## UC Berkeley

UC Berkeley Electronic Theses and Dissertations

## Title

Modeling, Estimation and Control of Traffic
Permalink
https://escholarship.org/uc/item/9570v0wi

## Author

Su, Dongyan
Publication Date
2014
Peer reviewed|Thesis/dissertation

# Modeling, Estimation and Control of Traffic 

by<br>Dongyan Su<br>A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy<br>in<br>Engineering - Mechanical Engineering<br>in the<br>Graduate Division of the<br>University of California, Berkeley<br>Committee in charge:<br>Professor Roberto Horowitz, Chair<br>Professor J. Karl Hedrick<br>Professor Pravin Varaiya

Fall 2014

# Modeling, Estimation and Control of Traffic 

Copyright 2014
by
Dongyan Su

Abstract<br>Modeling, Estimation and Control of Traffic<br>by<br>Dongyan Su<br>Doctor of Philosophy in Engineering - Mechanical Engineering<br>University of California, Berkeley<br>Professor Roberto Horowitz, Chair

This dissertation studies a series of freeway and arterial traffic modeling, estimation and control methodologies.

First, it investigates the Link-Node Cell Transmission Model's (LN-CTM's) ability to model arterial traffic. The LN-CTM is a modification of the cell transmission model developed by Daganzo. The investigation utilizes traffic data collected on an arterial segment in Los Angeles, California, and a link-node cell transmission model, with some adaptations to the arterial traffic, is constructed for the studied location. The simulated flow and the simulation travel time were compared with field measurements to evaluate the modeling accuracy.

Second, an algorithm for estimating turning proportions is proposed in this dissertation. The knowledge about turning proportions at street intersections is a frequent input for traffic models, but it is often difficult to measure directly. Compared with previous estimation methods used to solve this problem, the proposed method can be used with only half the detectors employed in the conventional complete detector configuration. The proposed method formulates the estimation problem as a constrained least squares problem, and a recursive solving procedure is given. A simulation study was carried out to demonstrate the accuracy and efficiency of the proposed algorithm.

In addition to addressing arterial traffic modeling and estimation problems, this dissertation also studies a freeway traffic control strategy and a freeway and arterial coordinated control strategy. It presents a coordinated control strategy of variable speed limits (VSL) and ramp metering to address freeway congestion caused by weaving effects. In this strategy, variable speed limits are designed to maximize the bottleneck flow, and ramp metering is designed to minimize travel time in a model predictive control frame work. A microscopic simulation based on the I-80 at Emeryville, California was built to evaluate the strategy, and the results showed that the traffic performance was significantly improved.

Following the freeway control study, this dissertation discusses the coordinated control of freeways and arterials. In current practice, traffic controls on freeways and on arterials are independent. In order to coordinate these two systems for better performance, a control strategy covering the freeway ramp metering and the signal control at the adjacent intersection is developed. This
control strategy uses upstream ALINEA, which is a well-known control algorithm, for ramp metering to locally maximize freeway throughput. For the intersection signal control, the proposed control strategy distributes green splits by taking into account both the available on-ramp space and the demands of all intersection movements. A microscopic simulation of traffic in an arterial intersection with flow discharge to a freeway on-ramp, which is calibrated using the data collected at San Jose, California, is created to evaluate the performance of the proposed control strategy. The results showed that the proposed strategy can reduce intersection delay by $8 \%$, compared to the current field-implemented control strategy.

Transportation mobility can be improved not only by traffic management strategies, but also through the deployment of advanced vehicle technologies. This dissertation also investigates the impact of Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC) on highway capacity. A freeway microscopic traffic simulation model is constructed to evaluate how the freeway lane flow capacity change under different penetration rates of vehicles equipped with either ACC or CACC system. This simulation model is based on a calibrated driver behavioral model and the vehicle dynamics of the ACC and CACC systems. The model also utilizes data collected from a real experiment in which drivers' selections of time gaps are recorded. The simulation shows that highway capacity can be significantly increased when the CACC vehicles reach a moderate to high market penetration, as compared to both regular manually driven vehicles and vehicles equipped with only ACC.

To My Family,
for their Support and Understanding

## Contents

Contents ..... ii
List of Figures ..... v
List of Tables ..... viii
Acknowledgements ..... $\mathbf{x}$
1 Introduction ..... 1
2 Related Works ..... 5
2.1 DATA COLLECTION ..... 5
Detection Technologies ..... 6
The Caltrans Performance Measurement System ..... 7
Next Generation Simulation ..... 8
2.2 TRAFFIC MODELS ..... 8
Macroscopic Models ..... 9
Microscopic Models ..... 15
3 Evaluation of An Arterial Link-Node Cell Transmission Model Using Traffic Data ..... 23
3.1 INTRODUCTION ..... 23
Link-Node Cell Transmission Model ..... 24
Related Works ..... 29
3.2 APPLYING THE LN-CTM ON ARTERIAL NETWORK ..... 30
Representing Arterial Networks with Signalized Intersections Using the LN-CTM ..... 30
Representing the Studied Site in the LN-CTM ..... 32
3.3 DESCRIPTION OF THE NGSIM LANKERSHIM STUDY SITE ..... 35
Data Set ..... 35
Demand Distribution ..... 36
Traffic at Each Intersection ..... 37
3.4 INPUTS TO THE SIMULATION MODEL ..... 39
Fundamental Diagrams ..... 39
Demands, Split Ratios and Initial Density Estimates ..... 40
Signal Control ..... 41
3.5 SIMULATION RESULTS AND ANALYSIS ..... 42
3.6 LESSONS LEARNED ..... 48
3.7 SUMMARY ..... 50
4 Estimation of Turning Proportions from Exit Counts in Arterial Traffic ..... 51
4.1 INTRODUCTION ..... 51
4.2 PROBLEM FORMULATION ..... 56
4.3 STEPS FOR SOLVING THE LEAST SQUARES PROBLEM ..... 60
4.4 INCORPORATING THE FORGETTING FACTOR AND COVARIANCE RESET- TING ..... 61
4.5 SIMULATION AND RESULTS ..... 61
Scenario I ..... 61
Scenario II ..... 66
Remark ..... 72
4.6 CONCLUSION ..... 72
5 Variable Speed Limits and Coordinated Ramp Metering ..... 73
5.1 INTRODUCTION ..... 73
Ramp Metering and Variable Speed Limits ..... 73
Microscopic Simulation ..... 78
5.2 I-80W Studied Site and Control Design ..... 79
I-80W PM Peak Traffic ..... 80
Coordinated VSL and Ramp Metering Control Design ..... 81
5.3 Simulation Construction ..... 88
Calibration of the Car-following Model ..... 88
Simulation Set Up and Control Scenarios. ..... 91
5.4 SIMULATION RESULTS ..... 96
5.5 CONCLUDING REMARKS AND FUTURE WORK ..... 101
6 Coordinated Freeway Ramp Metering and Street Intersection Signal Control ..... 103
6.1 INTRODUCTION ..... 104
Signal Control on Street Intersections ..... 104
Coordination of Freeway Ramp Metering and Intersection Signal Control ..... 113
6.2 DESCRIPTION OF STUDIED SITE AND MODEL CALIBRATION ..... 114
Studied Site ..... 114
Model Calibration ..... 116
6.3 CONTROL DESIGN ..... 120
On-ramp Metering ..... 121
Intersection Signal Optimization ..... 123
Controller Implementation in the Simulation Runs ..... 125
6.4 SIMULATION RESULTS ..... 127
6.5 LESSONS LEARNED ..... 131
6.6 CONCLUSION ..... 132
7 Evaluation of Impact of Adaptive Cruise Control and Cooperative Adaptive Cruise Control on Highway Capacity ..... 133
7.1 INTRODUCTION. ..... 133
ACC and CACC Systems ..... 133
Related Work ..... 135
7.2 VEHICLE TYPES TO BE SIMULATED ..... 136
7.3 THE MANUAL DRIVING MODEL ..... 137
7.4 CONTROL ALGORITHMS FOR ACC AND CACC VEHICLES ..... 138
7.5 SIMULATION CONSTRUCTION ..... 141
Simulation Network and Settings ..... 141
Vehicle Generation ..... 142
Driver Model Implementation ..... 144
7.6 SIMULATION RESULTS ..... 145
7.7 CONCLUSION ..... 157
8 Conclusion and Future work ..... 159
Bibliography ..... 163

## List of Figures

1.1 Speed on Bay Area Freeways, Taken from PeMS ..... 2
2.1 Different Shapes of Fundamental Diagrams ..... 10
2.2 Allowable Topologies in the Cell Transmission Model ..... 12
2.3 Demand and Supply in a Triangular Fundamental Diagram ..... 13
2.4 Aimsun Graphical User Interface ..... 17
2.5 Update Process of Aimsun Micro Simulator, Taken from Aimsun's Manual ..... 19
2.6 The Role of Aimsun API in Simulation, Taken from Aimsun's Manual ..... 20
2.7 Interaction between Aimsun Micro Simulator and Aimsun API, Taken from Aimsun's Manual ..... 21
3.1 Representation of Roads in the Link-Node Cell Transmission Model ..... 25
3.2 Topologies Used in the Link-Node Cell Transmission Model ..... 27
3.3 Representation of an Arterial Road Segment with Signalized Intersections Using the LN-CTM ..... 30
3.4 Map of Studied Site, Taken from Lankershim Data Analysis Report of the NGSIM Program ..... 33
3.5 Representing the Lankershim Network in the LN-CTM ..... 34
3.6 Speed Histograms on the NGSIM Lankershim Network Based on the Trajectory Data ..... 40
3.7 Comparison between the Simulated and Actual Vehicle Flows, for All Four Signalized Intersections along the Lankershim Road in the Northbound Direction during the 8:45-9:15 am Time Period46
3.8 Comparison between the Simulated and Actual Vehicle Flows, for All Four Signalized Intersections along the Lankershim Road in the Southbound Direction during the 8:45 -9:15 am Time Period ..... 47
3.9 Comparison between the Simulated and Actual Travel Times, for the Lankershim Net- work in both the Northbound and Southbound Directions during the 8:45-9:15 amTime Period48
4.1 Intersection with Entrance and Departure Detectors ..... 53
4.2 Intersection with Departure Detectors ..... 56
4.3 The RMSD Values of the RCLS and CLSAS Algorithms in Experiments \#1 to \#4 in Scenario I as Representatives ..... 65
4.4 Plots of Estimated Turning Proportions in Experiments \#1 to \#4 in Scenario II ..... 70
5.1 Traffic Light and Detectors Commonly Used in the Traffic Responsive Ramp Metering ..... 74
5.2 Displaying Variable Speed Limits Using Variable Message Signs ..... 77
5.3 The Studied Freeway Segment of I-80 Westbound Direction ..... 81
5.4 Network Representation of the Studied Freeway Segment in I-80 Westbound Direction ..... 82
5.5 The Road Sections in the VSL Control Design ..... 84
5.6 Traffic Speed Drop Probability Contour ..... 87
5.7 RMSPe Distribution resulting from the Model Calibration ..... 90
5.8 Comparison of the Speed and Trajectory of a Vehicle between Calibration Results and Actual Measurements ..... 91
5.9 Demand Profiles at all Network Entrances Used in the Freeway Simulation of the pro- posed Coordinated VSL and Ramp Metering Control Strategy ..... 92
5.10 The Structure of the Control Algorithm Implementation ..... 95
5.11 Comparison of the Simulated Accumulative Delay Using Different Control strategies in the I-80W Network: no control, all time VSL, all time combined VSL and ramp me-tering, switched VSL, switched combined VSL and ramp metering, switched metering,and all time VSL with 30\% compliance97
5.12 Comparison of the Simulated Accumulative Times Spent in Different Control Strate- gies of the I-80W Network: No Control, All Time VSL, All Time Combined VSL and Ramp Metering, Switched VSL, Switched Combined VSL and Ramp Metering,Switched Metering, and All Time VSL with 30\% Compliance98
5.13 Comparison of the Simulated Accumulative Travel Distances in Different Control Strategies of the I-80W Network: No Control, All Time VSL, All Time Combined VSL and Ramp Metering, Switched VSL, Switched Combined VSL and Ramp Meter- ing, Switched Metering, and All Time VSL with 30\% Compliance ..... 99
5.14 Comparison of the Simulated Vehicle Speed at the Diverge-point in the No Control and All Time VSL Strategies ..... 100
6.1 The Illustration of Phase Layout in Signal Control ..... 107
6.2 An Illustration of the Signal Phases and Stages Used in Signal Control ..... 108
6.3 The Dual-ring Structure Used in Signal Control ..... 108
6.4 The Illustration of Actuated Control ..... 112
6.5 Map of the Intersection at SR-87 and W. Taylor Street ..... 115
6.6 The Frame of the Implementation of the Coordinated Control Strategy Proposed for the Intersection of SR-87 and W. Taylor St ..... 126
7.1 The Structure of the Implementation of the Vehicle Generation Algorithm in the ACC/CACC Simulation ..... 143
7.2 The Frame of the Implementation of the Driving Models in the ACC/CACC Simulation ..... 145
7.3 Highway Lane Capacity as a Function of Changes in ACC Market Penetration Relativeto Manually Driven Vehicles, unit: veh/hr146
7.4 Highway Lane Capacity as a Function of Changes in CACC Market Penetration Rela-tive to Manually Driven Vehicles, unit: veh/hr . . . . . . . . . . . . . . . . . . . . . . 148
7.5 Highway Lane Capacity as a Function of Changes in CACC Market Penetration Rela- tive to "Here I Am" Vehicles, unit: veh/hr ..... 150
7.6 Prediction of Lane Capacity Effects of ACC and CACC Vehicles (With the RemainingVehicles Manually Driven, unit: veh/hr)152
7.7 Prediction of Lane Capacity Effects of HIA and CACC Vehicles (With the Remaining Vehicles Manually Driven, unit: veh/hr) ..... 154
7.8 Prediction of Lane Capacity Effects of HIA and CACC Vehicles (With the Remaining Vehicles Being ACC, unit: veh/hr) ..... 156

## List of Tables

3.1 The Volumes of Tracked Vehicles in the Studied Network during the Data Collecting Period (unit: number of vehicles) ..... 37
4.1 The Turning Proportions Used in the Simulation of Estimation Algorithm of Scenario I ..... 62
4.2 Performance Comparison between the CLSAS and RCLS Algorithms ..... 64
4.3 Estimated Turning Proportions Produced by the CLSAS Algorithm in Scenario I and the Error Percentages ..... 66
4.4 Estimated Turning Proportions Produced by the RCLS Algorithm in Scenario I and the Error Percentages ..... 66
4.5 Actual Turning Proportions Used in the Simulation of Estimation Algorithms in the ..... 67
4.6 Performance Comparison of the CLSAS, RCLS and RCLSFR Algorithms in Simula- tion Scenario II ..... 69
4.7 Estimated Turning Proportions Produced by the CLSAS Algorithm in Scenario II and the Error Percentages ..... 71
4.8 Estimated Turning Proportions Produced by the RCLS Algorithm in Scenario II and the Error Percentages ..... 71
4.9 Estimated Turning Proportions Produced by the RCLSFR Algorithm in Scenario II and the Error Percentages ..... 71
5.1 Estimated Driver Behavior Model Parameter Values ..... 90
5.2 Automobile Traffic Volumes between Each Origin and Destination during the First Hour ..... 93
5.3 Performance Comparison of Different Control Strategies in the Freeway Coordination Control Simulation: No Control, All Time VSL, All Time Combined VSL and Ramp Metering, Switched VSL, Switched Combined VSL and Ramp Metering, Switched Metering, and All Time VSL with 30\% Compliance ..... 101
6.1 Vehicle Volumes at the SR-87 \& Taylor St intersection during 4:15-7:00pm, Sep 05, 2012, unit: veh ..... 116
6.2 Calibration Results of the Intersection Flows ..... 119
6.3 Calibration Results of the Freeway Flow and Speed ..... 120
6.4 Comparison of the Traffic Performance between the Current Control Plans and the Proposed Control Strategy ..... 129
6.5 Comparison of the Queue Lengths between the Current Control Plans and the ProposedControl Strategy130
7.1 Highway Lane Capacity as a Function of Changes in ACC Market Penetration Relative to Manually Driven Vehicles, unit: veh/hr ..... 147
7.2 Highway Lane Capacity as a Function of Changes in CACC Market Penetration Rela-tive to Manually Driven Vehicles, unit: veh/hr149
7.3 Highway Lane Capacity as a Function of Changes in CACC Market Penetration Rela- tive to "Here I Am" Vehicles, unit: veh/hr ..... 151
7.4 Prediction of Lane Capacity Effects of ACC and CACC Vehicles (With the RemainingVehicles Manually Driven, unit: veh/hr)153
7.5 Prediction of Lane Capacity Effects of HIA and CACC Vehicles (With the Remaining Vehicles Manually Driven, unit: veh/hr) ..... 155
7.6 Prediction of Lane Capacity Effects of HIA and CACC Vehicles (With the Remaining Vehicles Being ACC, unit: veh/hr) ..... 157

## Acknowledgments

Graduate study at UC Berkeley has been a valuable experience. I sincerely thank my advisor, teachers, colleagues, and family for their support over the past six years.

Professor Roberto Horowitz has taught me a lot in various ways. He is enthusiastic, dedicated, and knowledgeable. He guided my study and research with rich experience, patience, and great support. Despite his many other responsibilities, he always conducts research with interest, and provides helpful recommendations and inspiring ideas to my study and research. I was impressed by his rigorous attitude toward research.

I am grateful to have had the chance to learn from and work with Professor Pravin Varaiya. Professor Varaiya has a deep understanding of transportation theory and many creative ideas. He has given me great guidance in conducting my research.

I would like to thank Professor J. Karl Hedrick, who was my teacher in ME237, and on the committee for both my dissertation and qualifying exam. I am also indebted to Dr. Xiao-Yun Lu, who has supported me and provided instruction for my research. I greatly appreciate his efforts. I also want to express my thanks to all of the professors in control and transportation. I have learned from them through lectures, seminars, and discussions.

I thank all my colleagues in the TOPL and the Connected Corridor groups. Over the past few years, they have been helped me a lot, particularly Gabriel Gomes, Alex Kurzhanskiy, Ajith Muralidharan, Rene Sanchez, and Gunes Dervisoglu, who mentored me in both my study and research. It has been a wonderful experience to work with them in the group.

Finally, I am grateful to my parents for their support and understanding. They love me and have encouraged me to pursue my dreams. Without them, I would never have come so far.

## Chapter 1

## Introduction

Most of the metropolitan areas in the US are experiencing traffic congestion during commuting hours. According to the Annual Urban Mobility Report produced by the Texas Transportation Institute [1], the nation-wide average delay of commuters rose from 15.5 hours per commuter per year in 1982 to 38.0 hours in 2011. In particular, commuters in 15 very large urban areas, which include the San Francisco-Oakland area, the New York-Newark area, and the Los Angeles-Long Beach-Santa Ana area, had a yearly delay of 52 hours per commuter in 2011, translating to 1,128 dollars of congestion cost per commuter on average. Fig. 1.1 shows the speeds on the freeways in the San Francisco Bay Area recorded by PeMS (Performance Measurement System) [2]. The speeds were measured at $5: 30 \mathrm{pm}$ on a typical Wednesday. In the figure, most of the freeways are plotted in yellow and red, representing moderate and heavy congestion, respectively. Traffic congestion not only causes travel delays, but also results in a large amount of fuel consumption and green-house gas emission.

Traffic congestion can be resolved by infrastructure expansions. However, there are some cases where this cannot be done or is unable to relieve congestion. Constructing more roads or widening existing roads requires more land, and expansion costs will probably be high. Moreover, such solutions may not be feasible in metropolitan areas because of lack of space, or may have a harmful effect on the environment. In addition, once roads are expanded and congestion is resolved, population will move from congested areas to less congested ones, and thus traffic demand for the latter will increase, which will in turn cause congestion again.

If road expansion is not able to eliminate traffic congestion, proper traffic operations are necessary to maintain network mobility. Some of the operations change traffic performance by shaping the traffic demand, for instance, through enforcing tolls, designating special lanes like HOV (high occupancy vehicle) and HOT (high occupancy/toll) lanes, or providing travelers with real-time information on traffic conditions. Some of the operations utilize roadside facilities to implement traffic control techniques in order to smooth traffic, increase network throughput and reduce travel


Figure 1.1 Speed on Bay Area Freeways, Taken from PeMS
delay, such as ramp metering and signal control.
Whatever traffic operations are considered, three critical elements are usually needed in order to accomplish an appropriate control design, as follows: a) knowledge of traffic features and dynamics, usually represented by traffic models; b) data to analyze traffic conditions, calibrate traffic models and evaluate the control design; and c) reasonable and implementable traffic control methodologies.

Understanding traffic features and dynamics is of great importance because it serves to identify the cause of traffic problems and to predict the influence of traffic operations. In traffic control design, traffic models, usually described by a set of mathematical formulas, are utilized to evaluate the proposed controls and/or formulate the control problem. There is no generic traffic model that can cover all traffic phenomena. Therefore, control designers have to select traffic models that can precisely represent the traffic conditions of interest and fulfill the design requirements.

Traffic data is necessary to support traffic analysis, operation designs, and real-time control execution. In particular, many control techniques used in the field are traffic-responsive nowadays, and real-time detection is required by those techniques. Traffic data can be obtained through surveys, manual collection, or detection with sensors, depending on the requirement and availability of collection techniques. It could happen that the demanded data is not directly measurable. In this case, feasible and reliable estimation methods are needed to obtain the data.

Control methods have significance in traffic control. A Traffic control strategy will have positive effects only when it has a proper logic. An appropriate control method must be reasonable,
feasible, and practical. Control designers not only need to justify the dynamics of the proposed strategies from theory, but also need to consider practical issues related to the implementation of the proposed techniques, including traffic safety, drivers' acceptance, feasibility of implementation, and accessibility of required facilities.

## Organization of this Dissertation

This dissertation is organized as follows:
In Chapter 2, we describe existing work and programs related to data collection and traffic models that are relevant to the materials discussed in this dissertation. The first part of this chapter, which has to do with data collection, introduces different detection technologies and two data collection programs, namely the Performance Measurement System (PeMS) and the Next Generation Simulation (NGSIM) program. In the second part of this chapter, different traffic models, including the Lighthill-Whitham-Richards Model (LWR), the Cell Transmission Model (CTM) and Aimsun simulation software, are described in detail, as these are germane to the studies presented in later chapters.

In Chapter 3, we describe a simulation study of urban street traffic using the Link-Node Cell Transmission Model (LN-CTM) and evaluate the accuracy of this simulation model. We first introduce the formulation of link-node cell transmission model, which is an extension of the cell transmission model, and review the related work on the cell transmission model and its extensions. Subsequently, we describe how the LN-CTM is adapted to urban street traffic, and give an example using the studied network of this chapter. After this, we describe the data used in the simulation study, the Lankershim data set from the NGSIM program. Following this, we present the inputs to the simulation model and how they were obtained from the data set in detail. This is followed by the simulation results and analysis.

In Chapter 4, we present a method to estimate turning proportions at an arterial signalized intersection based on traffic counts from departure detectors and signal information. We first review prior and related research on turning proportion estimation and describe the motivation for our study. We then describe a proposed algorithm for estimating turning proportions using exit counts. The estimation problem is cast as a least squares problem with boundary constraints. We give the detailed procedure to solve the problem using a recursive least squares method. We also suggest using a forgetting factor and covariance resetting in the least squares method if the turning proportions are time varying. Simulations of two scenarios, one of constant turning proportions while the other of time-varying ones, are utilized to evaluate the accuracy and computational effort of the proposed method.

Chapters 5 to Chapter 7 are related to traffic controls. In Chapter 5, we propose a freeway control strategy of variable speed limits (VSL) and ramp metering to address traffic congestion. Related work of VSL and ramp metering is reviewed in the first section of this chapter. The specific congestion problem discussed in this chapter is the evening congestion on the I-80W east
of the San Francisco Bay Area, California, where traffic congestion is caused by weaving. The studied location, its traffic condition, and the proposed control strategy are described in the second section. Here, the VSL is used to increase the freeway throughput, and ramp metering is implemented using model predictive control to minimize travel time. The proposed coordinated VSL and ramp metering control strategy is evaluated using a calibrated microscopic simulation. The model calibration results and the control strategy simulation results are presented in the third and fourth sections, respectively.

In Chapter 6, we discuss a method of coordinating the ramp metering control at a freeway on-ramp and the signal control at an adjacent arterial intersection. The study site selected in this chapter is an interchange in San Jose, California. Previous research on the coordination of the freeway system and urban street system is reviewed in the first section of this chapter. A microscopic model is constructed and accurately calibrated for this research. We use upstream ALINEA, a well-known local traffic-responsive ramp metering technique, with queue-overwrite for the ramp metering. The intersection signalization control is determined through a linear program, in which the green time for each phase is balanced according to its demand, and the on-ramp storage is considered to prevent queue spillover. We compare the performance of the proposed control with the current field-implemented control using the simulation.

In Chapter 7 , we evaluate the highway capacity change under different market penetration rates of vehicles equipped with Adaptive Cruise Control (ACC) or Cooperative Adaptive Cruise Control (CACC). This evaluation is based on a project that collected drivers' time gap selection when they used ACC or CACC. We perform a microscopic simulation study to estimate the change, in which the dynamics of manual driving, ACC, and CACC vehicles are carefully modeled to reflect the reality.

The last chapter of this dissertation summarizes the work and outlines possible directions for future research.

## Chapter 2

## Related Works

In this chapter, existent work in areas related to this dissertation is discussed. The chapter covers the topics of traffic data collection and traffic modeling. Detailed reviews of the literature related to the other material covered in this dissertation will be presented in the chapters where the specific material is discussed.

### 2.1 DATA COLLECTION

Traffic data is crucial in traffic analysis and control design. The richer and more detailed the data we have is, the better we can understand the traffic problem and evaluate the proposed control strategies. Historical data is frequently used by traffic engineers to plan new infrastructure, predict future traffic patterns, and test traffic models and controls. Real-time traffic data is required to implement traffic-responsive control, discover traffic incidents, and provide information to drivers.

Several terms are used in describing traffic states in this dissertation. These terms and their definitions are presented below. Reference to traffic terms can be found in [3, 4].

Volume Volume (or traffic count) is the number of vehicles passing a certain point in a given time interval.

Flow Flow is the number of vehicles per unit of time at a point and a specific time.
Speed Speed is the distance rate at which vehicles move. It should be noted that speed can be aggregated across space or time. The average taken at a specific location and interval of time is called a time-mean, while the average taken at an instant over a space interval is
called a space-mean. Vehicle speed measured by stationary sensors is the time-mean speed, sometimes called point speed.

Occupancy Occupancy is the fraction of time when an infrastructure, such as a loop detector, is occupied by vehicles.

Density Density is the number of vehicles per unit length at a point at a specific time.
Travel time Travel time is the amount of time taken by a vehicle to complete a trip or traverse a segment of road.

Travel distance Travel distance is the distance traversed in a trip or over a certain period of time.
Delay Delay is the additional travel time experienced by a driver, passenger, or pedestrian due to traffic congestion or traffic controls. It is measured as the time difference between actual travel time and free-flow travel time.

Free-flow Free-flow is a traffic condition under which drivers are able to travel at their desired speeds (or the posted speed limit) without being restricted by surrounding traffic. The opposite of free-flow is congestion.

Congestion Congestion is a traffic condition under which drivers move more slowly than their desired speeds (or the posted speed limit) and are restricted by their downstream (front) vehicles.

## Detection Technologies

Different detection technologies are utilized to determine real-time traffic conditions depending on the considerations of data accuracy, cost, and installation. [5, 6] provide reviews of detection technologies. In general, detection technologies can be classified into two types, namely, intrusive and non-intrusive. The former type includes inductive loop, wireless magnetometer and microloop. The latter type includes video, radar, microwave, sound, and infrared. In the following paragraphs, several technologies related to the data collection in the studies performed in this dissertation are introduced.

Inductive loops Inductive loops have been widely used in traffic state detection since the 1960s. They detect the presence of vehicles on the loops. From this, other traffic information such as traffic volume, occupancy, and speed (if dual loops are available) can be obtained. The accuracy of measurement from inductive loops is high. However, they have a significant drawback because they have to be embedded in the pavement for installation.

Magnetometer Like inductive loops, magnetometers detect the presence of vehicles and have reliable accuracy. Unlike inductive loops, however, a magnetometer is a small in-pavement
sensor that can be easily installed. Additionally, it offers wireless communication to the signal controller, which means that no wire is needed to connect detectors and controllers.

Radar Radar detects moving objects by sending radio waves and receiving waves reflected by objects. It is usually mounted overhead or roadside. It can detect vehicle type, volume, speed, and occupancy.

Video detection Video detection uses overhead cameras to collect images and convert them into traffic data by applying an image processing algorithm. It can detect incidents on the road, vehicle speed, vehicle types and counts. In order to obtain clear images covering the road segment of interest, it requires adequate light and cannot be blocked by other objects. Therefore, it does not perform well in the case of bad weather or vehicle occlusion.

## The Caltrans Performance Measurement System

The Performance Measurement System (PeMS) is a system designed to collect, filter, process, aggregate, and examine traffic data in the state of California [2]. It started in 1999 as a university research project and is currently maintained by the California Department of Transportation (Caltrans) and deployed across most California freeways. It provides real-time and archived traffic data to the public. As of 2013, the data sources in PeMS covered nine Caltrans districts out of 12, specifically, district 3 (North Central), 4 (Bay Area), 5 (Central Coast), 6 (South Central), 7 (Los Angeles and Ventura counties), 8 (San Bernardino and Riverside), 10 (Central), 11 (San Diego and Imperial), and 12 (Orange County). Over 35,000 detectors report traffic data every 30 seconds to PeMS, covering about 30,000 miles of freeways, and more than 12 TB of data is stored. Its sensors include inductive loops, radar, and magnetometers. In addition to stationary detectors, PeMS also collects and archives data sources from the Bay Area FasTrak (tag data), California Highway Patrol (incident data), and San Diego Metropolitan Transit System.

PeMS computes various traffic states and performance measures from raw data collected from the field. Raw data is given in flow and occupancy of 30 seconds and at a lane-by-lane level. The data processing procedure of PeMS can be described as follows. In the first step, 30 -second data from the districts is aggregated into 5 -minute data. In the second step, a number of parameters, including the g -factor, which is used to compute speed when only a single loop is available, and some imputation parameters are computed offline from the 5 -minute data. Data diagnostics is conducted for each detector every day. The 5-minute data might miss samples due to bad detectors or no data reported. In this case, missing data will be filled by some algorithms in the third step. After this, the 5-minute lane-by-lane data is aggregated across lanes and performance measures, such as vehicle hours traveled (VHT) and vehicle miles traveled (VMT), are computed.

In the study of Chapter 6, data from the PeMS is used in the traffic model construction and calibration.

## Next Generation Simulation

The Next Generation Simulation (NGSIM) program [7] is a project supported by United States Department of Transportation (US DOT) Federal Highway Administration (FHWA). The goal of this program is to develop a core of open behavioral algorithms for traffic simulation, mainly microscopic models, and to collect traffic data of high quality for research and evaluation of algorithms. Its products include vehicle trajectory data of high resolution and detailed supporting documents, which are open and free for the simulation community. Because it provides rich and high-quality data, it has been used by researchers, simulation software developers, and users since it was published.

The traffic data collected by the NGSIM program contains typical free-flow and congested traffic. Four data-sets are presented by this program. They were collected at I-80 in Emeryville, CA, US-101 in Los Angeles, CA, Lankershim Boulevard in Los Angeles, CA, and Peachtree Street in Atlanta, GA. The former two locations are freeway segments, and the latter two are surface street segments. The two freeway data-sets are 45 minutes long each and the two arterial data sets are 30 minutes long each. The vehicle trajectory data of NGSIM was collected through videos. Multiple cameras were mounted on buildings to obtain clear views covering the road segments of interest. Video images were taken every one-tenth of a second. Vehicles were automatically detected and tracked from video images using image processing algorithms. Because the trajectory data is of high detail and a broad range of supporting documents are offered, including signal and detector reports, the interactions of drivers and interactions presented to them from road facilities can be analyzed and used in support of the development of behavioral models.

Data from the NGSIM program is used in Chapter 3 and 5 of this dissertation to evaluate traffic models. In addition, one of the driver behavioral models used in Chapter 7 was calibrated using data from NGSIM. A more detailed description of the NGSIM data sets will be presented in the chapters where they are used.

### 2.2 TRAFFIC MODELS

Traffic simulation models can be divided into three categories, namely, microscopic, mesoscopic, and macroscopic [8]. Macroscopic traffic models describe traffic dynamics in terms of density, flow, and average velocity as a function of location and time. Vehicles are viewed as a homogeneous traffic stream and modeled by continuum equations. Interactions among vehicles are considered in terms of their aggregated effect rather than being modeled related to detail. Macroscopic models have advantages in computation efficiency, as well as ease of simulation. Microscopic models, on the other hand, describe traffic dynamics by delineating trajectories of individual vehicles in a network. They use driver behavior models to depict the positions and velocities of individual vehicles at every simulation step. The interaction between vehicles and their neighbors, and the interaction between vehicles and roadway or traffic controls are considered. Hence, traffic
behaviors or traffic controls that involve discrete events are easier to simulate in a microscopic simulation than in a macroscopic simulation, for instance, weaving, actuated control, and bus priority signals. Because microscopic simulation has to compute all individual vehicles with a small time step, it requires more input data, significantly more variables to calibrate, and larger computational effort than macroscopic simulation. Mesoscopic models stand in the middle of macroscopic and microscopic models. They do not trace individual vehicles, but still specify the behavior of individuals, for instance, in probabilistic or statistical terms. In mesoscopic models, traffic is considered by groups of objects, and the activities of groups and interactions among groups are described at a low detail level. In this dissertation, both macroscopic and microscopic models will be used.

## Macroscopic Models

As mentioned above, macroscopic models study traffic in terms of aggregated quantities such as density, flow and speed. Space-mean speed is used in these models, while vehicle interaction is not covered. In modeling freeway traffic, well-known examples of macroscopic models are the Cell Transmission Model [9, 10] and the METANET model [11]. In modeling urban street traffic, queue models [12], platoon dispersion models [13], and store-and-forward models [14] are used in addition to the Cell Transmission Model. Here, we will describe the Cell Transmission Model and its origin, the Lighthill-Whitham-Richards Model. In Chapter 3, an extension of the Cell Transmission Model is studied in detail.

## The Lighthill-Whitham-Richards Model

The Cell Transmission Model originated from the Lighthill-Whitham-Richards (LWR) model [15, [16]. The Lighthill-Whitham-Richards model was proposed in the 1950s, and uses two equations to describe traffic. The first equation, shown by Eq. (2.1), is a partial differential equation in terms of flow and density at a particular time and location. Here $\rho(x, t)$ represents the density at location $x$ and time $t$, and $f(x, t)$ represents flow. Eq. (2.1) is called the vehicle conversion equation, because it simply states that there are no vehicles created or disappearing on a road. The second equation, shown by Eq. (2.2), is a function describing the flow-density relation in steady state. In this equation, $\Phi(\rho)$ represents the curve on which flow and density should fall. This curve is also called the fundamental diagram in traffic engineering. Various shapes of fundamental diagrams have been suggested by researchers [17, 18], including Greenshields, triangular, and trapezoidal. Fig. 2.1 shows the plots for these three shapes of fundamental diagrams.

$$
\begin{gather*}
\frac{\partial \rho(x, t)}{\partial t}+\frac{\partial f(x, t)}{\partial x}=0  \tag{2.1}\\
f(x, t)=\Phi(\rho(x, t)) \tag{2.2}
\end{gather*}
$$



Figure 2.1 Different Shapes of Fundamental Diagrams

Despite the different shapes of fundamental diagrams that might be used, they all share the following characteristics [3, 18]:

1. $\Phi(\rho)$ is a concave function.
2. $\Phi(0)=\Phi\left(\rho_{J a m}\right)=0$, where $\rho_{J a m}$ is the jam density.
3. $\Phi(\rho)$ reaches a maximum at $\rho_{c}$, where $\rho_{c}$ is the critical density, and this maximum value of flow is called capacity, $Q$.

The fundamental diagram is separated into two regimes by the critical density, the free-flow regime ( $\rho \leq \rho_{c}$ ) and the congested regime ( $\rho \geq \rho_{c}$ ). If traffic is in free-flow regime, vehicle speed is dependent on the posted speed limit of a road or the desired speed of a driver, but not restricted by other vehicles. The speed in this regime is called the free-flow speed. On the other hand, the speed of a vehicle is restricted by surrounding vehicles if it is traveling in the congested regime.

## The Cell Transmission Model

The Cell Transmission Model (CTM) was proposed by Daganzo in [9, 10], and it can be viewed as a first order Godunov approximation of the LWR model [19]. In the cell transmission model, a one-way road segment is divided into small sections called cells. Each cell has a uniform length $L$, and traffic flow within each cell is viewed as a homogeneous stream. If the free-flow speed of this road segment is $v_{f}$ and the length of the simulation step is $\Delta T$, the length of a cell has to be $L=v_{f} \Delta T$. This requirement of cell length is put in place to guarantee convergence in numerical computations, or it can be interpreted as the condition that vehicles cannot move more than one
cell during a simulation time step. If the cells are numbered from upstream to downstream in increasing order deriving the evolution of density of cell $i$ from the conservation equation, shown by Eq. (2.3), is straightforward. In this equation, $n_{i}(k)$ is the number of vehicles in cell $i$ at time step $k$, and $y_{i, i+1}(k+1)$ is the volume from cell $i$ to cell $i+1$ between time step $k$ and time step $k+1$. Eq. (2.3) states that the increment of vehicles in cell $i$ equals the net flow into this cell.

$$
\begin{equation*}
n_{i}(k+1)=n_{i}(k)+\left(y_{i-1, i}(k+1)-y_{i, i+1}(k+1)\right) ; \tag{2.3}
\end{equation*}
$$

In the published paper on the cell transmission model [9, 10], the elementary network that is allowed has at most three links. Therefore, three types of network topologies are allowed, namely ordinary links (a node with only one entering link and one leaving link), diverge (a node with one entering link and two leaving links), and merge (a node with two entering links and one leaving link). Fig. 2.2illustrates these three types of topologies. If a network of interest contains topologies other than these three types, it has to be divided into smaller sub-networks that can be represented by allowable types of topologies before applying the cell transmission model.

(a) Ordinary Links

(b) Diverge Node

(c) Merge Node

Figure 2.2 Allowable Topologies in the Cell Transmission Model

In a network of ordinary links without any junctions, as shown in Fig. 2.2(a), volume $y_{i, i+1}(k+$ $1)$ is determined by the sending and receiving functions, also called demand and supply, respectively. When a triangular fundamental diagram is assumed, the sending and receiving functions can be represented by a dash line and solid line, respectively, as in Fig. 2.3. Moreover, $y_{i, i+1}$ can be computed by Eqs. (2.4) to (2.6).


Figure 2.3 Demand and Supply in a Triangular Fundamental Diagram

$$
\begin{align*}
S_{i} & =\min \left\{n_{i}, Q_{i}\right\}  \tag{2.4}\\
R_{i+1} & =\min \left\{Q_{i+1}, \delta\left(n_{i+1}-n_{J a m}\right)\right\}  \tag{2.5}\\
y_{i, i+1} & =\min \left\{S_{i}, R_{i+1}\right\} \tag{2.6}
\end{align*}
$$

In the above equations, $S_{i}(k)$ and $R_{i+1}(k)$ are used to denote the sending function of cell $i$ and receiving function of cell $i+1$ at time step $k$, respectively. $Q_{i}$ denotes the maximum throughput of cell $i$, and $\delta=\frac{v_{f}}{w}$, where $v_{f}$ is the free-flow speed and $w$ is the shockwave speed, as shown in Fig. 2.3. The sending function (demand) of cell $i$ is computed from its present number of vehicles and its capacity. The receiving function (supply) is computed from the number of vehicles and capacity of the downstream cell. The flow from cell $i$ to cell $i+1$ is the smaller of the demand from the upstream cell and supply of the downstream cell.

