STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

ADA Notice

For individuals with sensory disabilities, this document is available in alternate formats. For information call (916) 654-6410 or TDD (916) 654-3880 or write Records and Forms Management, 1120 N Street, MS-89, Sacramento, CA 95814.

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA17-2912		

4. TITLE AND SUBTITLE	5. REPORT DATE
CTM-based optimal signal control strategies in urban networks	June 2, 2017
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR	8. PERFORMING ORGANIZATION REPORT NO.
Wen-Long Jin and Qijian Gan; UCI and UCTC	UCTC

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NUMBER
The Regents of the University of California, Berkeley Sponsored Projects Office 2150 Shattuck Ave, Ste #300 Berkeley, CA 94704	11. CONTRACT OR GRANT NUMBER 65A0528 TO 013 Task 2912

12. SPONSORING AGENCY AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
California Department of Transportation Division of Procurements and Contracts 1737 30th St. MS#65 Sacramento, CA 95816	Final Report March 1, 2015 to January 30, 2017
	14. SPONSORING AGENCY CODE Caltrans

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This research introduces a novel analytical framework in deriving invariant averaged models for signalized intersections in urban networks, using the capability of the cell transmission model (CTM) to capture the detailed traffic dynamics such as the formation, propagation, and dissipation of congestion arising at network junctions. Generally, the CTM formulates the optimization problem as a mixed-integer-linear-programming (MILP) problem, which introduce many binary variables for large-scale urban networks and is difficult to solve. The approach aims to derive invariant averaged models to eliminate the binary variables introduced by the traffic signals. For the purpose of simplicity, the approach emphasizes on a signalized linear junction connecting one upstream link with one downstream link.

Using the Cell Transmission Model (CTM) simulation on a signalized ring road, we demonstrate that the invariant averaged model is a reasonable approximation to the original supply-demand model with binary signals.

Due to the existence of merging behaviors, we introduce two new terms while deriving the averaged model: Effective Demand and Merging Priority. With these two new terms, we follow similar procedures as those in the linear junction, and derive the corresponding invariant averaged model for the merging junction. We further show that the derived averaged model for the signalized linear junction is just one special case of the one for the signalized merging junction with empty demand in one of the upstream links.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Macroscopic Fundamental Diagram, Traffic Flow Models, Junction Models, Averaged models, Signal Control Strategy, Binary Variables, Signalized Junctions, Effective Demand and Merging Priority.	

19. SECURITY CLASSIFICATION *(of this report)*		21. COST OF REPORT CHARGED
Unclassified	97	N/A

Reproduction of completed page authorized.

DISCLAIMER STATEMENT

This document is disseminated in the interest of information exchange. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This publication does not constitute a standard, specification or regulation. This report does not constitute an endorsement by the Department of any product described herein.

For individuals with sensory disabilities, this document is available in alternate formats. For information, call (916) 654-8899, TTY 711, or write to California Department of Transportation, Division of Research, Innovation and System Information, MS-83, P.O. Box 942873, Sacramento, CA 94273-0001.

UC Transportation Center

CTM-based optimal signal control strategies in urban networks

Final Report

Wen-Long Jin, Ph.D. (UC Irvine)
Qijian Gan, Ph.D. (UC Irvine)

Sponsored by

Caltrans

2017 Berkeley *UC Davis* UC Irvine UCLA UC Riverside *UC Santa Barbar*

CTM-based optimal signal control strategies in urban networks

Friday, June 2, 2017

A report to the

University of California Transportation Center (UCTC)

Wen-Long Jin, Ph.D. (UC Irvine)

Qijian Gan, Ph.D. (UC Irvine)

Institute of Transportation Studies

California Institute for Telecommunications and Information Technology

Department of Civil and Environmental Engineering

University of California, Irvine

Irvine, CA 92697-3600

Executive Summary

Due to the rapid increase in travel demands, traffic in recent years is getting more and more congested (Papageorgiou et al., 2003). Congestion arising during peak hours has wasted travelers a massive amount of time, fuel, and money, and have caused a number of environmental and health problems (Schrank et al., 2012). To relieve traffic congestion, one of the primary recommendations is to upgrade the signal timing and control strategies where the current technology is deficient (Sorensen et al., 2008).

In reality, a number of control strategies have been developed to optimize the signal settings at urban intersections. For isolated intersections, examples of fixed-time signal control are SIGSET (Allsop, 1971a,b) and SIGCAP (Allsop, 1972, 1976), of which Webster's formula (Webster, 1958) is often used to calculate the traffic delay. Also, several other control strategies, such as the Vehicle Interval and Volume Density strategies (De la Breteque and Jezequel, 1979), have been designed for the traffic-responsive case. To optimize the signal settings of a set of intersections, examples of fixed-time coordination signal control can be found in MAXBAND (Little, 1966), MULTI-BAND (Gartner et al., 1991), and TRANSYT (Robertson, 1969). Furthermore, methods like SCOOT (Hunt et al., 1981, 1982) and OPAC (Gartner, 1983) have been proposed for the traffic-responsive case. However, as mentioned in (Papageorgiou et al., 2003), most of the prevailing signal control strategies were developed for unsaturated conditions, which as a result, may not work properly when traffic gets congested. In the literature, some efforts have been devoted to developing signal control strategies for oversaturated intersections. One of the examples is the store-and-forward model in (Gazis and Potts, 1963; Gazis 1964; D'ans and Gazis, 1976; Aboudolas et al., 2007, 2009). One of the key features in the store-and-forward model is that when the upstream demands are high, the cumulative junction outflow is approximated as a continuous time-dependent function with a slope of its averaged flow-rate. However, such an approximation is not reliable when the traffic demands are low or when downstream queue spillback exists. To the best of our knowledge, few traffic flow models used in existing signal control studies can work under various types of traffic conditions.

To model the traffic dynamics on road links, the LWR model in (Lighthill and Whitham, 1955; Richards, 1956), is widely used. However, the LWR model is a Partial Deferential Equation (PDE) and is

difficult to solve analytically. Therefore, in (Daganzo, 1994, 1995), the Cell Transmission Model (CTM) was proposed to numerically solve the LWR model using the Godunov method. Many following studies have shown that the CTM is able to capture detailed traffic dynamics such as the formation, propagation, and dissipation of traffic congestion arising on road links. Besides the CTM, there also exist other traffic flow models, e.g., the Link Transmission Model in (Yperman, 2007; Jin, 2014b), the Link Queue Model in (Jin, 2012b), and the Vertical Cell Model (Anderson et al., 2015). Since the CTM is the most popular one, we will use it as the traffic flow model in this project.

Earlier in (Lo, 1999; Lo et al., 2001), the CTM was introduced to serve as a traffic flow model in the optimization of signal settings in urban networks. Since then, a lot of improvements have proposed to enhance the urban signal design with the CTM: to model the platoon dispersion at signalized intersections in (Feldman and Maher, 2002), to develop junction models to take into account the merging and diverging behaviors in (Almasri and Friedrich, 2005; Su et al, 2013), to model more complicated signal settings in (Zhang et al., 2013), and to introduce the concept of sub-zones and sub-cells in (Li, 2010; Gao et al., 2015). In fact, in the literature, there have been a lot of junction models proposed for freeway junctions; examples are those in (Daganzo, 1995; Lebacque, 1996; Jin and Zhang, 2003; Jin et al., 2009; Jin, 2010, 2014a, 2012a). However, in order to mimic the cyclic pattern of signal control, binary variables are often introduced in the optimization problem. Due to the increasing number of binary variables in large-scale networks, the optimization problem becomes difficult to solve and thus, heuristic methods like the genetic algorithm are often used. Therefore, it is necessary to introduce new models into signalized intersections to reduce or eliminate the binary variables.

As one of the earliest attempts, the store-and-forward model in (Gazis and Potts, 1963; Gazis, 1964; D'ans and Gazis, 1976; Aboudolas et al., 2007, 2009) tries to simply the traffic dynamics at signalized intersections using a continuous averaged function under oversaturated conditions. However, its basic assumptions may not work under unsaturated conditions or when queue spillback from the downstream section occurs. Recently, in (Han et al., 2014), an averaged model with binary signal at a merging junction was proposed, and its approximation accuracy was analyzed under different combinations of traffic

conditions and traffic flow fundamental diagrams. However, this model is proposed heuristically, and there is no way to justify its correctness.

Therefore, in this project, we devote our efforts to developing a generic analytical framework to derive averaged models for signalized intersections. Using the supply-demand framework (Daganzo, 1995; Lebacque, 1996), we first propose three models with binary signals. The signal control is applied to the following three cases: (i) to both the upstream demands and the downstream supplies; (ii) to the upstream demands only; and (iii) to the downstream supplies only. These models can be applied to the CTM to calculate boundary fluxes from upstream demands, downstream supplies, turning proportions, and signal settings. They also can be extended to different types of junction with various merging and diverging behaviors (Jin and Zhang, 2003; Jin, 2010, 2014a, 2012a).

For a signalized linear junction, we analytically show that these three models are equivalent to each other when the binary signal is used, and it doesn't matter where to put the signal control to. We derive the averaged models by replacing the cyclic signal control with an averaged value, which is the effective green ratio. We apply them to the CTM for simulations, and find that they return inconsistent results under different traffic conditions. Therefore, we apply their local forms as entropy conditions to the signalized linear junction, and solve the arising Riemann problems with the framework of (Jin et al., 2009). After that, we derive their invariant forms. With the constraint of maximum average junction flux, we identify that only one of them is correct, and the other two are wrong since they fail to catch either the upstream or the downstream capacity constraint. That means invariance does not necessarily guarantee correctness. Furthermore, we find that different non-invariant averaged models can lead to the same invariant form, which shows the importance of deriving invariant models from their non-invariant forms. Using the CTM simulation in a signalized ring road, we analyze the approximation accuracy of the correct invariant averaged model under different settings of initial conditions, signal settings, and fundamental diagrams. We find that the invariant averaged model is a reasonable approximation to the original supply-demand model with binary signals, and its approximation accuracy is not sensitive to the types of fundamental diagrams but will degrade with long cycle lengths.

As a further extension, we apply the analytical framework to more complicated cases, e.g., a signalized merging junction. We first propose a supply-demand model with binary signals, in which the signal control is applied to both the upstream demands and downstream supplies. Then we obtain the corresponding averaged model by replacing the cyclic signal control at each upstream link with its effective green ratio. Different from the linear junction, merging behavior should be taken into account. Therefore, we introduce two new definitions: Effective Demand and Merging Priority. With these two new definitions, we follow the same procedures as those in the linear junction, and derive the invariant form. We further verify that the derived invariant model for the signalized linear junction is just a special case of the one for the signalized merging junction with zero demand in one of the upstream links.

In the future, we will introduce the developed analytical framework to more general signalized intersections, e.g., four-way intersections. In this case, more complicated driver's behaviors should be considered: merging and diverging. And also, with more upstream and downstream links, the derivation of invariant models will be more difficult since more combinations of traffic initial conditions should be considered. However, this work is very important for the following research tasks. First, once the derivation of invariant averaged models is done, we can combine the modeling of freeway and urban networks as a whole since the cyclic pattern of signal control no long exists. In such a case, we can run large-scale network simulations more efficiently within the CTM. Second, we can extend our study of network stationary states to larger urban networks since traffic dynamics at the signalized junction is now averaged over time. The modeling difficulty is significantly reduced. Third, since the development of the averaged models significantly reduced the number of binary variables in modeling the signalized junctions, it is possible to develop optimal signal settings for large-scale networks more efficiently. Fourth, on the planning side, we can fundamentally change the traffic model in the procedure of traffic assignment. We apply the invariant averaged model together with the prevailing traffic flow models as the basic simulation models to obtain more realistic estimations of travel times, queues, and etc.

Contents

Tables and Figures 4
Acknowledgements 5
About the Authors 6
Abstract 8
Part I: Introduction 10
PART II: Literature Review 17
1. Traffic flow models for urban networks 17
1.1 Cell transmission model 17
1.2 Other models for urban traffic flow 19
1.2.1 Store-and-forward model 19
1.2.2 Link transmission model 21
1.2.3 Link queue model 24
1.2.4 Vertical cell model 26
2. Traditional signal control methods 27
2.1 For isolated intersections 27
2.2 For coordinated intersections 29
3. CTM-based network representations and junction models at signalized intersections 30
3.1 Simple one-way streets 30
3.2 One-way streets with merging and diverging behaviors 31
3.3 A four-way roundabout 33
3.4 General road networks 33
3.4.1 One dimensional cell representation 33
3.4.2 Two dimensional cell representation: cells and sub-cells 38
4. CTM-based signal control strategies 42
4.1 Objective functions 42
4.1.1 Delay minimization 42
4.1.2 Minimization of Performance Index 43
4.1.3 Maximization of system throughput 44

4.1.4 Minimization of a combination of delay and early arrival flow 44
4.1.5 Minimization of mean excess delay 44
4.2 Optimization methods 45
4.2.1 Mixed-integer programing problems 45
4.2.2 Genetic Algorithms 45
4.2.3 Linear relaxation combined with heuristic methods 47
4.2.4 The hill-climbing method 47
PART III: Average models for signalized junctions 49
1. Models with binary signals and the averaged counterparts at a linear junction 49
1.1 Three models with binary signals 50
1.2 Averaged models and their properties 52
1.2.1 Averaged models 52
1.2.2 Properties 53
2. Invariant average models 55
2.1 Network kinematic wave model 55
2.1.1 Link models 55
2.1.2 Junction models 56
2.2 Derivation of invariant averaged models 56
2.2.1 Riemann Problems 56
2.2.2 Solutions 57
3. Comparison of averaged models 59
3.1 Constraint of maximum average junction flux 59
3.2 Another average model 62
4. Approximation accuracy on a signalized ring road 63
4.1 Impacts of initial densities and cycle lengths 64
4.2 Differences in the macroscopic fundamental diagram 67
5. Extension to a signalized merging junction 68
5.1 Model of binary signals at the merging junction and its averaged counterpart 69
5.1.1 Model with binary signals 69
5.1.2 The averaged counterpart 70
5.2 Invariant average model for the signalized merging junction 71
5.2.1 Riemann problems 71
5.2.2 Solutions 72

PART IV: Conclusions and future research directions 76
PART V: Appendices 80
Appendix A 80
Appendix B 82
PART VI: References 85

Tables and Figures

Figures

Figure 1 Cell representation inside a regular link 18
Figure 2 Arrival and discharging patterns at one intersection with two one-way streets. 21
Figure 3 Two types of signalized junctions: (a) Merging and (b) Diverging. 32
Figure 4 Three types of signalized junctions. 34
Figure 5 NEMA phase structure at a four-way intersection. 36
Figure 6 Link representation at a one-way road section. 38
Figure 7 Four zones at a signalized intersection approach. 39
Figure 8 Sub-cell representation of a diverging cell. 39
Figure 9 Sub-cell representation of a signalized intersection. 41
Figure 10 A signalized road link. 50
Figure 11 Properties of the three time-average models in Equation (69). 54
Figure 12 Solutions of stationary states for the three invariant averaged models. 62
Figure 13 A signalized ring road. 64
Figure 14 Junction fluxes with the same cycle length but different initial densities. 65
Figure 15 Junction fluxes with the same initial density but different cycle lengths. 66
Figure 16 Macroscopic fundamental diagrams of the signalized ring road 68
Figure 17 A signalized merging junction. 69

Acknowledgements

The authors would like to thank UCTC and Caltrans for the funding support of this research.

About the Authors

Qijian Gan received the Engineering Degree in Automatic Control from the University of Science and Technology of China in 2009, the M.S. and Ph.D. in Civil Engineering from the University of California, Irvine, in 2010 and 2014 respectively. He is now a Postdoctoral Researcher in the California PATH program at UC Berkeley. Prior to this position, he worked as an assistant project scientist in the Institute of Transportation Studies at UC Irvine. His main expertise includes network traffic flow theory, network modeling and simulation, traffic signal control, and data analysis.

Wenlong Jin received the Engineering Degree in Automatic Control from the University of Science and Technology of China in 1998, the M.S. and Ph.D in Applied Mathematics from the University of California, Davis, in 2000 and 2003 respectively. He is now an Associate Professor at the University of California, Irvine. Prior to this position, he was a Postdoctoral Researcher at the University of California, Irvine from 2003 to 2006, Associate Professor at the University of Science and Technology of China from 2006 to 2008, and Assistant Professor at the University of California, Irvine from 2008 to 2014. His main expertise includes network traffic flow theories, route choice behavior and transportation planning, freeway control systems, arterial control systems, and inter-vehicle communications and vehicle control.

This page left blank intentionally

Abstract

In the first part of this report, we provide a comprehensive review on the traffic flow models and signal control strategies in urban networks. We first introduce the theoretical formulations of existing traffic flow models that have been used or potentially can be used in urban networks, including the cell transmission model (CTM), the store-and-forward model, the link transmission model (LTM), the link queue model (LQM), and the vertical cell model (VCM). Then we provide a review on traditional signal control strategies which mostly rely on the formulation of Webster's delay or the bandwidth concept. Due to the capability in capturing detailed traffic dynamics such as the formation, propagation, and dissipation of congestion arising at network junctions, we also provide a detailed review on the network representations and network junction models used in existing studies of the CTM and the corresponding signal control strategies.

Generally, the optimization problem in the CTM can be formulated as a mixed-integer-linear-programming (MILP) problem, which introduce a lot of binary variables for large-scale urban networks and is difficult to solve. Therefore, in the second part of this report, we aim to derive invariant averaged models to eliminate the binary variables introduced by the traffic signals. For the purpose of simplicity, we first apply our study to a signalized linear junction connecting one upstream link with one downstream link. Starting with three equivalent supply-demand models with binary signals, we derive the averaged models by replacing the signal control with an average value, which is the effective green ratio. Then we apply the local forms of these averaged models as entropy conditions to the signalized linear junction. After solving the arising Riemann problems, we obtain their invariant forms, from which we are able to show that only one of them is correct and satisfies the constraint of maximum average junction flux. In addition, we find that different non-invariant averaged models can lead to the same invariant form. Using the Cell Transmission Model (CTM) simulation on a signalized ring road, we demonstrate that the invariant averaged model is a reasonable approximation to the original supply-demand model with binary signals, and the approximation accuracy is not sensitive to the types of traffic flow fundamental diagrams but will

degrade with long cycle lengths. As a further extension, we apply this analytical framework to more complicated cases, e.g., the signalized merging junction. Due to the existence of merging behaviors, we introduce two new term while deriving the averaged model: Effective Demand and Merging Priority. With these two new terms, we follow similar procedures as those in the linear junction, and derive the corresponding invariant averaged model for the merging junction. We further show that the derived averaged model for the signalized linear junction is just one special case of the one for the signalized merging junction with empty demand in one of the upstream links.

In summary, the contributions of this project are in two parts. First, it provides researchers and engineers a comprehensive review on the state-of-the-art traffic flow models for urban networks and the signal control strategies developed particularly in the framework of CTM. Second, it introduces a novel analytical framework in deriving invariant averaged models for signalized intersections. The insights obtained from this project can potentially help to develop more efficient and effective control and management schemes in urban networks in the future.

Part I: Introduction

For a transportation network, it consists of two major components: links and nodes. At the macroscopic (or even mesoscopic) level, traffic flow models are often used to model the evolution of traffic inside a link. At a junction, models are needed to distribute vehicle flows from the upstream links to the downstream ones. Unlike the uninterrupted traffic on freeways, traffic behaves differently in urban networks due to the existence of traffic signals to regulate conflicting traffic movements at intersections. Therefore, in the analysis of urban networks, a combination of traffic flow models, junction models, and signal control models is needed.

In the literature, there have been a number of studies on traffic flow models. At the macroscopic level, the widely-used one is the LWR model (Lighthill and Whitham, 1955; Richards, 1956), which is a kinematic wave model that considers traffic as a continuum media and incorporates the concept of traffic flow fundamental diagram. In (Daganzo, 1994, 1995), a so-called Cell Transmission Model (CTM) was proposed to numerically solve the LWR model under general traffic conditions. Since then, CTM has attracted a lot of attentions to simulate traffic dynamics not only in freeway networks but also on urban streets due to the fact that: (i) CTM can capture detailed traffic dynamics such as the formation, propagation, and dissipation of congestion arising at network junctions; (ii) CTM is a macroscopic traffic flow model which requires less parameters to calibrate, validate, and optimize. In TRANSYT 13 (https://trlsoftware.co.uk/support/products/transyt_13), CTM was introduced as an alternative traffic flow model to model the queue spillback effects of downstream links. Different from the CTM, another traffic flow model, which is called Link Transmission Model (LTM), was proposed to solve the LWR model in (Yperman, 2007; Jin, 2014b). It was argued that the LTM is more accurate than the CTM due to the numerical diffusion inside the cells in the CTM when shockwave exists. Since it is a new model, few studies have applied it for network traffic simulations. Besides the CTM and the LTM, there also been other queueing models proposed to model the traffic dynamics within the links. Examples are the store-and-

forward model in (Gazis 1964; D'ans and Gazis, 1976), the link queue model in (Jin, 2012b), and the vertical cell model in (Anderson et al., 2015).

Due to various types of junctions / intersections in our transportation network, a number of junction models have been proposed to distribute the vehicle flows from the upstream links to the downstream ones. Examples for linear, merging, diverging, and general junctions can be found in (Daganzo, 1995; Lebacque, 1996; Jin and Zhang, 2003; Jin et al., 2009; Jin, 2010, 2014a, 2012a). Some of these junction models have been incorporated into the CTM to simulate the traffic dynamics in both freeway and urban networks in the literature. However, we should note that due to the intrinsic characteristics of traffic flow models, it is not guaranteed that a junction model that can work in one traffic flow model can be applied to another one. It was shown in (Jin, 2014b) that non-invariant junction models cannot be used in the LTM since it may yield no conventional solution to the traffic statics problem under certain traffic conditions.

In reality, traffic signals have been installed at numerous intersections to regulate conflicting traffic movements, and as a result, various signal control strategies have been proposed. As mentioned in (Lo, 1999, 2001), signal control strategies can be categorized according to the traffic conditions they are applied to: unsaturated and oversaturated. For unsaturated conditions, the development of traffic signal mainly rely on the following assumptions: (i) traffic is in the uncongested regime and traffic state is nearly stationary, and therefore, it is sufficient to have a set of fixed timing plans; (ii) queueing is modeled through classical queueing theory, and measures such as average vehicle waiting time and average service rate are used. In such a case, Webster's delay formulas have been used to derive optimal signal settings, including cycle lengths and green splits, at isolated intersections (Webster, 1958): methods for fixed-time signals include SIGSET (Allsop, 1971a,b) and SIGCAP (Allsop, 1972, 1976); and methods for traffic-responsive signals include the Vehicle Interval and Volume Density strategies (De la Breteque and Jezequel, 1979). For coordinated traffic signals along an arterial corridor, the bandwidth concept of green waves is often used to determine the offsets: methods for coordinated fixed-time signals include MAXBAND (Little, 1966), MULTI-BAND (Gartner et al., 1991), and TRANSYT (Robertson, 1969); and methods for coordinated traffic responsive signals include SCOOT (Hunt et al., 1981, 1982) and OPAC (Gartner, 1983). However,

as pointed out in (Papageorgiou, 2003), most of the signal control strategies are mainly developed for uncongested traffic conditions. In the literature, there have been some efforts devoted to developing signal control strategies under oversaturated conditions. For example, studies in (Gazis and Potts, 1963; Gazis 1964; D'ans and Gazis, 1976; Aboudolas et al., 2007, 2009) introduced a store-and-forward method to develop optimal signal timing plans for oversaturated traffic conditions. In these studies, the departure flow-rate at the signalized intersections was approximated using a continuous time-dependent function with a slope of its average flow-rate, and vehicle queue is updated by a simple first order condition which utilizes the difference between the inflows and outflows as well as its initial value. However, few traffic flow models used in existing signal control studies can work under a wide range of traffic conditions, i.e., from unsaturated to oversaturated conditions.

Since the CTM can replicate real-world traffic dynamics under a wide range of traffic conditions, as one of the earliest attempts, studies in (Lo, 1999, 2001) proposed a new formulation for the traffic signal control for one-way streets under the framework of the CTM. Optimal signal control settings were obtained by solving the mixed-integer linear programing (MILP) problems. Along this line, (Lo et al., 2001; Lo and Chow, 2004) extended this framework to model more complicated one-way streets with merging and diverging behaviors. Since it is harder to solve the MILP problems as the network gets larger and more complicated, genetic algorithms were used in (Lo et al., 2001; Lo and Chow, 2004) to obtain a set of good, rather than optimal, signal settings.

