estimated start time on average of 87 seconds deviated from the actual blockage starting time, the highest among the three models. These findings suggest that the RNN, while effective in short-term prediction, struggles with capturing the exact timing of emerging congestion patterns.

- The EKF detected only 3 out of 12 blockage events (25% detection rate), but with no false alarms. Among the detected blockage events, Event 8 was correctly aligned with the ground truth, while Events 4 and 6 experienced minor timing delays in the detection of the blockage onset time. The EKF's estimated start times on average deviate about 40 seconds from the actual blockage presence time, reflecting its limited responsiveness to abrupt traffic state transitions, due in part to its reliance on local linearization and recursive filtering.

Overall, while all models demonstrated perfect performance in terms of avoiding false positives, only the developed model has achieved both a perfect detection rate and a precise estimate of the blockage onset time, rendering it particularly well-suited for real-time activation of adaptive signal control strategies.

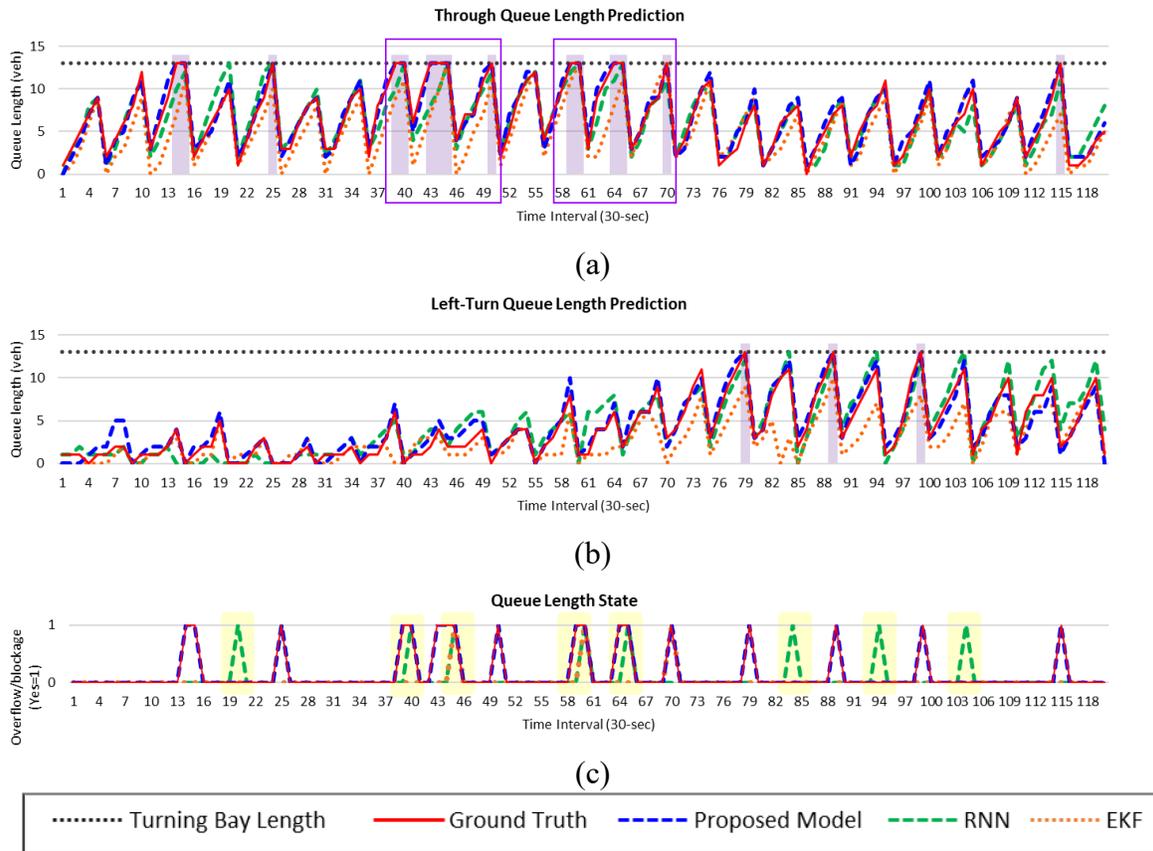


Figure 6.8 Comparison of predicted queue spillover/blockage events: (a) through movement prediction; (b) left-turn movement prediction; (c) binary spillover/blockage detection (1: blockage, 0: no blockage)

Estimation accuracy for the spillover/blockage duration

Figure 6.9 presents a comparison of the predicted versus actual duration for each of the 12 confirmed spillover/blockage events, allowing for further evaluation of the model’s ability to sustain accurate tracking of traffic states after the onset of a queue blockage. These findings are further supported by the evaluation results reported in Table 6.3:

- The developed model has perfectly captured the full duration of all 12 blockage events, with no deviation. This exact alignment reflects the model’s robustness in sustaining event state predictions across multiple time intervals and its ability to reflect the queue

presence and evolution. Such prediction reliability is especially important for proactive signal controls that depend on the accurate prediction of blockage duration to prioritize proactive strategies.

- Despite having detected 9 blockage events, the RNN tends to underestimate the duration, especially for those severe queue blockages. As illustrated with the dashed blue boxes in Figure 6.9, five of the nine detected blockage events were predicted to dissipate earlier than they actually occurred. On average, the RNN's predicted blockage durations were 20 seconds shorter than the observed durations, indicating a systematic bias toward premature termination of congested traffic queue states. Such performance likely arises from the RNN's inherent limitations in modeling the long-term temporal dependencies, which becomes more pronounced when congested queue flows span multiple cycles.
- The underestimation of blockage duration by the EKF model was even more pronounced, with an average of 30 seconds. Among the three events it successfully detected, the predicted durations for blockage Events 4 and 6 were significantly underestimated. This underperformance is likely due to the EKF's recursive smoothing process and its reliance on the dynamics of linear systems, which may limit its ability to accurately track sustained congestion patterns. As a result, the model tends to yield the prediction that traffic blockages will clear earlier than they actually take place, particularly under prolonged and saturated traffic states.

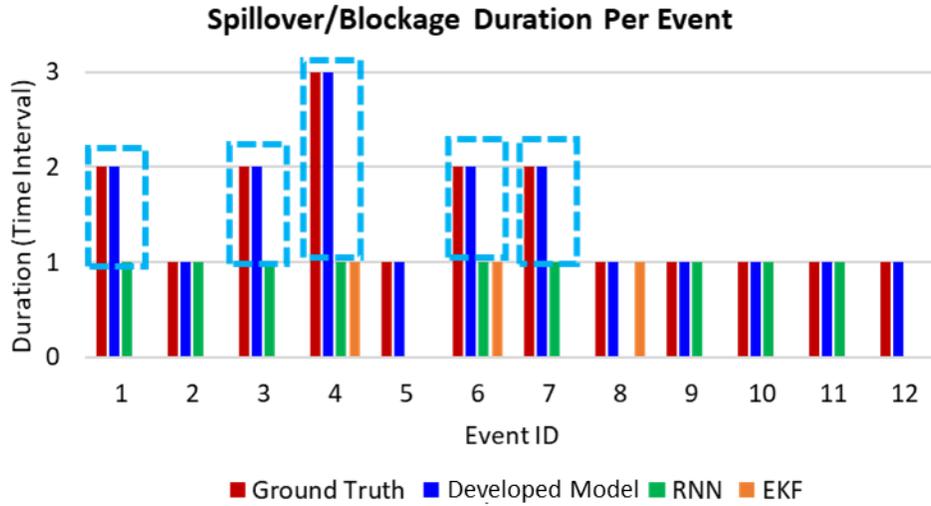


Figure 6.9 Comparison of Predicted and Actual Queue Blockage Events across Models (RNN, EKF, and the developed model)

The superior performance of the developed model can be attributed to its multi-branch LSTM architecture, which is specifically designed to fully capture the long-term dependencies across spatially distributed upstream traffic state information. By retaining temporal dynamics unique to each upstream intersection, the developed model is capable of maintaining a comprehensive and temporally coherent view of evolving traffic states. This architectural feature enables early detection of the onset time of predicted spillover/blockages, as well as accurate tracking of their persistent duration, even when congested queues surge rapidly or last over multiple time intervals—such as those illustrated within the purple boxes in Figure 6.8(a).

In contrast, the RNN’s reliance on a single memory stream constrains its ability to retain information over a long-time horizon, making it susceptible to losing critical contextual information as time progresses (i.e., vanishing gradient effects). As a result, it often underestimates the duration needed for blockage dissipation and fails to sustain its recognition of ongoing congested traffic queue states. As for EKF, it lacks the adaptability required to handle nonlinear

and abrupt changes in traffic patterns. Its filtering process, primarily optimized for noise reduction in stable systems, often leads to lagged detection and earlier than actual time for recovery, thereby limiting its potential for real-time traffic management.

In summary, the developed multi-branch LSTM model, with its promising performance in both the timely detection of blockage events and the accurate prediction of their durations, can offer all the information essential for the execution of the proactive signal controls, aimed at preventing severe traffic congestion caused by queue overflows and mutual blockages.