If a diverging junction exists and a road splits into two, as shown in Fig. 2.2 (b), flows turning to each downstream segment have to be determined. Using $i$ as the index of upstream road segment, and $j_{1}$ and $j_{2}$ as the indices of two downstream segments, according to the cell transmission model update equation, the flows crossing the cells are obtained by Eqs. (2.7) and (2.8). In these equations, $y$ is the total flow leaving cell $i$ and $\beta_{i, j}$ is the turning proportion from cell $i$ to cell $j$. $S$ and $R$ have the same meanings as in the ordinary link network described above. It should be noted
that the min operation between the term $R_{j_{1}} / \beta_{i, j_{1}}$ and $R_{j_{1}} / \beta_{i, j_{1}}$ in Eq. (2.7) enforces the First-In-First-Out (FIFO) law in computation. This law leads to a blockage phenomenon in congested traffic: If one of the downstream cells cannot accommodate all the vehicles turning to this cell, some of its vehicles will remain in the upstream cell and block a portion of vehicles turning to the other cell.

$$
\begin{gather*}
y=\min \left\{S_{i}, R_{j_{1}} / \beta_{i, j_{1}}, R_{j_{1}} / \beta_{i, j_{1}}\right\}  \tag{2.7}\\
y_{i, j_{1}}=\beta_{i, j_{1}} y \quad \text { and } \quad y_{i, j_{2}}=\beta_{i, j_{2}} y \tag{2.8}
\end{gather*}
$$

If a merge junction exists and two roads join into one, as shown in Fig. 2.2 (c), there are two flows competing to enter the downstream segment. Using $i_{1}$ and $i_{2}$ to denote the indices of two upstream segments, and $j$ for the downstream segment, the flows from upstream segments to downstream ones have to satisfy Inequation (2.9) and (2.10).

$$
\begin{gather*}
y_{i_{1}, j} \leq S_{i_{1}} \quad \text { and } \quad y_{i_{2}, j} \leq S_{i_{2}}  \tag{2.9}\\
y_{i_{1}, j}+y_{i_{2}, j} \leq R_{j} \tag{2.10}
\end{gather*}
$$

In the case that $S_{i_{1}}+S_{i_{2}} \leq R_{j}$ (the downstream cell can accept all upstream sending vehicles), derivation shows that $y_{i_{1}, j}$ and $y_{i_{1}, j}$ can be obtained by Eqs. (2.11) and (2.12).

$$
\begin{align*}
f_{i_{1}, j} & =S_{i_{1}}  \tag{2.11}\\
f_{i_{2}, j} & =S_{i_{2}} \tag{2.12}
\end{align*}
$$

In the case that $S_{i_{1}}+S_{i_{2}}>R_{j}$, the cell transmission model utilizes a concept of priority to describe which flow is favored in merging. Here, $p_{i_{1}}$ and $p_{i_{2}}$ are used to represent the priorities of two upstream segments. These are fractions such that $p_{i_{1}}+p_{i_{2}}=1$. If $p_{i_{1}}=1$, absolute priority will be given to upstream segment $i_{1}$ (vehicles in segment $i_{2}$ cannot advance until all the sending vehicles from segment $i_{1}$ have moved to the downstream segment). According to the cell transmission model, flows can be computed by Eqs. (2.13) and (2.14). In these two equations, mid means the middle value among the three.

$$
\begin{align*}
f_{i_{1}} & =\operatorname{mid}\left\{S_{i_{1}}, R_{j}-S_{i_{2}}, p_{i_{1}} R_{j}\right\}  \tag{2.13}\\
f_{i_{2}} & =\operatorname{mid}\left\{S_{i_{2}}, R_{j}-S_{i_{1}}, p_{i_{2}} R_{j}\right\} \tag{2.14}
\end{align*}
$$

One of the strengths of the cell transmission model is its simplicity. From its update equations, it can be seen that only a few parameters in the fundamental diagram need to be calibrated in this model. If a triangular fundamental diagram is assumed, it can be determined by any three parameters from capacity, jam density, free-flow speed, shockwave congestion speed, and critical
occupancy. These parameters all have physical meanings, and they can be easily determined from scattered flow versus density data plots. The cell transmission model updates traffic states by explicit equations, and is very efficient in computation. Even though it is simple, however, it still can reproduce freeway traffic condition accurately. Section 3.1 gives a discussion of research related to the cell transmission model and its extensions.

## Microscopic Models

In microscopic models, the behaviors of individual vehicles are modeled. Each vehicle in the network is treated and tracked separately. The positions and speeds of all simulated vehicles are updated at every simulation time step. The time step is usually chosen in the range of 0.1-1.5 seconds to obtain a satisfactory accuracy. Vehicle trajectories are usually calculated based on a car-following algorithm, lane-changing algorithm, and gap-acceptance algorithm. These are the core of the microscopic model and largely determine its performance.

The car-following algorithm describes the response of vehicles to other vehicles around them. It determines the speed that drivers maintain under a certain traffic condition, the headway or spacing between vehicles, and the acceleration or deceleration of vehicles when drivers need to adjust their speeds.

The lane-changing algorithm defines the decisions that drivers make when they change lanes, merge, and weave. Drivers can have reasons for changing lanes, which typically fall into the following two categories: mandatory lane change (changing lanes to get into the correct lane for the destination) and discretionary lane change (changing lanes to get gain in speed). Once drivers decide to make a lane change, they have to judge whether it is possible to change lanes and plan the actions to take in order to ensure a safe lane change. Defining the motivation for lane change, and the complex maneuvers of making the judgment of possibility and determining a course of action, are all controlled by the lane change algorithm. Lane changing behavior is the most difficult part to model in driver behavior, and is often the source of irregular simulation performance. In a traffic condition where heavy lane changes are involved (due to merging or weaving), it usually takes great effort for simulation program users to calibrate their model in order to match what they observe in the field.

The gap-acceptance algorithm models the behavior when vehicles turn into or across conflicting traffic flows. It defines the conditions when a gap is acceptable for a safe turn or crossing. In many programs, a minimum gap is used, and any gaps less than this minimum threshold are flagged as unsafe and rejected by drivers. This minimum gap can vary as waiting time grows.

Microscopic models often utilize a set of parameters to specify drivers' behavior. Some of these parameters are hard to observe or measure. For instance, reaction time, which accounts for the time a driver takes to perceive, recognize, and respond, is used in several car-following models. However, this variable is difficult to measure directly. Microscopic simulation tools also allow users to choose different parameter values for different groups of vehicles, and allow probability
distributions of model parameters. As a result, determining the parameter values is usually a timeconsuming endeavor in microscopic simulation. Furthermore, vehicle states are usually updated at an interval less than one second in microscopic models. Because the states of each vehicle have to be computed every step, the computational effort is often large.

Because microscopic models are hard to program, software packages have been developed for researchers and traffic engineers. Some popular ones are PARAMICS [20], VISSIM [21], Aimsun [22], Simtraffic [23], and CORSIM [24]. These software programs differ in their driver models, modeling capability, ease of use, and extensibility. Reviews and comparisons of different microscopic simulation programs can be found in [25, 26]. Aimsun is used in several experiments in this dissertation, and an introduction to this software is given below.

## Aimsun and Its Components

Aimsun (Advanced interactive Microscopic Simulator for Urban and Non-urban Networks) is a transportation simulation environment developed by TSS (Transport Simulation System) [22]. This software integrates macro, mesco and micro models in a single application. It provides a broad range of capabilities for modeling different networks and different vehicles types. Users can construct a network integrating various road types, such as freeway, urban street, pedestrian area, and any other road type define by users. Users are able to specify the features of any road type, like speed limit and lane width. A lane in a road section can also be designated as HOV lane or bus-only lane. Users are allowed to have different vehicle types in a simulation, such as automobile, truck and bus. There are several parameters for specifying a vehicle type, including maximum acceleration, maximum deceleration, vehicle length and width, and gap. Mean and deviation of these parameters are used to introducing randomness in simulation. Users have two ways to define the traffic demand in a simulation, Origin-destination matrices and boundary flows with turning proportions. Users can deal with a variety of traffic controls in Aimsun, including freeway ramp metering, pre-time and actuated signal control, stop control. Users can create any traffic incident and evaluate its impact, like lane closure and speed limit change. Users can access detailed performance in simulation. All frequently used variables like flow, speed, travel time, and etc., are offered in time series of values. If there is any performance or control type not provided by the software, users can program to collect data or implement their desired control through Aimsun's interface package.

Aimsun has the following major components: graphical user interface, microscopic simulator, mesoscopic simulator, hybrid simulator, macroscopic module, Aimsun Platform SDK (Software Development Kit), API, and microsimulator SDK. Graphical user interface, microscopic simulator, API and microsimulation SDK are used in the studies of Chapter 5 to Chapter 7. They will be briefly introduced in the following subsections. Detailed description on Aimsun's different components can be found in [27].

## Aimsun Graphical User Interface

Aimsun graphical user interface is a user-friendly interface to build a simulation network and monitor simulation. Any simulation elements can be configured in this interface. Users can draw the geometry of a network by placing and connecting road sections, or import from supporting data files. Aimsun supports importing from openstreemap, CONTRAM, PARAMICS, VISSIM, TRANSCAD, CUBE, VISUM, ROAD XML, GIS and SYNCHRO. Users can input traffic demand in network editor. Different types of vehicles and time-series of demand data can be used. Users can also configure controls, create traffic incident and set up management plans in user interface. During an animation simulation, users can watch vehicle movements and create plots of performance variables. This graphical user interface is a tool to help user construct a simulation and evaluate performance in an ease and straightforward way. Fig. 2.4 shows the window of user interface.


Figure 2.4 Aimsun Graphical User Interface

## Aimsun Micro Simulator

The Aimsun micro simulator [28] is the core component used to simulate traffic in a microscopic way in different types of networks. In every time step, it updates the control events, road conditions and states of simulated vehicles. It also generates new vehicles into network and collect traffic data. Aimsun accepts simulation steps to be between 0.1 second and 1.5 second. States of vehicles are computed according to the driver behavior chosen by users, either the model built in the software or user-written models generated through microsimulator SDK. The car-following model built in Aimsun is the Gipps' car-following model. Fig. 2.5 from Aimsun's manual illustrates the
update process of the micro-simulator. This process can be divided into five steps. In the first step, the simulator updates events that are independent of other activities, for instance traffic light changes of pre-timed signal control. After this, the simulator updates the states of each vehicle according to the selected driver's model. In the study of Chapter 7 , we use the Aimsun MicroSDK to program user-defined driver behavior. This program is called in this step. In the third step, the simulator generates new vehicles if space is available. In the fourth step, it refreshes graphical user interface display. Finally, if simulated detection data is asked to be generated and outputted and it is available, the simulator calls the outputting process.


Figure 2.5 Update Process of Aimsun Micro Simulator, Taken from Aimsun's Manual

## Aimsun API

The Aimsun API (Application Programming Interface) [29] is a powerful tool to extend the capabilities of traffic modeling and control. It allows users to program applications to dynamically interact with the micro-simulator during simulation execution. The applications can be written in C++ or python. Users can program an API application to collect traffic data, change traffic demand, modify parameters, or execute a control action which is not provided by the software. A
rich set of interface functions are available.
Fig. 2.6 depicts the interaction between the Aimsun simulation model and its API, and the interaction between the Aimsun API and an external application. The Aimsun API module can be viewed as the bridge between the simulation software and the user's application. In the studies of Chapter 5 to Chapter 7, user-defined control strategies are implemented in simulation through the use of an Aimsun API. The detailed interaction within the dashed box in Fig. 2.6 are shown by Fig. 2.7. There are six functions. The function AAPILoad() and AAPIUnLoad() are called when the API module is loaded and unloaded by the software, respectively. The function AAPIInit() is called before simulation starts, and it is used to initialize the external application and any initial states the simulation needs. The function AAPIManage() is called at the beginning of each simulation step while the function AAPIPostManage is after each step. The function AAPIFinish() is called when simulation terminates.


Figure 2.6 The Role of Aimsun API in Simulation, Taken from Aimsun's Manual


Figure 2.7 Interaction between Aimsun Micro Simulator and Aimsun API, Taken from Aimsun's Manual

## Aimsun Micro SDK

Aimsun micro SDK (Micro-simulator Software Development Kit) [30] is a tool to program userdefined driver behavior. Micro SDK provides the functionalities to access and modify vehicle states
during simulation run, thus users can overwrite the driver behavioral model offered by Aimsun by default. User-defined driver behavioral models are implemented through C++ plug-ins. This plugins is only activated when users select to update vehicle by user-defined model. In the update of every vehicle in every step, the user-defined driver model is called by the micro-simulator. If a vehicle is not updated by the user-defined driver model, Aimsun updates it by its default model. This means users can choose to apply their driver models globally to the whole network or locally to specific sections, to all simulated vehicles or a subset of vehicles. Users can also overwrite the complete driver model or partial of it. If the default attributes of vehicle states are insufficient to implement a new behavioral model, users are allowed to add new attributes.

## Chapter 3

## Evaluation of An Arterial Link-Node Cell

## Transmission Model Using Traffic Data

The cell transmission model is one of the most widely used macroscopic models. It is simple and of reliable accuracy in freeway modeling. This model and its extended models have been applied in both traffic simulation and control design, for both freeway traffic and urban street traffic. However, there are insufficient results to show the accuracy of this model and the extended models when they simulate urban street traffic. This chapter evaluates this problem using fieldcollected data. It is organized into six sections. Section 3.1 is an introduction to the studied model, the link-node cell transmission model, which is an extension of the cell transmission model, and a description of the related research on the cell transmission model and its extension. Section 3.2 describes in detail how the link-node cell transmission model is applied in arterial traffic modeling. Section 3.3 describes the data set used in the evaluation and draws some observations from this data set. Section 3.4 explains how the information in the data set is transformed to the inputs of the simulation model, which includes the selection of the fundamental diagram, demand, split ratio, initial density, and signal timing plans. Section 3.5 presents the simulation results. Summary of this chapter is presented in section 3.7 .

### 3.1 INTRODUCTION

Traffic simulation tools are widely used in traffic management and operation planning. They offer an economic way for evaluating transportation control strategies and help designers to find out potentially defects caused by the designed controls. To get a valid evaluation, simulation tools have to use traffic models that can accurately simulate traffic characteristics and reproduce traffic
conditions of interest.
Researchers have proposed a variety of traffic models in the past few decades. These models can be divided into three categories: macroscopic, mescoscopic and microscopic. A macroscopic traffic model considers traffic in terms of aggregated density and flow. Among different macroscopic models, the cell transmission model [9, 10] is one of the most widely used models, and it has been applied to both simulation and control design.

In Section 2.2, we described how traffic dynamics was modeled and computed in the cell transmission model. As mentioned in that section, the cell transmission model as originally proposed in [9, 10] only adopts cells of uniform length and three types of link-node topologies. This may be inconvenient when it is applied to general networks for two reasons. First, the lengths of all road segments might not be multiples of a common number. An approximation is needed if we stick to the requirement of uniform cell length. Second, it is hard or inconvenient to represent an arterial network by the three allowable topologies. Arterial networks usually contain closely spaced intersections, and one intersection commonly has more than two entering links and more than two leaving links. As a result, an intersection has to be divided into many sub-networks of merge- and diverge-type. For these two reasons, researchers extended the cell transmission model to allow a more flexible representation of the network. In this study, the Link-Node Cell Transmission Model [31], which is a modification of Daganzo's cell transmission model, will be used.

## Link-Node Cell Transmission Model

In this subsection, the basics of the link-node cell transmission model (LN-CTM) will be introduced. As its name suggested, the LN-CTM used links and nodes to represent a network. The following notations are used in this subsection to introduce the link-node cell transmission model. They have the same definitions as defined in the previous chapter where the cell transmission model was described. Through the introduction in the remaining part of this subsection, it can be seen that the LN-CTM shares many similarities of the CTM, while it allows more flexible application. Notice that, in this chapter, we will distinguish the words link and cell when we introduce the LN-CTM. The word link is used to represent a road segment connecting two nodes. The word cell is used to represent a relatively short road segment which is produced when we divide a long link into smaller segments for the purpose of obtaining necessary modeling accuracy. In this sense, we can consider a link as one or more cells connecting in series.
$i, j$ : index of a link.
$k$ : index of time step.
$\rho$ : density of a link.
$f$ : flow entering or exiting a link
$L$ : length of a cell.
$S$ : sending function, or called as demand.
$R$ : receiving function, or called as supply.
$v_{f}$ : free-flow speed of a link.
$w$ : shockwave speed of a link.
$Q:$ capacity of a link.
$p$ : priority factor of a link.
$\beta$ : split ratio (turning percentage) of a link.

The Link-Node Cell Transmission Model (LN-CTM) is an extension of Daganzo's basic cell transmission model. It allows non-uniform cell lengths and multiple entering and leaving links. In the LN-CTM, a traffic network is represented as a directed graph with links and nodes, as shown by Fig. 3.1. A link denotes a segment of one-way road, with a uniform fundamental diagram anywhere within this link. A node denotes a junction. It can be a physical junction where two roads merge or diverge, or it can be a virtual junction where the fundamental diagram changes. A long link can be further divided into smaller cells to satisfy the convergence requirement. If the free-flow speed of a link is $v_{f}$ and simulation time step is $\Delta T$, the requirement of cell length $L$ is that $L \geq v_{f} \Delta T$. This is slightly different from the cell transmission model, in which it is an equality constraint.


Figure 3.1 Representation of Roads in the Link-Node Cell Transmission Model

The LN-CTM is a density-based model. Similar to the CTM, it contains an equation to represent vehicle conservation, as shown by Eq. (3.1). This equation is only converting $n$ in Eq. (2.3) into density $\rho$. In the case when a node only connects one entering link and one leaving link (ordinary network) as shown in Fig. 3.2 (a), flow from the upstream link to the downstream link is computed in the same way as in the CTM, as shown by Eqs. (3.2) to (3.4). These three equations are very similar to Eqs. (2.4) to (2.6). In the basic cell transmission model, because the cell length is $L=v_{f} \Delta T$, even the last vehicle in a cell can traverse the whole cell in one time step
if it drives at free-flow speed. Hence, in Eq. (2.4), all the vehicles in the cell are demanded to be sent downstream. However, the link-node cell transmission model allows cell length to be larger than $v_{f} \Delta T$. Only the portion of vehicles that can reach the tail of a cell are able to be sent, which can be computed by $v_{f} \rho \Delta T$. The difference of allowable cell length also explains the difference between Eqs. (2.5) and (3.3).

The LN-CTM allows general junctions with multiple entering links and leaving links. To define how flows diverge and merge, split ratio (turning proportion) matrices are used to represent the turning proportions at junctions. A split ratio matrix is a $M \times N$ matrix, where $M$ is the number of entering links and $N$ is the number of leaving links. Each entry in the matrix specifies the percentage of the vehicles in an upstream link that are directing to a downstream link. Hence each row in the split ratio matrix should sum up to one.

In the case of pure diverging as shown in Fig. 3.2(b), the flow update equations for a junction are shown by Eqs. (3.5) and (3.6). This is very similar to the CTM, except for the use of multiple downstream segments. Hence, multiple receiving functions, are considered in the min operation. In the case of pure merging as shown in Fig. 3.2 (c), the flow is computed in a manner similar to the CTM, but the priority of each upstream segment is set to be proportional to its demand (as shown by Eq. (3.10)). Thus, the priority factor in the LN-CTM is time-varying if the demands are time-varying, while in the CTM it is a constant. Also, the priority factor in the LN-CTM is automatically computed in the model, while in the CTM it is a user-defined parameter. When the total upstream demand is no greater than the downstream supply ( $\sum_{i} S_{i} \leq R_{j}$ ), flow leaving an upstream segment is its demand, as indicated by Eq. (3.7). Otherwise, flows are computed by Eqs. (3.8) to (3.10). In the case of both diverging and merging occurring in a single node, computation of flows can be much more complicated. Since this case does not appear in the study discussed in this chapter, it will not be described here. More details of the link-node cell transmission model can be found in [31].

(a) ordinary links

(b) diverge network

(c) merge network

Figure 3.2 Topologies Used in the Link-Node Cell Transmission Model

Density update equation:

$$
\begin{equation*}
\rho_{i}(k+1)=\rho_{i}(k)+\left(f_{i-1, i}(k+1)-f_{i, i+1}(k+1)\right) / L \tag{3.1}
\end{equation*}
$$

Flow update equations in ordinary links:

$$
\begin{align*}
S_{i} & =\min \left\{v_{f} \rho_{i}, Q_{i}\right\}  \tag{3.2}\\
R_{i+1} & =\min \left\{Q_{i+1}, w\left(\rho_{i+1}-\rho_{J a m}\right)\right\}  \tag{3.3}\\
f_{i, i+1} & =\min \left\{S_{i}, R_{i+1}\right\} \tag{3.4}
\end{align*}
$$

Flow update equations in diverge network:

$$
\begin{align*}
f & =\min \left\{S_{i}, \min _{j}\left\{R_{j} / \beta_{i, j}\right\}\right\}  \tag{3.5}\\
f_{i, j} & =\beta_{i, j} f \tag{3.6}
\end{align*}
$$

Flow update equations in merge network:

$$
\begin{array}{lrl}
\text { (if } \left.\sum_{i} S_{i} \leq R_{j}\right) & f_{i, j} & =S_{i} \\
\text { (if } \left.\sum_{i} S_{i}>R_{j}\right) & f & =R_{j} \\
f_{i, j} & =p_{i} f \\
p_{i} & =\frac{S_{i}}{\sum_{i} S_{i}} \tag{3.10}
\end{array}
$$

From the update equations above, it can be seen that the LN-CTM is consistent with the CTM. The major difference of the LN-CTM to the CTM is that: 1) it allows non-uniform cell length, which is stated by the constraint of cell length $\left(L \geq v_{f} \Delta T\right)$; 2) it allows more general network topology, as partially shown by Figs. 3.2 (b) and (c); 3) the priority factor is time-varying and is proportional to the demand, as shown by Eq. 3.10 .

In Section 3.2, we will describe how the LN-CTM can be applied on arterial traffic modeling, with the example of the studied site in this chapter. But before that, it is good to review the related research in the following subsection.

## Related Works

The cell transmission model and its extension models are studied and utilized by many researchers. This subsection contains a brief introduction of related works.

One critical concern of a simulation model is its accuracy. Some researchers have demonstrated that the CTM and the LN-CTM, if properly calibrated, were able to offer reliable accuracy in simulation of freeway traffic. Lin et al. [32] validated the cell transmission model on a signal freeway link without on-ramp and off-ramp, using data from I-680 in California. Munoz et al. [33] calibrated the CTM with traffic data on a segment of I-210 in California, which contained on-ramps and off-ramps. In this calibration procedure, the free-flow speed, shockwave speed and jam density were selected by least-square fitting, and bottleneck capacity was estimated from downstream measured flow and on-ramp flow. Simulation results presented in this study showed that the CTM could match the collected measured travel time. Muralidharan et al. [34] calibrated the link-node cell transmission model in a case where ramp flow data was missing. In this calibration, freeway flow and density data were used to calibrate the fundamental diagram, and an adaptive iterative learning procedure was adopted to impute the missing ramp flow. The result showed that the simulation can re-create the traffic condition with good accuracy. The error in terms of density, speed, travel distance and travel time was attractively small.

Because of its simplicity and reliability, the cell transmission model was utilized in a wide range of control strategy designs. Ziliaskopoulos [35] and Golani et al. [36] applied it in traffic assignment. Waller et al. [37] used it in network design. Among the applications of the cell transmission model in urban street signal design, we can find examples in [38, 39, 40, 41] and many other papers. In [38], Almasri et al. modeled the traffic flow on urban streets with the cell transmission model and developed an online offset optimization method based on this method. This optimization tried to minimize the delay using a Genetic Algorithm. Similarly, Lo et al. in [39] used the cell transmission model to formulate the traffic dynamic and optimized signal timing by a mixed integer program. The optimization in the proposed method also took delay as the objective function. Ukkusuri et al. in [40] proposed a robust signal timing optimization accounting for the uncertainty. The formulation of the optimization problem employed an embedded cell transmission model for the traffic dynamics. In [41], TRANSYT, a transportation simulation tool developed by TRL (Transport Research Laboratory), offers the cell transmission model and platoon dispersion model in its traffic simulation and signal timing optimization.

However, compared with the work of calibrating and validating the cell transmission model using freeway data, there were insufficient results on testing its accuracy using urban traffic data, even though many control proposals on urban streets have already utilized this model in their control design and performance evaluation. The rest of this chapter tries to address the question of whether the cell transmission model is able to accurately simulate urban traffic. This model will be tested using field-collected data, and its simulation output will be compared with real measurements. The traffic data used is the NGSIM data collected in Lankershim Boulevard in Los Angeles, which is described in Section 3.3 .

### 3.2 APPLYING THE LN-CTM ON ARTERIAL NETWORK

## Representing Arterial Networks with Signalized Intersections Using the

## LN-CTM



Figure 3.3 Representation of an Arterial Road Segment with Signalized Intersections Using the LN-CTM

In previous section, we described the network representation and update equations of the link-node cell transmission model (LN-CTM). In this section we will describe how we represent the arterial network using the LN-CTM.

In the modeling in this study, we will represent a signalized intersection by a node and a oneway road section between two intersections will be represented by four links and a dummy node,
as shown by Fig. 3.3. Link \#1 denotes the road segment from the upstream intersection to the location where a left-turn bay starts. The remainder of the road segment is represented by another three links, denoted link \#2 to link \#4. Each of these three links represents the lane or lanes for one allowable movement, left-turning, through-traversing, or right-turning. In the case that a particular movement is prohibited, or actually there is no vehicle moving in a particular direction during the study time period, the corresponding link will be excluded. The node placed between the upstream link \#1 and downstream links \#2 to \#4 is called dummy node. The dummy node is not a physical junction. It is used to denote the approximate location where traffic flow split into different movements, so that we can apply the computation of Eqs. (3.5) to (3.6) in the simulation. As shown in Fig. 3.3. a standard four-way signalized intersection that has three allowable movements in each incoming way has twelve entering links and four leaving links. There is one dummy node in each direction between two signalized intersections.

Because the LN-CTM allows non-uniform cell lengths, as long as they meet the requirement of $L \geq v_{f} \Delta T$ as mentioned in the previous section, it is possible to have some long links in the simulation model without dividing them into smaller units (this means one link equals a cell). However, it is still good to keep the cell lengths in an appropriate range to gain good modeling accuracy. This is because the vehicles within a cell are considered to place evenly in the LN-CTM (also true in the CTM). Consider the situation that vehicles queue up at the intersections but the majority of the road segments between two intersections are empty. If long links are used without divided into smaller cells, the demands of those long links, which are computed by Eq. 3.3, are small due to the low average density. This leads to low exiting flows, which contradict with the reality that queuing vehicles leads to high exiting flows. As a result, it is necessary to divide long links into smaller cells of appropriate lengths to gain modeling accuracy.

To apply the signal timing in the LN-CTM, the links that terminate at the signalized intersection will have time-varying exiting flow capacities. If the signal light is green for a particular movement, the exiting flow capacity (the $Q_{i}$ in the Eq. 3.3) of associated link is set to saturation flow (the maximum traffic flow that the movement can reach). On the other hand, if the signal light is red, the capacity of the link associated with that signal light is set to zero to reflect the restriction of moving. The yellow light will be interpreted as red in the simulation of this study for modeling simplicity. In the case that a long link is divided into smaller cells, the time-varying exiting flow capacity is applied at the cell that immediately connects to the signalized intersection.

For modeling simplicity, a source link will be added to each origin in order to represent the most upstream segment. Flow demand profile will be used to specify the arrival flow of every source link. Source links in the LN-CTM have zero length, but they are able to store a queue of arbitrary length in the simulation if the generated demand fails to move to a downstream link. Hence, a source link can be viewed as a special link with known arrival flow and unlimited storage. The arrival flow and the queue of a source link jointly determine the actual demand of that source link, replacing the term $v_{f} \rho_{i}$ in Eq. 3.3. Similarly, a sink link will be added at each destination to represent the most downstream segment. A sink link also has zero length, and it can accept any arrival flow as long as the flow is no greater than the capacity of the link. This means the term
$w\left(\rho_{i+1}-\rho_{J a m}\right)$ is dropped for a sink link.
From the directed graph representation described above, it can be observed that traffic flow diverges at the dummy nodes, and merges at the signalized nodes. There are no nodes in this representation that have both diverging and merging simultaneously. Consequently, flows are computed by Eqs. (3.5) to (3.6) at the dummy nodes, while they are computed by Eqs. (3.7) to (3.10) at the signalized nodes.

The LN-CTM is implemented in Aurora [42], a simulation program developed in TOPL [43], in the same way of computation as described above. This software will be used as the tool for simulating traffic in the study discussed in this chapter.

## Representing the Studied Site in the LN-CTM

The data set used in the study presented in this chapter is the Lankershim data set generated by the Next Generation Simulation (NGSIM) project [7] of the Federal Highway Administration (FHWA). It is one of the four data sets produced by the NGSIM project. In Chapter 2.1 we have introduced this project. In Section 3.3, there is a detailed description of the information given in the data set used in the chapter. Here we will first describe the geometry of the studied site in this data set, and then explain how the geometry is represented in the simulation model.

The Lankershim data set was collected on a segment of Lankershim Boulevard between 8:28 and 9:00 am on June 16, 2005. This site is close to the Universal Studio in Los Angeles. Fig. 3.4 shows a map of the studied road segment. This arterial segment has a length of about 1600 feet, and it has three to four lanes in both directions. There are four signalized intersections within this segment, and they are labeled from south to north in ascending order in Fig. 3.4, as shown by the red numbers 1 to 4 in the rectangular box. There are left-turn bays at these four intersections in both directions of Lankershim Blvd. This road, from south to north, intersects with one off-ramp from US-101, Universal Hollywood Dr., and James Stewart Ave/Valleyheart Dr. In total, there are 11 entering roads and 10 exiting roads in this studied network. The entering roads are labeled with the pink numbers 101 to 111 in Fig. 3.4. The exiting roads are labeled with the pink numbers 201 to 211 (without 202) in the same figure. The studied Lankershim road segment is divided into 5 sections by the signalized intersections, and from south to north they are labeled in orange in an ascending order.

In the study of this chapter, we will follow the identifications of all intersections, entering and exiting roads, and sections, which are labeled in Fig. 3.4. Thus, intersection \#1 refers to the south most signalized intersection in this network. Origin 102 will refer to the off-ramp from freeway. Here origin is used as the entering road of this studied network, not necessary meaning the true starting location of a trip. Similarly for the use of destination. Destination 201 refers to the south most exiting road. Section 3 refers to the road segments between intersection \#2 and intersection \#3, both directions.


Figure 3.4 Map of Studied Site, Taken from Lankershim Data Analysis Report of the NGSIM Program


Figure 3.5 Representing the Lankershim Network in the LN-CTM

Fig. 3.5 shows the link-node representation of the Lankershim study network. This representation is obtained by applying the rules stated in the previous subsection. The four red nodes in this figure represent the four signalized intersections. The numbers in the rectangular box are the identifications of the origins and destinations, in the same order of Fig. 3.4. The dummy nodes, where traffic flows split, are shown by unfilled black circles. The road segments are represented by the direct links. Notice that the long links in the figure were divided into cells in the simulation, but this division is not shown by the figure in order to give a clear view. Because the free-flow speed on Lankershim Blvd is 35 mph , and the simulation time step was one second in this study, the length of each cell needed to be no less than 51.3 feet (as stated by $L \geq v_{f} \delta T$ ). When a long link was divided into cells in this study, the cells of the same link had the same length, and the division was to make the cell length as short as possible.

Notice that there are two driveways from two parking garages in the studied network, as shown by origin 104 and 106 in Fig. 3.4. To represent the junctions of the two driveways to Lankershim Blvd, two nodes were added to the network, which are represented by the blue circles in Fig. 3.5. These two junctions are modeled in the same way as signalized intersections, except that they do not have signal lights that change the exit flow capacity. Hence, both of these two nodes have links like link \#4 in Fig. 3.3 connecting to them to represent allowable turnings at corresponding junctions, but the flow capacities of those links are constant.

### 3.3 DESCRIPTION OF THE NGSIM LANKERSHIM

## STUDY SITE

## Data Set

In this section, we will describe the data presented in the NGSIM project, mainly focusing on the data files that were used in the study presented in this chapter. We also analyze the traffic condition at the studied site by analyzing the video and trajectory data given in the data set.

The complete package of Lankershim data set contains many different data files. It includes text files of vehicle trajectory data, video files of raw and processed video, image files of ortho photographs, CAD drawings, files of signal timing, text files of detector data, GIS files and files of data analysis. In the study discussed in this chapter, only vehicle trajectory data and signal timing plans were used for the inputs of the simulation.

The NGSIM vehicle trajectory data was collected by taking video from five cameras. Vehicles were detected and tracked from video images, and then vehicle trajectories were obtained through mapping the vehicle location in videos to map coordinates. The studied period was divided into two pieces: 8:28-8:45 am and 8:45-9:00 am. Thus, there are two files for vehicle trajectory. In
total, there are about 1.6 million rows of data in these two files. Each row of data contains vehicle identification, time, local/global $x$ and y position, vehicle length and width, speed, acceleration, origin and destination, the lane/intersection/section/direction that the vehicle was driving on, preceding and following vehicle, etc. This trajectory data will be used to obtain the fundamental diagram, demand, split ratios and initial density for the simulation discussed below.

In the signal timing files, signal timing sheets and real-time split monitor reports are presented. In a signal timing sheet for a particular intersection, the configuration of the signal plan is provided. It includes phase assignments, yellow and red time durations, maximum green lengths, gaps, cycle length and offset. According to the signal timing sheets, the intersections were in actuated control (Notice that the introduction of actuated control will be given in Section 6.1, where signal controls are introduced.). In a real-time monitor report, green durations for all the phases during the studied period are recorded, together with the start and end time of each cycle.

It needs to be noted that the video clock was not synchronized with the signals clock in the Lankershim data set. This can be determined by the information provided by the real-time signal report and observing the changes in the signal lights shown in the video. If the clock in the video (or trajectory data) is viewed as reference, the signal clocks at intersection \#1 to intersection \#4 were about 3 seconds behind, 28 seconds ahead, 17 seconds ahead, and 30 seconds ahead, respectively.

## Demand Distribution

As shown by Fig. 3.4, the studied network has 11 entering roads (labeled from 101 to 111) and 10 exiting roads (labeled from 201 to 211 but without 202). Table 3.1 shows the total volumes of tracked vehicles for each origin-destination pair during the $32-\mathrm{min}$ video. From this table, it can be observed that the largest flow in this small network was the one in the southbound direction from origin 108 to destination 201, which was the flow traveling from the most upstream origin to the most downstream destination in the southbound direction. The second largest flow was from origin 102 to destination 208, traveling in the northbound direction. This was the flow turning into Lankershim Blvd from the off-ramp of US-101. The third largest flow was from origins 108 to destination 203. This was the flow traveling in the southbound direction along Lankershim Blvd and turning to Universal Studio at intersection \#2. It should be noticed that the flow from origin 101 was also significant. The distribution of traffic demand indicates that most of the vehicles went straight at the intersections, and there was a large left-turn flow at intersection \#2.

Among the 2439 vehicles tracked in the video, 2367 of them were automobiles. They accounted for $97.05 \%$ of the total vehicles. These automobiles had an average vehicle length of 15.14 feet. 68 of the vehicles were buses and trucks, with an average vehicle length of 32.71 feet. Buses and trucks were about $2.79 \%$ of all the vehicles. There were 4 motorcycles, only $0.16 \%$ of all the tracked vehicles.

Table 3.1 The Volumes of Tracked Vehicles in the Studied Network during the Data Collecting Period (unit: number of vehicles)

|  | destination |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathbf{2 0 1}$ | $\mathbf{2 0 3}$ | $\mathbf{2 0 4}$ | $\mathbf{2 0 5}$ | $\mathbf{2 0 6}$ | $\mathbf{2 0 7}$ | $\mathbf{2 0 8}$ | $\mathbf{2 0 9}$ | $\mathbf{2 1 0}$ | $\mathbf{2 1 1}$ | total |
| $\mathbf{1 0 1}$ | 0 | 106 | 9 | 11 | 4 | 11 | 149 | 3 | 7 | 22 | 322 |
| $\mathbf{1 0 2}$ | 33 | 71 | 12 | 14 | 3 | 40 | 340 | 2 | 1 | 4 | 520 |
| $\mathbf{1 0 3}$ | 20 | 0 | 2 | 0 | 0 | 1 | 70 | 0 | 1 | 28 | 122 |
| $\mathbf{1 0 4}$ | 0 | 3 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 2 | 9 |
| $\mathbf{1 0 5}$ | 7 | 6 | 1 | 0 | 1 | 0 | 13 | 0 | 0 | 2 | 30 |
| $\mathbf{1 0 6}$ | 1 | 1 | 0 | 0 | 0 | 0 | 4 | 1 | 0 | 1 | 8 |
| $\mathbf{1 0 7}$ | 2 | 8 | 2 | 0 | 0 | 0 | 10 | 1 | 0 | 1 | 24 |
| $\mathbf{1 0 8}$ | 712 | 319 | 30 | 26 | 5 | 45 | 2 | 8 | 29 | 52 | 1228 |
| $\mathbf{1 0 9}$ | 8 | 3 | 0 | 0 | 0 | 0 | 9 | 0 | 0 | 5 | 25 |
| $\mathbf{1 1 0}$ | 2 | 2 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 10 |
| $\mathbf{1 1 1}$ | 7 | 57 | 7 | 11 | 0 | 16 | 43 | 0 | 0 | 0 | 141 |
| $\mathbf{t o t a l}$ | 792 | 576 | 63 | 63 | 13 | 114 | 647 | 15 | 39 | 117 | 2439 |

## Traffic at Each Intersection

Before we evaluated the CTM using the NGSIM data, we analyzed the traffic condition at each intersection from both video and trajectory data. The following paragraphs summarize some important factors that needed to be taken into consideration when we conducted our simulation study.

There is an intersection downstream of intersection \#1 in the southbound direction which was not covered in the video recording of the NGSIM data and, as a consequence, was not included in the CTM model. Southbound traffic can congest upstream of this unmodeled intersection and, as a consequence, queues can spillback into intersection \#1. Upstream queue spillback into intersection \#1 caused a decrease in the southbound exit vehicle flow, which can be observed from video
but may not be properly accounted for in the simulation with a free-flow boundary condition in the CTM model. Fortunately, this queue spillback phenomenon only took place a few instances during the studied time period. Moreover, when the queue spilled back, it dissipated very quickly. On northbound direction, there were a large portion of vehicles from the off-ramp turning right into Lankershim Blvd. As a consequence, the northbound road segment located upstream of intersection \#1 was often fully congested at intersection \#1 from the freeway off-ramp when the main northbound traffic light was red. Queue spillover from this segment was observed from about 8:49 to 9:00 am at this intersection.

At intersection \#2 in the southbound direction, there was a large left-turn flow towards US-101 North. Signal settings along both directions on this road had protected left turns (Description of protected left turn can be found in Section 6.1.). Consequently, the green duration for southbound left-turn was often long. It could be seen that there were occasional queue spillovers on the turn bays along the southbound direction. That is, the queue in the southbound left-turn bay was usually long and the bay was fully occupied sometimes. However, the queue spillover did not affect the flow of through vehicles much because there were four lanes on this section. The right-turn bay along the southbound direction was very short. Thus there were times that the right-turn vehicles failed to get into the bay as the through vehicles blocked their way. Queue spillover from any downstream intersections did not happened at this intersection during the studied time period. However, there was a bus stopping on the rightmost lane in the northbound downstream section around 8:43am, and this stop caused a small and temporary queue that reached the intersection. Similarly to intersection \#1, a significant amount of right-turn vehicles at this intersection were observed to turn during the red time. When the lights changed to green, turning vehicles had to wait for a couple seconds due to the large amount of pedestrian traffic at this intersection.