Since then, a lot of improvements have proposed to enhance the urban signal design with the CTM. To model the platoon dispersion at signalized intersections, a family of nonlinear speed-density relations, instead of the commonly-used triangular traffic flow fundamental diagram (Haberman, 1977), was introduced in (Feldman and Maher, 2002). Furthermore, studies have proposed various network junction models to take into account the merging and diverging behaviors at signalized intersections (Almasri and Friedrich, 2005; Su et al, 2013) and to model more complicated signal settings, such as the NEMA phase settings (Zhang et al., 2010). Under oversaturated conditions, it is possible to have lane blockages caused by the queue spillback from the downstream or by the conflicting left-turn and through traffic movements.

In (Li, 2010; Gao et al., 2015), the concepts of sub-zones and sub-cells were introduced to handle the lane blockage problems.

Generally, delay minimization is the major objective function to derive a set of optimal signal settings; examples can be found in (Lo, 1999; Lo, 2001; Almasri and Friedrich, 2005; Li, 2010). However, there also exist studies trying to maximize the system throughput (Li, 2010), to minimize the performance index (Feldman and Maher, 2002), to minimize a combination of delay and early arrival flow (He et al., 2010), or to minimize the mean of excess delay (Zhang et al., 2010). For small networks, commercial packages such as CPLEX can be used to solve the MILP problems based on the CTM formulation. However, as the study network gets larger and more complicated, it becomes more difficult to solve such MILP problems. Therefore, rather than finding an optimal solution, genetic algorithms have been used to find a set of reasonably good signal settings; examples can be found in (Lo et al., 2001; Lo and Chow, 2004; Almasri and Friedrich, 2005; Li, 2010; Zhang et al., 2010). To further improve the optimization process, in (He et al, 2010), linear relaxation and heuristic algorithms were used to find a feasible integer solution for the CTM-based traffic signal control models. Besides of these methods, studies in (Feldman and Maher, 2002) also introduced a hill-climbing method for the signal optimization based on CTM.

As discussed above, the major bottleneck in optimizing traffic signals at urban intersections is the increasing number of binary variables which are used to represent the green-red phases of traffic signals. To address this limitation, one of the earliest attempts is the store-and-forward model; examples can be found in (Gazis and Potts, 1963; Gazis, 1964; D'ans and Gazis, 1976; Aboudolas et al., 2007, 2009). However, this model relies on the assumption of high demand in the upstream and enough supply in the downstream, which can be violated when the traffic is either under free-flow conditions or is very congested with queue spillback from the downstream. Recently, in (Han et al., 2014), an averaged model of signal was proposed for a signalized merging junction, in which the approximation accuracy was analyzed in various aspects, e.g., under different traffic conditions and traffic flow fundamental diagrams. However, such a model is derived heuristically, which has no guarantee of its correctness. To the best of our knowledge, there is still a lack of a systematic and comprehensive study on: (i) deriving averaged models

for signalized intersections; and (ii) analyzing their properties and correctness under different combinations of road geometries, traffic conditions, and fundamental diagrams.

In this project, we attempt to fill this gap. For the purpose of simplicity, we take a linear signalized intersection connecting an upstream link to a downstream link as an example. We start with three equivalent supply-demand models with binary signals. Such supply-demand models were first introduced in the CTM to calculate boundary fluxes from upstream demands, downstream supplies, and turning proportions (Daganzo, 1995; Lebacque, 1996). They have been extended for different types of junctions with various merging and diverging behaviors (Jin and Zhang, 2003; Jin, 2010, 2014a, 2012a). These models can be simply extended with binary signals: for example, a signalized intersection is equivalent to a diverging junction during a phase.

Then for a linear junction, we derive the averaged models by replacing the signal control with an average value, which is the effective green ratio. The averaging method used in this study has been widely used in other systems with periodic forces (Krein et al., 1990; Sanders et al., 2007). We use these averaged models as entropy conditions for a network kinematic wave model. After solving the arising Riemann problems within the framework of (Jin et al., 2009), we obtain their invariant forms (Lebacque, 2005), from which we are able to show that only one of them is correct and satisfies the constraint of maximum average junction flux. In addition, we find that different non-invariant averaged models can lead to the same invariant form. Using the Cell Transmission Model (CTM) simulation on a signalized ring road, we demonstrate that the invariant averaged model is a reasonable approximation to the original supply-demand model with binary signals, and the approximation accuracy is not sensitive to the types of traffic flow fundamental diagrams but will degrade with long cycle lengths.

Furthermore, we extend this analytical framework to more complicated junctions, e.g., a signalized merging junction connecting two upstream links and one downstream one. We follow a similar procedure as in the signalized linear junction. We first provide a model of binary signals at the merging junction, in which the signal control is applied to both the upstream demands and downstream supplies. Then we derive its averaged counterpart by replacing the cyclic signal control in each phase with a constant value, which

turns out to be its effective green ratio. Different from the linear junction, we propose two new important definitions: Effective Demand and Merging Priority. Effective Demand of an upstream link takes into account not only its current demand, but also the reduced flow rate constrained by road geometries and signal settings. Merging Priority of an upstream link is a term computed as the percentage of its effective green time to the total effective green time. With these two new definitions, we apply the local form of this averaged model at the merging junction, solve the corresponding Riemann problems, and finally derive its invariant form. We further verify that the derived invariant averaged model for the signalized linear junction is just a special case of the one for the signalized merging junction with empty demand in one of the upstream links.

In summary, the analytical framework developed in this project provides fundamental bricks for future research developments. As already shown in this project, the current study framework is generic and can be extended to more complicated cases, such as four-way intersections. In such cases, more complicated driver's behaviors, such as merging and diverging, should be considered. In the literature, there have been models for freeway merging, diverging, and general (N-by-M) junctions; examples can be found in (Jin and Zhang, 2003; Jin 2010, 2014a, 2012a). For signalized intersections, traffic signals can be added into these junction models to mimic the evolutions of different traffic movements. However, it is not easy to derive correct averaged models for the corresponding intersections. Since more combinations of upstream demands and downstream supplies should be considered, solving the corresponding Riemann problems becomes difficult.

In the future, after the averaged models of signalized intersections are derived, it is straightforward to combine them with the state-of-the-art link models, which as a result forms a new framework of network kinematic wave models. Within this new framework, freeway and urban road networks can be modeled as a whole since the discrete signal control at urban intersections can be replaced by the continuous type of averaged models. This framework will become a very powerful tool in analyzing the static characteristics of traffic in large-scale networks, both theoretically and numerically. For example, traditional traffic assignment problems rely either on the link performance functions, which are unrealistic presentations of

congested traffic, or on more detailed microscopic simulation models, which are very complicated and time-consuming for large-scaled networks. This new framework will provide fundamental improvement on this part. On the one hand, its kinematic wave model provides a more realistic representation of traffic, which yields more accurate estimation of traffic performance metrics, e.g., vehicle queue, delay, speed, density, and flow-rate. On the other hand, its average models at signalized intersections significantly simplify the complexity of modeling signal control, which not only can dramatically improve the simulation speed but also enable the analytical studies of network traffic statics and dynamics. Furthermore, because the development of averaged models eliminates the binary variables used in traditional signal optimization algorithms, this new framework can be used as the baseline model to facilitate the optimization of traffic signals in large-scale networks.

PART II: Literature Review

Literature review in this part is organized as follows. In section 1, we provide a review on the traffic flow models including the CTM, the store-and-forward model, the LTM, the LQM, and the VCM. In section 2, we provide a review on traditional signal control strategies which mostly rely on Webster's delay formulation and the concept of maximum bandwidth. In Section 3, we provide a summary on the CTM-based network representations and junction models. In Section 4, we provide a summary on the CTM-based signal control strategies including the objective functions and the optimization methods. In Section 6, we draw our conclusions with some future research directions.

1. Traffic flow models for urban networks

1.1 Cell transmission model

In kinematic wave theories, traffic flow is considered as a continuous media. Three location-and-time dependent variables, speed $v(x,t)$, density $k(x,t)$, and flow-rate $q(x,t)$, are used to describe the traffic flow characteristics at point x and time t. For a road section without any entrances and exits, flow conservation is hold, which can be written as

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \tag{1}$$

In traffic flow theory, it is well known that there exists a fundamental relation between flow-rate (or speed) and density (Greenshields, 1935), i.e., $q(x,t) = Q(k(x,t))$ or $v(x,t) = V(k(x,t))$. Such a relation is known as the traffic flow fundamental diagram and can be validated using the vehicle loop detector data from freeways. Generally speaking, $Q(k)$ is a concave function and attains its capacity C at $k = k_c$, where k_c is the critical density. Introducing the fundamental diagram into Equation (1), the Lighthill-Whitham-Riahcrds (LWR) model (Lighthill and Whitham, 1955; Richards, 1956) is obtained.

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial Q(k(x,t))}{\partial x} = 0 \tag{2}$$

Equation (2) is a hyperbolic conservation law and is difficult to solve analytically under general initial and boundary conditions. Therefore, in (Daganzo, 1994, 1995), the cell transmission model (CTM) was introduced to numerically solve Equation (2). According to the Godunov method (Godunov, 1959), a link is equally divided into N cells with a length of Δx, and the whole time interval is partitioned into J time steps with an interval of Δt. In ***Figure 1***, cell representation inside a regular link is provided. Then the discrete version of Equation (2) can be written as

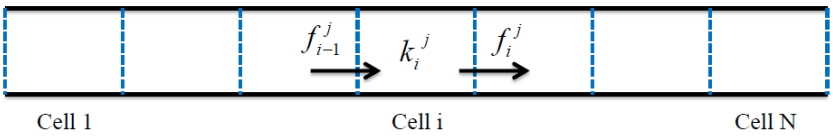


Figure 1 Cell representation inside a regular link

$$\frac{k_i^{j+1} - k_i^j}{\Delta t} - \frac{f_{i-1}^j - f_i^j}{\Delta x} = 0 \tag{3}$$

where k_i^j and k_i^{j+1} are the densities of cell i at time steps j and j+1, respectively, and f_{i-1}^j and f_i^j are the upstream and downstream boundary fluxes of cell i at time step j, respectively. Here, the choice of $\frac{\Delta t}{\Delta x}$ should follow the CFL condition (Courant et al., 1928), which requires a vehicle cannot travel across one cell at one time step. That is, $\frac{v_f \Delta t}{\Delta x} \leq 1$, where v_f is the free-flow speed of that link. Given densities and fluxes at time step j, the density at time step j+1 can be updated using the following equation:

$$k_i^{j+1} = k_i^j + \frac{\Delta t}{\Delta x}(f_{i-1}^j - f_i^j) \tag{4}$$

To obtain the fluxes crossing cell boundaries, the definitions of demand D and supply S (Daganzo, 1995; Lebacque, 1996) are introduced and can be calculated as

$$D = Q(\min\{k, k_c\}) \tag{5}$$

$$S = Q(\max\{k, k_c\}) \tag{6}$$

Therefore, the flux through a cell boundary can be calculated by taking the minimum of the upstream cell's demand and the downstream cell's supply, which is

$$f_{i-1}^{j} = \min\{D_{i-1}^{j}, S_i^{j}\} \tag{7}$$

where $D_{i-1}^{j}$is the demand of cell i-1, and S_i^{j} is the supply of cell i at time step j. For freeway networks, network junction models such as those in (Daganzo, 1995; Lebacque, 1996; Jin and Zhang, 2003; Jin et al., 2009; Jin, 2010, 2014a, 2012a) are needed to model the traffic dynamics at various types of junctions. For urban networks, besides the network junction models, signal control should be considered in order to manage the conflicting traffic movements at the intersections.

1.2 Other models for urban traffic flow

1.2.1 Store-and-forward model

The modeling framework of the store-and-forward model was first proposed by (Gazis and Potts, 1963) for an isolated intersection under oversaturated traffic conditions. After that, the study was extended to two consecutive intersections in (Gazis, 1964). Different from the graphical methods used in (Gazis and Potts, 1963; Gazis, 1964), the optimization of signal control was reduced to the linear programing problem with time discretization in (D'ans and Gazis, 1976). In addition, formulation of the optimal signal control problem under the store-and-forward framework was proposed for more complexed networks. In recent studies by (Aboudolas et al., 2007&2009), the store-and-forward method was applied to the control and optimization of signal control in large-scale urban networks.

In the store-and-forward model (D'ans and Gazis, 1976), an urban network is represented by a direct graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ is the set of links. Vehicles are assumed to travel with a constant speed inside a link before reaching the downstream node, and then, they are queued in the front of the node until it is allowed to move to the downstream destinations. For link i, the dynamics can be formulated as (Aboudolas et al., 2007&2009)

$$x_i^{j+1} = x_i^{j} + T[q_i^{j} - s_i^{j} + d_i^{j} - u_i^{j}] \tag{8}$$

where x_i^{j+1} and x_i^{j} are the numbers of vehicles in link i at time steps $j+1$, and j, respectively. q_i^{j} and u_i^{j} are the in-flow and out-flow at time step j, respectively; s_i^{j} and d_i^{j} are the exit and demand flows within

the link at time step j, respectively. T is the discrete time step which can be set to be equal to the cycle length.

A key component in the store-and-forward method is the approximation of the out-flow u_i^j under oversaturated traffic conditions. Let's consider a simple example of an intersection with two one-way streets (Gazis and Potts, 1963). During the peak periods when the road capacity is mostly needed, it is likely to have queues building up in one or both approaches if the following requirement is satisfied.

$$\frac{a_1(t)}{s_1} + \frac{a_2(t)}{s_2} > 1 - \frac{T_l}{C} \tag{9}$$

Here, $a_1(t)$ is the arrival rate for one approach, while $a_2(t)$ is for the other one. s_1 and s_2 are the saturation flow-rates in the downstream of the intersection. T_l is the total lost time, while C is the cycle length. Since the total traffic demand is high enough, it is possible to set the signal settings that the green times are fully used, and thus, the discharging flow-rates are equal to the saturation flow-rates. In ***Figure 2***, the arrival (blue solid lines) and the saw-toothed (red dashed lines) discharging patterns are provided. In (Gazis and Potts, 1963; Gazis, 1964), it was proposed to smooth the service curves using continuous functions (the green curves) by considering the fact that large delays are caused by queueing and the impact of the additional delay due to the saw-toothed pattern is limited. Therefore, for a general signalized intersection, the out-flow u_i^j can be calculated as (Aboudolas et al., 2007&2009)

$$u_i^j = \frac{G_i^j s_i}{C} \tag{10}$$

where G_i^j is the total effective green time in one cycle for link i at time step j.

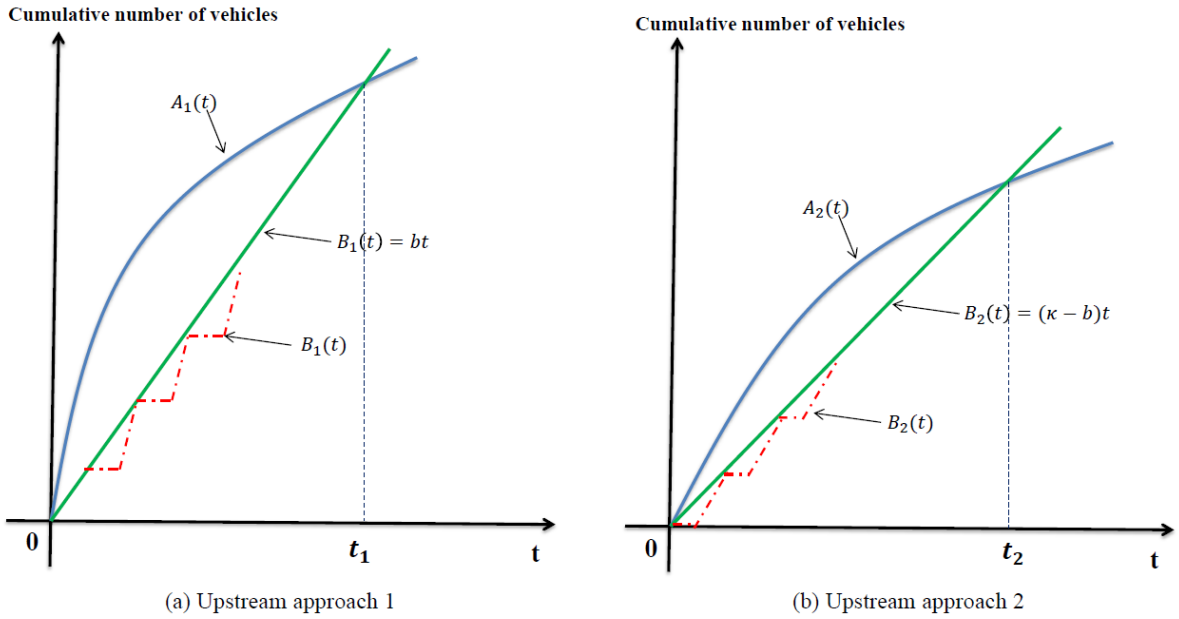


Figure 2 Arrival and discharging patterns at one intersection with two one-way streets.

1.2.2 Link transmission model

Traditionally, at a point (a, x_a) inside link a, the density $k_a(x_a, t)$, the speed $v_a(x_a, t)$, the flow-rate $q_a(x_a, t)$ are used as variables to describe the evolution of traffic flow. The flow conservation, $\frac{\partial k_a}{\partial t} + \frac{\partial q_a}{\partial x} = 0$, together with the traffic flow fundamental diagram, $q_a = Q_a(k_a)$, forms the LWR model. However, we also can use another type of state variable, which is the cumulative flow, $A_a(x_a, t)$, and is known as the Moskovitz function (Moskowitz, 1965). Since we have $k_a = -\frac{\partial A_a}{\partial x}$, and $q_a = \frac{\partial A_a}{\partial t}$, the flow conservation is automatically satisfied if we have $\frac{\partial^2 A_a}{\partial x \partial t} = \frac{\partial^2 A_a}{\partial t \partial x}$. Therefore, to solve the LWR model in Equation (2) is equivalent to solve the following Hamilton-Jacobi equation

$$\frac{\partial A_a}{\partial t} - Q_a\left(-\frac{\partial A_a}{\partial x}\right) = 0 \tag{11}$$

with the Hamiltonian $H\left(\frac{\partial A_a}{\partial x}\right) = -Q_a\left(-\frac{\partial A_a}{\partial x}\right)$. Besides the CTM, another new solution to the LWR model, which is called the Link Transmission Model (LTM), was proposed in recent studies. The discrete version can be found in (Yperman, 2007), while its continuous version can be referred to (Jin, 2014b).

Here, the triangular traffic flow fundamental diagram (Haberman, 1977), $q = Q(k) = \min\{v_f k, w(k_j - k)\}$, is used. The initial cumulative flow at $x_a \in [0, L_a]$ is denoted as $N_a(x_a)$. The cumulative in-flow and the in-flux at the upstream boundary are denoted as $F_a(t)$ and $f_a(t)$, respectively. The cumulative out-flow and the out-flux at the downstream boundary are denoted as $G_a(t)$ and $g_a(t)$, respectively. To describe the congestion pattern inside a link, two variables, the link queue size $\alpha_a(t)$ and the link vacancy size $\beta_a(t)$ are used and can be calculated as follows:

$$\alpha_a(t) = \begin{cases} N_a\left(L_a - v_{a,f}t\right) - G_a(t) & t \le \dfrac{L_a}{v_{a,f}} \\ F_a\left(t - \dfrac{L_a}{v_{a,f}}\right) - G_a(t) & t > \dfrac{L_a}{v_{a,f}} \end{cases}$$

$$\beta_a(t) = \begin{cases} N_a(w_a t) + k_{a,j} w_a t - F_a(t) & t \le \dfrac{L_a}{w_a} \\ G_a\left(t - \dfrac{L_a}{w_a}\right) + k_{a,j} L_a - F_a(t) & t > \dfrac{L_a}{w_a} \end{cases} \tag{12}$$

Initially, we have $\alpha(0) = 0$ and $\beta(0) = 0$. In the LTM, either cumulative flows or link queue and vacancy sizes can be used as stable variables to describe the evolution of traffic dynamics. If the cumulative flows, i.e., $F_a(t)$ and $G_a(t)$, are used, we have the following evolution equations:

$$\begin{aligned} \frac{d}{dt} F_a(t) &= f_a(t) \\ \frac{d}{dt} G_a(t) &= g_a(t) \end{aligned} \tag{13}$$

If the link queue and vacancy sizes, i.e., $\alpha_a(t)$ and $\beta_a(t)$, are used, we have the following evolution equations:

$$\frac{d\alpha_a(t)}{dt} = \begin{cases} k_a\left(L_a - v_{a,f}t, 0\right) v_{a,f} - g_a(t) & t \le \dfrac{L_a}{v_{a,f}} \\ f_a\left(t - \dfrac{L_a}{v_{a,f}}\right) - g_a(t) & t > \dfrac{L_a}{v_{a,f}} \end{cases} \tag{14}$$

$$\frac{d\beta_a(t)}{dt} = \begin{cases} -k_a(w_a t, 0)w_a + k_{a,j}w_a - f_a(t) & t \le \dfrac{L_a}{w_a} \\ g_a\left(t - \dfrac{L_a}{w_a}\right) - f_a(t) & t > \dfrac{L_a}{w_a} \end{cases}$$

To update the evolution functions in Equations (13) and (14), the in-fluxes and out-fluxes are needed to be calculated/updated first. Here we define an indicator function $H(y)$ for $y \ge 0$, which is formulated as follows:

$$H(y) = \lim_{\Delta t \to 0^+} \frac{y}{\Delta t} = \begin{cases} 0 & y = 0 \\ +\infty & y > 0 \end{cases} \tag{15}$$

Then the link demand $d_a(t)$ and link supply $s_a(t)$ are defined as

$$d_a(t) = \begin{cases} \min\{k_a(L_a - v_{a,f}t, 0)v_{a,f} + H(\alpha_a(t)), C_a\} & t \le \dfrac{L_a}{v_{a,f}} \\ \min\{f_a\left(t - \dfrac{L_a}{v_{a,f}}\right) + H(\alpha_a(t)), C_a\} & t > \dfrac{L_a}{v_{a,f}} \end{cases}$$

$$s_a(t) = \begin{cases} \min\{k_{a,j}w_a - k_a(w_a t, 0)w_a + H(\beta_a(t)), C_a\} & t \le \dfrac{L_a}{w_a} \\ \min\{g_a\left(t - \dfrac{L_a}{w_a}\right) + H(\beta_a(t)), C_a\} & t > \dfrac{L_a}{w_a} \end{cases} \tag{16}$$

At a junction j, macroscopic junction models are used to determine the in-fluxes and out-fluxes from the upstream link demands, downstream link supplies, and turning proportions, which in general can be written as follows:

$$\left(\boldsymbol{g}_j(t), \boldsymbol{f}_j(t)\right) = \boldsymbol{F}(\boldsymbol{d}_j(t), \boldsymbol{s}_j(t), \boldsymbol{\xi}_j(t))) \tag{17}$$

Here, $\boldsymbol{g}_j(t)$ is the set of in-fluxes, while $\boldsymbol{f}_j(t)$ is the set of out-fluxes. $\boldsymbol{d}_j(t)$ is the set of upstream link demands, while $\boldsymbol{s}_j(t)$ is the set of downstream link supplies. $\boldsymbol{\xi}_j(t)$ is a matrix that contains turning proportions from the upstream links to the downstream ones. As shown in (Jin, 2014b), non-invariant junction models cannot be used in the LTM, which may yield no conventional solution to the traffic statics problem under certain traffic conditions. A set of invariant junction models can be found in (Jin et al., 2009; Jin, 2010, Jin, 2012a; Jin, 2014a).

With Equations (16) and (17), the in-fluxes and out-fluxes can be calculated and then be introduced into Equations (13) or (14) to update the state variables. But note that, as shown in Equation (16), link demands and supplies depend on the historical data, and therefore, Equations (13) and (14) are systems of ordinary differential equations (ODEs) with delays. Once the cumulative in-flows $F_a(t)$ and the cumulative out-flows $G_a(t)$ are obtained, traffic states inside link a can be obtained. More details can be referred to (Jin, 2014b).

