Title
Increasing Freeway Capacity by Efficiently Timing its Nearby Arterial Traffic Signals

Author
Kan, Xingan David

Publication Date
2017

Peer reviewed|Thesis/dissertation

Increasing Freeway Capacity by Efficiently Timing its Nearby Arterial Traffic Signals

By

Xingan David Kan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Alexander Skabardonis, Chair
Professor Michael Cassidy
Professor Rhonda Righter

Summer 2017

Abstract

Increasing Freeway Capacity by Efficiently Timing its Nearby Arterial Traffic Signals

by

Xingan David Kan

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Alexander Skabardonis, Chair

On-ramp metering at freeway bottlenecks is an effective method of reducing a freeway system's delay because empirical studies have previously shown that metering can prevent capacity drop, or reduction of outflow as a result of queue formation. However, arterial traffic signals that facilitate access to freeway on-ramps operate independent of the traffic conditions on the freeway on-ramp. Consequently, the traffic signals employ long signal cycle lengths and thus long green durations to maximize capacity at the arterial signalized intersections, which result in long platoons of freeway-bound traffic advancing toward the on-ramps. This often causes queue spillback on the freeway on-ramps and the surface street network. Queue override, a function that terminates or significantly relaxes the on-ramp metering rate whenever a sensor placed at the entrance of the on-ramp detects a potential queue spillover of the on-ramp vehicles on the adjacent surface streets, has become a widely accepted method of resolving queue spillback on the freeway on-ramps and nearby surface streets. Unfortunately, queue override releases the queue into the freeway and negates the benefit of ramp metering during the peak hours with recurrent freeway congestion. Video data collected downstream of freeway/on-ramp merge in San Jose, California show that the bottleneck discharge flow diminishes when queue override is activated for a sustained period of time. Observations over a two-week period suggest that queue override reduces the bottleneck discharge flow by an average of 10%.

Recently, there has been significant interest in integrated corridor management (ICM) of facilities comprised of freeways and adjacent arterial streets. Significant benefits can be realized by preventing queue override and effectively storing the queued vehicles on the nearby arterial surface streets if the arterial traffic signals can account for traffic conditions on freeway on-ramps and avoid sending long platoons to the freeway on-ramp. A signal control strategy was developed and evaluated in this study. The algorithm takes available on-ramp storage and freeway ramp metering rate into account and dynamically reduces the cycle length and adjusts the green durations to prevent on-ramp queue spillback and mitigate unnecessary delay in the conflicting arterial directions. The proposed algorithm was tested through simulation and the results show that the proposed strategy reduces the freeway and system-wide delay even under fluctuations in traffic demand, at a modest penalty on the on-ramp bound traffic.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I have appreciated the learning experience at Berkeley throughout my Ph.D. career. It has been a challenging but at the same time rewarding experience. I greatly appreciate the support and guidance from my research advisor Alexander Skabardonis, especially during the toughest times of my academic career. I want to thank Prof. Michael Cassidy for his guidance on the queue override field study and Prof. Carlos Daganzo for his invaluable comments and suggestions for my presentation at the Institute of Transportation Studies (ITS) seminar. I would also like to extend my thanks to the exam and dissertation committee members Rhonda Righter and Joan Walker.

I would like to thank my colleagues at 416 McLaughlin Hall, Zahra Amini, Paul Anderson, Jean Doig, Lewis Lehe, Wei Ni, Nathalie Saade, and Joshua Seeherman, for their guidance and research recommendations.

I would also like to thank my Richmond Field Station colleagues, Dr. Xiao-yun Lu, Dr. Steven Shladover, Dr. Hao Liu, Dr. Hani Ramezani, and John Spring for their collaboration and technical guidance throughout the research project.

I am thankful to my parents, Shihai Kan and Shuping Wang, for their support and sacrifices to raise me and educate me. I could not have been where I am today without their help and guidance.

Lastly, I appreciate the collaboration and support from Zhongren Wang, David J. Wells, James Lau, and Ted Lombardi, of California Department of Transportation (Caltrans) Headquarter Division of Traffic Operations; the project manager Hassan Aboukhadijeh of Caltrans Division of Research, Innovation, and Systems Information (DRISI); Alan Chow, Lester Lee, Sean Coughlin, Einar Acuna, Stan Kung, Min Yin Lee, David Man, and Mazen Arabi, of Caltrans District 4; Lily Lim-Tsao and Joel Roque of the City of San Jose.

CHAPTER 1: INTRODUCTION

1.1 Motivation

Congestion on urban freeways has become a wide spread issue in many major metropolitan areas. Lost productivity and excessive fuel consumption are just two of the many negative consequences associated with peak hour congestion. Freeway congestion primarily occurs due to the presence of bottlenecks. A bottleneck is defined as a point where the traffic demand exceeds the normal capacity, resulting in formation of queues upstream of that location and free-flowing traffic downstream. The bottleneck is called "active" when traffic flow through the bottleneck is not affected by downstream restrictions (spillback from downstream bottlenecks). Recurrent bottlenecks occur on the same location and time periods of the day. Their behavior and characteristics are reproducible over many days. Typically the bottleneck remains active throughout the peak period(s). Traffic queues will dissipate from the back as traffic demand drops below the available capacity. On the other hand, non-recurrent bottlenecks due to incidents may have shorter duration, although some major incidents may last a long time. Non-recurrent bottlenecks are non-reproducible since incidents are random events and may occur anywhere in the freeway system. Furthermore, traffic queues will often dissipate from the front following the incident removal, i.e. when the normal capacity is restored.

There are various types of freeway recurrent bottlenecks depending on their location and causes. For example, at freeway merging areas, the vehicles entering from the on-ramp trigger traffic breakdown forming a bottleneck in the main lanes shortly downstream of the merge point. The capacity of an active freeway bottleneck is defined as the maximum sustained flow it discharges under prevailing traffic and roadway conditions. The capacity of a freeway system is heavily influenced by the amount of on-ramp traffic that merges onto the freeway. Queuing on the shoulder lane (near the merging area) as a result of high entry demand induces upstream lane change maneuvers that ultimately diminish the discharge flow of the freeway bottleneck (Cassidy and Rudjanakanoknad, 2005). This phenomenon is referred as "capacity drop". In response to this, freeway on-ramp metering has been extensively used as a traffic control strategy to regulate and restrict the entry of the on-ramp vehicles in order to prevent congestion and preserve the freeway capacity.

However, the nearby arterial traffic signals that facilitate freeway access operate independent of freeway ramp metering. The arterial traffic signals respond to the peak hour demand by employing long signal cycles, and thus long green durations, and by progressively coordinating traffic signals along the major arterial that facilitates access to freeway on-ramps, in order to maximize arterial capacity. This may lead to platoons of arterial traffic advancing to the freeway on-ramps that are metered during the peak hours. As a result, metering the large influx of on-ramp traffic requires sufficient space to store queued vehicles. Unfortunately, physical constraints in road geometry and the surrounding lane use do not often allow for the construction of long on-ramps. This leads to on-ramp oversaturation and queue spillback prior to the termination of the long green phase. As a result, a portion of the green phase cannot be used to serve the traffic demand, and therefore this can significantly reduce the intersection capacity. Most of the existing ramp metering systems therefore employ a "queue override" feature that is intended to prevent the on-ramp queue from obstructing traffic along the adjacent surface streets. The override is triggered whenever a sensor placed at the entrance of the on-ramp detects a potential queue spillover of the

on-ramp vehicles on the adjacent surface streets. This clears the on-ramp queue by temporarily turning off ramp metering or increasing the metering rate to the maximum allowable value (typically 900 veh/hour/lane). Unfortunately, this approach reduces the effectiveness of the ramp metering systems during the time of the highest traffic demand, when the ramp metering is most needed. So it may be useful to dispatch arterial traffic to the freeway on-ramp in smaller platoons instead of long platoons over a short period of time, in order to prevent on-ramp oversaturation and subsequent queue override. This can be achieved by reducing the cycle length of the arterial traffic signals adjacent to the freeways, which will be discussed in detail in the remainder of the thesis.

The rest of this chapter presents the research question and the research contribution, followed by an overview of the dissertation organization.

1.2 Research Questions

Recent concerns regarding urban congestion have prompted transportation agencies to investigate and explore the potential mobility benefits of coordination and integration of freeway traffic control systems such as ramp metering and the nearby arterial traffic signals, as well as the challenges both in the technical and institutional domains. Two major problems will be addressed in this research:

- Queue override can potentially diminish the freeway bottleneck discharge rate in an effort to mitigate on-ramp queue spillback and avoid interference on nearby arterial streets. This research is intended to verify that queue override can diminish the freeway bottleneck capacity via empirical observations, and will quantify the extent of reduction in bottleneck discharge flow.
- Nearby arterial traffic signals can help avoid queue spillback at freeway on-ramps if the signal control systems were integrated with those of the freeway ramp metering and if the traffic signals were timed effectively based on the ramp metering rate and on-ramp queue lengths. Thus, it may be useful to develop a viable signal control approach to alleviate or prevent queue spillback at freeway on-ramps, and help maintain freeway capacity. Furthermore, the new signal control approach should be readily implementable after simple adjustments such as system integration and software upgrades of the freeway ramp metering and arterial traffic signals, and should not require investments in new sensing technologies and infrastructural upgrades.

1.3 Research Contribution

There is currently a lack of literature regarding the impact of queue override on freeway bottleneck capacity, despite the extensive empirical studies on “capacity drop” and the effectiveness of freeway ramp metering to prevent it. The contributions of this dissertation include empirical evidence on the impact of queue override on capacity drop; and a signal control algorithm to manage the entry of vehicles on the on-ramp and mitigate queue spillback. The present empirical study is the first to quantify how much the freeway bottleneck capacity diminishes when queue override is activated. More importantly, the empirical study is conducted for the more widely accepted method of queue override, which maximizes the ramp metering rate instead of temporarily suspending ramp metering.

Recent literature on arterial signal timing near freeways has not effectively addressed arterial signals near metered freeway on-ramps. The proposed signal control algorithm dynamically adjusts the cycle lengths and green durations to avoid sending large platoons of vehicles onto the on-ramp. Unlike other signal control algorithms, the method presented in this dissertation reduces the cycle lengths rather than simply distributing the green durations without adjusting signal cycle. The latter leads to underutilized green times for some approaches and potential queue propagation and spillback on some arterial streets further upstream. Most importantly, the arterial signal control proposed in this dissertation is readily implementable after incremental changes in software integration of freeway ramp metering and arterial traffic signals, unlike other signal control strategies that require new detectors or other sensing technologies.

1.4 Dissertation Organization

Chapter 2 provides an overview of the literature in capacity drop and ramp metering, as well as recent research in the area of signal control of arterial intersections near freeways. Chapter 3 describes the selected test site, the field study that verifies the capacity drop induced by queue override, and the proposed arterial signal control strategy. Chapter 4 details the results from the microscopic simulation tests of the proposed signal control strategy. Lastly, Chapter 5 summarizes and concludes the study findings, followed by discussion of the limitation of this study and future work. Miscellaneous items are described in appendices. Appendix A documents the test site selection process and describes the selected test site in greater detail. Appendix B shows additional data obtained from the empirical study on the effect of queue override. Appendix C discusses the calibration of the preliminary microscopic simulation model. Appendix D details the model enhancements made to the simulation, followed by extensive calibration and validation of the enhanced driving behavior model in simulation.

CHAPTER 2: LITERATURE REVIEW

Capacity drop arises from queuing as a result of high entry demand. The high demand induces upstream lane change maneuvers that ultimately diminish the discharge flow of the freeway bottleneck (Cassidy and Bertini, 1999; Cassidy and Rudjanakanoknad, 2005). Many other empirical studies (Banks, 1991; Hall and Agyemang-Duah, 1991; Persaud et al., 1998; Sahin and Altun, 2008; Oh and Yeo, 2012; Srivastava and Geroliminis, 2013; Yuan et al., 2015) have documented capacity drop at merges. These earlier studies suggest that capacity drop typically entails a 5% to 15% reduction in the merge bottleneck discharge flow. In addition, many have proposed mathematical models to explain capacity drop (Leclercq et al., 2011; Parzani and Buisson, 2012; Jin et al., 2015; Yuan et al., 2015).

Freeway on-ramp metering (Papageorgiou and Kotsialos, 2002) has been extensively used as a traffic control strategy to regulate the entry of the on-ramp vehicles in order to avoid a freeway's capacity drop (Papageorgiou and Kotsialos, 2002; Smaragdis et al, 2003; Cassidy and Rudjanakanoknad, 2005; Horowitz et al, 2005; Zhang and Levinson, 2010; Al-Qbaedi and Yousif, 2012; Kim and Cassidy, 2012). Additional benefits of ramp metering include accident reduction, improved freeway travel time, and better travel time reliability (MnDOT, 2001).

However, there is currently very limited empirical evidence that quantifies the impact of queue override on freeway bottleneck capacity. Chilukuri et al. (2013) conducted an empirical study on the effect of turning off ramp metering for short intervals of 30 to 75 seconds. This is commonly practiced by the transportation agencies of U.S. states such as Georgia, Texas, and Washington. However, the methodology by Chilukuri et al. (2013) was flawed due to the absence of evidence suggesting that the bottleneck was isolated from the queue spillback of the downstream bottlenecks. Besides the field study attempted by Chilukuri et al. (2013), there has not been any empirical study that quantifies the impact of queue override, especially those that do not temporarily suspend ramp metering entirely but as an alternative, meter on-ramps at maximum allowable rate, an approach that has become widely practiced.

Some recent works have addressed arterial traffic signal timing near freeways as they pertain to recurrent on-ramp bottlenecks, freeway traffic diversion under non-recurrent conditions, and off-ramp bottlenecks. Representative approaches of each type are presented in the next three sections.

2.1 Control Strategies for Recurrent Congestion

There has been very limited research in the area of signal timing for arterial intersections near freeways on-ramp bottlenecks, and no generalizable control strategies have been implemented.

Tian et al (2005) developed an algorithm for diamond interchanges that reduces green durations for movements with on-ramp access to prevent on-ramp queue spillback, without changing the cycle length. This approach may cause spillback of on-ramp demand onto the upstream arterial, especially under long cycle lengths, and temporary activation of queue override.

Recker et al (2009) developed a system-wide optimization model for ramp metering and traffic signals from stochastic queuing theory. But the improvement observed after implementing

the control strategy at a network of freeways and arterials was a result of using a more efficient ramp metering control, rather than an improved signal timing plan. Moreover, the proposed approach requires solving non-linear optimization in real time, which is computationally intensive and not feasible in most real setting.

Other research efforts have focused only on control of isolated signalized intersections at or adjacent to freeway on-ramps. For example, Li and Tao (2011) proposed a signal optimization model for an arterial at an isolated freeway interchange using the cell transmission model, but did not consider ramp metering in their algorithm.

Su et al (2014) developed a signal optimization model for an isolated diamond interchange that takes the ramp meter rate and on-ramp queue length into account. A brief field test was conducted to show that coordination of freeway ramp metering and arterial traffic signals is technologically feasible and implementable in the real world. However, similar to the method by Tian et al (2005), the proposed algorithm simply reallocated green times without changing the cycle length. Therefore, it provided unnecessarily long green durations for the conflicting movements and disregarded the potential queue spillback into the upstream arterial intersection. Furthermore, the impact of queue override was not considered.

2.2 Control Strategies for Freeway Traffic Diversion

There are several studies on diverting freeway traffic onto adjacent arterials, mostly in non-recurrent conditions such as incidents on freeways. These studies typically explore how to effectively utilize the spare capacity of the adjacent arterials in the event of temporary freeway capacity reduction, in order to prevent severe congestion. They do not address how to efficiently time arterial traffic signals adjacent to the metered freeway on-ramps.

Recently, the Federal Highway Administration released a manual for coordinated freeway and arterial operation (Urbanik et al, 2006) but the document does not provide any control strategies for coordinated operation of the freeway and the arterial. The document outlines the practical issues such as institutional barriers, technological challenges, and integration of intelligent transportation systems. The manual also provides examples of freeway-arterial corridors that have implemented coordinated operation schemes, however, these examples only show how local arterials can be coordinated with the freeway in the event of an incident to divert some freeway traffic by utilizing the excess capacity on the local arterials.

Control strategies for traffic diversion adjust ramp metering rates and arterial traffic signal timing plans in order to facilitate high volumes of freeway traffic to exit the freeway, travel on the adjacent parallel arterial efficiently, and quickly return to the freeway immediately downstream of a capacity-constrained location such as an area with incidents. Tian et al (2002) proposed a traffic-responsive coordination strategy that extends the green times corresponding to the freeway off-ramps and parallel arterial and maximizes metering rates at the downstream on-ramps based on real time queue detection on the freeway. The idea was shown to be effective for freeway-arterial corridors with consecutive diamond interchanges. In addition, work by Zhang et al (2009) tested a similar approach at a corridor with various configurations of freeway interchanges. Other works in this area include: an optimization-based coordination strategy that minimizes corridor level delay during incident diversion (Zhang et al, 2012); an empirical study of the effect of dynamic

traveler information on the amount of freeway traffic diverted and the corridor-wide performance (Liu et al, 2013); and a control strategy for diverting traffic from the freeway to the adjacent arterials with significant spare capacity, in the event of periodic freeway capacity reductions (Gunther et al, 2012).

2.3 Control Strategies for Off-ramp Bottlenecks

Several studies investigated the queue spillback of off-ramp freeway traffic onto the freeway mainline. Off-ramp bottlenecks typically are created because of inefficient signal timing at the downstream end of the freeway off-ramp, where it intersects the adjacent arterial. Recently Yang et al (2014) proposed conditional signal priority for off-ramp traffic in order to mitigate the impact of off-ramp spillback on freeway performance. This was enhanced in (Yang et al, 2015) by incorporating downstream arterial signal progression to quickly discharge the off-ramp queue and further reduce the impact of off-ramp spillback.

CHAPTER 3: RESEARCH APPROACH

A freeway corridor with three recurrent on-ramp bottlenecks, along with its nearby arterials, was selected for the empirical study on queue override; and the evaluation of the proposed signal control strategy for nearby arterial intersections. The empirical study shows that queue override diminishes freeway bottleneck capacity. Three potential signal control strategies were proposed. The strategy ultimately selected was tested using microscopic simulation, which was enhanced and modified to help reproduce capacity drop and better replicate real world driving behavior.

3.1 Selected Study Site

The study site was selected based on the criteria that the site has recurrent bottlenecks, on-ramp metering, limited on-ramp queue storage, and actuated or adaptive signals that can modify cycle lengths and green times in real time. Detailed site selection criteria are discussed in Appendix A. After discussion with the California Department of Transportation (Caltrans) District 4, a three mile section of I-680 from Alum Rock Ave. to Berryessa Rd. in San Jose, California, shown in figure 3.1 was selected. There are three recurrent bottlenecks on this stretch of I-680; they are located near the on-ramps from Berryessa Rd., McKee Rd., and Alum Rock Ave. At all three bottlenecks, high on-ramp demand from of the westbound direction of Berryessa Rd., McKee Rd., and Alum Rock Ave., along with high on-ramp demand from both directions of Capitol Ave., result in high volumes of merging traffic entering the northbound freeway mainline during the morning peak (7:00-9:30 AM).

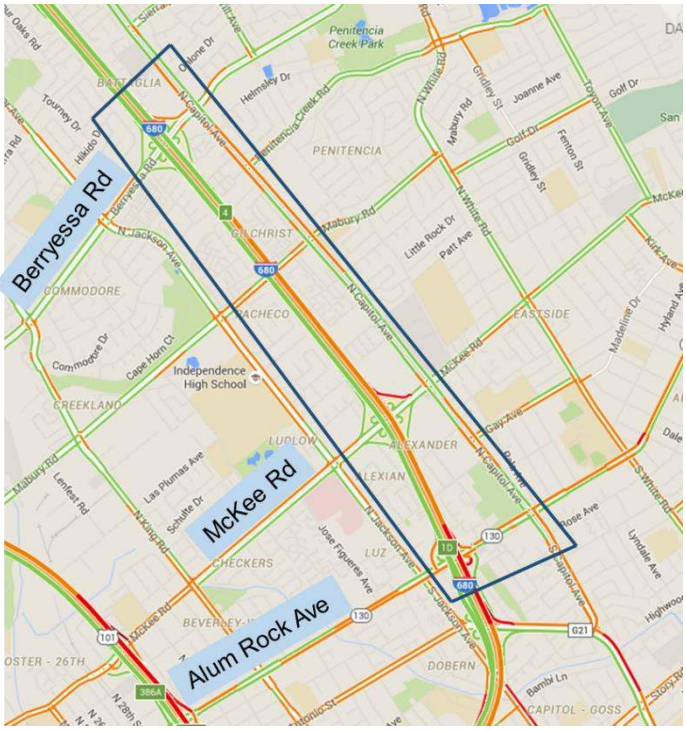


Figure 3.1 Map of study site.