At intersection \#3, the turning traffic was from and to a parking lot. This flow was small, therefore most of the vehicles went through at this intersection and the green time on Lankershim Blvd was very long. During the studied period, there was no downstream queue spillover appearing at this intersection. Even though there was a few seconds of protected green for left-turn vehicles on Lankershim Blvd, they usually arrived and turned during the permissive period. The left-turn bays usually had short, temporary queues, and there was not any spillover at the turning bays. Very similarly to other intersections on this road, a large number of right-turn vehicles from the parking lot were seen to turn into Lankershim Blvd during the red time after a stop.

At intersection \#4, there was very little cross street traffic. However, the green durations for the cross streets were still activated almost every cycle because there was often a pedestrian call or a vehicle call. Queue spillover from the downstream section or on the turning bay did not happen at this intersection. Even though one of the left-turns on Lankershim Blvd had a protected green time, it was seldom activated in the studied time period because left-turn vehicles usually came and left during the permissive time.

### 3.4 INPUTS TO THE SIMULATION MODEL

To conduct the simulation using the LN-CTM, it needs the inputs of road geometry, fundamental diagram to specify the property of the links, the traffic demands and turning ratios, and the signal control plans. The road geometry was represented by the link-node representation which is given in Section 3.2. In this section, we will explain how the other inputs are generated from the Lankershim data set.

## Fundamental Diagrams

The link fundamental diagram contains parameters that affect the accuracy of the link-node cell transmission model. Thus, these parameters need to be properly estimated. Fundamental diagrams of a triangular shape as shown in Fig. 2.1 were used in the simulation presented in this chapter. The parameters that need to be calibrated are the capacity (maximum discharge flow), free-flow speed and jam density.

Because the NGSIM Lankershim network is small and there are no significant changes in geometry along the road, free-flow speed and jam density were considered to be uniform for all the links in the simulation network. Jam Density was assumed to be 200vpm (vehicles per mile), which means about 26.4 feet per vehicle when vehicles stop in queue. As mentioned in section 3.3 , the average length of cars tracked in the video was 15.14 feet. Therefore, this choice of jam density is reasonable. From the simulation study, it was found that the simulation results were insensitive to the value of jam density. The value of the free-flow speed used in the simulation was 35 mph , which is the speed limit of the studied road. Fig. 3.6 shows the speed histograms of northbound and southbound traffic flows. These histograms were produced by taking all the speed values that were no less than 30 mph for any vehicles and at any time step. It can be seen that the speeds were highly concentrated at 35 mph .

The maximum exiting flow for the links representing turning movements (link \#2 to link \#4 in Fig. 3.3) are in general different from each other, because drivers adopt different time gaps depending on whether they are executing a left-turn, a straight advance or a right-turn. The maximum exiting flow of each of these links was estimated from headways extracted from the NGSIM trajectory data. We assumed that the headways generated from queue discharging were less than 3 seconds. From the trajectory data, the time stamps of vehicles passing an intersection can be obtained, which in turn provides the headways between successive vehicles. For each movement at each intersection, headways within 3 seconds were collected and aggregated. The maximum exiting flow was computed as 3600 divided by the mean headway. The estimated saturation flows were between 1800 to 2000 vph (vehicles per hour) for the through movement on Lankershim Blvd, and between 1400 to 1600 vph for the right-turn movement. For all left-turns and some right-turns, there were not enough samples of headways in the NGSIM Lankershim data set. Thus, we had to assume that the maximum exiting flows for the corresponding links were 1800 vph . The sample size for flows from cross street were not large enough either. Therefore, 1800 vph was also
used as the capacity for these cross street links. The value of 1800 vph was used because it was not too low to produce unwanted queues in the simulation. In the simulation study of this chapter, it was found out that the simulation results were insensitive to this value, because the traffic volumes of those links were low.


Figure 3.6 Speed Histograms on the NGSIM Lankershim Network Based on the Trajectory Data

## Demands, Split Ratios and Initial Density Estimates

Demands, split ratios (turning proportions) and initial density estimates must be defined in addition to the fundamental diagrams in order to perform simulations. As mentioned in the subsection
above, a demand profile has to be provided for each source link as the arrival flow. This profile was obtained by counting the vehicles appearing at each origin using the NGSIM vehicle trajectory data. Because the intersections on this road were under actuated control, the length of the green time was time-varying. It could be seen from the NGSIM video data that the green times for cross streets could be as small as 5 to 10 seconds in some cycles. To capture an accurate time profile of vehicle arrival, the demand profile was updated every 10 seconds in the simulation. As a result, the number of vehicles arriving at each origin was measured at every 10 seconds and this value was used as the input demand.

At the dummy nodes, traffic flow was split into through, left-turn and right-turn, and split ratios are needed to define how flow splits. Because the trajectory and destination of every vehicle are known from the NGSIM trajectory data, it is possible to tell which turn a vehicle was going to take when it passed the location of a dummy node. Split ratios were calculated by counting the number of vehicles that turned during a period of 10 seconds. Thus as in the demand profiles, split ratio profiles were updated every 10 seconds in the simulation.

Since the network was not empty at the initial time of the study period, it was necessary to estimate the initial density of each link. Otherwise, there would be fewer vehicles in the simulation than in reality. Initial density estimates were obtained by measuring the number of vehicles in each link at the start time of the studied period.

## Signal Control

The four signalized intersections in the NGSIM Lankershim network are under actuated control. When a signal is under actuated control, it will extend the green time when a vehicle passes a detector, if the length of green time has not yet reached its maximum limit. The cell transmission model is a macroscopic model, which considers the average flow and density over a period. Thus, it does not simulate microscopic discrete events such as the arrival of a vehicle. As a consequence, signals cannot be manipulated by the model as in reality. However, since precise information regarding the green lengths in each cycle is available from the data, actuated signals can be modeled as they were under pre-time control, while keeping the green lengths changing at every cycle. In this way, each movement can get the same green time in the simulation as in the data. We also know the lengths of yellow time from the signal timing sheets. We saw some vehicles passed the intersection during yellow time. To simplify the simulation, yellow times were treated as red times, which means no flow would discharge during this period.

In Section 3.3, we mentioned that a large percent of right-turn vehicles turned during the red time period, especially at intersection \#1. Our model does not have the functionality to model the right-turns in red time. Moreover, since we did not want to over-restrict the right-turn flow, any of the right-turn links that were immediately linked to a signalized intersection were not affected by the signal. This meant that right-turn links could discharge flow during red times in the same manner as during green times, as long as there was sufficient space to accept the flow in its down-
stream link. In reality, the discharge flow during a red time is usually less than that during a green time, because drivers need to make a complete stop and look for a safe gap before they turn. Nevertheless, it is hard to estimate the reduction in discharge flow during the red time from the data. Therefore, we decided to keep it unchanged. This would result in some discrepancies because the exiting flow was larger than it should have been, but we believe that such discrepancies would be small to the most part. It needs to be mentioned that, even though a right-turn link is not controlled by the signal in the simulation, there can still be a right-turn queue, if the downstream link does not have enough space to accommodate the vehicles that desire to get in, or if the queue of through movement is too long and blocks vehicles from entering the right-turn link. It also needs to be mentioned that there were a large amount of pedestrians at intersection \#2, and they affected the turning vehicle flow. However, we did not consider this influence in our simulation and assume that the turning flows were only affected by their corresponding demands and the available vehicle space in the downstream links.

### 3.5 SIMULATION RESULTS AND ANALYSIS

Vehicular Traffic in the Lankershim study area from 8:30am-9:00 am was simulated in this study. As indicated previously, the simulation step was one second. Flow and travel time were collected and compared to their corresponding data counterparts, in order to evaluate the accuracy of simulation. Fig. 3.7 shows the simulated and measured flows along the northbound direction, while Fig. 3.8 shows the corresponding southbound direction flows. Fig. 3.9 compares the travel time obtained from the simulation with the measured travel times along both directions at the Lankershim Blvd. Because the cycle length of the four signalized intersections was 100 seconds, an interval of 100 seconds was also chosen to be the time interval of aggregation in our comparisons. Thus, each point in Fig. 3.7 and Fig. 3.8 represents the aggregated flow passing the intersection over 100 seconds. Each point in Fig. 3.9 is the sum of time for all vehicles travelled in the corresponding direction during 100 seconds. The relative errors between the simulated and measured values were computed using Eq. (3.11). In this equation, $i$ is the index of data values, and $N$ is the total number of data values.

The values of flows and travel times used in the comparison were computed as follows. The simulated flow of each link at intersections was obtained by recording the flows in the update equations (3.7), (3.8), and (3.9), because the intersections were merging nodes. The total simulated flow passing an intersection of a particular direction during an aggregation interval was the sum of the simulated exiting flows of left-turn, through, and right-turn links during the same aggregation period. The measured flows at the intersection were obtained by counting the number of vehicles leaving the intersections presented in the NGSIM trajectory data. The simulated travel time of an aggregation interval was obtained by counting the numbers of vehicles presented in the network at each simulation time step, summing them up, and then times the length of a simulation time step, which was one second. The number of vehicles presented in the network at each time step
came from two sources. One was the numbers of vehicles presented in the network links, which could be computed as the products of link densities (given in Eqs. (3.2p) and the lengths of the corresponding links. The other was the queues at the source links, which were obtained from the source link states. The measured travel time of an aggregation interval was obtained in a similar way of the simulated travel time. During each time step, the number of vehicles presented in the network was counted by searching the NGSIM trajectory data. The travel time was the product of the sum of vehicle counts and the length of time step.

$$
\begin{equation*}
\text { Error } \%=\frac{\sum_{i=1}^{N} \mid \text { simulated value(i)-measured value(i) } \mid}{\sum_{i=1}^{N} \text { measured value(i) }} \times 100 \% \tag{3.11}
\end{equation*}
$$

Examining the northbound simulated and actual vehicle flows in Fig. 3.7, it can be seen that the simulated flows matched the corresponding measured flows very well. The relative errors in intersections \#1 to \#4 (upstream to downstream) were $8.08 \%, 6.23 \%, 12.37 \%$, and $10.61 \%$, respectively. The average relative error for the northbound flow was $9.41 \%$.

Notice that Intersection \#1 is the most upstream intersection in the northbound direction and its arrival flow was generated from the demand profile. Hence, one may wonder why its simulated flow did not match the actual flow well after 1,100 seconds, even though the demand and signal timing were accurate. A possible explanation can be deduced by observing from Fig. 3.7 that the simulated flow almost overlapped with measured flow up to 1,100 seconds, but the error became large after that. The large discrepancy after 1,100 seconds can be attributed to the fact that the rightturn flow from the off-ramp was not controlled by the signal at intersection \#1 in the simulation. As mentioned in Section 3.3, this right-turn incoming flow at intersection \#1 was very large. Also mentioned in Section 3.3, the downstream link of intersection \#1 in the northbound direction was not filled up before 9:00 am (900 seconds in the simulation), which was observed from the NGSIM video and trajectory data. Hence, the flows at intersection \#1 were not restricted by the density of the downstream link. After 9:00 am, the downstream link gradually became full, and queue spillover occurred. The simulated flow of the northbound most upstream link and the simulated flow of the right-turn traffic from the off-ramp shared the available downstream vehicle space. Remember the exiting flow was defined to be proportional to the corresponding by Eqs. (3.9) and (3.10). Due to the fact that the right-turn flow from the off-ramp was uncontrolled by the signal light in the simulation, and its demand was large, the simulated off-ramp exiting flow was larger than what we measured in the reality until the density of the northbound most upstream link accumulated to a point. As a consequence, the simulated flow of the northbound most upstream link temporarily became smaller than the measured flow, as indicated by the lower flow on the red dashed line to the solid blue line around 1,200 seconds. From then on, the simulated flow of intersection \#1 did not match the actual flow well. If the way to model right-turn at red is improved, this error can be reduced. However, this may introduce a discrete event in the model, and it is not in the scope of this study.

The simulated flow at intersection \#3 did not match the actual flow well. After carefully examining the simulation and the NGSIM data, we found that this was caused by variation of vehicle speed. Remember that in Section 3.3, it was mentioned that the green times on Lankershim Blvd at intersection \#3 were long, and few queues were presented at this intersection. As a consequence, it was expected in the simulation that vehicles traveled at the free-flow speed, which was 35 mph in the simulation, from intersections \#2 to \#4, without stopping at intersection \#3. However, it was observed from the NGSIM video and trajectory data that vehicles did not maintain 35 mph between intersection \#2 and \#4 in reality. The vehicle speeds varied among vehicles. The variation of vehicle speeds, led to the discrepancies of passing times at intersection \#3. Thus, there were discrepancies between the simulated flows and actual flows.

Fig. 3.8 shows the flow on the southbound direction from intersection \#1 to \#4 (downstream to upstream). The simulated flow again followed measured flow. The relative errors from intersections \#4 to \#1 were $2.96 \%, 9.31 \%, 8.77 \%$, and $15.50 \%$, respectively. The overall relative error on the southbound direction was $8.44 \%$. In this direction, the upstream intersection has a small error, and flow error propagates from upstream to downstream, which is the expectation.

Notice that the simulated flow at intersection \#1 did not match the actual flow well. This was a joint effect of the error propagation from upstream intersection and the unmodeled downstream traffic. Remember it was mentioned in Section 3.3 that there was an uncovered close downstream intersection of intersection \#1. The queues from this unmodeled downstream intersection occasionally reduced the speed and exiting flow of intersection \#1. As a consequence, the assumption of free-flow boundary condition in the simulation was true at intersection \#1. The flow discrepancies at this intersection could be reduced if the unmodeled intersection was included in the simulation.

Fig. 3.9 plots the travel time on the northbound direction and on the southbound direction. The simulated travel time on the northbound direction was less than the measured one, with a relative error of $19.83 \%$. The simulated travel time on the southbound direction was higher than the measured one, with a relative error of $30.16 \%$. A possible explanation of the travel time errors was the speed variations, or the non-homogeneity feature of arterial traffic. The vehicle speeds on arterial were frequently interrupted by the signal lights and intersection queues. Thus, arterial traffic flow was much less homogeneous than freeway traffic flow. Further, vehicles had finite acceleration and declaration. They could not change from zero speed to free-flow speed, or verse vice, in zero time as assumed in the simulation. As a consequence, the simulated travel time did not match the actual travel time, even though the simulated flows passing each intersection matched well.

Based on the simulation results, we think that the link-node cell transmission model is able to simulate the traffic on arterial and to give acceptable accuracy, given the following considerations. First, arterial traffic is not as homogeneous as freeway traffic. The flow and travel time errors of macroscopic simulations using freeway data were about $5 \%$ [33, 44]. It can be seen that the simulated flow error in the simulation of this chapter was comparable to the errors of freeway traffic simulation. The fact that travel time error in the simulation of this chapter was larger than the errors of freeway traffic simulations was expected for the inhomogeneous traffic feature. Second,
there were unmodeled traffic dynamics in the simulation of this chapter, such as the influence of pedestrian, the right-turns on red, and the uncovered intersection close to intersection \#1. Third, the aggregation interval used in the simulation was relatively small. The error equation (3.11) computed the absolute values of the differences between the simulated values and actually measured values. Any small variation was capture by the error equation because the small aggregation interval. If the aggregation interval had increased, the error percentage would be smaller, because the variation was smoothed.


Figure 3.7 Comparison between the Simulated and Actual Vehicle Flows, for All Four Signalized Intersections along the Lankershim Road in the Northbound Direction during the 8:45-9:15 am Time Period


Figure 3.8 Comparison between the Simulated and Actual Vehicle Flows, for All Four Signalized Intersections along the Lankershim Road in the Southbound Direction during the 8:45-9:15 am


Figure 3.9 Comparison between the Simulated and Actual Travel Times, for the Lankershim
Network in both the Northbound and Southbound Directions during the 8:45-9:15 am Time Period

### 3.6 LESSONS LEARNED

The following points were learned from the simulation study presented in this chapter.
First, the non-homogeneity feature of arterial traffic makes macroscopic simulations of arterial traffic less accurate compared to macroscopic freeway simulations. The interruption of traffic flow caused by signal control and the variation of vehicle speed both contribute to simulation error.

Second, the traffic dynamics of arterial traffic are difficult to model using macroscopic simulations. In the simulation study of this chapter, we did not model the influence of pedestrian, the right-turn on red, and the permissive left-turn. In addition, the actuate control was modeled using a time-varying pre-timed control. All these events are discrete and are hard to include in macroscopic models. If one tries to model them in macroscopic models, the model complexity will increase,
and the advantage of simpleness of macroscopic models may lose. However, These discrete events influence the modeling accuracy of arterial traffic. This is a trade-off between model accuracy and model simpleness.

Third, the update frequencies of demand profiles and split ratios need to be selected wisely. The arterial traffic flow is much less smooth than freeway traffic flow, and the variation is larger. If we use a large update interval (low update frequency) of traffic demand profiles and split ratios, we may lose the traffic pattern. When we conducted the simulation study presented above, we found that using an update interval of one minute caused significant error in simulated flow. This may be due to the fact that the variation in traffic flow of the studied network was large and we simulated a relative short period of time. In order to obtain good simulation results, it is good to use small update interval. However, the existing traffic detection facilities in most area provide traffic measurements in minutes. Hence, data of high resolution is not accessible.

Fourth, if the update interval is small in the simulation, the location for measuring traffic data need to consider carefully. Boundary flow is best to measure at the location that will not be impact by traffic congestion (maintain free-flow condition during the study period). In the simulation presented in this chapter, the measurement of boundary flow was limited by the scope of video cameras. Therefore, boundary queues were not captured well. Further, the split ratios were measured at the approximate location where traffic flows split in this simulation. If we measured them at intersection stop lines, it was often that the traffic was moving near the split location but stopped near the intersection stop line, due to the small update interval. This meant there would be bias in the split ratio measurements, and as a consequence, bias in the simulated flows.

Finally, the challenge of arterial traffic modeling, both macroscopic and microscopic, is the lack of necessary traffic data. As we can see from the simulation study presented in this chapter, some of the inputs to a simple macroscopic model were still inaccessible or directly measurable. We had vehicle trajectory data and video data in this simulation study, thus we can get most of the model inputs from the NGSIM data. However, the time period of the NGSIM data is still short and the network coverage is small, which does not allow us to carry out a more comprehensive simulation study to include more traffic phenomena over a larger scope and time period. Generally, only traffic data measured from point detectors (like loop detectors) will not be available in most of the signalized intersections, which causes model inputs such as initial density and split ratios cannot be obtained. In some circumstances, flow data of a portion of the modeled road segments is not available, because detectors may not be installed on unimportant roads. In this case, boundary flow of the network and maximum exiting flows of some links cannot be measured. In order to construct an arterial traffic model or design signal control plans, detectors are needed to provide necessary and accurate measurements.

### 3.7 SUMMARY

This chapter simulated arterial traffic using the link-node cell transmission model and the NGSIM arterial data. The link-node cell transmission model is an extension of the cell transmission model. It is consistent with the cell transmission model but has more flexibility in modeling arterial traffic. The simulation results showed that the link-node cell transmission model can simulate arterial traffic, and the error was comparable to the simulation results of freeway traffic simulation. Because arterial traffic was not very homogeneous and the effects of pedestrian crossing and right-turn on red were simplified in the simulation, we think the simulation error was reasonable. The model accuracy can be improved if such events are modeled in more detail. In this simulation study, demand and split ratios were updated at an interval of 10 seconds. Such data may not be available in reality and methods are needed to estimate the necessary input data to the simulation model.

## Chapter 4

## Estimation of Turning Proportions from

## Exit Counts in Arterial Traffic

### 4.1 INTRODUCTION

In Chapter 3, it was shown that turning proportions were inputs to the cell transmission model (CTM). In fact, real-time knowledge of turning volumes, or turning proportions, is a frequent requirement for any macroscopic simulation model and advanced intersection signal operation. This information is usually described by so-called origin-destination (OD) matrices. In the past decades, great attention was paid to estimating OD matrices from traffic counts collected by sensors, and many methods have been proposed [45].

Consider the following problem. Suppose detectors are placed at all the entrance and exit legs at an intersection to obtain vehicle counts, as shown by Fig. 4.1. Use $f_{1}$ to $f_{4}$ to denote the counts measured at the detectors at the entrance legs, and $D_{1}$ to $D_{4}$ to denote the counts at the exit legs. Denote $b_{1}$ to $b_{12}$ as the turning proportions of the flows entering an intersection, as indicated by Fig. 4.1. The problem of estimating turning proportions is to find $b_{1}$ to $b_{12}$ from full or partial measurements of the vehicle counts $f_{1}$ to $f_{4}$ and $D_{1}$ to $D_{4}$.

Suppose every vehicle passing an entrance detector must also pass an exit detector in the study period, Eq. (4.1) can be derived. Because each $b_{i}$ is a percentage, it must be non-negative. This is shown by constraint (4.2). In addition, because all the $b_{i}$ 's of the same leg must sum up to one, constraint (4.3) has to be satisfied. Both constraint (4.2) and constraint (4.3) restrict the $b_{i}$ to be between 0 and 1 , which is natural because each $b_{i}$ is a percentage. Constraints of the type of (4.2) is called the non-negative constraints, while constraints of the type of (4.3) is called the unity
constraints. In the problem formed by Eq. (4.1), constraints (4.2) and (4.3), variables $f_{i}$ and $D_{i}$ are measurements, and $b_{i}$ are turning proportions needed to estimate. There are 12 unknowns, 12 inequalities and 8 equalities. Thus, it is an under-determined problem and multiple solutions exist.

$$
\begin{gather*}
D=B \cdot F  \tag{4.1}\\
b_{i} \geq 0 \text { for } i=1, \ldots, 12  \tag{4.2}\\
\sum_{i} b_{i}=1 \text { for } i=\{1,2,9\} \text {, or } i=\{3,4,10\} \text {, or } i=\{5,6,11\} \text {, or } i=\{7,8,12\}  \tag{4.3}\\
\text { where } D=\left[D_{1}, D_{2}, D_{3}, D_{4}\right]^{T}, B=\left[\begin{array}{cccc}
0 & b_{4} & b_{11} & b_{7} \\
b_{1} & b_{10} & 0 & b_{8} \\
b_{2} & 0 & b_{5} & b_{12} \\
b_{9} & b_{3} & b_{6} & 0
\end{array}\right], \text { and } F=\left[f_{1}, f_{2}, f_{3}, f_{4}\right]^{T}
\end{gather*}
$$



Figure 4.1 Intersection with Entrance and Departure Detectors

To avoid under-determination, time series of data can be used to obtain additional equations. In this case, Eq. (4.1) will be changed to Eq. (4.4), where $k$ represents the time index of measurements. In this equation, there is an underlying assumption that turning proportions are constant. If the variations in vehicle counts are sufficiently rich to form enough linearly independent equations, a unique solution can be obtained.

$$
\begin{equation*}
D(k)=B \cdot F(k) \tag{4.4}
\end{equation*}
$$

However, because turning proportions might not be constant over time, and the flow measurements are noisy, it is more common for the problem formed by (4.2) to (4.4) to become an overdetermined problem rather than there existing a unique solution. For this reason, the estimation of
the matrix B will be posed as a constrained least squares problem. This means we find $b_{i}$ 's that minimize the objective function of (4.5). In this sense, the problem becomes a least squares problem with linear equality and inequality constraints. The inequality constraints give the boundary of the range the unknowns must fall in, they are also called the boundary constraints.

$$
\min \sum_{k}(D(k)-B \cdot F(k))^{T}(D(k)-B \cdot F(k))
$$

s.t.

$$
\begin{align*}
& b_{i} \geq 0 \text { for } i=1, \ldots, 12  \tag{4.5}\\
& \sum_{i} b_{i}=1 \text { for } i=\{1,2,9\}, \text { or } i=\{3,4,10\}, \text { or } i=\{5,6,11\}, \text { or } i=\{7,8,12\}
\end{align*}
$$

If the problem is formulated as a constrained least squares problem, it can be solved by a variety of constrained quadratic optimization algorithms. However, direct solutions of constrained least squares problems through optimization can be computationally expensive, and therefore inappropriate for real-time implementations. With this concern, researchers have proposed different recursive approaches. Early examples can be found in [46, 47]. Such recursive methods usually included two major steps. In the first step, an unconstrained least squares estimator was used, while in the second step, the parameter estimates were constrained using some correction methods, such as normalization and truncation. Because of the operation of normalization or truncation, there is no guarantee that the resulting estimates will be optimal. Motivated by this concern, Bell [48] proposed a simple algorithm for the least squares problem with inequality boundary constraints. Unlike the previous methods, constraints were enforced by an iteration process in this algorithm. The convergence of this algorithm was proven, and it was stated that the imposition of the inequality constraints in the update iteration could improve the accuracy. Based on this algorithm, Li et al. [49] proposed a recursive method that further incorporated both equality and boundary inequality constraints in the least squares update. This method was able to achieve good estimation with a much smaller computational effort, as compared to an optimization approach.

Even though recursive approaches have been shown to be computationally efficient and performed well, there are still two concerns about these methods. The first concern is that a complete configuration of detectors may not be available. The detector layout shown by Fig. 4.1 is called a complete configuration, because it has detectors at all legs of the intersection. In such a configuration, a typical four-way intersection requires at least eight detectors, whose price might be expensive. In order to make intersection detection more economical, the so called "Cordon" configuration, which groups several intersections and deletes one or more intermediate detectors, was proposed in [50, 51]. When an incomplete detector configuration is used, it is important to examine its possibility of still obtaining turning proportions estimates and to evaluate the reliability of the estimates. [50] examined the feasibility of estimating turning proportions from an incomplete set of traffic counts and gave the condition under which it is possible to uniquely obtain turning
proportion estimates from incomplete information obtained from the cordon configuration. However it was also mentioned in [50] that an increase in the estimates variability could happen when counts were incomplete, since fewer traffic counts were accessible. Subsequently, [51] suggested that the condition number of the Jacobian matrix, which was derived from predicted output counts with respect to the parameters, could be used to evaluate whether an incomplete information was sufficient to provide "good" estimates. In the numerical example given in that paper, it was shown that the Jacobian matrix condition number enlarged dramatically as detector saving grew (more detectors were removed from the complete configuration), which might lead to a greater potential of poor parameter estimation. Additionally, real-time estimation strategies for addressing incomplete detector configuration can be found in [52, 53, 54]. These references indicated that parameter estimation accuracy may degrade if a complete configuration was not available.

The second concern of some previously proposed methods is that they might fail to detect changes in estimated parameters when they are time varying. Most of the recursive least squares estimation methods were designed to estimate dynamic turning proportions in real time. To achieve this, time-series of data were used. However, most of these methods, including most of the methods mentioned above, were only evaluated with static turning proportions in the paper they presented. The performance of these methods was not shown for the case when turning proportions changed during simulation. It is well known that the recursive least squares estimator parameter adaptation gain decreases as time progresses. Thus, an ordinary least squares estimator may fail to follow parameter changes in a dynamic scenario. The techniques of forgetting factor and covariance resetting have been suggested by researchers for tracking time-varying parameters. Description of these techniques can be found in [55, 56].

In the remainder of this chapter, a new method for real-time estimation of turning proportions is presented. Compared with previous research, the algorithm proposed in this chapter only requires detectors to be installed at exit legs. Hence, it saves half the detectors of a complete detector configuration. However, real-time signal timing information is used to relate the portion of traffic volumes to its corresponding phase. The estimation problem is formulated as a constrained least squares problem, and then solved recursively. Therefore, it is called Recursive Constrained Least Squares (RCLS) here. An extended version of the RCLS algorithm is presented for the case of time-varying turning proportions. This version utilizes the forgetting factor and covariance resetting. The extended version is called Recursive Constrained Least Squares with Forgetting Factor and Covariance Resetting (RCLSFR).

The reminder of this chapter is organized as follows: Section 4.2 describes the mathematical formulation of the estimation problem. Section 4.3 explains the steps used to solve the problem. Section 4.4 discusses how to introduce the forgetting factor and covariance resetting in the parameter adaptation algorithm. Section 4.5 evaluates the algorithm using simulation. The conclusions are presented in Section 4.6.

### 4.2 PROBLEM FORMULATION



Figure 4.2 Intersection with Departure Detectors

Fig. 4.2 shows a four-way signalized intersection with detectors placed at the exit legs. The northbound and southbound arrival volumes are respectively denoted as $f_{1}$ and $f_{2}$. Let $b_{1}$ and $b_{2}$ respectively represent the northbound left-turn and through flow turning proportions, and $b_{3}$ and $b_{4}$ respectively represent the southbound left-turn and through flow turning proportions. As a consequence, the northbound and southbound right-turn percentages, denoted as $b_{9}$ and $b_{10}$, equal $1-b_{1}-b_{2}$ and $1-b_{3}-b_{4}$, respectively. Similar notations are used for the eastbound and westbound directions, as shown by the figure. Assuming that there is no right-turn on red, and the vehicle volume measurements are accurate, during the period that the northbound and the southbound traffic has green, the numbers of vehicles passing each detector are denoted $D_{1}$ to $D_{4}$. Following the notations in Fig. 4.2 and utilizing flow conservation, obtaining Eqs. (4.6) to (4.9) is
straightforward.

$$
\begin{align*}
f_{2} b_{4} & =D_{1}  \tag{4.6}\\
f_{1} b_{1}+f_{2}\left(1-b_{3}-b_{4}\right) & =D 2  \tag{4.7}\\
f_{1} b_{2} & =D_{3}  \tag{4.8}\\
f_{1}\left(1-b_{1}-b_{2}\right)+f_{2} b_{3} & =D 4 \tag{4.9}
\end{align*}
$$

In the period that the eastbound and the westbound traffic has green, another set of traffic counts should be collected. One can easily obtain the same equations as Eqs. (4.6) to (4.9) to represent the flow conservation on the East/West Bound, except that the subscripts will change. Therefore, the ideas discussed in the paragraphs below can be applied to both North/South Bound traffic and East/West Bound traffic, with the only difference being subscripts. Hence, in the discussion that follows, the North/South Bound traffic will be taken as representative and East/West Bound traffic will not be discussed further.

It should be notice that no assumption is made regarding the lane configuration of each intersection, or the type of signal control. Whether there is a designated lane for a particular movement, or whether pre-timed or actuated control is used, will not change Eqs. (4.6) to (4.9). However, Eqs. (4.6) to (4.9) are true only when north/south bound and east/west bound traffic do not receive green at the same time. This is the case most of the time because most of the intersections do not allow north/south bound and east/west bound traffic go simultaneously in order to prevent traffic conflict. Therefore, the turning proportion estimation method presented in this chapter, which is derived from Eqs. (4.6) to (4.9), can be used at most of the cases.

Using Eqs. (4.6) and (4.8) to substitute $f_{1}$ and $f_{2}$ in Eqs. (4.7) and (4.9), we obtain

$$
\begin{align*}
& D_{3} \frac{b_{1}}{b_{2}}+D_{1}\left(\frac{1}{b_{4}}-\frac{b_{3}}{b_{4}}-1\right)=D_{2}  \tag{4.10}\\
& D_{3}\left(\frac{1}{b_{2}}-\frac{b_{1}}{b_{2}}-1\right)+D_{1} \frac{b_{3}}{b_{4}}=D_{4} \tag{4.11}
\end{align*}
$$

Because all the $b_{i}$ 's must fall in the interval of $[0,1]$, we have the constraints

$$
\begin{array}{r}
\frac{1}{b_{2}}, \frac{1}{b_{4}} \geq 1 \\
\frac{1}{b_{1}} \geq \frac{b_{1}}{b_{2}} \geq 0 \tag{4.13}
\end{array}
$$

$$
\begin{equation*}
\frac{1}{b_{4}} \geq \frac{b_{3}}{b_{4}} \geq 0 \tag{4.14}
\end{equation*}
$$

In addition, because $b_{1}+b_{2} \leq 1$ and $b_{3}+b_{4} \leq 1$, there are another two constraints

$$
\begin{align*}
& \frac{b_{1}}{b_{1}}-\frac{1}{b_{2}} \leq-1  \tag{4.15}\\
& \frac{b_{3}}{b_{4}}-\frac{1}{b_{4}} \leq-1 \tag{4.16}
\end{align*}
$$

Constraints (4.12) - (4.16) can be rewritten as

$$
\begin{align*}
& \frac{1}{b_{2}}, \frac{1}{b_{4}} \geq 1  \tag{4.17}\\
& \frac{b_{1}}{b_{2}}, \frac{b_{3}}{b_{4}} \geq 0  \tag{4.18}\\
& \frac{1}{b_{2}}-\frac{b_{1}}{b_{1}} \geq 1  \tag{4.19}\\
& \frac{1}{b_{4}}-\frac{b_{3}}{b_{4}} \geq 1 \tag{4.20}
\end{align*}
$$

Eqs. 4.10, 4.11) and constraints (4.17) to 4.20 define the estimation of parameters $b_{1}$ to $b_{4}$ problem. There are four unknowns, two equations, and six inequalities. This is an underdetermined problem, and can be solved in the least squares sense.

Since (4.10), (4.11), and (4.17) to (4.20) are not in a good form to apply least squares estimation methods, they are reorganized into the matrix form as shown in Eqs. 4.21) and (4.22).

$$
\begin{align*}
{\left[\begin{array}{cccc}
0 & D_{3} & D_{1} & -D_{1} \\
D_{3} & -D_{3} & 0 & -D_{1}
\end{array}\right] } & {\left[\begin{array}{c}
\frac{1}{b_{2}}-1 \\
\frac{b_{1}}{b_{2}} \\
\frac{1}{b_{4}}-1 \\
\frac{b_{3}}{b_{4}}
\end{array}\right]=\left[\begin{array}{l}
D_{2} \\
D_{4}
\end{array}\right] }  \tag{4.21}\\
& {\left[\begin{array}{c}
\frac{1}{b_{2}}-1 \\
\frac{b_{1}}{b_{2}} \\
\frac{1}{b_{4}}-1 \\
\frac{b_{3}}{b_{4}}
\end{array}\right] \geq 0 } \tag{4.22}
\end{align*}
$$

Let $\beta=\left[\begin{array}{llll}\frac{1}{b_{2}}-1, & \frac{b_{1}}{b_{2}}, & \frac{1}{b_{4}}-1, & \frac{b_{3}}{b_{4}}\end{array}\right]^{T}, X=\left[\begin{array}{cccc}0 & D_{3} & D_{1} & -D_{1} \\ D_{3} & -D_{3} & 0 & -D_{1}\end{array}\right]$ and $Y=\left[\begin{array}{l}D_{2} \\ D_{4}\end{array}\right]$, the problem of (4.21) and (4.22) can be represented by

$$
\begin{equation*}
Y=X \beta \tag{4.23}
\end{equation*}
$$

Subject to

$$
\begin{equation*}
\beta \geq 0 \tag{4.24}
\end{equation*}
$$

In practice, time-series of data are used. Hence, it is necessary to add a time index in Eq. (4.23). In addition, sensor noise, vehicle traverse delay, and right-turn-on-red vehicles will make detector measurements not as ideal as in the assumption. The influences of these factors can be represented by an error term in the formulation. As a result, a more exact expression of Eq. 4.23) should be

$$
\begin{equation*}
Y(k)=X(k) \beta+v(k) \tag{4.25}
\end{equation*}
$$

Here, $k$ represents the time index and $v$ represents measurement error. To solve for $\beta$ in the least squares sense, one minimizes the object function (4.26).

$$
\begin{equation*}
\min _{\hat{\beta}} \sum_{k=1}^{N}(Y(k)-X(k) \hat{\beta})^{T}(Y(k)-X(k) \hat{\beta}) \tag{4.26}
\end{equation*}
$$

The hat on the variable $\beta$ denotes that it is an estimated value, and $N$ is the number of intervals of data. Eqs. (4.24) to (4.26) define a least squares problem with inequality boundary constraints. This can be solved by an algorithm proposed by Bell [48] and Li [49]. Notice that Constraints (4.15) and (4.16) are excluded in (4.22), because they will bring in linear constraints other than boundary constraints, and consequently make the solution of the problem more computationally difficult. Instead, the problem will be solved as follows: in the first step, an ordinary least squares estimator is used to solve the problem of (4.25) and (4.26); in the second step, Bell's correction developed in [48] will be applied if constraint (4.24) is not met; in the third step, $\hat{b}_{i}$ 's are obtained from $\hat{\beta}$. Within the third step, a normalization and/or truncation will be conducted if constraints (4.15) and 4.16) are not satisfied. Details of the solution procedure are described in the next section.

### 4.3 STEPS FOR SOLVING THE LEAST SQUARES

## PROBLEM

Step 0: Initialize $\hat{\beta}_{0}$ and $P_{0}$. Here, $\hat{\beta}_{0}$ is a $4 \times 1$ non-negative vector. It can be initialized via prior knowledge or random values. Moreover, $P_{0}$ is a $4 \times 4$ positive-definite matrix. It is usually selected to be the identity matrix.

Step 1: Update by an ordinary least squares estimator. After collecting new traffic counts, construct $X_{k}$ and $Y_{k}$, and recursively the estimate $\hat{\beta}_{k}$ by

$$
\begin{equation*}
\hat{\beta}_{k}=\hat{\beta}_{k-1}+K_{k}\left(Y_{k}-X_{k} \hat{\beta}_{k-1}\right) \tag{4.27}
\end{equation*}
$$

Where

$$
\begin{align*}
K_{k} & =P_{k-1} X_{k}^{T} S_{k}^{-1}  \tag{4.28}\\
S_{k} & =X_{k} P_{k-1} X_{k}^{T}+I  \tag{4.29}\\
P_{k} & =\left(I-K_{k} X_{k}\right) P_{k-1} \tag{4.30}
\end{align*}
$$

Here, $I$ is the $4 \times 4$ identity matrix.
Step 2: If any of the elements in $\hat{\beta}_{k}$ is negative, apply Bell's correction [48] to guarantee that the non-negative constraint is met. Otherwise, skip this step.
2.1 Let $\tilde{\beta}_{k}=\hat{\beta}_{k}$ and $\mu=0$.
2.2 Compute $D=\operatorname{diag}\left\{1 / p_{i i}\right\}$, where $P_{k}=\left[p_{i j}\right]_{4 \times 4}$.
2.3 Compute $\mu=T F\left(\mu-D \tilde{\beta}_{k}\right)$. Here, $T F(\bullet)$ is a truncation function which drives all negative values to zeros.
$2.4 \tilde{\beta}_{k}=\hat{\beta}_{k}+P_{k} \mu$.
Repeat 2.3 and 2.4 until $\tilde{\beta}_{k}$ converges. If step 2 is executed, $\hat{\beta}_{k}$ will be replaced by $\tilde{\beta}_{k}$.
Step 3: Solve for $\hat{b}_{i}$ s from $\hat{\beta}_{k}$. If any of the elements in $\hat{b}_{i}$ is negative or greater than one, respectively truncate it to zero or one. If $\hat{b}_{1}+\hat{b}_{2}>1$, normalize $\hat{b}_{1}$ and $\hat{b}_{2}$ so that $\hat{b}_{1}+\hat{b}_{2}=1$. Normalization is done by setting $\hat{b}_{1}=\frac{\hat{b}_{1}}{\hat{b}_{1}+\hat{b}_{2}}$ and $\hat{b}_{2}=\frac{\hat{b}_{2}}{\hat{b}_{1}+\hat{b}_{2}}$. Also apply normalization if $\hat{b}_{3}+\hat{b}_{4}>1$. Compute the turning proportions for right-turns $\hat{b}_{9}$ and $\hat{b}_{10}$ by $1-\hat{b}_{1}-\hat{b}_{2}$ and $1-\hat{b}_{3}-\hat{b}_{4}$, respectively.

The above steps, Steps 0 to 3, are the solving procedure in the RCLS algorithm. Steps 1 to 3 are repeated if new measurements are available.