1.2.3 Link queue model

In (Jin, 2012b), a so-called link queue model was proposed to consider vehicles in a link as a single queue. We denote the set of regular links as A, the set of origins as O, and the set of destinations as R. Then, the state variable for a single link $a \in A$ is its average density, k_a, while it is the link volume, K_o, for an origin $o \in O$. The link queue model incorporates two important features: (i) traffic flow fundamental diagram at the link level is used to define the link-based supply and demand; (ii) the junction fluxes are calculated based on the link demands in the upstream and the link supplies in the downstream. The link queue model of network traffic flow in (Jin, 2012b) can be formulated as

$$\begin{aligned} \frac{dk_a(t)}{dt} &= \frac{1}{L_a}\big(f_a(t) - g_a(t)\big), \qquad a \in A \\ \frac{dK_o(t)}{dt} &= f_o(t) - g_o(t), \qquad o \in O \end{aligned} \tag{18}$$

where L_a is the link length, and $f_a(t)$ is the in-flux and $g_a(t)$ is the out-flux of link a. $f_o(t)$ is the arrival rate, and $g_o(t)$ is the out-flux at origin o.

In order to update the densities or queued vehicles in Equation (18), the in-flows and out-flows should be calculated first. For a regular link, $a \in A$, its demand and supply are defined as

$$\begin{aligned} D_a(t) &= Q_a(\min\{k_a(t), k_{a,c}\}) \\ S_a(t) &= Q_a(\max\{k_a(t), k_{a,c}\}) \end{aligned} \tag{19}$$

Here, $Q_a(k_a)$ is the traffic flow fundamental diagram of link a. Different from a regular link, the demand for an origin $o \in O$ is defined as

$$D_o(t) = f_o(t) + I_{K_o(t)\geq 0} = \begin{cases} +\infty & K_o(t) > 0 \\ f_o(t) & K_o(t) = 0 \end{cases} \tag{20}$$

The above equation shows that the demand at an origin is infinity if a queue exists, while it is the same as the inflow when the queue disappears. For a destination $r \in R$, the supply is defined as $S_r(t)$, which is set to be $+\infty$ when the downstream is not blocked. Since each link is considered as a whole in the link queue model, we only need to calculate the inflows and outflows at the junctions. Therefore, macroscopic junction models are needed for different types of junctions. Generally, given upstream link demands and downstream link supplies, a junction model at junction j can be formulated as

$$\left(\boldsymbol{G}_j(t), \boldsymbol{F}_j(t)\right) = \text{FF}(\boldsymbol{D}_j(t), \boldsymbol{S}_j(t), \boldsymbol{\xi}_j(t)) \tag{21}$$

where $\boldsymbol{G}_j(t)$ is the set of out-fluxes, $\boldsymbol{F}_j(t)$ the set of in-fluxes, $\boldsymbol{D}_j(t)$ the set of upstream link demands, and $\boldsymbol{S}_j(t)$ the set of downstream link supplies. $\boldsymbol{\xi}_j(t)$ is the matrix of turning proportions indicating vehicle flows from the upstream links to the downstream ones. Examples of the junction models can be found in (Daganzo, 1995; Lebacque, 1996; Jin and Zhang, 2003; Jin et al., 2009; Jin, 2010, Jin, 2012a; Jin, 2014a).

The link queue model in Equation (18) cannot be analytically solved under general initial and boundary conditions, and therefore, numerical methods should be introduced to obtain its approximate solutions. In (Jin, 2012b), an explicit Euler method was used to obtain the discrete version of Equation (18). The whole time period is equally divided into J time steps with a size of Δt. At time step j, the average density on link a is denoted as k_a^j, and the average queue at origin o is denoted as K_o^j. The link demand and supply are denoted as D_a^j and S_a^j, respectively. For origin o, the demand is denoted as D_o^j and is calculated as

$$D_o^j = \frac{K_o^j}{\Delta t} + f_o^j \tag{22}$$

Then the density and queue length at time step $j+1$ can be updated by

$$k_a^{j+1} = k_a^j + \frac{\Delta t}{L_a}(f_a^j - g_a^j) \tag{23}$$

$$K_o^{j+1} = K_o^j + (f_o^j - g_o^j)\Delta t$$

1.2.4 Vertical cell model

In (Anderson et al., 2015), a vertical queueing model called Vertical Cell Model (VCM) was proposed. In VCM, new features such as link transit time and finite queue capacity are incorporated into the queueing dynamics. In the network representation of VCM, a network is represented as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ is the set of links. Physical roads are divided into several parallel links, and each link contains vehicle flows aiming to the same downstream roadway. Mid-link split is used to divide a roadway into two to take into account the cases of shared lanes and turn bays. Therefore, the link set $\mathcal{L}$ contains three types of links: internal links, entry links, and exit links. For an internal link, the in-flux and out-flux are constrained by the number of vehicles it contains. For an entry link, it serves as an entry for external demands. For an exit link, it serves as a reservoir which has unlimited storage. The node set $\mathcal{N}$ contains two types of nodes: intermediate nodes and intersection nodes. For an intermediate node, it acts as a diverging junction in which a single demand flow is divided into several downstream demand flows with a set of specified turning ratios. For an intersection node, it takes into account the conflicts between vehicle flows and distributes them according to the dictated turning ratios.

In the VCM, fluxes crossing a node are limited by the so-called sending constraints of its upstream links, and the receiving constraints of its downstream links. The out-flux departing from an upstream link $l \in I_n$ at node n can be calculated as

$$d_l(t) = G_l(t)\min\{S_l(t), \min_{z \in Out(l)}\{\frac{1}{\beta_n^{l,z}} R_z(t)\}\} \tag{24}$$

where $Out(l)$ is the set of downstream links of link l, $S_l(t)$ the sending constraint of link l, $R_z(t)$ the receiving constraint of link z, $\beta_n^{l,z}$ the turning ratio from link l to link z, and $G_l(t)$ an indicator to permit link l to discharge according to the signal settings. The in-flux entering a downstream queue $m \in O_n$ is calculated as

$$a_m(t) = \sum_{k \in I_n} \beta_n^{k,m} d_k(t) \tag{25}$$

Similar to the CTM, a VCM link is divided into several cells, e.g., τ_l cells. After a vehicle enters a VCM link, it first travels at the free-flow speed across $\tau_l - 1$ transit cells before reaching the terminal queueing cell τ_l. The terminal queueing cell τ_l is used to store the queued vehicles. Therefore, there are τ_l state variables inside each link: (i) $f_{l,i}(t),\ i = 1, \dots, \tau_l - 1$, is the flux entering link l at time $t - i$; and (ii) $q_l(t)$ is the number of queued vehicles inside the link. Then, receiving constraints in the VCM can be calculated as

$$R_l(t) = \gamma_l \min\{c_l \Delta t, \kappa_l - q_l(t) - \sum_{i=1,\dots,(\tau_l - 1)} f_{l,}(t)\} \tag{26}$$

where c_l is the maximum flow-rate, κ_l is a fixed queueing capacity, and γ_l is the number of lanes on link l. The sending constraints can be calculates as

$$S_l(t) = \gamma_l \min\{c_l \Delta t, q_l(t)\} \tag{27}$$

With Equations from (24) to (27), the VCM model is complete and is ready for urban network traffic simulations.

2. Traditional signal control methods

For signal control at arterial intersections, it can be classified into the following two types: for isolated intersections only and for coordinated intersections. In the literature, there have been a number of signal control strategies proposed for each category. In the following subsections, we provide a review of some prevailing strategies.

2.1 For isolated intersections

According to (Papageorgiou et al., 2003), fixed-time control strategies for a single intersection can be stage-based or phase-based. For stage-based strategies, the stage settings are fixed, and the proposed strategies are developed to find optimal splits and cycle lengths by minimizing the total delay or maximizing the total throughput at the intersection. To calculate vehicle's average delay, the delay formulation proposed by Webster (Webster, 1958) has been widely used in the literature, which can be formulated as follows:

$$d = \frac{\frac{1}{2}C\left(1-\frac{g}{C}\right)^2}{1-\frac{g}{C}X} + \frac{X^2}{2v(1-X)} - 0.65\left(\frac{C}{v^2}\right)^{\frac{1}{3}} X^{2+5\frac{g}{C}} \quad (28)$$

where

C =cycle length,

g =effective green time,

v =arrival flow-rate,

c =capacity of the intersection approach,

s =saturation flow-rate,

$X = \frac{v}{c} = \frac{vC}{sg}$, the degree of saturation.

The above delay calculation consists of three parts: the first part is the uniform delay; the second part is the random delay; and the third part is the empirical adjustment. In (Webster, 1958), optimal cycle lengths were obtained by minimizing the total delay at the intersection under given arrival flow-rates. In (Miller, 1963b), to obtain optimal settings of splits and cycle lengths, various arrival patterns were taken into account in the calculation of random delay. Since it is possible that an approach may have right of way in more than one stage within a cycle, SIGSET was proposed in (Allsop, 1971a; Allsop, 1971b) to take into account such a case, and Webster's delay formula was used in the delay estimation. By minimizing the total delay with the capacity, cycle lengths, and minimum green time constraints, optimal settings of cycle length and effective green time for each stage were obtained. Under similar constraints as those in SIGSET, another program called SIGCAP was proposed in (Allsop, 1972; Allsop, 1976) to maximize the practical capacity at signalized intersections.

Different from stage-based control strategies, phase-based control strategies are developed to further consider optimal stage settings. One example can be found in (Improta and Cantarella, 1984), in which the constraint of fixed staging was released. Instead, incompatibility of traffic streams was introduced as a constraint in the optimization problem. By either minimizing the total delay or maximizing the intersection capacity, optimal settings of splits, cycle lengths, and stage settings can be obtained. In (Improta

and Cantarella, 1984), the optimization problem was formulated as a binary-mixed-integer-linear-programming (BMILP) problem, and solutions were obtained using a branch-and-bound method.

Besides fixed-time control strategies, there also exist traffic-responsive control strategies that utilize the real-time loop detector data in the field. In (De la Breteque and Jezequel, 1979), examples such as the Vehicle Interval strategy, the Volume Density strategy, and Miller's algorithm were provided. In the Vehicle Interval strategy, each stage has a set of pre-specified minimum and maximum green times. If a vehicle is detected to cross the intersection, a critical interval (CI) will be used to extend the green time to allow that vehicle to pass. A similar control logit was used in the Volume Density strategy. But it further takes into account queue lengths and vehicles' waiting times during the red phases while deciding the switching time instants. In (Miller, 1963a), a computer program was used to determine whether to switch the signal immediately or to delay the switch for a user-defined time interval at every time step. Such a decision is made based on the evaluation of the time gain in postponing the switch. If the time gain is negative, the signal is switched immediately; otherwise, it remains unchanged for the next time step.

2.2 For coordinated intersections

If traffic signals in an arterial are close enough, the dissipation of vehicles is usually in platoons. Therefore, it is possible to synchronize the signals so as to allow vehicles travel along the arterial from the beginning to the end without stopping. In this case, bandwidth in one traffic direction is defined as the time difference between the first and the last vehicles that satisfy the above requirement. In the literature, there have been studies trying to maximize the bandwidths along the arterial. For example, with given cycle and speed ranges, MAXBAND was introduced in (Little, 1966) to obtain optimal offset settings so as to maximize the total bandwidths of a two-way arterial. The optimization problem was formulated as a mixed-integer-linear-programing (MILP) problem, and a branch-and-bound method was used to solve it. Later in (Gartner et al., 1991), MULTI-BAND was proposed to add new features such as determination of left-turn phases and different bandwidths among the links into the optimization problem. In (Robertson, 1969), TRANSYT (TRAffic Network StudY Tool) was proposed to obtain multi-directional green waves so as to minimize the total delay. Such a model consists of two parts: (i) with given network information such as

road geometries, turning ratios at intersection, and demands, a platoon dispersion model is used to describe vehicle's progression inside a link; (ii) a "hill-climbing" method is used to solve the optimization problem. Performance Index (PI) is introduced to evaluate the improvements at each optimization step. The program will stop when a (local) minimum is found.

Due to the fact that demands and turning movements at intersections are changing as time elapses, traffic-responsive coordinated strategies have also been proposed in the literature. SCOOT (Split, Cycle and Offset Optimization Technique), which is a traffic-responsive version of TRANSYT, was proposed in (Hunt et al., 1982; Hunt et al., 1981). While keeping similar optimization structure as in TRANSYT, SCOOT works in a real-time fashion: it utilizes real-time measurements of flows and occupancies from vehicle loop detectors to predict delay and stops; the signal optimizer works in real time, and new signal settings are implemented directly on the street. Besides SCOOT, another algorithm called OPAC (Optimization Policies for Adaptive Control), which is a model-based traffic-responsive strategy, was proposed in (Gartner, 1983). In OPAC, splits, offsets, and cycles are not explicitly considered. A rolling horizon approach is used for real-time applications: at time t, the optimization method calculates an optimal switching scheme for the time interval $[t-h, t+H-h]$ $(H > h)$ based on the data in the time interval $[t-h, t]$ and applies it to the time interval $[t, t+h]$; then the optimization time horizon moves to the next step, $t+h$. Note that since OPAC employs complete enumeration in the optimization, it is not real-time feasible for multiple intersections (Papageorgiou et al., 2003).

3. CTM-based network representations and junction models at signalized intersections

Even though the formulation of CTM is similar in existing studies, the network representations and junction models at the signalized intersections can be different due to the fact that the network topologies in the study networks can vary a lot. In this section, we will categorize existing studies according to their study networks and summarize their network representations as well as the corresponding network junction models.

3.1 Simple one-way streets

In (Lo, 1999, 2001), a novel traffic signal control which is based on the CTM was proposed. In these studies, links are categorized into three types: source link, exit link, and intermediate link. The cells inside a link is labeled from its upstream direction. For a source link, the first cell is modeled as a parking lot to store the total demand that intends to enter the network. For an exit link, only one cell is used, and its inflow capacity is modeled to reflect the impact of signal control: it equals to the saturation flow-rate, s, when the traffic light is green, and zero otherwise. Its holding capacity is set to be infinite to serve as a reservoir. For an intermediate link, the first cell serves as a traffic signal. For example, for link l, the inflow capacity of the first cell, $Q_{l,1}(t)$, at time t can be calculated as

$$Q_{l,1}(t) = \begin{cases} s & t \in \text{green phase} \\ 0 & \text{otherwise} \end{cases} \quad (29)$$

Since simple one-way streets were considered in (Lo, 1999, 2001), the signal control was simple without merging and diverging behaviors. For the signal control, a constant cycle length with no yellow time was used. For signal coordination, initial offsets for the major approaches were introduced at the beginning of the analysis time period. Such a simple one-way street network was also used in (He et al., 2010) to study different heuristic algorithms to solve the 0-1 mixed-integer-linear-programing (MILP) formulations of traffic signal control.

3.2 One-way streets with merging and diverging behaviors

In (Lo et al., 2001; Lo and Chow, 2004), studies were extended to more complicated one-way streets with merging and diverging behaviors. The effect of a traffic signal was still modeled as in Equation (29). Besides the model for the linear junction, models for merging and diverging junctions were also introduced. In ***Figure 3***, a signalized merging junction with two upstream cells and one downstream cell, and a signalized diverging junction with one upstream cell and two downstream cells are provided.

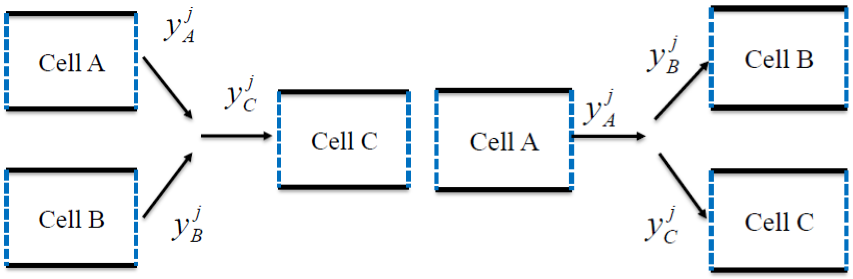


Figure 3 Two types of signalized junctions: (a) Merging and (b) Diverging.

The fluxes through the merging junction can be calculated as

$$\begin{aligned} y_A^j &= \min\{n_A^j, Q_A^j, \delta[N_C - n_C^j]\}, \\ y_B^j &= \min\{n_B^j, Q_B^j, \delta[N_C - n_C^j]\}, \\ y_C^j &= y_A^j + y_B^j. \end{aligned} \tag{30}$$

Here, n_i^j, $i = A, B, C$, is the number of vehicles in cell i at time step j. N_C is the holding capacity in cell C. Q_i^j, $i = A, B$, is the maximum number of vehicles that can enter cell C at time step j, which is calculated from Equation (29) to reflect the impact of signal control. δ is the ratio of the shockwave speed to the free-flow speed, i.e., $\delta = \frac{w}{v_f}$. For the diverging junction, the following equations are used:

$$\begin{aligned} y_B^j &= \beta_B y_A^j \\ y_C^j &= \beta_C y_A^j \\ y_A^j &= \min\begin{cases} n_A^j \\ Q_A^j \\ \dfrac{\min\{Q_B^j, \delta[N_B - n_B^j]\}}{\beta_B} \\ \dfrac{\min\{Q_C^j, \delta[N_C - n_C^j]\}}{\beta_C} \end{cases} \end{aligned} \tag{31}$$

where β_B and β_C are the turning proportions from A to B and C, respectively, and $\beta_B + \beta_C = 1$. Note that with the diverging model in Equation (31), we can see that flows in the upstream cell will be restricted

when one of the downstream cells cannot accommodate the assigned number of vehicles and thus becomes congested.

3.3 A four-way roundabout

In (Feldman and Maher, 2002), traffic dynamics at a four-way roundabout was studied under the framework of the CTM. At a roundabout, blocking queues can occur due to the limited storage space. It was argued in (Feldman and Maher, 2002) that with the commonly-used triangular or trapezoidal traffic flow fundamental diagram, the CTM cannot model the actual platoon dispersion at signalized intersections. Therefore, the following family of non-linear speed-density relations is used to capture the real platoon dispersion patterns at the intersections:

$$v = v_0 \left[1 - \left(\frac{k}{k_j} \right)^{l-1} \right]^{\frac{1}{1-m}} \tag{32}$$

where l and m are parameters to produce a variety of different shapes for the $v - k$ relations.

3.4 General road networks

3.4.1 One dimensional cell representation

In (Almasri and Friedrich, 2005), three types of networks junctions are considered: linear (1×1), merging (2×1) and diverging (1×2). Examples can be found in ***Figure 4*** with $U = 2$ and $D = 2$ in this study. Let $\bar{D}_u^j = \min\{n_u^j, Q_u^j\}$, and $\bar{S}_d^j = \min\{Q_d^j, \frac{w}{v_f}(N_d - n_d^j)\}$. The number of vehicles passing through a linear junction can be calculated as

$$y_u^j = \min\{\bar{D}_u^j, \bar{S}_d^j\} \tag{33}$$

For the diverging junction, the numbers of vehicles passing through the junction boundaries can be calculated as

$$\begin{aligned} y_u^j &= \min\{\bar{D}_u^j, \frac{\bar{S}_{d1}^j}{\beta_{u,d_1}}, \frac{\bar{S}_{d2}^j}{\beta_{u,d_2}}\} \\ y_{d_1}^j &= \beta_{u,d_1} y_u^j \end{aligned} \tag{34}$$

$$y_{d_2}^j = \beta_{u,d_2} y_u^j$$

where β_{u,d_1} and β_{u,d_2} are the turning proportions from u to d_1 and d_2, respectively. For the merging junction, there are several cases to calculate the numbers of vehicles passing through the junction boundaries:

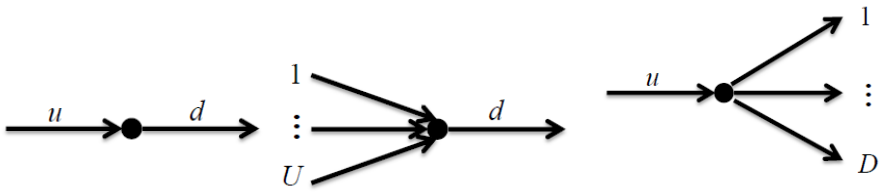
(i) When $\bar{S}_d^j \geq \bar{D}_{u_1}^j + \bar{D}_{u_2}^j$,

$$\begin{aligned} y_d^j &= \bar{D}_{u_1}^j + \bar{D}_{u_2}^j \\ y_{u_1}^j &= \bar{D}_{u_1}^j \\ y_{u_2}^j &= \bar{D}_{u_2}^j \end{aligned} \tag{35}$$

(ii) When $\bar{S}_d^j < \bar{D}_{u_1}^j + \bar{D}_{u_2}^j$,

$$\begin{aligned} y_{u_1}^j &= mid\{\bar{D}_{u_1}^j, \bar{S}_d^j - \bar{D}_{u_2}^j, p_{u_1}\bar{S}_d^j\} \\ y_{u_2}^j &= mid\{\bar{D}_{u_2}^j, \bar{S}_d^j - \bar{D}_{u_1}^j, p_{u_2}\bar{S}_d^j\} \\ y_d^j &= y_{u_1}^j + y_{u_2}^j \end{aligned} \tag{36}$$

where p_{u_1} and p_{u_2} are the proportions of vehicles that come from u_1 and u_2. For the signal control, it can be done in two ways: (i) restrict the capacity of the controlling cell at the road junction; (ii) introduce a time variant $p(t)$ that takes the value of 0 or 1.



(a) Linear junction (b) Merging junction (c) Diverging junction

Figure 4 Three types of signalized junctions.

In (Zhang et al., 2010), an extension to Lo's CTM was proposed which includes the modeling of two-way traffic, optimization of phase sequences, and a new formulation of the CTM model. The junctions in a signalized network were categorized into the following groups: ordinary, origin, destination, non-

signalized diverging, signalized diverging, and signalized merging. For the ordinary junction, the calculation of flows is formulated as

$$y_i^j = \min\{n_i^j, Q_i^j, Q_{i+1}^j, \frac{w}{v_f}[N_{i+1} - n_{i+1}^j]\} \tag{37}$$

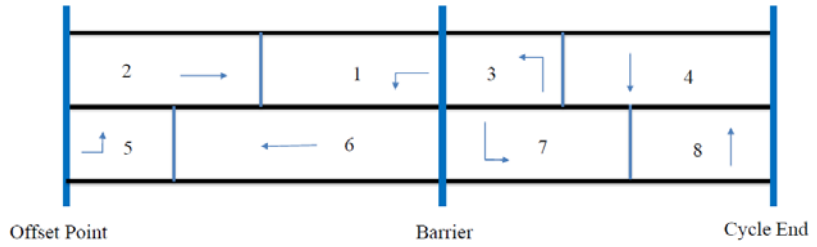
which is essentially a linear conditional piecewise function (CPF) and can be translated into the following constraints with two binary variables ξ_1 and ξ_2 and a sufficiently large negative constant U^-:

$$\begin{gathered}
(\xi_1 + \xi_2)U^- \le y_i^j - n_i^j \le 0 \\
(1 + \xi_1 - \xi_2)U^- \le y_i^j - Q_i^j \le 0 \\
(1 - \xi_1 + \xi_2)U^- \le y_i^j - Q_{i+1}^j \le 0 \\
(2 - \xi_1 - \xi_2)U^- \le y_i^j - \frac{w}{v_f}[N_{i+1} - n_{i+1}^j] \le 0
\end{gathered} \tag{38}$$

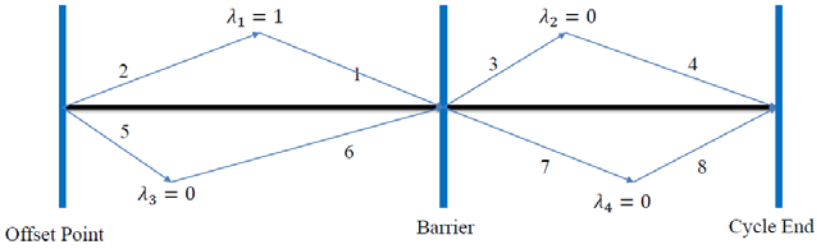
Similar translations are used to change the minimization functions in the diverging and merging junctions to different sets of constraints with binary variables. Since the relation between flow and traffic signal follows an "if-then" pattern (e.g., Equation (29)), such a relation is translated to a set of equality and inequality constraints using two binary variables $z_1(p,t)$ and $z_2(p,t)$.

$$\begin{gathered}
-U \times z_1(p,t) + \epsilon \le t - e(p) \le U \times [1 - z_1(p,t)] \\
-U \times z_2(p,t) \le b(p) - t \le U \times [1 - z_2(p,t)] - \epsilon \\
z_1(p,t) + z_2(p,t) = z(p,t) + 1 \\
Q_i(t) = (z_1(p,t) + z_2(p,t) - 1) \times s \\
\sum_p (z_1(p,t) + z_2(p,t)) = 2
\end{gathered} \tag{39}$$

Here, U is a sufficient large positive number, and ϵ is an arbitrary small number. $b(p)$ and $e(p)$ are the beginning time and the ending time of phase p, respectively.