One of the main reasons for morning peak demand in this corridor is the increasing number of commuters to the neighboring cities of Fremont and Milpitas. There are many trips generated

from the densely populated residential areas surrounding San Jose in the south to the employment centers in Fremont and Milpitas in the north, during the morning peak period.

This section of the freeway has 4 lanes in each direction, whereas the parallel arterial Capitol Ave., as well as Alum Rock Ave. and Berryessa Rd. each has 2 lanes in each direction. McKee Rd. has 3 lanes in each direction. Typically, merging traffic causes the average freeway speed to decrease to about 20 mph near the Alum Rock on-ramp, 30 mph near the McKee on-ramp, and 40 mph near the Berryessa on-ramp. Refer to Figures A.7-A.9 in Appendix A for flow and speed time series of each bottleneck during a typical morning peak.

All of the on-ramps in this corridor are metered and the meters operate under the local traffic responsive demand-capacity approach in which the metering rates are assigned based on various thresholds of freeway mainline occupancies immediately upstream of the merging or weaving area, in order to ensure that the capacity downstream of the merge is not exceeded. The metering rates and their respective occupancy thresholds for each on-ramp are shown in table A.1 and table A.2 of Appendix A. Under the current practice, the ramp meters employ "queue override" in case queue spillback is detected by the loop detector at the on-ramp entrance. On a typical day, queue override is activated from 7:30 AM to 8:00 AM, though it sometimes lasts until 8:30 AM. The queue override algorithm at this site increases the ramp metering rate by 100 veh/hour/lane as soon as queue spillover is detected at the entrance of the on-ramp, and by another 100 veh/hour/lane if queue spillover continues in each of the next 30 second cycle, until the rate reaches the maximum of 900 veh/hour/lane.

The signalized intersections of this corridor are actuated and coordinated. Each time period of the day uses its unique timing plan. For the weekday morning peaks, the existing cycle lengths are relatively long (130 to 160 seconds) and the signal timing plan provides progression to the northbound direction, which is the one with higher demand. Refer to figures A.10 to A.13 in Appendix A for turning movement volumes and signal timing plans of the four major signalized intersections during a typical weekday morning peak.

3.2 Field Study: Impact of Queue Override

This section describes an empirical study at a merge bottleneck to assess the effect of queue override on freeway capacity. I-680 northbound at the McKee Rd. on-ramp is an active recurrent bottleneck located in the middle of the selected corridor, as described in the previous section 3.1. Loop detector data of the surrounding area suggests that this bottleneck is typically isolated from exogenous restrictions such as queue spillback from the downstream bottleneck (Caltrans PeMS, 2016). Figure 3.2 shows the detailed lane configuration of the bottleneck selected and the camera locations for this empirical study. According to the figure, the on-ramp consists of two lanes upstream of the ramp meter, and the two lanes merge into a single lane before reaching the freeway mainline. The ramp meter restricts the flow of on-ramp merging traffic and ensures smooth merging operation of the two on-ramp lanes by alternating the green times assigned to each on-ramp lane. The maximum discharge flow of this bottleneck is typically observed during the morning peak (7:00 AM – 9:30 AM). Based on the information provided by Caltrans District 4, the operating agency, the prescribed ramp metering rate during the morning peak is summarized as in table A.2 in Appendix A.

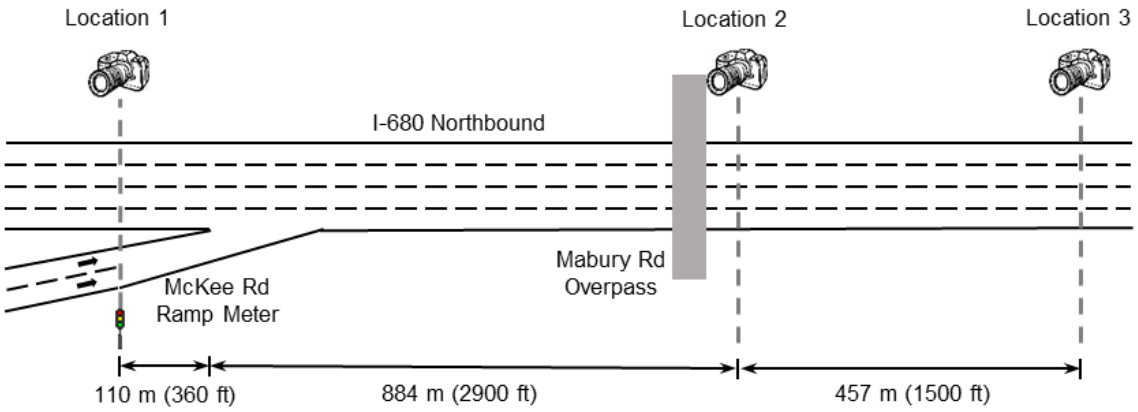


Figure 3.2 I-680 merge bottleneck at McKee Rd. on-ramp.

Video cameras were placed upstream and downstream of the McKee Rd. on-ramp merge during the study periods of May 9, 2016 to May 13, 2016 and May 16, 2016 to May 20, 2016, and the camera locations are shown in figure 3.2. The camera placed upstream recorded all four mainline lanes, as well as the McKee Rd. on-ramp. The third camera was placed to ensure the absence of exogenous restrictions such as queue spillback from the bottlenecks further downstream. The records of the frequency and duration of queue override activation were provided by Caltrans District 4, the agency that operates the freeway ramp meters at this site. Lastly, there were no major incidents or weather events during the selected study period.

Vehicle count at each location and each 30 second interval was extracted from the video data. Bottleneck discharge rates during periods of active and inactive queue override were compared.

Figure 3.3 shows the oblique curves for cumulative vehicle count of the mainline lanes $O(t)$, obtained at all three cameras locations shown in figure 3.2, vs time, t. The curves were plotted to display virtual departures as a function of time at location 3 in figure 3.2 (Daganzo, 1997). The vertical displacement between the curves is the excess vehicle accumulation. The area between the curves indicates the total delay of the freeway segment. Figure 3.4 shows the curves for cumulative vehicle count of the on-ramp vs time, t. The data presented in the figures 3.3 and 3.4 were collected on Tuesday, May 10, 2016.

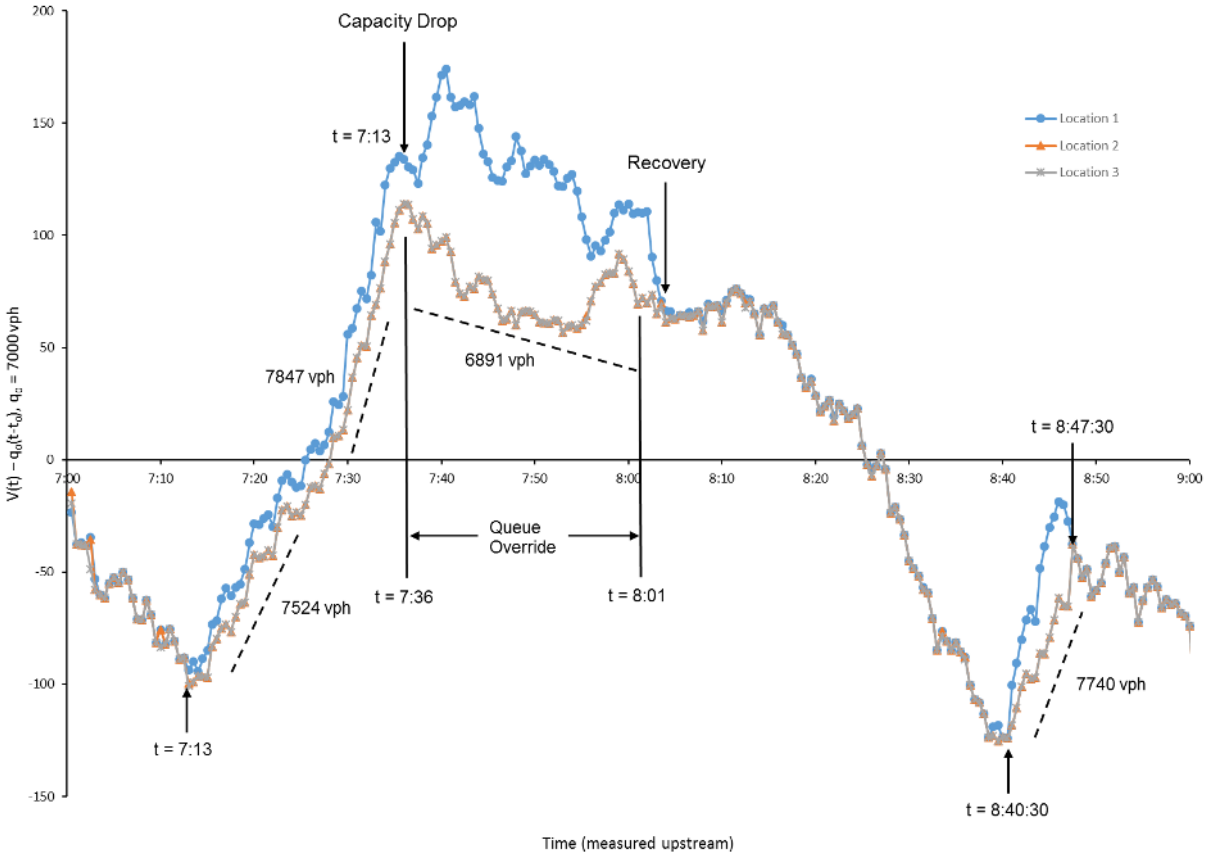


Figure 3.3 May 10, 2016: $O(t)$ curves for locations 1 through 3.

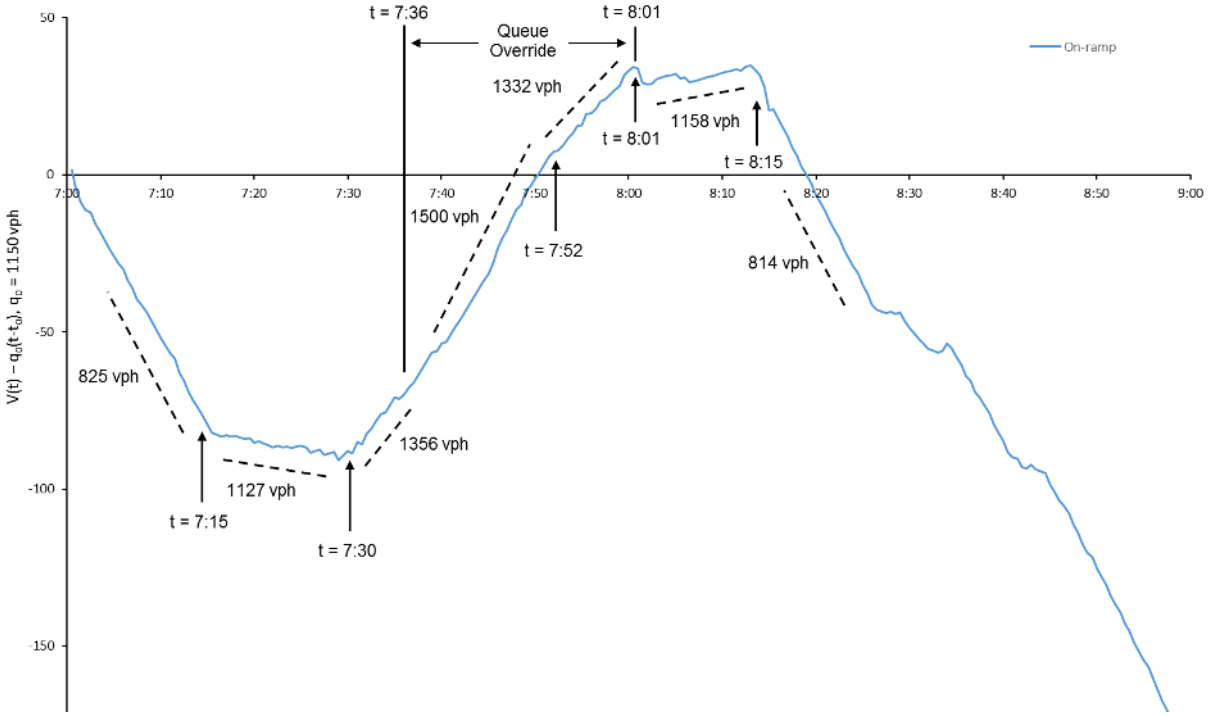


Figure 3.4 May 10, 2016: $O(t)$ curves for McKee Rd. on-ramp.

The vertical scales in figures 3.3 and 3.4 were modified by plotting on the oblique coordinate system, in order to make the excess accumulation (vertical displacement) clearly noticeable by visual inspection (Munoz and Daganzo, 2002). $O(t)$, the oblique coordinate transformation of the cumulative vehicle count, $V(t)$, is described by the following:

$$O(t) = V(t) - q_0(t - t_0) \tag{3.1}$$

where q_0 is the specified reference value of flow and t_0 is the specified reference value of initial time.

The $O(t)$ curves shown in figure 3.3 reveal that the arrival rate at location 1 was relatively low and the freeway was free-flowing (all three curves overlap) from $t = 7{:}00$ to $t = 7{:}13$. Video data from location 1 also show that the observed on-ramp flow is relatively low at about 825 vph, as described by the $O(t)$ curve in figure 3.4. This corresponds approximately to the prescribed restrictive metering rate of 400 vph/lane for the period of $t = 7{:}00$ to $t = 7{:}15$. The variation in actual ramp flow can be attributed to variability in green times, perception reaction time, etc.

Immediately after $t = 7{:}13$, the curves for locations 1 and 2 shown in figure 2 began to diverge as the freeway transitions from free-flow condition to queueing. At $t = 7{:}15$, the prescribed ramp meter rate increased from 400 vph/lane to 600 vph/lane, as indicated by the ramp flow of 1127 vph in figure 3.4. Despite the increase in on-ramp merging traffic, the bottleneck outflow remained high at 7524 vph during the initial period of queueing, as shown by the dashed lines.

Queuing continued at $t = 7{:}30$, when the prescribed ramp meter rate increased to 700 vph/lane (indicated by the ramp flow of 1356 vph in figure 3.4). Under the less restrictive ramp meter rate, the outflow of the bottleneck slightly increased to 7847 vph, shown in figure 3.3 by the dashed line.

However, the high outflow persisted only until $t = 7{:}36$, when sufficient on-ramp queue spillback prompted the activation of queue override based on records obtained from the Caltrans District 4. As indicated by the dashed line in figure 3.4, the on-ramp flow exceeded the expected 1400 vph after $t = 7{:}36$, at 1500 vph. The observed on-ramp flow was less than expected value of 1800 vph under the maximum meter rate because queues already formed at and near the merging area physically restricted the number of vehicles entering the freeway from the on-ramp. As shown in figure 3.3, queuing persisted after $t = 7{:}36$. The arrival rate remained high but the outflow of the bottleneck diminished, indicated by the downward trending $O(t)$ curve at location 2.

Queue override continued but the on-ramp flow began to diminish from 1500 vph to 1352 vph at $t = 7{:}52$ because queue spillback occurred less frequently therefore queue override was not constantly activated. This explains the slight increase in the bottleneck outflow, indicated by the curve for location 2.

As shown in figure 3.4, queue override ended at $t = 8{:}01$ and the on-ramp flow returned to 1158 vph; this corresponds to the prescribed meter rate of 600 vph/lane for $t = 8{:}00$ to $t = 8{:}15$. Despite the relatively high on-ramp flow, the overall arrival rate at location 1 was relatively

low, which led to free-flow conditions. The free-flow condition persisted after $t = 8:15$, when the on-ramp flow reduced to 814 vph due to the change in prescribed ramp meter rate; except for a brief period ($t = 8:40:30$ to $t = 8:47:30$) of demand surge that resulted in queuing and a high outflow of 7740 vph.

Further inspection of the $O(t)$ curves for location 2 and location 3 reveals that the segment between these locations remained free-flowing for the entire study period (both curves always overlapped). Thus the bottleneck was isolated and located between location 1 and location 2. Furthermore, queue persisted during the period of queue override. Therefore, the observed reduction in the bottleneck outflow during $t = 7:36$ to $t = 8:01$ was not a result of a reduction in traffic demand nor the result of an exogenous downstream restriction but the result of queue override. According to figure 3.3, the bottleneck outflow during queue override diminished to an average of 6891 vph, a reduction of 12.18% in comparison with the bottleneck outflow immediately before queue override was activated.

Moreover, the $O(t)$ curves corresponding to the next day (Wednesday, May 11, 2016) exhibit similar trends in comparison with those from Tuesday, May 10, 2016. The only difference is observed during the period following queue override deactivation. Demand upstream of the bottleneck on Tuesday May 10, 2016 subsided after the termination of queue override at $t = 8:01$, as indicated by all three downward sloping and overlapping $O(t)$ curves. On the other hand, figure 3.5 shows that the high bottleneck discharge rate was observed again once queue override terminated at $t = 8:09$, and queuing persisted until the end of the morning peak, which occurred at $t = 9:22:30$. Similarly, $O(t)$ curves from other days showed diminished bottleneck discharge rate during the period of queue override activation but fluctuations in demand during the period after queue override was deactivated. Details can be found in Appendix B.

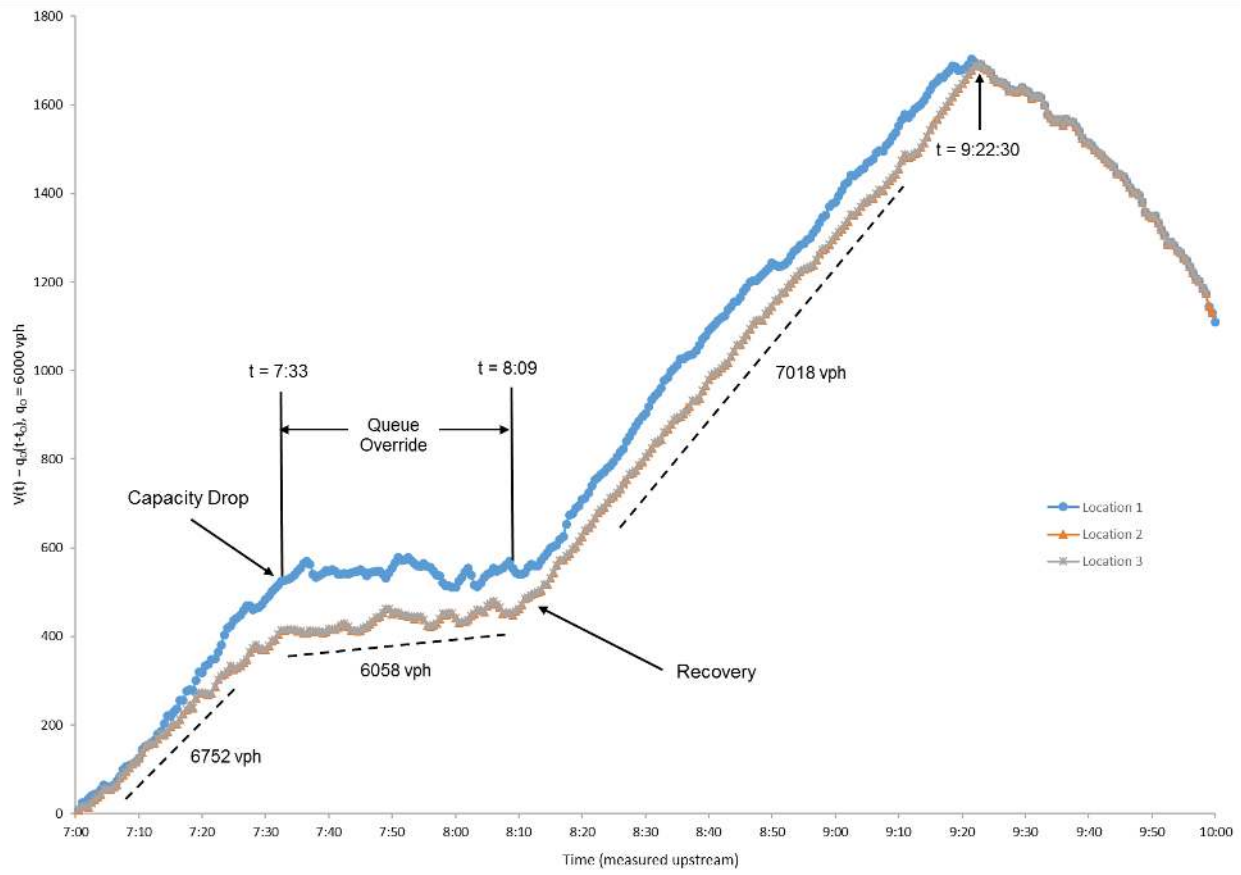


Figure 3.5 May 11, 2016: $O(t)$ curves for locations 1 through 3.