6.3.1.4.4 Sensitivity of upstream traffic information coverage on the prediction accuracy

To further evaluate the impact of including a different number of upstream intersections in the multi-branch LSTM model on the resulting prediction accuracy, this section reports the analysis results with the following three candidate model structures:

- Three-branch model structure: Includes data from Intersections 1 through 3 (i.e., two upstream intersections).
- Four-branch model structure: Includes data from Intersection 1 through 4 (i.e., three upstream intersections). This is the developed model.
- Five-branch model structure: Includes data from Intersections 1 through 5 (i.e., four upstream intersections).

To ensure consistency in evaluation, all models were tested over the PM peak hour period, during which congestion is most severe. Table 6.4 summarizes the performance of each model across the selected MOEs.

Table 6.4 Effect of spatial input scope on prediction accuracy: three-, four-, and five-branch LSTM comparison

Metric	Three-branch Model	Developed Model (Four-branch)	Five-branch Model
Event Detection Rate	75%	100%	100%
False alarm rate	0%	0%	7.7%
Average Start Time Deviation (sec/event)	27 ^a	0	-2.5 ^a
Average Duration Deviation (sec/event)	-20 ^b	0	2.5 ^b

Note:

a) A positive (negative) value indicates that the detected event is predicted to take place later (earlier) than its actual onset time.

b) A positive (negative) value implies an overestimation (underestimation) of the blockage duration.

Blockage event detection

Figure 6.10 illustrates the model prediction results across the three configurations. Figure 6.10(a) and Figure 6.10(b) show the predicted queue lengths for the through and left-turn movements, respectively, and Figure 6.10(c) presents the results of the binary queue state classification. Since the developed four-branch model has achieved perfect alignment with the ground truth—demonstrating a 100% event detection rate, no false alarms, and no deviation for both the predicted start times of blockage events and their lasting durations—the ground truth line is omitted from the figure.

Key findings are summarized as follows:

- Both the four-branch and five-branch models successfully detected all 12 spillover/blockage events, as reflected in the 12 matching peaks in Figure 6.10(c).
- The three-branch model detected only 9 out of the 12 events—failing to capture 3 events entirely. Among the 9 detected events, 4 exhibited some delays in detection (highlighted in yellow in Figure 6.10(c), corresponding to Events 1, 6, 7, and 10), implying diminishing sensitivity to upstream congestion due to limited spatial input.

- Despite a lower detection rate (75%), the three-branch model has not produced any false positives.
- The five-branch model, while matching the four-branch model in detection rate (100%), has introduced one false positive prediction (highlighted in red in Figure 6.10(c)). This suggests that incorporating excessive data from more distant upstream intersections may introduce some noise or weakly correlated information, thus increasing the risk of misclassification.

With respect to the average deviation of the predicted blockage start time, the three-branch model exhibits a significant detection lag (27 seconds per event), reflecting its insufficient upstream input to capture the queue dynamics at the target intersection. In contrast, the predicted starting times with the five-branch model are on average 2.5 seconds ahead of the actual starting times, potentially due to overfitting or reliance on weakly relevant long-range input.

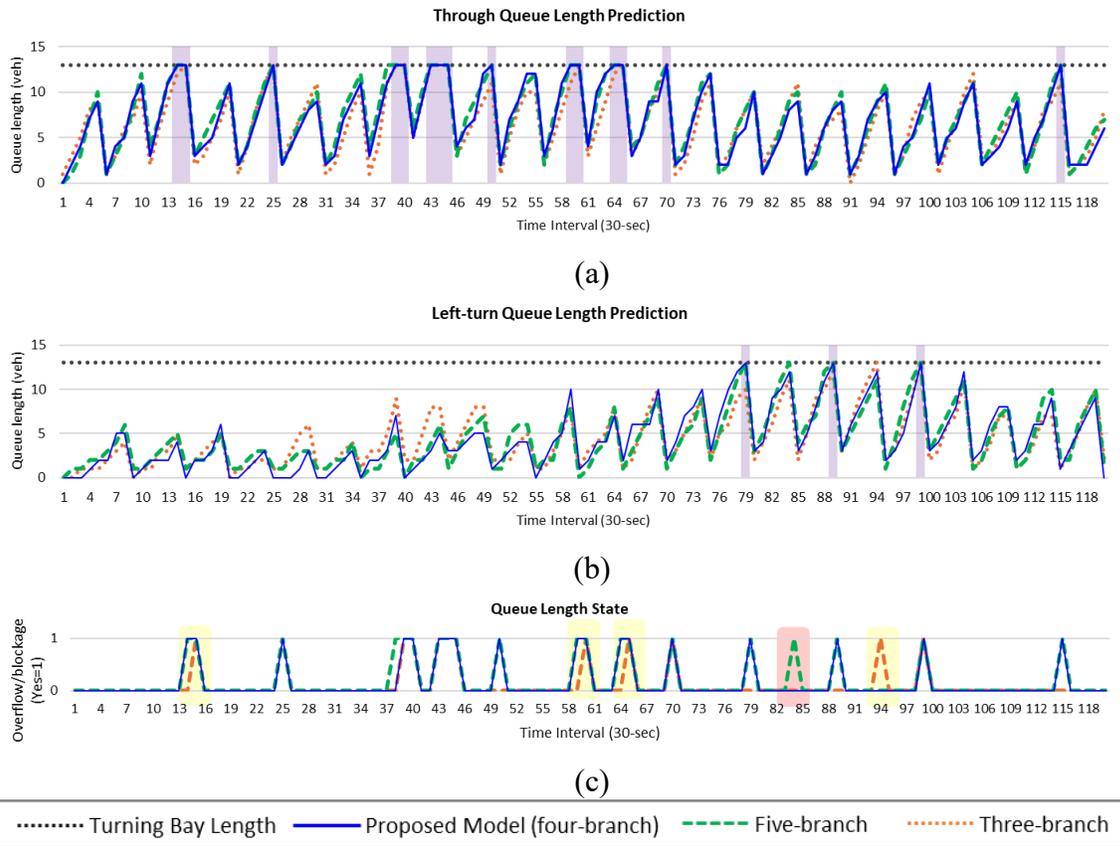


Figure 6.10 Comparison of predicted queue spillover/blockage events using three-, four-, and five-branch LSTM models: (a) through movement prediction; (b) left-turn movement prediction; (c) binary spillover/blockage detection (1: blockage, 0: no blockage)

Spillover/blockage duration accuracy

Figure 6.11 compares the predicted blockage durations across the three models:

- The three-branch model has yielded underestimated duration for 3 out of its 9 detected events (Events 1, 6, and 7, marked with blue dashed boxes), likely due to the absence of sufficient upstream information to capture prolonged queuing patterns.

- In contrast, the five-branch model has produced an overestimated duration for blockage Event 3 (marked with a red dashed box), potentially due to noise incurred by the inclusion of a more distant fifth intersection.

Overall, the average blockage duration with the three-branch model is underestimated by 20 seconds, and overestimated by 2.5 seconds with the five-branch model. Only the four-branch model has yielded the desirable results of a perfect match with the ground truth data for the experimental analyses.

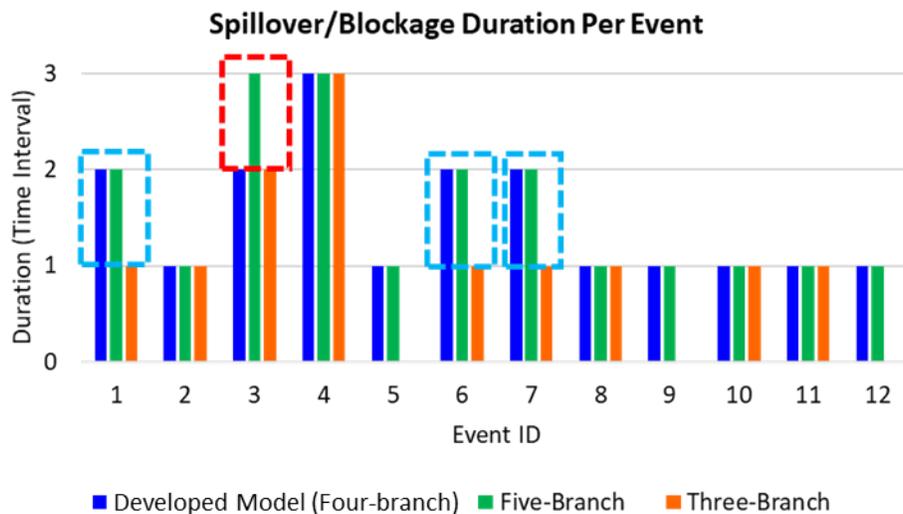


Figure 6.11 Predicted spillover/blockage duration for each event across three-, four-, and five-branch LSTM models

In summary, among the three different model configurations, the four-branch model has achieved the optimal balance of spatial coverage and prediction accuracy. It has consistently detected all blockage events with respect to the accurate onset timing and lasting duration, demonstrating its ability to capture relevant upstream impacts without the burden of excessive data. The five-branch model, despite producing a comparable detection rate, has been interfered by some data noise from the fifth intersection, as evidenced by a false positive prediction and the

slightly overestimated blockage duration. This finding aligns with the Granger causality test results, which showed that the fifth upstream intersection does not contribute significantly to the prediction accuracy. Conversely, the three-branch model, with its limited spatial scope to include critical upstream traffic information, has failed to detect several events and yielded underestimated durations of detected blockage events.