### 4.4 INCORPORATING THE FORGETTING FACTOR AND

## COVARIANCE RESETTING

Because the matrix $S_{k}$ in Eq. (4.29) keeps growing, the gain ( $K_{k}$ in Eq. (4.27)) in an ordinary least squares estimator would become smaller and smaller as the number of steps grows. Therefore, an ordinary least squares estimator might fail to track time-varying turning proportions. If the turning proportions change during a studied period, the extended version of the RCLS algorithm, the RCLSFR algorithm, should be used. In the RCLS algorithm, the forgetting factor and covariance resetting proposed in [55, 56] are applied in the covariance update. In this case, Eq. (4.31] is used instead of (4.30) in Step 1.

$$
\begin{equation*}
P_{k}=\frac{1}{\lambda}\left(I-K_{k} X_{k}\right) P_{k-1}+\varepsilon I-\delta P_{k-1}^{2} \tag{4.31}
\end{equation*}
$$

In Eq. (4.31), the parameter $\lambda$ is the forgetting factor, and it usually falls in [0.9, 1]. The parameters $\varepsilon$ and $\delta$ are small positive values to adjust the covariance resetting process. They are usually in $[0,0.1]$.

It should be noted that, if Eq. (4.31) is used in Step 1, the matrix $P_{k}$ used in Step 2 should still be updated through Eqs. (4.28) to (4.30). As a result, two different sets of matrices $K_{k}, S_{k}$, and $P_{k}$ have to be maintained during the solution procedure.

### 4.5 SIMULATION AND RESULTS

In this section, the accuracy and computational efficiency of the proposed method will be evaluated by simulation. To do so, the estimated turning proportions computed via the proposed approach, either using the RCLS or RCLSFR algorithm, was compared with those obtained from an optimization approach. The optimization approach chosen in this comparison was the Constrained Least Squares by Active-Set algorithm (CLSAS) in Matlab 2013a. Two scenarios, one with static turning proportions and the other with time-varying proportions, were simulated. In each scenario, each tested method involved 10 simulation runs with different random seeds.

## Scenario I

In this scenario, static turning proportions were used to compare the RCLS and CLSAS algorithms. Consider a four-way intersection with the turning proportions shown in Table 4.1 (data source from [53]). Denote the turning proportions as $p_{i, j}$, where $i$ is the index of the entering leg while $j$ is the index of the exiting leg. The arrival volume at each entering leg of the intersection, denoted as $f_{i}$ following the notations in Fig. 4.2, is a random variable generated from a Poisson distribution
with mean 100 . Turning volumes, denoted as $t_{i, j}$, are multinomial random numbers generated from the arrival volumes $f_{i}$ and the turning proportions $p_{i, j}$ defined in Table 4.1. The vehicle volume measured at each exit leg is the sum of the vehicle volume entering that leg, denoted as $q_{i}$, and the measurement noise, denoted as $v_{i}$. The vehicle volume entering an exit leg is the sum of all turning volumes directing to that leg. Measurement noise at a detector is modeled as a zero mean Gaussian distribution with a standard deviation equal to $10 \%$ of the true vehicle volume. Hence, the traffic counts obtained at a detector $D_{i}$ are computed as

$$
\begin{equation*}
D_{i}(k)=q_{i}(k)+v_{i}(k) \tag{4.32}
\end{equation*}
$$

where $q_{i}=\sum_{i} t_{i, j}$ and $v_{i}(k)=N\left(0,0.1 q_{i}\right) . D_{i}$ will be rounded if it is not a non-negative integer.

Table 4.1 The Turning Proportions Used in the Simulation of Estimation Algorithm of Scenario I

|  | Left | Straight | Right |
| :---: | :---: | :---: | :---: |
| Northbound | 0.23 | 0.414 | 0.356 |
| Southbound | 0.29 | 0.352 | 0.358 |
| Eastbound | 0.149 | 0.8 | 0.051 |
| Westbound | 0.083 | 0.843 | 0.074 |

Traffic counts for 10 intervals were generated in the simulation. A constrained least squares problem defined by Eqs. (4.10), (4.11) and Constraints (4.17) to (4.20) was solved at every interval by the CLSAS optimization approach. This optimization took as input all the traffic counts generated from the beginning to the current interval. The RCLS method was also executed at every interval. Its inputs were the latest traffic counts and the estimated variables from the previous interval. The elements of the initial estimate $\hat{\beta}_{0}$ in the RCLS algorithm were set to random positive values. Results of the simulation are displayed in Table 4.2 .

To evaluate the accuracy of the estimation algorithms, the error between the estimated turning proportions and the real ones was computed at the end of each simulation run. Root-Mean-Square Deviation (RMSD), which is defined in Eq. (4.33), was chosen as the indicator of error. Turning proportions of all movements were included in the computation of RMSD. Because there were 12 turning proportions, the number of data points in the computation of RMSD at each simulation time step, which is the symbol $N$ in Eq. (4.33), was 12 . The RMSD shown in Table 4.2 was computed at the end of the simulation. The time for executing an algorithm on 10 intervals of data was also collected, in order to compare the computational efficiency. Furthermore, the frequencies of applying corrections (step 2 and step 3) was used to indicate the effort the RCLS algorithm makes
to satisfy the constraints (4.2) and (4.3) after its update by the least squares estimator (step 1). The terms of "steps requiring iteration", "number of iterations", and "steps requiring truncation" were computed in the simulation and display in the results. "Steps requiring iteration" was defines as the number of times when step 2 was applied in a simulation run. "Number of iterations" was the total number of iterations executed in a run, which was, how many number of times was step 2.3 needed to correct the elements in $\hat{\beta}_{k}$ from negative to positive. Remember steps 2.3 and 2.4 are repeated until $\hat{\beta}_{k}$ converges. A large number of iterations means the estimation algorithm is computational inefficient. "Steps requiring truncation" computed the number of times when any truncation or normalization was applied in step 3. Remember that the estimates are not optimal any more when truncation or normalization is applied. Thus, the ideal value of "Steps requiring truncation" is zero.

$$
\begin{equation*}
R M S D=\sqrt{\frac{\sum_{i=1}^{N}\left(b_{i, \text { estimated }}-b_{i, \text { real }}\right)^{2}}{N}} \tag{4.33}
\end{equation*}
$$

Table 4.2 gives the performance comparison between the CLSAS and RCLS algorithms, while Fig. 4.3 plots the RMSD curves in the first four experiments as representatives. It can be seen from Table 4.2 that the RMSD values in the CLSAS and RCLS algorithms were very close. In 6 out of 10 experiments, the RMSD values in the RCLS algorithm were the same as in the CLSAS algorithm, and in the remaining 4 experiments, the RMSD values produced by the RCLS algorithm were only different from those produced by the CLSAS algorithm in the fourth digit after decimal point. It is obvious from Fig. 4.3 that, after the first few steps, the RMSD curve of the RCLS algorithm almost overlapped with that of the CLSAS algorithm in each experiment. Notice that, in contrast, the execution time of the RCLS algorithm was only about $3 \%$ of that of the CLSAS algorithm. This means that the RCLS algorithm is as effective an estimator as the CLSAS algorithm, but is significantly more computationally efficient. It is also important to remark that, on average, the iteration in step 2 was only executed 0.9 times out of 20 steps (two directions, and 10 intervals of data for each direction), and only 3.3 iterations are applied in one experiment. Truncation or normalization was conducted 2.4 times out of 20 steps. All of these values indicate that the ordinary least squares estimator in the RCLS algorithm estimates the turning proportions while meeting the constraints most of the time, and that the iteration procedure in its algorithm that are necessary to maintain the estimates within the constraints are executed infrequently. Therefore they do not detrimentally affect real-time efficiency.

Tables 4.3 and 4.4 provide the estimated turning proportions at the last interval by the CLSAS and RCLS algorithms, respectively. The proportional values are the averages over 10 simulation runs. The percentages in parentheses are the error percentages taken with respect to the real turning proportions in Table 4.1. Notice that the values in Tables 4.3 and 4.4 are identical. This also supports the observation that the RCLS algorithm generates for the most part the same estimates as the CLSAS algorithm after a few iterations. From the percentages in Table 4.4, it can be seen that the estimated values of all movements were very close to the real ones after the estimation had settled, as already indicated by the small RMSD values.

Table 4.2 Performance Comparison between the CLSAS and RCLS Algorithms

|  | CLSAS |  | RCLS |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\begin{aligned} & 2 \pi \\ & 2 \\ & 0 \end{aligned}$ |  | $\begin{aligned} & \pi \\ & 3 \\ & 0 \end{aligned}$ |  |  | 7 2 0 0 0 0 0 0 0 0 0 0 0 |  |
| 1 | 0.0083 | 0.0919 | 0.0083 | 0.0032 | 1 | 2 | 3 |
| 2 | 0.0359 | 0.0938 | 0.0358 | 0.0032 | 1 | 2 | 1 |
| 3 | 0.0247 | 0.1438 | 0.0247 | 0.0022 | 0 | 0 | 3 |
| 4 | 0.0340 | 0.0989 | 0.0338 | 0.0027 | 2 | 4 | 3 |
| 5 | 0.0090 | 0.0962 | 0.0091 | 0.0031 | 2 | 19 | 5 |
| 6 | 0.0142 | 0.0982 | 0.0142 | 0.0025 | 1 | 2 | 1 |
| 7 | 0.0235 | 0.0893 | 0.0235 | 0.0027 | 0 | 0 | 1 |
| 8 | 0.0132 | 0.0967 | 0.0134 | 0.0022 | 1 | 2 | 2 |
| 9 | 0.0155 | 0.0904 | 0.0154 | 0.0022 | 1 | 2 | 2 |
| 10 | 0.0254 | 0.0963 | 0.0254 | 0.0031 | 0 | 0 | 3 |
| Mean | 0.0203 | 0.0996 | 0.0203 | 0.0027 | 0.9 | 3.3 | 2.4 |



Figure 4.3 The RMSD Values of the RCLS and CLSAS Algorithms in Experiments \#1 to \#4 in Scenario I as Representatives

Table 4.3 Estimated Turning Proportions Produced by the CLSAS Algorithm in Scenario I and the Error Percentages

|  | Left | Straight | Right |
| :---: | :---: | :---: | :---: |
| Northbound | $0.230(-0.12 \%)$ | $0.418(1.07 \%)$ | $0.352(-1.17 \%)$ |
| Southbound | $0.286(-1.39 \%)$ | $0.350(-0.54 \%)$ | $0.364(1.66 \%)$ |
| Eastbound | $0.146(-1.95 \%)$ | $0.799(-0.08 \%)$ | $0.055(7.00 \%)$ |
| Westbound | $0.083(0.22 \%)$ | $0.849(0.75 \%)$ | $0.068(-8.77 \%)$ |

Table 4.4 Estimated Turning Proportions Produced by the RCLS Algorithm in Scenario I and the Error Percentages

|  | Left | Straight | Right |
| :---: | :---: | :---: | :---: |
| Northbound | $0.230(-0.12 \%)$ | $0.418(1.07 \%)$ | $0.352(-1.17 \%)$ |
| Southbound | $0.286(-1.39 \%)$ | $0.350(-0.54 \%)$ | $0.364(1.66 \%)$ |
| Eastbound | $0.146(-1.95 \%)$ | $0.799(-0.08 \%)$ | $0.055(7.00 \%)$ |
| Westbound | $0.083(0.22 \%)$ | $0.849(0.75 \%)$ | $0.068(-8.77 \%)$ |

## Scenario II

In this simulation scenario, the RCLS, RCLSFR, and CLSAS turning proportions estimation algorithms were compared by estimating time-varying turning proportions. The following simulation was conducted for the comparison: consider the same intersection used in Scenario I. The initial turning proportions were defined as in Table 4.1 and kept unchanged for 20 intervals. Subsequently, the turning proportions defined in Table 4.5 were used, and these were kept unchanged for another 20 intervals. The process mentioned in Scenario I was used to generate arrival volumes, exit volumes, and detector counts. The RCLS, RCLSFR, and CLSAS algorithms were tested in the same manner as in Scenario I. However, an adjustment had to be made to test the CLSAS algorithm in this scenario. This was due to the fact that, if all the traffic counts generated from the beginning to current interval were entered into the CLSAS optimization algorithm, as had been done in the previous scenario, the CLSAS algorithm would compute the aggregated turning proportions in the
second period, rather than the new ratios. On the other hand, using only the latest traffic counts was not a good idea either, because the estimates will be significantly affected by randomness or noise. With such concerns, how many pieces of data should be fed into the optimization had to be decided. From some empirical observations, it was found that using the data generated in the latest eight time intervals produced the best estimate in this test. As a result in the experiments of this part, at each step when we computed the turning proportions via the CLSAS algorithm, we only took the data generated in latest eight time intervals. The parameters $\lambda, \varepsilon$, and $\delta$ used in the RCLSFR algorithm also needed to be specified. After some tests, $\lambda=0.995, \varepsilon=0.0005$ and $\delta=0.0005$ were chosen because they gave the best estimates in the simulation test.

Table 4.5 Actual Turning Proportions Used in the Simulation of Estimation Algorithms in the
Second Period (The turning Proportions Used in the First Period are given in Table 4.1

|  | Left | Straight | Right |
| :---: | :---: | :---: | :---: |
| Northbound | 0.32 | 0.544 | 0.136 |
| Southbound | 0.16 | 0.471 | 0.369 |
| Eastbound | 0.33 | 0.54 | 0.13 |
| Westbound | 0.2 | 0.5 | 0.3 |

The same performance variables used in Table 4.2 were also used in this part of simulation. Table 4.6 shows the performance of the CLSAS, RCLS, and RCLSFR algorithms in this scenario, while Fig. 4.4 plots the RMSD curves in the first four experiments of this scenario as representatives. It should be remarked that the over RMSD in the RCLS algorithm was about three times larger than that in the CLSAS. Moreover, from Fig. 4.4, it can also be seen that, even though the RCLS algorithm provided good turning proportions estimates in the first period, its RMSD jumped to a very high value after the turning proportions changed and dropped very slowly afterwards. The RMSD values in the table and the pattern of the RMSD curve in the figure indicate that the RCLS algorithm can hardly detect the change in turning proportions that took place after 20 intervals. On the other hand, the RMSD value of the RCLSFR algorithm was much smaller than that of the RCLS algorithm in the second period (after turning proportions changed), yet it was still larger than the RMSD value of the CLSAS algorithm. It is shown by Fig. 4.4 that the RMSD curve of the RCLSFR algorithm decreased faster than that of CLSAS after the turning proportions changed, but fluctuated at a later point and ended up with a larger value than the CLSAS, which means the estimated turning proportions produced by the RCLSFR algorithm had larger error that those produced by the CLSAS algorithm. The pattern of the RMSD curves demonstrates that the RCLSFR algorithm is able to quickly detect changes in the turning proportions, but it has higher fluctuations when the turning proportions are constant. The computational time of the RCLSFR algorithm was
about 3\% that of the CLSAS algorithm, and only slightly larger than that of the RCLS algorithm. The values for "steps requiring iteration", "number of iterations" and "steps requiring truncation" were 1.1, 3.4, and 3.2, respectively. As indicated by these values, the RCLSFR algorithm does not require much computational effort to satisfy the constraints after its update from the least squares estimator. Given its fast response to change and small computational effort, the RCLSFR algorithm appears to be a competitive algorithm as compared to the CLSAS algorithm, even accounting for parameter estimates fluctuations during the period when the turning parameters are constant.

Tables 4.7 to 4.9 present the estimated turning proportions produced by the CLSAS, RCLS and RCLSFR algorithms at the end of the simulation. The numbers in these tables were obtained in the same way as those in Table 4.4. The results shown by these tables again support the observation that the CLSAS algorithm produces the closest estimates, followed by the RCLSFR algorithm. The estimation error of the RCLS algorithm is large, due to the fact that it cannot track time varying turning proportions.
Table 4．6 Performance Comparison of the CLSAS，RCLS and RCLSFR Algorithms in Simulation Scenario II

| $$ | Steps requiring truncation | m | $\bigcirc$ | $m$ | $\bigcirc$ | － | － | $m$ | $\sim$ | $\sim$ | $\sim$ | $\stackrel{N}{\text { N }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Number of iterations | 0 | 三 | $\sim$ | $\checkmark$ | $\checkmark$ | 0 | 0 | $\checkmark$ | 0 | 0 | $\stackrel{+}{\dot{m}}$ |
|  | Steps requiring iteration | $\bigcirc$ | in | － | $\sim$ | $\sim$ | $\bigcirc$ | $\bigcirc$ | $\checkmark$ | $\bigcirc$ | $\bigcirc$ | $亏$ |
|  | Execution time | $\begin{aligned} & \text { No } \\ & \text { ò } \\ & \hline 0 \end{aligned}$ | $\stackrel{\underset{\sim}{J}}{\underset{O}{\circ}}$ | $\begin{aligned} & 0 \\ & 0 \\ & 0 \end{aligned}$ | $\begin{aligned} & \Xi \\ & \vdots \\ & 0 \end{aligned}$ | $\begin{aligned} & \hat{O} \\ & \text { O} \\ & \hline 0 \end{aligned}$ | $\begin{aligned} & \infty \\ & \stackrel{\infty}{\circ} \\ & 0 \\ & \hline \end{aligned}$ | $\stackrel{\Im}{O}$ | $\begin{aligned} & 8 \\ & 0 \\ & 0 \\ & 0 \end{aligned}$ | $\begin{aligned} & 8 \\ & 0 \\ & 0 \\ & 0 \end{aligned}$ | $\begin{aligned} & \text { N} \\ & 0 \\ & 0 \end{aligned}$ | $\stackrel{N}{3}$ |
|  | RMSD | $\begin{aligned} & \stackrel{\rightharpoonup}{\mathrm{a}} \\ & \text {. } \end{aligned}$ | $\begin{aligned} & \text { in } \\ & 0 \\ & 0 \end{aligned}$ | $\begin{aligned} & 8.8 \\ & \stackrel{\circ}{\circ} \\ & \hline \end{aligned}$ | $\begin{gathered} \text { N} \\ \\ \hline 0 \end{gathered}$ | $\stackrel{O}{0}$ | $\begin{aligned} & \text { B} \\ & 0 \\ & 0 \\ & 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \text { O. } \\ & 0 . \end{aligned}$ | $\begin{aligned} & \text { N} \\ & \text { O. } \\ & 0 \end{aligned}$ | $\begin{gathered} \sigma \\ 0 \\ 0 \end{gathered}$ | $\begin{aligned} & \hat{N} \\ & \underset{o}{0} \end{aligned}$ | ¢ |
| ${\underset{\sim}{0}}_{0}^{2}$ | Steps requiring truncation | $\bigcirc$ | $m$ | $\sim$ | $\sim$ | $\sim$ | － | － | $\cdots$ | － | － | $\stackrel{+}{-}$ |
|  | Number of iterations | $\bigcirc$ | $\cdots$ | － | $\sim$ | $\bigcirc$ | $\bigcirc$ | 0 | $\wedge$ | $\sim$ | $\bigcirc$ | $\stackrel{\infty}{\text { i }}$ |
|  | Steps requiring iteration | 0 | $\sim$ | $\sim$ | － | $\bigcirc$ | $\bigcirc$ | 0 | － | － | $\bigcirc$ | $\bigcirc$ |
|  | Execution time | $$ | $\begin{aligned} & \hat{\infty} \\ & \substack{0 \\ 0 \\ \hline} \end{aligned}$ | $\begin{aligned} & 0 \\ & \text { 合. } \\ & 0 . \end{aligned}$ | $\begin{aligned} & \infty \\ & \stackrel{N}{8} \\ & 0 \end{aligned}$ | $\stackrel{\infty}{\stackrel{\infty}{8} .}$ | $\begin{aligned} & 20 \\ & 0 \\ & 0 \\ & 0 \end{aligned}$ | $$ | $\begin{aligned} & \infty \\ & \hline 8 . \\ & \hline 0 \end{aligned}$ | $$ | $\begin{aligned} & 8 . \\ & 8 . \\ & \hline 0 \end{aligned}$ | 「 |
|  | RMSD | $\stackrel{N}{n} \underset{0}{0}$ | $\begin{aligned} & \underset{\sim}{0} \\ & \underset{0}{2} \end{aligned}$ | $\begin{aligned} & \infty \\ & \underset{0}{0} \\ & \hline \end{aligned}$ | $\frac{0}{3}$ | $\stackrel{N}{\infty} \underset{\substack{0}}{\infty}$ | $\frac{2}{3}$ | $\begin{aligned} & \text { N } \\ & \underset{\sim}{\mathrm{T}} \end{aligned}$ | $\stackrel{t}{0}$ | $\begin{aligned} & \text { ò } \\ & \text { ob } \end{aligned}$ | $\frac{0}{8}$ | $\stackrel{\circ}{\square}$ |
| $\begin{aligned} & \text { è } \\ & \frac{1}{3} \end{aligned}$ | Execution time | $\begin{aligned} & \overline{\mathrm{N}} \\ & \stackrel{\rightharpoonup}{\circ} \end{aligned}$ | $\stackrel{N}{\underset{\sim}{i}}$ | $\begin{aligned} & \stackrel{\rightharpoonup}{6} \\ & \text { m} \end{aligned}$ | $\begin{aligned} & \text { n } \\ & \underset{\sim}{0} \\ & \vdots \end{aligned}$ | $\begin{aligned} & \dot{\infty} \\ & \stackrel{y}{\circ} \\ & \vdots \end{aligned}$ | $\stackrel{\infty}{\underset{0}{3}}$ | $\begin{aligned} & \text { n} \\ & \text { N} \\ & \text { N} \end{aligned}$ | $\begin{aligned} & \text { à } \\ & \text { è } \end{aligned}$ | $\begin{gathered} \infty \\ \underset{0}{0} \\ \vdots \end{gathered}$ | $\begin{gathered} \text { Y } \\ \text { Nu } \end{gathered}$ | $\cdots$ |
|  | RMSD | $\begin{aligned} & \text { ơ } \\ & \text { c. } \\ & 0 . \end{aligned}$ | $\stackrel{\rightharpoonup}{\mathrm{i}}$ | $\stackrel{N}{\infty}_{\infty}^{\infty}$ | $\underset{\substack{\text { O} \\ \hline \multirow{2}{c}{\hline}\\ \hline}}{\text {. }}$ | $\begin{aligned} & \text { 칭 } \\ & \stackrel{0}{0} \end{aligned}$ | $\underset{\text { 중 }}{2}$ | $\begin{aligned} & \text { n } \\ & \text { on } \\ & \text { on } \end{aligned}$ | $$ | $\begin{aligned} & 0 \\ & 0 \\ & 0 \\ & 0 \end{aligned}$ | $\begin{aligned} & \overrightarrow{3} \\ & 0 . \end{aligned}$ | $\stackrel{2}{2}$ |
|  | Experiment No． | － | $\sim$ | $m$ | － | in | $\bigcirc$ | $\checkmark$ | $\infty$ | の | $\bigcirc$ | 滣 |



Figure 4.4 Plots of Estimated Turning Proportions in Experiments \#1 to \#4 in Scenario II

Table 4.7 Estimated Turning Proportions Produced by the CLSAS Algorithm in Scenario II and the Error Percentages

|  | Left | Straight | Right |
| :---: | :---: | :---: | :---: |
| Northbound | $0.320(-0.00 \%)$ | $0.542(-0.33 \%)$ | $0.138(1.32 \%)$ |
| Southbound | $0.162(0.95 \%)$ | $0.486(3.11 \%)$ | $0.353(-4.38 \%)$ |
| Eastbound | $0.327(-1.05 \%)$ | $0.542(0.28 \%)$ | $0.132(1.50 \%)$ |
| Westbound | $0.205(2.41 \%)$ | $0.505(0.95 \%)$ | $0.290(-3.19 \%)$ |

Table 4.8 Estimated Turning Proportions Produced by the RCLS Algorithm in Scenario II and the Error Percentages

|  | Left | Straight | Right |
| :---: | :---: | :---: | :---: |
| Northbound | $0.280(-12.58 \%)$ | $0.479(-11.89 \%)$ | $0.241(77.15 \%)$ |
| Southbound | $0.218(36.21 \%)$ | $0.425(-9.87 \%)$ | $0.358(-3.11 \%)$ |
| Eastbound | $0.251(-24.03 \%)$ | $0.647(19.87 \%)$ | $0.102(-21.55 \%)$ |
| Westbound | $0.115(-42.71 \%)$ | $0.753(50.50 \%)$ | $0.133(-55.70 \%)$ |

Table 4.9 Estimated Turning Proportions Produced by the RCLSFR Algorithm in Scenario II and the Error Percentages

|  | Left | Straight | Right |
| :---: | :---: | :---: | :---: |
| Northbound | $0.333(4.07 \%)$ | $0.541(-0.64 \%)$ | $0.126(-7.02 \%)$ |
| Southbound | $0.151(-5.89 \%)$ | $0.476(1.09 \%)$ | $0.373(1.16 \%)$ |
| Eastbound | $0.331(0.27 \%)$ | $0.551(2.00 \%)$ | $0.118(-9.02 \%)$ |
| Westbound | $0.217(8.36 \%)$ | $0.524(4.73 \%)$ | $0.260(-13.45 \%)$ |

## Remark

From the simulation results above, it can be seen that, by choosing the proper update algorithm for time invariant and time varying turning proportion case, the estimation error of the proposed estimation algorithm was small while the computation was fast. The beauty of the proposed estimation method is that accurate estimates can be obtained with fewer detectors compared to the previously used complete detector configuration. The saving of the detectors is achieved by using the intersection signal timing information to separate the vehicle volumes into different movement groups. To do this, it requires the sensors have the capability to measure vehicle volumes in flexible time intervals. This is because the green lengths for different movements are not the same at most of the signalized intersections, and the time interval of detection needs to be adjusted accordingly. For instance, if the green lengths are time varying, as at actuated controlled intersections, the detectors should be able to measure in whatever time intervals the signal controllers determine. It should be noted that current sensors installed at signalized intersections only provide measurements at fix time intervals, say like 30 seconds. At some intersections it is even worse that sensors can only provide vehicle presence rather than vehicle volumes, as observed in most queue detectors or detectors utilized in actuated control. In the case of measurement interval is fix, previous estimation algorithms using complete detector configuration can apply, but the proposed algorithm cannot. In the case only vehicle presence is available, neither previous algorithms nor the proposed algorithm can be used.

### 4.6 CONCLUSION

Recursive methods, namely the RCLS and RCLSFR algorithms, were proposed in this chapter to estimate the traffic flow turning proportions at signalized arterial intersections in real time. By utilizing real-time signal timing information, these methods only require exit counts, consequently saving half of the detectors used in a complete detector configuration. The estimation problem was posed in the form of a constrained least squares estimation problem, and a recursive algorithm was proposed to solve it. Simulation results showed that the proposed methods estimated the turning proportions as accurately as the CLSAS optimization approach, while only requiring $3 \%$ of the CLSAS computation effort. The extended RCLSFR algorithm, which incorporates forgetting factor and covariance resetting, was suggested for estimating time-varying turning proportions. It was found that it was able to respond to changes in turning proportions very quickly. However, although it exhibited a tolerable larger parameter error variance than the CLSAS algorithm when the turning proportions were not constant, it required significantly less computational effort.

## Chapter 5

## Variable Speed Limits and Coordinated

## Ramp Metering

Traffic congestion appears along freeways in many metropolitan areas. Traffic control is commonly viewed by traffic engineers as an effective means to relieve congestion. In this chapter, we present a traffic control method that combines ramp metering and variable speed limits to address freeway congestion. A calibrated microscopic simulation is used to evaluate the proposed control.

### 5.1 INTRODUCTION

## Ramp Metering and Variable Speed Limits

## Ramp Metering

Ramp metering [57] is a freeway traffic control that regulates the traffic flow entering freeways. Traffic lights are placed at the ramp entrances to provide signals that usually change every a few seconds in order to regulate the traffic flow entering the freeway. During rush hour, traffic lights switch between red and green, and drivers approaching a light are required to stop and wait for a green signal before they can enter the freeway. In practice, one or two cars are allowed to pass per green. Therefore, the flow of vehicles entering the freeway can be controlled by the frequency of the green signal, which is called the metering rate.

The metering rate can be set according to a time-of-day plan, or be traffic responsive and
dynamically change based on real-time traffic conditions. The former method uses a fixed rate regardless of traffic conditions, while the latter can adapt to traffic variations. Therefore, traffic responsive ramp metering is more commonly used at present. In a traffic responsive deployment, detectors need to be installed on the freeway mainline, usually a short distance upstream or downstream from an on-ramp, to provide real-time vehicle occupancy measurements. These detectors are called mainline detectors. Detectors are also placed along on-ramps; depending on their locations, these are called demand detectors, passage detectors, or queue detectors, as illustrated by Fig. 5.1. When a vehicle stops on a demand detector on a particular lane, a message is sent to the controller to request a green signal for that lane. The controller determines the time to turn on the green light based on the configured metering rate. When this vehicle passes the passage detector behind the stop line, another signal is sent to notify the ramp metering controller that the vehicle has left and the traffic light can safely change back to red. Queue detectors are placed near the trail of an on-ramp to detect whether the on-ramp queue reaches the on-ramp storage limit and will spill back to the neighboring arterial streets.


Figure 5.1 Traffic Light and Detectors Commonly Used in the Traffic Responsive Ramp Metering

It has been demonstrated that ramp metering, if appropriately operated, can reduce congestion and accidents near the merge area. Without ramp metering, it is often observed that a platoon of vehicles tries to merge simultaneously, and, as a consequence, vehicles on the freeway mainline have to slow down to let them enter. Since traffic at the freeway mainline is often of high density during rush hours, these slower speeds will quickly propagate upstream and bring speed disruptions
to traffic flow. With ramp metering, only one or two vehicles merge at a time. As a consequence, merging is smoother and less likely to cause disruptions. This means that the freeway mainline can be operated at a higher traffic density flow and faster speed, and consequently, travel delay can be reduced. Multiple states in the U.S. and countries worldwide have reported that traffic can benefit by deploying ramp metering [58, 59]. For instance, Kansas City [60] implemented ramp metering on seven interchanges along a five-mile corridor of the I-435 in 2010, and an evaluation covering 12 months after implementation indicated that ramp metering decreased system-wide accidents by $64 \%$, and accidents that could likely be attributed to merging by $81 \%$. The Minnesota Department of Transportation (DOT) also evaluated ramp metering in 2001 [61]. It was found that, during six weeks of deactivating all ramp meters in Twin Cities, freeway throughput decreased by 9\%, travel time increased by $22 \%$, and crashes increased by $26 \%$. This led to an approximate annual savings of 40 million dollars produced by ramp metering.

Ramp metering can also have detrimental effects. Drivers at on-ramps must queue and wait for the green signal. If the traffic demand for an on-ramp is large and the metering rate is inappropriately low, it is likely that the on-ramp queue will spill back and block arterial intersections upstream. Therefore, it is of great importance to balance the gain obtained on the freeway mainline with the detrimental impacts produced on on-ramps and surface streets in ramp metering applications.

The first ramp metering in the United States was implemented in 1963 in Chicago. Up to 2002, there were 2,100 ramp meters deployed in 29 metropolitan areas within the United States. Ramp metering algorithms in the early years used pre-timed or time-of-day strategies, but most of them have become traffic-responsive in recent decades [57, 62]. Pre-timed algorithms are designed based on historical data and require periodic updates. On the other hand, traffic-responsive algorithms require real-time detection. They cost more than pre-timed algorithms, but usually generate better performance. Ramp metering algorithms can be classified into two types, namely local or coordinated. A local ramp metering algorithm is also a decentralized algorithm. It regulates the metering rate based on the measurements collected in the vicinity of an individual ramp. There is no communication between upstream and downstream ramps. Since the operation decision is made independently of other controllers, it is incapable of responding to downstream congestion that will possibly propagate, or planning ahead for a large arrival flow from upstream. A coordinated algorithm, on the other hand, considers system-wide traffic conditions and coordinates the actions for multiple metered ramps. Hence, it is more flexible than a local algorithm and it can address problems in which multiple ramps are involved.

Over the years, a variety of ramp metering algorithms have been proposed [57, 63]. Some have been tested or deployed in the field. Examples of these include ALINEA, ZONE, BOTTLENECK, SWARM, and HERO. ALINEA (Asservissement Linaire d'Entre Autoroutire) is a local feedback ramp metering algorithm that has been deployed in several European countries [64]. It maximizes freeway throughput by an integral feedback control action, which drives the traffic state toward a target value, usually chosen to be critical occupancy. ZONE [65] is a coordinated algorithm used in Minnesota. In this algorithm, a freeway corridor is divided into multiple zones. Each zone can
comprise multiple off-ramps, on-ramps, and freeway-to-freeway connectors, depending on the geometry and traffic condition of the freeway. The total metering rate of a zone is chosen for the sake of balancing traffic volumes entering and exiting a metering zone, and then it is distributed to each metered on-ramp or freeway-to-freeway connector according to pre-defined factors. BOTTLENECK [66] is a coordinated algorithm adopted in Seattle. Under normal conditions, it computes a local metering rate according to the measured mainline occupancy and predefined occupancymetering rate relationship. When the mainline occupancy exceeds a threshold and vehicles are found to be continuously stored in a section, bottleneck rates will be calculated. A bottleneck rate is determined by the traffic volume on a downstream segment of the ramp. Since an on-ramp can be linked to multiple downstream segments, there could be multiple bottleneck rates computed for a single on-ramp. The BOTTLENECK algorithm chooses the most restrictive meter rate from the local metering rate and the bottleneck rates. SWARM [67] stands for "system-wide adaptive ramp metering." As its name suggests, it is a system-wide (coordinated) metering algorithm, and it was tested in Los Angeles area. Similar to BOTTLENECK, SWARM computes a local metering rate and a system-wide decision rate, and chooses the more restrictive rate to apply. Unlike BOTTLENECK, SWARM predicts bottlenecks based on recorded data, rather than identifying them from measured data. The accuracy of the prediction largely influences the performance of this algorithm. HERO (HEuristic Ramp metering coOrdination) is a coordinated algorithm that extends ALINEA; it has been successfully applied in Australia [68]. In this algorithm, each ramp computes a local ramp metering rate according to the ALINEA algorithm. The coordination among ramps is called when mainline bottlenecks are identified and there is any on-ramp having a queue that potentially reaches its limit. In this case, the upstream ramps of the concerned ramp will be recruited as slave ramps, and the ramp metering rate will be adjusted to increase storage.

Besides the ramp metering algorithms that have been evaluated in the field, a number of proposed algorithms still await field testing. Among the published coordinated ramp metering methods, most are based on certain traffic dynamic models and optimization techniques. These methods usually assume that traffic demand is known or can be accurately predicted. They feed the dynamic model with real-time detection and future demand, and search for control actions by optimizing a cost function, often chosen to be travel delay. Examples can be found in [69, 70, 71]. In [69], the Asymmetric Cell Transmission Model was used to represent traffic dynamics, and the metering rate was computed by minimizing an objective function, which was a linear combination of total travel time and total travel distance. Although this optimization was nonlinear, a near-global optimal solution can be obtained by linear programming when certain conditions are met. In [70], a second-order model was adopted and nonlinear optimization with the total time spent as the cost function was used to obtain the metering rate. In [71], a second-order model was again used and a model predictive control was formulated for the ramp metering problem.

## Variable Speed Limits

Variable Speed Limits (VSL) is a control strategy enforced on the freeway mainline to regulate traffic speeds. Overhead or roadside electronic signs, called variable message signs (VMS), are
placed to display speed limits that drivers should travel at (Fig. 5.2). Speed limits are dynamically determined based on road, traffic, and weather conditions. In some applications, the posted speed limits are advisory, which means it is optional to follow them, while in other applications, speed limits are mandatory and drivers must comply with them just like static speed limit signs. Displayed speed limits can be uniform across lanes, or different on general lanes and special lanes, such as HOV and HOT lanes. Variable speed limits are often used as part of incident management, congestion management, and weather advisory in Intelligent Transportation System (ITS).


Figure 5.2 Displaying Variable Speed Limits Using Variable Message Signs

VSL have been deployed for more than 40 years, and are used in parts of Europe, Australia, and the United States [72]. Primarily, VSL systems are used to improve traffic safety, because they can enforce a lower speed in bad weather conditions, or notify drivers about speed changes ahead to avoid sudden braking. In many VSL operations, video surveillance is used to monitor road traffic conditions, and traffic engineers adjust speed limits based on their priori knowledge. In recent years, traffic sensors and environmental sensors have been utilized and speed limits can be automatically or semi-automatically selected in some deployments [73].

Although VSL have been applied in many cities, well-documented reports about the impacts of VSL on traffic flow are very limited. The Michigan DOT [74] deployed a prototype VSL system on a segment of the I-96 in 2002, and it was found that after the system was activated, average speed increased in the work zone. England's Highways Agency implemented mandatory VSL and dynamic use of the shoulder lane in 2006 [75]. An evaluation of safety using 36 months of data after this system was introduced indicated that the number of Personal Injury Accidents was reduced by $55.7 \%$, and crashes were less severe. The Washington State DOT [76] installed 15 sign bridges along a 7.5 mile stretch of the I-5 to display dynamic message signs. Data collected over 3.5 months, although a bit limited, showed that collisions were reduced by $65-75 \%$. The

Missouri DOT [77] installed VSL signs in 2008 on the I-270/I-255 around St. Louis. One year after deployment, the data revealed a 4.5 to $8 \%$ crash reduction. However, it was also found that only limited benefit of mobility was gained in some segments in this evaluation. The desired improvement of system-wide mobility was not achieved.

Although only limited reports on the VSL impacts have been presented, researchers and engineers still believe that VSL has the potential to improve traffic conditions, not only by enhancing safety but also by benefiting mobility. Different approaches, including simulation models and real-time data, have been utilized to show how VSL can benefit traffic flow, and various control strategies have been proposed for a realistic and beneficial implementation. In [78], a simulation was constructed to evaluate VSL's potential for crash reduction. The results showed that VSL was beneficial in medium-to-high-speed traffic, and based on this observation, it gave some recommendations for the VSL implementation. In [79], real-time data were collected on a European motorway to analyze the impact of VSL on the fundamental diagram. It was found that VSL would change the curve of the fundamental diagram by shifting the critical occupancy to higher values. When VSL was applied in the condition where occupancy was lower than the critical occupancy, it decreased the slope in the fundamental diagram. In [80], two online control algorithms were proposed to increase work zone throughput and reduce queues. Simulation was used to examine the effectiveness of these algorithms. In [81], a coordinated VSL strategy based on model predictive control was presented to suppress shock waves. Shock waves would lead to sudden changes in vehicle speed and therefore unsafe situations and longer travel time. This strategy showed by simulation that travel time can be reduced by $17.4 \%$.

## Integrated Control of Variable Speed Limits and Ramp Metering

VSL is a relatively new control tool in handling freeway congestion, while ramp metering has been studied and utilized for many years in multiple regions. Many researchers have proposed an integrated control of the two to gain more flexibility and better performance in control design. In [82], a coordinated method of variable speed limits and ramp metering was compared with their individual use in a micro-simulation, in terms of safety improvement and travel time reduction. In [83, 84], control methods for integrated ramp metering and variable speed limits were proposed to minimize travel time. They both formulated the traffic dynamics by a second-order model. While [83] is a simple open-loop optimal control, [84] represents a model predictive control approach. Due to the nature of the traffic model, both methods have to solve a nonlinear optimization. Muralidharan et. al [85] proposed a model predictive control algorithm based on the link-node cell transmission Model, which utilized a linear program and hence can be solved efficiently.

## Microscopic Simulation

Microscopic simulation is used by many researchers to evaluate transportation management designs. Different microscopic traffic models have been established and investigated in the past few
years. In a microscopic traffic model, car dynamics are described through a car-following model, lane-changing model and gap-acceptance model. The basic inputs of a car-following model are usually the current speed of the considered vehicle and its front vehicle (leading vehicle), the distance between the two vehicles, reaction time, acceleration, and deceleration. Some car-following models use a set of explicit equations to define the speed in the next steps [86], for example, the Gazis-Herman-Rothery model [87], Gipps' model [88], and Newell's model [89], etc. Some other car-following models are described by multiple driving models, like Michaels' model [90]. The drivers have different types of reactions in different modes, and these modes are separated by certain thresholds. There are also car-following models using fuzzy logic to specify the driver response [91, 92].