Table 3.1 provides a summary of the observed freeway bottleneck capacities prior to and during queue override for the two week study period. There were slight variations in the percentages of capacity drop observed in different days. The observed capacities prior to and after the activation of queue override vary by the day of the week, for instance, the observed capacities on Tuesday May 10, 2016 and May 19, 2016 were higher than those of the other days. Furthermore, the duration of queue override and capacity drop was about 25 to 30 minutes on average, with the exception of a 15 minute duration on Tuesday May 17, 2016 and a 40 minute duration on Wednesday May 18, 2016. In addition to the day to day variation, the capacity drop can be slight more severe during the first few minutes of queue override, for example, on Thursday May 19, 2016. Overall, the observations suggest that queue override diminishes the bottleneck outflow by an average of 10%.

Table 3.1 Freeway bottleneck capacities during morning peaks.

	Freeway bottleneck outflow (vph)		
	Before queue override	After queue override	% difference
Week 1			
May 9, 2016 (Monday)		Not activated	
May 10, 2016 (Tuesday)	7847	6891	-12.81
May 11, 2016 (Wednesday)	6752	6058	-10.28
May 12, 2016 (Thursday)		Downstream spillback	
May 13, 2016 (Friday)		Not activated	
Week 2			
May 16, 2016 (Monday)		Not activated	
May 17, 2016 (Tuesday)	7214	6672	-7.51
May 18, 2016 (Wednesday)	7109	6493	-8.67
May 19, 2016 (Thursday)	7532	6612	-12.21
May 20, 2016 (Friday)		Not activated	
Overall	----	----	-10.30

3.3 Development of Arterial Signal Control Strategy

Three signal control strategies were developed for arterial intersections that are adjacent to and facilitate on-ramp access. The most appropriate and readily implementable control strategy was selected for extensive simulation tests and future field implementation.

3.3.1 Control Strategy I

This strategy is an extension of the signal control strategy developed by Su et al. (2014) in order to resolve the issue with queue spillback on freeway on-ramps. The proposed approach maintains the existing freeway ramp metering algorithm in practice and is not limited to any specific freeway ramp metering algorithm.

The signal control proposed by Su el al (2014) balances the demand and supply of the green time of each phase while taking the on-ramp queue storage capacity into account, using a linear optimization model. The model assumes uniform arrivals, fixed phase sequence, constant lane capacities, and sufficient queue storage space downstream of the intersection unless a freeway on-ramp is located downstream. Green distributions are updated in real time at the beginning of every cycle. The objective function of the optimization model is shown in equation 3.2.

$$Min \sum_i \left| \frac{g_i(T)}{\frac{q_i(T-1) + d_i(T) \cdot C}{f_{sat,i}}} \right| + \delta \cdot \left(\sum_r \sum_{i \in R} \left| f_{sat,i} \cdot g_i(T) \cdot \beta_i - (RA_r + RA_l) \right| \right) \quad (3.2)$$

where,

$g_i(T)$: decision variable; optimal green time of phase i and the T-th cycle
C: cycle length
$d_i(T)$: uniform traffic demand (arrival rate) during the T-th cycle
$q_i(T)$: residual excess accumulation at the end of the T-th cycle
$f_{sat,i}$ saturation flow of phase i
R: indicator that the downstream section of phase i has access to an on-ramp
β_i: proportion of the traffic in phase i that accesses the on-ramp (based on historical data)
RA_r: available on-ramp storage space of ramp r
RA_l: available storage space of the arterial lanes with on-ramp access
δ: weighting factor for the penalty function.

The cycle length is fixed and assumed to be known in advance and is typically the network-wide optimal cycle length used in the coordinated actuated arterial traffic signals. $q_i(T)$, RA_r, and RA_l can be estimated by computing the difference between the number of arrivals upstream and the number of departures downstream during the previous cycle.

The expression

$$\frac{q_i(T-1) + d_i(T) \cdot C}{f_{sat,i}}$$

is the demand of green time of phase i in cycle T, and $g_i(T)$ is the supply of green time of phase i in cycle T.

The expression $\sum_{i \in R} f_{sat,ik} \cdot g_i(t) \cdot \beta_i$ is the maximum flow discharged toward freeway on-ramp *r* from the adjacent intersection. The objective function penalizes on the difference between the total flow discharged from the adjacent intersection and the sum of RA_r and RA_l, as illustrated in figure 3.6. This penalty function is intended to reduce green times to the turning movements that discharge traffic onto on-ramp r when on-ramp r and the adjacent arterial lanes for on-ramp access can no longer accommodate the demand. This would reallocate the green times to the conflicting directions. Furthermore, this would mitigate the queue spillback of the freeway on-ramp and its immediate upstream areas thus prevent on-ramp traffic from blocking the intersection and therefore the need for queue override. As a result, the absence of queue override would prevent any further negative impact on the freeway performance and maintain high freeway bottleneck capacity, based on the conclusions of the queue override field study.

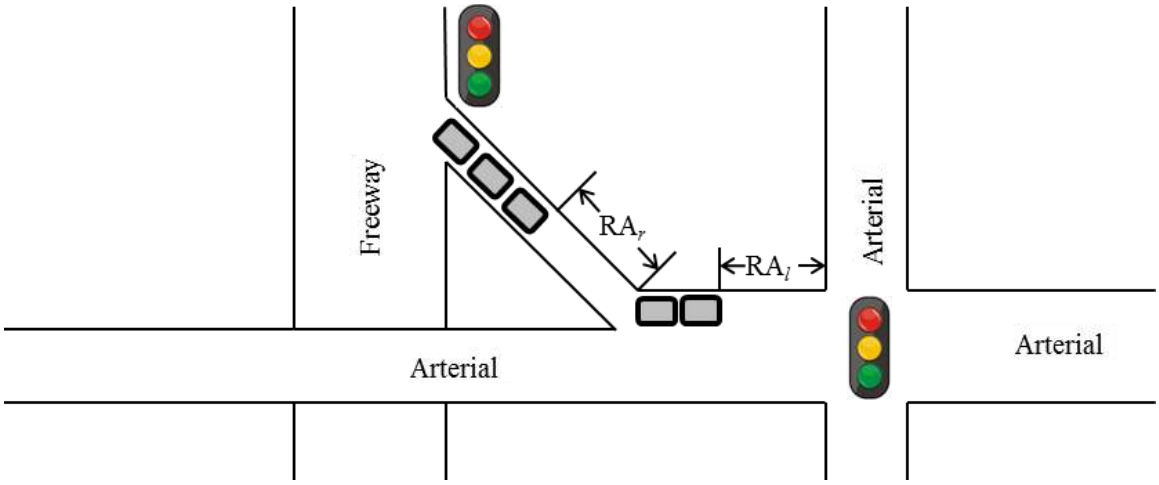


Figure 3.6 Arterial and freeway on-ramp queues.

Equations 3.3 to 3.5 are the constraints that address several practical limitations; equation 3.3 ensures that the minimum green time $G_{i,min}$ of phase i related to traffic safety is satisfied, and equations 3.4 and 3.5 ensures the dual ring structure is satisfied if a signal timing plan is developed for the typical four-leg intersection with actuated signals and protected left turning each direction shown in figure 3.7.

$$g_i(T) \geq G_{i,min} \tag{3.3}$$

$$g_1(T) + g_2(T) = g_5(T) + g_6(T) \tag{3.4}$$

$$\sum_{i=1-4\ or\ 5-8} g_i(t) = C \tag{3.5}$$

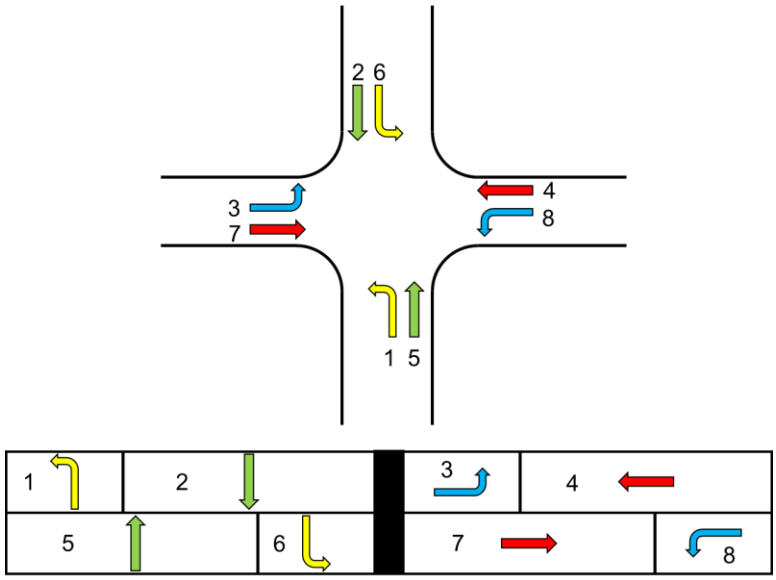


Figure 3.7 Dual ring actuated signal timing plan for a four-leg intersection.

Modifications

This signal control approach was initially developed for an isolated arterial signalized intersection and disregarded the arterial queue storage downstream of the intersection. The method in Su et al. (2014) was extended and modified to account for limited queue storage downstream of an arterial intersection, in the case of multiple arterial intersections near a freeway with multiple on-ramps accessed by arterial traffic. Thus, a new penalty term was added to the objective function presented in Phase I. The additional penalty term is the following:

$$\varepsilon \cdot \left(\sum_{i \epsilon AR} \left| f_{sat,ik} \cdot [o_{ik}(T) + l_{acc}] - \left[s_L - \sum_{i \epsilon L} q_{i,k+1}(T) \right] \right| \right) \tag{3.6}$$

where,

$o_{ik}(T)$: offset between phase *i* of intersection *k* and the corresponding phase in the downstream intersection during the *T*-th cycle
l_{acc}: start-up lost time
s_L: length of the downstream link (in number of vehicles)
$q_{i,k+1}(T)$: queue length of phase *i* of the downstream intersection at the end of cycle *T*
AR: indicator that the downstream section is an arterial
L: indicator for the downstream arterial link
ε: weighting factor for the penalty function

Figure 3.8 provides an illustration of the notations shown in Equation 4.8.

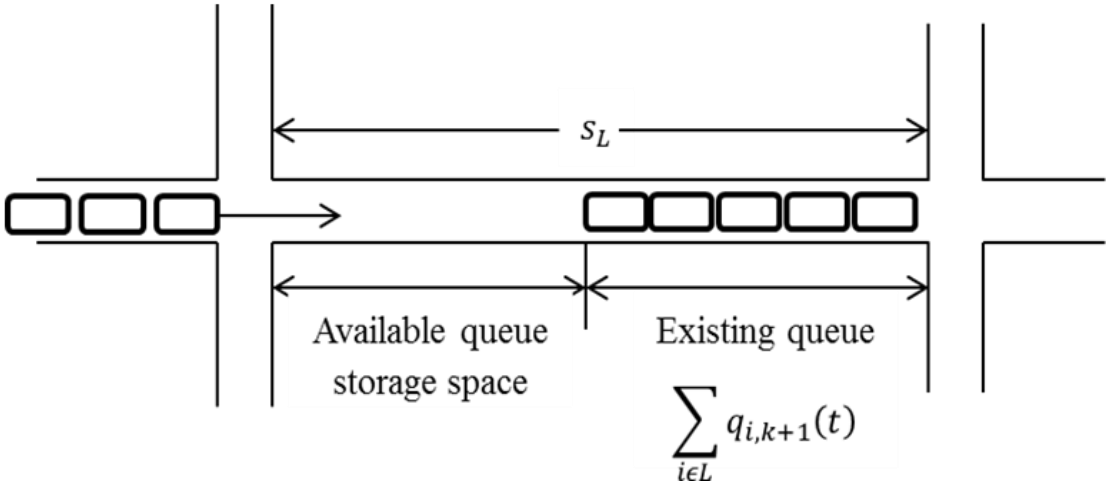


Figure 3.8 Queues at adjacent signalized intersections.

For adjacent signalized intersections, the offset between the corresponding phases of the upstream and downstream intersections causes the downstream intersection to initiate green at a later time, thus during the time between the beginning of green of phase *i* of intersection *k* and the beginning of green of the corresponding phase in the downstream intersection, the flow through phase *i* of intersection *k* is $f_{sat,ik} \cdot [o_{ik}(T) + l_{acc}]$. In order to reduce the upstream green time when the downstream section cannot accommodate the flow discharged from the upstream intersection,

the control strategy penalizes on the difference between the flow discharged upstream and the amount of queue storage space downstream so that more green time would be allocated to the conflicting directions instead. Thus, the new optimization model is formulated as the following:

$$
\begin{aligned}
Min \sum_{i \neq j} \mu_{ij} & \left| \frac{g_i(T)}{\frac{q_i(T-1) + d_i(T) \cdot C}{f_{sat,i}}} - v_{ij} \left(\frac{g_j(T)}{\frac{q_j(T-1) + d_j(T) \cdot C}{f_{sat,j}}} \right) \right| \\
& + \delta \left| \sum_r \left(\sum_{i \in R} f_{sat,i} \cdot g_i(T) \cdot \beta_i - (RA_r + RA_l) \right) \right| \\
& + \varepsilon \left| \sum_{i \in AR} \left(f_{sat,ik} \cdot [o_{ik}(T) + l_{acc}] - \left[s_L - \sum_{i \epsilon L} q_{i,k+1}(T) \right] \right) \right|
\end{aligned}
\tag{3.7}
$$

Subject to:

$$g_i(T) \geq G_{i,min} \tag{3.8}$$

$$g_1(T) + g_2(T) = g_5(T) + g_6(T) \tag{3.9}$$

$$\sum_{i=1-4 \; or \; 5-8} g_i(t) = C \tag{3.10}$$

However, this signal control strategy has its limitation because the arterial traffic signals are optimized under the less realistic assumption that vehicles arrive uniformly at the intersection. With signalized intersections nearby, this assumption no longer holds due to the fact that the time and distribution of arrivals from the upstream intersection depend heavily on the time at which the green phase begins and the duration of the green phase at the upstream intersection. Typically, traffic from the upstream intersection arrives in platoons during a short time interval instead of uniformly over a longer period of time. As a result, the misrepresentation of arrival pattern could have increased the delay of each turning movement at the selected intersection. This limitation would lead to less optimal results in the analysis of an arterial corridor as well. Furthermore, it is not feasible to update and modify offsets in real time, especially at a frequency of every cycle. In order to address this limitation, a new signal control strategy was developed and is discussed in the next sub-section.

3.3.2 Control Strategy II

A new control strategy for the signalized intersections near the freeway on-ramps was developed to account for the more realistic assumption that traffic arrive in platoons from the upstream arterial traffic signals. The proposed approach maintains the existing freeway ramp metering algorithm in practice and is not limited to any specific freeway ramp metering algorithm. According to this strategy, signal timings are optimized for all signalized intersections along the arterial and updated in real time at the beginning of every cycle, in order to minimize delay. Similar to the approach previously proposed, the cycle is fixed and assumed to be known in advance.

The following notations will be used:

u: intersection index
$g_{i,T}^{u}$: green time allocated to phase i in cycle T at intersection u [sec]
$g_{i,min}^{u}$: minimum green time of phase i at intersection u [sec]
$g_{i,max}^{u}$: maximum green time of phase i at intersection u [sec]
I^{u}: total number of phases in a cycle for intersection u
L^{u}: total lost time at intersection u [sec]
C: cycle length [sec]
$R_j^{(1)u}(g_{i,T}^{u})$: red time from the beginning of the cycle to the beginning of the green for the subject lane group
$G_j^{e,u}(g_{i,T}^{u})$: duration of effective green time for the subject lane group
$R_j^{(2)u}(g_{i,T}^{u})$: red time from the end of the green to the end of the cycle for the subject lane group

$$R_j^{(1)u}(g_{i,T}^{u}) = \sum_{i=1}^{k_j^u-1} g_{i,T}^{u} + \sum_{i=1}^{k_j^u-1} y_i^u$$

$$G_j^{e,u}(g_{i,T}^{u}) = \sum_{i=k_j^u}^{l_j^u} g_{i,T}^{u} + \sum_{i=k_j^u}^{l_j^u-1} y_i^u$$

$$R_j^{(2)u}(g_{i,T}^{u}) = \sum_{i=l_j^u+1}^{I^u} g_{i,T}^{u} + \sum_{i=l_j^u}^{I^u} y_i^u$$

y_i^u: yellow time interval after phase i at intersection u
$g_{i,T}^{u}$: green time for phase i in cycle T at intersection u
k_j^u: the first phase in a cycle that can serve lane group i at intersection u
l_j^u: the last phase in a cycle that can serve lane group i at intersection u

The new optimization model is the following:

Minimize

$$\sum_{u=1}^{N_i}\sum_{j=1}^{J_u}\left(D_{j,T}^{H,u} + D_{j,T}^{T,u}\right) + \sum_{u=1}^{N_i}\sum_{j=1}^{J_u} D_{j,T}^{Q,u} + \sum_{r=1}^{N_r} D_T^r \qquad (3.11)$$

Subject to:

$$g_{i,min}^{u} \le g_{i,T}^{u} \le g_{i,max}^{u} \qquad (3.12)$$

$$\sum_{i=1}^{I^u} g_{i,T}^u + L^u = C \tag{3.13}$$

$$N_T^r \leq RS_r \tag{3.14}$$

The objective function, shown in equation 3.11, consists of the delay at the signalized intersections and the delay at the freeway on-ramps. The delay at the signalized intersection includes the delay for vehicles that travel in platoons on incoming links and the delay for vehicles in residuals queues. $D_{j,T}^{H,u}$ and $D_{j,T}^{T,u}$ denote the delays caused by stopping the head and tail of the platoon, respectively, $D_{j,T}^{Q,u}$ is the delay experienced by vehicles in the residual queues, D_T^r is the delay experienced by vehicles on the on-ramps, N_i is the number of signalized intersections along the arterial, N_r is the number of on-ramps in the system, and J_u is the number of lane groups at intersection u. These delays are expressed as functions of the optimal green times, which are the decision variables. The optimal green times also determine the beginning of the coordinated phase, which is typically the arterial's through movement. Thus, the offsets are effectively optimized.

There are several practical constraints. equation 3.12 is a constraint for maximum and minimum allowable green time, equation 3.13 is a constraint that ensures the sum of all green times equals the cycle length, and equation 3.14 is a constraint that ensures the number of vehicles in the residual queue of on-ramp r at the end of cycle T, N_T^r, does not exceed RS_r, which is the number of vehicles that on-ramp r can accommodate at jam density.