6.3.2 Model 2: Prediction Monitoring and Mode Switching

To ensure the reliability of proactive signal control decisions, Model 2 performs a critical supervisory function by monitoring the accuracy of queue length state predictions generated by Model 1. Even a well-trained predictive model may not produce the desired level of accuracy during some periods of highly volatile traffic conditions, and unexpected disruptions such as incidents. As such, a mechanism is required to continuously assess the predictive accuracy and to determine whether the system should keep operating under proactive control or revert to the responsive mode.

6.3.2.1 Role with the control architecture

Model 2 is situated between the prediction state (i.e., Model 1) and the downstream control models (i.e., Models 3 and 4). It serves as a gatekeeper, ensuring that only reliably predicted queue states are used to trigger the proactive signal control strategies. If significant and sustained deviations are detected between the predicted and the observed queue lengths, the system will temporarily deactivate the proactive control path and redirect the operations back to the responsive control module, which operates with the real-time detected data.

6.3.2.2 Monitoring logic and switching criteria

Given that queue length state predictions are produced at several consecutive time intervals (e.g., every 30 seconds) within a cycle, but control decisions are applied on a cycle-by-cycle basis (e.g., every 150 seconds), the monitoring logic is designed to evaluate the results of multiple predictions within each control cycle.

Let $\hat{Q}_{i,t}^n$ represents the predicted queue length for movement n at intersection i and time step t , and $Q_{i,t}^n$ denotes the corresponding detected (actual) queue length. The normalized absolute prediction error $\epsilon_{i,t}^n$ is defined as:

$$\epsilon_{i,t}^n = \left| \frac{\hat{Q}_{i,t}^n - Q_{i,t}^n}{q_i^n} \right| \quad n \in \{LT, TH\} \quad (6-10)$$

where q_i^n is the maximum physical storage capacity (typically the length of the turning bay or approach lane) for movement n at intersection i . This normalization facilitates fair comparisons across movements and intersections with different geometric configurations.

Monitoring Rules

To ensure reliable signal control, the following three rules are developed to assess the prediction results within each control cycle over a prediction window W , corresponding to the length of one signal cycle.

- **Rule 1: Instantaneous Threshold Violation**

If the prediction error $\epsilon_{i,t}^n$ exceeds a predefined upper bound θ_1 for any movement n at any time step $t \in W$, a warning flag is raised:

$$\exists t \in W: \epsilon_{i,t}^n > \theta_1$$

- **Rule 2: Temporal Violation Count Rule**

Let k be the allowable number of time steps within one cycle where the prediction error may exceed a moderate threshold θ_2 . If this count is exceeded by any movement n , a second violation flag is triggered:

$$\sum_{t \in W} \mathbf{1}(\epsilon_{i,t}^n > \theta_2) > k$$

where $\mathbf{1}(\cdot)$ is the indicator function that returns 1 if the condition is true and 0 otherwise.

- **Rule 3: Consecutive Error Violation Rule**

If the normalized error $\epsilon_{i,t}^n$ exceeds a stricter threshold θ_3 over two or more consecutive time steps, this persistent deviation indicates an ongoing degradation in the prediction quality:

$$\exists t \in W: \epsilon_{i,t}^n > \theta_3 \text{ and } \epsilon_{i,t+1}^n > \theta_3$$

Note that the thresholds are set such that $\theta_1 > \theta_2 > \theta_3$, enabling the module to distinguish between sharp outliers, moderate error frequencies, and persistent deviations, respectively.

Switching Decision Logic

At the end of each control cycle, the module evaluates the results from all three rules. If any two out of the three are violated by any movement n at any monitored intersection i , the system will deactivate the proactive mode and switch to the responsive control module detailed in Chapter 5. This switching mechanism prevents improper adoption of proactive control when the accuracy of the predictive information is not up to an acceptable level.

After back to the responsive mode, the system will continue to monitor the prediction accuracy. If the prediction errors have stabilized (i.e., all three rules are not violated for a predefined number of consecutive cycles), the system may re-activate the proactive operations. The logical flows of this process are illustrated in Figure 6.12.

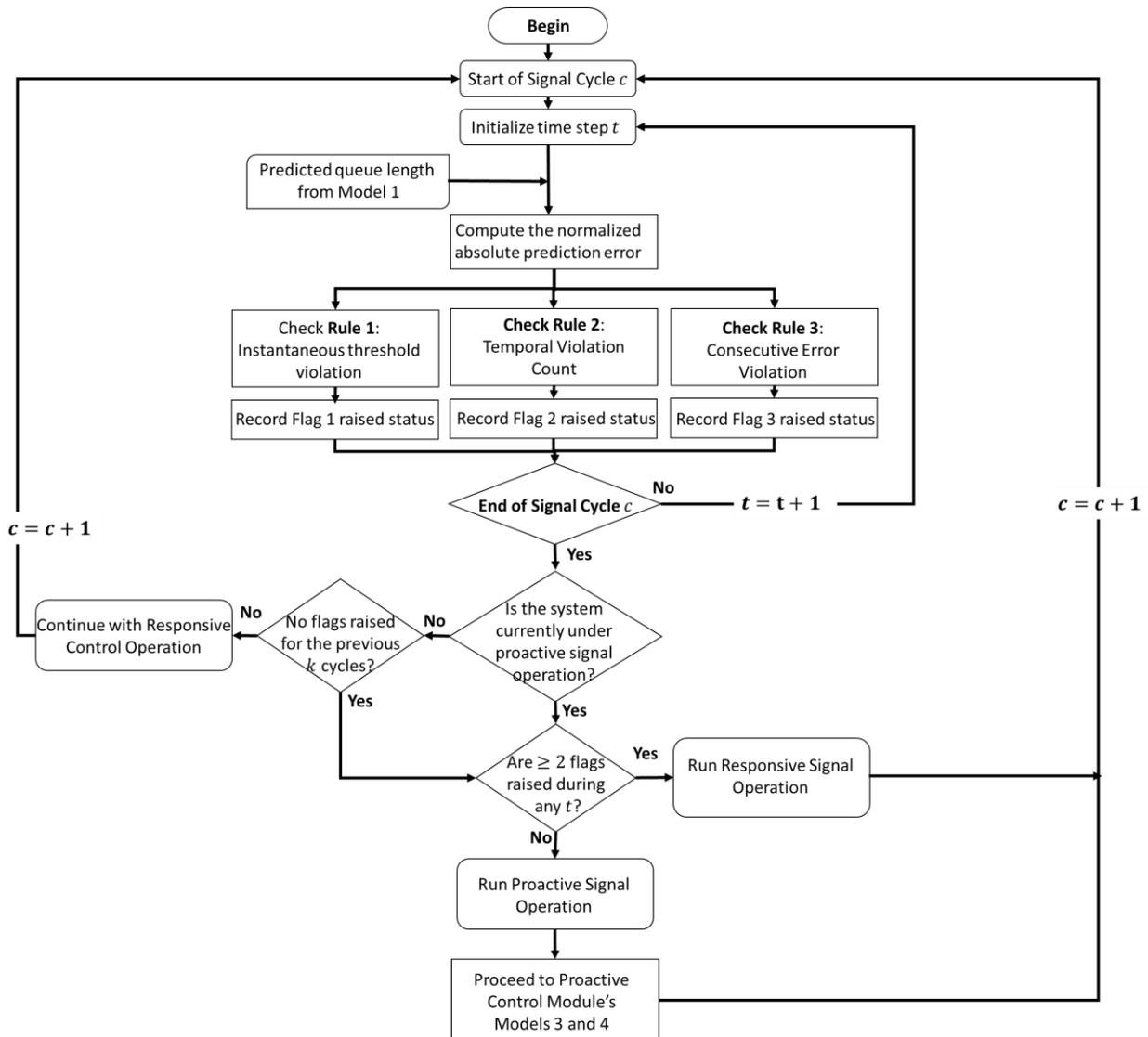


Figure 6.12 Flowchart for prediction monitoring and mode switching logic

6.3.3 Model 3: Congestion Level Assessment

Model 3 is responsible for evaluating the predicted queue blockage patterns and classifying the overall congestion level across the arterial. This classification provides the critical input for determining which proactive signal strategy (Model 4) should be triggered. While structurally consistent with the logic used in the responsive control, this model uniquely operates on predicted instead of detected queue states.

Using the same multi-dimensional rule set introduced in Section 5.3.2, the predicted matrix Λ is evaluated based on spatial, temporal, and spatiotemporal criteria. Each element $\lambda_{i,w} \in \{0,1\}$ represents whether a predicted mutual blockage is expected at intersection i during the w^{th} signal cycle. This binary matrix, constructed over a monitoring window of W cycles and I intersections, encapsulates the evolving queue spillovers and blockage patterns throughout the arterial.

$$\Lambda = \begin{bmatrix} \lambda_{1,1} & \dots & \lambda_{1,W} \\ \vdots & & \vdots \\ \lambda_{i,1} & \ddots & \lambda_{i,W} \\ \vdots & & \vdots \\ \lambda_{I,1} & \dots & \lambda_{I,W} \end{bmatrix}$$

Each column in Λ represents the predicted congestion state across all intersections in a single cycle, while each row reflects the congestion trend over time at a specific location. This structure enables the system to analyze not only the current traffic state but also its spatiotemporal progression and potential propagation.