(a)



(b)

Figure 5 NEMA phase structure at a four-way intersection.

To model the NEMA phase structure shown in ***Figure 5***(a), four binary variables, e.g., λ_i, $i = 1, 2, 3, 4$, are used to identify the phase sequence of the eight traffic movements in ***Figure 5***(b). For example, λ_1 is used to construct the phase sequence of 1 and 2. Let's denote o as the offset point of an intersection, m as the intersection ID, c as the cycle ID, h as the barrier point. Then the relations between phase 1 and phase 2 can be modelled as follows:

$$b(m,1,c) = \lambda_1 \times o(m) + \lambda_1 \times l \times (c-1) + (1-\lambda_1) \times e(m,2,c)$$

$$b(m,2,c) = (1-\lambda_1) \times o(m) + (1-\lambda_1) \times l \times (c-1) + \lambda_1 \times e(m,1,c)$$

$$e(m,1,c) = b(m,1,c) + g(m,1) \tag{40}$$

$$e(m,2,c) = b(m,2,c) + g(m,2)$$

$$g(m,1) + g(m,2) = h(m)$$

where $b(m,i,c)$ and $e(m,i,c)$, $i = 1, 2$, are the beginning and ending times for phase i, respectively.

In (Su et al, 2013), the study site in the NGSIM project (*http://ngsim-community.org/*), which is a segment of the Lankershim Boulevard in Los Angeles, California, was studied under the CTM framework. The study network contains different types of junctions, e.g., linear, merge, and diverge, which are shown

in ***Figure 4***. Therefore, junction models are introduced to calculate the corresponding fluxes through the boundaries. The flux f_u^j through the linear boundary is calculated as

$$f_u^j = \min\{D_u^j, S_d^j\} \tag{41}$$

where D_u^j is the demand (sending flow), and S_d^j is the supply (receiving flow) at time step j. The fluxes through a merging junction can be calculated as

$$\begin{aligned} f_d^j &= \min\{\sum D_u^j, S_d^j\}, \\ f_u^j &= p_u f_d^j \end{aligned} \tag{42}$$

where p_u is proportional to the link demand D_u^j at the merging junction. The fluxes through a diverging junction can be calculated as

$$\begin{aligned} f_u^j &= \min_d \left\{ D_u^j, \frac{S_d^j}{\beta_{u,d}} \right\}, \\ f_d^j &= \beta_{u,d} f_u^j \end{aligned} \tag{43}$$

where $\beta_{u,d}$ is the turning proportion from cell u to cell d.

Because there exists left-turn bays in real signalized road networks, a one-way road section was divided by three links according to the allowable movements in (Su et al, 2013), which is shown in ***Figure 6***. The dividing point is the location where the left-turn bay starts. The ends of the three links (Links 2 to 4) are controlled by a signal. Therefore, the link capacity is the saturation flow-rate when the traffic light is green, and zero when it is red. However, due to the modeling limitation, in (Su et al, 2013) the model was relaxed to allow the discharging flow of right-turn vehicles during the red time to be the same as that during the green time. And also, the impact of pedestrians on the turning traffic was not considered in (Su et al, 2013).

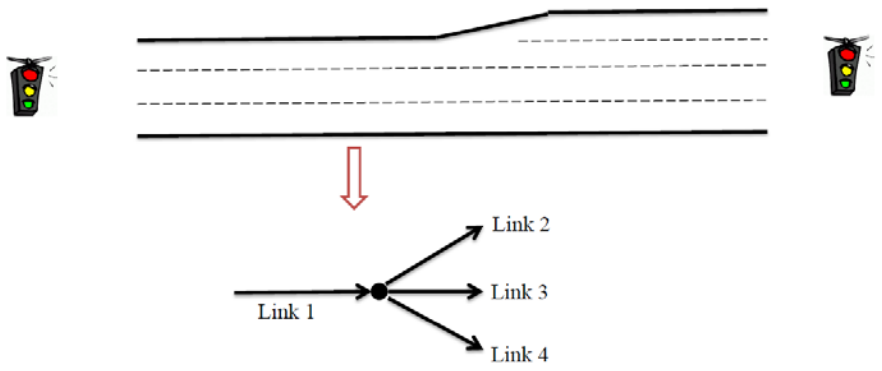


Figure 6 Link representation at a one-way road section.

3.4.2 Two dimensional cell representation: cells and sub-cells

In (Li, 2010), each link was divided into four zones: merging, propagation, diverging, and departure, which is shown in ***Figure 7***. Junctions are categorized into three types: ordinary, merging and diverging. In the merging zone, there are three upstream cells for through, left-turn, and right-turn vehicles, denoted as TH, LT, and RT, respectively. The downstream cell is denoted as d. The junction model is formulated as

$$y_{d,i}^{j} = \min\{n_i^j, Q_i^j, \delta[N_d - n_d^j]\}, \quad i = TH, LT, RT \tag{44}$$

where $\delta = 1$ if $n_i^j \le Q_i^j$, and otherwise, $\delta = \frac{w}{v_f}$. In the propagation zone, the junction model is formulated as

$$y_i^j = \min\left\{n_i^j, Q_i^j, \frac{w}{v_f}[N_{i+1} - n_{i+1}^j]\right\} = \min\{\bar{D}_i^j, \bar{S}_{i+1}^j\} \tag{45}$$

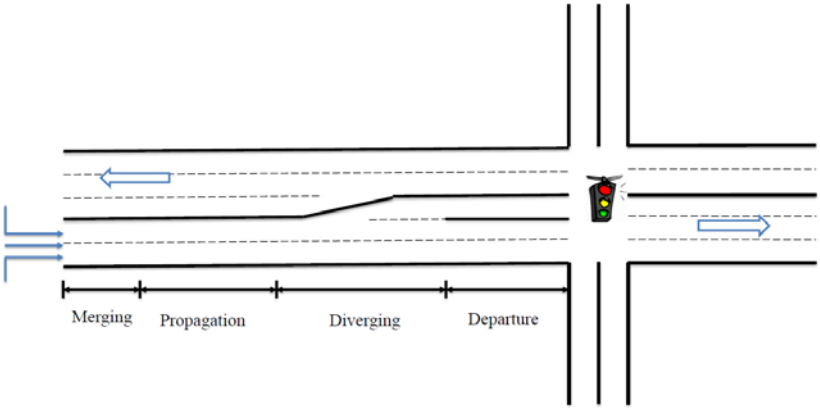


Figure 7 Four zones at a signalized intersection approach.

In the diverging zone, it is represented by a single cell i+1. Inside this zone, it is possible to have vehicle blockages caused by the conflicting movements between left-turn and through vehicles. Therefore, cell i+1 is further divided into two sub-cells: one for left-turn vehicles while the other for through vehicles. In the diverging zone, it is further divided into three sub-zones: one is designated for left-turn vehicles only, one for through vehicles only, and the upstream one shared by both left-turn and through vehicles. Detailed sub-cell representation is provided in ***Figure 8***.

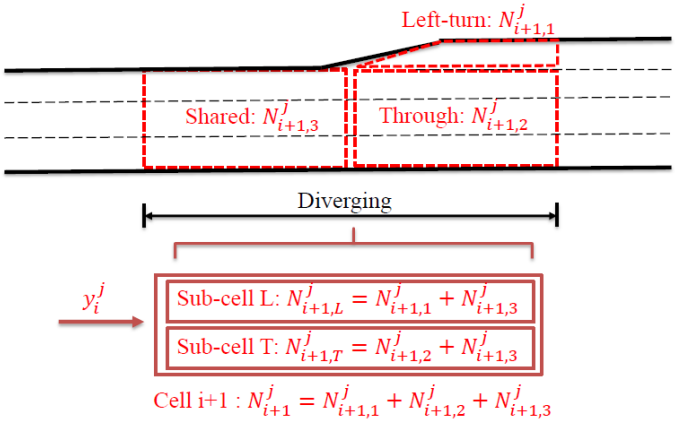


Figure 8 Sub-cell representation of a diverging cell.

To explicitly consider the turning bay effects, the capacities for the two sub-cells are calculated as

$$N_{i+1,L}^{j} = N_{i+1,1}^{j} + N_{i+1,3}^{j}$$
$$N_{i+1,T}^{j} = N_{i+1,2}^{j} + N_{i+1,3}^{j} \quad (46)$$

Then the diverging model can be calculated as

$$y_i^j = \min\{\bar{D}_i^j, \bar{S}_{i+1}^j, \frac{\bar{S}_{i+1,L}^j}{\beta_{i+1,L}^j}, \frac{\bar{S}_{i+1,T}^j}{\beta_{i+1,T}^j}\}$$
$$y_{i,L}^j = y_i^j \beta_{i+1,L}^j \quad (47)$$
$$y_{i,T}^j = y_i^j \beta_{i+1,T}^j$$

where $\beta_{i+1,L}^j$ and $\beta_{i+1,T}^j$ are the ratios of left-turn and through vehicles inside the diverging zone, respectively. In the departure zone, the flow capacity Q_i^j depends on two factors: the saturation flow-rate s and the green time of the corresponding movement g_i^j, which can be formulated as

$$Q_i^j = s g_i^j \quad (48)$$

In (Gao et al., 2015), due to the existence of multiple lane groups at a signalized intersection, the traditional cell representation was further enhanced with the introduction of sub-cells. An example of the sub-cell representation is provided in ***Figure 9***. At the intersection, there are virtual cells A and E to handle diverging and merging traffic at approaches and exits. Virtual cell A contains cell a, and virtual cell E contains cell e. Cell A is a two-dimensional cell which contains three sub-cells for the left-turn, through, and right-turn movements. Cell E is a one dimensional cell which holds the merging flows to exit the intersection. In addition, there is an overlapping cell C to handle conflicts between left-turn and through movements when the left-turn is permissive. Under this cell presentation, the path of cells for the through and left-turn movements is $a \to A \to C \to E \to e$, while it is $a \to A \to E \to e$ for the right-turn movements. Vehicle movements between cell A and a, or between E and e, are implemented immediately since cells A and E are virtual cells. Therefore, to move from cell a to cell e, it takes two time steps for through and left-turn vehicles, and only one time step for right-turn vehicles.

With lane channelization, the number of vehicles in each virtual diverging sub-cell can be calculated as follows:

$$n_{A_i}^j = \sum_k \beta_{i,k}^j n_{a_k}^j$$

$$N_{A_i}^j = \sum_k \beta_{i,k}^j N_{a_k}^j \tag{49}$$

$$Q_{A_i}^j = \sum_k \beta_{i,k}^j Q_{a_k}^j$$

Here, $n_{A_i}^j$ is the number of vehicles in the virtual sub-cell A_i, i =left-turn, through, and right-turn. $\beta_{i,k}^j$ is the turning proportion of movement i from lane k at the approach a. $N_{A_i}^j$ and $Q_{A_i}^j$ are the maximum number of vehicles and the flow capacities for the sub-cell A_i, respectively. $N_{a_k}^j$ and $Q_{a_k}^j$ are the maximum number of vehicles and the flow capacities for the lane k at the approach a, respectively.

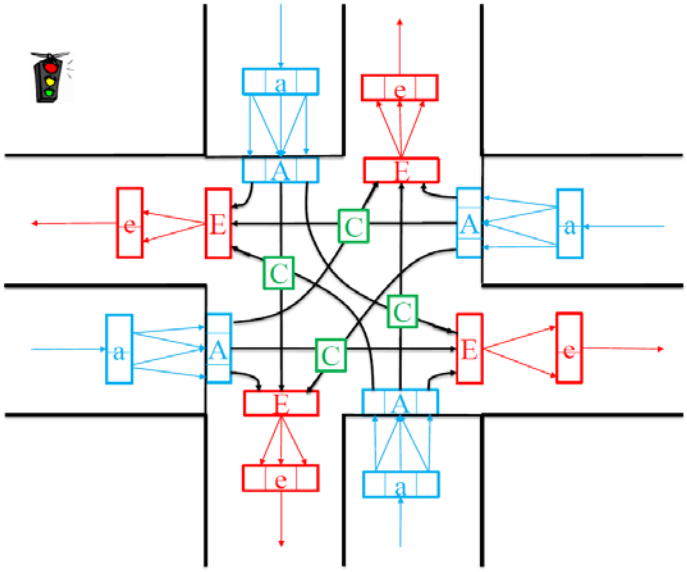


Figure 9 Sub-cell representation of a signalized intersection.

For the merging behavior at cell E, its upstream is three virtual sub-cells A_i, $i = 1, 2, 3$, while its downstream is cell e with M multiple lanes, e.g., M=3. Then the calculation of vehicle flows is formulated as follows:

Characteristics of cell E:

$$Q_E^j = \sum_m^M Q_{e_m}^j \tag{50}$$

$$N_E^j = \sum_m^M N_{e_m}^j$$

$$n_E^j = \sum_m^M n_{e_m}^j$$

If $S_E^j \geq \sum_{i=1}^3 D_{A_i}^j$

$$y_{A_i}^j = D_{A_i}^j$$

$$y_E^j = \sum_{i=1}^3 y_{A_i}^j \quad (51)$$

If $S_E^j < \sum_{i=1}^3 D_{A_i}^j$

$$y_{A_i}^j = mid\{D_{A_i}^j, S_E^j - \sum_{w=1,w\neq i}^3 D_{A_w}^j, p_{A_i} S_E^j\}$$

$$y_E^j = \sum_{i=1}^3 y_{A_i}^j \quad (52)$$

Flow distribution at cell e:

$$y_{e_i}^j = \frac{S_{e_i}^j}{S_E^j} y_E^j \quad (53)$$

Furthermore, to handle the conflicts between left-turn and through movements when the left-turn is permissive, an overlapping cell C is introduced. The constraints on the left-turn and through movements in cell C can be formulated as

$$\begin{aligned} n_{C,L}^j \times n_{C,Th}^j &= 0 \\ y_{C,L}^{j+1} \times y_{C,Th}^{j+1} &= 0 \end{aligned} \quad (54)$$

Whether a left-turn or a through vehicle to occupy the overlapping cell C is determined using a random draw with probability p_L for the left-turn movement and $1 - p_L$ for the through movement. The impact of traffic signal is modelled as Equation (29).

4. CTM-based signal control strategies

4.1 Objective functions

4.1.1 Delay minimization

In (Lo, 1999, 2001), the delay was defined as the additional time that a vehicle stays in a cell. Then the delay in cell i at time step j, d_i^j, can be calculated as

$$d_i^j = n_i^j - y_i^j \tag{55}$$

The total vehicle delay in a network can be calculated as

$$J = \sum_i \sum_j d_i^j \tag{56}$$

Then the objective function is to minimize the total network delay in Equation (56).

In (Almasri and Friedrich, 2005), it was assumed that during an updating time step, the relationship between a vehicle's delay and its speed follows a linear trend with a negative slope. That is, the maximum delay for a vehicle is the updating time step when it is fully stopped, while the minimum delay is zero when it travels at the free-flow speed. With this assumption, the delay calculation is the same as those in Equations (55) and (56) and thus omitted here. Simulation results demonstrated that compared with the Kimber-Hollis and Akcelik delay calculations, CTM provides good estimation accuracy under both uncongested and congested conditions.

In (Li, 2010), it was proposed to minimize the total system delay. But different from Equation (55), a weighted coefficient was introduced to modify the relative importance of each cell. Then the objective function can be formulated as

$$\min[total\ delay = \Delta T \sum_j^J \sum_u w_u (n_u^j - \sum_{d \in \Gamma(u)} y_d^j) \tag{57}$$

where w_u is the weighted coefficient, and $\Gamma(u)$ is the set of downstream cells of cell u.

4.1.2 Minimization of Performance Index

In TRANSYT, Performance Index (PI), which is a weighted combination of the delay and stops on all links in the network, is used as the measure of performance. A similar concept was then introduced in (Feldman and Maher, 2002). The performance index in CTM, PI_{CTM}, is defined as the average network

occupancies during the cycle time, which is calculated using the sum of occupancies in each cell during the cycle time divided by the cycle time. If a cycle has J time steps, then $\mathrm{PI}_{\mathrm{CTM}}$ can be calculated as

$$\mathrm{PI}_{\mathrm{CTM}} = (\sum_{j=1}^{J}\sum_{i=1}^{N} n_i^j)/\mathrm{M} \tag{58}$$

Then the optimization problem is to find a set of signal settings that generates the lowest $\mathrm{PI}_{\mathrm{CTM}}$.

4.1.3 Maximization of system throughput

In (Li, 2010), it was also proposed to optimize the signal settings to maximize the system throughput. The network throughput are defined as the sum of throughputs at all sink cells, and the objective function is formulated as

$$\max(Throughput = \sum_{j}^{J}\sum_{d\in\Phi}\sum_{u\in\Gamma^-(d)} y_u^j) \tag{59}$$

Where Φ is the set of sink cells, $\Gamma^-(d)$ is the set of upstream cells of cell d, and J is the total time steps.

4.1.4 Minimization of a combination of delay and early arrival flow

In (He et al., 2010), a different objective function was proposed to minimize the network delay as well as the early arrival flow, which can be formulated as follows:

$$\min(\alpha\sum_{i\in\Phi}\sum_{j}\mathrm{j}\times y_i^j + \beta\sum_{i\notin\Phi}\sum_{j}\mathrm{j}\times\ y_i^j) \tag{60}$$

where Φ is the set of destination cells. α is the coefficient for delay, while β is the coefficient for the early arrival flow to avoid vehicle holding at a cell even though there is capacity available downstream. In the minimization problem, constraints were imposed on the regular cells, origin cells, destination cells, as well as the cells adjacent to the signalized intersections. Other constraints such as minimum and maximum green times were also considered. In addition, it was required that the flow conservation should be hold so that after a simulation time period of T, all vehicles in the network are cleared.

4.1.5 Minimization of mean excess delay

Traditionally, demand is considered constant at the modeling of the CTM. But in reality, demand changes from time to time. In (Zhang et al., 2010), stochastic demands at the origins were considered. A

set of demand scenarios, $\Omega = \{1, 2, 3, \dots, \varphi\}$, are considered, and each scenario represents a set of demand combinations, $d^{\omega} = \{d_1^{\omega}, d_2^{\omega}, \dots, d_r^{\omega}, \dots\}$, $\forall r \in O$, which will happen with a probability of p_{ω}. Due to the randomness in the origin demands, (Zhang et al., 2010) aimed to minimize the excess delay which is caused by those high-consequence scenarios with collective probability of occurrence of $1-\alpha$, where α is a specified confidence level (e.g., 80%). Let's denote the delay for scenario ω as L_{ω}. Then, for $L_1 < L_2 < L_{\omega_a - 1} < L_{\omega_a} < \cdots < L_{\omega}$, the definition of α can be explained as $\sum_1^{\omega_\alpha} p_{\omega} \geq \alpha \geq \sum_{\omega}^{\omega_\alpha - 1} p_{\omega}$. That also means there only can be a chance of $1-\alpha$ to have delay greater than or equal to L_{ω_α}. Then the mean excess delay can be formulated as

$$Z_{\omega_\alpha} = \frac{1}{1-\alpha}\left[\left(\sum_{\omega}^{\omega_\alpha} p_{\omega} - \alpha\right) L_{\omega_\alpha} + \sum_{\omega_\alpha + 1}^{\varphi} p_{\omega} L_{\omega}\right] \tag{61}$$

This study tried to find a set of robust timing plans to minimize the mean excess delay.

4.2 Optimization methods

4.2.1 Mixed-integer programing problems

In the minimization problem in (Lo, 1999, 2001), constraints on minimum and maximum green times, and offsets were introduced, and the offset and the green times were used as decision variables. To eliminate the minimum sign in the calculation of boundary flux y_i^j, the "less than or equal to" constraints are used. Furthermore, the mixed-integer programing technique with the introduction of two binary variables is introduced to model the "if-then" conditions in Equation (29). To solve the mixed-integer-linear-programing (MILP) problems, commercial software packages such as CPLEX can be used.

4.2.2 Genetic Algorithms

Since the size of MILP problem increases dramatically as the network size increases, later in (Lo et al., 2001), a so-called dynamic intersection signal control optimization (DISCO) was proposed. The heuristic approach is based on the Genetic Algorithm (GA) and aims to find a good, rather than an optimal solution. In the GA problem, the durations of effective green and red phases for the intersection are coded as a series of chromosomes. Two types of GA methods are proposed: (i) network-wide GA (Net-GA) which

considers the timing plans of all intersections in the network simultaneously; (ii) sequential GA (S-GA) which only consider the timing plan for one intersection at a time. Analysis results showed that DISCO performs better than TRANSYT in optimizing the traffic signals with a reduction in delay up to 33% under a wide range of demand patterns. In (Lo and Chow, 2004), DISCO was extended to optimize the network with more flexible signal settings. Three different signal control strategies were analyzed: fixed-time (FT) plan, variable green fixed cycle (VGFC), and variable green no cycle (VGNC). Again, the GA algorithm was used. Analysis results showed that all DISCO plans outperform the existing plan. It was interesting to find that the FT plan performs better than the VGFC and VGNC plans, which may be due to the large variable space introduced by these two dynamic plans and the limitation of the GA algorithm to find the truly global solution. However, the study also showed that delay can be further reduced if the output from the FT plan is used as an initial solution in the VGFC and VGNC plans.

In (Almasri and Friedrich, 2005), the GA algorithm was used to optimize the offsets in the network. Two types of GA algorithms were used: (i) parallel GA (PGA) with a simultaneous search over all offset times; (ii) serial GA (SGA) with only one offset value and only one part of the chromosome updated each time. Simulation studies showed that SGA can reduce the CPU time significantly and performs as well as PGA when the study network has low interdependencies among the links. But it was argued that the performance of SGA may decay as the network gets larger with increasing multiple dependencies of the offsets. The C++ GAlib (Genetic algorithm library) written by Matthew Wall at MIT was used in this study.

In (Li, 2010), the GA method was used to obtain a near-optimal signal timing plan. The decision variables are the cycle length, the green time split, and the offset of each signal. Instead of having a common network cycle length, this study allowed the cycle length at certain intersections to be half of the common one. Simulation results demonstrated that the proposed model performs far better than TRANSYT-7F under a wide range of traffic conditions with less total delay and higher network throughput, especially at high traffic volumes.

Similar to the previous studies, the GA method was also used in (Zhang et al., 2010) to find a robust timing plan.

4.2.3 Linear relaxation combined with heuristic methods

It was argued in (He et al., 2010) that it becomes impossible to solve the MILP problem in the CTM with commercial software as the network scales up. At the same time, it also cannot guarantee the commonly-used GA method can reach a local optimal given the analysis time frame. Therefore, the Linear Relaxation (LR) method was introduced in the MILP problem in (He at al., 2010) to allow the 0-1 variables to take continuous values within the region of $[0, 1]$. After the problem is solved, non-integer solutions are obtained. Then three heuristic methods are used to convert these infeasible solutions into feasible ones: the dive-and-fix method, the ratio-cluster method, and the cumulative-departure method.

(i) The dive-and-fix method: the idea of the method is to find those fractional variables in the solutions from the LR problem, and try to round them to the nearest integers, either 0 or 1. Each time after the LR problem is solved, the most integer-like variable is rounded to be 0 or 1, and the LR problem is solved again. Such a process is repeated until all variables are integers of 0 or 1.

(ii) The ratio-cluster method: The fractional ratios obtained from the LR method can be considered as the green time allocations for a given cycle length. The ratio-cluster method tries to utilize the green time ratios obtained from the LR method to generate a feasible MILP solution. The method first keeps track of the changes of green times from the LR solution and divides the total time steps into several clusters. Then suitable integer cycles are then introduced into each cluster to match the average green ratios and minimize the cumulative green time error.