The delays at the signalized intersections are estimated under the assumption of deterministic vehicle arrivals, fixed phase sequences, constant lane capacities, and negligible platoon dispersion. Thus, the vehicles arrive and are served at the intersection at capacity, and that all vehicle trajectories are parallel assuming that kinematic wave theory (Lighthill and Whitman, 1955; Richards, 1956) holds. This implies that the first and last vehicle that get stopped in a platoon will experience the same delay, thus the total delay of all vehicles can be easily calculated using only the arrival time of the first vehicle in a platoon at the back of its lane group's queue at intersection u, $t_{j,T}^u$, the size of the platoon, $P_{j,T}^u$, and the traffic conditions at the approach as expressed by the size of the residual queue of lane group j at the end of the previous cycle $T-1$, $N_{j,T-1}^u$. The size of the platoon in lane group j at intersection, u, during cycle T, $P_{j,T}^u$ can be determined using the sum of the flows of all movements that discharge traffic toward intersection u during the last cycle, which can be measured by the stop line detectors on the corresponding lanes of the upstream intersection, assuming there are reasonable estimates of the percentages of traffic choosing the left turn, thru, and right turn lanes at intersection u. For simplicity and ease of computation, the delay calculations are done under the assumption of single platoon, however, the control strategy could be extended to include multiple platoons arriving within a cycle. For the arrival time of the first vehicle in a platoon at the back of lane group j's queue at intersection u, $t_{j,T}^u$, the sum of the time at which the first vehicle of the upstream intersection's first phase or turning movement that discharges vehicle toward intersection u leaves the stop bar detector and the free-flow travel time from the upstream intersection to intersection u would provide an accurate estimate. With an accurate estimate of the platoon arrival time, the effect of disruption of the progression can be accounted for in the delay estimation. However, in order to optimize multiple signals at the same time, the arrival time of the platoon and the start time of the cycle

must be normalized by the free-flow travel time from each intersection to the critical intersection, a method known as coordinate transformation proposed by Newell (1989). Lastly, the size of the residual queue of lane group j at the end of the previous cycle $T-1$, $N_{j,T-1}^{u}$, can be calculated using the measured parameters. In summary, the delay estimation requires minimal field measurements, and the measurements can be obtained using existing loop detectors commonly placed at the stop bars of the signalized intersections.

The delay estimations for signalized intersections vary on a case-by-case basis, and can be calculated based on the details shown below. In order to determine which of the cases is applicable, a binary decision variable and the related constraints for time of platoon arrival and residual queue can be introduced, and the summation of the binary variables for a lane group should be constrained to one in order to apply only one delay equation for each lane group of a signalized intersection.

Case 1: Arrival before residual queue served, entire platoon served in green

A platoon of size $P_{j,T}^{u}$ that belongs to lane group j of intersection u arrives at the back of its lane group's queue during cycle T at time $t_{j,T}^{u}$ before the time that the corresponding residual queue of j from the previous cycle $T-1$, $N_{j,T-1}^{u}$, would have finished being served if there was enough green time available. There is enough available green time to serve the residual queue, and spare green time to serve all $P_{j,T}^{u}$ vehicles in the platoon. These conditions are summarized as:

$$t_{j,T}^{u} \leq t_{T}^{u} + R_{j}^{(1)u}\left(g_{i,T}^{u}\right) + \frac{N_{j,T-1}^{u}}{s_{j}^{u}}$$

$$N_{j,T-1}^{u} \leq G_{j}^{e,u}\left(g_{i,T}^{u}\right)s_{j}^{u}$$

$$P_{j,T}^{u} \leq G_{j}^{e,u}\left(g_{i,T}^{u}\right)s_{j}^{u} - N_{j,T-1}^{u}$$

where s_{j}^{u} is the saturation flow for lane group j at intersection u and t_{T}^{u} is the beginning of cycle T for intersection u, normalized by the free-flow travel time between intersection u and the critical intersection, which is determined as:

$$t_{T}^{u} = t_{T} + O_{T}^{u} - \frac{d_{cr}^{u}}{v_{f}}$$

$t_{T} = (T-1)C$ is the beginning of cycle T at the critical intersection which is the first one to be optimized and O_{T}^{u} is the difference between the starting time of cycle T at intersection u and the critical intersection, normalized by the free-flow travel time of between intersection u and the critical intersection. The number of vehicles in the residual queue, $N_{j,T-1}^{u}$, is calculated as:

$$N_{j,T-1}^{u} = \max\left\{P_{j,T-1}^{u} + N_{j,T-2}^{u} - G_{j}^{e,u}\left(g_{i,T-1}^{u}\right)s_{j}^{u}, 0\right\}$$

As shown in figure 3.9, all vehicles in the platoon experience delay caused by stopping the head of the platoon, $D_{j,T}^{H,u}$:

$$D_{j,T}^{H,u} = P_{j,T}^{u}\left(t_T^u + R_j^{(1)u}\left(g_{i,T}^u\right) + \frac{N_{j,T-1}^u}{s_j^u} - t_{j,T}^u\right)$$

but no delay caused by stopping the tail of the platoon, $D_{j,T}^{T,u}$.

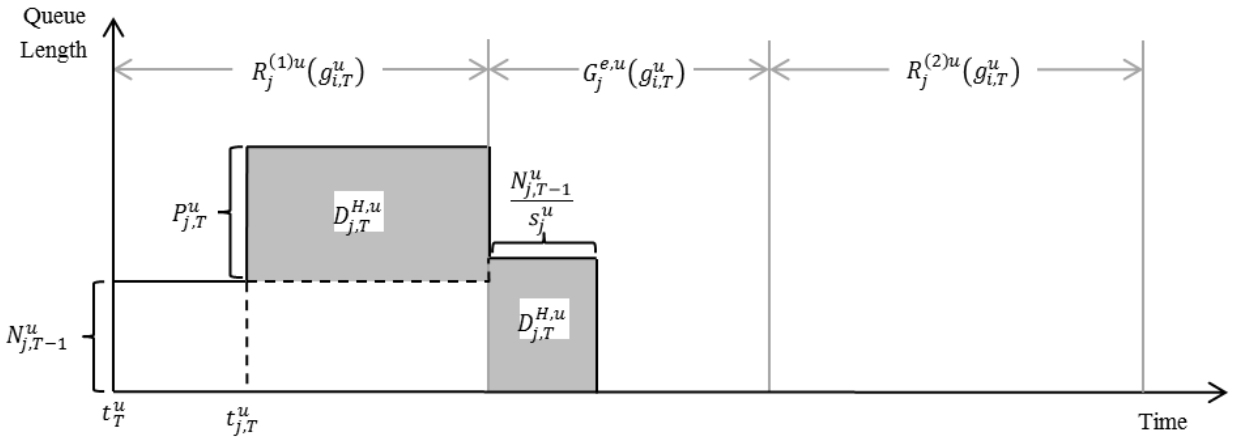


Figure 3.9 Delay of platoon in Case 1.

Case 2: Arrival before residual queue served, insufficient green to serve entire platoon

A platoon of size $P_{j,T}^u$ arrives at the back of its lane group's queue at time $t_{j,T}^u$ before the time that the corresponding residual queue $N_{j,T-1}^u$, would have finished being served. There is enough available green time to serve the residual queue, but there is not enough spare green time to serve all $P_{j,T}^u$ vehicles. These conditions are summarized as:

$$t_{j,T}^u \le t_T^u + R_j^{(1)u}\left(g_{i,T}^u\right) + \frac{N_{j,T-1}^u}{s_j^u}$$

$$N_{j,T-1}^u \le G_j^{e,u}\left(g_{i,T}^u\right)s_j^u$$

$$P_{j,T}^u \ge G_j^{e,u}\left(g_{i,T}^u\right)s_j^u - N_{j,T-1}^u$$

According to figure 3.10, all vehicles in the platoon experience delay caused by stopping the head of the platoon, $D_{j,T}^{H,u}$, and a portion of the vehicles experience delay caused by stopping the tail of the platoon, $D_{j,T}^{T,u}$:

$$D_{j,T}^{H,u} = P_{j,T}^{u}\left(t_T^u + R_j^{(1)u}\left(g_{i,T}^u\right) + \frac{N_{j,T-1}^u}{s_j^u} - t_{j,T}^u\right)$$

$$D_{j,T}^{T,u} = \left(P_{j,T}^u - G_j^{e,u}\left(g_{i,T}^u\right)s_j^u + N_{j,T-1}^u\right)\left(C - \frac{N_{j,T-1}^u}{s_j^u}\right)$$

For linearity, the delay caused by stopping the tail of the platoon is equal to the cycle length minus the time it takes to serve the residual queue. Although this overestimates the delay caused by stopping the tail of the platoon by $R_j^{(1)u}(g_{i,T}^u)$, it maintains computational simplicity.

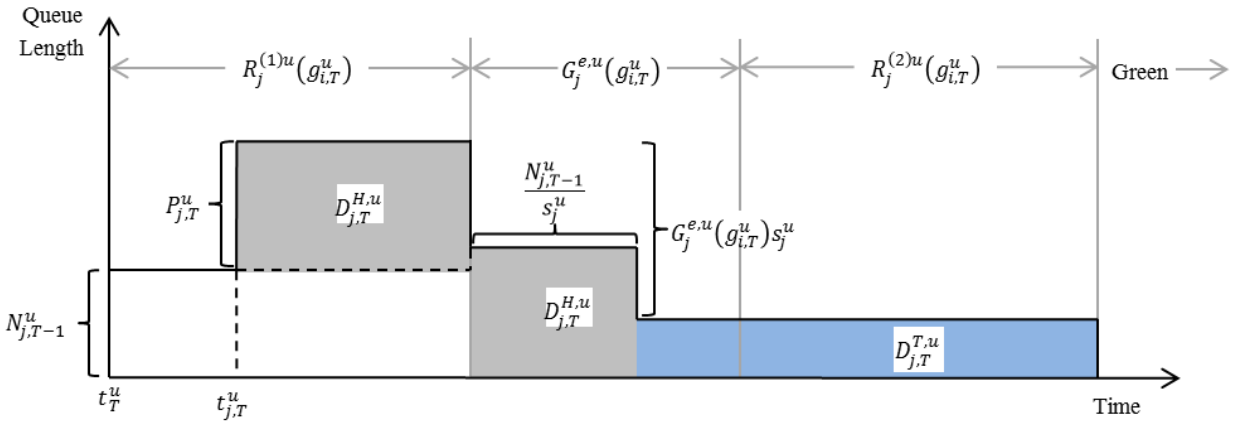


Figure 3.10 Delay of platoon in Case 2.

Case 3: Arrival before end of green, insufficient green to serve residual queue

A platoon of size $P_{j,T}^u$ arrives at the back of its lane group's queue at time $t_{j,T}^u$, before the end of the phase that can serve it, but there is not enough available green time to serve all $N_{j,T-1}^u$ vehicles in the residual queue. These conditions are summarized as:

$$t_{j,T}^u \leq t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u)$$

$$N_{j,T-1}^u \geq G_j^{e,u}(g_{i,T}^u)s_j^u$$

As shown in figure 3.11, all vehicles in the platoon experience delay caused by stopping the head of the platoon. $D_{j,T}^{H,u}$, and by stopping the tail of the platoon, $D_{j,T}^{T,u}$:

$$D_{j,T}^{H,u} = P_{j,T}^u\left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u\right)$$

$$D_{j,T}^{T,u} = P_{j,T}^u\left(R_j^{(2)u}(g_{i,T}^u)\right)$$

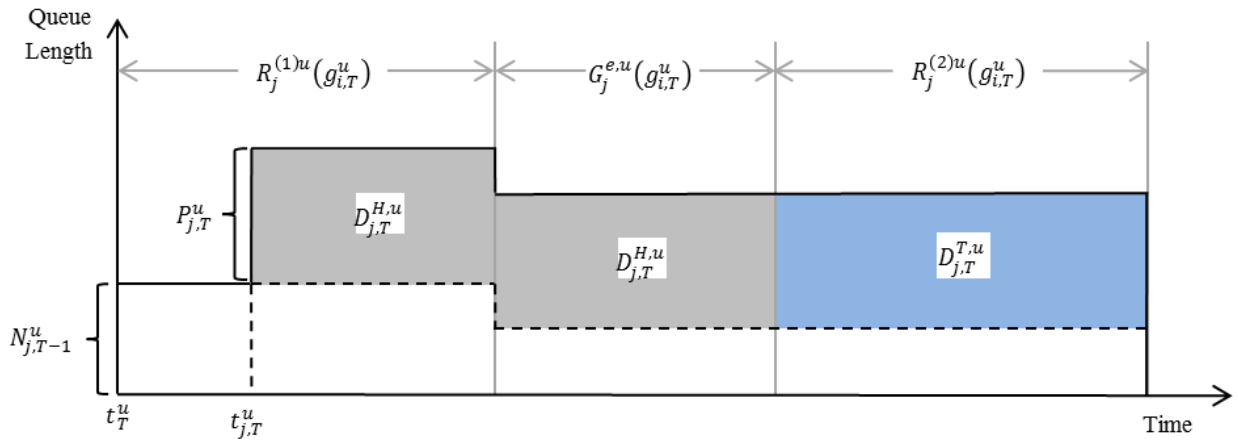


Figure 3.11 Delay of platoon in Case 3.

Case 4: Arrival after residual queue served, entire platoon served in green

A platoon of size $P_{j,T}^u$ arrives at the back of its lane group's queue at time $t_{j,T}^u$ after the time that the corresponding residual queue $N_{j,T-1}^u$, would have finished being served. There is enough available green time to serve the residual queue, and there is enough spare green time to serve all $P_{j,T}^u$ vehicles. These conditions are summarized as:

$$t_{j,T}^u \geq t_T^u + R_j^{(1)u}(g_{i,T}^u) + \frac{N_{j,T-1}^u}{s_j^u}$$

$$N_{j,T-1}^u \leq G_j^{e,u}(g_{i,T}^u)s_j^u$$

$$P_{j,T}^u \leq \left(t_T^u + R_j^{(1)u}(g_{i,T}^u) + G_j^{e,u}(g_{i,T}^u) - t_{j,T}^u\right)s_j^u$$

In this case, vehicles in the platoon do not experience any delay at intersection u. As a result, both the delay caused by stopping the head of that platoon, $D_{j,T}^{H,u}$, and by stopping the tail of the platoon, $D_{j,T}^{T,u}$, are zero, as illustrated in figure 3.12.

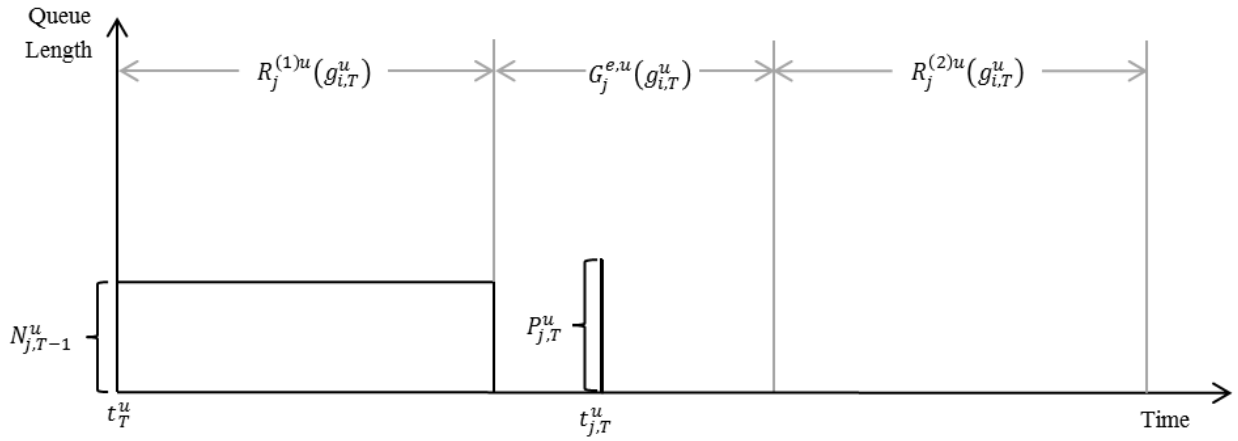


Figure 3.12 Delay of platoon in Case 4.

Case 5: Arrival after residual queue served, insufficient green to serve entire platoon

A platoon of size $P_{j,T}^u$ arrives at the back of its lane group's queue at time $t_{j,T}^u$ after the time that the corresponding residual queue, $N_{j,T-1}^u$, would have finished being served. There is enough available green time to serve the residual queue, but there is not enough spare green time to serve all $P_{j,T}^u$ vehicles. These conditions are summarized as:

$$t_{j,T}^u \geq t_T^u + R_j^{(1)u}\left(g_{i,T}^u\right) + \frac{N_{j,T-1}^u}{s_j^u}$$

$$N_{j,T-1}^u \leq G_j^{e,u}\left(g_{i,T}^u\right)s_j^u$$

$$P_{j,T}^u \geq \left(t_T^u + R_j^{(1)u}\left(g_{i,T}^u\right) + G_j^{e,u}\left(g_{i,T}^u\right) - t_{j,T}^u\right)s_j^u$$

A portion of vehicles in the platoon experience delay caused by stopping the tail of the platoon, $D_{j,T}^{T,u}$:

$$D_{j,T}^{T,u} = \left[P_{j,T}^u - \left(t_T^u + R_j^{(1)u}\left(g_{i,T}^u\right) + G_j^{e,u}\left(g_{i,T}^u\right) - t_{j,T}^u\right)s_j^u\right]\left(t_{T+1}^u - t_{j,T}^u\right)$$

but no delay caused by stopping the head of the platoon, $D_{j,T}^{H,u}$, shown in figure 3.13. This portion of vehicles in the platoon will also experience residual queue delay in the next cycle.

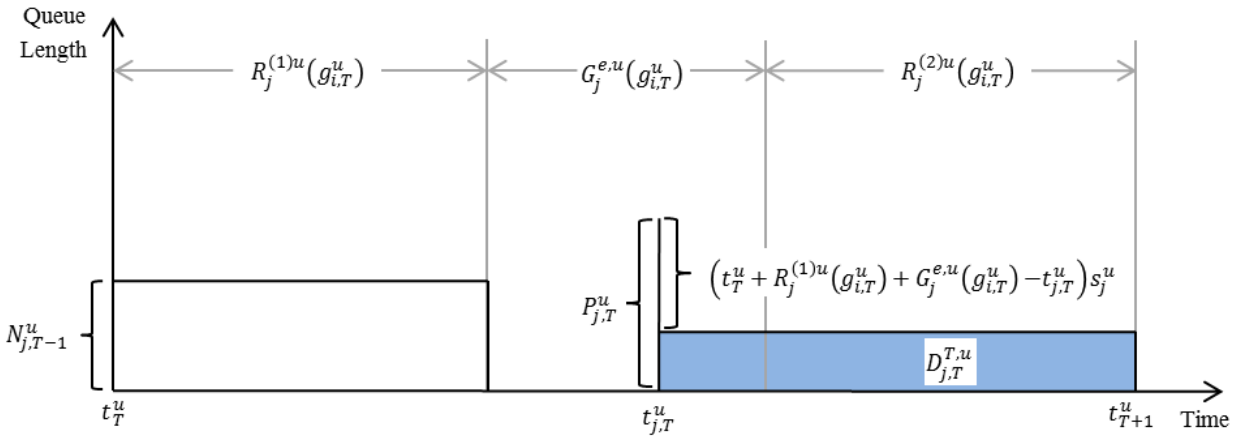


Figure 3.13 Delay of platoon in Case 5.

Case 6: Arrival after the green

A platoon of size $P_{j,T}^u$ arrives at the back of its lane group's queue at time $t_{j,T}^u$ after the end of the phase that can serve it. This case captures all arrivals not satisfying the conditions of cases 1 through 5, and it can all be expressed as

$$t_{j,T}^u > t_T^u + R_j^{(1)u}\left(g_{i,T}^u\right) + G_j^{e,u}\left(g_{i,T}^u\right)$$

All vehicles in the platoon experience delay caused by stopping the tail of the platoon, $D_{j,T}^{T,u}$:

$$D_{j,T}^{T,u} = P_{j,T}^{u}\left(t_{T+1}^{u} - t_{j,T}^{u}\right)$$

but no delay caused by stopping the head of the platoon, $D_{j,T}^{H,u}$, as illustrated in figure 3.14.