The module then applies a structured rule set to classify the overall congestion level. These rules span from detecting isolated events at single intersections to identifying persistent congestion propagation across multiple locations and cycles. The classification results help determine whether localized, arterial-wide, or escalating congestion is present—each requiring a tailored control response. Rule-based classification of predicted congestion levels is summarized in Table 6.5.

Table 6.5 Rule-Based Classification of Predicted Congestion Levels

Congestion Level	Condition Description	Key Rule Indicator
Level-N	No predicted mutual blockage across all intersections and cycles	$\sum_{i=1}^I \sum_{w=1}^W \lambda_{i,w} = 0$
Level-1	Fewer than half of intersections exhibit mutual blockages in any cycle	$0 < \sum_{i=1}^I \lambda_{i,w} < 0.5I; \forall w$
Level-2	Two or more consecutive intersections show blockages in a cycle; or persistent blockage over time	$\lambda_{i,w} + \lambda_{i+1,w} = 2; \forall w, i$ $\lambda_{i,w} + \lambda_{i+1,w+1} = 2$ or $\sum_{i=1}^I \lambda_{i,w} \geq 0.5 I \forall w$
Level-3	Blockage propagation across upstream intersections; or consistently high blockage coverage (i.e., Level 2) across cycles	

Once the congestion level is identified, this information is passed to Model 4, which then selects and applies the appropriate proactive signal control strategy. By leveraging both the spatial footprint and the temporal evolution of predicted congestion, the model ensures that the adopted control decisions are aligned with the traffic conditions before the emergence of excessive delays or spillovers.

6.3.4 Model 4: Proactive Strategy Generator

Model 4 serves as the final component within the proactive signal control module. Based on the predicted congestion level identified with Model 3, this model is responsible for selecting and implementing a signal control strategy aimed at preemptively mitigating the predicted traffic congestion.

Once a congestion level (Level-1 through Level-3) is classified, Model 4 will initiate a set of predefined proactive strategies. Each strategy is designed to address the severity and spatial coverage of the projected congestion pattern, with adjustments ranging from localized green

extensions to a full-scale arterial-wide re-optimization. Notice that such strategies are not triggered by observed traffic disruptions, but rather by the predicted queue spillovers and mutual blockages—enabling the system to take preemptive control actions.

The core signal control logic within each proactive strategy mirrors the optimization procedures established in Chapter 4, including reallocation of the green splits, offsets, and phase sequences to reduce the projected delays in the primary commuting direction while preserving arterial coordination in the opposite direction.

To summarize, the control actions to be executed by Model 4 under different congestion levels are as follows:

Table 6.6 Proactive Control Actions Based on Predicted Congestion Level

Congestion Level	Proactive Control Action
Level-N	No adjustment; keep the current signal plan.
Level-1	Apply minor green extensions for selected movements (e.g., left-turn and/or through queues).
Level-2	Re-optimize offsets and phase sequences at intersections within the study scope.
Level-3	Implement the cycle length and green splits adjustments, then re-optimize the offsets and phase sequences.

Should Model 2 determine that the prediction reliability has degraded below the preset level, the proactive strategies in Model 4 will be deactivated, and the system will transfer the control task back to the responsive module and make the decisions based on real-time detected traffic states.

6.4 Case Study

The developed proactive real-time control module has been evaluated with the same simulation platform and network configuration as introduced in Chapter 5. The testbed comprises a highly congested segment of MD 355 Rockville Pike in Bethesda, Maryland, spanning five signalized intersections. As previously described, this arterial corridor is characterized by highly asymmetric

traffic flows during peak hours, and frequent turning-bay overflows as well as mutual queue blockages.

The evaluation is focused on comparing the performance of proactive control and the responsive control modules under identical traffic and roadway scenarios.

To ensure consistency between comparisons, both control strategies were tested over the same one-hour simulation duration, corresponding to 24 signal cycles (150 seconds per cycle). All simulation experiments were conducted using PTV VISSIM, a traffic simulation tool. As previously noted, a uniform increase of 2.55% was applied to each 15-minute interval of the original evening peak-hour volumes, collected by the Maryland State Highway Administration's Internet Traffic Monitoring System (ITMS, as shown in Figure 5.4(b)).

6.4.1 Effectiveness in Preventing Mutual Blockage Events through Proactive Control Module

To assess the benefits of preemptively mitigating mutual queue blockages, this section compares their frequencies and impacts under the proactive and the responsive controls.

Table 6.7 presents the cycle-by-cycle identification of mutual blockage types across the arterial under both the proactive and responsive strategies. Blockage events are categorized based on the blockage types defined in Chapter 4, and the observed pattern per cycle. Key findings from Table 6.7 are as follows:

- **Reduction in Blockage Frequency:** Under the proactive module, most minor blockages observed under the responsive control were successfully prevented. For example, Intersections 1 and 3, which previously experienced intermittent minor blockages, exhibited zero blockage events under the proactive control, highlighting the advantage of early intervention and preventive action with the queue prediction information.

- **Significant Improvement at Critical Intersections:** Intersection 2, which features the shortest turning bay and link lengths, experienced the highest blockage frequency under the responsive strategy, with eight blockage events recorded during the 24-cycle period (i.e., Cycles 13-16, 18-19, 22, and 24). In contrast, with the proactive control, the same intersection encounters only two such blockages (i.e., Cycles 16 and 19), demonstrating the advantage of preventive actions in mitigating the impacts of queue blockages and their mutual interference (see the red box in Table 6.7) on the traffic conditions.
- **Improved Control over Complex Events:** At Intersection 4, a particularly challenging mutual blockage scenario was observed during Cycle 19. Under the responsive control, this event was classified as Blockage Type 3, where the through movement initially blocked the left-turning vehicles, and later in the same cycle, the resulting left-turn queue blocked the through movement. In contrast, under the proactive control, the same cycle experienced only Blockage Type 2, where only the left-turn queue interfered with the through movement. Since Blockage Type 3 involves two sequential blockages within a single cycle, it typically leads to more severe operational disruptions. The proactive control's ability to reduce the event to a single, less complex interaction demonstrates its effectiveness in mitigating the compounding effects of mutual blockages.

These results collectively demonstrate that the proactive control module, informed by queue state prediction, can effectively suppress the formation and escalation of mutual blockages, particularly at intersections with geometric constraints and high imbalance in the two opposite directional flows.

Table 6.7 Blockage Type Observations under Responsive vs. Proactive Control Strategies

Cycle	Intersection 1		Intersection 2		Intersection 3		Intersection 4	
	Responsive Control Module	Proactive Control Module						
1-12	-	-	-	-	-	-	-	-
13	-	-	S_1	-	-	-	-	-
14	S_1	-	S_1	-	-	-	-	-
15	-	-	S_1	-	S_1	-	-	-
16	-	-	S_1	S_1	-	-	-	-
17	-	-	-	-	S_1	-	-	-
18	S_1	-	S_1	-	-	-	S_1	-
19	S_1	-	S_1	S_1	S_1	-	S_3	S_2
20	-	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-	-
22	-	-	S_1	-	-	-	-	-
23	-	-	-	-	-	-	S_1	-
24	-	-	S_1	-	S_1	-	-	-

Note: S_i denotes the observed Blockage type- i ; the cell without any letter represents no mutual blockage observed

6.4.2 Overall Module Performance Evaluation

As shown in Table 6.8, further comparison of these two control modules’ effectiveness includes the average delay and average number of stops per vehicle per hour for through vehicles over the arterial, as well as the average delays per vehicle per hour for through and left-turn vehicles from upstream traffic flows on individual links. Note that the northbound movement represents the major commuting direction during the evening peak period.

The results from the performance comparison confirm the effectiveness of the proactive control module with respect to the selected MOEs. Key findings from the experimental analyses are summarized below:

- The arterial’s northbound delay was significantly reduced under the proactive control, decreasing from 217.01 to 204.12 seconds per vehicle per hour—a 5.94% reduction. This confirms that early detection and adjustment based on predicted queue formation can

meaningfully reduce excessive delays during the peak commuting period. As observed with the responsive control strategy, the southbound direction—representing the non-commuting flow—remained unaffected (0.02% reduction, not statistically significant), demonstrating that prioritizing the control for the dominant traffic direction does not degrade the quality of flows in the opposite direction.