(iii) The cumulative-departure method: The smooth cumulative departure curve obtained from the LR solution should be the optimal one since it has less restrictions on the constraints of variables. Therefore, this method tries to find a feasible binary solution to generate a cumulative curve to better match the one obtained from the LR.

4.2.4 The hill-climbing method

Similar to TRANSYT, a hill-climbing method was also used in (Feldman and Maher, 2002) to compare the performance with the GA method. Given an initial set of signal timings, the program first calculates the PI for the whole network. Then the program changes the offset (and/or green times) of one

of the signals successively with predefined step sizes until a minimum PI is found. Each offset (and/or green times) is optimized in turn in this way, and a final set of signal settings is obtained by repeating this process a number of times. It was shown in (Feldman and Maher, 2002) that the hill-climbing method performs reasonably well if the time budget is low, while the GA algorithm performs better if the time budget is high.

PART III: Average models for signalized junctions

In this part, we introduce an analytical approach to derive invariant averaged models for signalized intersections. As a starting point, we focus on a simpler case, a signalized linear junction. We first introduce three models with binary signals and derive their averaged counterparts by replacing the cyclic signal control with its effective green ratio in Section 1. In Section 2, we apply the local forms of these three averaged models as entropy conditions at the junction. We solve the corresponding Riemann problems and derive the invariant forms of these three averaged models. In Section 3, we identify the correctness of these three invariant averaged models using the constraint of maximum averaged junction flux. We will show that only one of them can correctly catch such a constraint under various traffic conditions. We will also show that different non-invariant averaged models can lead to the same invariant form. In Section 4, we analyze the approximation accuracy of the correct invariant averaged model in a signalized ring road, considering the impact of initial densities, cycle lengths, and traffic flow fundamental diagrams. In Section 5, we extend our analytical framework to a more complicated case, a signalized merging junction. Following similar procedures, we derive the corresponding invariant averaged model for the signalized merging junction.

1. Models with binary signals and the averaged counterparts at a linear junction

For a signalized linear junction shown in ***Figure 10***, let's assume the traffic signal is installed at $x = 0$. The upstream section ($x < 0$) is denoted as link 1 with length of L_1, while the downstream section ($x > 0$) is denoted as link 2 with length of L_2. The traffic flow fundamental diagram is of the same type for both links and is denoted as $q_a = Q(k_a)$, $a \in \{1,2\}$.

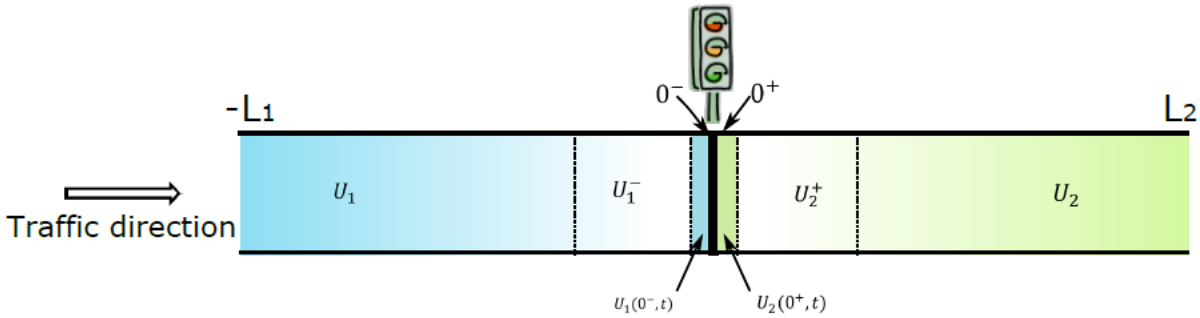


Figure 10 A signalized road link.

1.1 Three models with binary signals

On link$a$, the following definitions of demand and supply were introduced in (Lebacque, 1996) (similar to the maximum sending and receiving flows in (Daganzo, 1995)):

$$\begin{aligned} \mathrm{D_a(x,t)} &= \mathrm{Q(\min\{k_a(x,t), k_{a,c}\})} \\ S_a(x,t) &= Q(\max\{k_a(x,t), k_{a,c}\}) \end{aligned} \tag{62}$$

where $k_{a,c}$ is the critical density of link a. Then the flow-rate $q_a(x,t)$, capacity C_a, density $k_a(x,t)$, and speed $v_a(x,t)$ can be uniquely determined through its demand and supply using the following equations:

$$\begin{aligned} \mathrm{q_a(x,t)} &= \min\{\mathrm{D_a(x,t)}, \mathrm{S_a(x,t)}\} \\ \mathrm{C_a} &= \max\{\mathrm{D_a(x,t)}, \mathrm{S_a(x,t)}\} \\ \mathrm{k_a(x,t)} = \boldsymbol{\mathcal{R}}(\mathrm{D_a(x,t)/S_a(x,t)}) &= \begin{cases} \mathbf{D}^{-1}(\mathrm{C_a D_a(x,t)/S_a(x,t)}) & \mathrm{D_a(x,t) \le S_a(x,t)} \\ \mathbf{S}^{-1}(\mathrm{C_a S_a(x,t)/D_a(x,t)}) & \mathrm{D_a(x,t) > S_a(x,t)} \end{cases} \\ \mathrm{v_a(x,t)} &= \mathrm{q_a(x,t)/k_a(x,t)} \end{aligned} \tag{63}$$

Thus, different from traditional approaches that use density as the state variable, the demand and supply pair,$U_a(x,t) = (D_a(x,t), S_a(x,t))$, can be used to represent the traffic conditions (Jin, 2009). Initially, vehicles are assumed to be uniformly distributed along each link, and therefore, the initial conditions can be simply written as

$$\mathrm{U_a = U_a(x, t = 0) = (D_a, S_a)}, \quad \mathrm{a \in \{1,2\}} \tag{64}$$

where D_a and S_a are constant values.

For the traffic signal at $x = 0$, the cycle length is T, and the effective green time is ηT with $\eta \in (0,1)$. An indicator function $\delta(t)$ is introduced to describe the binary control of signals, which is formulated as

$$\delta(\mathrm{t}) = \begin{cases} 1 & \mathrm{t} \in [\mathrm{nT}, \mathrm{nT} + \eta\mathrm{T}) \\ 0 & \mathrm{t} \in [\mathrm{nT} + \eta\mathrm{T}, (\mathrm{n}+1)\mathrm{T}) \end{cases} \text{ with } \mathrm{n} = 0,1,2,\ldots \tag{65}$$

For constant loading problems (or with infinitely-long links), the junction flux q can be determined from the global settings of initial demand and supply as well as the binary signals, and can be written in the following three forms:

$$\begin{aligned} \mathrm{q} &= \mathrm{F}_1\big(\mathrm{U}_1, \mathrm{U}_2; \delta(\mathrm{t})\big) = \delta(\mathrm{t})\min\{\mathrm{D}_1, \mathrm{S}_2\} && \text{(a)} \\ \mathrm{q} &= \mathrm{F}_2(\mathrm{U}_1, \mathrm{U}_2; \delta(\mathrm{t})) = \min\{\delta(\mathrm{t})\mathrm{D}_1, \mathrm{S}_2\} && \text{(b)} \\ q &= F_3(U_1, U_2; \delta(t)) = \min\{D_1, \delta(t)S_2\} && \text{(c)} \end{aligned} \tag{66}$$

However, at the junction ($x = 0$), the following forms with the local settings of supply and demand should be used:

$$\begin{aligned} \mathrm{q} &= \mathrm{f}_1(\mathrm{U}_1(0^-, \mathrm{t}), \mathrm{U}_2(0^+, \mathrm{t}); \delta(\mathrm{t})) = \delta(\mathrm{t})\min\{\mathrm{D}_1(0^-, \mathrm{t}), \mathrm{S}_2(0^+, \mathrm{t})\} && \text{(a)} \\ \mathrm{q} &= \mathrm{f}_2(\mathrm{U}_1(0^-, \mathrm{t}), \mathrm{U}_2(0^+, \mathrm{t}); \delta(\mathrm{t})) = \min\{\delta(\mathrm{t})\mathrm{D}_1(0^-, \mathrm{t}), \mathrm{S}_2(0^+, \mathrm{t})\} && \text{(b)} \\ \mathrm{q} &= \mathrm{f}_3(\mathrm{U}_1(0^-, \mathrm{t}), \mathrm{U}_2(0^+, \mathrm{t}); \delta(\mathrm{t})) = \min\{\mathrm{D}_1(0^-, \mathrm{t}), \delta(\mathrm{t})\mathrm{S}_2(0^+, \mathrm{t})\} && \text{(c)} \end{aligned} \tag{67}$$

Definition 1

(Global and local flux functions)

Functions like $F(U_1, U_2)$ are called global flux functions since initial traffic conditions U_1 and U_2 are used. Functions like $f(U_1(x = 0^-, t), U_2(x = 0^+, t))$ are called local flux functions since only local traffic conditions such as $U_1(0^-, t)$ and $U_2(0^+, t)$ at the junction are used.

Definition 2

(Invariance)

For any global flux function $F(U_1, U_2)$, it is called invariant if the same global form can be derived by introducing its local from $f(U_1(0^-, t), U_2(0^+, t))$ as an entropy condition at the junction and solving the arising Riemann problems. Otherwise, it is non-invariant.

Theorem 3

The three models with binary signals in Equation (66) are invariant and are equivalent to each other.

Proof: During the red time period, $q = 0$, and thus the flux function is automatically invariant. During the green time period, we have $F_i(U_1, U_2; \delta(t) = 1) = \min\{D_1, S_2\}$ for $i = \{1,2,3\}$. In (Jin et al., 2009), it has been shown that such a flux function is invariant. Therefore, even though we have binary signals at the linear junction, the three models in Equation (66) are invariant and are equivalent to each other. ■

1.2 Averaged models and their properties

Due to the existence of traffic signal, the junction flux in Equation (66) periodically switches between zero and certain nonzero values, for example, the saturation flow-rate. Here, we first derive the averaged models from Equation (66) to eliminate such a cyclic pattern, then we analyze their properties using CTM simulations.

1.2.1 Averaged models

According to (Sanders et al., 2007), when a parameter in a system equation is periodic, we can simplify the system dynamics by averaging the parameter over its period. Since $\delta(t)$ in Equation (66) is T-periodic, its average can be computed as

$$\bar{\delta} = \frac{1}{T}\int_0^T \delta(s)ds = \eta. \tag{68}$$

Therefore, the three averaged models can be derived from Equation (66), which are formulated as follows:

$$\begin{aligned} q &= \tilde{F}_1(D_1, S_2; \eta) = \eta \min\{D_1, S_2\} && \text{(a)} \\ q &= \tilde{F}_2(D_1, S_2; \eta) = \min\{\eta D_1, S_2\} && \text{(b)} \end{aligned} \tag{69}$$

$$q = \tilde{F}_3(D_1, S_2; \eta) = \min\{D_1, \eta S_2\} \quad \text{(c)}$$

Correspondingly, their local forms are

$$q = f_1(D_1(0^-, t), S_2(0^+, t); \eta) = \eta \min\{D_1(0^-, t), S_2(0^+, t)\} \quad \text{(a)}$$

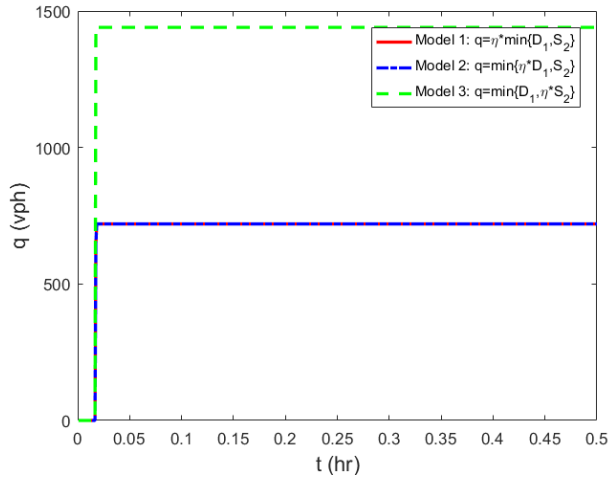
$$q = f_2(D_1(0^-, t), S_2(0^+, t); \eta) = \min\{\eta D_1(0^-, t), S_2(0^+, t)\} \quad \text{(b)} \quad (70)$$

$$q = f_3(D_1(0^-, t), S_2(0^+, t); \eta) = \min\{D_1(0^-, t), \eta S_2(0^+, t)\} \quad \text{(c)}$$

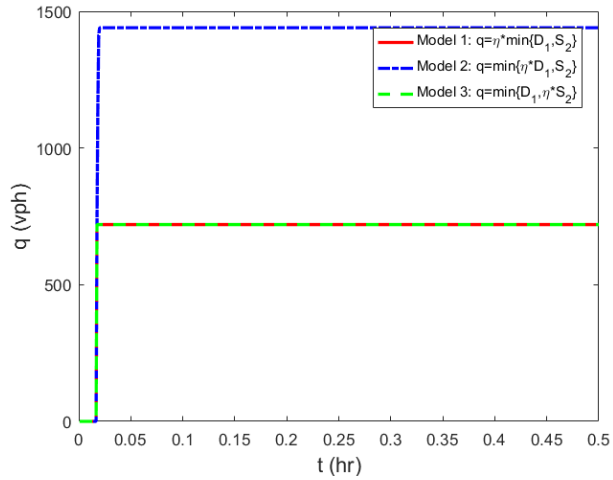
With Equation (70), the signalized linear junction is changed into an un-signalized one with a different local flux function $q = f_i(D_1(0^-, t), S_2(0^+, t); \eta)$, $i \in \{1,2,3\}$, at $x = 0$.

1.2.2 Properties

In ***Figure 11***, we provide average junction fluxes for the three averaged models in Equation (69) under constant loading scenarios (with fixed boundary demand and supply). The CTM simulation is used. Each link is 0.5mile long and is partitioned into 10 cells. The free-flow speed v_f is 60 mph, and thus the updating time interval Δt is 3 seconds with $CFL = 1$. The critical density k_c is 30 vpmpl and the jam density is 150 vpmpl, and therefore, the capacity C is 1800 vphpl. The simulation time is 0.5 hr, which is long enough to allow traffic to reach a stationary state. Initially, all links are empty.



(a) $2C_1 = C_2 = 3600$vph, $D_1 = 0.9C_1$, $S_2 = 0.5C_2$, and $\eta = 0.4$



(b) $C_1 = 2C_2 = 3600$vph, $D_1 = 0.5C_1$, $S_2 = 0.9C_2$, and $\eta = 0.4$

Figure 11 Properties of the three time-average models in Equation (69).

The firs case is when the upstream link has one lane but the downstream one has two. The upstream demand is $0.9C_1$, and the downstream supply is $0.5C_2$. At the signalized junction, the green ratio η is 0.4. In this case, we can find that both the upstream demand and downstream supply are greater than the maximum number of vehicles that can enter the intersection, i.e., $\min\{D_1, S_2\} > \eta \max\{C_1, C_2\}$. As illustrated in ***Figure 11***(a), under stationary traffic conditions, the third model (Equation (69c)) provides a different average junction flux from the one provided by the first (Equation (69a)) and the second models (Equation (69b)): the average junction flux q is 1440 vph ($\eta\ C_2$) for the third model, while it is 720 vph (ηC_1) for the first and the second ones. However, a different situation occurs when the upstream link has two lanes but the downstream one has only one. In this case, the upstream demand is $0.5C_1$, and the downstream supply is $0.9C_2$. The green ratio remains 0.4. As illustrated in ***Figure 11***(b), the average junction flux of the second model converges to $q = \eta C_1 = 1440$ vph, while it converges to $q = 720$ vph for the first and the third models. Obviously, given different road geometries and traffic conditions, these three models return different average junction fluxes. More importantly, it is hard to tell which one is correct since they are derived from the three forms with binary signals in Equation (66), which have been proven to be equivalent in **Theorem 3**.

2. Invariant average models

In this section, we will first introduce the network kinematic wave model, which include both link models and junction models. Within this framework, we will solve the Riemann problems arising at the junctions and obtain the invariant averaged models.

2.1 Network kinematic wave model

2.1.1 Link models

Generally speaking, link models for freeways can be applied to arterial networks. In kinematic wave theory, traffic is modeled as a continuous media. Traffic flow variables such as flow-rate $q_a(x,t)$, speed $v_a(x,t)$, and density $k_a(x,t)$ are usually used to describe the state at (x,t) of link a. With the assumption of a traffic flow fundamental diagram,$q_a = Q(k_a)$, the following LWR model (Lighthill and Whitham, 1955; Richards, 1956) is used to describe the traffic dynamics on link a:

$$\frac{\partial \mathrm{k_a(x,t)}}{\partial \mathrm{t}} + \frac{\partial \mathrm{Q\big(k_a(x,t)\big)}}{\partial \mathrm{x}} = 0 \tag{71}$$

In (Daganzo, 1994), a so-called Cell Transmission Model (CTM) was introduced to numerically solve Equation (71) using the Godunov method (Godunov, 1959): (i) a link is partitioned into cells with equal length of Δx, and (ii) the time is divided into intervals with equal duration of Δt. The selection of Δx and Δt strictly follows the CFL condition (Courant et al., 1928) that requires a vehicle cannot travel across a cell (longer than Δx) during time Δt. A cell's density at every time step Δt is updated according to the difference of its boundary fluxes.

However, using the discrete version of the CTM (Daganzo, 1994, 1995), it is impossible to obtain analytical results, for example, traffic stationary states, in the network. Therefore, in (Jin et al., 2009; Jin, 2010, 2014a, 2012a), a continuous version of the CTM that allows the cell size Δx and the time step Δt approach to zero was proposed. Different from the discrete version in (Daganzo, 1994, 1995), the demand and supply pair, $U_a(x,t) = (D_a(x,t), S_a(x,t))$, is used as the state variable. In the continuous formulation, Riemann problems with jump initial conditions are found and analytically solved in the supply-demand space. Through the analysis, it is shown that stationary states exist and will eventually dominate the traffic

conditions on the link after some time. Meanwhile, there exist interior states that take infinitesimal space in the continuous formulation and only occupy one cell in the numerical solution.

2.1.2 Junction models

For signalized intersections, time-dependent binary variables are usually used to mimic the control logic of green-red intervals. Then for a given green interval, junction models for freeways can be applied to assign the traffic from the upstream approaches to the downstream ones. Examples of such junction models are provided in Equation (66) with the local forms in Equation (67). Furthermore, if we consider the averaged performance of traffic, we can use averaged models to eliminate the cyclic patterns of signal control. Examples of such junction models are provided in Equation (69) with the local forms in Equation (70).

The above two types of models can be applied at the junction together with the link models to update the boundary fluxes and the corresponding densities, which as a result forms the network kinematic wave model.

2.2 Derivation of invariant averaged models

In this subsection, we will combine the continuous CTM (Jin et al., 2009; Jin, 2010, 2014a, 2012a) together with the averaged models in Equation (69). From that, we will solve the Riemann problems and obtain the invariant forms of the averaged models.

2.2.1 Riemann Problems

In the continuous CTM, as demonstrated in (Jin et al, 2009), there exist three types of traffic states on an infinitely-long road link:

- Initial states: $U_1 = (D_1, S_1)$ and $U_2 = (D_2, S_2)$;
- Stationary states: $U_1^- = (D_1^-, S_1^-)$ and $U_2^+ = (D_2^+, S_2^+\})$;
- Interior states: $U_1(0^-, t) = (D_1(0^-, t), S_1(0^-, t))$ and $U_2(0^+, t) = (D_2(0^+, t), S_2(0^+, t))$.

As an example, the locations of these traffic states at the signalized linear junction are provided in ***Figure 10***. Correspondingly, we have the following three types of Riemann problems:

- Type I: the Riemann problem between U_1 and U_1^-, or between U_2 and U_2^+;
- Type II: the Riemann problem between $U_1(0^-,t)$ and U_1^-, or between $U_2(0^+,t)$ and U_2^+;
- Type III: the Riemann problem between $U_1(0^-,t)$ and $U_2(0^+,t)$.

To uniquely solve the above Riemann problems, we need to introduce the following two entropy conditions. The first one is the constraints on wave directions inside a link: (i) the Riemann problem between U_1 and U_1^- can not have positive waves, while the Riemann problem between U_2 and U_2^+ can not have negative waves; (ii) the Riemann problem between $U_1(0^-,t)$ and U_1^- can not have negative waves, while the Riemann problem between $U_2(0^+,t)$ and U_2^+ can not have positive waves. The second entropy condition is driver's macroscopic behaviors such as fair merging and First-In-First-Out (FIFO) at the junction, e.g., Equation (70).

2.2.2 *Solutions*

Lemma 4

With Equation (70a) applied as the entropy condition at the signalized linear junction, stationary and interior states can take the following values:

[1] When $min\{D_1, S_2\} > \eta\, min\{C_1, C_2\}$,

$U_1^- = U_1(0^-,t) = (C_1, \eta\, min\{C_1, C_2\})$, *and* $U_2^+ = U_2(0^+,t) = (\eta\, min\{C_1, C_2\}, C_2)$.

[2] When $D_1 < S_2$ *and* $D_1 \le \eta\, min\{C_1, C_2\}$,

$U_1^- = (D_1, C_1)$, $U_1(0^-,t) = (D_1/\eta, C_1)$, *and* $U_2^+ = U_2(0^+,t) = (D_1, C_2)$.

[3] When $D_1 > S_2$ *and* $S_2 \le \eta\, min\{C_1, C_2\}$,

$U_1^- = U_1(0^-,t) = (C_1, S_2)$, $U_2^+ = (C_2, S_2)$, *and* $U_2(0^+,t) = (C_2, S_2/\eta)$.

[4] When $S_2 = D_1 \le \eta\, min\{C_1, C_2\}$, $U_1^- = (D_1, C_1)$, *and* $U_2^+ = (C_2, S_2)$.

a) If $U_2(0^+,t) = (C_2, S_2/\eta)$, $U_1(0^-,t) = (D_1(0^-,t), S_1(0^-,t))$ *with* $S_1(0^-,t) \ge D_1$ *and* $D_1(0^-,t) \ge D_1/\eta$.

b) If $U_1(0^-,t) = (D_1/\eta, C_1)$, $U_2(0^+,t) = (D_2(0^+,t), S_2(0^+,t))$ *with* $S_2(0^+,t) \ge S_2/\eta$ *and* $D_2(0^+,t) \ge S_2$.

The proof is provided in **Appendix A**.

At the signalized linear junction, the junction flux can be calculated from the stationary states. Then according to **Lemma 4**, we have the following theorem:

Theorem 5

With Equation (70a) applied as the entropy condition at the signalized linear junction, the junction flux can take the following values:

[1] When $min\{D_1, S_2\} > \eta\, min\{C_1, C_2\}$, $q = \eta\, min\{C_1, C_2\}$.

[2] When $D_1 \le min\{S_2, \eta\, min\{C_1, C_2\}\}$, $q = D_1$.

[3] When $S_2 \le min\{D_1, \eta\, min\{C_1, C_2\}\}$, $q = S_2$.

Therefore, the global averaged model can be written as

$$q = \hat{F}_1(D_1, S_2; \eta) = min\{D_1, S_2, \eta C_1, \eta C_2\} \qquad (72)$$

The proof is simple and thus is omitted here.

Corollary 6

The global averaged model in Equation (72) is invariant.

Proof: If we apply the following local averaged model as the entropy condition at the signalized junction,

$$\mathrm{q} = \hat{f}_1(\mathrm{D}_1(0^-, \mathrm{t}), \mathrm{S}_2(0^+, \mathrm{t}); \eta) = \min\{\mathrm{D}_1(0^-, \mathrm{t}), \mathrm{S}_2(0^+, \mathrm{t}), \eta \mathrm{C}_1, \eta \mathrm{C}_2\} \qquad (73)$$

We can derive the same averaged model in Equation (72). According to **Definition 1**, Equation (72) is invariant since it has the same structure as Equation (73). ■

Theorem 7

With Equations (70b) and (70c) applied as the entropy conditions at the signalized linear junction, we can derive the following invariant averaged models:

$$\textit{For Equation (70b)} \qquad q = \hat{F}_2(D_1, S_2; \eta) = min\{D_1, S_2, \eta C_1\} \qquad \textit{(a)} \quad (74)$$

For Equation (70c) $q = \hat{F}_3(D_1, S_2; \eta) = min\{D_1, S_2, \eta C_2\}$ *(b)*

Proof: Similar to the proof in **Lemma 4**, we can derive Equation (*74*) from Equations *(70b) and (70c)*. Furthermore, if we apply the following local averaged models as entropy conditions at the signalized junction,

$$\begin{aligned} \mathrm{q} &= \hat{f}_2(\mathrm{D}_1(0^-,\mathrm{t}), \mathrm{S}_2(0^+,\mathrm{t}); \eta) = \min\{\mathrm{D}_1(0^-,\mathrm{t}), \mathrm{S}_2(0^+,\mathrm{t}), \eta \mathrm{C}_1\} \\ \mathrm{q} &= \hat{f}_3(\mathrm{D}_1(0^-,\mathrm{t}), \mathrm{S}_2(0^+,\mathrm{t}); \eta) = \min\{\mathrm{D}_1(0^-,\mathrm{t}), \mathrm{S}_2(0^+,\mathrm{t}), \eta \mathrm{C}_2\} \end{aligned} \tag{75}$$

we can derive the same averaged models in Equation (*74*). Therefore, Equation (*74*) is invariant.■

Corollary 8

For the averaged models in Equation (69), they are non-invariant but have the same invariant forms in Equations (72) and (74).