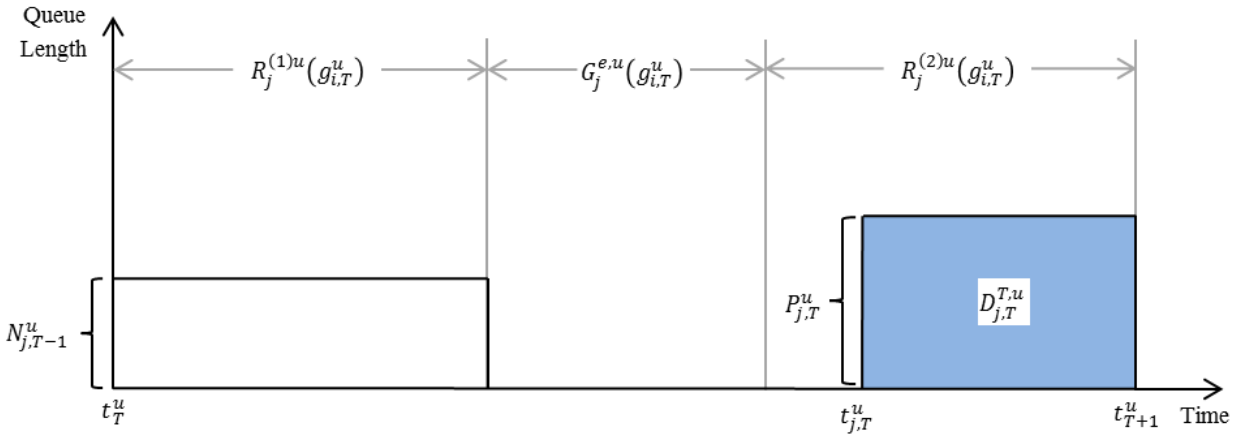


Figure 3.14 Delay of platoon in Case 6.

The delay for vehicles in residual queues at signalized intersections are estimated based on the size of the residual queue and whether or not it can be entirely served during cycle T. Two cases arise which are described next along with the corresponding delay equations for an intersection u:

Case A: Residual queue served in green

The residual queue of a lane group j, $N_{j,T-1}^{u}$, can be entirely served during cycle T:

$$N_{j,T-1}^{u} \leq G_j^{e,u}\left(g_{i,T}^{u}\right)s_j^u$$

Shown in figure 3.15, the total delay experienced by all vehicles in the residual queue, $D_{j,T}^{Q,u}$, is

$$D_{j,T}^{Q,u} = N_{j,T-1}^{u} R_j^{(1)u}\left(g_{i,T}^{u}\right)$$

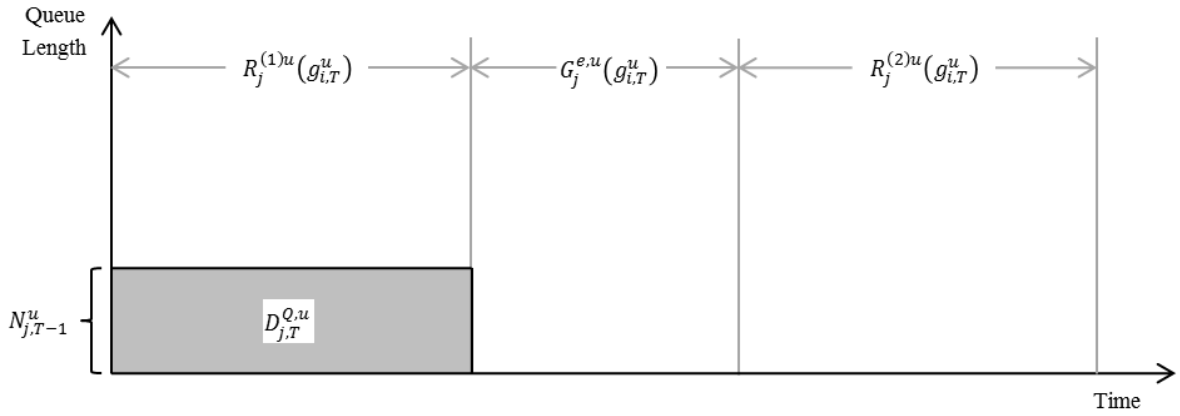


Figure 3.15 Residual queue delay of Case A.

Case B: Residual queue not entirely served in green

The residual queue of a lane group j, $N_{j,T-1}^{u}$, cannot be entirely served during cycle T:

$$N_{j,T-1}^{u} \geq G_j^{e,u}\left(g_{i,T}^{u}\right)s_j^{u}$$

Shown in figure 3.16, the total delay experienced by all vehicles in the residual queue, $D_{j,T}^{Q,u}$, is:

$$D_{j,T}^{Q,u} = N_{j,T-1}^{u}R_j^{(1)u}\left(g_{i,T}^{u}\right) + \left(N_{j,T-1}^{u} - G_j^{e,u}\left(g_{i,T}^{u}\right)s_j^{u}\right)C$$

because the vehicles that do not get served will have to wait for an extra cycle before they start being served.

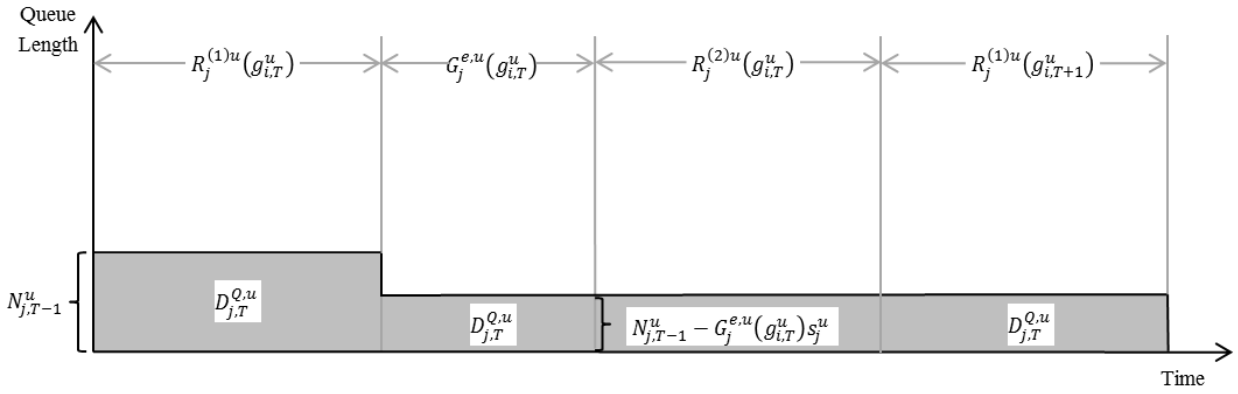


Figure 3.16 Residual queue delay of Case B.

The delay incurred by vehicles at the freeway on-ramps is determined by the number of vehicles discharged at the ramp meter, the number of vehicles already queuing on the freeway on-ramp, and the number of vehicles arrived at the ramp meter, and those that cannot be discharged during cycle T incur delays equivalent to the duration of the entire cycle. The delay at the freeway on-ramp, D_T^r can be expressed as:

$$D_T^r = \left[\left(\sum_{j\in ramp} G_j^{e,u}\left(g_{i,T}^{u}\right)s_j^{u}\right) + N_{T-1}^{r} - RM_T^r \cdot C\right]C$$

where $\sum_{j\in ramp} G_j^{e,u}\left(g_{i,T}^{u}\right)s_j^{u}$ is the number of vehicles discharged from the corresponding lane groups at the upstream signalized intersection to the freeway on-ramp r, during cycle T. RM_T^r is the ramp meter rate of freeway on-ramp r during cycle T, and C is the duration of the cycle T. The residual queue from the last cycle N_{T-1}^r can be calculated using the following expression:

$$N_{T-1}^{r} = \left(\sum_{j\in ramp} P_{j,T-1}^{u}\right) + N_{j,T-2}^{r} - RM_{T-1}^{r} \cdot C$$

where $\sum_{j \in ramp} P_{j,T-1}^{u}$ is the overall platoon size or total flow of vehicles discharged from the upstream intersection to the freeway on-ramp during the last cycle, $N_{j,T-2}^{r}$ is the number of vehicles in residual queue from the cycle before the last cycle, and RM_{T-1}^{r} is the ramp meter rate used in the last cycle.

The ramp meter rate is an input rather than a decision variable, thus the green times that are assigned to the phases that discharge traffic from the arterials to the freeway on-ramps must be reduced and assigned to the conflicting phases in order to reduce total delay. The total reduction in delay at the ramp meter and in the conflicting directions of the arterial traffic would outweigh the additional delay imposed on the on-ramp traffic due to limiting its corresponding green time at the signalized intersection upstream, thereby reducing the overall delay and preventing on-ramp queue spillbacks and queue overrides.

Despite the more realistic assumption that vehicles arrive in platoons from signalized intersections upstream, this new control strategy is not feasible for real-world implementation because it requires solving mix-integer linear programming (MILP) every cycle in order to determine which of the above cases apply, which requires computation times that may be longer than the cycle length. More importantly, this approach still maintains the predetermined cycle length, similar to the previous control strategy. This still does not address the major flaw of the signal control presented in previous literature (Tian et al, 2005; Su et al, 2014). Predetermined signal cycle lengths are long and intended to maximize intersection capacity under the assumption that the on-ramp downstream of the intersection has unlimited storage for queued vehicles. Similar to the approach in the previous sub-section, this approach does not address the importance of selecting the appropriate cycle length to avoid queue spillback when the freeway on-ramp has queue storage constraints but reallocates green times instead. The reduced green times for phases with on-ramp access can result in significant queue spillback towards the upstream arterial segment and intersection, and the additional green allocated to the conflicting directions may be underutilized if too much green is redistributed to these conflicting phases.

3.3.3 Control Strategy III

A simple and readily implementable control strategy is presented to resolve the inefficient control of freeway ramp metering and arterial traffic signals, which fail to recognize oversaturation of metered on-ramps by using long cycle lengths and providing progression to the heavier direction, as well as activating queue override. The control strategy is intended to manage the on-ramp queues at most freeway corridors with metered on-ramps and adjacent arterials primarily used to facilitate freeway access.

The proposed approach maintains the existing freeway ramp metering algorithm in practice and is not limited to any specific freeway ramp metering algorithm. For the adjacent arterial traffic signals, this approach recommends that the signal timing plans for arterials adjacent to freeway on-ramps should be developed similar to signal timings of over-saturated arterials. Recent research in this area (Gettman et al, 2012) suggests that long cycle lengths and long green durations should be avoided, and instead, the cycle lengths and green durations should consider the on-ramp queue storage space and should be designed to avoid queue spillback.

Consider a freeway on-ramp and the adjacent signalized intersection with 3 phases (Figure 3.17). The on-ramp is metered with a rate *r(t)* with queue storage capacity Q_r (in number of vehicles). In this example, signal phases 1 and 2 serve the turning movements feeding the on-ramp, and the remaining phase does not have on-ramp access.

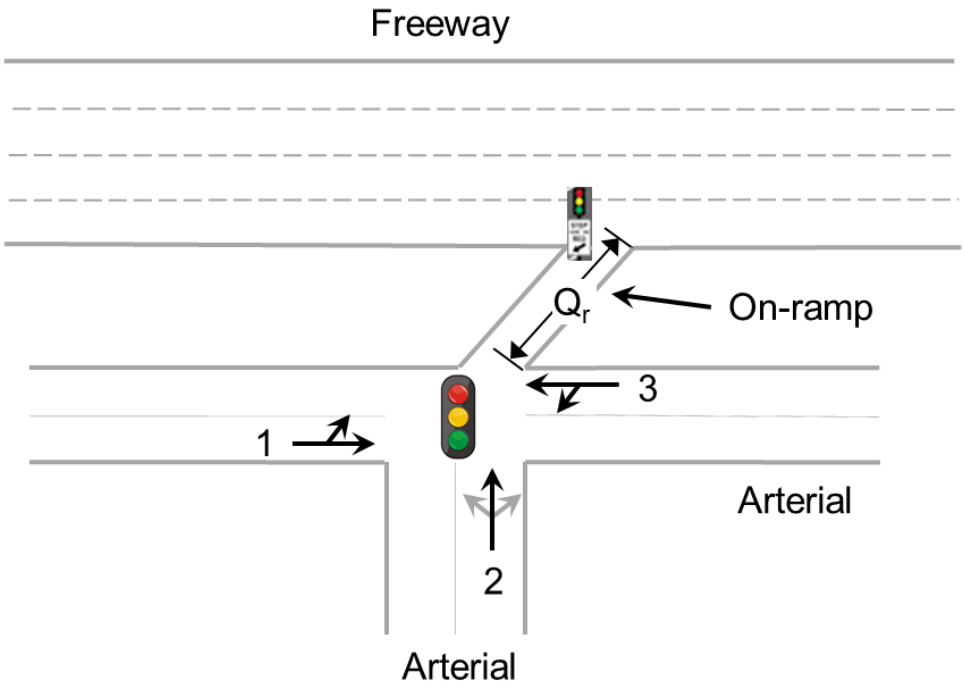


Figure 3.17 Example of a signalized intersection with on-ramp access.

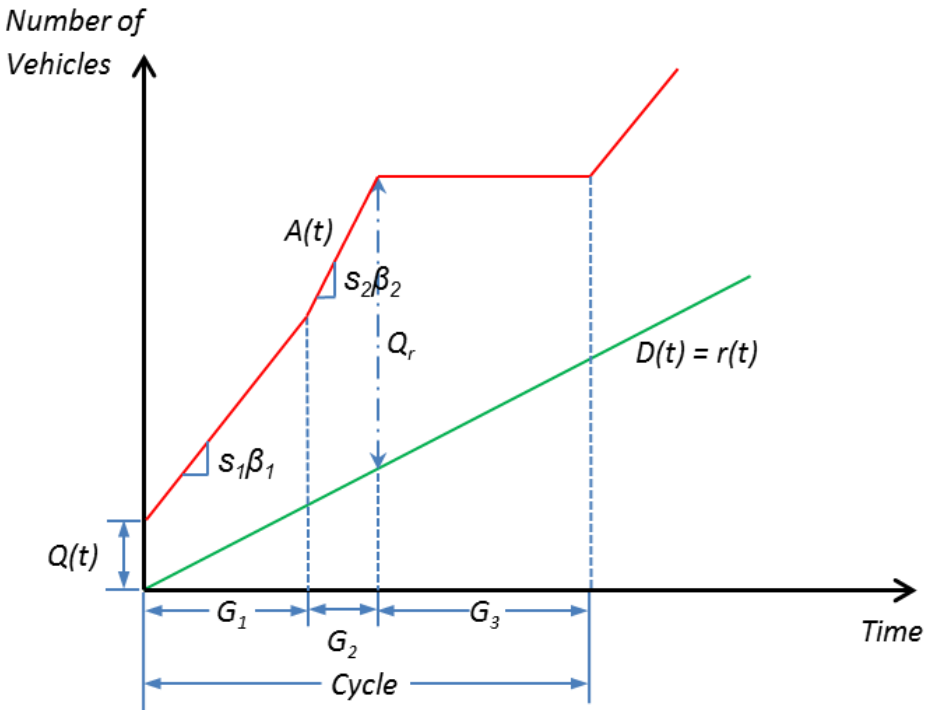


Figure 3.18 Queueing diagram of freeway on-ramp during a signal cycle.

The queuing diagram in figure 3.18 illustrates the upstream arrival pattern, downstream departure rate, and the excess accumulation at the freeway on-ramp when ramp metering is active. When the vehicle arrivals *A(t)* from phases 1 and 2 of the upstream signalized intersection are greater than the discharge flow *D(t),* then we have excess accumulation *Q(t)* on the on-ramp. When *Q(t)* exceeds the storage capacity Q_r (in number of vehicles), we have queue spillback that causes the activation of queue override.

Since the arrival rates (equal to saturation flows) from phases 1 and 2 exceed the ramp metering rate, the corresponding phase green times must terminate before the excess accumulation reaches the on-ramp queue storage capacity. The remaining signal phase, which serves the intersection approach that does not feed the on-ramp, can be served earlier, and at that time the on-ramp queue dissipates. As a result, vehicles from phases 1 and 2 can enter the on-ramp in the next signal cycle.

Under this strategy phases 1 and 2 receive shorter green times which may cause spillback on the arterial, as it was reported in the literature (Tian et al, 2005; Su et al, 2014). Such shortcoming can be remedied by maintaining the existing green distributions but using a shorter cycle length, as explained in the following sections.

The mathematical expressions for this control strategy are described as follows. First, the excess accumulation of the on-ramp must be estimated. Using the ramp meter rate $r(t)$ that is updated each time step *t*, the on-ramp excess accumulation *Q(t)* at time step *t* can be determined based on the following process:

$$Q(0) = 0$$

$$Q(1) = Q(0) + A(1) - D(1)$$

$$Q(2) = Q(1) + A(2) - D(2)$$

$$\vdots$$

$$Q(t) = Q(t-1) + A(t) - D(t)$$

Where $A(t)$ is the number of arrivals at the on-ramp during time step *t*, and $D(t)$ is the number of departures from ramp meter during time step *t*. The arrivals and departures can be measured by the loop detectors at the upstream and downstream ends of the freeway on-ramp, respectively. The on-ramp excess accumulation should be updated at the end of every cycle in order to perform real time control.

Equation 3.15 is developed using the queuing diagram in figure 3.18, to ensure that the green time for phase 1 and phase 2 must terminate at or before the excess accumulation reaches the maximum on-ramp queue storage capacity Q_r, therefore imposes an upper limit for the cycle length:

$$Q(t-1) + g_1 \cdot s_1 \cdot \beta_1 + g_2 \cdot s_2 \cdot \beta_2 - g_1 \cdot r(t) - g_2 \cdot r(t) - 2l \cdot r(t) \leq Q_r \qquad (3.15)$$

where,

g_1: effective green time of phase 1
g_2: effective green time of phase 2
s_1: saturation flow of phase 1
s_2: saturation flow of phase 2
β_1: percentage of demand of phase 1 that access the on-ramp
β_2: percentage of demand of phase 2 that access the on-ramp
$Q(t-1)$: residual on-ramp queue from the previous cycle.
$r(t)$: ramp metering rate
l: lost time of each phase (sec)

The effective green times can be expressed as functions of cycle length, based on commonly practiced approach derived from an *M/D/1* queuing model (Newell, 1989). Thus, g_1 and g_2 are expressed as the following:

$$g_1 = \frac{y_1}{Y} \cdot (C - 3l) \tag{3.16}$$

$$g_2 = \frac{y_2}{Y} \cdot (C - 3l) \tag{3.17}$$

where,

y_1: ratio of arrival rate and saturation flow of phase 1
y_2: ratio of arrival rate and saturation flow of phase 2
Y: sum of y's of the cycle
C: cycle length (sec)

Substitute equations 3.16 and 3.17 into equation 3.15, equation 3.15 can be expressed in terms of cycle length. Solving for cycle length in terms of the rest of the variables, the upper limit of cycle length is the following:

$$C \leq \frac{[Q_r - Q(t-1) + r(t) \cdot 2l] \cdot Y + 3l \cdot \left[\sum_{i=1,2} s_i \beta_i y_i - \sum_{i=1,2} r(t) y_i\right]}{\left[\sum_{i=1,2} s_i \beta_i y_i - \sum_{i=1,2} r(t) y_i\right]} \tag{3.18}$$

The upper limit of the cycle length should be updated at the end of every cycle in order to perform real time control that coordinates with freeway ramp metering.

Although the shorter cycle may introduce additional lost time, this ensures that all movements can clear the intersection at saturation flow rate therefore prevent loss in intersection capacity caused by queue spillback that impedes the flow of traffic during green phases.

The above selected cycle length does not provide for the maximum bandwidth of through traffic on the parallel arterial, which is appropriate for close to saturation arterial facilities that primarily provide access to multiple freeway on-ramps. The appropriate signal settings in this case, consist of shorter cycles and offsets that prevent long queues on the intersection approaches (Lieberman et al., 2000).

The proposed strategy is applicable to real-world multiphase signalized intersections by appropriately adjusting equations 1 through 4. Also, the proposed strategy does not require any new surveillance technologies or infrastructural changes and can be accomplished using the existing infrastructure and detection capabilities. It only requires modification of signal and ramp meter controller settings (i.e. cycle length, queue override) and communication between the two controllers.

Additionally, many signalized intersection are timed using the dual ring structure. A typical example shown in figure 3.7 is a four-leg signalized intersection with 8 phases, and each phase represents a turning movement. Phases 1 and 3 correspond to the leading left turns while phases 6 and 8 correspond to the lagging left turns, and the rest of the phases represent the through movements. There is a center barrier to prevent conflicts among turning movement of the north-south and the east-west directions.

As an example, phase 1 and phase 2 correspond to the turning movements feeding the on-ramp. Similarly, figure 3.19 illustrates the upstream arrival patterns, downstream departure rate, and the excess accumulation of the freeway on-ramp when ramp metering is active and the demand from the upstream arterial exceeds the capacity of the metered on-ramp. The methodology for determining the upper limit of the cycle length is the same as before, with the exception of how the effective green times are computed.