- The average number of stops per hour for northbound through vehicles also showed significant improvement, decreasing by 5.31% (from 6.22 to 5.89 stops). This indicates smoother progression and fewer interruptions caused by the mutual blockages along the commuting arterial, and the benefit of using proactive control to mitigate impending blockages.
- Through-movement delays from upstream approaches showed substantial reductions across all evaluated links, ranging from 7.70% to 12.94%, with the largest improvement observed on the segment approaching Intersection 2—a location known for frequent and severe blockages due to a short turning-bay and link length. These results evidence that the proactive module is highly effective in minimizing congestion where spillovers and queue interactions are most pronounced.
- Left-turning vehicle delays along the northbound approaches also improved under the proactive control strategy, with reductions between 6.98% and 11.44%. These gains highlight the module’s ability to reduce the compounding effects of turning-bay overflows and mutual blockages by acting ahead of the onset of any blockage.
- At the network level, the proactive module achieved a 5.75% reduction in average delay per vehicle per hour (from 124.00 to 116.87 seconds) and a 4.85% decrease in the average number of stops per vehicle per hour (from 3.09 to 2.94). These system-wide

improvements highlight the benefits of prediction-informed control decisions, not only for the congested arterial but also for the entire network.

In summary, the simulation results support the operational advantages of the proposed proactive real-time control module in mitigating delays, improving traffic progression, and minimizing stop frequency. Compared to the responsive control, the proactive module's ability to predict the formation of congestion and adjust signal timing in advance of queue blockage presence can indeed yield additional benefits, particularly in highly congested arterial systems where queue spillbacks and mutual interferences are likely to occur frequently.

Table 6.8 Simulation results for performance comparison

Direction		Proactive real-time control module	Responsive real-time control module	% change	
Average delay of through vehicles along the arterial per vehicle per hour (sec)					
Northbound	⑤→①	204.12	217.01	-5.94%	*
Southbound	①→⑤	141.43	141.46	-0.02%	
Average number of stops of through vehicles along the arterial per vehicle per hour					
Northbound	⑤→①	5.89	6.22	-5.31%	*
Southbound	①→⑤	1.99	1.98	0.51%	
Average delay of through vehicles from all upstream streams on individual links per vehicle per hour (sec)					
Northbound	②→①	22.09	24.19	-8.68%	*
	③→②	40.16	46.13	-12.94%	*
	④→③	37.72	41.98	-10.15%	*
	⑤→④	34.26	37.12	-7.70%	*
Average delay of left-turn vehicles from all upstream streams on individual links per vehicle per hour (sec)					
Northbound	②→①	36.09	38.80	-6.98%	*
	③→②	50.14	56.62	-11.44%	*
	④→③	67.97	75.20	-9.61%	*
	⑤→④	26.84	30.09	-10.80%	*
Average delay per vehicle per hour (sec)					
Network		116.87	124.00	-5.75%	*
Average number of stops per vehicle per hour					
Network		2.94	3.09	-4.85%	*

Note: * indicates the significance level of at least 0.05

6.5 Conclusions

This chapter has introduced and evaluated a proactive real-time arterial signal control module, developed to complement the responsive control presented in Chapter 5. Building upon the same architecture, the proactive module is incorporated with a predictive capability for queue state estimation, enabling the system to take preventative actions before the emergence of excessive queues or mutual blockages. In addition, a monitoring mechanism was developed to ensure the control robustness, allowing the system to revert to detection-based responsive control when the prediction accuracy deteriorates to the predefined level.

The results of performance evaluation with simulated experiments based on a congested segment of MD 355 Rockville Pike in Bethesda, Maryland, have demonstrated the operational advantages of the developed proactive module. Compared to the responsive control module, the proactive approach can more effectively mitigate the formation and propagation of mutual blockages, reduce the overall vehicle delays, and increase the arterial's throughput. Notably, the prediction-driven signal plans are able to preemptively adjust signal timings ahead of the congestion formation, yielding the expected performance under high-demand traffic conditions.

In conclusion, the developed proactive control module can enhance the responsiveness and adaptability of an arterial's signal operations with its early intervention and dynamic switching between prediction-driven and detection-responsive strategies. This dual-mode architecture ensures that the system can maintain its optimal performance across a wide range of real-time traffic conditions on congested commuting arterials.

Chapter 7 Conclusions

7.1 Research Summary

This dissertation was motivated by the persistent and unique challenges in signalized urban arterials during peak commuting periods—particularly for those with highly asymmetric directional flows and short turning bays. Despite significant progress in signal optimization and adaptive systems over the past decades, the traffic community still struggles to effectively cope with mid-block congestion dynamics, cross-lane interferences, and time-varying queue interactions, especially when mutual queue blockages occur between through and turning traffic streams. The research objectives were thus focused on developing an integrated signal control system that can responsively and proactively mitigate such traffic disruptions.

The following developments and contributions have been accomplished in this study to fulfill the research objective:

- **Development of a Time-of-Day Signal Control Module Under the Constraints of Asymmetric Volume Distributions, Short turning bays, and Link Lengths:**

The foundation of this study is a pre-timed signal coordination model specifically designed for arterials experiencing highly directional traffic flows and geometric constraints (i.e., short turning bays). Rather than assuming uniform arrival patterns, this module is formulated to account for the delays at each intersection approach, contributed from all time-varying upstream arriving flows, which can be effectively minimized with the phase sequences and offsets.

To best contend with congestion incurred by excessive queues, this pretimed control module has also been designed with a function to identify and classify various types of mutual traffic blockages, including scenarios where left-turn bays are blocked by through queues and vice versa, and a series of mutual blockages within one signal cycle. The adverse impacts of

such queue blockages include long residual queues, increased queue accumulation speed, excessive vehicle delays, reduced progression band widths for through movements, and green time loss for blocked movements. Hence, estimating the impacts of such excessive congestion patterns on the traffic state at each intersection and along the entire arterial is also one of the module's embedded functions. By accounting for such contributors to delay accumulation and queue formation, the module can offer an effective time-of-day signal control plan for commuting arterials plagued by traffic overflows and congested mutual queue blockages.

- **Design of a Responsive Real-Time Control Module that can Dynamically Implement the Most Effective Signal Control Strategy Based on the Detected Traffic Queue and Blockage Patterns**

To enhance the adaptability under evolving congestion conditions, this study has advanced the pretimed module to responsive real-time control that can dynamically respond to the detected traffic congestion patterns and queue distributions with the most effective signal control plan. Using lane-specific detector data, the responsive module can continuously monitor the temporal-spatial queue evolution over the entire arterial and identify the nature and type of detected mutual blockages, based on the customized rules for congestion severity assessment. The control module, based on the assessment results, will then select the most effective control strategies, such as localized adjustments (e.g., green extensions) to the phase sequences and offsets re-computation. By limiting signal interventions to only those intersections plagued by excessive queue impacts, the responsive control module can effectively mitigate traffic congestion without frequent arterial-wide re-optimization unless necessary, thus ensuring timely and targeted responses to maintain the arterial's overall efficiency.

- **Development of a Proactive Real-Time Control Module with Queue State Prediction**

The third main contribution of this research is the development of a proactive real-time signal control module, which centers on a multi-branch, multi-head Long Short-Term Memory (LSTM) prediction model to perform the forward-looking traffic management over congested commuting arterials. This prediction model is designed to predict upcoming queue states and mutual blockage events for a target intersection by learning spatiotemporal traffic patterns across multiple upstream intersections. It is designed with a unique function for capturing the dynamic lagging impacts of upstream traffic on the downstream traffic conditions, based on distances between intersections, coordination quality, and prevailing congestion levels. The predicted outcomes include the presence, types of mutual blockages, and their estimated durations.

These predicted queue states will then feed into a congestion assessment model and a strategy generator for executing proactive control actions, both adapted from the responsive module but reconfigured to work with the forecasted traffic and queue conditions. Such sequential and integrated operations allow the signal control system to preemptively classify congestion severity and activate control actions such as green splits adjustments or offsets and phase sequences re-optimization prior to the emergence of traffic disruptions. To safeguard the operational reliability, the proactive signal control module features its embedded rule-based monitoring mechanism, which can continuously evaluate the prediction accuracy and trigger a fallback to the responsive control mode when pre-defined quality thresholds are violated. This dual-layered design for real-time control operations can ensure that the proactive strategies are deployed only under reliable conditions, enabling the system to maintain adaptability and stability in contending with dynamically evolving congested patterns over the arterial.

7.2 Future Research

While the developed control system with different modules for different operating environments has demonstrated promising potential in mitigating congestion and mutual blockage impacts on commuting arterials, several technical gaps remain open for future enhancement. For example, despite that the current system can offer effective and robust performance under typical peak-hour conditions and geometric constraints, it may need further refinement to expand its applicability to multimodal traffic environments. Moreover, integration of the developed system with other components of Advanced Traffic Management Systems (ATMS) shall present additional opportunities to advance arterial-wide control efficiency.

7.2.1 System Enhancements

Key research directions to enhance the developed system are listed as follows:

- **Customizing Solution Algorithms for Real-Time Cycle Length Optimization**

The current control structure employs a rule-based approach to incrementally adjust cycle lengths in response to different levels of detected traffic congestion. While this method provides timely and computationally light-weight interventions, it lacks the flexibility and precision of optimization-based approaches, especially under highly heterogeneous traffic conditions. Future system development could incorporate embedded optimization models to dynamically recompute the cycle length and green splits in real time. These models could be formulated as mathematical programming problems (e.g., MILP) or data-driven surrogates using reinforcement learning or metaheuristic algorithms such as genetic algorithms.