The credibility of a microscopic simulation relies on its accuracy. To make the microscopic model reflect actual driver behavior, the simulation model has to be carefully calibrated. Compared to macroscopic models, microscopic models have much more parameters; hence, calibration is more difficult. Different methods have been proposed to achieve a reliable microscopic model calibration, and much of work has been done on the calibration of different microscopic models. Because the Gipps' model will be used here, we only discuss calibration methods and results related to this model. There are three main methods for calibrating the Gipps' model: trajectorybased, aggregated-measurement-based, and steady-state-behavior-based. In the trajectory-based method, which is widely used for calibrating car-following models, model parameters are chosen to make simulated vehicle trajectories close to real vehicle trajectories. One example of this method is [93]. In [93], Ossen used trajectory data extracted from digital camera photos, and tried to minimize both the speed and position errors between simulated and real trajectories. In the aggregated-measurement-based method, aggregated speed, flow or density date collected from simulation is compared with field-collected measurements. One example of this method is [94], where Brockfeld used Berkeley Highway Lab data [95], which were double loop measurements, to calibrate and compare different microscopic models. He averaged the five lane data into one lane, and tried to minimize the error between the simulated vehicle speeds generated by the model with field-measured speeds. In the steady-state-behavior-based method, a relationship between microscopic parameters (reaction time, minimum distance, etc.) and macroscopic parameters (speed, flow, etc.) is derived by analyzing vehicle dynamics in the steady state. Based on the derived relation, microscopic parameters can be obtained from macroscopic measurement. One example of this method is presented in [96, 97]. In [96], Wilson analyzed the steady-state solutions and the stability of Gipps' model. In [97], Rakha presented a model calibration procedure based on Wilson's results.

### 5.2 I-80W Studied Site and Control Design

In this section, we present the I-80W studied site and the control design to address its traffic congestion problem. The first subsection describes the site, while the second subsection presents a
control strategy that includes variable speed limits and ramp metering.

## I-80W PM Peak Traffic

The site in this study is a freeway segment of I-80 westbound direction at the east of San Francisco Bay in California. We will consider the segment from the Carlson Blvd. on-ramp in Richmond to the diverge-point of the I-80 and I-580 in Emeryville, as shown in Fig. 5.3. There are seven on-ramps, seven off-ramps, and one freeway-to-freeway connector within this segment. At the end of this segment, the diverge-point, the five-lane freeway splits into two three-lane freeways (I-580EB/I-880SB and I-80WB). The congestion in the afternoon peak usually starts from the diverge-point and propagates upstream. Upstream of Carlson Blvd. on-ramp is often in free-flow. There are several factors that cause traffic congestion in this segment. First, the traffic demand for this segment is usually high during afternoon peak hours, because there is a great amount of vehicles going to I-580EB/I-880SB. Second, there is a lane drop upstream of the Powell St. on-ramp. Third, the on-ramp flow from Powell St. is large. Finally, some drivers going to the I-580EB/I-880SB take advantage of less traffic on the I-80WB during the afternoon peak in the right-most lanes and change to the desired (left) lanes at the last possible moment. Thus, from the Powell on-ramp to the diverge-point of the I-80WB and I-580EB/I-880SB, there are a large number of vehicles changing lanes. The combination of freeway geometry, high demand, and vehicle traffic weaving lead to severe congestion. Vehicle traffic weaving near the Powell St. on-ramp in particular greatly reduces freeway capacity. Since the freeway exhibits high density during peak hours, vehicles attempting to change lanes have to slow down to wait for a gap. This "slow down" causes the vehicles behind them to brake and creates smaller gaps. The disruption of vehicle speed propagates upstream and produces congestion shock waves. When congestion is formed, freeway throughput is significantly reduced.

Currently, there is no meter control on this stretch of freeway. The detectors installed on this freeway are upstream of the on-ramps, and there is no ramp detector to provide ramp flow. There is one HOV (high occupancy vehicle) lane designated within the studied stretch.


Figure 5.3 The Studied Freeway Segment of I-80 Westbound Direction

## Coordinated VSL and Ramp Metering Control Design

The objective of the control design in this study is to reduce overall travel time in the network while avoiding spillback at the on-ramps. The traffic operation strategies adopted are VSL and ramp metering. As mention in the previous subsection, the congestion in the studied site is caused by high traffic demand and the vehicle traffic weaving. Since traffic demand cannot be reduced, it is necessary to reduce the negative impact of weaving in order to increase freeway mainline throughput. This can be done by enforcing VSL. To prevent insufficient feeding flow to the freeway and spillback at the on-ramps, it is also necessary to regulate on-ramp flow through ramp metering. Consequently, the control design includes two parts, the VSL strategy and the ramp metering strategy. In the two subsections below, we will present these two control strategies.

Before we describe the control design, we discretize the studied network into multiple segments
to facilitate the description of the control design. These segments are represented by directed links. A link starts some distance upstream of an on-ramp and terminates upstream of the next on-ramp. The length of a link is between 100 and 200 meters. This distance was chosen so that a detector would be covered by its corresponding link. The segment from the most downstream on-ramp to the diverge-point was divided into two links for control design purposes. Fig. 5.4 shows an illustration of the network representation.


Figure 5.4 Network Representation of the Studied Freeway Segment in I-80 Westbound Direction

The following notation was used through the remaining subsections:
$u_{i}(k)$ : design vehicle speed limit at link $i$ at time step $k$, from upstream to downstream; link indices are from 1 to M .
$\bar{u}_{i}(k)$ : measured vehicle speed at link $i$ at time step $k$.
$q_{i}(k)$ : aggregated measured mainline flow into link $i$ at time step $k$.
$r_{i}(k), s_{i}(k)$ : aggregated on-ramp and off-ramp flows of link $i$ at time step $k$.
$d_{i}(k)$ : on-ramp flow demand for link $i$ at time step $k$.
$Q_{i}, Q_{m, o}$ : aggregated mainline capacity and on-ramp capacity for link $i$.
$L_{i}, L_{i, o}:$ link $i$ mainline and on-ramp length.
$\rho_{i}(k)$ : aggregated mainline vehicle density of link $i$ at time step $k$.
$\rho_{J}$ : aggregated mainline jammed density for all lanes.
$w_{i}(k)$ : on-ramp queue length for link $i$ at time step $k$.
$\lambda_{i}$ : number of lanes of link $i$.
$\varsigma_{1}, \varsigma_{1}$ : control gains.
$o_{d}$ : desired occupancy.
$\bar{o}(k)$ : measured occupancy of discharge section at time step $k$.
$T$ : simulation time step length.

## Variable Speed Limit Controller Design

In the VSL control design, the studied corridor is divided into four sections, as shown in Fig. 5.5, For the purposes of weakening the negative impact of weaving, we create a special section immediately upstream of the bottleneck to facilitate lane changing. This section is called the discharge section and is about 500 to 700 meters long. The length of this section is selected such that vehicles can accelerate from stopping at the start of this section to free-flow speed at the diverge-point. In this section, vehicle density should be maintained low so that the spacing between two successive vehicles remains sufficiently large to allow vehicles to change lanes without braking. In order to accomplish this, the speed limit at the upstream section will be regulated, while allowing free-flow speed in the discharge section. The section upstream of the discharge section is called the critical section, because the speed limit enforced in this section largely determines the vehicle flow throughput of this system. The speed limit in this section, called the critical speed limit, is the lowest speed limit in the system, and it is regulated such that the downstream discharge section will not be congested while the throughput flow is not reduced too much. If the demand is high, traffic may become highly congested in the near upstream vicinity of the critical section, because that is where the stop-and-go queue forms. Density in these sections may approach the jam density. Supposing that the links involved in this section are from link $t$ to link $M-1$, the section containing these links is called the highly congested section. At the beginning of the studied period, congestion may not exist on the freeway; thus, there is no link in the highly congested section in that period. In addition, the length of the highly congested section may change during the studied period, as it depends on the traffic demand and the exit flow of the system. The section covering the most upstream link of interest (denoted as link 1) to link $t-1$ is called the speed decrease section. The speed limits in this section are decreased gradually from upstream to downstream to prevent the congestion downstream from growing too fast and blocking the flow through ramps. Usually, the most upstream link is in free flow. Again, the length of the speed decrease section may change, depending on the traffic demand. The freeway segment upstream of the speed decrease section has no speed limit control. That is, speed limits there are free-flow speed and stay unchanged.


Figure 5.5 The Road Sections in the VSL Control Design

The diverge-point is the bottleneck of the corridor. In order to reduce the total travel time, we need to adjust the critical speed limit such that throughput at the discharge section is near optimal. To do this, we use an integral control that drives occupancy in the discharge section to preferred values. Therefore, detectors are needed in the discharge section, and the critical speed limit is computed by Eq. (5.1).

$$
u_{M}(k)=u_{M}(k-1)+\left\{\begin{array}{l}
\varsigma_{1}\left(o_{d}-\bar{o}(k-1)\right), \text { if } \bar{o}(k-1)<o_{d}  \tag{5.1}\\
\varsigma_{2}\left(o_{d}-\bar{o}(k-1)\right), \text { if } \bar{o}(k-1)>o_{d}
\end{array}\right.
$$

The two gains in the above control, $\varsigma_{1}$ and $\varsigma_{2}$, should be selected with the consideration of both fast response and acceptable speed change over time. They can be different to obtain flexibility of operation in the case of free-flow and congested traffic, because the change rate of occupancy with respect to flow is different in free flow and congestion. Here, $\bar{o}$ and $o_{d}$ are, respectively, the measured and desired occupancy at the discharge section. The desired occupancy, $o_{d}$, can be estimated in real-time to determine the value at which the flow throughput in the discharge section is maximum. Alternatively, it can be selected to be the critical occupancy. Eq. (5.1) is interpreted as follows: If the discharge section is congested, the speed limit at the critical section is reduced to lower the incoming flow and clear congestion. On the other hand, if the discharge section is able to accommodate a larger flow, the speed limit at the critical section should be increased to release more vehicles, in order to increase its efficiency.

Once the critical speed limit is determined, it is applied to both the critical section and the highly congested section. The reason for the highly congested section to adopt the critical speed limit is to keep the upstream feeding flow consistent with the discharge flow at the critical section. The control equation that expresses this idea is shown by Eq. (5.2).

$$
\begin{equation*}
u_{i}(k)=u_{M}(k), \text { for } i=t, \ldots, M-1 \tag{5.2}
\end{equation*}
$$

In the speed decrease section, the speed limit is designed to enable a smooth change of speed limits. By doing this, hard braking can be avoided and traffic safety guaranteed. The speed limit
design also needs to balance the congestion among links so that the on-ramp or off-ramp in a particular link will not be blocked. Speed limits in the speed decrease section are calculated through Eqs. (5.3) to (5.9).

For $i=2, \ldots, \mathrm{t}-1$

$$
\begin{align*}
u_{1}(k) & =\bar{u}_{1}(k)  \tag{5.3}\\
u_{i}(k) & =u_{i-1}(k)+\max \left\{-10, \min \left\{\left(\eta \alpha_{i}(k)+(1-\eta) \beta_{i}\right)\left[u_{t}(k)-\bar{u}_{1}(k)\right], 0\right\}\right\}  \tag{5.4}\\
\alpha_{i}(k) & =H\left(Q_{i}-\bar{q}_{i}(k)\right)  \tag{5.5}\\
\beta_{i} & =H\left(1 / L_{i, o}\right)  \tag{5.6}\\
\bar{q}_{i}(k) & =q_{i-1}(k)+R_{i}(k)-s_{i}(k)  \tag{5.7}\\
R_{i}(k) & =\min \left\{d_{i}(k), Q_{i, o}, Q_{i}-q_{i-1}(k-1)\right\}  \tag{5.8}\\
H\left(x_{i}\right) & =\frac{1 / x_{i}^{2}}{\sum_{i=1}^{n} 1 / x_{i}^{2}} \tag{5.9}
\end{align*}
$$

Eq. (5.4) defines the posted speed limit for link $i$ by subtracting the speed limit of its immediate upstream link, link $i-1$, by a certain amount. The second term in Eq. (5.4) is always non-positive, because the term $u_{t}(k)$ (speed limit of the last link in the highly congested section) is less than the term $\bar{u}_{1}(k)$ (speed limit of the first link in the network, usually free-flow speed). The max operator used in Eq. (5.4) is to avoid large speed limit differences over space. Here, the speed limit change from the upstream link to the downstream link is restricted to be no greater than $10 \mathrm{~km} / \mathrm{h}$. The parameters $\alpha_{i}(k)$ and $\beta_{i}(k)$ in the second term of Eq. (5.4) are used to balance the traffic over all links, and they change depending on traffic conditions. Here, $\alpha_{i}(k)$ and $\beta_{i}(k)$ are computed by Eqs. (5.5) and (5.6), respectively. The function $H(\cdot)$ used in these two equations is defined by Eq. (5.9). According to this equation, if $x_{i}$ is small in all $x_{i}$ 's, the computed value of this function is large. This means this function gives a larger value of $\alpha$ to a link that has less available space on the mainline and a larger value of $\beta$ to a link that has a shorter on-ramp. As a result, the parameter $\alpha_{i}(k)$ and $\beta_{i}(k)$ in Eq. (5.4) is larger for a link that is more congested; hence, the speed limit for that link drops more. In addition, all $\alpha_{i}(k)$ 's (or $\beta_{i}(k)$ 's) sum up to one by the definition of Eq. (5.9). Therefore, the sum of speed drops of link 1 to $t-1$ will be exactly equal to $\bar{u}_{1}(k)-u_{t}(k)$, if we do not enforce the maximum speed drop of $10 \mathrm{~km} / \mathrm{h}$. Here, $\eta$ is a weighing factor between 0 and 1. If $\eta$ is larger than 0.5 , traffic on the mainline has more emphasis than the queue on the on-ramp of that link.

## Coordinated ramp metering controller design

Once the speed limits for all the links, $u_{i}(k)$ 's, are determined, we can design the metering rate $r_{i}(k)$ 's to further coordinate the traffic among all links. To do this, we compute the metering rate by optimizing system performance, that is, by minimizing total time spent (TTS) and maximizing total travel distance (TTD). We choose to use Model Predictive Control. Here, a model is used to represent system dynamics and the control action is solved by optimizing an object function over a finite time horizon. The model used for prediction is shown by Eqs. (5.10) and (5.11).

$$
\begin{align*}
\rho_{i}(k+1) & =\rho_{i}(k)+\frac{T}{L_{i} \lambda_{i}}\left(\lambda_{i-1} \rho_{i-1}(k) u_{i-1}(k)-\lambda_{i} \rho_{i}(k) u_{i}(k)+r_{i}(k)-s_{i}(k)\right)  \tag{5.10}\\
w_{i}(k+1) & =w_{i}(k)+T\left[d_{i}(k)-r_{m}(k)\right] \tag{5.11}
\end{align*}
$$

In Eq. (5.10), $\lambda_{i-1} \rho_{i-1}(k) u_{i-1}(k)$ and $\lambda_{i} \rho_{i}(k) u_{i}(k)$ are the mainline flows entering and leaving link $i$, respectively, and $r_{i}(k)$ and $s_{i}(k)$ are the on-ramp flow and off-ramp flow, respectively. Hence, Eq. (5.10) is an equation representing flow conservation on the freeway mainline. In Eq. (5.11), $d_{i}(k)$ is the demand (arrival flow) and $r_{m}(k)$ is the metering rate. Thus, Eq. (5.11) represents vehicle conservation on the on-ramp. It is noted that $u_{i}(k)$ is known from the VSL control design, and we assume it is obeyed by the drivers. The dynamics model in Eqs. (5.10) and (5.11) is linear in density and on-ramp queue length.

The objective function is a combination of total time spent (TTS) and weighed total travel distance (TTD). This is represented by (5.12). In this objective function, $\sigma$ is used to balance the values of total time spent and total travel distance. $N_{p}$ is the number of steps (time horizon) for prediction. Total time spent includes the time vehicles spend on the freeway mainline and the time they wait in the on-ramp queues. Consequently, it has two terms in its computation. In contrast, total travel distance only counts the distance vehicles drive on the freeway mainline.

$$
\begin{equation*}
\min J=T T S-\sigma T T D \tag{5.12}
\end{equation*}
$$

Where

$$
\begin{align*}
T T S & =T \sum_{j=1}^{N p} \sum_{i=1}^{M} L_{i} \lambda_{i} \rho_{i}(k+j)+T \sum_{j=1}^{N p} \sum_{i=1}^{M} w_{i}(k+j) \\
T T D & =\sum_{j=1}^{N p} \sum_{i=1}^{M} \lambda_{i} L_{i} q_{i}(k+j) \tag{5.13}
\end{align*}
$$

In practice, the designed ramp metering rate needs to satisfy a set of constraints. First, the on-ramp queue should not exceed the on-ramp storage limit, which is determined by the on-ramp length and jam density. Second, the metering rate should be less than the effective demand, onramp maximum discharge flow, and estimated mainline acceptable flow. Finally, it is desired that the mainline density stays in the region that avoids traffic speed drops. The region can be described by curves of a specified traffic speed drop probability contour, $\varphi\left(u_{i}(k)\right)$, as indicated in Fig. 5.6 Details of these curves are investigated and described in [98]. Once speed $u_{i}(k)$ is known, $\varphi\left(u_{i}(k)\right)$ is a determined number, and we need the freeway density to be less than this number. The three constraints are expressed by 5.14 to 5.16 .

$$
\begin{align*}
& 0 \leq w_{i}(k+j) \leq L_{i, o} \rho_{j}  \tag{5.14}\\
& 0 \leq r_{i}(k+j) \leq \min \left\{d_{i}(k)+w_{i}(k) / T, Q_{i, o}, Q_{i}-q_{i-1}(k), \lambda_{i} \omega_{i}(k)\left[\rho_{J}-\rho_{i}(k)\right]\right\}  \tag{5.15}\\
& 0 \leq \rho_{i}(k+j) \leq \min \left\{\rho_{j}, \varphi\left(u_{i}(k)\right)\right\} \tag{5.16}
\end{align*}
$$



Figure 5.6 Traffic Speed Drop Probability Contour

The decision variables in the optimization are metering rates. The dynamics equations (5.10) and (5.11), the objective function (5.12) and the constraints (5.14) to (5.16) are linear in the metering rates, freeway densities, and on-ramp queues. The optimization problem is a linear program. It can be solved by any linear programming algorithm. Linear Programming is efficient and suitable for real-time implementation, because it can find a global optimum in a limited number of computations.

### 5.3 Simulation Construction

The control strategy described in the above section was evaluated using the Aimsun traffic simulation software [22]. Aimsun is a traffic simulation environment developed by TSS (Transport Simulation Systems). It contains a microscopic simulator to simulate traffic on the individual vehicle level. It also offers the Aimsun Application Programming Interface (API), which is an extension tool for advanced investigation. Aimsun API allows users to program plug-ins to incorporate any application program or implement user-defined control actions. The car-following model implemented in Aimsun is Gipps' model. Details of this software were given in Section 2.2.

A simulation network was built in Aimsun, which duplicated the geometry of the studied site. Each freeway section and ramp in the simulated network has exactly the same length and number of lanes as in reality. The currently speed limit, which is $65 \mathrm{mile} / \mathrm{hr}(120 \mathrm{~km} / \mathrm{hr})$, is also set in the simulation.

## Calibration of the Car-following Model

The following notations are used in this subsection to describe the car-following model used in Aimsun.
$T$ : reaction time. $a(n)$ : the maximum acceleration of the $n$th vehicle. $V^{*}(n)$ : the desired speed of the $n$th vehicle. $d(n)$ : the maximum deceleration of the $n$th vehicle. $d^{\prime}(n-1)$ : the deceleration estimation of the $(n-1)$ th vehicle by the driver of the $n$th vehicle. $\alpha$ : the sensitivity factor. $s(n-1)$ : the length of the $(n-1)$ th vehicle. $x(n, t)$ : the position of the $n$th vehicle at time $t$.

The car-following model provided by Aimsun, and hence used in this study, is the Gipps' carfollowing model [88]. The dynamics of this model are shown by Eqs. (5.17) to (5.20). These equations show that the speed of the $n^{\text {th }}$ vehicle at time $t+T, V(n, t+T)$, is dependent on the states of the preceding $(n-1)^{t h}$ vehicle and itself. The calibration variables are $T, a(n), V^{*}(n)$, $d(n)$, and $\alpha$. A detailed explanation and derivation of this model can be found in [88].

$$
\begin{align*}
V_{a}(n, t+T) & =V(n, t)+2.5 a(n) T\left(1-\frac{V(n, t)}{V^{*}(n)}\right) \sqrt{0.025+\frac{V(n, t)}{V^{*}(n)}}  \tag{5.17}\\
V_{b}(n, t+T) & =d(n) T+\sqrt{d(n)^{2} T^{2}-d(n)\left[\begin{array}{l}
2(x(n-1, t)-s(n-1)-x(n, t)) \\
-V(n, t) T-\frac{V(n-1, t)^{2}}{d^{\prime}(n-1)}
\end{array}\right]}  \tag{5.18}\\
V(n, t+T) & =\min \left\{V_{a}(n, t+T), V_{b}(n, t+T)\right\}  \tag{5.19}\\
d^{\prime}(n-1) & =d(n-1) \alpha \tag{5.20}
\end{align*}
$$

We choose to calibrate the car-following model using the trajectory-based method. The data used for calibration is the I-80 data set in the Next Generation Simulation (NGSIM) data [7]. These data, as part of the product of the Federal Highway Administration (FHWA) NGSIM project, was collected on a freeway segment of the I-80 eastbound direction in Emeryville, California on April 13, 2005. It is at the same location but the opposite direction to our studied site. The segment in the NGSIM data is a six-lane freeway segment which covers 1650 feet and includes an onramp. Vehicle trajectories were derived from camera video, and the data presented included speed and position information of each vehicle at 0.1 second resolution. In the calibration, we only chose vehicles that did not change lanes during their travel within the covered segment, and whose leading vehicle also stayed in the same lane. By doing this, we eliminated the lane changing effect on the car-following model. In total, the trajectories of 754 vehicles were chosen in the calibration.

Because the Gipps' model is described in terms of vehicle speed, we decided to minimize speed error in the calibration. Hence, the objective function in the calibration was chosen to be the root mean square percentage error (RMSPe) between the simulated speed relative to the observed speed for all vehicles. The equation of RMSPe is shown by equation (5.21). In this equation, $v^{o b s}(i)$ and $v^{s i m}(i)$ are the $i$ th data points of observed (real) speed and simulated speed, respectively. Automobiles and trucks are viewed as two different types of vehicles in the calibration; thus, they are calibrated separately.

$$
\begin{equation*}
R M S P e=\sqrt{\frac{1}{N} \sum_{i=1}^{N}\left(\frac{V^{o b s}(i)-V^{\operatorname{sim}}(i)}{V^{o b s}(i)}\right)^{2}} \tag{5.21}
\end{equation*}
$$

Even though Aimsun allows users to use a normal distribution to randomize the car-following parameters, we assume here that their distribution is uniform within a vehicle group, for calibration simplicity. Table 5.1 shows the calibrated parameters, and Fig. 5.7 shows the RMSPe distribution of all the calibrated vehicles. In this calibration, $80.5 \%$ of the vehicles have an RMSPe less than $0.1,77.45 \%$ of them are less than 0.05 , and $71.62 \%$ less than 0.01 . The obtained calibrated values in Table 5.1 are reasonable and consistent with our empirical observation. The calibrated parameters for trucks might be less reliable than those for automobiles, because there were only 25 trucks in the data set, compared with more than 700 automobiles. Fig. 5.8 is an example plot showing both the simulated and observed speeds and the simulated and observed trajectories for one vehicle with an RMSPe of less than 0.01 . As shown in Fig. 5.8, can see that the simulated and real trajectories almost overlap.

Table 5.1 Estimated Driver Behavior Model Parameter Values

|  | Automobile | Truck |
| :---: | :---: | :---: |
| $T$ (second) | 1.12 | 1.05 |
| $a\left(\mathrm{~m} / \mathrm{s}^{2}\right)$ | 1.57 | 1.28 |
| $V^{*}(\mathrm{~km} / \mathrm{h})$ | 108 | 109 |
| $d\left(\mathrm{~m} / \mathrm{s}^{2}\right)$ | -2.29 | -2 |
| $\alpha$ | 1.45 | 1.23 |



Figure 5.7 RMSPe Distribution resulting from the Model Calibration


Figure 5.8 Comparison of the Speed and Trajectory of a Vehicle between Calibration Results and Actual Measurements

## Simulation Set Up and Control Scenarios

The studied site only has freeway mainline detectors installed in the field. These detectors are placed upstream of the on-ramps. Due to the lack of detectors on the ramps, we were not able to obtain actual on-ramp demand data and split ratio for the off-ramps. In addition, the split ratio for the diverge-point was not accessible due to the lack of detectors in this segment. Therefore, virtual demand data was used in the simulation study. The simulation period was 5 hours. In our
simulation, flow demands during the first hour was very high, and in the remaining four hours, $25 \%$ of the demand of the first hour was used for all the on-ramps and the most upstream of the mainline. In the first hour, about 7700 vehicles needed to go through the diverge-point, and $60 \%$ of them go to the I-580EB/I-880SB. Fig. 5.9 shows the accumulative demand of all on-ramps and most the upstream section of the mainline, while Table 5.2 shows the demand distribution of automobiles during the first hour. The selected automobile to truck ratio in the demand was 20:1. Using this demand, it was observed in the simulation that many vehicles need to change lanes after the on-ramp at Powell. This section of freeway has one HOV lane, but this was ignored in the simulation study for simplicity. It should be noted that, as defined by Table 5.2, a simulated vehicle could only exit the freeway at the second or later off-ramp after it entered the freeway.


Figure 5.9 Demand Profiles at all Network Entrances Used in the Freeway Simulation of the proposed Coordinated VSL and Ramp Metering Control Strategy
Table 5.2 Automobile Traffic Volumes between Each Origin and Destination during the First Hour

|  |  | Destination |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Carlson Blvd | Central Ave | Cleveland Ave | $\begin{gathered} \text { Gilman } \\ \mathrm{St} \end{gathered}$ | University Ave | Ashby St | Powell St | $\begin{gathered} \mathrm{I}- \\ 580 \mathrm{~EB} \\ / \mathrm{I}- \\ 880 \mathrm{SB} \end{gathered}$ | I-80WB | Totals |
| $\frac{E_{0}}{0_{0}}$ | mainline <br> most <br> upstream | 51.2 | 51.2 | 51.2 | 102.4 | 102.4 | 102.4 | 153.6 | 1843.2 | 1228.8 | 3686.4 |
|  | Carlson Blvd | 0 | 0 | 25.6 | 25.6 | 25.6 | 25.6 | 20.48 | 276.48 | 184.32 | 583.68 |
|  | Central Ave | 0 | 0 | 0 | 20.48 | 20.48 | 20.48 | 20.48 | 276.48 | 184.32 | 542.72 |
|  | Buchanan St | 0 | 0 | 0 | 0 | 20.48 | 20.48 | 20.48 | 276.48 | 184.32 | 522.24 |
|  | I-580EB | 0 | 0 | 0 | 0 | 0 | 25.6 | 20.48 | 276.48 | 184.32 | 506.88 |
|  | Gilman St | 0 | 0 | 0 | 0 | 0 | 0 | 20.48 | 276.48 | 184.32 | 481.28 |
|  | University Ave | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 276.48 | 184.32 | 460.8 |
|  | Ashby St | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 276.48 | 184.32 | 460.8 |
|  | Powell St | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 645.12 | 430.08 | 1075.2 |
|  | Totals | 51.2 | 51.2 | 76.8 | 148.48 | 168.96 | 194.56 | 256 | 4423.68 | 2949.12 | 8320 |

The calibration results described in the previous subsection were used to define the vehicle characteristics. In all of the results that will be presented, the simulated network had no vehicles at the beginning of the simulation. In the simulation, detectors were placed upstream of each onramp, at the same location as the currently field-installed detectors reported in the Performance Measurement System (PeMS) [2]. By doing so, the proposed control strategy computes the variable speed limits and ramp metering rates using the traffic measurements accessible from existing detectors, and there is no need to install new detectors for future field testing or implementation of the proposed strategy. The link of the discharge section had a length of 500 meters in the simulated network. An advisory speed limit sign was placed at the merge point of each on-ramp. Therefore, speed instructions were activated from one on-ramp to the next. Ramp metering was deployed at each on-ramp, but not at the freeway-to-freeway connector (the I-580 and I-80 merge-point). As a result, the arrival flow from each on-ramp was completely controlled, but the vehicle flows into the most upstream link of the mainline and into I-580 were not regulated.

The proposed VSL and ramp metering control strategies were programmed through the Aimsun API. Fig. 5.10 shows the program structure. Every one minute, the API program collected measurements from each detector and calculates the speed limits and ramp metering rates. In the computation of the metering rate, the demands for on-ramps and most upstream, and flows of offramps were assumed to be known exactly. In practice, this data could be obtained from historical data or the demand could be estimated using real-time demand prediction techniques [99]. The control actions were updated every minute in the simulation to prevent frequent change. It should be observed that the equations for speed limit computation are explicit and can be executed very quickly. The optimization used the ramp metering in rate computation is a linear program, which can be solved very efficiently by existing algorithms, even for a large-scale network. Hence, the proposed control algorithm can be computed sufficiently fast for real-time implementation.

It needs to be mentioned that the on-ramps in the simulation network were extended, in order to accommodate all the queued vehicles generated from the demand and the imposed ramp meter rates, and to include the waiting time of all the queued vehicles in the computation of total time spent (TTS). In practice, the vehicles stored in the extended section of the on-ramps would have spilled back into arterials if the current on-ramp length were used.

To prevent on-ramp queue spillback into neighboring arteries, queue-overwrite control was used in ramp metering. In the queue-overwrite control used in this simulation study, ramp metering at an on-ramp was switched off (display all green) if the corresponding on-ramp queue was too long. Every 20 seconds, the length of each on-ramp queue was measured. If the on-ramp queue exceeded a pre-defined limit, the ramp metering stop lights displayed a continuous green, allowing all the on-ramp traffic to merge into the mainline until the on-ramp queue is shorter than the prescribed limit.

Also for practical considerations, the displayed variable speed limit was rounded to the nearest multiple of 5 , such as 60 or 65 . Moreover, it was assumed in the simulation study that all the drivers followed the advisory speed limit in the ramp metering control design, while in reality this is not necessarily true. Hence, it is necessary to evaluate the control impact for incomplete


Figure 5.10 The Structure of the Control Algorithm Implementation
compliance. The simulation results in the next section show that $30 \%$ speed limit compliance has a similar outcome to $100 \%$ compliance.

Six control strategies were simulated in the simulation study, as follows: 1) All time VSL, in which the variable speed limits control is on from the beginning of the simulation but the ramp metering is deactivated; 2) all time combined VSL and ramp metering, in which both the VSL and the ramp metering are activated from the beginning of simulation; 3) switched VSL, in which ramp metering is off all the time and variable speed limits are switched on only when the traffic congested (defined as when the detected occupancy in the discharge section is over a specified threshold) and is switched off when traffic recovers to free flow; 4) switched combined VSL and ramp metering, in which both VSL and ramp metering are only turned on when congestion is detected; 5) all time ramp metering, in which only ramp metering is implemented and it is utilized from the beginning of simulation; 6) all time VSL with $30 \%$ compliance, in which only the variable speed limits is activated and only $30 \%$ of drivers follow the posted speed limit. Scenarios 1 to 5 assume $100 \%$ driver compliance with the posted values of speed limits.

### 5.4 SIMULATION RESULTS

The performance of the proposed control strategies are shown in Figs. 5.11 to 5.13 and Table 5.3. It can be observed that ramp metering alone improves traffic conditions in the studied freeway segment, but more improvement is achieved if VSL is used (alone or together with ramp metering). Table 5.3 shows that if we use variable VSL, total travel time can decrease by more than a half, and delay drops to around one third. Fig. 5.14 shows that average vehicle speed at the diverge-point is greatly increased by VSL control.

There are several observations that can be drawn from Figs. 5.11 to 5.13 and Table 5.3. First, the difference between using VSL all the time and using combined VSL and ramp metering all the time is not significant. Likewise, switched combined VSL and ramp metering does not decrease congestion significantly when compared to switched VSL. Second, it is observed that all time control performs significantly better than the corresponding switched control. A possible explanation for this observation is that in this simulation study the control may have been switched on too late and, as a consequence, it takes more time for the discharge section to recover to free flow once it congests. Therefore, the choice of threshold for turning on the traffic management control needs further consideration. Finally, the results show that $30 \%$ driver compliance regarding advisory speed limit yields similar results to $100 \%$ compliance. This may be explained as follows: Due to the density being rather high for saturated traffic, if $30 \%$ of the drivers follow the posted VSL, other drivers have to follow, since there is no space for other vehicles to wiggle around, and pass the slower vehicles.


Figure 5.11 Comparison of the Simulated Accumulative Delay Using Different Control strategies in the I-80W Network: no control, all time VSL, all time combined VSL and ramp metering, switched VSL, switched combined VSL and ramp metering, switched metering, and all time VSL with $30 \%$ compliance


Figure 5.12 Comparison of the Simulated Accumulative Times Spent in Different Control Strategies of the I-80W Network: No Control, All Time VSL, All Time Combined VSL and Ramp Metering, Switched VSL, Switched Combined VSL and Ramp Metering, Switched Metering, and All Time VSL with 30\% Compliance


Figure 5.13 Comparison of the Simulated Accumulative Travel Distances in Different Control Strategies of the I-80W Network: No Control, All Time VSL, All Time Combined VSL and Ramp Metering, Switched VSL, Switched Combined VSL and Ramp Metering, Switched Metering, and All Time VSL with 30\% Compliance


Figure 5.14 Comparison of the Simulated Vehicle Speed at the Diverge-point in the No Control and All Time VSL Strategies

Fig. 5.14 compares the simulated vehicle speed at the diverge-point, where I-80 and I-580 split, for the cases when VSL is active at all times and when no control is active. As shown in the figure, when VSL is active, the speed at the diverge point exhibits oscillatory from the beginning of the simulation to about 2.2 hours ( 8000 seconds), which is the period of congestion. This phenomenon is caused by the response delay in control and in drivers' response. The control delay is due to the fact that the posted speed limits are updated only every 1 minute in the simulation. When congestion began to form near the diverge-point, the VSL control system has to wait for the next computation cycle to collect detection and set new speed limits. During this one-minute period, vehicles will accumulate near the diverge-point. After the speed limits are updated, drivers from upstream have to slow down, and as a consequence flow to the diverge-point is gradually reduced. However, it takes some time for the congestion to clear up. When congestion is cleared, the flow to the diverge-point is usually less than capacity, and the VSL control algorithm will
increase the posted speed limits until the detected occupancy near the diverge-point is no less than the target value. However, when the controller detects the occupancy again exceeds the target value, congestion will have generally started because the detection only happens at a fixed interval. Two methods may be used to smooth the speed oscillations. One is to use a smaller gain in the computation of critical speed limit (Eq. (5.1)). This would reduce the variation of speed but also slow the system response. The other method is to utilize a prediction model to estimate the change of occupancy. This can reduce the delay of control action and avoid over-increasing or over-reducing the flow to the diverge-point.

Table 5.3 Performance Comparison of Different Control Strategies in the Freeway Coordination Control Simulation: No Control, All Time VSL, All Time Combined VSL and Ramp Metering, Switched VSL, Switched Combined VSL and Ramp Metering, Switched Metering, and All Time VSL with 30\% Compliance

|  | $\begin{aligned} & z \\ & 0 \\ & 0 \\ & 0 \\ & 0 \\ & 0 \end{aligned}$ |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| TTS ( $\times 10^{3}$ veh $\cdot \mathrm{h}$ ) | 13.595 | 5.4719 | 5.7129 | 6.6299 | 6.4305 | 11.006 | 6.1423 |
| $\frac{\text { TTS with control }}{\text { TTS without control }} \%$ | - | 40.25 | 42.02 | 48.77 | 47.30 | 80.96 | 45.18 |
| Delay ( $\times 10^{3}$ veh $\cdot \mathrm{h}$ ) | 12.4 | 3.5948 | 3.9525 | 4.7891 | 4.5556 | 9.5601 | 4.2333 |
| $\frac{\text { Delay with control }}{\text { Delay without control }} \%$ | - | 28.99 | 31.88 | 38.62 | 36.74 | 77.1 | 34.14 |

### 5.5 CONCLUDING REMARKS AND FUTURE WORK

This chapter presented a control strategy for mitigation of bottlenecks caused by weaving effects on a freeway. The strategy included a variable speed limits (VSL) design and a coordinated ramp metering design. A discharge section was created by the critical speed limit in the VSL design to maximize the bottleneck flow. The ramp metering design tried to minimize total travel time and maximize total travel distance in a finite predictive time horizon. This strategy was tested by
microscopic simulation after calibrating the car-following model using the NGSIM data. Different control methods, including all time VSL, all time combined (VSL and ramp metering), switched VSL, switched combined, switched metering, and all time VSL with $30 \%$ compliance, were simulated and compared with no control case. The simulation results showed that the proposed control strategy can improve traffic performance significantly. In particular, the VSL alone control is more effective than the ramp metering alone control, as in the switched VSL case the system delay was reduced about $60 \%$ while in the switched metering case only $23 \%$. It was also found that the VSL with $30 \%$ driver compliance yielded similar performance of the VSL with $100 \%$ driver compliance. Overall, the performance difference between the all-time control cases and the switched control cases were not large.

## Chapter 6

## Coordinated Freeway Ramp Metering and

## Street Intersection Signal Control

Urban street signal control and freeway ramp metering are the two traffic operations that we encounter most often in our daily commutes. In the current practice of traffic operations in most areas, traffic control on freeways and that on urban streets work independently. These controls only look at the traffic conditions within their respective systems and actions independently of each other. However, the freeway and urban street systems are physically connected and exhibit a strong interaction with each other in two ways. First, drivers access and leave freeways through ramps that discharge at intersections close to the freeway. Second, if the freeway is congested due to an incident, drivers may reroute to a neighboring parallel signalized arterial. In this case, the arterial parallel to the freeway will be subjected to a larger traffic flow than normal, due to freeway congestion. If the signals in the parallel arterial are not adjusted to accommodate the increase flow due to the directed freeway traffic, it may also congest and in turn induced even more freeway traffic congestion. For these two reasons, the independent controls of the two systems may reduce the performance on both roadways due to bad coordination.

In this chapter, we will present a novel coordinated control of freeway ramp metering and street signal control on a small-scale system, namely one freeway on-ramp and its adjacent intersection. This chapter is organized as follows: Section 6.1 provides an introduction to signal control and the related works on coordinating signal control and ramp metering; Section 6.2 describes the studied site used in this chapter and a microscopic model used to simulate the traffic at the studied site; Section 6.3 explains the control strategy that we propose for the studied site; Section 6.4 presents the results of a simulation study in which the proposed control strategy was compare to the currently field-implemented control plans; and the final section summarizes this chapter.

### 6.1 INTRODUCTION

## Signal Control on Street Intersections

## Signal Control Basics

Before introducing different types of intersection signal control methods that have been utilized or proposed, we introduce some terminologies that will be used throughout this chapter. Note that, even though some of the terminologies that will be introduced are used for both vehicular and pedestrian movements in other research, we only consider vehicular movements in this chapter.

Direction Direction is used to point out where the traffic goes toward. In a standard four-way intersection, the directions usually are eastbound, southbound, westbound, and northbound.

Movement Movement is used to refer the turning that a driver takes at an intersection. Usually, there are three movements for each direction, left-turn, through traversing, and right-turn. U-turn is not considered in this chapter, as it occurs much less frequent than the other three.

Protected turning A left-turn is protected if a green light is given for this movement and no green light is given to any conflict traffic to this left-turn movement simultaneously. A protected left-turn is usually indicated by a green arrow.

Permissive turning In contrast to a protected left-turn, a permissive left-turn receives non-exclusive green lights. Vehicles involved in a permissive left-turn should yield to upcoming traffic from the opposite direction, and initiate a turning only when there is a sufficiently large time gap to make a safe turn.