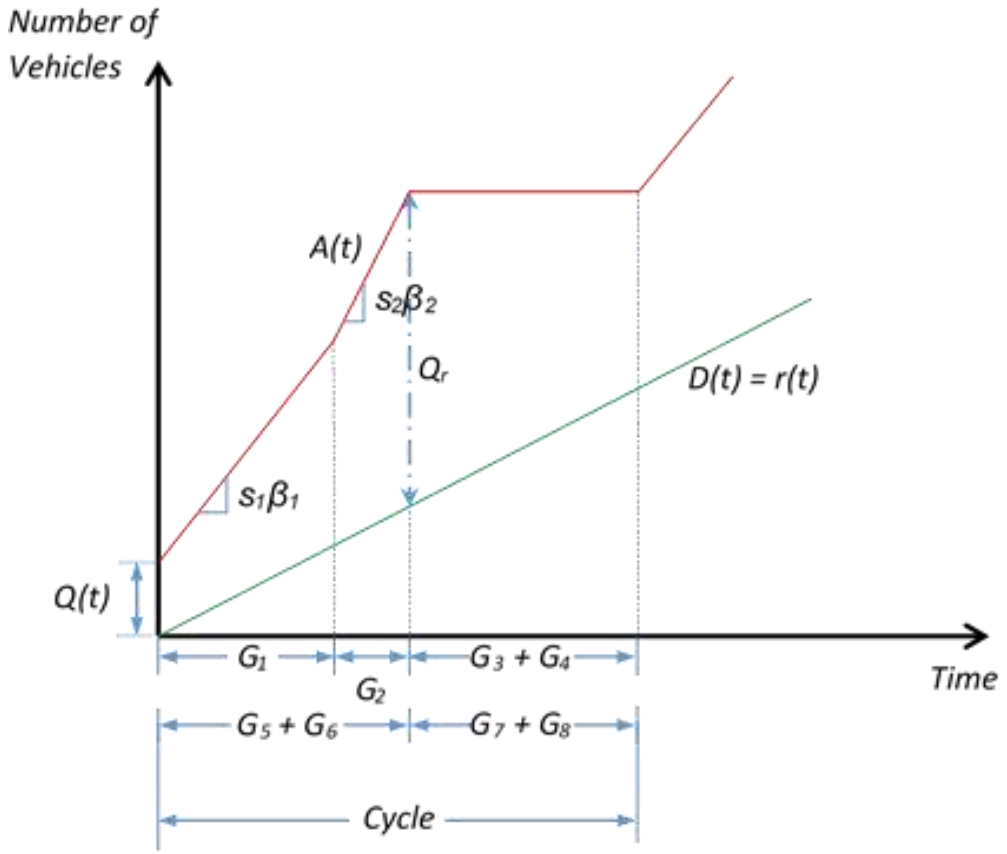


Figure 3.19 Queueing diagram of freeway on-ramp during a signal cycle (Dual Ring).

To formulate the expression for the effective green times, let y_L denote the ratio of arrival rate and saturation flow of all phases to the left of the barrier, similarly, let y_R denote the ratio of arrival rate and saturation flow of all phases to the right of the barrier. They are defined as the following:

$$y_L = \max(y_1 + y_2, y_5 + y_6) \tag{3.19}$$

$$y_R = \max(y_3 + y_4, y_7 + y_8) \tag{3.20}$$

Where y_i denotes the ratio of arrival rate and saturation flow of phase i, for $i = 1,2,\dots,7,8$.

Then, the effective green times for phase 1 and phase 2 are defined as the following:

$$g_1 = \frac{y_1}{y_1 + y_2} \cdot \frac{y_L}{y_L + y_R} \cdot (C - 4l) \tag{3.21}$$

$$g_2 = \frac{y_2}{y_1 + y_2} \cdot \frac{y_L}{y_L + y_R} \cdot (C - 4l) \tag{3.22}$$

For the remaining phases, the effective green times are:

$$g_3 = \frac{y_3}{y_3 + y_4} \cdot \frac{y_R}{y_L + y_R} \cdot (C - 4l) \tag{3.23}$$

$$g_4 = \frac{y_4}{y_3 + y_4} \cdot \frac{y_R}{y_L + y_R} \cdot (C - 4l) \tag{3.24}$$

$$g_5 = \frac{y_5}{y_5 + y_6} \cdot \frac{y_L}{y_L + y_R} \cdot (C - 4l) \tag{3.25}$$

$$g_6 = \frac{y_6}{y_5 + y_6} \cdot \frac{y_L}{y_L + y_R} \cdot (C - 4l) \tag{3.26}$$

$$g_7 = \frac{y_7}{y_7 + y_8} \cdot \frac{y_R}{y_L + y_R} \cdot (C - 4l) \tag{3.27}$$

$$g_8 = \frac{y_8}{y_7 + y_8} \cdot \frac{y_R}{y_L + y_R} \cdot (C - 4l) \tag{3.28}$$

Substitute equations 3.21 and 3.22 into equation 3.15, equation 3.15 can be expressed in terms of cycle length. Solving for cycle length in terms of the rest of the variables, the upper limit of cycle length is the following:

$$C \leq \frac{[Q_r - Q(t-1) + r(t) \cdot 2l] \cdot (y_L + y_R)}{\left[\frac{(s_1 \cdot \beta_1 - r(t)) \cdot y_1 y_L}{y_1 + y_2} + \frac{(s_2 \cdot \beta_2 - r(t)) \cdot y_2 y_L}{y_1 + y_2}\right]} \tag{3.29}$$

If phase 2 and phase 3, instead of phase 1 and phase 2, correspond to the turning movements feeding the on-ramp, then the upper limit of the cycle length becomes:

$$C \leq \frac{[Q_r - Q(t-1) + r(t) \cdot 2l] \cdot (y_L + y_R)}{\left[\frac{(s_2 \cdot \beta_2 - r(t)) \cdot y_2 y_L}{y_1 + y_2} + \frac{(s_3 \cdot \beta_3 - r(t)) \cdot y_2 y_L}{y_3 + y_4}\right]} \quad (3.30)$$

Similar to previous discussions, the upper limit of the cycle length must be updated at the end of every cycle in order to perform real time control that coordinates with freeway ramp metering. The effective green times can be computed using equations 3.21 through 3.28.

Overall, the above approaches are intended to prevent queue spillback at the on-ramp and further upstream at the arterial, in order to mitigate any unnecessary penalty imposed on the conflicting directions of the arterial traffic and eliminate the need for queue override at the metered on-ramps.

3.4 Evaluation: Simulation Tests

In order to evaluate the proposed signal control strategy (control strategy III), a microscopic simulation model was built in the AIMSUN microscopic simulation model (TSS, 2017) using the most up to date road geometry, lane configurations, and speed limits of the freeway and the arterial at the selected site.

3.4.1 Simulation Model Development

The simulation model covered three miles of northbound I-680, the on-ramps and off-ramps at Alum Rock Ave., McKee Rd., and Berryessa Rd., the parallel arterial Capitol Ave., and 15 signalized intersections. Freeway and arterial data obtained from the morning peak (7:00 AM to 9:30 AM) of September 23, 2015 were used for the inputs in demand and turning percentages. During the morning peak of September 23, 2015, freeway ramp meters and arterial traffic signals were functioning properly and there were no incidents within the selected corridor.

For the freeway, 5-minute interval loop detector data for flow were obtained from PeMS (Caltrans PeMS, 2016) and used as the demand input at the most upstream location of the simulation network and as the turning percentages at any applicable mainline-off-ramp split. Ramp metering rates and algorithms were obtained from Caltrans District 4 and modeled in microscopic simulation via AIMSUN API (Application Programming Interface). Ramp metering has been active since September 1, 2015.

For the arterial, video cameras were placed at the signalized intersections and arterial-on-ramp splits during the morning peak of September 23, 2015. Turning movement flows were recorded every 5 minutes, and they were used as the demand input at the entry points of the corridor (i.e. Southbound Capitol north of Berrryessa Rd.) and as turning percentages at the signalized intersections and the arterial-on-ramp splits. Field data suggest that majority of the on-ramp bound traffic come from the northbound left turns on Capitol Ave. The signal timing plans for the coordinated actuated signals were provided by the city of San Jose.

The model was calibrated to existing conditions prior to the evaluation of the proposed control strategy. Twenty replications of the simulation model runs with different random number seeds were made using the existing demand, turning percentages, ramp metering algorithm, and signal timing plans. Calibration results showed satisfactory agreement between the simulated and field observed freeway flow and speed and arterial flow. Detailed calibration results are shown in Appendix C.

Simulation tests were conducted to determine the effectiveness of the recommended control strategy III described in this chapter. First, the existing conditions, which employs long signal cycles and queue override at on-ramps were simulated as the reference case for comparison. Next, the proposed signal control with shorter signal cycles was simulated and queue override was not active during simulation. Twenty replications with different random seeds were run in each of the scenarios described above, and paired t-tests were conducted to determine the statistical significance of changes in performance after introducing the proposed signal control. The following performance metrics were used to analyze the simulation results:

- **Freeway mainline**: Total delay and vehicle miles of travel (VMT) of the entire peak period. Comparison of total delay is an indication of changes in speed and travel time on the freeway mainline and comparison of VMT implies any capacity changes at the freeway bottlenecks.

- **Parallel Arterial:** Average delays of all turning movements corresponding to the parallel arterial (Capitol Ave.) at each signalized intersection over the entire peak period. The sum of the average delays at each signalized intersection will be presented.

- **Cross Streets:** Average delays of all turning movements corresponding to the cross streets (Alum Rock Ave., McKee Rd., and Berryessa Rd.), across the entire peak period.

- **Total system:** Total delay of freeway and arterial during the entire peak.

Despite the satisfactory calibration results, the driving behavior model, especially the criteria pertaining to lane changing in congested conditions do not represent realistic driving behavior. Based on the preliminary results, the proposed control strategy eliminated the activation of the queue override which improved the freeway operating conditions. The total delay on the freeway mainline was reduced by 17.9%. Based on the field study presented in this chapter, the freeway capacity is expected to increase once queue override is avoided. However, there was not any statistically significant increase in freeway capacity, as indicated by the negligible change in total distance travelled shown in table 3.2. Therefore, the preliminary simulation results suggest that avoiding queue override simply redistributes delay from the freeway mainline to the on-ramp. As a result, the simulation model cannot properly simulate capacity drop, which implies that the simulated freeway performance may not be accurate.

Table 3.2 Preliminary freeway performance before and after proposed signal control.

	Before		After		% Difference	
Freeway Mainline						
	Total Delay (veh-hr)	Total Distance Traveled (veh-mile)	Total Delay (veh-hr)	Vehicle Miles of Travel (VMT)	Change in Total Delay	Change in VMT
I-680 NB	799.06	47235.16	655.81	46619.69	**-17.93%**	1.30%

3.4.2 Simulation Model Enhancement

An enhanced freeway driving behavior model was developed (Lu et al, 2017) to better simulate real world driving behavior and capacity drop on freeways. This external driving behavior model can be programmed into AIMSUN simulation via the MicroSDK (Mico-software development kit) feature, which allows the enhanced driving behavior model to override the driving behavior model provided by the AIMSUN simulation software. Detailed descriptions and formulations can be found in Appendix D. Furthermore, the enhanced model was calibrated using field data from a complex freeway corridor, and the model calibration is documented in Appendix D.

In order to demonstrate that the enhanced freeway driving behavior model is able to simulate capacity drop at on-ramp merge bottlenecks, a simple simulation experiment was built in the AIMSUN microscopic simulation framework (TSS, 2017) using the road geometry shown in figure 3.20. This test site consists of four freeway mainline lanes, an on-ramp, and an acceleration lane in the merging area. Vehicle counts were collected at 30-second intervals downstream of the merging area, as shown by the dotted line in figure 3.20.

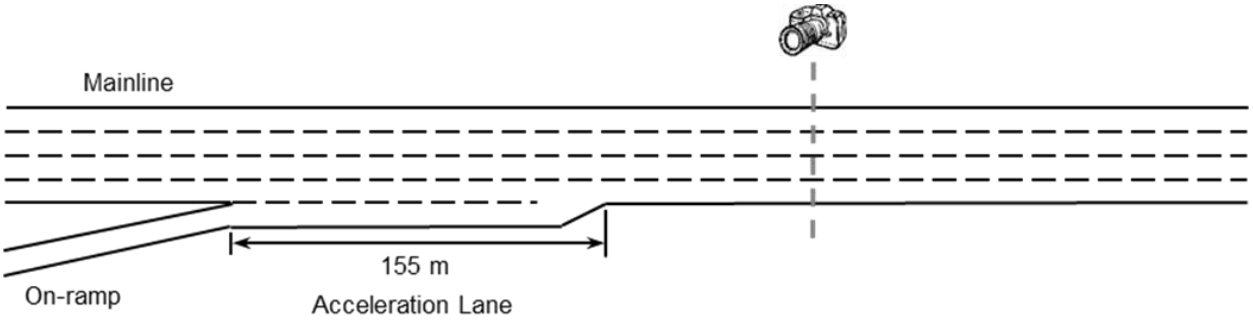


Figure 3.20 Multilane freeway merge bottleneck.

First, simulation experiments were run to determine the freeway capacity without any input from the on-ramp. The capacity is defined as the maximum 15-minute moving average flow rate, according to the Highway Capacity Manual (HCM, 2010). The experiment began with simulating a constant and relatively low traffic volume for one hour. If the freeway remained free-flowing, then subsequent simulations were conducted with slightly higher volume input (e.g., plus 1000 veh/hr), until congested conditions were observed and the highest observed 15-minute moving average flow no longer increased as the input became larger. This 15-minute moving average flow is the capacity of the 4-lane freeway mainline, based on five replications with different random seeds.

The maximum mainline traffic volume obtained from the previous simulation was then simulated with additional on-ramp flow to determine if the capacity previously observed can be sustained following the introduction of merging traffic. The simulation began with a 10-minute warm up, then a constant mainline input equal to the observed mainline capacity for 20 minutes, followed by a 40-minute period with an on-ramp input of 300 veh/hr in addition to the mainline input. A 20-minute period of no demand immediate followed, in order to dissipate any residual queues. Subsequently, the cycle repeated (without the warm up period) itself but for on-ramp demands of 600 veh/hr, 900 veh/hr, 1200 veh/hr, and 1500 veh/hr, each with different random seeds for five replications.

Figure 3.21 and table 3.3 reveal how freeway merge bottleneck capacity changes as on-ramp demand increases. As expected, the bottleneck capacity remained nearly the same when the on-ramp was relatively low (600 veh/hr or less), and capacity drop was observed at higher on-ramp demand. Overall, the enhanced model was able to simulate 5% to 8% capacity drop that arise from high on-ramp demand entering the freeway mainline. The percentages of capacity drop simulated are similar to those presented in the empirical studies (Banks, 1991; Hall and Agyemang-Duah, 1991; Persaud et al., 1998; Cassidy and Bertini, 1999; Cassidy and Rudjanakanoknad, 2005; Sahin and Altun, 2008; Oh and Yeo, 2012; Srivastava and Geroliminis, 2013; Yuan et al., 2015), thus the microscopic simulation can approximately reproduce the real world driving behavior on freeways.

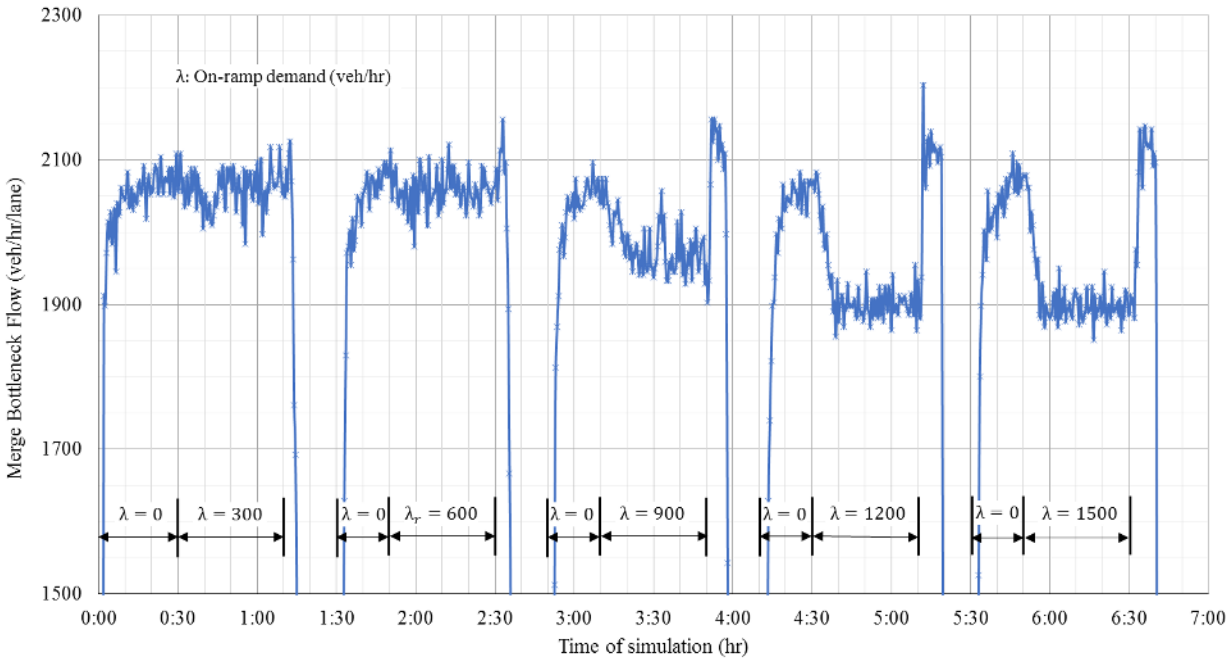


Figure 3.21 Multilane freeway merge bottleneck capacities.

Table 3.3. Merge bottleneck capacity under maximum mainline demand and varying levels of on-ramp demand.

λ: on-ramp demand (veh/hr)	Merge Bottleneck Capacity (veh/hr) (% change vs. λ = 0)
0	8280
300	8260 (-0.21 %)
600	8260 (-0.25 %)
900	7924 (-4.28 %)
1200	7640 (-7.70 %)
1500	7612 (-8.07 %)

CHAPTER 4: SIMULATION RESULTS

Comparison of simulation results prior to and after implementing the proposed signal control described in chapter 3 is summarized in table 4.1.

- **Freeway Mainline**: The proposed control strategy prevented the activation of the queue override which improved the freeway performance. The total delay on the freeway mainline was reduced by 11.13%, and the vehicle-miles of travel (VMT) was improved by 3.92%. Paired t-tests suggest that both improvements are statistically significant, as indicated by the numerical values in bold. The significant delay reduction can be attributed to the increase in mainline speed and the capacity of the freeway bottlenecks. The higher freeway capacity is indicated by the increase in VMT once queue override is not active, and this conforms to the observations of the empirical study in the previous chapter. The relatively small increase in VMT can be explained by the fact that queue override typically affects only 30 to 60 minutes of the 2.5 hour morning peak, rather than the entire period.

- **Parallel Arterial (Capitol Ave.):** Both directions of the parallel arterial experienced statistically significant change in average delay. The northbound direction, which consists of more than 60% left turn traffic intended to access the freeway on-ramps, experienced additional delay because it no longer benefits from the relaxed metering rate after queue override was avoided. However, the southbound direction of the parallel arterial remained mostly unaffected, due to the low demand for on-ramp access.

- **Cross Streets:** On the cross streets, delays were reduced because the shorter cycle lengths mitigated the long delays imposed on the cross streets. There were statistically significant delay reductions at McKee Rd. and Berryessa Rd. but almost none at Alum Rock Ave. This is because there is sufficient space for on-ramp queue storage near the Alum Rock Ave. on-ramp so the signal timing remained the same. The queue storage space were limited at the McKee Rd. and Berryessa Rd. on-ramps therefore the cycle lengths were reduced at the corresponding signalized intersections, which reduced the average delay on those cross streets. The westbound directions of the cross streets benefited much less than the eastbound directions because there is significant on-ramp bound arterial traffic in the westbound direction. As discussed earlier, the on-ramp bound traffic were penalized after avoiding the queue override's relaxed metering rate, thus negated some of the delay reduction experienced by those not intended to access the on-ramp (in the other lanes of the same westbound approach).