Proof: For the averaged models in Equation (69), their local forms are provided in Equation (70). According to **Theorem 5** and **Theorem 7**, the derived invariant forms are different from those in Equation (69), and therefore, the models in Equation (69) are non-invariant.■

3. Comparison of averaged models

3.1 Constraint of maximum average junction flux

After deriving the invariant averaged models, we are able to identify their correctness by applying the constraint of maximum average junction flux. When the capacities of the upstream and downstream links are the same, the invariant averaged models in Equations (72) and (74) are identical. However, they have different properties when the upstream and downstream capacities are different. Then we have the following theorem.

Theorem 9

The invariant averaged models in Equation (74) are incorrect since they fail to meet the constraint of maximum average junction flux when the capacities of the upstream and downstream links are different. That is to say, the invariance property of an average model does not guarantee its correctness.

Proof: For the signalized linear junction, we have the following two capacity combinations for the upstream and downstream links.

- **C1: $C_1 < C_2$.**

  In this case, Equation (72) and is the same as Equation (74a). In ***Figure 12***(a) and ***Figure 12***(b), we provide the solutions of stationary states for the three invariant averaged models with $C_1 < C_2$ in the $D_1 - S_2$ space. Red dots indicate initial states, while blue ones indicate stationary states. We find that when $D_1 \leq \eta C_1$ or $S_2 \leq \eta C_1$, the solutions of stationary states (i.e., the average junction fluxes) are the same for these three models. However, when both D_1 and S_2 are greater than ηC_1, the stationary states for the three models are different. It is found that the average junction flux is bounded by the upstream capacity constraint ηC_1 for the models in Equations (72) and (74a), while it is bounded by $\min\{\eta C_2, C_1\}$ for the model in Equation (74b).

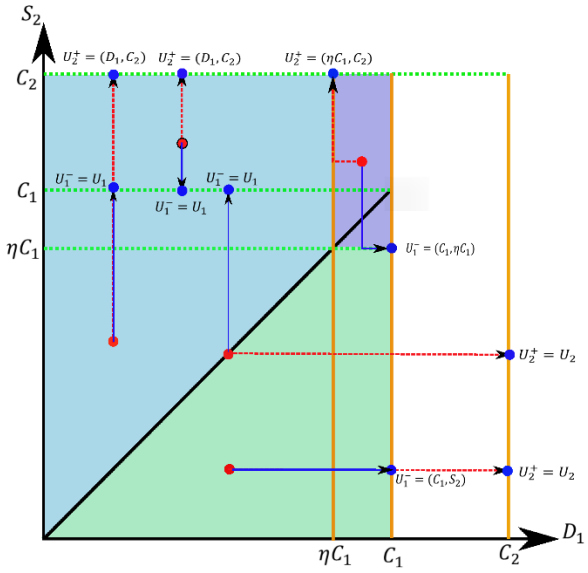
  With the signal control in Equation (66), when the downstream supply is high enough to accommodate all upstream vehicles during the effective green time, the maximum junction flux is equal to the upstream capacity. But it reduces to zero when the traffic light turns red. Therefore, the maximum average junction flux can only be ηC_1. The average junction flux from Equation (74b) is higher than this value, which means Equation (74b) fails to capture the upstream capacity constraint.

- **C2: $C_1 > C_2$.**

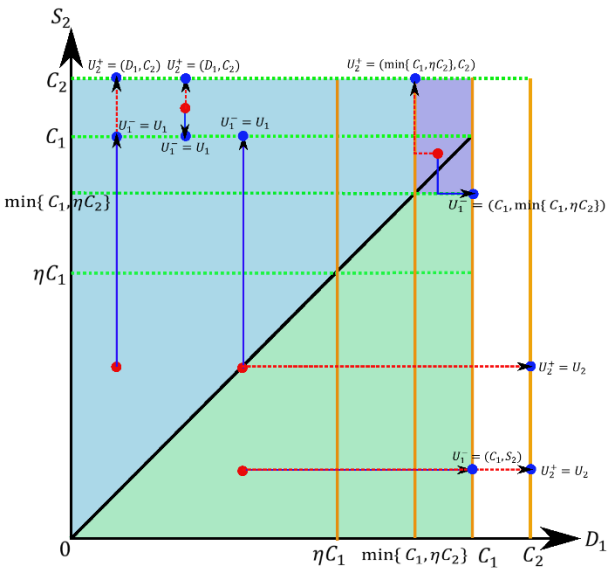
  In this case, Equation (72) is the same as Equation (74b). Similarly, we provide the solutions of stationary states for the three averaged models with $C_2 < C_1$ in ***Figure 12***(c) and ***Figure 12***(d). We find that these three models are the same when $D_1 \leq \eta C_2$ or $S_2 \leq \eta C_2$. However, they are different when both D_1 and S_2 are greater than ηC_2. In this case, the average junction flux is bounded by $\min\{\eta C_1, C_2\}$ for the model in Equation (74a) and by the downstream capacity constraint ηC_2 for the models in Equations (72) and (74b).

With the signal control in Equation (66), when the upstream demand is high enough to fully use the effective green time in each cycle, the maximum junction flux is equal to the downstream capacity. Again, due to the existence of red interval, the junction flux is zero for $1-\eta$ of the cycle. Therefore, the maximum average junction flux can only be ηC_2. The average junction flux from Equation (74a) is higher than this value, which means Equation (74a) fails to capture the downstream capacity constraint.

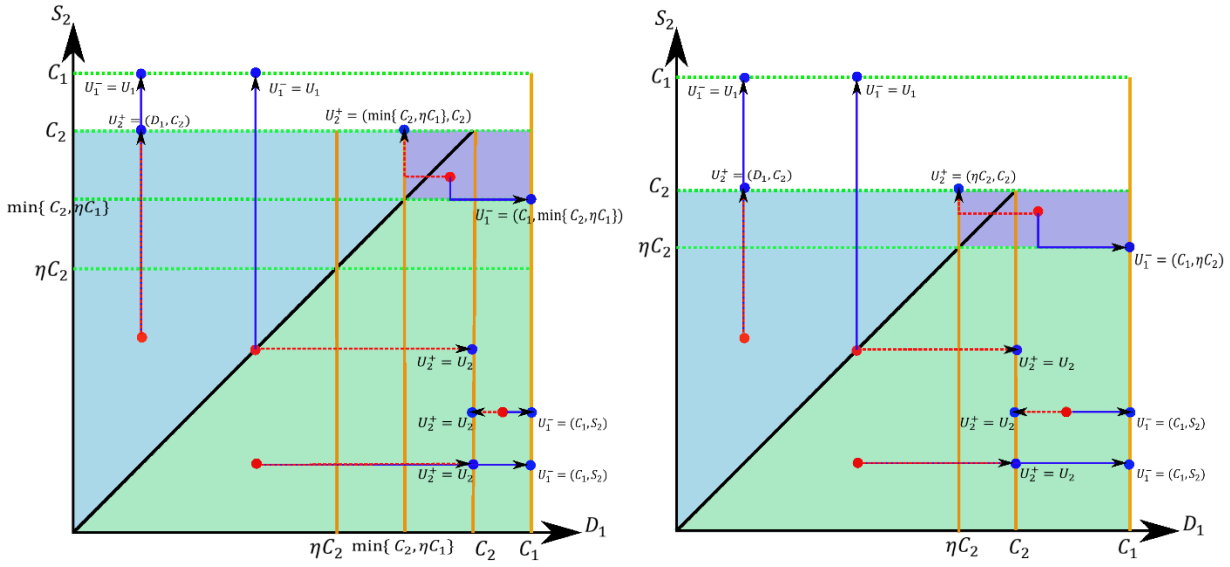
From the above analysis, we find that only the invariant averaged model in Equation (72) can meet the constraint of maximum average junction flux at the signalized linear junction. The invariant averaged models in Equation (74) are incorrect since they fail to meet this constraint when the capacities of the upstream and downstream links are different. ■



(a) $C_1 < C_2$ and $q = \min\{D_1, S_2, \eta C_1\}$



(b) $C_1 < C_2$ and $q = \min\{D_1, S_2, \eta C_2\}$



(c) $C_1 > C_2$ and $q = \min\{D_1, S_2, \eta C_1\}$ (d) $C_1 > C_2$ and $q = \min\{D_1, S_2, \eta C_2\}$

Figure 12 Solutions of stationary states for the three invariant averaged models.

3.2 Another average model

In (Han et al., 2014), an averaged model was proposed for a signalized merging junction with the consideration of effective supplies in the downstream. If one of the upstream links is empty and has zero demand, the signalized merging junction is changed into the signalized linear junction shown in ***Figure 10***. Then the averaged model in (Han et al., 2014) is simplified as

$$\mathrm{q} = \tilde{F}_4(\mathrm{D}_1, \mathrm{S}_2; \eta) = \min\{\mathrm{D}_1, \eta\, \mathrm{S}_2'\} = \min\{\mathrm{D}_1, \eta \mathrm{S}_2, \eta \mathrm{C}_1\} \tag{76}$$

Here, S_2' is the effective supply and is defined as $S_2' = \min\{S_2, C_1\}$. Correspondingly, its local form at $x = 0$ is

$$\mathrm{q} = \mathrm{f}_4(\mathrm{D}_1(0^-, \mathrm{t}), \mathrm{S}_2(0^+, \mathrm{t}); \eta) = \min\{\mathrm{D}_1(0^-, \mathrm{t}), \eta \mathrm{S}_2(0^+, \mathrm{t}), \eta \mathrm{C}_1\} \tag{77}$$

Theorem 10

The averaged model in Equation (76) is non-invariant but has the same invariant form as Equation (72). That means different non-invariant averaged models can have the same invariant form.

Proof: With the entropy condition in Equation (77) applied at the signalized junction, we can get the following stationary and interior states:

- When $\min\{D_1, S_2\} > \eta \min\{C_1, C_2\}$,

  $U_1^- = U_1(0^-, t) = (C_1, \eta \min\{C_1, C_2\})$, and $U_2^+ = U_2(0^+, t) = (\eta \min\{C_1, C_2\}, C_2)$.

- When $D_1 < S_2$ and $D_1 \leq \eta \min\{C_1, C_2\}$,

  $U_1^- = U_1(0^-, t) = (D_1, C_1)$, and $U_2^+ = U_2(0^+, t) = (D_1, C_2)$.

- When $D_1 > S_2$ and $S_2 \leq \eta \min\{C_1, C_2\}$,

  $U_1^- = U_1(0^-, t) = (C_1, S_2)$, $U_2^+ = (C_2, S_2)$, and $U_2(0^+, t) = (C_2, \frac{S_2}{\eta})$.

- When $S_2 = D_1 \leq \eta \min\{C_1, C_2\}$, $U_1^- = (D_1, C_1)$, and $U_2^+ = (C_2, S_2)$.

  a) If $U_2(0^+, t) = (C_2, \frac{S_2}{\eta})$, $U_1(0^-, t) = (D_1(0^-, t), S_1(0^-, t))$ with $S_1(0^-, t) \geq D_1$ and $D_1(0^-, t) \geq D_1$.

  b) If $U_1(0^-, t) = (D_1, C_1)$, $U_2(0^+, t) = (D_2(0^+, t), S_2(0^+, t))$ with $S_2(0^+, t) \geq \frac{S_2}{\eta}$ and $D_2(0^+, t) \geq S_2$.

  c) If $D_1 = S_2 = \eta C_1$, $U_1(0^-, t) = (D_1(0^-, t), S_1(0^-, t))$ with $S_1(0^-, t) \geq D_1$ and $D_1(0^-, t) \geq D_1$; $U_2(0^+, t) = (D_2(0^+, t), S_2(0^+, t))$ with $S_2(0^+, t) \geq \frac{S_2}{\eta}$ and $D_2(0^+, t) \geq S_2$.

Based on the stationary states, we can easily derive the invariant averaged model which is the same as Equation (*72*). Note that Equations (69a) and *(76)* are two different non-invariant averaged models, but they have the same invariant form in Equation (*72*). ■

4. Approximation accuracy on a signalized ring road

To evaluate the approximation accuracy of the invariant averaged model in Equation (*72*), we consider a signalized ring road shown in ***Figure 13***, which is formed by connecting the downstream exit of the signalized junction to its upstream entrance in ***Figure 10***. The ring road is a one-lane roadway with a length of L. The downstream and the upstream to the signal are labeled as $x = 0$ and $x = L$, respectively. Initially, vehicles are uniformly distributed along the ring.

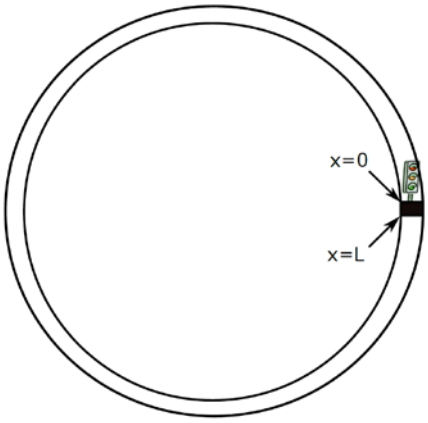


Figure 13 A signalized ring road.

4.1 Impacts of initial densities and cycle lengths

Due to the existence of signal control, it is hard to obtain analytical solutions of traffic stationary states on the signalized ring road. Therefore, the CTM simulation (Daganzo, 1994, 1995) is used. The ring is one mile long and is equally divided into 150 cells, and thus the updating time step Δt is 0.4s with $CFL = 1$. The triangular fundamental diagram,$q = Q_t(k) = \min\{v_f k, w(k_j - k)\}$, with $v_f = 60$ mph, $w = 15$ mph, and $k_j = 150$ vpm is used, and therefore, the capacity C is 1800 vph. The total simulation time is 2 hours, which is considered long enough to allow traffic to reach a stationary state. For the traffic signal, the effective green ratio is constant with $\eta = 0.5$.

To unveil the impact of initial densities, we set the cycle length to be constant, e.g., $T = 60s$. In ***Figure 14***, we show the evolution patterns of junction fluxes under four different initial densities. The blue solid lines represent the junction fluxes with signal control. It is clear to see that regardless of the initial densities, junction fluxes become periodic after some time, which indicates traffic in the signalized ring road has reached a stationary state (Jin et al., 2013). Therefore, average junction fluxes are obtained by averaging the periodic ones over the last four cycles and are provided as blue dashed-dotted lines in the figure.

Different from the case with signal control, it is possible to analytically derive all stationary states on the signalized ring road with the invariant averaged model in Equation (*72*). According to (Jin, 2012c), there exist three types of stationary states on a road link: strictly under critical (SUC), over critical (OC),

and zero-speed shockwave (ZS). When $C_1 = C_2 = C$, the entropy condition (Equation (73) at the signalized junction is now changed to $q = \min\{D(L,t), S(0,t), \eta C\}$. Since initially vehicles are uniformly distributed, we have $U(x,0) = (D(x,0), S(x,0)) = (D(L,0), S(0,0))$ for $x \in [0,L]$. When traffic on the signalized ring road is stationary after $t > t_0 \geq 0$, we can have the following possible stationary states:

- If $D(L,0) \leq \min\{S(0,0), \eta C\}$, only SUC stationary states can exist with $q = D(L,0)$. That means $U(x,t) = (D(x,t), S(x,t)) = (D(L,0), C)$ for $x \in [0,L]$ and $t > t_0 = 0$.
- If $S(0,0) \leq \min\{D(L,0), \eta C\}$, only OC stationary states can exist with $q = S(0,t)$. That means $U(x,t) = (D(x,t), S(x,t)) = (C, S(0,0))$ for $x \in [0,L]$ and $t > t_0 = 0$.
- If $\eta C < \min\{D(L,0), S(0,0)\}$, only ZS stationary states can exists with $q = \eta C$. In this case, for $t > t_0 > 0$, we have $U(x,t) = (D(x,t), S(x,t)) = (\eta C, C)$ for $x \in [0, \alpha L]$, and $U(x,t) = (D(x,t), S(x,t)) = (C, \eta C)$ for $x \in (\alpha L, L]$, where $\alpha \in (0,1)$.

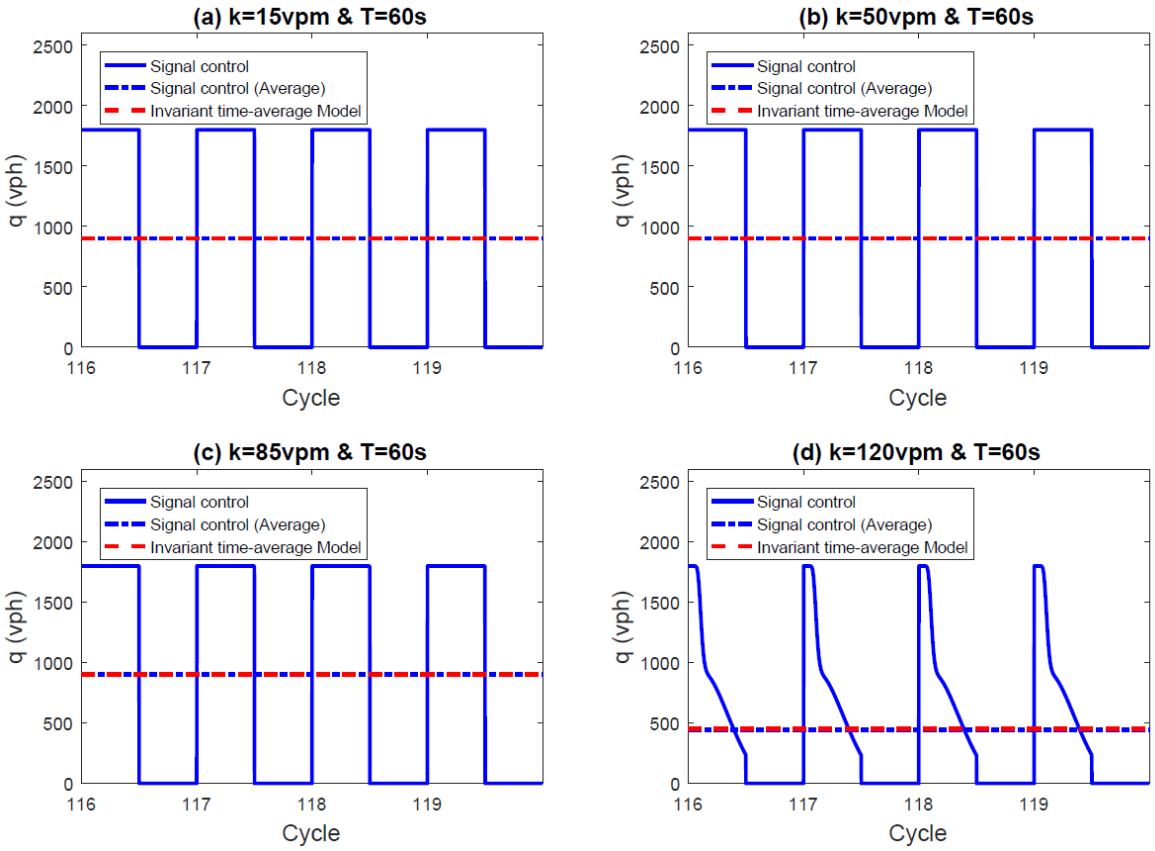


Figure 14 Junction fluxes with the same cycle length but different initial densities.

We provide the average junction fluxes derived from the invariant averaged model as red dashed lines in ***Figure 14***. From the figure, we find that under our current settings, the derived average junction fluxes are the same as the averaged ones with signal control.

Similarly, to unveil the impact of cycle lengths, we set the initial density to be constant, e.g., $k = 15$vpm. In ***Figure 15***, we show the junction fluxes with four different cycle lengths. From the figure, we can find that regardless of the cycle lengths, the junction fluxes under signal control become periodic after some time, which confirms that traffic on the signalized ring road has reached a stationary state. However, different from ***Figure 14***, we find that with the same initial density, the average junction fluxes derived from the invariant averaged model deviate from the ones with signal control as the cycle length increases. Similar patterns are also found with other initial densities, which indicates the approximate accuracy is impacted by the cycle lengths: long cycle lengths can reduce the approximation accuracy.

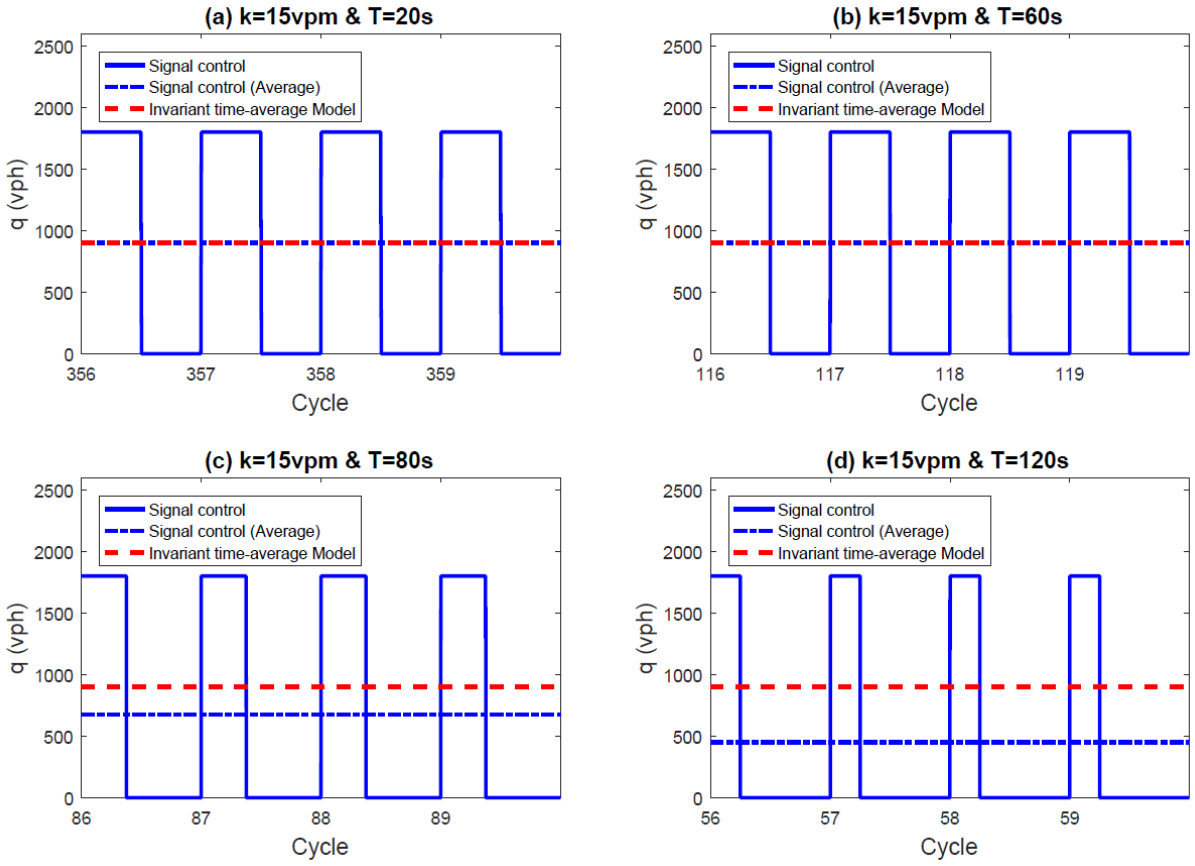
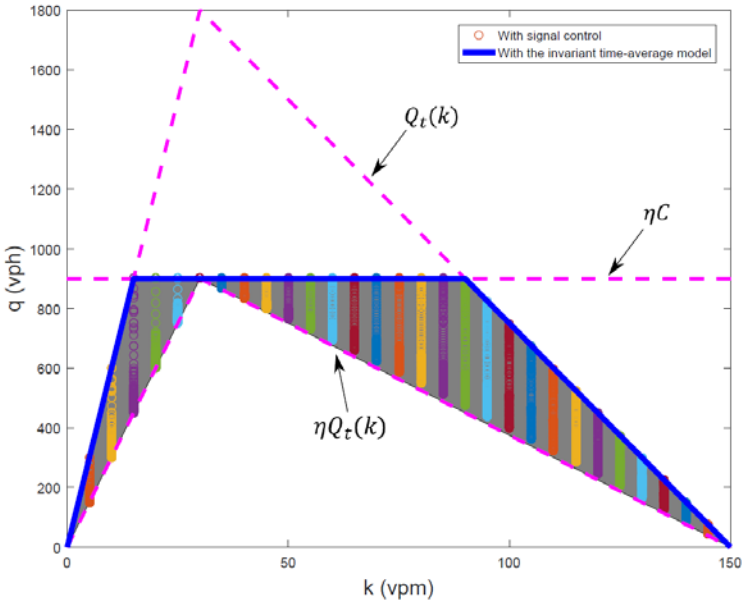


Figure 15 Junction fluxes with the same initial density but different cycle lengths.