Phase In signal control, a phase is a combination of one or more movements that receive the green light simultaneously. In general, a movement should only be assigned to a phase. For example, the left-turn, through and right-turn of the northbound direction in Fig. 6.1 (a) are assigned to phase 2. In contrast, in Fig. 6.1(b), phase 2 contains only the through and rightturn of the northbound direction, while the left-turn is assigned to phase 5. In this figure and the other figures in this section, the symbol $\phi$ is used to represent phase. Thus, $\phi 2$ means phase 2.

Stage A stage is a combination of one or more phases that receive the green light simultaneously. The definition of stage appears to be similar to that of phase. However, a phase can be assigned to one or more stages, while a movement is only assigned to one phase. For instance, the first stage in Fig. 6.2 contains phases 1 and 5, which are the southbound left-turn and northbound left-turn, respectively. The second stage in that figure contains phases 1 and 6, where phase 6 represents northbound through and right-turn movements.

Cycle The signal lights are given periodically at intersections. In one cycle, all the phases that demand green signals will be served once. The order of service is defined by the phase sequence.

Cycle Length The time duration that it takes to complete a cycle is called cycle length or cycle time. The cycle length can be a constant or time-varying, depending on which type a signal control plan is used (the signal control types will be discussed in next subsection). Usually, a phase is given the green light once per cycle, but it could also be skipped or given more than once, depending on the type of signal control being used.

Split The percentage of a cycle length that is given to a phase is call split or green split.
Offset Offset is used in coordinated signal control. It is the time difference (seconds or percent of cycle length) between a defined point in an intersection cycle and a global system reference point. Or it can be view as the time difference between the intersection clock and the system clock.

Detector configuration Sensors at intersections that detect the presence of vehicles stopped at a red light and/or measure that flow of vehicles going through an intersection are called detectors. Detector configuration is the layout of detectors at an intersection. In signal control, a detector can be classified as a stop-bar, an approach, a departure detector, or some other types, depending on its location. A stop-bar detector is installed immediately upstream of an intersection (right before the stop-line), while a departure detector is immediately downstream of the intersection. An approach detector is installed a certain distance upstream of the stop line.

NEMA The National Electrical Manufacturers Association (NEMA) standard is a standard widely used to configure signal controls. The structure of the NEMA standard will be described in the paragraphs below.

Dual-ring The Dual-ring structure is a specification of the phase sequence (order). It will be discussed in the paragraphs below.

The NEMA standard is used by traffic engineers, controller manufacturers, and many other people to avoid misunderstandings when specifying signal control configurations. The NEMA standard specifies a functional configuration that can efficiently transit between phases while preventing conflicts. In the NEMA structure, even phases are assigned to through movements and the associated right-turns, and odd phases are assigned to left-turn movements if the left-turn movements are separated from the through phases of the same directions. The odd phase that is tied to a particular even phase is fixed. The phase layouts in Fig. 6.1 actually follows the NEMA standard. Phase 1 is always the left-turn phase of the same direction of phase 6 . Similarly for phases 3,5 , and 7. The even phases are laid down in the clock-wise direction on the map. In addition, phases 2 and 6 are usually used to represent the through movements on the major roads, while phase 4 and 8 represent the through movements on the cross streets.

Once the phase numbers are assigned to the movements according to the NEMA standard, the so-called dual-ring can be used to figure out the phase sequence in a signal control. The dualring, as its name suggests, has an upper and a lower ring. The upper ring contains phases 1 to 4 , while the lower ring contains phases 5 to 8 . If a particular phase is missing at an intersection, it is excluded from the dual-ring. Fig. 6.3 provides one example of the dual-ring layout. The time used to go through phase 1 to phase 4 in the upper ring, and that used to go through phase 5 to phase 8 in the lower ring, both equal the cycle length. There are two barriers in the dual-ring, the one located after the first half of the ring (after phases 2 and 6 in Fig. 6.3) and the other after the second half (after phase 4 and 8 in the figure). The dual-ring structure requires that the upper and lower rings cross the barrier (move to the following phases) simultaneously. Referring to Fig. 6.1 (a), this requirement means that the north and south (or east and west) bound traffic stop at the same time. The restrictions imposed by the barriers prevent traffic conflicts. The phase sequence (order) between two barriers can be altered by traffic engineers. For example, the order of phases 1 and 2 can be exchanged in Fig. 6.3. The Same can be done for phases 3 and 4, phases 5 and 6, and phases 7 and 8 .

(a) The Phase Layout, Example 1

(b) The Phase Layout, Example 2

Figure 6.1 The Illustration of Phase Layout in Signal Control


Figure 6.2 An Illustration of the Signal Phases and Stages Used in Signal Control


Figure 6.3 The Dual-ring Structure Used in Signal Control

## Signalized Intersection Control Methods

Signal control can be seen at major intersections on arterials, commercial areas, and residential areas [100]. Arterials are the main roads through the cities that carry the major of traffic. Commercial areas are usually downtown or business area. Signal control on surface streets can be divided into two categories, namely isolated and coordinated. In isolated signal control, an intersection signal is determined without considering neighboring signalized intersections. This is often used at an intersection that is remote from other signalized intersections, or where the traffic at that intersection is not influenced much by the traffic at other intersections. In coordinated signal control, signals of a group of adjacent intersections are coordinated to provide progressive traffic flow for the favored direction(s). This is often used at intersections that are closely spaced, so that vehicles can pass through along the favored direction without frequent stops and large waiting delays.

Signal control can also be classified into three types, namely pre-timed, actuated, and advanced. In pre-timed control, signal timing parameters are fixed over a period of time, including cycle length, phase sequence, green splits, offset, and so on. In actuated control, on the other hand, the amount of green time in each phase varies based on traffic demand. Detectors are utilized in actuated control to detect the arrivals of vehicles. Generally, stop-bar detectors and/or approach detectors are used. Stop-bar detectors are sensors that are installed immediately upstream of the stop line. They are used to detect the presence of vehicles (to detect if there is any vehicle waiting for a green signal). In contrast, approach detectors are installed a certain distance from the stop line. They are used to detect whether there is a vehicle coming. A phase is actuated when the corresponding detectors detect the presence of vehicles. The green time of an actuated phase is extended in response to the arrivals of successive vehicles. Actuated control is classified as fully actuated or semi-actuated. If all phases need actuation to activate, the operation is called fully actuated. Otherwise, it is called semi-actuated, in which at least one phase is guaranteed to be served in every cycle. In the next subsection, the algorithm for actuated control will be described in detail.

Pre-timed control is suitable to use if intersections are densely located, and traffic demand is predictable and constant during a period. Hence, it is often used in commercial areas and on arterials, and it is usually coordinated. Fully actuated control is often used in isolated intersections. Semi-actuated control can be used when both flexible green splits and coordination are needed. Thus, it is seen on arterials. Both pre-timed and actuated control use pre-configured timing parameters derived from historical traffic data. Hence, they must be re-evaluated and updated periodically.

Advanced signal control has more capability to change the way green splits are given than pre-timed control or actuated control, and thus has larger potential to adapt to traffic variations and to improve traffic performance. Depending on its freedom to change signal timings, an advanced signal control method is classified as traffic adjusted, traffic responsive, or traffic adaptive. Traffic adjusted (also called plan selection by some researchers) methods automatically select preconfigured timing plans according to the detected traffic conditions. The time interval between two
decision points for determining whether to change from current signal timing plan to another is relatively large, usually more than 15 minutes. On the other hand, both traffic responsive and traffic adaptive methods deploy online computation to continuously and rapidly adjust control settings. Traffic responsive methods can update green splits, cycle lengths, and offsets every cycle or every few minutes, but the changes in each update are relatively small and smooth. Traffic adaptive methods have even more freedom than traffic responsive methods, in that they allow larger and more frequent adjustments. The update interval for a traffic adaptive method can be as small as a few seconds. Traffic responsive methods usually employ certain traffic models and optimization strategies, while traffic adaptive methods often additionally utilize prediction of traffic demand. Both traffic responsive and traffic adaptive methods must have reliable detection systems. Well-known examples of advanced signal control strategies include SCOOT (Split Cycle Offset Optimisation Technique) [101], SCATS (the Sydney Coordinated Adaptive Traffic System) [102], RHODES (Real-time Hierarchical Optimization Distributed Effective System) [103], and OPAC (Optimized Policies for Adaptive Control) [104]. The former two are traffic responsive and the latter two are traffic adaptive.

## Actuated Control

In actuated control, the green light of a phase is activated (changed from red to green) and extended by the actuation of the associated detectors. Stop-bar and/or approach detectors are usually used. A detector is used to activate the phase, extend the green light, or both, depends on the signal settings. For illustration purpose, we will assume that a phase is activated by the actuation of stop-bar detectors, while it is extended by the actuation of approach detectors in the description here. Fig. 6.4 illustrates how the actuated control works. During the red time of a particular phase, named phase A, a vehicle stops on a stop-bar detector and generates an actuation. The control system detects the presence of the vehicle and calls for the green light of phase A to be activated. After the phase prior to phase A in the dual-ring completes, the green light of phase A is then displayed, and its duration is initially set to a pre-defined amount of time, called the minimum green. After the minimum green finishes, a clock starts to count down from a certain amount of time, called the time gap. The time gap is usually 2 to 3 seconds, depending on the setting. If a vehicle passes on an approach detector of phase A during the countdown, the actuation will reset the countdown clock to the set time gap, and the displayed green light will be extended. On the other hand, if there is no vehicle passing on the approach detectors of phase A during the countdown, its green light is terminated (change to yellow) and the signal will move to the next phase in the dual ring, as shown in Fig. 6.4 (a). Termination caused by no actuation during the time gap is called gap-out. If phase A continuously receives actuation during the countdown, the countdown clock will be reset multiple times, and its green time will reach a maximum amount of time allowed, which is called maximum green. Then, the green signal will be terminated regardless of actuation, and this type of termination is called max-out, as shown in Fig. 6.4(b). By extending the green light based on vehicle actuation, the signal lights adapt to the traffic volume variation.

If all the phases must have vehicle actuation to generate a call for the green signal, this actuated
control is classified as being fully actuated. Detectors must be installed at all the roads approaching to the intersection in fully actuated control. In contrast in semi-actuated control, detectors are only installed on the cross streets and the left-turns of major streets (if left-turn is defined to be a separate phase). Thus, the green lights of major street through phases, which usually labeled as phases 2 and 6 in the dual-ring, are activated at every cycle in semi-actuated control (through phases on major streets are guaranteed to serve).

From the description above, it can be seen that the green length of a phase under actuated control varies from cycle to cycle (the time point when the green light terminates is floating), depending on how many vehicles are detected. The larger traffic volume a phase has, the longer green it receives. Remember that the dual-ring structure requires that the upper and lower rings must cross the barriers simultaneously. As a consequence, the phase that terminates earlier will hold its green light until the other phase also terminates. Moreover, there is no way to coordinate multiple intersections if the cycle lengths change over time, and the time points when the green lights start and terminate are floating. A fix starting time point of a cycle is needed in coordinated control, for the purpose of maintaining the offset synchronized with the global system clock. As a consequence, the cycle lengths of coordinated intersections are fixed, as well as the time instance when the dual-ring crosses the barrier after the first half portion. In the dual-ring example of Fig. 6.3. this means that the green lights of phases 2 and 6 are forced to turn off at a fixed time point, regardless of vehicle actuation.

(a) The Illustration of Gap-out in Actuated Control

(b) The Illustration of Max-out in Actuated Control

Figure 6.4 The Illustration of Actuated Control

## Coordination of Freeway Ramp Metering and Intersection Signal Control

In Section 5.1, we introduced freeway ramp metering and its related works. In this subsection, research on coordinated control of ramp metering and signal control is described.

Strategies of coordinated control between freeway ramp meters and intersection signals can be divided into two groups, model-based or non-model-based. Model-based methods usually utilize traffic dynamics models and determine the control actions through the solutions of optimization problems. Examples of this type are [105, 106]. In [105], Papageorgiou presented a unified design approach of integrated control for traffic corridors based on a store-and-forward model. This method could formulate a traffic network control problem of any topology and various traffic control approaches, such as ramp metering, signal control and route guidance, as a linear optimization problem. In [106], van den Berg et al. proposed a control method for a mixed network of freeways and urban streets, which used the METANET model to represent freeway traffic, a queue model to represent urban street traffic, and used the Model Predictive Control frame work to determine the control actions. In a model-based design, traffic dynamics and other factors that affect control can be included in the control synthesis design because of the use of a model. Hence, the design can be applied to different topologies and various scales of network, mixed control techniques and different control constraints. However, knowledge of boundary traffic demand and turning proportions at intersections/junctions, as well as calibrated models, is required in such an approach.

In non-model-based strategies, a common method is to switch among different signal control plans according to real-time traffic situations. Examples of this type can be found in [107, 108, [109]. In [107], Kwon et al. proposed an adaptive control to coordinate the ramp meter and the adjacent intersection signal without requiring the prediction of demand. In this approach, vehicle counts and presence determined by loop detectors were used to compute a congestion index for each link on the freeway and arterial. Based on the congestion index, the metering rate was adjusted to balance the congestion level of the freeway and arterial, and the green duration of a phase was determined in a similar structure of actuated control, with the congestion index replacing vehicle actuation. In [108], Tian et al. developed an integrated control strategy of a system in which the freeway and street were connected by a diamond interchange. In this strategy, the metering rate was selected based on freeway occupancy and the maximum green of intersection actuated control was changed if rerouting occurred. A similar idea appeared in [109]. In that paper, Zhang et al. proposed a simple locally synchronized control by utilizing actuated signal control and ramp metering method ALINEA. In the proposed control, ramp metering rate was computed according to the ALINEA algorithm, but it would be switched off when intersection was congested and queue-overwrite was needed. The intersection control was actuated control, but the maximum green of each phase was adjusted according to the detected traffic situation.

The coordinated control strategy presented in this chapter is proposed to use at a freeway onramp and the adjacent intersection. It is a non-model-based strategy. Similar to [109], the proposed control strategy will utilize ALINEA algorithm to determine the ramp metering rate. However, the ramp metering will not be switched off when queue-overwrite is need. Instead, it will use a pre-
defined ramp metering rate to release on-ramp queue. The intersection signal control is determined by an optimization, taking into account the demand of each phase and the available storage space of the on-ramp. This control strategy will be described in detail in Section6.3.

### 6.2 DESCRIPTION OF STUDIED SITE AND MODEL

## CALIBRATION

## Studied Site

The site selected in this study is an interchange located at the SR-87 and W. Taylor St. in San Jose, California. This is about 2 miles south of the San Jose International Airport. Fig. 6.5 shows the map of this location. During afternoon peak hours, there is a large traffic flow on SR-87 SB from Taylor St. The freeway mainline has two general lanes and one high-occupancy vehicle (HOV) lane. The southbound on-ramp is about 200 meters long, and has two general lanes and one HOV lane. The on-ramp flow in peak hours can be 1300 vehicles per hour (vph) or higher. The downstream on-ramp, the Julian St. on-ramp, is about 800 meters from the Taylor St. on-ramp. The merge section for the Taylor St. on-ramp is very short because the acceleration lane becomes an exit-only lane soon after the Taylor St. on-ramp. Thus, drivers have to merge to the left lanes very quickly after they get on the freeway. The freeway road geometry at this location, together with the high mainline and on-ramp traffic demands, makes this location an on-ramp bottleneck. According to the measurement collected from loop detectors installed upstream of the on-ramp, flow recorded there during typical weekday afternoon peak hours is only about 1300 to 1500 vehicles per hour per lane, while it can reach 1600 to 1800 vehicles per hour per lane during non-peak hours. Speed at peak hours is about 20 mph (data source from the Performance Measurement System (PeMS) [2]).


Figure 6.5 Map of the Intersection at SR-87 and W. Taylor Street

Table 6.1 shows the traffic volumes on the interchange from the data collected on Wednesday, September 5th, 2012. The data were collected from 4:15 p.m. to 7:00 p.m., during which time the freeway was congested. This data were obtained through the Miovision system [110]. In this system, overhead cameras were set up at the intersection to record videos, and traffic volumes were obtained by image processing. The data collection date was a typical workday, and the weather was fine. From the table, it can be observed that the westbound left-turn (LT) and the eastbound right-turn (RT), the two movements towards the southbound on-ramp, had significant flows at the intersection. Due to congestion in the merge area on the freeway, there was usually a long queue at the on-ramp. Queue spillover at the on-ramp can be observed from the videos during peak hours.

On-ramp spillover further caused long westbound left-turn and eastbound right-turn queues at the intersection. Because the Taylor St. \& San Pedro St. intersection is very close to the studied Taylor St. \& SR-87 intersection, the westbound left-turn bay cannot accommodate the queue occasionally. This queue was frequently seen to propagate and block the upstream intersection(s) from the video. From data recorded by PeMS, the volume of the middle freeway lane was 4093 vehicles during the same period that intersection data was collected, and the HOV lane (leftmost lane) traffic volume was 2357 vehicles. The rightmost lane did not have measurements because the detector was not functioning. From the video, we observed that the right-most freeway lane had close traffic volume to the middle lane. As a consequence, we assumed the right-most lane flow was the same as the middle lane flow in the simulation we conducted.

Table 6.1 Vehicle Volumes at the SR-87 \& Taylor St intersection during 4:15-7:00pm, Sep 05,
2012, unit: veh

|  | EB | EB | EB | WB | WB | LB Left | NB <br> Left | NB <br> Through |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Right | Left | Through | SB |  |  |  |  |  |
| Cehicle | 240 | 1037 | 1732 | 2051 | 1009 | 302 | 1209 | 1738 |

## Model Calibration

An Aimsun microscopic simulation model was built in this study, in order to evaluate the performance of the proposed control method. An introduction to the Aimsun simulation software can be found in Section 2.2. This simulation model covered one freeway segment and two arterial intersections. The freeway segment is from about 1800 meters north of Taylor St. to the Julian St on-ramp (the Julian St on-ramp is southbound downstream ramp of the Taylor St on-ramp). Both directions of freeway traffic and all the ramps within this region were included in the simulation model. The two arterial intersections modeled were the interchange of SR-87 and Taylor St. and the intersection of Taylor St. and San Pedro St. It should be noted that, even though the intersection of Taylor St. and San Pedro St was included in the simulation network, the proposed control design is only for the SR-87 \& Taylor St. interchange in this study. The Taylor St. \& San Pedro St. intersection was covered in the simulation model in order to evaluate queue spillovers. The geometry and speed limits of the roads in the simulation model were the same as in reality.

To accurately re-create traffic conditions, field-implemented traffic control plans and fieldcollected traffic volumes were utilized in the simulation model. Traffic control plans were obtained from maintenance agencies. According to those control plans, both ramp meter and signal control are both time-of-day plans. The ramp metering rate is selected from a set of values based on free-
way mainline occupancy. The two arterial intersections are in actuated control, but they are not coordinated. Traffic volumes were obtained from two sources. One data source was the intersection data from the Miovision system mentioned in the previous subsection. The data covered the period from 4:15 to 7:00pm during Sep. 05, 2012, and volumes were given at every minute. The volumes of the westbound right-turn and southbound right-turn were missing in the intersection data, because they were not captured by the cameras. However, the missing data should not affect the accuracy of the simulation model. This is because the demands for these two movements were much lower than the others, about five vehicles every 2 minutes for each movement. The westbound right-turn had a very long green time and the southbound right-turn was stop controlled (there was no traffic light but a stop sign). There was usually no queue present in these two movements, and the interaction between the vehicles of these two movements and the vehicles of the other movements of the intersection was very limited. The other data source was the freeway data from PeMS [2]. An introduction to PeMS can be found in Section 2.1. Detectors were placed upstream of the on-ramps, and PeMS traffic data contain flow, occupancy, speed, and so on for the freeway mainline. There is no ramp data provided by PeMS, but this can be computed from the intersection data. The intersection data and freeway data were aggregated over 5 minutes as the demand profile in the simulation model. Remember that in Chapter 3, we updated demand profile every 10 seconds in order to capture an accurate time profile of vehicle arrival and gain necessary simulation accuracy. The choice of 5-minute update period used in this chapter did not contradict what we did in Chapter 3 for two reasons. First, unlike the traffic flow observed in Chapter 3, the vehicle arrival flow at the studied location of this chapter was observed to change very slowly during the study period. Second, there were usually long queues at the studied intersections of this chapter, which guaranteed that the intersections always had sufficient discharge flow. The model calibration results presented later in this subsection will demonstrate that the 5-minute update period was adequate to provide required model accuracy.

Since the micro-simulation is used as a test bed, it has to be sufficiently accurate to represent real traffic conditions. Model calibration was conducted before we used the model to evaluate the proposed control. The criterion for calibration used here is selected from [111, 112]. The two sources give microscopic simulation criterions that are widely accepted by traffic engineers. Basically, they require the aggregated flow, speed, and other traffic performance metrics of microscopic simulation to be match the reality within some specified errors. In order to capture the dynamic pattern, we not only utilized the error range specified in the references, but also compared the flow and speed on the freeway mainline and link flows at intersections with real traffic measurements every 10 minutes. For flows, at least $85 \%$ of the flows are required to have acceptable error and $\mathrm{GEH}<5$, as we now describe. A simulated flow quantity is said to have acceptable error if it satisfies the requirement below.

Link flow quantity

1. If $700 \mathrm{vph}<$ real flow $<2700 \mathrm{vph}$, simulated flow has an error within $15 \%$;
2. If real flow $<700 \mathrm{vph}$, simulated flow has an error within 100 vph ;
3. If real flow $>2700 \mathrm{vph}$, simulated flow has an error within 400 vph .

In this sense, the percentage of flows with an acceptable error is that of simulated flows falling within the allowable error range mentioned above. The GEH statistic is a statistic value frequently used in traffic engineering to compare different sets of traffic volumes. It can be computed by Eq. (6.1).

$$
\begin{equation*}
G E H(k)=\sqrt{\frac{2(M(k)-C(k))^{2}}{M(k)+C(k)}} \tag{6.1}
\end{equation*}
$$

In this equation, $M(k)$ is the simulated flow and $C(k)$ is the corresponding flow measured in the field at time $k$. A satisfactory calibration requires that the simulated flow quantities can satisfy the condition $G E H(k)<5$ for least $85 \%$ of time points $k$. For speed, at least $85 \%$ of simulated speed values are required to achieve an error within 5 mph .

The simulation model has been run with 20 different random seeds. Table 6.2 shows a summary of calibration results for intersection movements at SR-87 and Taylor St. It can be seen that, on average, all movements have more than $85 \%$ of the time points at which the flow is within the required error range. The percentages of flows for all movements with GEH $<5$ also reach $85 \%$ or higher. Row 3 and Row 4 show the standard deviation of the percentages. Notice that one data point accounts for $6.25 \%$ in the simulation. It is observed that the deviation is in an acceptable range. Table 6.3 shows the result for the freeway. This illustrates that the percentage of flows meeting the allowable error and the percentage of flows with GEH $<5$ are again higher than $85 \%$. In addition, more than $85 \%$ of the speed values exhibit less than 5 mph error. In summary, Table 6.2 and Table 6.3 indicate that the model calibration is satisfactory.
Table 6.2 Calibration Results of the Intersection Flows

| EB Left | EB <br> Through | EB <br> Right | WB <br> Left | WB <br> Through | SB Left | NB Left | NB <br> Right |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| with acceptable <br> error | $100.0 \%$ | $100.0 \%$ | $92.1 \%$ | $89.2 \%$ | $98.3 \%$ | $100.0 \%$ | $100.0 \%$ | $96.7 \%$ |
| Mean of the <br> percentage of flows <br> with GEH $<\mathbf{5}$ | $100.0 \%$ | $100.0 \%$ | $97.1 \%$ | $92.1 \%$ | $99.2 \%$ | $93.8 \%$ | $100.0 \%$ | $99.2 \%$ |
| Standard deviation <br> of the percentages of <br> flows with acceptable <br> error | $0.0 \%$ | $0.0 \%$ | $6.2 \%$ | $5.8 \%$ | $2.8 \%$ | $0.0 \%$ | $0.0 \%$ | $5.0 \%$ |
| Standard deviation <br> of the percentages of <br> flows with GEH $<5$ | $0.0 \%$ | $0.0 \%$ | $3.9 \%$ | $4.2 \%$ | $2.1 \%$ | $4.0 \%$ | $0.0 \%$ | $2.1 \%$ |

Table 6.3 Calibration Results of the Freeway Flow and Speed

|  | Flow | Speed |
| :---: | :---: | :---: |
| Mean of the percentages of flows with acceptable error | 93.4\% | NA |
| Mean of the percentages of flows with GEH $<5$ | 90.3\% | NA |
| Standard deviation of the percentages of flows with acceptable error | 1.4\% | NA |
| Standard deviation of the percentages of flows with $\mathbf{G E H}<5$ | 3.1\% | NA |
| Mean of the percentages of speeds with error $<5 \mathrm{mph}$ | NA | 89.7\% |
| Standard deviation of the percentages of speeds with error $<5 \mathrm{mph}$ | NA | 3.1\% |

### 6.3 CONTROL DESIGN

The goal of this study is to coordinate the control of two subsystems, namely on-ramp metering and intersection signal control at the adjacent intersection, with a proper strategy that can potentially improve overall traffic conditions. In this section, the control method will be explained in detail.

As mentioned in Section 6.2, the intersection congestion is usually caused by the freeway onramp queue spillback, which is due to the congestion on the freeway. Improving freeway traffic conditions can help to reduce intersection queue lengths. Therefore, it is necessary to have a ramp metering design that could prevent freeway breakdown and maximize throughput flow. However, the ramp metering rate cannot be too low to allow absolute priority to freeway traffic, as this would lead to queue propagation at the interchange, and even further, queue spillovers or blockage in the
arterial.
At the interchange, the westbound left-turn and eastbound right-turn are two major flows, and they usually have long queues. Hence, delay at these two movements is the major part of intersection overall delay. It is preferable that the signal intersection controller to provide long green durations to these two turning movements, if the on-ramp can accommodate this. By doing so, insufficient on-ramp feed flow can be avoided, on-ramp storage can be effectively used, and intersection queues can be reduced. As a result, intersection delay can be improved. Once the on-ramp becomes full, it is inappropriate to distribute long green to these two movements because vehicles cannot advance and green time will be wasted. In this case, the green durations for these two movements should be reduced and extra green should be given to other movements. In this sense, the signal control has to wisely distribute green times to all intersection movements, taking into account the condition at the on-ramp. The actuated signal control that is currently operating in the field fails to do this because it extends green as long as there is a vehicle actuating the detector (until max out), disregarding whether the on-ramp can take the vehicle or not.

The proposed control strategy has two parts, specifically ramp metering control and interchange signal control. The metering rate is updated every 30 seconds, at the same frequency as the realtime detection measurement. The ramp metering control adopted in this study is UP ALINEA with queue-overwrite. In interchange signal control, green durations for each movement are determined through the solution of an optimization problem, taking into account the on-ramp metering rate and queue length, in order to coordinate the interchange signal controller with the freeway on-ramp metering controller. Details about the two control strategies are given in the following subsections.

## On-ramp Metering

The following notations are used in this subsection to explain the metering control.
$k$ : time step index.
$r(k)$ : metering rate at the $k$-th interval, unit in veh/hr.
$\hat{o}$ : target occupancy, usually takes the value of critical occupancy.
$o_{\text {in }}(k)$ : freeway vehicle occupancy measured at upstream of on-ramp at the $k$-th interval, unit in percentage.
$o_{\text {out }}(k)$ : freeway vehicle occupancy measured at the merge area (downstream of on-ramp) at the $k$-th interval.
$\tilde{o}_{\text {out }}(k)$ : estimation of $o_{\text {out }}(k)$.
$K_{R}$ : controller integral action gain.
$q_{r}(k)$ : on-ramp flow at the $k$-th interval, unit in veh/hr.
$q_{\text {in }}(k)$ : freeway mainline flow measured at upstream of on-ramp at the $k$-th interval.
$\lambda_{i n}$ : number of lanes at upstream of on-ramp.
$\lambda_{\text {out }}$ : number of lanes at downstream of on-ramp.
$w$ : shock wave speed, unit in $\mathrm{km} / \mathrm{hr}$.
$\alpha, \gamma$ : tuning parameters.
$L$ : section length.

The metering algorithm used in this study is UP ALINEA. Details of this algorithm can be found in [113]. Eq. (6.2) shows how the metering rate is updated in the ALINEA algorithm. This update law is an integral control. $K_{R}$ is a positive value. If the measured freeway occupancy is lower than the target occupancy, the metering rate will be increased; otherwise, it will be decreased. The control update equation drives the freeway occupancy in the merging area toward a target value. The freeway should be operated at an optimal condition at the target value. Hence, it is usually selected to be the critical occupancy, which leads to maximum freeway throughput.

$$
\begin{equation*}
r(k)=r(k-1)+K_{R}\left[\hat{o}-o_{\text {out }}(k-1)\right] \tag{6.2}
\end{equation*}
$$

Since freeway detectors are placed upstream of the on-ramp at the studied site, occupancy in the merging area is not directly measurable. Therefore, ALINEA's extended version, UP ALINEA, is adopted here. UP ALINEA estimates the occupancy of the merging area (downstream of the onramp) from upstream occupancy and flows of the mainline and the on-ramp. If upstream occupancy is not greater than the critical occupancy ( $o_{i n}(k) \leq o_{c r}$ ), downstream occupancy can be estimated by Eq. 6.3).

$$
\begin{equation*}
\text { if } o_{\text {in }}(k) \leq o_{c r}, \quad \tilde{o}_{\text {out }}(k)=\alpha o_{\text {in }}(k)\left(1+\frac{q_{r}(k)}{q_{\text {in }}(k)}\right) \frac{\lambda_{\text {in }}}{\lambda_{\text {out }}} \tag{6.3}
\end{equation*}
$$

Otherwise, it will be estimated through Eqs. (6.4) and (6.5).

$$
\begin{align*}
& \text { if } o_{\text {in }}(k)>o_{\text {cr }}, \quad \tilde{o}_{\text {out }}(k)=\gamma \tilde{o}_{\text {out }}^{\prime}(k)+(1-\gamma) \tilde{o}_{\text {out }}^{\prime}(k-1)  \tag{6.4}\\
& \tilde{o}_{\text {out }}^{\prime}(k)=o_{\text {in }}(k) \frac{\lambda_{\text {in }}}{\lambda_{\text {out }}}+\frac{100 L}{w \lambda_{\text {out }}} q_{r}(k) \tag{6.5}
\end{align*}
$$

In this study, $\alpha=1, \gamma=0.2, w=-15 \mathrm{~km} / \mathrm{h}$ and $\hat{o}=25 \%$ are chosen.
Because a very low or very high metering rate is not practical, the calculated rate from UP ALINEA will be truncated if it is outside the range of 400-900 vehicles per hour per lane. In addition, it is necessary to release vehicles to avoid on-ramp queue spillback to the intersections. Therefore, if queue spillback is detected, the on-ramp metering rate computed by Eq. (6.3) will be overwrite, and a fixed metering rate of 700 vehicles per hour per lane will be used.

## Intersection Signal Optimization

The following notations are used in this subsection to explain the signal control design at the interchange.
$i$ : phase index.
$g_{i}(k)$ : the green length distributed to phase $i$, parameter to be designed for signal control, unit in seconds.
$G_{m i n, i}:$ minimum green length of phase $i$.
$R$ : the set of indices for phases directing to freeway (SR-87) SB on-ramp.
$\beta_{i}(k)$ : turning ratio to the SB on-ramp of phase $i$.
$R A(k)$ : the number of vehicles that the on-ramp of interest can accommodate, unit in number of vehicles.
$C$ : cycle length, unit in seconds.
$q_{i}(k)$ : queue length of phase $i$, unit in number of vehicles.
$q_{\text {on }}(k)$ : on-ramp queue length, unit in number of vehicles.
$Q_{o n}$ : maximum number of vehicles that the on-ramp can hold, unit in number of vehicles.
$d_{i}(k)$ : demand of phase $i$, unit in veh/sec.
$f_{\text {sat }, i}$ : saturation flow of phase $i$, unit in veh/sec.
$\mu_{i j}, v_{i j}, \delta$ : non-negative tuning parameters, which are used in optimization criterion given by Eq. (6.6).

The goal in signal control design is to determine the green durations $g_{i}$ 's in each cycle with two objectives. First, we want to balance the assigned green times of each cycle. The situation that one phase gets too much green length, while another gets too little, and consequently resulting in a long intersection queue, should not happen. Second, we want to effectively use the on-ramp storage space, which means we want to feed as much vehicle as possible to the on-ramp. As to the first objective, we will distribute the whole cycle to each phase according to its desired green time. Desired green time refers to the green time that a phase demands based on its present queue length and incoming vehicle volume. Given the cycle length $C$ and a phase $i$, demand rate $d_{i}(k)$, present queue $q_{i}(k)$, and discharge (saturation) flow rate $f_{\text {sat }, i}$, the signal desired green time $g_{d e s, i}$ can be calculated by $g_{\text {des }, i}(k)=\left(d_{i}(k) \cdot C+q_{i}(k)\right) / f_{\text {sat }, i}$. If we minimize the first term in objective function (6.6), with $\mu_{i j}=v_{i j}=1$, all the phases will have the same ratio between their respective assigned green length and desired green length. In this way, the demand and supply of each phase are balanced in the green distribution. As to the second objective, we will match the vehicle flow to the on-ramp and the on-ramp storage. In the second term of objective function 6.6), the term $\sum_{i \in R} f_{\text {sat }, i} \cdot \beta_{i} \cdot g_{i}(k)$ can be viewed as the maximum potential traffic volume into the on-ramp. The number of incoming vehicles that the on-ramp can accommodate, $R A(k)$, can be estimated by $R A(k)=Q_{o n}-q_{\text {on }}(k-1)+r(k) \dot{C}$. In this equation, $Q_{o n}$ is the maximum number of vehicles that the on-ramp can hold, which is determined by the length of the on-ramp. $q_{o n}(k-1)$ is the on-ramp
queue length (unit in number of vehicles) measured at the end of the $k-1$ interval $r(k)$ is the ramp metering rate computed from the ramp metering control algorithm, and $C$ is the intersection signal control cycle length. The second term in (6.6) penalizes the differences between the feeding volume and the available space at the on-ramp. By adding this term to the objective function, the signal control decision is made to maximize the use of available on-ramp space. Here, $\mu_{i j}$ and $v_{i j}$ are used to represent the priority of phases and $\delta$ is used to scale the second absolute value (unit in vehicles) to the level of the first term in (6.6), which is a ratio.

There are several practical constraints to be considered in the signal control design, which are respectively expressed by Eqs. (6.7) to (6.10). Constraint (6.7) is the minimum green of each phase. It is enforced for traffic safety. Meanwhile, Constraints (6.8) and (6.9) are two requirements related to the dual ring structure utilized by the signal at the interchange. Constraint (6.8) requires the green durations to add up to one cycle, while constraint 6.9) implies that the upper ring and the lower ring must cross the barrier at the same time. Constraint 6.10) limits the total number of vehicles advancing to the on-ramp such that it is no greater than the on-ramp available space. This constraint can help to avoid queue spillover. However, Constraints 6.7) and 6.10) cannot be satisfied at the same time if traffic is highly congested and the on-ramp is fully occupied. In this case, Constraint (6.10) will be dropped once it conflicts with Constraint 6.7).

$$
\begin{align*}
& \min _{g_{i}, i=1 \text { to } 8} \sum_{i=1 \text { to } 8, j=i \text { to } 8} \mu_{i j}\left|g_{i}(k) /\left(\frac{q_{i}(k)+d_{i}(k) \cdot C}{f_{\text {sat }, i}}\right)-v_{i j} \cdot g_{j}(k) /\left(\frac{q_{j}(k)+d_{j}(k) \cdot C}{f_{\text {sat }, j}}\right)\right| \\
& \quad+\delta\left|\sum_{i \in R} f_{\text {sat }, i} \cdot \beta_{i}(k) \cdot g_{i}(k)-R A(k)\right| \tag{6.6}
\end{align*}
$$

subject to

$$
\begin{gather*}
g_{i}(k) \geq G_{\text {min }}  \tag{6.7}\\
g_{1}(k)+g_{2}(k)=g_{5}(k)+g_{6}(k)  \tag{6.8}\\
\sum_{i=1 \sim 4} g_{i}(k)=C  \tag{6.9}\\
\text { or } i=5 \sim 8 \\
\sum_{i \in R} f_{\text {sat }, i} \cdot \beta_{i}(k) \cdot g_{i}(k) \leq R A(k) \tag{6.10}
\end{gather*}
$$

The discussion in the two paragraphs above assumes that the cycle length is known in advance. However, cycle length can be a decision variable in some signal control designs. In this study, cycle length can be freely chosen and time-varying, because the controlled intersection is not coordinated with other intersections. Therefore, cycle length needs to be selected before the optimization described above is applied. Here, we use a simple method to determine the cycle
length. Before a new cycle starts, the critical flow ratio from the demand of all phases is calculated. The computation of critical flow ratio is shown by Eq. (6.11). If the critical flow ratio is lower than 0.5 , it is believed that the intersection traffic demand is small and 70 seconds is selected as the cycle length. Otherwise, the intersection is considered to have a large traffic demand and 95 seconds will be selected instead. Threshold value of 0.5 , and two cycle length values of 70 seconds and 95 seconds, were chosen based on observed results of several simulation runs. In particular, the threshold was selected to be 0.5 because this value will not cause the cycle length to fluctuate frequently. In the conducted simulation runs, the cycle length is 95 seconds during the congested period, and a cycle length of 75 seconds occurs near the end of the simulation when the traffic demand drops and the system recovers to free-flow.

$$
\begin{equation*}
\text { critical flow ratio }=\sum_{i} \frac{d_{i}}{f_{s a t, i}} \quad \text { (phase } i \text { is in critical lane groups) } \tag{6.11}
\end{equation*}
$$

## Controller Implementation in the Simulation Runs

The control methods of freeway ramp metering and intersection signal control described above were implemented in simulation through the Aimsun Application Programming Interface (API). Fig. 6.6 shows the frame of implementation. Every 30 seconds, the API program collects measurements from detectors on the freeway and calculates the new ramp metering rate by the UP ALINEA algorithm. As mentioned in the subsection of on-ramp metering, if the metering rate is outside the allowable range, it will be truncated. If the queue at the on-ramp exceeds the limit, the computed metering rate will be replaced by the pre-defined metering rate to release the on-ramp queue. The updating of green splits is not executed at fixed intervals. At every step of the simulation, the API program checks if the intersection signal cycle terminates. If it does, the program collects the detection at the intersection and on-ramp. The new cycle length and green splits are subsequently computed according to the algorithm presented above and assigned to the intersection controller. The optimization in the signal control is a linear program, which can be solved very rapidly. As a result, it will not cause time delay problems in real time.


Figure 6.6 The Frame of the Implementation of the Coordinated Control Strategy Proposed for the Intersection of SR-87 and W. Taylor St

### 6.4 SIMULATION RESULTS

In this section, the proposed control strategy and the currently field-implemented control plans are compared through microscopic traffic simulation. The calibrated microscopic model described in Section 6.2 was used, and the proposed control was implemented in the simulation by the Aimsun API. Twenty simulation runs with different random seeds were conducted to obtain an average performance. Delay and total travel distance were chosen to compare the performance at the two control schemes. These quantities were collected in the simulation runs from individual movements at the interchange and the overall system.

Table 6.4 shows the simulation result. Columns 2-4 present the performance aggregated over 20 runs, and columns 5-7 present the standard deviation of the performance measures in 20 runs. For each entry, the value without parentheses is delay in hours, while the value inside the parentheses is travel distance in kilometers. It can be observed that the difference in travel distance was very small when the proposed control was applied, compared with the current field-implemented plans. This means there was no vehicle lost or withheld in the simulation when the proposed control was used. In the control strategy proposed in this chapter, total delay in the system was reduced by $1.37 \%$ when compared to the controller that is currently implemented. However, the intersection delay was reduced by $8.11 \%$, which is much more significant. Eastbound (EB) right-turn, WB left-turn, WB through and NB right-turn had significant reductions in delay. Even though SB leftturn showed a $17.05 \%$ increase in delay, the traffic flow for this movement was very small and the increase is acceptable as an absolute value. From this table, it can be seen that the delay of the movements with significant contributions to the total delay was decreased by the proposed control. This reduction overweighs the small delay increase of the movements with small contributions to the total delay.