- **Total System:** The proposed control strategy reduced the system-wide delay by more than 5%.

Table 4.1 Performance before and after proposed signal control.

	Before		After		% Difference	
Freeway Mainline						
	Total Delay (veh-hr)	Total Distance Traveled (veh-mile)	Total Delay (veh-hr)	Vehicle Miles of Travel (VMT)	Change in Total Delay	Change in VMT
I-680 NB	833.41	43104.13	740.64	44792.95	**-11.13%**	**3.92%**
Arterial						
Average Delay on Main Parallel Arterial (min/veh)						
Capitol Ave NB	8.63		10.51		**21.84%**	
Capitol Ave SB	5.72		5.91		**3.33%**	
Average Delay of Cross Street (sec/veh)						
Alum Rock WB	48.05		47.33		-1.43%	
Alum Rock EB	37.27		37.82		1.47%	
McKee WB	56.76		52.34		**-7.79%**	
McKee EB	28.92		16.51		**-42.91%**	
Berryessa WB	47.27		39.26		**-16.73%**	
Berryessa EB	50.50		37.55		**-34.48%**	
Total System						
	Total Delay (veh-hr)		Total Delay (veh-hr)		Change in Total Delay	
Freeway & Arterial	2881.37		2727.19		**-5.65%**	

Bold (% Difference): Statistical significance

Sensitivity analyses were performed to evaluate the performance of the proposed signal control under fluctuations in demand during recurrent morning peaks. Simulation results corresponding to a 5% increase in both freeway and arterial demand are shown in table 4.2.

- **Freeway Mainline**: The proposed control strategy prevented the activation of the queue override which improved the freeway operating conditions. The total delay on the freeway mainline was reduced by 15.21%, and the vehicle-miles of travel (VMT) was improved by 5.15%. Similar to the results in table 4.1, paired t-tests suggest that both improvements are statistically significant, as indicated by the numerical values in bold. The significant delay reduction can be attributed to the increase in mainline speed and the capacity of the freeway bottlenecks. The higher freeway capacity is indicated by the increase in VMT once queue override is not active, and this conforms to the observations of the empirical study in the previous chapter. The relatively small increase in VMT can be explained by the fact that queue override typically affects only 30 to 60 minutes of the 2.5 hour morning peak, rather than the entire period.

- **Parallel Arterial (Capitol Ave.):** Both directions of the parallel arterial experienced statistically significant change in average delay. The northbound direction, which consists of more than 60% left turn traffic intended to access the freeway on-ramps, experienced additional delay because it no longer benefits from the relaxed metering rate after queue override was avoided. However, the southbound direction of the parallel arterial remained mostly unaffected, due to the low demand for on-ramp access.

- **Cross Streets:** Similar to the results in table 4.1, there was no significant change at Alum Rock Ave due to sufficient on-ramp queue storage near the Alum Rock Ave. on-ramp. On the other hand, the cross street delays were significantly reduced because of the shorter cycle lengths at the intersections with McKee Rd. and Berryessa Rd., but only for the eastbound direction. The westbound direction of McKee Rd. experienced no delay savings while the westbound direction of Berryessa Rd. was penalized with additional delay because there is significant on-ramp demand in the westbound direction that no longer benefit from the queue override's relaxed metering rate. This implies that the penalty imposed on on-ramp bound traffic negated some of the time savings experienced by those not intended to access the on-ramp (in the other lanes of the same westbound approach).

- **Total System:** The proposed control strategy increased the system-wide delay by more than 3%. However, higher VMT indicates that more trips were served and the average delay was lower.

Table 4.2 Performance before and after proposed signal control (+5% demand).

	Before		**After**		**% Difference**	
Freeway Mainline						
	Total Delay (veh-hr)	Total Distance Traveled (veh-mile)	Total Delay (veh-hr)	Vehicle Miles of Travel (VMT)	Change in Total Delay	Change in VMT
I-680 NB	1086.57	43238.12	921.30	45463.63	**-15.21%**	**5.15%**
Arterial						
Average Delay on Main Parallel Arterial (min/veh)						
Capitol Ave NB	12.24		16.57		**35.37%**	
Capitol Ave SB	6.02		7.06		**17.38%**	
Average Delay of Cross Street (sec/veh)						
Alum Rock WB	61.92		68.23		10.19%	
Alum Rock EB	43.43		48.15		10.87%	
McKee WB	58.67		58.27		-0.67%	
McKee EB	29.55		13.62		**-53.90%**	
Berryessa WB	49.80		59.46		**19.38%**	
Berryessa EB	52.97		39.49		**-34.15%**	
Total System						
	Total Delay (veh-hr)		Total Delay (veh-hr)		Change in Total Delay	
Freeway & Arterial	3693.72		3809.35		**3.04%**	

Bold (% Difference): Statistical significance

Simulation results corresponding to a 10% increase in both freeway and arterial demand are shown in table 4.3.

- **Freeway Mainline**: The proposed control strategy prevented the activation of the queue override which improved the freeway performance. The total delay on the freeway mainline was reduced by 19.53%, and the vehicle-miles of travel (VMT) was improved by 6.39%. Similar to the results in table 4.2, paired t-tests suggest that both improvements are statistically significant, highlighted by the bold percentage values. Higher freeway mainline speed and bottleneck capacity contributed to the delay saving. The higher freeway capacity is indicated by the increase in VMT once queue override is not active, and this agrees with the conclusions of the empirical study on queue override.

- **Parallel Arterial (Capitol Ave.):** Similar to the results in Table 4.2, both directions of the parallel arterial experienced statistically significant increase in average delay because the parallel arterial, which serves significant demand for on-ramp access, no longer receives the benefits of the less restrictive metering rate used in queue override. However, 10% increase on demand resulted even higher delay. This severely restricted the flow of on-ramp bound traffic on the arterial and resulted in oversaturation.

- **Cross Streets:** Similar to the results in tables 4.1 and 4.2, there was no significant change at Alum Rock Ave due to sufficient on-ramp queue storage near the Alum Rock Ave. Conversely, the cross street delays were significantly reduced because of the shorter cycle lengths at the intersections with McKee Rd. and Berryessa Rd., but only for the eastbound direction. As discussed in +5% demand scenario, there is significant on-ramp demand in the westbound direction that no longer benefit from queue override after the new control is implemented, which implies that the penalty imposed on on-ramp bound traffic negated some of the time savings experienced by those not intended to access the on-ramp (in the other lanes of the same westbound approach). Consequently, the westbound directions of McKee Rd. and Berryessa Rd. experienced additional delay.

- **Total System:** The proposed control strategy increased the system-wide delay by more than 6%, despite the similar magnitude of VMT increase. This suggests that the proposed signal control alone may not necessarily be sufficient if demand were to increase 10% or higher. More efficient ramp metering algorithm or higher freeway and arterial capacity may be needed.

Table 4.3 Performance before and after proposed signal control (+10% demand).

	Before		After		% Difference	
Freeway Mainline						
	Total Delay (veh-hr)	Total Distance Traveled (veh-mile)	Total Delay (veh-hr)	Vehicle Miles of Travel (VMT)	Change in Total Delay	Change in VMT
I-680 NB	1219.73	43246.01	981.52	46008.00	**-19.53%**	**6.39%**
Arterial						
Average Delay on Main Parallel Arterial (min/veh)						
Capitol Ave NB	14.86		23.28		**56.65%**	
Capitol Ave SB	6.39		9.66		**51.29%**	
Average Delay of Cross Street (sec/veh)						
Alum Rock WB	95.98		90.70		-5.50%	
Alum Rock EB	45.02		49.23		9.36%	
McKee WB	60.52		64.65		**6.82%**	
McKee EB	31.01		13.38		**-56.85%**	
Berryessa WB	52.90		73.27		**38.52%**	
Berryessa EB	54.72		42.79		**-27.87%**	
Total System						
	Total Delay (veh-hr)		Total Delay (veh-hr)		Change in Total Delay	
Freeway & Arterial	4154.35		4436.89		**6.37%**	

Bold (% Difference): Statistical significance

Simulation results corresponding to a 5% decrease in both freeway and arterial demand are shown in table 4.4.

- **Freeway Mainline**: The proposed control strategy prevented the activation of the queue override which improved the freeway performance. The total delay on the freeway mainline was reduced by 10.26%, and the vehicle-miles of travel (VMT) was improved by 2.98%. Paired t-tests suggest that both improvements are statistically significant, indicated by the bold percentage values. Similar to the results in table 4.1, the significant delay reduction is a result of the increase in mainline speed and capacity of the freeway bottlenecks. The higher freeway capacity is indicated by the increase in VMT once queue override is not active. The relatively small increase in VMT can be explained by the fact that queue override typically affects only 30 to 60 minutes of the 2.5 hour morning peak, rather than the entire period.

- **Parallel Arterial (Capitol Ave.):** Both directions of the parallel arterial experienced nearly no change in delay. The on-ramp bound traffic on the parallel arterial experienced longer delay after the less restrictive metering rate (in queue override) was avoided. However, the penalty was not large enough and offset by the delay savings experienced by those not intended to access the on-ramp, as a result of short cycle length.

- **Cross Streets:** Similar to the result in table 4.1, there were statistically significant delay reductions at McKee Rd. and Berryessa Rd., but almost no change in delay at Alum Rock Ave. due to the fact that sufficient on-ramp queue storage space near the Alum Rock Ave left the signal timing unchanged at Alum Rock Ave. The limited queue storage space at the McKee Rd. and Berryessa Rd. on-ramps reduced the signal cycle lengths and therefore the average delay on those cross streets. Similarly, the variation between the westbound and eastbound directions of the cross streets can be attributed to the significant on-ramp demand in the westbound direction that is penalized by the absence of queue override.

- **Total System:** The proposed control strategy reduced the system-wide delay by about 11%.

Table 4.4 Performance before and after proposed signal control (-5% demand).

	Before		After		% Difference	
Freeway Mainline						
	Total Delay (veh-hr)	Total Distance Traveled (veh-mile)	Total Delay (veh-hr)	Vehicle Miles of Travel (VMT)	Change in Total Delay	Change in VMT
I-680 NB	487.30	41680.20	437.30	42922.27	**-10.26%**	**2.98%**
Arterial						
Average Delay on Main Parallel Arterial (min/veh)						
Capitol Ave NB	7.03		6.86		-2.44%	
Capitol Ave SB	5.28		5.25		**-0.61%**	
Average Delay of Cross Street (sec/veh)						
Alum Rock WB	28.03		28.53		1.78%	
Alum Rock EB	31.32		31.57		0.79%	
McKee WB	53.84		47.49		**-11.79%**	
McKee EB	27.19		13.87		**-48.99%**	
Berryessa WB	44.68		33.81		**-13.14%**	
Berryessa EB	47.58		42.18		**-12.82%**	
Total System						
	Total Delay (veh-hr)		Total Delay (veh-hr)		Change in Total Delay	
Freeway & Arterial	2039.17		1832.25		**-11.29%**	

Bold (% Difference): Statistical significance

Simulation results corresponding to a 10% decrease in both freeway and arterial demand are shown in table 4.5.

- **Freeway Mainline**: The proposed control strategy prevented the activation of the queue override which improved the freeway performance. The total delay on the freeway mainline was reduced by 3.62%, and the vehicle-miles of travel (VMT) was improved by 2.38%. Paired t-tests suggest that both improvements are statistically significant, indicated by the bold percentage values. The relatively smaller delay savings can be attributed to the lower demand input and consequently shorter peak period. Similar to the results in table 4.1, the significant delay reduction is a result of the increase in mainline speed and capacity of the freeway bottlenecks. The higher freeway capacity is indicated by the increase in VMT once queue override is not active. The relatively small increase in VMT can be explained by the fact that queue override typically affects only 30 to 60 minutes of the 2.5 hour morning peak, rather than the entire period.

- **Parallel Arterial (Capitol Ave.):** Similar to the results in table 4.4, both directions of the parallel arterial experienced nearly no change in delay. The penalty imposed on on-ramp bound traffic as a result of the absence of queue override was relatively small and offset by the delay savings experienced by those not intended to access the on-ramp.

- **Cross Streets:** Similar to the result in table 4.1, there were almost no change in delay at Alum Rock Ave due to the fact that sufficient on-ramp queue storage space near the Alum Rock Ave left the signal timing unchanged at Alum Rock Ave. The limited queue storage space at the McKee Rd. on-ramp helped reduce the signal cycle lengths and therefore the average delay in both directions of McKee Rd. The variation between the westbound and eastbound directions of the cross streets can be explained by the penalty experienced by the on-ramp bound trips in the westbound direction due to the absence of queue override. Lastly, the lower demand at Berryessa Rd. was not influenced by the signal timing changes.

- **Total System:** The proposed control strategy reduced the system-wide delay by more than 10%.

Table 4.5 Performance before and after proposed signal control (-10% demand).

	Before Coordination		**After Coordination**		**% Difference**	
Freeway Mainline						
	Total Delay (veh-hr)	Total Distance Traveled (veh-mile)	Total Delay (veh-hr)	Vehicle Miles of Travel (VMT)	Change in Total Delay	Change in VMT
I-680 NB	248.55	39570.84	239.56	40512.63	**-3.62%**	**2.38%**
Arterial						
Average Delay on Main Parallel Arterial (min/veh)						
Capitol Ave NB	6.42		6.18		-3.72%	
Capitol Ave SB	5.02		4.96		**-1.14%**	
Average Delay of Cross Street (sec/veh)						
Alum Rock WB	24.81		23.29		-6.14%	
Alum Rock EB	28.84		28.62		-0.76%	
McKee WB	51.93		41.18		**-20.69%**	
McKee EB	26.41		13.12		**-50.33%**	
Berryessa WB	43.71		43.67		-0.11%	
Berryessa EB	45.92		45.50		-0.91%	
Total System						
	Total Delay (veh-hr)		Total Delay (veh-hr)		Change in Total Delay	
Freeway & Arterial	1617.27		1462.72		**-10.57%**	

Bold (% Difference): Statistical significance

Lastly, figures 4.1 and 4.2 summarize how freeway and arterial performance change as demand varies. Figure 4.1 shows that the proposed control strategy further reduces freeway delay and improves freeway capacity as the demand become higher. Figure 4.2 shows that the proposed control strategy further increases arterial delay as demand grows. The penalty imposed on arterial traffic can be severe at high traffic demand.

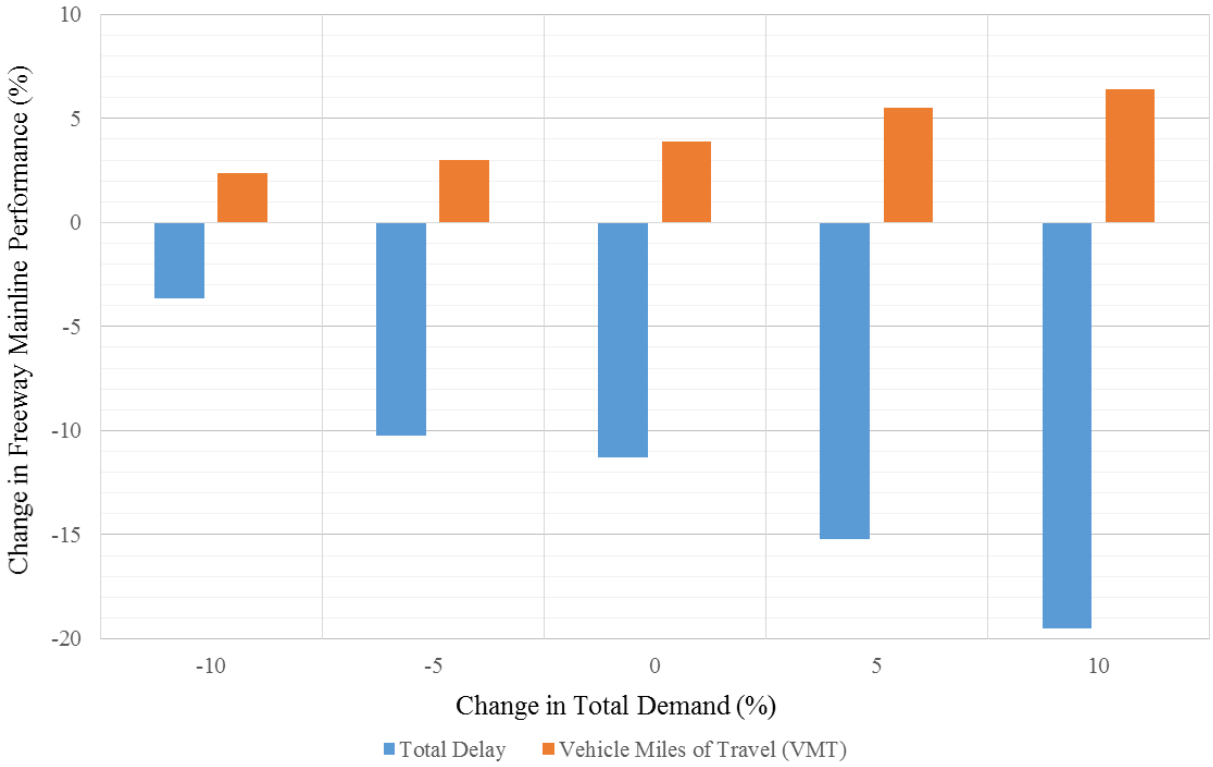


Figure 4.1 Freeway performance for varying levels of demand.

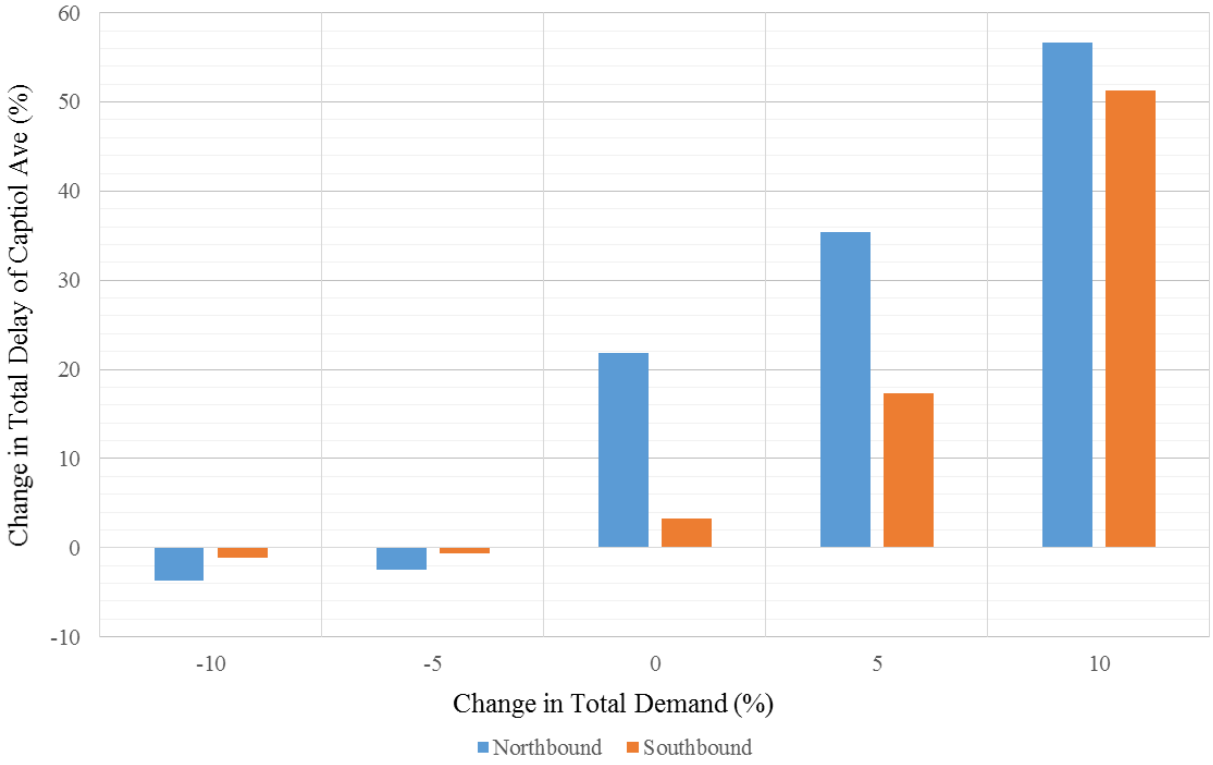


Figure 4.2 Total delay of Capitol Ave. for varying levels of demand.