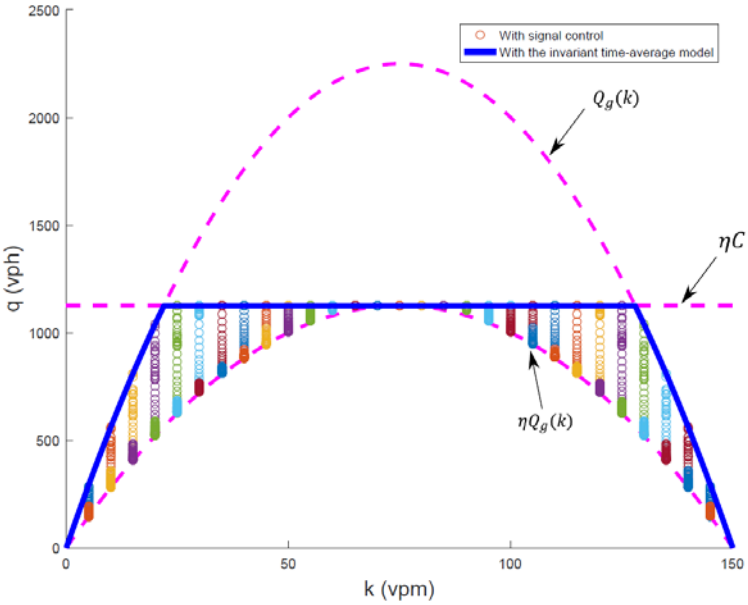
4.2 Differences in the macroscopic fundamental diagram

In this subsection, we want to analyze the difference between signal control (Equation (66)) and the invariant averaged model (Equation (*72*)) on the macroscopic fundamental diagram (MFD) (Geroliminis and Daganzo, 2008). With signal control, the CTM simulation is used, and the cycle length varies from 4s to 720s with $\eta = 0.5$. The simulation results are provided as circles in ***Figure 16***(a). From the figure, we have the following observations: (i) consistent with ***Figure 15***, for a given average network density, the average junction flux can take different values, which is related to the cycle length; and (ii) the average junction fluxes are bounded inside the shaded region formed by $Q_t(k)$, $\eta Q_t(k)$, and ηC, which is also shown in the figure. As a comparison, the derived MFD with the invariant averaged model is shown as the blue solid line in the figure. We find that the average junction fluxes derived from the invariant averaged model are on the upper bound of the MFD with signal control.

Furthermore, we analyze the property of the invariant averaged model with the Greenshields' fundamental diagram (Greenshield et al., 1935), $q = Q_g(k) = v_f k(1 - \frac{k}{k_j})$, which is a strictly concave function. In the CTM simulation, the free-flow speed is $v_f = 60$ mph, and the jam density is $k_j = 150$ vpm. The cycle length ranges from 4s to 360s. In ***Figure 16***(b), we provide the MFDs with both signal control and the invariant averaged model. From the figure, we find that the invariant averaged model is not sensitive to the types of fundamental diagrams since similar patterns in ***Figure 16***(a) can be observed in ***Figure 16***(b).



(a) With a triangular fundamental diagram



(b) With the Greenshields' fundamental diagram

Figure 16 Macroscopic fundamental diagrams of the signalized ring road.

5. Extension to a signalized merging junction

In this section, we will extend the proposed analytical framework to more complicated signalized intersections, e.g., a signalized merging junction. In ***Figure 17***, a 2×1 signalized merging junction is

provided. Traffic signal is installed at $x = 0$. The two upstream links are denoted as links 1 and 2, respectively, with $x < 0$, and the downstream link is denoted as link 3 with $x > 0$.

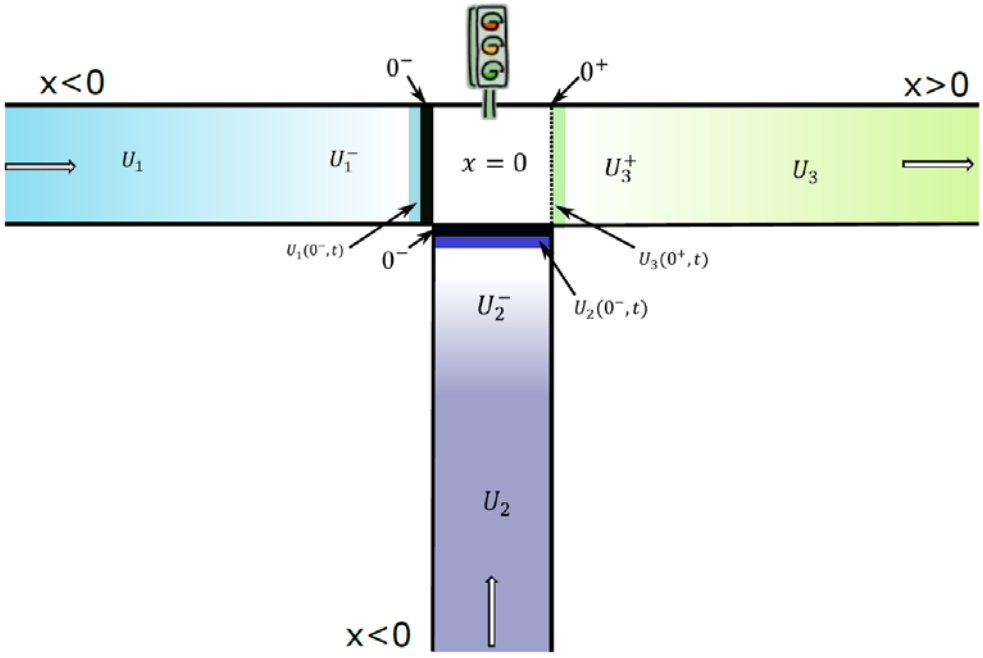


Figure 17 A signalized merging junction.

5.1 Model of binary signals at the merging junction and its averaged counterpart

5.1.1 Model with binary signals

For the signal settings at the merging junction, two phases are assigned to the two upstream links: phase 1 for link 1 and phase 2 for link 2. The cycle length is denoted as T, and each phase has the same lost time Δ. The green ratio is denoted as π_1 for link 1 and π_2 for link 2. If we consider the yellow and all red period is equal to the lost time period, the effective green time is $\pi_1 T$ for link 1 and $\pi_2 T$ for link 2, and $(\pi_1 + \pi_2)T = T - 2\Delta$. Then the following two indicators are used to describe the periodic signal regulation at the signalized merging junction.

$$\delta_1(t; T, \Delta, \pi_1) = \begin{cases} 1 & t \in [nT, nT + \pi_1 T) \\ 0 & \text{otherwise} \end{cases} \quad n \in \mathbb{N}_0,$$
$$\delta_2(t; T, \Delta, \pi_1) = \begin{cases} 1 & t \in [nT + \Delta + \pi_1 T, (n+1)T - \Delta) \\ 0 & \text{otherwise} \end{cases} \quad n \in \mathbb{N}_0, \tag{78}$$

where $\mathbb{N}_0 = \{0,1,2,3,\ldots\}$. With infinitely long links or under the case of fixed boundary demands and supplies, the junction fluxes are globally determined by the initial conditions as well as the signal control. Therefore, with Equation (78), the junction fluxes can be computed as:

$$\begin{gathered} q_i = F^i\left(U_i, U_j, U_3, \delta_i(t), \delta_j(t)\right) = \delta_i(t)\min\{D_i, S_3\}, \text{for } i,j = 1,2 \text{ and } i \neq j \\ q_3 = q_1 + q_2 \end{gathered} \tag{79}$$

At $x = 0$, the following local form is applied:

$$\begin{aligned} q_i &= f^i\left(U_i(0^-,t), U_j(0^-,t), U_3(0^+,t), \delta_i(t), \delta_j(t)\right) \\ &= \delta_i(t)\min\{D_i(0^-,t), S_3(0^+,t)\} \text{ for } i,j = 1,2 \text{ and } i \neq j \\ q_3 &= q_1 + q_2 \end{aligned} \tag{80}$$

Note that, similar to Equation (66a), signal control is applied to both the upstream demands and downstream supplies in Equations (7982) and (80).

5.1.2 The averaged counterpart

Due to the existence of signal control, the out-fluxes cyclically switch between zero and other non-zero values, which makes the analytical study of traffic dynamics very difficult. Here, we are going to simply this problem by averaging the impact of signal control over time and derive the averaged model for the signalized merging junction.

Since the two indicators $(\delta_1(t), \delta_2(t))$ are T-periodic, traffic control at the merging junction can be simplified by averaging them over time T (Sanders et al., 2007). Therefore, we have

$$\begin{aligned} \bar{\delta_1} &= \frac{1}{T}\int_0^T \delta_1(s)ds = \pi_1 \\ \bar{\delta_2} &= \frac{1}{T}\int_0^T \delta_2(s)ds = \pi_2 \end{aligned} \tag{81}$$

Then we can derive the following averaged model by replacing $(\delta_1(t), \delta_2(t))$ with (π_1, π_2) in Equation (79):

$$\begin{gathered} q_i = \tilde{F}^i\left(U_i, U_j, U_3, \pi_i, \pi_j\right) = \pi_i \min\{D_i, S_3\} \text{ for } i,j = 1,2 \text{ and } i \neq j \\ q_3 = q_1 + q_2 \end{gathered} \tag{82}$$

And its local forms at $x = 0$ is

$$\begin{aligned} q_i &= \tilde{f}^i\left(U_1(0^-,t), U_2(0^-,t), U_3(0^+,t), \pi_i, \pi_j\right) \\ &= \pi_i \min\{D_i(0^-,t), S_3(0^+,t)\} \quad \text{for } i,j = 1,2 \text{ and } i \neq j \\ q_3 &= q_1 + q_2 \end{aligned} \tag{83}$$

5.2 Invariant average model for the signalized merging junction

For the signalized merging junction in ***Figure 17*** , we assume each link has the same traffic flow fundamental diagram, e.g., $q_a = Q(k_a),\ a = \{1,2,3\}$, which is a concave function and attains its capacity at the critical density $k_{a,c}$. The flow-rate q_a vanishes when the road is empty with $k_a = 0$ or jammed at the jam density $k_a = k_{a,j}$. To describe the traffic dynamics on these links ($x \neq 0$), the LWR model in Equation (71) is used. Initially vehicles are uniformly distributed along each link, and thus, the initial condition in the demand-supply space can be formulated as

$$\begin{aligned} U_1(x, t=0) &= U_1 = (D_1, S_1), \quad x < 0 \\ U_2(x, t=0) &= U_2 = (D_2, S_2), \quad x < 0 \\ U_3(x, t=0) &= U_3 = (D_3, S_3), \quad x > 0 \end{aligned} \tag{84}$$

5.2.1 Riemann problems

According to (Jin et al., 2009), traffic on a link can have three different types of states: initial state, stationary state, and interior state. For the three links connected to the merging junction, the initial states are denoted as U_i, $i \in \{1,2,3\}$. But as time elapses, stationary states, which are denoted as U_1^-, U_2^-, and U_3^+, will gradually dominate the traffic conditions on these links and propagate upstream to Links 1 and 2 and downstream to Link 3. Meanwhile, interior states, which are denoted as $U_1(0^-,t)$, $U_2(0^-,t)$, and $U_3(0^+,t)$ and theoretically only take infinitesimal space, may appear at the merging junction. The locations of these three types of states at the merging junction are provided in ***Figure 17***.

With the existence of these three types of traffic states, we have to solve the following three types of Riemann problems:

- **Type I:** the Riemann problems between the initial states and the stationary states, e.g., U_1 and U_1^-, U_2 and U_2^-, and U_3 and U_3^+.
- **Type II:** the Riemann problems between the stationary states and the interior states, e.g., U_1^- and $U_1(0^-,t)$, U_2^- and $U_2(0^-,t)$, and U_3^+ and $U_3(0^+,t)$.

- **Type III:** the Riemann problem among the interior states, e.g., $U_1(0^-,t)$, $U_2(0^-,t)$ and $U_3(0^+,t)$.

To uniquely solve the Riemann problems, we need two entropy conditions. The first entropy condition is the junction model such as Equation (83) to regulate traffic movements at the junction. The second entropy condition is used to restrict the wave speeds on the upstream and downstream links: (i) the Riemann problems between U_i and U_i^-, $i = \{1,2\}$, should not have positive waves, while the Riemann problem between U_3 and U_3^+ should not have negative waves; (ii) the Riemann problems between U_i^- and $U_i(0^-,t)$, $i = \{1,2\}$, should not have negative waves, while the Riemann problem between U_3^+ and $U_3(0^+,t)$ should not have positive waves. In this case, the admissible conditions for stationary and interior states in (Jin, 2010) are valid in this study.

5.2.2 Solutions

When the demand for an upstream link $i \in \{1,2\}$ is low, the maximum sending flow is D_i. However, as the demand increases, a queue may form on that link, and thus, the maximum sending flow is $\pi_i \min\{C_i, C_3\}$. Therefore, we have the following definition of effective demand.

Definition 11

Effective demand

The effective demand for an upstream link $i \in \{1,2\}$ is defined as

$$\overline{D}_i = min\{D_i, \pi_i C_i, \pi_i C_3\} \quad (85)$$

If both upstream links are congested, vehicles from the upstream will compete for the space in the downstream link 3. Due to the existence of traffic signals, we have the following definition of merging priority.

Definition 12

Merging priority

The merging priority α_i for link i is defined as

$$\alpha_i = \frac{\pi_i}{\pi_i + \pi_j} \quad for\ i,j = \{1,2\}, and\ i \neq j, \tag{86}$$

which is the percentage of its green time to the total green time.

While solving the Riemann problems arising at the merging junction, possible stationary and interior states can be obtained. Then we have the following Lemma.

Lemma 13

With Equation (83) applied at the signalized merging junction, we can have the following stationary and interior states under different traffic conditions.

[1] When $\overline{D}_1 + \overline{D}_2 < S_3$, link 3 is SUC with $U_3^+ = U_3(0^+, t) = (q_3, C_3)$, where $q_3 = q_1 + q_2$. For the upstream link $i \in \{1,2\}$, we have:

(i) If $D_i \leq \pi_i\, min\{C_i, C_3\}$,

link i is UC with $q_i = D_i$, and $U_i^- = (D_i, C_i)$, $U_i(0^-, t) = \left(\frac{D_i}{\pi_i}, C_i\right)$.

(ii) If $D_i > \pi_i\, min\{C_i, C_3\}$,

link i is SOC with $q_i = \pi_i\, min\{C_i, C_3\}$, and $U_i^- = U_i(0^-, t) = (C_i, \pi_i\, min\{C_i, C_3\})$.

[2] When $\sum_i \overline{D}_i \geq S_3$ and $\overline{D}_i \geq \alpha_i S_3$ for $i = 1,2$, link 3 is OC with $q_3 = S_3$. We have $U_3^+ = (C_3, S_3)$, and $U_3(0^+, t) = (C_3, \frac{S_3}{\pi_1 + \pi_2})$. For the upstream link $i \in \{1,2\}$, it is SOC with $q_i = \alpha_i S_3$ and $U_i^- = U_i(0^-, t) = (C_i, q_i)$.

[3] When $\overline{D}_i + \overline{D}_j \geq S_3$ and $\overline{D}_i < \alpha_i S_3$ for $i,j = 1,2$ and $i \neq j$, link 3 is OC with $q_3 = S_3$. We have $U_3^+ = (C_3, S_3)$, and $U_3(0^+, t) = (D_3(0^+, t), S_3(0^+, t))$ with $D_3(0^+, t) \geq S_3$.

For the upstream link i, we have

(i) When $D_i \leq \pi_i\, min\{C_i, C_3\}$, it is UC with $q_i = D_i$. Then $U_i^- = (D_i, C_i)$, and $U_i(0^-, t) = (D_i(0^-, t), S_i(0^-, t))$ with $D_i(0^-, t) \geq \frac{D_i}{\pi_i}$ and $S_i(0^-, t) \geq D_i$.

(ii) *When* $D_i > \pi_i\, min\{C_i, C_3\}$, *it is SOC with* $q_i = \pi_i\, min\{C_i, C_3\}$. *Then* $U_i^- = U_i(0^-, t) = (C_i, \pi_i\, min\{C_i, C_3\})$.

For the upstream link j, *we have*

(i) *When* $D_j = S_3 - \overline{D}_i \le \pi_i\, min\{C_i, C_3\}$, *it is UC with* $q_j = D_j$. *Then* $U_j^- = (D_j, C_j)$, *and* $U_j(0^-, t) = (D_j(0^-, t), S_j(0^-, t))$ *with* $D_j(0^-, t) \ge \frac{D_j}{\pi_j}$ *and* $S_j(0^-, t) \ge D_j$.

(ii) *When* $D_j \ge S_3 - \overline{D}_i$ *and* $D_j > \pi_j\, min\{C_j, C_3\}$, *it is SOC with* $q_j = S_3 - \overline{D}_i$. *Then* $U_j^- = U_j(0^-, t) = (C_j, S_3 - \overline{D}_i)$.

The proof is provided in **Appendix B.** According to **Lemma 13**, we can have the following Theorem.

Theorem 14

Based on the stationary states derived under different traffic conditions in ***Lemma 13****, we can derive the following averaged model:*

$$q_i = \hat{F}^i(U_i, U_j, U_3, \pi_i, \pi_j) = min\{\overline{D}_i, max\{S_3 - \overline{D}_j, \alpha_i S_3\}\}$$
$$q_3 = \sum_{i=1}^{2} q_i \qquad (87)$$

where $i, j = \{1,2\}$, *and* $i \ne j$.

The proof is simple and thus is omitted here.

Corollary 15

The averaged model in Equation (87) is invariant.

Proof: The local form of Equation (87) can be written as

$$\begin{aligned} q_i &= \hat{f}^i(U_i(0^-, t), U_j(0^-, t), U_3(0^+, t), \pi_i, \pi_j) \\ &= \min\{\overline{D}_i(0^-, t), \max\{S_3(0^+, t) - \overline{D}_j(0^-, t), \alpha_i S_3(0^+, t)\}\} \\ q_3 &= \sum_{i=1}^{2} q_i \end{aligned} \qquad (88)$$

where $i, j = 1,2$ and $i \ne j$. Similar to **Lemma 13** and ***Theorem* 14**, we can derive the same averaged model as in Equation (87). According to **Definition 2**, the averaged model in Equation (87) is invariant.■

Corollary 16

The global averaged model in Equation (82) is non-invariant since a different global form in in Equation (87) is derived from its local form in Equation (83).

The proof is simple and thus is omitted here.

Corollary 17

The invariant averaged model for the signalized linear junction (Equation (72)) is a special case of the one for the signalized merging junction (in Equation (87)) with empty demand in one of the upstream links.

The proof is simple and thus is omitted here.

PART IV: Conclusions and future research directions

In the first part of this report, we provided a comprehensive review on the traffic flow models and signal control strategies for urban traffic networks. We reviewed several different traffic flow models, including the CTM, the store-and-forward model, the LTM, the LQM, and the VCM. We decided to pick the CTM as the simulation tool in our study since as a discrete version of the LWR model, a number of studies on both freeway and urban networks have shown that the CTM can replicated the real-world traffic dynamics such as the formation, propagation, and dissipation of queues arising at the network junctions. We provided a review on traditional signal control strategies which mostly relied on the formulation of Webster's delay or the bandwidth concepts. But we focused more on the signal control strategies on the CTM with a summary on the following aspects: the network representations, the junction models, the objective functions, and the optimization methods.

In reality, in order to handle multiple conflicting traffic movements, the network topology at signalized intersections can be very complicated and vary a lot. Therefore, the cell presentations and junction models in the CTM vary a lot in existing studies. Some consider simple road networks, and thus the regular one-dimension cell presentation and simple network junction models are enough. Others consider more complicated traffic behavior such as land blockages caused by queue spillbacks or by the conflicting left-turn and through movements. In this case, the concepts of sub-zones and sub-cells are introduced to model more detailed traffic movements at the signalized intersection.

For the application of signal control with the CTM, the objective function also varies in existing studies. Most of the studies are focused on the delay minimization such as the network delay, excess delay, or a combination of delay and early arrival flow, while others consider to minimize the performance index or to maximize the network throughput. The optimization problem in the CTM in signalized networks normally can be formulated as a MILP problem. For small networks, commercial application software such as CPLEX can be used to solve such a problem. However, as the network becomes larger and more complicated, it becomes more difficult to solve the MILP problem. Therefore, most of the existing studies

use heuristic methods to solve it. The genetic algorithm is the most popular heuristic method in existing studies to find a good, rather than an optimal set of signal settings. But there also have some studies using other methods such as the linear relaxation and the hill-climbing method to solve the optimization problem.

According to our review, the major issue in the signal optimization of large-scale networks is the increasing amount of binary variables, which are used to mimic the on-and-off pattern of signal control. Therefore, in the second part of this report, we provided a systematic and comprehensive study on deriving and analyzing invariant averaged models for signalized junctions. Particularly, as the starting point, we used the signalized linear junction as an example. We first introduced three models with binary signals and derived their averaged counterparts by replacing the cyclic, binary signal control with an average ratio, which is the effective green ratio. However, simulations in the CTM demonstrated that these time-average models return different average junction fluxes under different road geometries and traffic conditions. Therefore, we derived their invariant models by applying their local forms as entropy conditions at the junction and solving the arising Riemann problems. Using the derived invariant averaged models, we were able to identify their correctness under the constraint of maximum average junction flux. We showed that only one of the three invariant forms is correct since the other two fail to capture the upstream or the downstream capacity constraint. That also showed that invariance does not necessarily guarantee correctness. We also found that different non-invariant models can lead to the same invariant form. Furthermore, we ran CTM simulations on a signalized ring road under different settings of cycle lengths, traffic conditions, and fundamental diagrams. Results showed that the approximation accuracy is not sensitive to the types of fundamental diagrams, and the invariant averaged model provides a reasonable proxy to the macroscopic fundamental diagram. However, we found that long cycle lengths degrade the approximation accuracy.

As a further extension, we applied the proposed analytical framework to more complicated signalized junctions, e.g., a signalized merging junction. We followed the same procedures as those in the signalized linear junction, and proposed one model with binary signals applied to both the upstream demands and downstream supplies. Then we derived its averaged model by replacing the cyclic signal

control with a constant value, which is the effective green ratio of each upstream link. Different from the linear junction, merging behaviors should be considered at the merging junction, and therefore, we proposed two new definitions, Effective Demand and Merging Priority. Effective Demand takes into account not only the upstream demand, but also the reduced maximum flow-rate caused by the signal control at the junction. For the Merging Priority, it is computed as the ratio between the effective green times of the two upstream links. With these two new definitions, the derivation process is significantly simplified. By introducing the local form of the averaged model as the entropy condition at the merging junction and solving the arising Riemann problems, we were able to derive the invariant form of the averaged model. It is easy to show that the derived invariant averaged model for the signalized linear junction in Equation (*72*) is a special case of the one in Equation (*87*) for the signalized merging junction with empty demand in one of the upstream links.