The small delay reduction in the overall system and significant improvement at the interchange can be explained as follows. The freeway on-ramp bottleneck causes the major delay in the network. Given the high traffic demand in the freeway mainline and at the on-ramp, it is unlikely that the on-ramp metering controller can significantly reduce freeway congestion without causing onramp queue spillback into the interchange if ramp metering is the only control influencing freeway flow. The intersection delay only accounts for a small portion of the overall system delay due to its relatively small volume and short distance of the road that was modeled. Therefore, its significant reduction does not make a large change in the total delay. To get larger improvement in the overall system, the metering at multiple neighboring on-ramps need to be coordinated. However, this is beyond the project scope in this study.

Table 6.5 presents the maximum queue lengths recorded in the simulation. The queue lengths were collected for individual movements at the intersection and they were aggregated over lanes. The unit for queue length is number of vehicles. Columns 2-4 give the average values in 20 simulation runs and columns 5-7 present the standard deviation. The changes in the mean maximum queue lengths were not significant. In the simulation, queue spillover at WB left-turn was still observed in the proposed control method, but it was slightly better than that in the current
field-implemented control. This spillover is difficult to avoid because of the large traffic demand of this movement and the congestion on the on-ramp due to freeway congestion and the on-ramp metering. Queue spillovers for other movements did not appear in the simulation.
Table 6.4 Comparison of the Traffic Performance between the Current Control Plans and the Proposed Control Strategy

|  | Mean |  |  | Standard deviation |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Total delay <br> (travel <br> distance) in <br> current <br> control plans | Total delay <br> (travel <br> distance) in <br> proposed <br> control plans | Change in <br> delay (travel <br> distance) | Total delay <br> (travel <br> distance) in <br> current <br> control plans | Total delay <br> (travel <br> distance) in <br> proposed <br> control plans | Change in <br> delay (travel <br> distance) |
| EB Left | 2.52 | 2.62 | $3.83 \%$ | 0.06 | 0.09 | $4.22 \%$ |
| EB Through | $(58.13)$ | $(58.05)$ | $(-0.15 \%)$ | $(0.77)$ | $(0.77)$ | $(0.61 \%)$ |
| EB Right | $(254.30)$ | 18.95 | $(256.08)$ | 17.48 | $-0.23 \%$ | 0.21 |
|  | $(1.89 \mathrm{e}+03)$ | $(1.89 \mathrm{e}+03)$ | $(-0.09 \%)$ | $(3.72 \%)$ | $(1.49)$ | $(1.70)$ |

Table 6.5 Comparison of the Queue Lengths between the Current Control Plans and the Proposed Control Strategy

|  | Mean |  |  | Standard deviation |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Maximum <br> queue length <br> in current <br> control plans | Maximum <br> queue length <br> in proposed <br> control plans | Change in <br> maximum <br> queue length | Maximum <br> queue length <br> in current <br> control plans | Maximum <br> queue length <br> in proposed <br> control plans | Change in <br> maximum <br> queue length |
| EB Left | 4.67 | 4.67 | $0.00 \%$ | 0.58 | 0.47 | $17.29 \%$ |
| EB Through | 12.89 | 12.83 | $-0.43 \%$ | 1.20 | 0.69 | $9.25 \%$ |
| EB Right | 29.11 | 29.67 | $1.91 \%$ | 2.42 | 3.87 | $14.08 \%$ |
| WB Left | 45.06 | 44.89 | $-0.37 \%$ | 4.61 | 4.77 | $15.98 \%$ |
| WB Through | 9.83 | 8.83 | $-10.17 \%$ | 0.69 | 1.26 | $14.89 \%$ |
| SB Left | 5.56 | 6.00 | $8.00 \%$ | 1.01 | 1.05 | $24.30 \%$ |
| NB Left | 16.50 | 16.83 | $2.02 \%$ | 1.01 | 1.64 | $11.42 \%$ |
| NB Right | 18.22 | 15.28 | $-16.16 \%$ | 1.96 | 1.76 | $10.44 \%$ |

### 6.5 LESSONS LEARNED

Three points were learned when we conducted the study presented in this chapter.
First, the calibration of a microscopic simulation model is difficult, particularly when lane changing is frequent. Microscopic simulation models have many parameters that control the driver's behavior. This in one perspective allows flexibility in modeling. However, it is a timeconsuming work to configure all those parameters. In addition, the effect of a particular parameter in the simulation performance is not sufficiently straightforward. In macroscopic simulation models, the parameters of fundamental diagram directly influence the flow and density in the model. Hence users can calibrate the macroscopic models in a straightforward way. In contrast, the parameters used in microscopic models define the behavior of individual drivers. The overall system performance is a jointly effect of many microscopic parameters. The microscopic simulation calibration criteria are to match the simulation flow, speed, and other macroscopic performance index. Users need a great amount of work and experience to find out a good combination of parameter values. Furthermore, the lane-changing models in most of the existent macroscopic models are still insufficiently satisfactory, because they have difficulties in producing traffic situations of highly congestion and frequent lane-changing that can match the reality well. As in the study of this chapter, the simulated site was an on-ramp bottleneck that had a large on-ramp traffic volume. We spent a large portion of time on tuning the merging behavior in the simulation to make the calibration results to be in an acceptable error.

The second lesson learned in the study of this chapter is that the effect of local or sole ramp metering control is limited in some situations. As in the studied site of this chapter. The currently field-implemented ramp metering control plan failed to resolve the on-ramp bottleneck. The onramp queue spilled back to the intersections. With the proposed local coordinated control strategy, the intersection delay was reduced and the queue spillovers occurred less. However, the freeway was still congested and the overall system performance improved very marginally. The results of this study showed that coordinated control strategy of multiple control techniques (signal control, ramp metering, variable speed limits, and so on) and larger control scope (with more ramps involved) was needed to get further improvement.

The third lesson needed our attention is the problem of insufficient or inaccurate detection. In fact, this problem is not only encountered during the study in this chapter, but also caused research difficulties through all the work presented in this thesis. As already mentioned in Chapter 3 which was about arterial traffic modeling, getting all the necessary data for traffic modeling was challenging, and most of time we had to rely on estimation based on limited data. This led to the study presented in Chapter 4, which proposed an algorithm to estimate turning proportions at signalized intersection. In Chapter [5, we proposed a control strategy of VSL and coordinated ramp metering, and simulated this strategy using microscopic simulation. In the simulation we failed to use real traffic demand at network boundary and real split ratios at freeway junctions because the studied location did not have sufficient detectors to provide complete measurements. A better prediction of the benefit of the proposed control strategy can be obtained if we are able to get
real traffic data for the simulation. In this chapter, the studied location again did not give all the necessary data by on-site detectors, because some detectors were not installed at the right location within a road section/lane, or the data was not sufficiently accurate. However, we had the luck to get support from Caltrans (California Department of Transportation) to use the Miovision system to record the traffic on one day. Hence we can get a good model calibration. Furthermore, we chose the UP ALINEA instead of the basic ALINEA as our metering control algorithm because we did not have the occupancy detection at the merging area near the on-ramp. The trade-off of using sensors upstream of the on-ramp instead of at the merging area is that some estimation parameters in UP ALINEA must be well tuned in order to achieve reliable occupancy estimation. In both Chapter $[5$ and this chapter, our control strategies were restricted by the available sensors at the studied site in that we needed to limit the extra sensors required by the proposed control. If more sensors were allowed to use, there was more freedom to choose the control strategies. In general, we strongly feel through the work we conducted in this thesis that, if more detectors can be installed at the studied locations, we can have better knowledge of the traffic situation, get necessary traffic data to construct the simulation model, and have more choice to design control strategies.

### 6.6 CONCLUSION

In this chapter, a strategy of intersection traffic signal control algorithm and ramp metering has been developed to integrate two sub-systems, the freeway and the adjacent intersection. UP ALINEA with on-ramp queue-overwrite was chosen for the metering control. A signal optimization approach was utilized for green distribution, with the intention of balancing the green time for each movement according to its demand while taking into account the available on-ramp space. The proposed control was compared with the current field-implemented control plan through a wellcalibrated microscopic simulation. The simulation results showed that even though the overall system delay only decreased slightly, intersection delay had an $8 \%$ reduction on average.

## Chapter 7

## Evaluation of Impact of Adaptive Cruise

## Control and Cooperative Adaptive Cruise

## Control on Highway Capacity

### 7.1 INTRODUCTION

In Chapter 5 and 6, it was shown that traffic performance was affected by control strategies at the system operation level. It is also possible to improve traffic performance by adopting in-vehicle driving assistance tools. Vehicle manufacturers have developed, or are developing, various driving assistance systems, such as cruise control, lane keeping, and collision avoidance systems. Some of these systems can not only enhance driving comfort, but also have the potential to improve driving safety or traffic mobility. In this chapter, we discuss the influence of Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC) on freeway capacity and mobility.

## ACC and CACC Systems

ACC [114, 115] is a cruise control system embedded in vehicles for driving assistance. It is an extension of the conventional cruise control, which can only be commanded to follow a target speed. In contrast, ACC is able to automatically adjust both speed and driving distance. The ACC system relies on range sensors installed in the vehicle for real-time measurement. It does not require communication between a controlled vehicle and other facilities or vehicles on the road. CACC,
a further extension of the ACC system, controls a vehicle based on measurements obtained not only from on-board sensors, but also through vehicle-to-infrastructure (V2I) or vehicle-to-vehicle (V2V) communication. ACC and/or CACC are considered important component of intelligent cars and have received great attention from both researchers and vehicle manufacturers.

The operation of an ACC system can be described as follows. A radar or laser sensor is installed in the middle of the front apron of the control vehicle to measure its distance to its preceding vehicle. A controller collects measurements from the distance sensor and computes acceleration or deceleration based on the driver's selection of time gap and the control logic pre-configured in the controller. The maximum acceleration or deceleration rate is generally $2 \mathrm{~m} / \mathrm{s}^{2}$ ( $20 \%$ of gravitational acceleration). The engine's throttle and brake are automatically adjusted based on the computed acceleration or deceleration. Drivers can set their preferred time gaps in the instrument panel.

The ACC system has two control states, namely speed control and time gap control. If a vehicle is not detected in the front of the path of the controlled vehicle, or the distance between the front vehicle and controlled vehicle is large, the ACC operates in the speed control state. The speed control state is the same as the typical cruise control that has been available in the vehicles sold in the market. In this state, the ACC system regulates the vehicle speed to the target speed. On the other hand, if a front vehicle is present and the distance between two vehicles is short, the ACC system operates in time gap control state. In this state, the time gap between the controlled vehicle and the front vehicle is regulated to the user-defined time gap, under the restriction that vehicle speed will not exceed the user-defined maximum speed.

One critical problem of the ACC system is that it is hard to achieve string stability [116]. This means the disruption of the front vehicle speed will propagate increasingly in a platoon of vehicles under ACC control. String instability could lead to crashes involving multiple vehicles. To solve this problem, CACC with V2V communication is needed. The CACC system operates in almost the same way as the ACC system, except that it can obtain the information (position, speed, acceleration, etc.) of the front vehicle through V2V communication rather than by measuring it through the distance sensor used in the ACC system. The V2V communication is achieved by using dedicated short-range communications (DSRC), which enables vehicles to send out its own information to nearby vehicles and receive information from others within a certain range. After receiving information from nearby vehicles, the CACC control system needs to figure out the information from which vehicle should be used to determine its control action. Due to the usage of direct communication between vehicles, the speed and acceleration information used by the CACC system will be more accurate, and the time delay in obtaining the measurements will be shorter.

The maximum traffic flow in the roadway is a function of the time gaps that are maintained between successive vehicles. The time gaps are generally determined by drivers' choices of comfortable time gap and the time gaps allowed by the vehicles' control systems if there are any. In manual driving, there is no driving assistant system. Thus, the time gaps are only determined by the drivers' perception of what is safe and his or her reaction time. The driver's perception of safety is influenced by the driver's experiences, including expectations about the behaviors of other
drivers, particularly the driver of the front vehicle. Vehicles with ACC or CACC system usually have discrete time gap settings, which the driver can select based on his or her perceptions of the capabilities of the system. If drivers have sufficient confidence that ACC or CACC system can react faster and better than human, and select time gaps significantly smaller than the gaps they use in manual driving, it is possible to increase traffic throughput.

## Related Work

The research presented in this chapter is part of a project conducted at California Partners for Advanced Transportation TecHnology (California PATH). About 10 years ago, California PATH carried out a simulation study [117, 118] showing that the capacity could be increased to as high as 4,400 vehicles per hour per lane, if all vehicles on road were equipped with a CACC system and drivers were comfortable using this system at a time gap of 0.5 second. This result indicated that the CACC system had the potential to improve mobility and relieve congestion. However, a field study was needed to determine which time gap (seconds) the drivers would actually accept. As a consequence, California PATH undertook a project to collect drivers' preference in driving with the ACC or CACC system in live traffic. The results from this field test have been used for a new evaluation of capacity, as described in the remaining sections of this chapter.

The primary goal of the field test was to study the human factor in driving with ACC or CACC system. Two Infinity FX45 cars, originally equipped with an ACC system from the manufacturer, were retrofitted with the CACC system developed by California PATH. Sixteen drivers randomly selected from the general public drove the cars on a road, and were encouraged to choose their preferred time gaps for the ACC and CACC systems. Time gaps, together with many other parameters, were recorded during driving. The details and results from this test can be found in [119, 115, 120, 121].

This field test offered quantitative results on the time gap selections of the drivers involved in the test. These can be used to estimate the preferences of the driving population, and furthermore, the road capacity resulting from these selections. The micro-simulation discussed in the next few sections will use the control algorithm and recorded time gaps from the field test to evaluate the influence of the road capacity offered by the ACC or CACC system under such time gap selections.

In addition to California PATH's project of collecting drivers' preference and estimating its impact on road capacity, there have been other research projects from different organizations that have evaluated the effects of ACC or CACC systems on road capacity. Some early efforts were reviewed in [117, 118]. Research from the past 10 years can be found in [122, 123, 124, 125]. van Arem, Driel and Visser [122] used the MIXIC microscopic simulation to investigate the traffic throughput and stability impacts of CACC, incorporating good vehicle dynamics and driver behavior models. They studied a freeway lane drop as the disturbance to induce a shock wave to limit capacity, and found that the shock wave effect could be mitigated and the average speed increased with higher market penetrations of CACC. Their predictions of CACC effects are generally con-
sistent with the results that we will show here, with significant capacity increases at higher market penetration levels. Kesting, et. al [123, 124] simulated ACC with infrastructure-determined set speeds, a form of infrastructure-cooperative ACC rather than vehicle-cooperative ACC. Their intelligent driver model (IDM) model showed that a $25 \%$ market penetration of these ACC vehicles could eliminate congestion (and even 5\% could produce noticeable improvements in travel times) for the specific peak-period traffic scenario that they chose for simulation. However, this congestion benefit appears to be attributable to the variable speed limit strategy that they adopted rather than to the car-following dynamics of the ACC system. They did not directly consider the effects that ACC would have on the achievable capacity. Schakel, van Arem, and Netten [125] used a modified version of IDM (which they called IDM+) to explore the traffic flow stability implications of CACC, and an Acceleration Advice Controller (AAC), which advises the driver when to accelerate and decelerate rather than doing this automatically. They reported the results of a field experiment using 50 vehicles equipped with the AAC, showing reductions in variability of speeds and gaps between vehicles. Their CACC design was focused on improving traffic flow stability rather than on increasing capacity, therefore they did not directly address the capacity issue.

In this chapter, the maximum flow capacity of a road way is estimated by conducting a vehicle traffic flow microscopic simulation study using Aimsun. The flow capacity is estimated for different combinations of vehicles, manually driven, ACC and CACC. The parameters to specify the driving model of manually driven are obtained from previous calibration work conducted in [126, 127], and the parameters to specify the driving models of ACC and CACC are based on the field test mentioned above. This chapter is organized as follows: Section 7.2 describes the vehicle types used in the simulation; Section 7.3 presents the models for manual driving; Sections 7.4 presents the dynamics of the ACC and CACC models; Section 7.5 covers the construction of simulation; Section 7.6 shows the results of the simulation; and finally, Section 7.7 provides conclusions.

### 7.2 VEHICLE TYPES TO BE SIMULATED

The objective of this research is to estimate the impact of ACC/CACC vehicles on freeway capacity. It will evaluate the maximum traffic flow that can be achieved in stationary traffic conditions under different penetration rates of ACC/CACC vehicles. In this study, four types of vehicles are considered to accommodate all possible vehicle combinations. The driving characteristics of these vehicles are modeled to simulate the interactions between vehicles. These four types of vehicles are as follows:
(a) Manual vehicle: driven manually by a driver, with car following behavior represented by the oversaturated flow model [126, 127] (described in Section 7.3).
(b) Adaptive cruise control (ACC): car following is determined based on a simple first-order control law representing the behavior of a typical ACC system, with relatively slow, gentle re-
sponses to changes of the car ahead (described in Section 7.4).
(c) Here I am! (HIA): driven manually just like the manual vehicle, but it is equipped with a dedicated short-range communications (DSRC) radio that frequently broadcasts a "here I am" message giving its location and speed. If an HIA vehicle is being followed by a CACC vehicle, that CACC vehicle can use the CACC control capability.
(d) Cooperative adaptive cruise control (CACC): if it is following an HIA vehicle or another CACC vehicle, it can use its CACC car-following capability. If it is following a manual vehicle or an ACC vehicle, it acts like an ACC vehicle. The CACC car-following capability includes a faster response to changes of the car ahead and permits following at significantly shorter time gaps.

### 7.3 THE MANUAL DRIVING MODEL

The manual driving behavior model used in this study is the oversaturated freeway flow model developed by Yeo [126, 127]. This model was calibrated by two NGSIM freeway data sets, the I-80E data set which was collected on I-80E in Emeryville, California and the US-101 data set which was collected on US-101 in Los Angeles, California [7]. The calibration results showed that this model matched well with the drivers' behavior and could recover traffic conditions [128]. The calibrated parameters from [128] are used in this study. The NGSIM oversaturated freeway flow model comprises a car-following model and a lane-changing model. However, only the car-following model is used here, as the simulation does not cover lane changing or merging. The basic car-following model computes vehicle position and speed of the next simulation time step based on the positions and speeds of the studied vehicle and its front vehicle. The following notations are used to explain the manual driving model in this section.
$x_{n}^{U}(t+\Delta t)$ and $x_{n}^{L}(t+\Delta t)$ : the upper bound and lower bounds for the position of the $n$th vehicle at time $(t+\Delta t)(\mathrm{m})$.
$\tau$ : the wave travel time (seconds).
$g_{n}^{j a m}$ : the jam gap (m) for the $n$th vehicle.
$l:$ the length of the vehicle (m).
$v$ : the speed of the vehicle $(\mathrm{m} / \mathrm{s})$.
$x$ : the position of the vehicle (m).
$v^{f}$ : the free flow speed $(\mathrm{m} / \mathrm{s})$.
$\Delta t$ : the reaction time (seconds).
$a_{n}^{U}$ and $a_{n}^{L}$ : the maximum acceleration and deceleration of the $n$th vehicle.

The basic car-following model in oversaturated freeway flow model comes from Newell's linear model. Eqs. (7.1) to (7.5) describe its update formulas. In these equations, $\Delta t$ is the driver's reaction time, which represents the time a driver needs to react to changes. It can be viewed as
total time for perception, recognition, decision and response. Hence, the position and speed at current time $t$ will have an impact at time $t+\Delta t$. Eq. (7.2) computes the upper bound of position (can be viewed as the furthest position) that the studied vehicle can reach, which is the most restricted of four terms. The first term in the min operation is the location that the $n$ vehicle has to maintain if it comes to a stop. The second term is the location that the vehicle can reach if it fully accelerates. The third term is the location if the vehicle travels at free-flow (maximum) speed. The last term is the furthest location the vehicle can go without crashing into its front vehicle. Eq. (7.3) illustrates the lower bound of position (can be viewed as the nearest position) that the studied vehicle has to pass, which is the larger one of two terms. The first term in the max operation is the location that the vehicle will reach if it fully decelerates. The second term is the current position of the vehicle. This term is in the equation in order to prevent the vehicle from going backwards. Because a vehicle always advances as much as it can, its position in the next step will be the larger one of its upper and lower bound of position, which is expressed by Eq. (7.1).

$$
\begin{align*}
x_{n}(t+\Delta t) & =\max \left(x_{n}^{U}(t+\Delta t), x_{n}^{L}(t+\Delta t)\right)  \tag{7.1}\\
x_{n}^{U}(t+\Delta t) & =\min \binom{x_{n-1}\left(t+\Delta t-\tau_{n}\right)-l_{n-1}-g_{n}^{j a m}, x_{n}(t)+v_{n}(t) \Delta t+a_{n}^{U} \Delta t^{2},}{x_{n}(t)+v_{n}^{f} \Delta t, x_{n}(t)+\Delta x_{n}^{s}(t+\Delta t)}  \tag{7.2}\\
x_{n}^{L}(t+\Delta t) & =\max \left(x_{n}(t)+v_{n}(t) \Delta t+a_{n}^{L} \Delta t^{2}, x_{n}(t)\right)  \tag{7.3}\\
\Delta x_{n}^{s}(t+\Delta t) & =\Delta t\left(a_{n}^{L} \tau_{n}+\sqrt{\left(a_{n}^{L} \tau_{n}\right)^{2}-2 a_{n}^{L}\left(x_{n-1}(t)-x_{n}(t)-\left(l_{n=1}+g_{n}^{j a m}\right)+d_{n-1}(t)\right)}\right.  \tag{7.4}\\
d_{n-1}(t) & =\frac{v_{n-1}^{2}(t)}{2 a_{n-1}^{L}} \tag{7.5}
\end{align*}
$$

### 7.4 CONTROL ALGORITHMS FOR ACC AND CACC

## VEHICLES

The control algorithms for ACC and CACC vehicles used in this study are from the actual controls implemented in the field test vehicles. The ACC car-following rules are proprietary to Nissan, while the CACC car-following rules have been described in [129]. The complete rules consider many aspects for driving safety, cruise control efficiency and drivers' comfort. A full implementation of these rules can make the simulation computationally expensive, since these rules have to be executed at every simulation time step for each ACC/CACC vehicle. Some of these rules, like those related to drivers' comfort, have little influence on traffic flow dynamics, and hence they do
not need to be included in the simulation. Therefore, simplified but still accurate representations of these rules are needed. The following variables are used to define the car-following control algorithms for ACC or CACC vehicles in this section.
$v$ : the speed of the controlled ACC/CACC vehicle ( $\mathrm{m} / \mathrm{s}$ ).
$v_{d}$ : the desired speed set by the driver, or the speed limit of the road ( $\mathrm{m} / \mathrm{s}$ ).
$v_{l}$ : the speed of the leading (front) vehicle ( $\mathrm{m} / \mathrm{s}$ ).
$v_{e}$ : the speed error ( $\mathrm{m} / \mathrm{s}$ ).
$v_{r}$ : the relative speed ( $\mathrm{m} / \mathrm{s}$ ).
$a_{s c}$ : the acceleration by speed control $\left(\mathrm{m} / \mathrm{s}^{2}\right)$.
$a_{g c}$ : the acceleration by time gap control $\left(\mathrm{m} / \mathrm{s}^{2}\right)$.
$a$ : the acceleration adopted by the vehicles $\left(\mathrm{m} / \mathrm{s}^{2}\right)$.
$s$ : the spacing between the controlled vehicle and its leading (front) vehicle (meters).
$s_{d}$ : the desired spacing (meters).
$s_{e}$ : the spacing error (meters).
$h_{d}$ : the desired time gap (seconds).

ACC and CACC vehicles have very similar control algorithms, with the difference being in their desired time gaps and detection delay. There are two modes, speed control and gap control, in the ACC/CACC control algorithm. The goal of speed control is to keep the vehicle speed close to the speed limit, and that of gap control is to maintain the gap between the controlled vehicle and its front vehicle to be the desired gap. Speed control is activated when the spacing to the front vehicle in the same lane is larger than 120 meters, and gap control is activated when the spacing is smaller than 100 meters. If the spacing is between 100 and 120 meters, the controlled vehicle retains the previous control mode to provide hysteresis and avoid dithering between the two strategies.

In speed control, the control law is listed by Eqs. (7.6) to (7.8). This control law is an integral control with saturation. Eq. (7.6) computes the difference between vehicle speed and desired speed. If it is not zero, the acceleration computed by Eq. (7.7), an integral form of control, will drive the difference to zero. In this equation, the function $\operatorname{bound}()$ is defined as $\operatorname{bound}\left(x, x_{u b}, x_{l b}\right):=\max \left(\min \left(x, x_{u b}\right), x_{l b}\right)$. The values +2 and -2 are the maximum acceleration and deceleration of the vehicle under ACC/CACC control in units of $\mathrm{m} / \mathrm{s}^{2}$. Eq. (7.8) simply passes the acceleration calculated by speed control law to the vehicle. In summary, this control law tries to eliminate the error between the vehicle speed and the set speed if the vehicle is in the speed control mode.

$$
\begin{equation*}
v_{e}=v-v_{d} \tag{7.6}
\end{equation*}
$$

$$
\begin{align*}
a_{s c} & =\operatorname{bound}\left(-0.4 v_{e}, 2,-2\right)  \tag{7.7}\\
a & =a_{s c} \tag{7.8}
\end{align*}
$$

In gap control, the control law is described by Eqs. (7.10) to (7.16). Eqs. (7.10) and (7.11) are the same as Eqs. (7.6) and (7.7). Eq. (7.12) computes the desired distance between controlled vehicle and its front vehicle under the current speed and set time gap. Eq. (7.13) obtains the distance error, which is defined as the difference of current distance and desired distance. Eq. (7.14) calculates the relative speed between the controlled vehicle and its front vehicle. Once the distance error and relative speed are obtained, the acceleration of the time gap control is computed from Eq. (7.15). In this equation, +2 and -2 have the same meaning as in speed control. The influence of the term $0.25 s_{e}-v_{r}$ can be described as follows. If relative speed is zero (the controlled vehicle has the same speed as its front vehicle) and the distance error is positive (the current distance is larger than the desired distance), the acceleration of the time gap control is positive and the controlled vehicle will speed up to reduce the distance. If the relative speed is positive and distance error is zero, which means that the current distance meets the requirement but vehicles are approaching, the acceleration of the time gap control is negative and the controlled vehicle will slow down. Hence, the term $0.25 s_{e}-v_{r}$ considers both the current distance error and its tendency. Eq. (7.16) chooses the smaller of acceleration of speed control and that of time gap control as the designed acceleration of the vehicle. By this equation, the controlled vehicle is forced to approach its desired time gap set in gap control, but still obey the set speed in speed control.

$$
\begin{align*}
v_{e} & =v-v_{d}  \tag{7.10}\\
a_{s c} & =\operatorname{bound}\left(-0.4 v_{e}, 2,-2\right)  \tag{7.11}\\
s_{d} & =h_{d} * v  \tag{7.12}\\
s_{e} & =s-s_{d}  \tag{7.13}\\
v_{r} & =v-v_{l}  \tag{7.14}\\
a_{g c} & =\operatorname{bound}\left(0.25 s_{e}-v_{r}, a_{s} c,-2\right)  \tag{7.15}\\
a & =\min \left(a_{g} c, a_{s} c\right) \tag{7.16}
\end{align*}
$$

Although ACC and CACC have same control algorithm, they have differences in time gap selections and detection delay. The time gaps are discussed in the next section. The detection delay of ACC is larger than that of CACC, because ACC uses a distance sensor, while CACC
uses wireless communication. The minimum simulation step in this study is 0.1 second. CACC is assumed to have no detection delay in simulation and ACC has one step ( 0.1 second) delay.

### 7.5 SIMULATION CONSTRUCTION

In this study, a micro-simulation was conducted on the Aimsun platform, which is a transportation simulation environment developed by Transport Simulation Systems (TSS). An introduction to this simulation software can be found in Section 2.2. In the simulation study present in this chapter, we used the Aimsun microscopic simulator, application programming interface (API) and microsimulator software development kit (MicroSDK) to build the simulation.

## Simulation Network and Settings

This study aims to estimate the road capacity under different penetration rates of ACC or CACC vehicles. Road capacity should be measured in a stationary traffic condition. Because we only want to evaluate the impact of time gaps collected from the field test, and we do not want any lane changing or merging behavior to disturb the capacity estimation, only car-following was tested in the simulation. The simulated network was a one-lane straight freeway segment with the speed limit of $105 \mathrm{~km} / \mathrm{h}(65 \mathrm{mph})$. This freeway segment was 6.5 kilometers long and 3.5 meters wide, and there was a detector located 6 km from the entrance. This location was selected to ensure that all the flow measurements obtained are in steady state. The freeway was empty before the simulation starts. During the simulation, the entering of new vehicles was controlled by the algorithm written in the API file, which will be described in next subsection. The four types of vehicles mentioned in Section 7.2 were simulated, using the driver models presented in Sections 7.3 and 7.4 . The implementation of driver models can be found in the final subsection of the present section.

To evaluate the effects brought in by the different market penetration of each type of vehicle, different simulation scenarios have been defined to represent various combinations of manually driven, ACC, CACC and HIA vehicles. For each scenario, three simulations were run with different random number seeds. The results of these three simulation runs were aggregated to produce the estimates of achievable traffic flow.

In each simulation run, the total simulation length was 1 hour, and the simulation step was 0.1 second. Traffic flow was recorded at intervals of 5 minutes, but the first measurement was discarded because the first 5 minutes were viewed as the warm-up time. The capacity was the average flow over the remaining 55 minutes.

The four types of vehicles had the same physical characteristics in the simulation, with a length of 4.7 meters and width of 1.9 meters. The maximum acceleration was $2 \mathrm{~m} / \mathrm{s}^{2}$, and maximum deceleration was $-2 \mathrm{~m} / \mathrm{s}^{2}$.

## Vehicle Generation

In this study, the Aimsun API application was programmed to generate vehicles and record simulation data. The introduction to Aimsun API can be found in Section 2.2. During the simulation, the type of the next entering vehicle was randomly chosen, but followed the percentages defined in the simulation scenarios we wanted to test. The desired headway of the entering vehicle was also random. Note that there is a difference between the headway and the time gap. The time gap is the time a vehicle needs to travel the distance from its front bumper to the back bumper of the front vehicle. In contrast, the headway adds the time needed to travel the vehicle length. For manual driving, there is no explicit expression of headway or time gap defined in the model. In fact, the time gap is indirectly controlled by the variable $g_{n}^{j a m}$. From experience, we knew that the maximum flow for manually driven vehicles on this type of simple freeway link should be about $2,200 \mathrm{veh} / \mathrm{hr}$, hence we assumed the desired headway for manual driving was 1.64 seconds ( 3600 seconds/2200 vehicles per hour). For ACC or CACC vehicles, the time gap was a variable defined in the control rules, and we obtained its value in field test. Therefore, the desired time gaps of the ACC or CACC vehicles have been selected based on the field test results [119, 115, 120, 121]. The time gaps used and the corresponding percentages were as follows:

ACC: $31.1 \%$ at 2.2 seconds time gap, $18.5 \%$ at 1.6 seconds time gap, $50.4 \%$ at 1.1 seconds time gap.
CACC: $12 \%$ at 1.1 seconds time gap, $7 \%$ at 0.9 seconds time gap, $24 \%$ at 0.7 seconds time gap, $57 \%$ at 0.6 seconds time gap.

The desired entering headways for the ACC and CACC vehicles were chosen based on these time gaps, with the addition of the time increment to account for vehicle length. The gap for manual driving was selected randomly during the simulation, within a $+/-10 \%$ error range of 1.64 seconds, that is, from 1.48 to 1.8 seconds. At each simulation step, we checked the travel time from the entrance to the location of the last entering vehicle, based on the speed of that vehicle at that step. If this travel time was larger than the desired entering time gap, we let a new vehicle enter the freeway at the same speed as its front vehicle at that step. Using this algorithm, the vehicles enter the freeway at an interval and speed that did not generate a measured maximum flow lower than the real capacity due to insufficient demand, while preventing collisions associated with entering at too high a speed or too small a time gap. The frame of vehicle generation is illustrated in Fig. 7.1.


Figure 7.1 The Structure of the Implementation of the Vehicle Generation Algorithm in the

Because the entering time gap for the manually driven vehicles in the simulation usually did not match their desired time gap, the manually driven vehicles needed to adjust their speeds after they entered the freeway. This caused the vehicles following them, whether they were manually driven, ACC, or CACC, to need to adjust speeds as well. Therefore, vehicles at the entrance usually were drive at maximum speed (the speed limit), but they can gradually accelerate to maximum speed after they adjusted to their comfortable time gaps. Near the exit of the network, all vehicles were able to maintain their maximum speeds and preferred time gaps. In this sense, the capacity we measured did not represent a purely free-flow condition. It was the maximum flow that was achievable and stable in traffic with small disturbances upstream.

## Driver Model Implementation

The implementation of driver models in the simulation was carried out using the Aimsun MicroSDK. The introduction of this tool is presented in Section 2.2. The frame of implementation is illustrated in Fig. 7.2. The elements in the left and right rectangles are identical, except that the former is part of the ACC control model, while the latter is part of the CACC control model. At each simulation time step, the vehicle type of the vehicle that needed to be updated was initially identified. If the vehicle was not an ACC or CACC vehicle, it was updated according to the manually driven model, which is the oversaturated freeway flow model. Otherwise, it was updated by the ACC or CACC control algorithm, depending on the vehicle types of the vehicle to update and its front vehicle. A vehicle was updated by the CACC model only when it was a CACC vehicle and had a CACC or HIA vehicle in front. In the ACC and CACC models, the distance between the update vehicle and its front vehicle was checked to determine the control state. Speed control was chosen if the distance was larger than 120 meters, while gap control was selected if the distance was smaller than 100 meters. When the distance was between 100 and 120 meters, the vehicle used the previous control state.


Figure 7.2 The Frame of the Implementation of the Driving Models in the ACC/CACC Simulation

### 7.6 SIMULATION RESULTS

In this section we will describe the simulation results. Different combinations of traffic flow were simulated in this study to evaluate the maximum lane flow. As mentioned in previous section, the results present here are aggregated over three simulation runs with different random seeds.

The all-manually-driven case has already been referenced as a base case with a nominal capacity of 2,200 vehicles per hour per lane. When basic ACC vehicles were incorporated into the traffic stream, the achievable traffic flow appeared to be remarkably insensitive to the market penetration of ACC vehicles, as shown in Table 7.1 and Fig. 7.3. Note that the flow remains within the narrow range of 2,030 to 2,100 vehicles per hour per lane, regardless of the market penetration. This is because the driver preferences for ACC time gap settings were very similar to the time gaps that they adopt when they drive manually. According to the 2,200 vehicles per hour per lane capacity of the manually driven vehicles, the average headway of manually driven
vehicles is 1.636 seconds. The time needed to drive through the average vehicle length in this simulation study, which was 4.7 meters, was 0.141 second, if the vehicles were driving at the 65 miles per hour speed limit. This means the average time gap of the manually driven vehicles was 1.495 seconds ( $=1.636-0.141$ ). On the other side, according to the time gap selections of the ACC vehicles mentioned in previous section, the average time gap of the ACC vehicles was 1.534 seconds $(=31.1 \% \times 2.2+18.5 \% \times 1.6+50.4 \% \times 1.1)$. The results of this simulation refute the assumptions in some published papers that ACC could substantially increase highway capacity.

Flow


Figure 7.3 Highway Lane Capacity as a Function of Changes in ACC Market Penetration Relative to Manually Driven Vehicles, unit: veh/hr

Table 7.1 Highway Lane Capacity as a Function of Changes in ACC Market Penetration Relative to Manually Driven Vehicles, unit: veh/hr

|  | Flow (veh/hr) |
| :---: | :---: |
| $\mathbf{1 0 \%} \mathbf{A C C}$ | 2026 |
| $\mathbf{2 0 \%} \mathbf{A C C}$ | 2043 |
| $\mathbf{3 0 \%} \mathbf{A C C}$ | 2042 |
| $\mathbf{4 0 \%} \mathbf{A C C}$ | 2050 |
| $\mathbf{5 0 \%} \mathbf{A C C}$ | 2057 |
| $\mathbf{6 0 \%} \mathbf{A C C}$ | 2066 |
| $\mathbf{7 0 \%} \mathbf{A C C}$ | 2072 |
| $\mathbf{8 0 \%} \mathbf{A C C}$ | 2080 |
| $\mathbf{9 0 \%} \mathbf{A C C}$ | 2060 |
| $\mathbf{1 0 0 \%} \mathbf{A C C}$ | 2067 |

If we consider only the combinations of manually driven and CACC vehicles, the trend in highway lane capacity with respect to CACC market penetration is as shown in Table 7.2 and Fig. 7.4. This has an obvious quadratic shape. A CACC vehicle can only use its CACC capability when it is following another CACC or HIA vehicle, but when it is following a manual vehicle it must revert to conventional ACC control. As a result, the capacity grows very slowly until the CACC market penetration becomes substantial, and then grows much more rapidly. If all vehicles in a lane were equipped with CACC capability and the drivers chose the same distribution of CACC time gaps as they did in the field test, the lane capacity can increase to 3,970 vehicles per hour, representing a dramatic increment of road capacity.

Flow


Figure 7.4 Highway Lane Capacity as a Function of Changes in CACC Market Penetration
Relative to Manually Driven Vehicles, unit: veh/hr

Table 7.2 Highway Lane Capacity as a Function of Changes in CACC Market Penetration Relative to Manually Driven Vehicles, unit: veh/hr

|  | Flow (veh/hr) |
| :---: | :---: |
| $\mathbf{1 0 \%}$ CACC | 2036 |
| $\mathbf{2 0 \%}$ CACC | 2083 |
| $\mathbf{3 0 \%} \mathbf{C A C C}$ | 2136 |
| $\mathbf{4 0 \%} \mathbf{C A C C}$ | 2224 |
| $\mathbf{5 0 \%} \mathbf{C A C C}$ | 2340 |
| $\mathbf{6 0 \%} \mathbf{C A C C}$ | 2528 |
| $\mathbf{7 0 \%} \mathbf{C A C C}$ | 2687 |
| $\mathbf{8 0 \%} \mathbf{C A C C}$ | 2940 |
| $\mathbf{9 0 \%} \mathbf{C A C C}$ | 3373 |
| $\mathbf{1 0 0 \%} \mathbf{C A C C}$ | 3971 |

One of the strategies being proposed in the Connected Vehicles initiative to improve the performance of cooperative systems at low market penetrations is to equip as many existing vehicles as possible with a simple and inexpensive aftermarket positioning and communication on-board unit (OBU) that can broadcast an "Here I Am" (HIA) message. This message provides the basic global positioning system (GPS) coordinates and vehicle speed and heading information so that the OBUs on other vehicles can detect the trajectory of the vehicle. This information, if it is sufficiently accurate, would enable an HIA-equipped vehicle to be the leader for a CACC vehicle to follow at a short time gap. The effects of replacing the manually driven vehicles in the simulation cases of Fig. 7.4 with HIA vehicles are shown in Table 7.3 and Fig. 7.5. In this case, all the vehicles that did not have CACC were equipped with HIA devices and can therefore serve as leaders for the CACC vehicles. With this change, the quadratic growth in Fig. 7.4 becomes more nearly linear, and the capacity of the highway lane can be increased more significantly even at modest CACC market penetrations. At a $20 \%$ market penetration, the HIA addition increased capacity by $7 \%$; at $30 \%$ market penetration, it increased by more than $10 \%$; and in the $50 \%$ to $60 \%$ market penetration range, the increase was in the range of $15 \%$ compared to the cases without HIA devices.