- **Multimodal Integration with Transit Flows**

In urban arterials, frequent bus operations introduce intermittent disruptions to the traffic flows due to their stop-and-go movements at curbside stops, dwell times for passenger exchange, and occasional signal priority, all of which can alter lane utilization and impede traffic progression. Accordingly, future research should extend the signal controls to multimodal traffic flows by explicitly modeling transit vehicle interactions within the broader traffic streams. Key issues to be addressed include: 1) how to mathematically characterize and capture the dynamic interactions between transit vehicles and general traffic; and 2) how to quantify the resulting spatial-temporal delays caused by transit operations and integrate them into the signal control logic. Importantly, while the current control structure primarily addresses mutual blockages between through and left-turn movements, the inclusion of transit flows introduces a new class of spatial conflicts—most notably in the rightmost lane—where stopped buses may obstruct both through and right-turning vehicles. Future control strategies must explicitly capture and mitigate these multimodal interference patterns to ensure more equitable and efficient signal operations for all roadway users.

- **Accounting for Non-Recurring Events and Incident Disruptions**

The current system is primarily designed to accommodate recurring traffic patterns. However, traffic conditions can deviate significantly from regular patterns due to non-recurring events such as traffic incidents. These atypical events may induce abrupt changes in queue formation, travel time, and flow distribution. Future research should focus on equipping the control system with the capability to respond to such disruptions in real time. Examples of such tasks may involve: 1) developing anomaly detection models trained to distinguish between recurrent

congestion and incident-induced traffic disturbance; 2) adjusting signal control logic dynamically under non-recurring traffic disruptions.

- **Integration with Connected and Autonomous Vehicle (CAV) Technologies**

As connected and autonomous vehicles become more prevalent, there is a growing opportunity to incorporate vehicle-to-infrastructure (V2I) data into the real-time control system. Future research may explore the use of CAV-based traffic state information—such as precise queue positions and acceleration profiles—to refine the prediction model and enhance the granularity of control decisions. Additionally, proactive control strategies could be adapted to communicate the recommended speed profiles or green-wave advisories to connected vehicles, thus facilitating smoother progression and reducing stop-and-go behaviors.

7.2.2 Potential Applications and Integration with ATMS Systems

Aside from technical enhancements, another vital direction for extension involves integrating the developed signal control system into the broader Advanced Traffic Management System (ATMS) infrastructure. Such integration would enable more coordinated and scalable deployments in real-world traffic networks. Two key application areas where this system may provide the critical inputs and coordination mechanisms are as follows:

- **Queue Warning Systems Integration**

The developed system's fine-grained prediction of queue states can directly support queue warning systems deployed along arterial roadways. By leveraging the proactive module's real-time and short-term forecasts of queue formation, the system can dynamically trigger warning messages on Dynamic Message Signs (DMS) upstream of congested arterial segments. These alerts may include queue length estimations, congestion severity levels, and expected

dissipation times, allowing drivers to take precautionary measures, reduce speeds, or consider alternate routes. Such predictive warnings can significantly reduce the risk of rear-end collisions and improve overall safety in high-density commuting corridors.

- **Conditional Transit Signal Priority (TSP)**

The system's responsive and proactive control logic may also be adapted to support condition-based Transit Signal Priority strategies. Specifically, by incorporating real-time bus detection data—such as vehicle location, schedule adherence, and passenger loading level—into the control framework, the system can dynamically assess whether a priority signal adjustment (e.g., early green or green extension) is warranted. Moreover, congestion state predictions from the proactive module can be used to determine if granting priority would result in disproportionate disruption to general traffic or trigger downstream blockages. This data-informed and context-sensitive TSP logic will offer a balance between improving transit reliability and an arterial's general traffic efficiency. When integrated into ATMS platforms, such enhancements can also facilitate the coordinated multimodal operations across different agencies and jurisdictions.

Bibliography

- Achterberg, T. (2023). What's new in Gurobi 11.0. Webinar Talk URL: <https://www.gurobi.com/events/whats-new-in-gurobi-11-0/>
- Alghamdi, T., Elgazzar, K., Bayoumi, M., Sharaf, T., & Shah, S. (2019, June). Forecasting traffic congestion using ARIMA modeling. In 2019 15th international wireless communications & mobile computing conference (IWCMC) (pp. 1227-1232). IEEE.
- Arshad, J., Muhammad, T. R., Hassan, M. A.-A., Irfan, U. & Muhammad, Z. (2020). Intelligent Intersection control for delay optimization: Using meta-heuristic search algorithms. *Sustainability*, 12(5), 1896.
- Bede, Z., & Péter, T. (2014). Optimal control with the dynamic change of the structure of the road network. *Transport*, 29(1), 36-42.
- Cai, C., Wong, C. K., & Heydecker, B. G. (2009). Adaptive traffic signal control using approximate dynamic programming. *Transportation Research Part C: Emerging Technologies*, 17(5), 456-474.
- Cai, L., Zhang, Z., Yang, J., Yu, Y., Zhou, T., & Qin, J. (2019). A noise-immune Kalman filter for short-term traffic flow forecasting. *Physica A: Statistical Mechanics and its Applications*, 536, 122601.
- Cetin, M., & Comert, G. (2006). Short-term traffic flow prediction with regime switching models. *Transportation Research Record*, 1965(1), 23-31.

- Chang, S. Y., Wu, H. C., & Kao, Y. C. (2023). Tensor extended Kalman filter and its application to traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 24(12), 13813-13829
- Chang, T. H., & Lin, J. T. (2000). Optimal signal timing for an oversaturated intersection. *Transportation Research Part B: Methodological*, 34(6), 471-491.
- Chang, T. H., & Sun, G. Y. (2004). Modeling and optimization of an oversaturated signalized network. *Transportation Research Part B: Methodological*, 38(8), 687-707.
- Chaudhary, N. A., & Messer, C. J. (1993). PASSER IV: A program for optimizing signal timing in grid networks. *Transportation research record*, 1421, 82-93.
- Chen, S., Wang, H., & Meng, Q. (2021). An optimal dynamic lane reversal and traffic control strategy for autonomous vehicles. *IEEE transactions on intelligent transportation systems*, 23(4), 3804-3815.
- Chen, Y. H., Cheng, Y., & Chang, G. L. (2019). Concurrent progression of through and turning movements for arterials experiencing heavy turning flows and bay-length constraints. *Transportation research record*, 2673(9), 525-537.
- Chen, Y. H., Cheng, Y., & Chang, G. L. (2021). Design of an arterial signal progression plan for multi-path flows with only intersection turning counts. *Transportation Research Part C: Emerging Technologies*, 130, 103322.
- Chen, Y. H., Cheng, Y., & Chang, G. L. (2022). Incorporating delay minimization in design of the optimized arterial signal progression. *Transportation research record*, 2676(4), 649-668.

- Chen, Z., Lu, Z., Chen, Q., Zhong, H., Zhang, Y., Xue, J., & Wu, C. (2022). Spatial–temporal short-term traffic flow prediction model based on dynamical-learning graph convolution mechanism. *Information Sciences*, 611, 522-539.
- Christofa, E., Ampountolas, K., & Skabardonis, A. (2016). Arterial traffic signal optimization: A person-based approach. *Transportation Research Part C: Emerging Technologies*, 66, 27-47.
- Cohen, S. L. (1982). Concurrent use of the MAXBAND and TRANSYT signal timing programs for arterial signal optimization. Federal Highway Administration, Urban Transportation Management Division.
- Cohen, S. L., & Liu, C. C. (1986). The bandwidth-constrained TRANSYT signal optimization program. *Transportation Research Record*, 1057, 1-7.
- Synchro Studio 11 (2019). Cubic ITS Trafficware
- Dell, P. A. O. L. O., & Mirchandani, B. (1995). REALBAND: An approach for real-time
- Duan, P., Mao, G., Liang, W., & Zhang, D. (2018). A unified spatio-temporal model for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(9), 3212-3223.
- Feng, X., Ling, X., Zheng, H., Chen, Z., & Xu, Y. (2018). Adaptive multi-kernel SVM with spatial–temporal correlation for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(6), 2001-2013.
- Gartner, N. H. (1990). OPAC: Strategy for demand-responsive decentralized traffic signal control. *IFAC Proceedings Volumes*, 23(2), 241-244.

- Gartner, N. H., Assman, S. F., Lasaga, F., & Hou, D. L. (1991). A multi-band approach to arterial traffic signal optimization. *Transportation Research Part B: Methodological*, 25(1), 55-74.
- Ghasempour, T., & Heydecker, B. (2020). Adaptive railway traffic control using approximate dynamic programming. *Transportation Research Part C: Emerging Technologies*, 113, 91-107.
- Ghosh, B., Basu, B., & O'Mahony, M. (2009). Multivariate short-term traffic flow forecasting using time-series analysis. *IEEE transactions on intelligent transportation systems*, 10(2), 246-254
- Guo, J., Huang, W., & Williams, B. M. (2014). Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transportation Research Part C: Emerging Technologies*, 43, 50-64.
- Hajjema, R., & Hendrix, E. M. (2014). Traffic responsive control of intersections with predicted arrival times: a Markovian approach. *Computer-Aided Civil and Infrastructure Engineering*, 29(2), 123-139.
- Hausknecht, M., Au, T. C., Stone, P., Fajardo, D., & Waller, T. (2011, October). Dynamic lane reversal in traffic management. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (pp. 1929-1934). IEEE.
- Henry, J. J., Farges, J. L., & Tuffal, J. (1984). The PRODYN real time traffic algorithm. In *Control in transportation systems* (pp. 305-310). Pergamon.
- Hunt, P. B., Robertson, D. I., Bretherton, R. D., & Royle, M. C. (1982). The SCOOT on-line traffic signal optimisation technique. *Traffic Engineering & Control*, 23(4).