In the future, we can continue our study in the following directions. First, regarding to the application of the CTM to more general signalized intersections, currently there is a lack of a general rule in discretizing the road links into cells and sub-cells. Also, there is a lack of a general guideline for the use of network junction models with signals.

Second, regarding to the development of invariant averaged models, the current analytical framework can be applied to more complicated intersections, e.g., four-way intersections with different turning movements. In this case, both merging and diverging behaviors should be properly captured in the invariant averaged models. Also, not only the green times, but also the signal phase sequence should be considered. Another important issue is to analyze the impact of signal settings and initial conditions on the approximation accuracy of the averaged model.

Third, after the derivation of invariant averaged models, we can form a new framework by introducing them into prevailing traffic flow models, e.g., the CTM, the LTM, and even the LQM. Numerically, we can run simulations to study the dynamic properties of large-scale urban networks. Analytically, we also can try to derive the network stationary states under various traffic states and signal settings. Based on these insights, we should be able to develop optimal signal control strategies to

dynamically change the signal settings at intersections so as to improve the overall network performance. At a more aggregated level, with the application of the averaged model, both freeway and urban networks can be modeled as a whole, which enables the potential of integrated traffic management schemes to reduce traffic congestion. At the planning level, this new framework can be served as the based network traffic flow model to overwrite the traditional link performance functions used in the step of traffic assignment. It is expected to have more realistic results and higher computation speeds with this new framework.

PART V: Appendices

Appendix A

***Proof of* Lemma 4**

We denote $q(U_1^-)$ and $q(U_2^+)$ as the fluxes for the Riemann problem of Type I, $q_1(0^-,t)$ and $q_2(0^+,t)$ as the fluxes for the Riemann problem of Type II, and q as the flux for the Riemann problem of Type III. According to the traffic conservation, we have the following equation

$$q = q_1(0^-,t) = q_2(0^+,t) = q(U_1^-) = q(U_2^+) \tag{89}$$

To determine the stationary and interior states arising at the signalized linear junction, we use the admissible conditions in Section 4 in (Jin et al, 2009). With the entropy condition $q = \eta \min\{D_1(0^-,t), S_2(0^+,t)\}$ applied at the signalized junction, we can have the following combinations of stationary and interior states:

[1] When $D_1 < S_2 \leq C_2$, we have $U_2^+ = U_2(0^+,t) = (D_2^+, C_2)$ since the downstream link is SUC, which leads to $q = \eta \min\{D_1(0^-,t), C_2\}$.

(a) If the upstream link is SOC, i.e.,$U_1^- = U_1(0^-,t) = (C_1, S_1^-)$, we can have $q = S_1^- = D_2^+ = \eta \min\{C_1, C_2\}$, and $D_1 > \eta \min\{C_1, C_2\}$.

(b) If the upstream link is UC, i.e., $D_1 \leq S_1^- = C_1$, we have $U_1^- = (D_1, C_1)$ and $U_1(0^-,t) = (D_1(0^-,t), S_1(0^-,t))$ with $S_1(0^-,t) \geq D_1^- = D_1$. Therefore, we have $q = D_2^+ = D_1 = \eta \min\{D_1(0^-,t), C_2\}$, which leads to $D_1(0^-,t) = \frac{D_1}{\eta}$ and $S_1(0^-,t) = C_1$. Since $D_1(0^-,t) \leq C_1$, we have $D_1 \leq \eta \min\{C_1, C_2\}$.

[2] When $S_2 < D_1 \leq C_1$, the upstream link is SOC. Thus we have $U_1^- = U_1(0^-,t) = (C_1, S_1^-)$, which leads to $q = \eta \min\{C_1, S_2(0^+,t)\}$.

(a) If the downstream link is SUC, i.e., $U_2^+ = U_2(0^+,t) = (D_2^+, C_2)$, we have $q = D_2^+ = S_1^- = \eta \min\{C_1, C_2\}$. In this case, $S_2 > \eta \min\{C_1, C_2\}$.

(b) If the downstream link is OC, $U_2^+ = (C_2, S_2)$ and $U_2(0^+, t) = (D_2(0^+, t), S_2(0^+, t))$ with $D_2(0^+, t) \geq S_2$. Thus we have $q = S_2 = S_1^- = \eta \min\{C_1, S_2(0^+, t)\}$, which leads to $S_2(0^+, t) = \frac{S_2}{\eta}$ and $D_2(0^+, t) = C_2$. Since $S_2(0^+, t) \leq C_2$, we can have $S_2 \leq \eta \min\{C_1, C_2\}$.

[3] When $D_1 = S_2$, we have $U_2^+ = U_2(0^+, t) = (D_2^+, C_2)$ if the downstream link is SUC, which leads to $q = \eta \min\{D_1(0^-, t), C_2\}$ and $S_2 > D_2^+$.

(a) If the upstream link is SOC, i.e., $U_1^- = U_1(0^-, t) = (C_1, S_1^-)$, then $q = S_1^- = D_2^+ = \eta \min\{C_1, C_2\}$. In this case, we have $D_1 = S_2 > \eta \min\{C_1, C_2\}$.

(b) If the upstream link is UC, i.e., $D_1 \leq S_1^-$, we have $U_1^- = (D_1, C_1)$, which leads to $U_1(0^-, t) = (D_1(0^-, t), S_1(0^-, t))$ with $S_1(0^-, t) \geq D_1^-$. Since $q(U_1^-) = D_1$, we have $q = q(U_1^-) = D_1$, which leads to $D_2^+ = D_1$. But it is impossible since $S_2 > D_2^+ = D_1$, which contradicts $D_1 = S_2$.

If the downstream link is OC, $U_2^+ = (C_2, S_2)$, and $U_2(0^+, t) = (D_2(0^+, t), S_2(0^+, t))$ with $D_2(0^+, t) \geq S_2$. Then we have $q = \eta \min\{D_1(0^-, t), S_2(0^+, t)\}$.

(c) If the upstream link is SOC, i.e., $U_1^- = U_1(0^-, t) = (C_1, S_1^-)$, then $q = S_1^- = S_2$, and $D_1 > S_1^-$. But it is impossible since we have $D_1 > S_2$, which contradicts $D_1 = S_2$.

(d) If the upstream link is UC, i.e., $D_1 \leq S_1^-$, we have $U_1^- = (D_1, C_1)$, which leads to $U_1(0^-, t) = (D_1(0^-, t), S_1(0^-, t))$ with $S_1(0^-, t) \geq D_1^-$. Because $q(U_1^-) = D_1$ and $q(U_2^+) = S_2$, we have $q = D_1 = S_2$. If $q = \eta D_1(0^-, t)$, we have $S_1(0^-, t) = C_1$ and $\eta S_2(0^+, t) \geq S_2$. If $q = \eta S_2(0^+, t)$, we have $D_2(0^+, t) = C_2$ and $\eta D_1(0^-, t) \geq D_1$. Since $D_1(0^-, t) \leq C_1$ and $S_2(0^+, t) \leq C_2$, we have $D_1 = S_2 \leq \eta \min\{C_1, C_2\}$.■

Appendix B

Proof of Lemma 13

With the entropy condition in Equation (83) applied at the junction, we can have the following cases:

[1] When the total effective demand is less than the supply, i.e., $\sum_{i=1}^{2} \overline{D}_i < S_3$, the downstream link 3 is SUC since it can accommodate all vehicles from the upstream links. Therefore, we have $U_3^+ = U_3(0^+, t) = (q_3, C_3)$. For the entropy condition, it can be written as

$$\mathrm{q_i} = \pi_\mathrm{i} \min\{\mathrm{D_i}(0^-, \mathrm{t}), \mathrm{C_3}\} \quad \mathrm{i} = 1{,}2.$$

Now, let's consider possible states in the upstream links.

(i) When $D_i \le \pi_i \min\{C_i, C_3\}$ for $i = 1{,}2$, we have $\overline{D}_i = D_i$ and $\sum_{i=1}^{2} D_i < S_3$. In this case, both upstream links are UC with $U_i^- = (D_i, C_i)$, and $U_i(0^-, t) = (\frac{D_i}{\pi_i}, C_i)$. Therefore, $q_i = D_i$, and $q_3 = \sum_{i=1}^{2} D_i$.

(ii) When $D_i > \pi_i \min\{C_i, C_3\}$ for $i = 1{,}2$, we have $\overline{D}_i = \pi_i \min\{C_i, C_3\}$ and $\sum_{i=1}^{2} \pi_i \min\{C_i, C_3\} < S_3$. In this case, both upstream links are SOC with $U_i^- = U_i(0^-, t) = (C_i, \pi_i \min\{C_i, C_3\})$. Therefore, $q_i = \pi_i \min\{C_i, C_3\} < D_i$, and $q_3 = \sum_{i=1}^{2} \pi_i \min\{C_i, C_3\}$.

(iii) When $D_i \le \pi_i \min\{C_i, C_3\}$ and $D_j > \pi_j \min\{C_j, C_3\}$ for $i \ne j$ and $i, j = 1{,}2$, we have $\overline{D}_i = D_i$, $\overline{D}_j = \pi_j \min\{C_j, C_3\}$, and $D_i + \pi_j \min\{C_j, C_3\} < S_3$. In this case, link i is UC with $U_i^- = (D_i, C_i)$ and $U_i(0^-, t) = (\frac{D_i}{\pi_i}, C_i)$, and link j is SOC with $U_j^- = U_j(0^-, t) = (C_j, \pi_j \min\{C_j, C_3\})$. Therefore, $q_i = D_i$, $q_j = \pi_j \min\{C_j, C_3\}$ and $q_3 = D_i + \pi_j \min\{C_j, C_3\}$.

[2] When $\overline{D}_i \ge \alpha_i S_3$ and $\overline{D}_j \ge \alpha_j S_3$, the total upstream effective demand is higher than the downstream supply, i.e., $\sum_{i=1}^{2} \overline{D}_i \ge S_3$.

First, let's show $q_3 = S_3$. If $q_3 < S_3$, link 3 is SUC with $U_3^+ = U_3(0^+, t) = (q_3, C_3)$. Also, at least one of the upstream links (e.g., link i) has $q_i < \overline{D}_i$. Otherwise, we have $q_3 = q_1 + q_2 =$

$\overline{D}_1 + \overline{D}_2 \geq S_3$. Without loss of generality, we assume link i is $q_i < \overline{D}_i$. Then we have $U_i^- = U_i(0^-, t) = (C_i, \pi_i \min\{C_i, C_3\})$. However, it is impossible to have $q_i = \pi_i \min\{C_i, C_3\} < \overline{D}_i = \min\{D_i, \pi_i C_i, \pi_i C_3\}$. Therefore, $q_3 = S_3$.

Second, let's show $q_i = \alpha_i S_3$, for $i = 1,2$. If $q_i < \alpha_i S_3 \leq \overline{D}_i$, we have $q_j > \alpha_j S_3$ since $q_i + q_j = S_3$. For link j, we have

$$\mathrm{q_j} = \pi_\mathrm{j} \min\{\mathrm{D_j}(0^-, \mathrm{t}), \mathrm{S_3}(0^+, \mathrm{t})\} > \alpha_\mathrm{j} \mathrm{S_3},$$

which leads to $S_3(0^+, t) > \frac{S_3}{\pi_i + \pi_j}$. For link i, it is SOC, and $U_i^- = U_i(0^-, t) = (C_i, q_i)$. Because $S_3(0^+, t) > \frac{S_3}{\pi_i + \pi_j}$ and $q_i < \alpha_i S_3 \leq \overline{D}_i$, we have

$$q_i = \pi_i \min\{C_i, S_3(0^+, t)\} = \pi_i C_i < \alpha_i S_3 \leq \overline{D}_i = \min\{D_i, \pi_i C_i, \pi_i C_3\},$$

which is impossible. Therefore, $q_i = \alpha_i S_3$, for $i = 1,2$, and $q_3 = S_3$.

For the stationary and interior states in link i, we have $U_i^- = U_i(0^-, t) = (C_i, \alpha_i S_3)$. For the stationary and interior states in link 3, we have $U_3^+ = (C_3, S_3)$ and $U_3(0^+, t) = (C_3, \frac{S_3}{\pi_i + \pi_j})$.

[3] When $\overline{D}_i + \overline{D}_j \geq S_3$ and $\overline{D}_i < \alpha_i S_3$ for $i \neq j$ and $i, j = 1,2$.

First, let's show that $q_3 = S_3$. If $q_3 < S_3$, link 3 is SUC with $U_3^+ = U_3(0^+, t) = (q_3, C_3)$.

(a) If $q_i = \overline{D}_i$ for $i = 1,2$, we have $q_3 = \sum_{i=1}^{2} \overline{D}_i \geq S_3$, which is impossible since we assume $q_3 < S_3$.

(b) If $q_i < \overline{D}_i$ for $i = 1,2$, link i is SOC with $U_i^- = U_i(0^-, t) = (C_i, q_i)$. For the entropy condition, we have

$$q_i = \pi_i \min\{C_i, C_3\} < \overline{D}_i = \min\{D_i, \pi_i C_i, \pi_i C_3\},$$

which is impossible.

Therefore, $q_3 = S_3$, and link 3 is OC with $U_3^+ = (C_3, S_3)$, and $U_3\ (0^+, t) = (D_3(0^+, t), S_3(0^+, t))$ with $D_3(0^+, t) \geq S_3$.

Second, let's show that $q_i = \overline{D}_i$. If $q_i < \overline{D}_i$, that means link i is SOC with $U_i^- = U_i(0^-, t) = (C_i, q_i)$. For the entropy condition, we have

$$q_i = \pi_i \min\{C_i, S_3(0^+, t)\} < \overline{D}_i = \min\{D_i, \pi_i C_i, \pi_i C_3\} < \alpha_i S_3,$$

which leads to $S_3(0^+, t) < \frac{S_3}{\pi_i + \pi_j}$. For link j, we have

$$q_j = \pi_j \min\{D_j(0^-, t), S_3(0^+, t)\} < \frac{\pi_j S_3}{\pi_i + \pi_j} = \alpha_j S_3.$$

Therefore, we have $q_3 = \sum_{i=1}^{2} q_i < S_3$, which contradicts $q_3 = S_3$. Therefore, we have $q_i = \overline{D}_i$. For link j, we have $q_j = S_3 - \overline{D}_i$.

Next, let's discuss about the traffic states on links i and j with $q_i = \overline{D}_i$ and $q_j = S_3 - \overline{D}_i$.

(i) When $D_i \leq \pi_i \min\{C_i, C_3\}$, link i is UC. Then $U_i^- = (D_i, C_i)$, $U_i(0^-, t) = (D_i(0^-, t), S_i(0^-, t))$ with $D_i(0^-, t) \geq \frac{D_i}{\pi_i}$ and $S_i(0^-, t) \geq D_i$.

(ii) When $D_i > \pi_i \min\{C_i, C_3\}$, link i is SOC. Then $U_i^- = U_i(0^-, t) = (C_i, \pi_i \min\{C_i, C_3\})$.

(iii) When $D_j = S_3 - \overline{D}_i \leq \pi_j \min\{C_j, C_3\}$, link j is UC. Then $U_j^- = (D_j, C_j)$, $U_j(0^-, t) = (D_j(0^-, t), S_j(0^-, t))$ with $D_j(0^-, t) \geq \frac{D_j}{\pi_j}$ and $S_j(0^-, t) \geq D_j$.

(iv) When $D_j > S_3 - \overline{D}_i$ and $D_j > \pi_j \min\{C_j, C_3\}$, link j is SOC. Then $U_j^- = U_j(0^-, t) = (C_j, S_3 - \overline{D}_i)$. ■

PART VI: References

Aboudolas, K., Papageorgiou, M., Kosmatopoulos, E., 2007. Control and optimization methods for traffic signal control in large-scale congested urban road networks. In: American Control Conference, 2007. ACC'07. IEEE, pp. 3132–3138.

Aboudolas, K., Papageorgiou, M., and Kosmatopoulos, E., 2009. Store-and-forward based methods for the signal control problem in large-scale congested urban road networks. Transportation Research Part C: Emerging Technologies, 17(2), 163-174.

Allsop, R. E., 1971a. Delay-minimizing settings for fixed-time traffic signals at a single road junction. IMA Journal of Applied Mathematics 8 (2), 164–185.

Allsop, R.E., 1971b. SIGSET: a computer program for calculating traffic signal settings. Traffic Engineering & Control.

Allsop, R.E., 1972. Estimating the traffic capacity of a signalized road junction. Transportation Research, 6(3):245-255.

Allsop, R.E., 1976. SIGCAP: A computer program for assessing the traffic capacity of signal controlled road junctions. Traffic Engineering & Control, 17(Analytic)

Almasri , Essam and Bernhard Friedrich, 2005. Online offset optimisation in urban networks based on cell transmission model. ITS Hanover.

Courant, R., K. Friedrichs, and H. Lewy, 1928. Über die partiellen Differenzengleichungen der mathematischen Physik. Mathematische Annalen, Vol. 100, No. 1, pp. 32–74.

Daganzo, C. F., 1994**.** The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B: Methodological, Vol. 28, No. 4, pp. 269–287.

Daganzo, C.F., 1995. The cell transmission model, part II: network traffic. Transportation Research Part B: Methodological, Vol. 29, No. 2, pp. 79–93.

D'ans, G., Gazis, D., 1976. Optimal control of oversaturated store-and-forward networks. Transportation Science 10 (1), 1–19.

De la Breteque, L., Jezequel, R., 1979. Adaptive control at an isolated intersection-a comparative study of some algorithms. Traffic Engineering & Control 20 (7).

Feldman, Olga, and Mike Maher, September, 2002. The Application of the Cell Transmission Model to the Optimization of Signals on Signalized Roundabouts. PROCEEDINGS OF THE AET EUROPEAN TRANSPORT CONFERENCE, HOMERTON COLLEGE, CAMBRIDGE, UK-CD ROM.

Gartner, N. H., 1983. OPAC: A demand-responsive strategy for traffic signal control. Transportation Research Record, (906).

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 25 (1), 55–74.

Gao, Yunfeng, Yue Liu, Hua Hu, and Ying-en Ge, 2015. An Enhanced Cell Transmission Model for Traffic Operations within a Signalized Intersection. In Transportation Research Board 94th Annual Meeting, no. 15-2695.

Gazis, D. C., 1964. Optimum control of a system of oversaturated intersections. Operations Research 12 (6), 815–831.

Gazis, D. C., and R. B. Potts, 1963. The oversaturated intersection. Technical report.

Geroliminis, N., Daganzo, C., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological 42 (9), 759–770.

Godunov, S., 1959. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. Matematicheskii Sbornik, Vol. 89, No. 3, pp. 271–306.

Greenshields, B. D., 1935. A study of highway capacity. Proceedings of highway research record, 14, 448–477.

Haberman, R., 1977. Mathematical models. SIAM.

Han, K., Gayah, V. V., Piccoli, B., Friesz, T. L., Yao, T., 2014. On the continuum approximation of the on-and-off signal control on dynamic traffic networks. Transportation Research Part B: Methodological 61, 73–97.

He, Qing, Wei-Hua Lin, Hongchao Liu, and K. Larry Head, 2010. Heuristic algorithms to solve 0–1 mixed integer LP formulations for traffic signal control problems. In Service Operations and Logistics and Informatics (SOLI), 2010 IEEE International Conference. pp. 118-124.

Hunt, P., D. Robertson, R. Bretherton, and M. Royle, 1982. The SCOOT on-line traffic signal optimisation technique. Traffic Engineering & Control, 23(4).

Hunt, P., D. Robertson, R. Bretherton, and R. Winton, 1981. SCOOT-a traffic responsive method of coordinating signals. Technical report.

Improta, G., and Cantarella, G., 1984. Control system design for an individual signalized junction. Transportation Research Part B: Methodological 18 (2147–167.

Jin, W.-L., 2010. Continuous kinematic wave models of merging traffic flow. Transportation research part B: methodological, 44(8):1084-1103.

Jin, W.-L., 2012a. A kinematic wave theory of multi-commodity network traffic ow. Transportation Research Part B: Methodological, 46(8):1000-1022.

Jin, W.-L., 2012b. A link queue model of network traffic flow. arXiv preprint arXiv:1209.2361.

Jin, W.-L., 2012c. The traffic statics problem in a road network. Transportation Research Part B 46 (10), 1360–1373.

Jin, W.-L., 2014a. Analysis of kinematic waves arising in diverging traffic flow models. Transportation Science, 49(1), pp.28-45.

Jin, W.-L., 2014b. Continuous formulations and analytical properties of the link transmission model. Transportation Research Part B: Methodological, 74, pp.88-103.

Jin, W.-L., L. Chen, and E. G. Puckett, 2009. Supply-demand diagrams and a new framework for analyzing the inhomogeneous Lighthill-Whitham-Richards model. In Transportation and Traffic Theory 2009: Golden Jubilee, pages 603-635. Springer.

Jin, W.-L., Gan, Q.-J., Gayah, V. V., 2013. A kinematic wave approach to traffic statics and dynamics in a double-ring network. Transportation Research Part B: Methodological 57, 114–131.

Jin, W.-L., Zhang, H. M., 2003. On the distribution schemes for determining flows through a merge. Transportation Research Part B: Methodological 37 (6), 521–540.

Krein, P. T., Bentsman, J., Bass, R. M., Lesieutre, B. L., 1990. On the use of averaging for the analysis of power electronic systems. IEEE Transactions on Power Electronics 5 (2), 182–190.

Lebacque, J., 1996. The Godunov scheme and what it means for first order traffic flow models. In Internaional symposium on transportation and traffic theory, pp. 647–677.

Li, Zichuan, 2010. Modeling arterial signal optimization with enhanced cell transmission formulations. Journal of Transportation Engineering 137, no. 7: 445-454.

Lighthill, M. and G. Whitham, 1955. On kinematic waves II: A theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, Vol. 229, No. 1178, pp. 317–345.

Little, J.D., 1966. The synchronization of traffic signals by mixed-integer linear programming. Operations Research, 14(4):568-594.

Lo, H. K., 1999. A novel traffic signal control formulation. Transportation Research Part A: Policy and Practice, 33(6), 433-448.

Lo, H. K., 2001. A cell-based traffic control formulation: strategies and benefits of dynamic timing plans. Transportation Science, 35(2), 148-164.

Lo, H. K., Chang, E., & Chan, Y. C., 2001. Dynamic network traffic control. Transportation Research Part A: Policy and Practice, 35(8), 721-744.

Lo, H. K., & Chow, A. H., 2004. Control strategies for oversaturated traffic. Journal of Transportation Engineering, 130(4), 466-478.

Miller, A. J., 1963a. A computer control system for traffic networks.

Miller, A. J., 1963b. Settings for fixed-cycle traffic signals. Journal of the Operational Research Society, 14(4), pp.373-386.

Moskowitz, K., 1965. Discussion of freeway level of service as influenced by volume and capacity characteristics by DR Drew and CJ Keese. Highway Research Record, 99:43-44.

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., Wang, Y., 03. Review of road traffic control strategies. Proceedings of the IEEE 91 (12), 2043–2067.

Richards, P., 1956. Shock waves on the highway. Operations research, pp. 42–51.

Robertson, D. I., 1969. “TRANSYT" method for area traffic control. Traffic Engineering & Control, 11(6).

Sanders, J. J. A., Verhulst, F., Murdock, J. A., 2007. Averaging methods in nonlinear dynamical systems. Vol. 59. Springer.

D. Schrank, B. Eisele, and T. Lomax. TTI 's 2012 urban mobility report. Texas A&M Transportation Institute. The Texas A&M University System, 2012.

P. Sorensen, M. Wachs, E. Y. Min, A. Kofner, and L. Ecola. Moving Los Angeles: Short-term policy options for improving transportation. RAND Corporation, 2008.

Su, Dongyan, Alex Kurzhanskiy, and Roberto Horowitz, 2013. Simulation of Arterial Traffic Using Cell Transmission Model. In 92nd Transportation Research Board Annual Meeting. No. 13-2387.

Webster, F. V., 1958. Traffic signal settings. Technical Report.

Yperman, I., 2007. The link transmission model for dynamic network loading. PhD dissertation.

Zhang, Lihui, Yafeng Yin, and Yingyan Lou, 2010. "Robust signal timing for arterials under day-to-day demand variations." Transportation Research Record: Journal of the Transportation Research Board, 156-166.