Flow


Figure 7.5 Highway Lane Capacity as a Function of Changes in CACC Market Penetration Relative to "Here I Am" Vehicles, unit: veh/hr

Table 7.3 Highway Lane Capacity as a Function of Changes in CACC Market Penetration Relative to "Here I Am" Vehicles, unit: veh/hr

|  | Flow (veh/hr) |
| :---: | :---: |
| $\mathbf{1 0 \%} \mathbf{C A C C}$ | 2114 |
| $\mathbf{2 0 \%}$ CACC | 2231 |
| $\mathbf{3 0 \%} \mathbf{C A C C}$ | 2362 |
| $\mathbf{4 0 \%} \mathbf{C A C C}$ | 2505 |
| $\mathbf{5 0 \%} \mathbf{C A C C}$ | 2699 |
| $\mathbf{6 0 \%} \mathbf{C A C C}$ | 2904 |
| $\mathbf{7 0 \%} \mathbf{C A C C}$ | 3052 |
| $\mathbf{8 0 \%} \mathbf{C A C C}$ | 3304 |
| $\mathbf{9 0 \%} \mathbf{C A C C}$ | 3624 |
| $\mathbf{1 0 0 \%} \mathbf{C A C C}$ | 3971 |

In the scenarios above, combinations of only two types of vehicles were studied in the simulation. In the following scenarios, combinations of three types are considered. If traffic was a mixture of ACC, CACC, and manually driven vehicles, the simulation result based on the time gaps that drivers actually chose in the field test is as shown in Fig. 7.6 and Table 7.4. The pattern of capacity increment is similar to the case of manually driven and CACC vehicles, which could be explained by the observation in Fig. 7.3 that ACC vehicles do not significantly change road capacity.


Figure 7.6 Prediction of Lane Capacity Effects of ACC and CACC Vehicles (With the Remaining Vehicles Manually Driven, unit: veh/hr)

Table 7.4 Prediction of Lane Capacity Effects of ACC and CACC Vehicles (With the Remaining Vehicles Manually Driven, unit: veh/hr)

|  |  | Percentage of CACC Vehicles |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 10\% | 20\% | 30\% | 40\% | 50\% | 60\% | 70\% | 80\% | 90\% |
|  | 10\% | 2065 | 2090 | 2170 | 2265 | 2389 | 2458 | 2662 | 2963 | 3389 |
|  | 20\% | 2065 | 2110 | 2179 | 2265 | 2378 | 2456 | 2671 | 2977 | - |
|  | 30\% | 2077 | 2127 | 2179 | 2269 | 2384 | 2487 | 2710 | - | - |
|  | 40\% | 2088 | 2128 | 2192 | 2273 | 2314 | 2522 | - | - | - |
|  | 50\% | 2095 | 2133 | 2188 | 2230 | 2365 | - | - | - | - |
|  | 60\% | 2101 | 2138 | 2136 | 2231 | - | - | - | - | - |
|  | 70\% | 2110 | 2084 | 2155 | - | - | - | - | - | - |
|  | 80\% | 2087 | 2101 | - | - | - | - | - | - | - |
|  | 90\% | 2068 | - | - | - | - | - | - | - | - |

The capacity effects of different combination of CACC vehicles and HIA vehicles (with the rest being manually driven) are shown in Fig. 7.7 and Table 7.5. As the market penetration of CACC increases, the increasing capacity attributable to the additional HIA vehicles can be seen, but this is a relatively subtle effect. For completeness, the analogous results for different combinations of CACC vehicles and HIA vehicles (with the rest being conventional ACC vehicles) are shown in Fig. 7.8 and Table 7.6. Since the effects on capacity of ACC and manually driven vehicles are very similar, these results do not differ much from the previous results.


Figure 7.7 Prediction of Lane Capacity Effects of HIA and CACC Vehicles (With the Remaining Vehicles Manually Driven, unit: veh/hr)

Table 7.5 Prediction of Lane Capacity Effects of HIA and CACC Vehicles (With the Remaining Vehicles Manually Driven, unit: veh/hr)

|  | Percentage of CACC Vehicles |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Percentage of HIA Vehicles |  | 10\% | 20\% | 30\% | 40\% | 50\% | 60\% | 70\% | 80\% | 90\% |
|  | 10\% | 2045 | 2110 | 2179 | 2288 | 2447 | 2576 | 2760 | 3111 | 3624 |
|  | 20\% | 2054 | 2125 | 2211 | 2323 | 2512 | 2671 | 2893 | 3303 | - |
|  | 30\% | 2064 | 2148 | 2246 | 2378 | 2519 | 2787 | 3041 | - | - |
|  | 40\% | 2073 | 2165 | 2282 | 2434 | 2611 | 2891 | - | - | - |
|  | 50\% | 2084 | 2187 | 2318 | 2503 | 2685 | - | - | - | - |
|  | 60\% | 2097 | 2206 | 2362 | 2545 | - | - | - | - | - |
|  | 70\% | 2102 | 2227 | 2395 | - | - | - | - | - | - |
|  | 80\% | 2114 | 2252 | - | - | - | - | - | - | - |
|  | $\mathbf{9 0 \%}$ | 2123 | - | - | - | - | - | - | - | - |



Figure 7.8 Prediction of Lane Capacity Effects of HIA and CACC Vehicles (With the Remaining Vehicles Being ACC, unit: veh/hr)

Table 7.6 Prediction of Lane Capacity Effects of HIA and CACC Vehicles (With the Remaining Vehicles Being ACC, unit: veh/hr)

|  | Percentage of CACC Vehicles |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 10\% | 20\% | 30\% | 40\% | 50\% | 60\% | 70\% | 80\% | 90\% |
|  | 10\% | 2086 | 2132 | 2168 | 2278 | 2443 | 2567 | 2831 | 3108 | 3624 |
|  | 20\% | 2135 | 2164 | 2207 | 2366 | 2446 | 2669 | 2941 | 3303 | - |
|  | 30\% | 2137 | 2193 | 2291 | 2364 | 2533 | 2775 | 3041 | - | - |
|  | 40\% | 2128 | 2206 | 2302 | 2439 | 2588 | 2891 | - | - | - |
|  | 50\% | 2139 | 2220 | 2324 | 2499 | 2685 | - | - | - | - |
|  | 60\% | 2134 | 2239 | 2373 | 2545 | - | - | - | - | - |
|  | 70\% | 2137 | 2245 | 2395 | - | - | - | - | - | - |
|  | 80\% | 2132 | 2252 | - | - | - | - | - | - | - |
|  | 90\% | 2123 | - | - | - | - | - | - | - | - |

### 7.7 CONCLUSION

This chapter reported the predictions of the effects of ACC and CACC on highway lane capacity. The predictions were based on real experimental data from drivers who had driven suitably equipped vehicles and selected time gap settings with which they were comfortable. These results show that conventional ACC is unlikely to produce any significant change in the capacity of highways, but CACC has the potential to substantially increase highway capacity when it reaches a moderate to high market penetration.

These results showed a maximum lane capacity of about 4,000 vehicles per hour if all vehicles were equipped with CACC. If the vehicle population consisted of CACC and HIA vehicles, meaning that all vehicles had been equipped with DSRC radios, the lane capacity increased approximately linearly from 2,000 to 4,000 as the percentage of CACC vehicles increased from zero
to one hundred. On the other hand, if the vehicle population consisted of manual and CACC vehicles, without any mandate for non-CACC vehicles to be equipped with DSRC, the increase in lane capacity followed a quadratic profile, lagging significantly behind at the intermediate market penetration values. Therefore, the capacity benefits of CACC can be accelerated, or obtained at somewhat lower market penetrations, if the rest of the vehicle population is equipped with HIA devices so that they can serve as the front vehicles for CACC vehicles.

## Chapter 8

## Conclusion and Future work

This dissertation has investigated specific topics in traffic modeling, estimation, and control. It studied the problem of simulating urban street traffic using the Link-Node Cell Transmission Model (LN-CTM), proposed a method to estimate turning proportions at intersections, developed control strategies for traffic congestion, and evaluated the impact of Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC) on road capacity.

In Chapter 2, existing research and programs related to traffic data collection and traffic modeling were introduced. In the first part of this chapter, we listed several detection technologies commonly utilized. The accuracy and disadvantages of each detection technology were presented. Two data collection programs, the Performance Measurement System (PeMS) and Next Generation Simulation (NGSIM), were described. These were the data source for the research in the later chapters. In the second part of Chapter 2, we looked at both macroscopic and microscopic traffic models and their differences. For macroscopic models, the Lighthill-Whitham-Richards (LWR) Model and the Cell Transmission Model (CTM) were explained in detail. In terms of microscopic models, the simulation software Aimsun was introduced, as this was used several times in this dissertation.

In Chapter 3, we simulated arterial traffic using the Link-Node Cell Transmission Model and evaluated its accuracy in simulating an arterial road segment and comparing the simulation results with real traffic data, the NGSIM data. We used this simulation because there was insufficient research to demonstrate how accurately the Cell Transmission Model and its extensions can be applied to urban street traffic, although they have been utilized by many researchers. In this chapter, we first introduced the link-node cell transmission model and reviewed related work on the cell transmission model and its extensions. The link-node cell transmission model is consistent with the cell transmission model, with modifications to allow more general network topologies. The data we used in the simulation were the Lankershim data-set from the NGSIM program. This fieldcollected data was presented in high resolution and with rich supporting files, and consequently
allowed us to construct the simulation with detailed inputs and to carry out a solid analysis of the result. The simulation results showed that the accuracy of the model is acceptable. The average the northbound and southbound directions flow errors were $9.41 \%$ and $8.44 \%$, respectively. The travel time errors were a bit larger at $19.83 \%$ and $30.16 \%$ for the northbound and southbound directions, respectively.

In Chapter 4, we developed a recursive method to estimate turning proportions at intersections. Knowledge of turning proportions is often required by traffic simulation or control design. The proposed method only used departure detectors, together with signal information. As a result, it requires half of the number of detectors that conventional methods utilize. The estimation problem was formulated as a least squares problem, with linear equations and boundary constraints. An algorithm that can solve the problem recursively was described. We also proposed incorporating a forgetting factor and covariance resetting in the least squares estimation algorithm for the case of time-varying turning proportions. Two simulation scenarios, one with static turning proportions and the other with time-varying proportions, were used to compare the proposed method with an optimization method. The simulation results showed that, in the scenario with static turning proportions, the proposed method can obtain as accurate estimates as the optimization method, but only requiring $3 \%$ computational effort of the latter. In the scenario of time-varying turning proportions, the proposed method with forgetting factor and covariance resetting was able to trace the change of turning proportions. Even though it had slightly larger error than the optimization method, it took significantly less computation effort.

In Chapter 5 , we proposed a control strategy that combines variable speed limits and ramp metering to address freeway congestion. The studied freeway segment in this chapter was a segment of the I-80W between Richmond and Emeryville in the east of San Francisco Bay Area, where congestion is caused by weaving. In the first step of this strategy, the speed limits were calculated. In the most downstream segment, we used a critical speed limit to create a discharge section in order to facilitate lane changing. For the speed limits in upstream segments, we gradually decreased the speed limits from the free-flow speed to the critical speed limit, balancing the congestion in each segment. After the speed limits were determined, we adopted a predictive control model to compute the ramp metering rates, with the objective of minimizing total time spent and maximizing total travel distance. The optimization is a linear program and could be computed very rapidly. We tested the proposed strategy using calibrated microscopic simulation. The results from different scenarios demonstrated the following: 1) if only ramp metering was adopted, the resulting delay reduction is marginal; 2 ) if variable speed limits were adopted, the resulting delay reduction became significant; and 3) $30 \%$ driver compliance to speed limits exhibited similar improvements to $100 \%$ driver compliance.

In Chapter 6, we presented a control strategy for freeway ramp metering and signal control at the adjacent intersection. The objective of this control strategy is to integrate the controls of two sub-systems, namely the freeway and the intersection. The studied site for this research was the interchange of SR-87 and W. Taylor Street in San Jose, California. In the control strategy, upstream ALINEA with queue-overwrite was utilized to maximize the freeway throughput. The intersection
signal was calculated through a linear program that balances the green time of each movement according to its demand and tries to prevent queue spillovers at the on-ramp. We compared the performance of the proposed strategy and that of the current field-implemented control plans in a well-calibrated microscopic simulation. The simulation results showed that the delay at the intersection was reduced by $8 \%$ using the proposed method, assuming that queue length in both the roadways and the freeway on-ramp could be measured.

In Chapter 7, we estimated the influence of Adaptive Cruise Control and Cooperative Adaptive Cruise Control on road capacity. This study used results obtained from an previous project that collected drivers' time gap selections when they used Adaptive Cruise Control (ACC) or Cooperative Adaptive Cruise Control (CACC) on-road capacity. We evaluated the change in road capacity with different ACC and CACC penetration rates through microscopic simulation. The behavioral models that represented manual driving, Adaptive Cruise Control and Cooperative Adaptive Cruise Control were programmed in the simulation. The simulation predicted the following: 1) Conventional Adaptive Cruise Control would not change the road capacity significantly; 2) Cooperative Adaptive Cruise Control was able to increase capacity substantially when its market penetration became moderate or high, and the capacity can reach about 4000 vehicles per hour per lane if all vehicles were equipped with CACC; and 3) the HIA devices can accelerate the increase in capacity when the CACC penetration rate was low.

One issue that should be noted through this dissertation is that the lack of sufficient and reliable detection seriously curtails the performance and safety gains that can be potentially harvest through the use any of the traffic management and estimation techniques presented in this dissertation. From Chapter 3 to Chapter 6, we have seen various difficulties in obtaining the necessary traffic data to adequately implement the proposed traffic management strategies: no sensors installed at the required location, sensors not being well placed within a lane, inaccurate measurement, and insufficient detection resolution. It can be said that the necessary traffic data remains one of the largest challenges in transportation management. In Chapter 3, where we evaluated the link-node cell transmission model, we could only find one data set that can give all the required data for the simulation, the NGSIM data. However, this data set only captured a very short time period at a small scope of network, which did not allow us to do a more comprehensive simulation. In most of the intersections we see in practice, the measurements we can get are not enough for arterial traffic simulation. In Chapter 4 , we presented an algorithm to estimate turning proportions. However, the proposed algorithm, together with other algorithms reviewed in that chapter, is hard to deploy in reality because most of the signalized intersections do not have the required sensors, or the measurement resolution is low. In both Chapter 5 and Chapter 6, we needed to restrict the choices of control strategies because the deployment of extra sensors was not possible by the requirements of the research projects. If we had the freedom to install new sensors, there would have been less restriction in control strategy design, and it would have been possible to implement a more effective control strategy. In Chapter 5, we were not able to get an accurate estimate of the benefit of the proposed control strategy using at the studied site because we failed to obtain real traffic demand for model construction. In Chapter 6, the on-site sensors were not able to provide the adequate traffic data for modeling because the measurements were not sufficiently
accurate. In addition, the best measurement of the traffic congestion near an on-ramp should be measured at the merging area. However, we have to estimate downstream congestion because only detectors installed upstream of the on-ramp were available. Generally, if more detectors with reliable measurements are available, we can achieve better traffic modeling, less dependence on reliable traffic estimation, and larger freedom to choose an effective traffic control strategy.

## Future Work

There are multiple possible directions for future research on the topics covered in this dissertation. First, for the simulation of urban street traffic discussed in Chapter 3, we can consider the following: 1) enhancing the model with the capability to model right-turn-on-red and permissive left-turn; 2) comparing different traffic models with the same data to find out which is the most accurate or the conditions under which one model is better than the others; 3) testing the sensitivity of the Cell Transmission Model to uncertainty. Second, it would be worthwhile to evaluate the estimation method presented in Chapter 4 with field-collected data. In addition, since departure flows and turning proportions can be obtained from the detector configuration used in Chapter 4 , we could think about estimating other traffic states, such as queue length and travel delay, and propose a control method based on the estimates. Third, for the control strategies proposed in Chapters 5 and 6, field tests are needed to further demonstrate their effects on traffic. Moreover, a control strategy that includes variable speed limits, ramp metering, signal control, and probably other factors should be considered to further coordinate the freeway system and arterial system. Finally, for the study in Chapter 7, we can extend it to include the scenario of lane changing, in order to obtain more complete predictions regarding the impact of ACC or CACC.

## Bibliography

[1] David Schrank, Bill Eisele, and Tim Lomax. Ttis 2012 urban mobility report. Texas A\&M Transportation Institute. The Texas A\&M University System, 2012.
[2] California Department of Transportation. Caltrans pems, 2013. http://pems.dot. ca.gov/, accessed Noverber 10, 2013.
[3] Highway Capacity Manual. Transportation research board. National Research Council, Washington, DC, 113, 2000.
[4] Fred L. Hall. Traffic stream characteristics. Traffic Flow Theory. US Federal Highway Administration, 1996.
[5] Luz Elena Y. Mimbela and Lawrence A. Klein. Summary of vehicle detection and surveillance technologies used in intelligent transportation systems. 2000.
[6] Brian Sowers Kevin Aguigui. Detector technology, 2013. http://www.mtc.ca. gov/services/arterial_operations/downloads/3-1-10/Detector_ Technology-Kimley-Horn_Associates.pdf, accessed Noverber 10, 2013.
[7] Federal Highway Administration. Next generation simulation (ngsim), 2013. http:// ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm, accessed July 27, 2013.
[8] Serge P. Hoogendoorn and Piet H.L. Bovy. State-of-the-art of vehicular traffic flow modelling. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 215(4):283-303, 2001.
[9] Carlos F Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B: Methodological, 28(4):269-287, 1994.
[10] Carlos F. Daganzo. The cell transmission model, part ii: network traffic. Transportation Research Part B: Methodological, 29(2):79-93, 1995.
[11] Apostolos Kotsialos, Markos Papageorgiou, Christina Diakaki, Yannis Pavlis, and Frans Middelham. Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool metanet. Intelligent Transportation Systems, IEEE Transactions on, 3(4):282-292, 2002.
[12] H.M. Zhang, Yu Nie, and Zhen Qian. Modelling network flow with and without link interactions: the cases of point queue, spatial queue and cell transmission model. Transportmetrica B: Transport Dynamics, 1(1):33-51, 2013.
[13] Charles E. Wallace, K.G. Courage, D.P. Reaves, G.W. Schoene, and G.W. Euler. Transyt-7f user's manual. Technical report, 1984.
[14] Denos C. Gazis and Renfrey B. Potts. The oversaturated intersection. Technical report, 1963.
[15] Michael J. Lighthill and Gerald Beresford Whitham. On kinematic waves. ii. a theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 229(1178):317-345, 1955.
[16] Paul I. Richards. Shock waves on the highway. Operations research, 4(1):42-51, 1956.
[17] Monograph on traffic flow theory. Federal Highway Administration, 1997. http:// www.fhwa.dot.gov/publications/research/operations/tft/, accessed Noverber 10, 2013.
[18] Gabriel Clemente Gomes. Optimization and Microsimulation of On-ramp Metering for Congested Freeways. PhD thesis, UNIVERSITY OF CALIFORNIA, BERKELEY, 2004.
[19] J.P. Lebacque. The godunov scheme and what it means for first order traffic flow models. In Internaional symposium on transportation and traffic theory, pages 647-677, 1996.
[20] Quadstone Paramics Ltd. Quadstone paramics, 2013. http://www. paramics-online.com/, accessed Noverber 10, 2013.
[21] PTV Group. Ptv vissim, 2013. http://vision-traffic.ptvgroup.com/ en-us/products/ptv-vissim/, accessed Noverber 10, 2013.
[22] TSS. Aimsun, 2013. http://www.aimsun.com/wp/, accessed Noverber 10, 2013.
[23] Trafficware. Planning \& analysis software, 2013. http://www.trafficware.com/ products/planninganalysis-software/, accessed Noverber 10, 2013.
[24] McTrans. Corsim, 2013.http://www-mctrans.ce.ufl.edu/featured/TSIS/ Version5/corsim.htm, accessed Noverber 10, 2013.
[25] Steven L. Jones, Andrew J. Sullivan, Naveen Cheekoti, Michael D. Anderson, and D. Malave. Traffic simulation software comparison study, volume 2217. University Transportation Center for Alabama, 2004.
[26] Loren Bloomberg and Jim Dale. Comparison of vissim and corsim traffic simulation models on a congested network. Transportation Research Record: Journal of the Transportation Research Board, 1727(1):52-60, 2000.
[27] Transport Simulation Systems. Aimsun 7 Users' Manual, May 2012.
[28] Transport Simulation Systems. Aimsun 7 Dynamic Simulators Users Manual, October 2012.
[29] Transport Simulation Systems. Aimsun 7 Microsimulator API Manual, October 2012.
[30] Transport Simulation Systems. Aimsun Microscopic Simulator Behavioural Models SDK for Aimsun 7, November 2011.
[31] Alexandr A. Kurzhanskiy. Modeling and software tools for freeway operational planning. PhD thesis, University of California, Berkeley, 2007.
[32] Wei-Hua Lin and Dike Ahanotu. Validating the basic cell transmission model on a single freeway link. PATH technical note; 95-3, 1995.
[33] Laura Muñoz, Xiaotian Sun, Dengfeng Sun, Gabriel Gomes, and Roberto Horowitz. Methodological calibration of the cell transmission model. In American Control Conference, 2004. Proceedings of the 2004, volume 1, pages 798-803. IEEE, 2004.
[34] Ajith Muralidharan, Gunes Dervisoglu, and Roberto Horowitz. Freeway traffic flow simulation using the link node cell transmission model. In American Control Conference, 2009. ACC'09., pages 2916-2921. IEEE, 2009.
[35] Athanasios K Ziliaskopoulos. A linear programming model for the single destination system optimum dynamic traffic assignment problem. Transportation science, 34(1):37-49, 2000.
[36] Hina Golani and S. Travis Waller. Combinatorial approach for multiple-destination user optimal dynamic traffic assignment. Transportation Research Record: Journal of the Transportation Research Board, 1882(1):70-78, 2004.
[37] S. Travis Waller and Athanasios K. Ziliaskopoulos. Stochastic dynamic network design problem. Transportation Research Record: Journal of the Transportation Research Board, 1771(1):106-113, 2001.
[38] Essam Almasri and Bernhard Friedrich. Online offset optimisation in urban networks based on cell transmission model. ITS Hanover, 2005.
[39] Hong K. Lo. A novel traffic signal control formulation. Transportation Research Part A: Policy and Practice, 33(6):433-448, 1999.
[40] Satish V. Ukkusuri, Gitakrishnan Ramadurai, and Gopal Patil. A robust transportation signal control problem accounting for traffic dynamics. Computers \& Operations Research, 37(5):869-879, 2010.
[41] TRANSYT 13 User Guide.
[42] TOPL. Topl software, 2012. http://gateway.path.berkeley.edu/topl/ software.html, accessed Noverber 10, 2013.
[43] TOPL. Tools for operations planning website, 2012. http://gateway.path. berkeley.edu/topl/, accessed Noverber 10, 2013.
[44] Ajith Muralidharan. Tools for modeling and control of freeway networks. Master's thesis, University of California, Berkeley.
[45] Luis G. Willumsen. Estimation of an od matrix from traffic counts-a review. 1978.
[46] M. Cremer and H. Keller. A new class of dynamic methods for the identification of origindestination flows. Transportation Research Part B: Methodological, 21(2):117-132, 1987.
[47] Nancy L. Nihan and Gary A. Davis. Recursive estimation of origin-destination matrices from input/output counts. Transportation Research Part B: Methodological, 21(2):149-163, 1987.
[48] Michael G.H. Bell. The estimation of origin-destination matrices by constrained generalised least squares. Transportation Research Part B: Methodological, 25(1):13-22, 1991.
[49] Baibing Li and Bart De Moor. Recursive estimation based on the equality-constrained optimization for intersection origin-destination matrices. Transportation Research Part B: Methodological, 33(3):203-214, 1999.
[50] Chang-Jen Lan and Gary A. Davis. Real-time estimation of turning movement proportions from partial counts on urban networks. Transportation Research Part C: Emerging Technologies, 7(5):305-327, 1999.
[51] Chang-Jen Lan. Sufficiency of detector information under incomplete configuration for intersection od estimation. In Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE, pages 398-403. IEEE, 2001.
[52] Chang-Jen Lan and Gary A. Davis. Real-time estimation of turning movement proportions from partial counts on urban networks. Transportation Research Part C: Emerging Technologies, 7(5):305-327, 1999.
[53] Baibing Li and Bart De Moor. Dynamic identification of origin-destination matrices in the presence of incomplete observations. Transportation Research Part B: Methodological, 36(1):37-57, 2002.
[54] Gang-Len Chang and Xianding Tao. Estimation of time-dependent turning fractions at signalized intersections. Transportation Research Record: Journal of the Transportation Research Board, 1644(1):142-149, 1998.
[55] Svante Gunnarsson. Combining tracking and regularization in recursive least squares identification. In Decision and Control, 1996., Proceedings of the 35th IEEE, volume 3, pages 2551-2552. IEEE, 1996.
[56] Mario E. Salgado, Graham C. Goodwin, and Richard H. Middleton. Modified least squares algorithm incorporating exponential resetting and forgetting. International Journal of Control, 47(2):477-491, 1988.
[57] Lisa Nelson Doug Sallman Leslie Jacobson, Jason Stribiak. Ramp Management and Control Handbook. Federal Highway Administration, 2006. http://www.fhwa.dot.gov/ publications/research/operations/tft/, accessed Noverber 10, 2013.
[58] Seungmin Kang and David W. Gillen. Assessing the benefits and costs of intelligent transportation systems: ramp meters. California PATH Program, Institute of Transportation Studies, University of California at Berkeley, 1999.
[59] E.D. Arnold Jr. Ramp metering: a review of the literature. Contract, 21535:6940, 1998.
[60] The Kansas and Missouri Departments of Transportation. Maximizing the flow, ramp metering 2011 evaluation report, 2011.
[61] Cambridge Systematics. Twin Cities Ramp Meter Evaluation. Cambridge Systematics, Incorporated, 2001.
[62] FHWA Operations. Ramp metering, 2013. http://www.ops.fhwa.dot.gov/bn/ resources/rampmetering.htm, accessed Sep 4, 2013.
[63] Michael Zhang, Taewan Kim, Xiaojian Nie, Wenlong Jin, Lianyu Chu, and Will Recker. Evaluation of on-ramp control algorithms. 2001.
[64] Markos Papageorgiou, Habib Hadj-Salem, and F. Middelham. Alinea local ramp metering: Summary of field results. Transportation Research Record: Journal of the Transportation Research Board, 1603(1):90-98, 1997.
[65] Doug Lau. Minnesota department of transportation: Stratified metering algorithm. Technical report, Internal Report. Minnesota Department of Transportation, Minneapolis, Minnesota, 2001.
[66] Leslie N. Jacobson, Kim C. Henry, and Omar Mehyar. Real-time metering algorithm for centralized control. Number 1232. 1989.
[67] G. Paesani, J. Kerr, P. Perovich, and F.E. Khosravi. System wide adaptive ramp metering (swarm). In Merging the Transportation and Communications Revolutions. Abstracts for ITS America Seventh Annual Meeting and Exposition, 1997.
[68] Ioannis Papamichail, Markos Papageorgiou, Vincent Vong, and John Gaffney. Heuristic ramp-metering coordination strategy implemented at monash freeway, australia. Transportation Research Record: Journal of the Transportation Research Board, 2178(1):10-20, 2010.
[69] Gabriel Gomes and Roberto Horowitz. Optimal freeway ramp metering using the asymmetric cell transmission model. Transportation Research Part C: Emerging Technologies, 14(4):244-262, 2006.
[70] Apostolos Kotsialos, Markos Papageorgiou, Morgan Mangeas, and Habib Haj-Salem. Coordinated and integrated control of motorway networks via non-linear optimal control. Transportation Research Part C: Emerging Technologies, 10(1):65-84, 2002.
[71] Andreas Hegyi, Bart De Schutter, Hans Hellendoorn, and T Van Den Boom. Optimal coordination of ramp metering and variable speed control-an mpc approach. In American Control Conference, 2002. Proceedings of the 2002, volume 5, pages 3600-3605. IEEE, 2002.
[72] Mark D. Robinson. Examples of variable speed limit applications. 2000.
[73] Davey Warren. Variable speed limits. In FHWA Speed Management Workshop. Dallas, Texas, 2000.
[74] R.W. Lyles, William C. Taylor, and Jeff Grossklaus. Field Test of Variable Speed Limits in Work Zones (in Michigan). Available through Michigan Department of Transportation, 2003.
[75] Highways Consultancy Group/Highways Research Group. M42 mm monitoring and evaluation, 2001.
[76] Paula J. Hammond. The 2010 congestion report, 2010.
[77] Ghulam H Bham, Suzanna Long, Hojong Baik, Tom Ryan, Lance Gentry, Khushboo Lall, Mahdi Arezoumandi, Daxiao Liu, Tao Li, and Brian Schaeffer. Evaluation of variable speed limits on i-270/i-255 in st. louis. Technical report, 2010.
[78] Mohamed Abdel-Aty, Jeremy Dilmore, and Albinder Dhindsa. Evaluation of variable speed limits for real-time freeway safety improvement. Accident analysis \& prevention, 38(2):335-345, 2006.
[79] Markos Papageorgiou, Elias Kosmatopoulos, and Ioannis Papamichail. Effects of variable speed limits on motorway traffic flow. Transportation Research Record: Journal of the Transportation Research Board, 2047(1):37-48, 2008.
[80] Pei-Wei Lin, Kyeong-Pyo Kang, and Gang-Len Chang. Exploring the effectiveness of variable speed limit controls on highway work-zone operations. In Intelligent transportation systems, volume 8, pages 155-168. Taylor \& Francis, 2004.
[81] Andreas Hegyi, Bart De Schutter, and J. Hellendoorn. Optimal coordination of variable speed limits to suppress shock waves. Intelligent Transportation Systems, IEEE Transactions on, 6(1):102-112, 2005.
[82] Mohamed A. Abdel-Aty and Albinder Dhindsa. Coordinated use of variable speed limits and ramp metering for improving safety on congested freeways. In Transportation Research Board 86th Annual Meeting, number 07-0008, 2007.
[83] Ioannis Papamichail, Katerina Kampitaki, Markos Papageorgiou, and Albert Messmer. Integrated ramp metering and variable speed limit control of motorway traffic flow. In 17th IFAC World Congress, Seoul, Korea, pages 14084-14089, 2008.
[84] Andreas Hegyi, Bart De Schutter, and Hans Hellendoorn. Model predictive control for optimal coordination of ramp metering and variable speed limits. Transportation Research Part C: Emerging Technologies, 13(3):185-209, 2005.
[85] Ajith Muralidharan and Roberto Horowitz. Optimal control of freeway networks based on the link node cell transmission model. In American Control Conference (ACC), 2012, pages 5769-5774. IEEE, 2012.
[86] Mark Brackstone and Mike McDonald. Car-following: a historical review. Transportation Research Part F: Traffic Psychology and Behaviour, 2(4):181-196, 1999.
[87] Denos C. Gazis, Robert Herman, and Richard W. Rothery. Nonlinear follow-the-leader models of traffic flow. Operations Research, 9(4):545-567, 1961.
[88] Peter G. Gipps. A behavioural car-following model for computer simulation. Transportation Research Part B: Methodological, 15(2):105-111, 1981.
[89] Gordon Frank Newell. A simplified car-following theory: a lower order model. Transportation Research Part B: Methodological, 36(3):195-205, 2002.
[90] R.M. Michaels. Perceptual factors in car following. In Proceedings of the 2nd International Symposium on the Theory of Road Traffic Flow (London, England), OECD, 1963.
[91] Shinya Kikuchi and Partha Chakroborty. Car-following model based on fuzzy inference system. Number 1365. 1992.
[92] K. Yikai, J.I. Satoh, N. Itakura, N. Honda, and A. Satoh. A fuzzy model for behaviour of vehicles to analyze traffic congestion. In Proceedings of the International Congress on Modelling and Simulation, 1993.
[93] Saskia Ossen, Serge P. Hoogendoorn, and Ben G.H. Gorte. Interdriver differences in carfollowing: a vehicle trajectory-based study. Transportation Research Record: Journal of the Transportation Research Board, 1965(1):121-129, 2006.
[94] Elmar Brockfeld, Reinhart D. Kühne, and Peter Wagner. Calibration and validation of microscopic models of traffic flow. Transportation Research Record: Journal of the Transportation Research Board, 1934(1):179-187, 2005.
[95] Berkeley University of California. Berkeley highway lab, 2013. http://bhl.path. berkeley.edu/, accessed Noverber 10, 2013.
[96] R. Eddie Wilson. An analysis of gipps's car-following model of highway traffic. IMA journal of applied mathematics, 66(5):509-537, 2001.
[97] Hesham Rakha and Weidong Wang. Procedure for calibrating gipps car-following model. Transportation Research Record: Journal of the Transportation Research Board, 2124(1):113-124, 2009.
[98] A. Chow, X. Lu, Z. Qiu, S. Shladover, and H. Yeo. An empirical study of traffic breakdown for variable speed limit and ramp metering control. In 89th TRB Annual Meeting, 2009.
[99] Ajith Muralidharan and Roberto Horowitz. Imputation of ramp flow data for freeway traffic simulation. Transportation Research Record: Journal of the Transportation Research Board, 2099(1):58-64, 2009.
[100] Peter Koonce, Lee Rodegerdts, Kevin Lee, Shaun Quayle, Scott Beaird, Cade Braud, Jim Bonneson, Phil Tarnoff, and Tom Urbanik. Traffic signal timing manual. Technical report, 2008. http://ops.fhwa.dot.gov/publications/fhwahop08024/, accessed Noverber 10, 2013.
[101] R.D. Bretherton and G.T. Bowen. Recent enhancements to scoot-scoot version 2.4. In Road Traffic Control, 1990., Third International Conference on, pages 95-98. IET, 1990.
[102] PR Lowrie. Scats, sydney co-ordinated adaptive traffic system: A traffic responsive method of controlling urban traffic. 1990.
[103] K Larry Head, Pitu B Mirchandanai, Steve Shelby, Tucson University of Arizona, Department of Systems, Industrial Engineering, and AZ 85721 USA Tucson. The RHODES prototype: a description and some results. Transportation Research Board, 1998.
[104] Nathan H Gartner, Farhad J Pooran, and Christina M Andrews. Implementation of the opac adaptive control strategy in a traffic signal network. In Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE, pages 195-200. IEEE, 2001.
[105] Markos Papageorgiou. An integrated control approach for traffic corridors. Transportation Research Part C: Emerging Technologies, 3(1):19-30, 1995.
[106] Monique van den Berg, Andreas Hegyi, Bart De Schutter, and J Hellendoorn. Integrated traffic control for mixed urban and freeway networks: A model predictive control approach. European journal of transport and infrastructure research, 7(3):223-250, 2007.
[107] Eil Kwon, Ravi-Praveen Ambadipudi, and John Bieniek. Adaptive coordination of ramp meter and intersection signal for optimal management of freeway corridor. In Proc. 82th Annu. Meeting Transp. Res. Board, 2003.
[108] Zong Z. Tian, Kevin Balke, Roelof Engelbrecht, and Larry Rilett. Integrated control strategies for surface street and freeway systems. Transportation Research Record: Journal of the Transportation Research Board, 1811(1):92-99, 2002.
[109] H. Michael Zhang, Jingtao Ma, and Yu (Marco) Nie. Local synchronization control scheme for congested interchange areas in freeway corridor. Transportation Research Record: Journal of the Transportation Research Board, 2128(1):173-183, 2009.
[110] Miovision Technologies. Rethink traffic - miovision, 2013. http://miovision.com/, accessed Noverber 10, 2013.
[111] Wisconsin Department of Transportation. Model calibration, 2013. http: //www.wisdot.info/microsimulation/index.php?title=Model_ Calibration, accessed Noverber 10, 2013.
[112] Michael Zhang, Jingtao Ma, and Hu Dong. Developing calibration tools for microscopic traffic simulation final report part ii: Calibration framework and calibration of local/global driving behavior and departure/route choice model parameters. California Path Program, Institute of Transportation Studies, University of California, Berkely, Tech. Rep, 2008.
[113] Markos Papageorgiou, Habib Hadj-Salem, and Jean-Marc Blosseville. Alinea: A local feedback control law for on-ramp metering. Transportation Research Record, (1320), 1991.
[114] Greg Marsden, Mike McDonald, and Mark Brackstone. Towards an understanding of adaptive cruise control. Transportation Research Part C: Emerging Technologies, 9(1):33-51, 2001.
[115] Christopher Nowakowski, Steven E Shladover, Delphine Cody, Fanping Bu, Jessica OConnell, John Spring, Susan Dickey, and David Nelson. Cooperative Adaptive Cruise Control: Testing Drivers' Choices of Following Distances. California PATH Program, Institute of Transportation Studies, University of California at Berkeley, 2010.
[116] Xiao-Yun Lu, J Karl Hedrick, and Michael Drew. Acc/cacc-control design, stability and robust performance. In American Control Conference, 2002. Proceedings of the 2002, volume 6, pages 4327-4332. IEEE, 2002.
[117] Joel Vander Werf, Steven E. Shladover, Mark A. Miller, and Natalia Kourjanskaia. Effects of adaptive cruise control systems on highway traffic flow capacity. Transportation Research Record: Journal of the Transportation Research Board, 1800(1):78-84, 2002.
[118] Steven Shladover, Joel VanderWerf, Mark A. Miller, Natalia Kourjanskaia, and Hariharan Krishnan. Development and performance evaluation of avcss deployment sequences to advance from today's driving environment to full automation. 2001.
[119] S. E. Shladover, X. Lu, D. Cody, C. Nowakowski, Z. Qiu, A. Chow, J. OConnell, J. Nienhuis, and D. Su. Development and Evaluation of Selected Mobility Applications for VII. California PATH Program, Institute of Transportation Studies, University of California at Berkeley, 2011.
[120] Steven E. Shladover, Christopher Nowakowski, Jessica O’Connell, and Delphine Cody. Cooperative adaptive cruise control: Driver selection of car-following gaps. In 17th ITS World Congress, 2010.
[121] Christopher Nowakowski, Jessica O’Connell, Steven E Shladover, and Delphine Cody. Cooperative adaptive cruise control: Driver acceptance of following gap settings less than one second. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 54, pages 2033-2037. SAGE Publications, 2010.
[122] Bart Van Arem, Cornelie J.G. van Driel, and Ruben Visser. The impact of cooperative adaptive cruise control on traffic-flow characteristics. Intelligent Transportation Systems, IEEE Transactions on, 7(4):429-436, 2006.
[123] Arne Kesting, Martin Treiber, Martin Schönhof, and Dirk Helbing. Adaptive cruise control design for active congestion avoidance. Transportation Research Part C: Emerging Technologies, 16(6):668-683, 2008.
[124] Arne Kesting, Martin Treiber, Martin Schönhof, and Dirk Helbing. Extending adaptive cruise control to adaptive driving strategies. Transportation Research Record: Journal of the Transportation Research Board, 2000(1):16-24, 2007.
[125] Wouter J. Schakel, Bart van Arem, and Bart D. Netten. Effects of cooperative adaptive cruise control on traffic flow stability. In Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on, pages 759-764. IEEE, 2010.
[126] Hwasoo Yeo. Asymmetric microscopic driving behavior theory. PhD thesis, University of California, Berkeley, 2008.
[127] Hwasoo Yeo, Alexander Skabardonis, John Halkias, James Colyar, and Vassili Alexiadis. Oversaturated freeway flow algorithm for use in next generation simulation. Transportation Research Record: Journal of the Transportation Research Board, 2088(1):68-79, 2008.
[128] H. Yeo and A. Skabardonis. Parameter estimation for ngsim freeway flow algorithm. ASCE AATT 2008, 2008.
[129] Fanping Bu, Han-Shue Tan, and Jihua Huang. Design and field testing of a cooperative adaptive cruise control system. In American Control Conference (ACC), 2010, pages 46164621. IEEE, 2010.