- Ji, L., & Cheng, W. (2022). Method of Bidirectional Green Wave Coordinated Control for Arterials under Asymmetric Release Mode. *Electronics*, 11(18), 2846.
- Kang, D., Lv, Y., & Chen, Y. Y. (2017, October). Short-term traffic flow prediction with LSTM recurrent neural network. In *2017 IEEE 20th international conference on intelligent transportation systems (ITSC)* (pp. 1-6). IEEE.
- Karlaftis, M. G., and E. I. Vlahogianni. 2011. "Statistical Methods Versus Neural Networks in Transportation Research: Differences, Similarities and Some Insights." *Transportation Research Part C: Emerging Technologies* 19 (3): 387–399
- Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., & Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research Part B: Methodological*, 46(10), 1393-1403.
- Keyvan-Ekbatani, M., Papageorgiou, M., & Papamichail, I. (2013). Urban congestion gating control based on reduced operational network fundamental diagrams. *Transportation Research Part C: Emerging Technologies*, 33, 74-87.
- Kim, S., Hajbabaie, A., Williams, B. M., & Roupail, N. M. (2016). Dynamic bandwidth analysis for coordinated arterial streets. *Journal of Intelligent Transportation Systems*, 20(3), 294-310.
- Kumar, S. V., & Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review*, 7, 1-9.
- Lertworawanich, P., Kuwahara, M., & Miska, M. (2011). A new multiobjective signal optimization for oversaturated networks. *IEEE Transactions on Intelligent Transportation Systems*, 12(4), 967-976.

- Lee, S. S., Lee, S. H., Oh, Y. T., & Choi, K. C. (2002). Development of degree of saturation estimation models for adaptive signal systems. *KSCE Journal of Civil Engineering*, 6, 337-345.
- Li, M., Luo, D., Liu, B., Zhang, X., Liu, Z., & Li, M. (2022). Arterial coordination control optimization based on AM-BAND-PBAND model. *Sustainability*, 14(16), 10065.
- Little, J. D. (1966). The synchronization of traffic signals by mixed-integer linear programming. *Operations Research*, 14(4), 568-594.
- Little, J. D., Kelson, M. D., & Gartner, N. H. (1981). MAXBAND: A versatile program for setting signals on arteries and triangular networks.
- Liu, C. C. (1988). Bandwidth-constrained delay optimization for signal systems. *ITE Journal*, 58(12), 21-26.
- Li, X., Chen, J., & Wang, H. (2013). Study on flow direction changing method of reversible lanes on urban arterial roadways in China. *Procedia-Social and Behavioral Sciences*, 96, 807-816.
- Li, X., Wang, H., & Chen, J. (2013). Dynamic lane-use assignment model at signalized intersections under tidal flow. In *ICTE 2013: Safety, Speediness, Intelligence, Low-Carbon, Innovation* (pp. 2673-2678).
- Liu, Y., & Chang, G. L. (2011). An arterial signal optimization model for intersections experiencing queue spillback and lane blockage. *Transportation research part C: emerging technologies*, 19(1), 130-144.

- Liu, Y., Zheng, H., Feng, X., & Chen, Z. (2017, October). Short-term traffic flow prediction with Conv-LSTM. In 2017 9th international conference on wireless communications and signal processing (WCSP) (pp. 1-6). IEEE.
- Lowrie, P. R. (1982). The Sydney coordinated adaptive traffic (SCAT) system-principles, methodology, algorithm. In Proc. of the Second International Conference on Road Traffic Signaling, IEE (pp. 67-70).
- Luk, J. Y. K., & Sims, A. G. (1982). Selecting Offsets for Sub-area Linkage in SCATS. Australian Road Research, 12(2).
- Lu, K., Tian, X., Jiang, S., Lin, Y., & Zhang, W. (2023). Optimization model of regional green wave coordination control for the coordinated path set. IEEE Transactions on Intelligent Transportation Systems.
- Lu, K., Zeng, X., Li, L., & Xu, J. (2012). Two-way bandwidth maximization model with proration impact factor for unbalanced bandwidth demands. Journal of Transportation Engineering, 138(5), 527-534.
- Luyanda, F., Gettman, D., Head, L., Shelby, S., Bullock, D., & Mirchandani, P. (2003). ACS-Lite algorithmic architecture: Applying adaptive control system technology to closed-loop traffic signal control systems. Transportation Research Record, 1856(1), 175-184.
- Lv, Y., Y. Duan, W. Kang, Z. Li, and F. Y. Wang. 2015. "Traffic Flow Prediction with big Data: A Deep Learning Approach." IEEE Trans. Intelligent Transportation Systems 16 (2): 865–873.

- Ma, X., Z. Tao, Y. Wang, H. Yu, and Y. Wang. 2015. "Long Short-Term Memory Neural Network for Traffic Speed Prediction Using Remote Microwave Sensor Data." *Transportation Research Part C: Emerging Technologies* 54: 187–197.
- Mackenzie, J., Roddick, J. F., & Zito, R. (2018). An evaluation of HTM and LSTM for short-term arterial traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(5), 1847-1857.
- Mao, L., Li, W., Hu, P., Zhou, G., Zhang, H., & Dai, J. (2020). Design of real-time dynamic reversible lane in intelligent cooperative vehicle infrastructure system. *Journal of Advanced Transportation*, 2020.
- McCann, B. (2014, September). A review of SCATS operation and deployment in Dublin. In *Proceedings of the 19th JCT traffic signal symposium & exhibition*.
- McTrans. (2008). "TRANSYT-7F Users Guide", University of Florida.
- Messer, C. J., Whitson, R. H., Dudek, C. L., & Romano, E. J. (1973). A variable-sequence multiphase progression optimization program. *Highway Research Record*, 445(1973), 24-33.
- Mirchandani, P., & Head, L. (2001). A real-time traffic signal control system: architecture, algorithms, and analysis. *Transportation Research Part C: Emerging Technologies*, 9(6), 415-432.
- Morgan, J. T., & Little, J. D. (1964). Synchronizing traffic signals for maximal bandwidth. *Operations Research*, 12(6), 896-912.

- Mou, Luntian, Pengfei Zhao, Haitao Xie, and Yanyan Chen. "T-LSTM: A long short-term memory neural network enhanced by temporal information for traffic flow prediction." *Ieee Access* 7 (2019): 98053-98060.
- Narmadha, S., & Vijayakumar, V. (2023). Spatio-Temporal vehicle traffic flow prediction using multivariate CNN and LSTM model. *Materials today: proceedings*, 81, 826-833.
- Noaen, M., Mohajerpoor, R., Far, B. H., & Ramezani, M. (2021). Real-time decentralized traffic signal control for congested urban networks considering queue spillbacks. *Transportation research part C: emerging technologies*, 133, 103407.
- Ojeda, L. L., Kibangou, A. Y., & De Wit, C. C. (2013, June). Adaptive Kalman filtering for multi-step ahead traffic flow prediction. In *2013 American Control Conference* (pp. 4724-4729). IEEE.
- Peng, X., & Wang, H. (2023). Coordinated control model for arterials with asymmetric traffic. *Journal of Intelligent Transportation Systems*, 27(6), 752-768.
- Qiao, S., Sun, R., Fan, G., & Liu, J. (2017, November). Short-term traffic flow forecast based on parallel long short-term memory neural network. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)* (pp. 253-257). IEEE.
- Quinn, D. J. (1992). A review of queue management strategies. *Traffic Engineering+ Control*, 33(11), 600-5.
- Rahman, R., & Hasan, S. (2021). Real-time signal queue length prediction using long short-term memory neural network. *Neural Computing and Applications*, 33, 3311-3324.

- Redhu, P., & Kumar, K. (2023). Short-term traffic flow prediction based on optimized deep learning neural network: PSO-Bi-LSTM. *Physica A: Statistical Mechanics and its Applications*, 625, 129001.
- Robertson, D. I. (1969). *TRANSYT: a traffic network study tool*.
- Roger P. Roess, S. Elena Prassas, and William R. McShane, *Traffic Engineering*. 3rd Edition, Pearson Prentice Hall, 2004, ISBN 0-13-142471-8)
- Shao, H., & Soong, B. H. (2016, November). Traffic flow prediction with long short-term memory networks (LSTMs). In *2016 IEEE region 10 conference (TENCON)* (pp. 2986-2989).
- Shepherd, S. (1994). Traffic control in over-saturated conditions. *Transport Reviews*, 14(1), 13-43.
- Shih, C. Y., Chang, C. M., Wu, B. F., Chang, C. H., & Hwang, F. N. (2024). Data-driven numerical simulation with extended Kalman filtering and long short-term memory networks for highway traffic flow prediction. *Journal of Mechanics*, 40, 31-43.
- Shu, W., Cai, K., & Xiong, N. N. (2021). A short-term traffic flow prediction model based on an improved gate recurrent unit neural network. *IEEE Transactions on Intelligent Transportation Systems*, 23(9), 16654-16665.
- Sims, A. G., & Dobinson, K. W. (1980). The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits. *IEEE Transactions on vehicular technology*, 29(2), 130-137.
- Skabardonis, A., & May, A. D. (1985). Comparative analysis of computer models for arterial signal timing. *Transportation Research Record*, 1021, 45-52.
- Smith, S., Barlow, G., Xie, X. F., & Rubinstein, Z. (2013, June). Smart urban signal networks: Initial application of the surtrac adaptive traffic signal control system. In *Proceedings of*

- the International Conference on Automated Planning and Scheduling (Vol. 23, pp. 434-442).
- Stevanovic, A., P. T. Martin, and J. Stevanovic. VisSim-based genetic algorithm optimization of signal timings. *Transportation Research Record*, 2007. 2035(1), 59-68.
- Sun, W., Wang, Y., Yu, G., & Liu, H. X. (2015). Quasi-optimal feedback control for a system of oversaturated intersections. *Transportation Research Part C: Emerging Technologies*, 57, 224-240.
- Tan, M. C., Wong, S. C., Xu, J. M., Guan, Z. R., & Zhang, P. (2009). An aggregation approach to short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 10(1), 60-69.
- Tchrakian, T. T., Basu, B., & O'Mahony, M. (2011). Real-time traffic flow forecasting using spectral analysis. *IEEE Transactions on Intelligent Transportation Systems*, 13(2), 519-526.
- Tian, Y., & Pan, L. (2015, December). Predicting short-term traffic flow by long short-term memory recurrent neural network. In *2015 IEEE international conference on smart city/SocialCom/SustainCom (SmartCity)* (pp. 153-158).
- Tian, Y., Zhang, K., Li, J., Lin, X., & Yang, B. (2018). LSTM-based traffic flow prediction with missing data. *Neurocomputing*, 318, 297-305.
- Van Hinsbergen, C. P., Schreiter, T., Zuurbier, F. S., Van Lint, J. W. C., & Van Zuylen, H. J. (2011). Localized extended kalman filter for scalable real-time traffic state estimation. *IEEE transactions on intelligent transportation systems*, 13(1), 385-394.
- Varaiya, P. (2013). Max pressure control of a network of signalized intersections. *Transportation Research Part C: Emerging Technologies*, 36, 177-195.

- Wallace, C. E., & Courage, K. G. (1982). Arterial progression-New design approach. *Transportation Research Record*, 881, 53-59.
- Wallace, C. E., Courage, K. G., Reaves, D. P., Schoene, G. W., & Euler, G. W. (1984). TRANSYT-7F user's manual (No. UF-TRC-U32 FP-06/07).
- Williams, B. M. (2001). Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling. *Transportation Research Record*, 1776(1), 194-200.
- Williams, B. M., Durvasula, P. K., & Brown, D. E. (1998). Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record*, 1644(1), 132-141.
- Wolshon, B., & Lambert, L. (2006). Reversible lane systems: Synthesis of practice. *Journal of Transportation Engineering*, 132(12), 933-944.
- Wong, S. C. (1996). Group-based optimisation of signal timings using the TRANSYT traffic model. *Transportation Research Part B: Methodological*, 30(3), 217-244.
- Wong, S. C., Wong, W. T., Leung, C. M., & Tong, C. O. (2002). Group-based optimization of a time-dependent TRANSYT traffic model for area traffic control. *Transportation Research Part B: Methodological*, 36(4), 291-312.
- Xiao, Y., & Yin, Y. (2019). Hybrid LSTM neural network for short-term traffic flow prediction. *Information*, 10(3), 105.
- Xie, X. F., Smith, S. F., Lu, L., & Barlow, G. J. (2012). Schedule-driven intersection control. *Transportation Research Part C: Emerging Technologies*, 24, 168-189.

- Xie, Y., Zhang, Y., & Ye, Z. (2007). Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition. *Computer-Aided Civil and Infrastructure Engineering*, 22(5), 326-334.
- Yang, D., Li, S., Peng, Z., Wang, P., Wang, J., & Yang, H. (2019). MF-CNN: traffic flow prediction using convolutional neural network and multi-features fusion. *IEICE TRANSACTIONS on Information and Systems*, 102(8), 1526-1536
- Yang, I., & Jayakrishnan, R. (2015). Real-time network-wide traffic signal optimization considering long-term green ratios based on expected route flows. *Transportation Research Part C: Emerging Technologies*, 60, 241-257.
- Yang, X., Cheng, Y., & Chang, G. L. (2015). A multi-path progression model for synchronization of arterial traffic signals. *Transportation research part C: emerging technologies*, 53, 93-111.
- Yao, R., Zhang, X., Wu, N., & Song, X. (2018). Modeling and control of variable approach lanes on an arterial road: A case study of Dalian. *Canadian Journal of Civil Engineering*, 45(11), 986-1003.
- Yao, T., Zhang, C., Zhao, J., Gupta, A., & Mondal, S. (2023). Adaptive signal control for overflow prevention at isolated intersections based on fuzzy control. *Transportation research record*, 2677(5), 1387-1401.
- Yu, C., Feng, Y., Liu, H. X., Ma, W., & Yang, X. (2018). Integrated optimization of traffic signals and vehicle trajectories at isolated urban intersections. *Transportation research part B: methodological*, 112, 89-112.

- Yu, H., Wu, Z., Wang, S., Wang, Y., & Ma, X. (2017). Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors*, 17(7), 1501.
- Yun, I. (2006, December). Application of stochastic optimization method for an urban corridor. In *Proceedings of the 2006 Winter Simulation Conference* (pp. 1493-1499). IEEE.
- Yu, Q., & Tian, R. (2014). Research on reversal lane application method of urban road network based on the bi-level programming. In *Practical Applications of Intelligent Systems: Proceedings of the Eighth International Conference on Intelligent Systems and Knowledge Engineering, Shenzhen, China, Nov 2013 (ISKE 2013)* (pp. 983-992). Springer Berlin Heidelberg.
- Zhang, C., Xie, Y., Gartner, N. H., Stamatidis, C., & Arsava, T. (2015). AM-band: an asymmetrical multi-band model for arterial traffic signal coordination. *Transportation Research Part C: Emerging Technologies*, 58, 515-531.
- Zhang, W., Yu, Y., Qi, Y., Shu, F., & Wang, Y. (2019). Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. *Transportmetrica A: Transport Science*, 15(2), 1688-1711.
- Zhang, Y., Zhang, Y., & Haghani, A. (2014). A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. *Transportation Research Part C: Emerging Technologies*, 43, 65-78
- Zhao, J., Yao, T., Zhang, C., & Shafique, M. A. (2024). Signal control for overflow prevention at intersections using partial connected vehicle data. *Transportmetrica A: Transport Science*, 1-31.

- Zeng, B., & Li, C. (2018). Improved multi-variable grey forecasting model with a dynamic background-value coefficient and its application. *Computers & Industrial Engineering*, 118, 278-290.
- Zheng, L., Feng, M., Yang, X., & Xue, X. (2023). Stochastic simulation-based optimization method for arterial traffic signal coordination with equity and efficiency consideration. *IET Intelligent Transport Systems*, 17(2), 373-385.
- Zheng, L., & Li, X. (2023). Simulation-based optimization method for arterial signal control considering traffic safety and efficiency under uncertainties. *Computer-Aided Civil and Infrastructure Engineering*, 38(5), 640-659.
- Zheng, L., Rao, H., Pu, S., Jiang, W., Hao, Y., Li, W., ... & Zhang, W. (2023, September). Traffic State Identification Based on Graph Neural Networks: Taking Tidal Traffic as an Example. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 2157-2162). IEEE.
- Zheng, X., & Recker, W. (2013). An adaptive control algorithm for traffic-actuated signals. *Transportation Research Part C: Emerging Technologies*, 30, 93-115.
- Zhou, G., Mao, L., Bao, T., Dai, J., & Bao, X. (2023). Construction of real-time dynamic reversible lane safety control model in intelligent vehicle infrastructure cooperative system. *Applied Artificial Intelligence*, 37(1), 2177009.
- Zhou, H., Hawkins Jr, H. G., & Zhang, Y. (2017). Arterial signal coordination with uneven double cycling. *Transportation Research Part A: Policy and Practice*, 103, 409-429.

Zhou, T., Han, G., Xu, X., Han, C., Huang, Y., & Qin, J. (2019). A learning-based multimodel integrated framework for dynamic traffic flow forecasting. *Neural Processing Letters*, 49, 407-430.

Zhou, T., Jiang, D., Lin, Z., Han, G., Xu, X., & Qin, J. (2019). Hybrid dual Kalman filtering model for short-term traffic flow forecasting. *IET Intelligent Transport Systems*, 13(6), 1023-1032.