CHAPTER 5: CONCLUSIONS

5.1 Summary of the Findings

On-ramp metering has become an effective and widely practiced method of mitigating freeway congestion and maintaining high freeway capacity. However, arterial traffic signal near freeway on-ramps disregard the on-ramp queues and employ long cycle lengths to maximize capacity, which sends long platoons of on-ramp bound vehicles toward the oversaturated on-ramps and cause queue spillbacks onto the adjacent arterial intersections. Queue override has been a popular method to alleviate the spillback by relaxing the metering rate. However, queue override diminishes the freeway bottleneck discharge rate. This was demonstrated by a two-week long empirical study conducted at the McKee Rd. on-ramp bottleneck of northbound Interstate 680 in San Jose, California. The empirical study suggests that queue override reduces the freeway capacity by approximately 10% percent.

However, nearby arterial traffic signals can help avoid queue spillback at freeway on-ramps if the signal control systems were integrated with those of the freeway ramp metering and if the traffic signals were timed effectively based on the current ramp metering rate and on-ramp queue lengths. This study developed a signal control algorithm that reduces the signal cycle length and maintains the same green time distribution of the arterial intersections near freeway on-ramps. This can prevent on-ramp queue spillback and subsequently avoid queue override by sending on-ramp bound traffic in smaller but more frequent platoons, without imposing significant penalties.

A section of the I-680 Northbound freeway with Capitol Ave. as the parallel arterial in the city of San Jose, California was selected as the test site to evaluate the proposed signal control. The AM peak was selected as the analysis period. Traffic operations at the site were simulated using the AIMSUN microscopic model after extensive data collection and model calibration. Additional enhancements were made to the driving behavior model of the simulation in order to better replicate real world conditions and especially the phenomenon of capacity drop.

The simulation results show that the proposed signal control strategy eliminated queue spillback on the metered on-ramps and prevented queue override. This resulted in 11.13% reduction in freeway mainline delay. Furthermore, the freeway capacity improved, as indicated by a 3.92% increase in vehicle miles of travel (VMT) throughout the entire 2.5 hour morning peak. The delay on the parallel arterial was increased due to the penalties imposed on the approaches for on-ramp bound trips as a result of absence of queue override. Nevertheless, the rest of the signalized approaches experienced significant delay savings.

Moreover, sensitivity analyses suggest that less than 10% increase or decrease in freeway and arterial demand would not significantly impact the performance of the freeway and nearby arterial, as improvements can still be observed despite minor demand fluctuations.

Lastly, the proposed algorithm is simple and readily implementable at most freeway corridors with metered on-ramps and adjacent arterials primarily used to facilitate freeway access. The proposed approach is not limited to any specific freeway ramp metering algorithm, and requires only software integrations between the on-ramp metering and arterial traffic signals. The implementation does not require any investment in new sensing or signal control technologies.

5.2 Future Work

Real world field tests of the proposed signal control strategy is recommended in order to better estimate the potential benefits. Field implementation of the proposed control strategy is relatively simply under existing technologies and infrastructures. Software integration of arterial traffic signals and freeway ramp metering will be conducted and fine-tuned for real world implementation. The field test will be performed from early March 2018 to late May 2018. Real world implementation will help us identify the practical constraints and issues that were not shown in simulation. The real world performance of the proposed signal control strategy will serve as a guideline for transportation agencies when designing arterial traffic signal timing plans near freeways.

Simulation studies and subsequent field tests are recommended for freeway corridors with different types of freeway-arterial interchanges such as diamond interchange and partial cloverleaf interchange, in addition to the cloverleaf interchange presented in this study. Arterial traffic signal timings plans typically vary by the type of interchange, and simulation and field tests will allow further evaluation of the effectiveness and adaptability of the proposed signal control strategy, as well as challenges and limitations that are unique to each type of interchange geometry. This can be valuable information for transportation agencies that would like to reconfigure their signal timing plans for the intersections near freeway access.

Furthermore, simulation and field studies are suggested for freeway corridors that employ different ramp metering algorithms. Some freeway ramp metering algorithms are effective at mitigating queue spillback and thus may not require more efficient signal timing for the nearby arterial intersections, while other ramp metering algorithms may cause significant on-ramp queue spillback. Therefore such simulation and field study will help the planning and operating agencies determine the suitability of the proposed signal control approach, should they consider updating their signal timing plans for the arterial intersections nearby freeway on-ramps.

Lastly, further tests and performance evaluations are recommended for large scale freeway corridors or networks prior to introducing the proposed signal timing plan at the regional level. The adaptability and effectiveness of the proposed signal control may vary across a metropolitan area due to distinct demand patterns, road geometries, etc.

REFERENCES

Al-Qbaedi, J., Yousif, S., 2012. Microsimulation model for motorway merges with ramp metering control. IEEE Transaction on Intelligent Transportation Systems 13 (1), 296–306.

Banks, J.H., 1991. Two-capacity phenomenon at freeway bottlenecks: a basis for ramp metering. Transportation Research Record 1320, 83–90.

Caltrans PeMS. http://pems.dot.ca.gov/. Accessed on March 31, 2016.

Cassidy, M.J., Bertini, R.L., 1999. Observations at a freeway bottleneck. Proceedings of the 14th International Symposium on Transportation and Traffic Theory, 107–146.

Cassidy, M.J., Rudjanakanoknad, J., 2005. Increasing the capacity of an isolated merge by metering its on-ramp. Transportation Research Part B 39 (10), 896–913.

Chilukuri, B.R., Laval, J.A., Chen, D., 2013. Some traffic features during on-ramp queue flush. Transportation Research Board 92nd Annual Meeting.

Daganzo, C.F., 1997. Fundamentals of Transportation and Traffic Operations. Elsevier, New York.

Gettman, D., G. Madrigal, S. Allen, T. Boyer, S. Walker, J. Tong, S. Phillips, H. Liu, X. Wu, H. Hu, M. Abbas, Z. Adam, and A. Skabardonis., 2012. Operation of Traffic Signal Systems in Oversaturated Conditions. National Cooperative Highway Research Program.

Gunther, G., Coeymans, J., Munoz, J., Herrera, J., 2012. Mitigating Freeway Off-ramp Congestion: A Surface Street Coordinated Approach. Transportation Research Part C: Emerging Technologies 20 (1), 112-124.

Hall, F.L., Agyemang-Duah, K., 1991. Freeway capacity drop and the definition of capacity. Transportation Research Record 1320, 91–98.

HCM 2010: Highway Capacity Manual. 2010. Transportation Research Board of the National Academies, Washington, D. C.

Horowitz, R., A. May, A. Skabardonis, and P. Varaiya, 2005. Design, Field Implementation and Evaluation of adaptive Ramp Metering Algorithms. PATH Research Report UCB-ITS-PRR-2005-2, Institute of Transportation Studies, University of California, Berkeley.

Jin, W., Gan, Q., Lebacque, J., 2015. A kinematic wave theory of capacity drop. Transportation Research Part B 41 (1), 82–95.

Kim, K., Cassidy, M.J., 2012. A capacity-increasing mechanism in freeway traffic. Transportation Research Part B 46 (2012), 1260–1272.

Leclercq, L., Laval, J.A., Chiabaut, N., 2011. Capacity drop at merges: an endogenous model. Transportation Research Part B 45 (2011), 1302–1311.

Li, Z., Tao, R., 2011. Integrate Control of Freeway Interchange Model based on Enhanced Cell Transmission Model. Transportation Research Record 2259, 179-191.

Lieberman, E.B., Chang, J., and Prassas, E.S., 2000. Formulation of Real-Time Control Policy for Oversaturated Arterials. Transportation Research Record 1727, 77-78.

Lighthill, M. and G. Whitham, 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 229 (1178), 317-345.

Liu, X., Zhang, G., Kwan, C., Wang, Y., Kemper, B., 2013. Simulation-based, Scenario-driven Integrated Corridor Management Strategy Analysis. Transportation Research Record 2396, 38-44.

Lu, X., Kan, X., Shladover, S.E., Wei, D., Ferlis, R.A., 2017. An enhanced microscopic traffic simulation model for application to connected automated vehicles. Transportation Research Board 96th Annual Meeting, Washington, DC.

Minnesota Department of Transportation (2001), Twin cities ramp meter evaluation, Cambridge Systematics, Inc., Cambridge, MA, 2001. http://www.dot.state.mn.us/rampmeterstudy.

Munoz, J.C., Daganzo, C.F., 2002. Fingerprinting traffic from static freeway sensors. Cooperative Transportation Dynamics 1, 1-11.

Newell, G., 1989. Theory of Highway Traffic Signals. Institute of Transportation Studies Research Report, UCB-ITS-RR-89-7.

Oh, S., Yeo, H., 2012. Microscopic analysis on the causal factors of capacity drop in highway merging sections. Transportation Research Board 91st Annual Meeting, Washington, DC.

Papageorgiou, M., Kotsialos, A., 2002. Freeway ramp metering: an overview. IEEE Transactions on Intelligent Transportation Systems 3 (4).

Parzani, C., Buisson, C., 2012. Second order model and capacity drop at merge. Transportation Research Record 2315, 25–34.

Persaud, B.N., Yagar, S., Brownlee, R., 1998. Exploration of the breakdown phenomenon in freeway traffic. Transportation Research Record 1634, 64–69.

Recker, W., X. Zheng, and L. Chu., 2009. Development of an Adaptive Corridor Traffic Control Model. California PATH Research Report.

Richards, P., 1956. Shock waves on the highway. Operations Research 4 (1), 42-51.

Sahin, I., Altun, I., 2008. Empirical study of behavioral theory of traffic flow. Transportation Research Record 2088, 109–116.

Smaragdis, E., Papageorgiou, M., Kosmatopoulos, E., 2003. A flow-maximizing adaptive local ramp metering strategy. Transportation Research Part B 38, 251–270.

Srivastava A., Geroliminis, N., 2013. Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model. Transportation Research Part C 30 (2013), 166-177.

Su, D., XY. Lu, R. Horowitz, Z. Wang., 2014. Coordinated Ramp Metering and Intersection Signal Control. International Journal of Transportation Science and Technology 3 (2), 179-191.

Tian, Z., Balke, K., Engelbrecht, R., Rilett, L., 2002. Integrated Control Strategies for Surface Street and Freeway Systems. Transportation Research Record 1811, 92-99.

Tian, Z., Messer, C., Balke, K., Urbanik, T., 2005. Integration of Diamond Interchange and Ramp Metering Operations. Transportation Research Record 1925, 100-111.

TSS|Aimsun. http://www.aimsun.com/. Accessed on March 19, 2017.

Urbanik, T., D. Humphreys, B. Smith, and S. Levine., 2006. Coordinated Freeway and Arterial Operations Handbook. Federal Highway Administration FHWA-HRT-06-095.

Yang, X., Cheng, Y., Chang, G., 2014. Dynamic Signal Control Strategy to Mitigate Off-ramp Queue Spillback to Freeway Mainline Segment. Transportation Research Record 2438, 1-11.

Yang, X., Cheng, Y., Chang, G., 2015. Integrating off-ramp Spillback Control with a Decomposed Arterial Signal Optimization Model. Transportation Research Record 2487, 112-121.

Yuan, K., Knopp, V.L., Hoogendoorn, S.P. 2015. Capacity drop: relationship between speed in congestion and the queue discharge rate. Transportation Research Record 2491, 72–80.

Yuan, K., Knopp, V.L., Schreiter, T., Hoogendoorn, S.P., 2015. A hybrid kinematic wave model incorporating capacity drops. IEEE 18th International Conference on Intelligent Transportation Systems, 72–80.

Zhang, L., Gou, J., Jin, M., 2012. Model of Integrated Corridor Traffic Optimization. Transportation Research Record 2311, 179-191.

Zhang, L., Levinson, D., 2010. Ramp metering and freeway bottleneck capacity. Transportation Research Part A 44 (2010), 218–235.

Zhang, H., Ma, J., Nie, Y., 2009. Local Synchronization Control Scheme for Congested Interchange Areas in Freeway Corridor. Transportation Research Record 2128, 173-183.

APPENDIX A: TEST SITE SELECTION

This chapter describes the process of test site selection and the characteristics of the site selected for testing the proposed signal control. The selected site should be representative of typical freeway corridor segments (3 to 6 mile long) with at least one adjacent arterial facilitating freeway access. The selected corridor should satisfy the following criteria:

A.1 Site Selection Criteria

(1) The freeway corridor should have 3 to 5 freeway-arterial interchanges;

(2) The freeway corridor must not contain any freeway-to-freeway interchanges;

(3) At least one recurrent bottleneck must be observed during either the morning or the evening peak hours, preferable in only one direction of the freeway;

(4) The recurrent bottleneck(s) must be caused by the high on-ramp demand;

(5) Under recurrent conditions, the bottlenecks observed along the freeway corridor must be isolated (free-flow conditions at the upstream and downstream ends of the corridor);

(6) The physical capacity of a section is fixed except for lane reduction caused by lane closure due to incident/accident;

(7) The freeway corridor must have low frequency of incidents that contribute to non-recurrent delay;

(8) The length of the freeway on-ramps should not be too short or too long (ideally, they should accommodate 30 to 50 queued vehicles);

(9) The corridor must contain at least one parallel arterial adjacent to the freeway;

(10) The parallel arterial(s) must connect the arterials that have interchanges with and are perpendicular to the freeway;

(11) The parallel and perpendicular arterials adjacent to the freeway should be primarily used to facilitate freeway access;

(12) High demand from arterial to freeway should be the main cause of arterial congestion;

(13) No more than 5 major signalized intersections along the parallel arterial;

(14) There should not be high concentration of pedestrians crossing the arterial or bicyclists impeding the arterial traffic flow;

(15) No active work zones on the freeway and the arterial;

(16) Satisfactory detector health and properly functioning ramp meters and traffic signals;

(17) Cooperation between the jurisdictions responsible for the operation and maintenance of the freeway ramp metering and arterial traffic signals control systems;

(18) The selected site must be supported by centralized data acquisition and control system in order to integrate freeway ramp-metering and arterial traffic signals for field implementation;

(19) The selected site must be representative of freeway corridors with adjacent arterial(s) facilitating freeway access.

A.2 Candidate Sites

Six candidate test sites were identified based on extensive data analysis and suggestions from Distract 4 of California Department of Transportation (Caltrans). The maps of the candidate sites are shown in Figure A.1 through A.6. Comments regarding whether the site selection criteria were satisfied can be found below the respective figures. Unsatisfied criteria are highlighted in red.

A.2.1 Candidate Site #1: I-80 Northbound PM Peak

Segment of interest: Central Ave. to Pinole Valley Rd. (9 miles)

Parallel arterial: San Pablo Ave. (15 major signalized intersections)

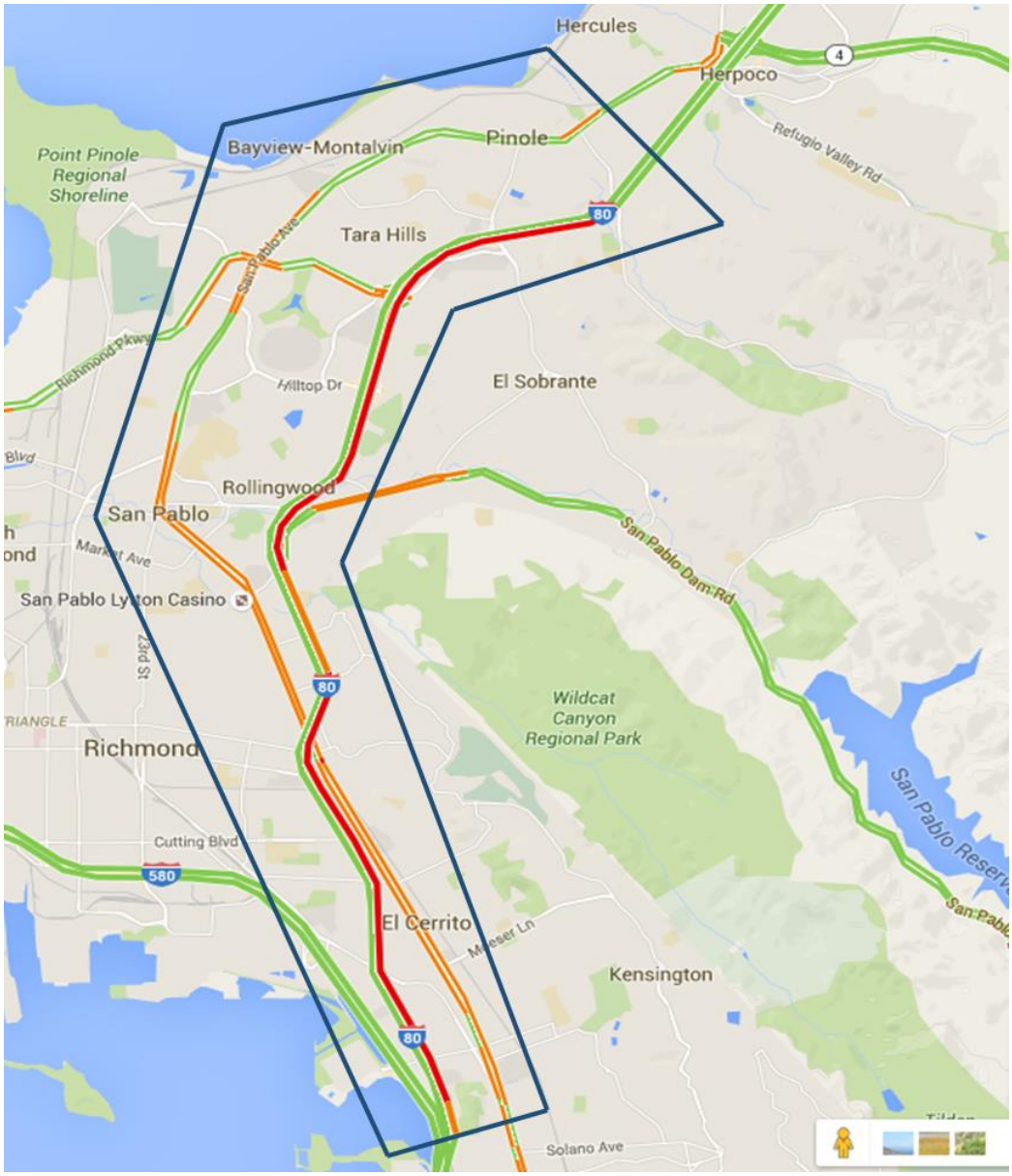


Figure A.1 Map of I-80 Northbound PM peak and San Pablo Ave.

(1) There are 10 freeway-arterial interchanges.
(2) No freeway-to-freeway interchange.
(3) Multiple recurrent bottlenecks.
(4) Recurrent congestion is mostly caused by high on-ramp demand but also high off-ramp demand at a few locations (i.e. San Pablo Dam Rd.).
(5) Free-flow conditions can be observed upstream of Central Ave. (up to the I-580 split) and downstream of Pinole Valley Rd.
(6) The freeway capacity is fixed.
(7) There is relatively high frequency of incidents/accidents that contribute to significant non-recurrent delay.
(8) The on-ramp lengths fit the requirement with the exception of San Pablo Dam Rd. and Solano Ave. on-ramps.
(9) San Pablo Ave. is a major arterial near the freeway.
(10) San Pablo Ave. connects the perpendicular arterials with freeway access but the road geometry is less than ideal because San Pablo Ave is too far from I-80 downstream of San Pablo Dam Rd.
(11) San Pablo Ave. is not primarily used to facilitate freeway access.
(12) High demand for freeway access is not the main cause of oversaturation on San Pablo Ave.
(13) There are roughly 15 major signalized intersections along San Pablo Ave.
(14) There is limited pedestrian and bicyclist traffic due to the suburban setting.
(15) There were no active work zones at the time of site selection but the schedule of future construction activities on the corridor conflict with the timelines for field data collection and implementation.
(16) Very good detector health (almost 100% observed), and the traffic signals are properly functioning.
(17) Cooperation among different jurisdictions is uncertain. Multiple agencies (i.e. City of El Cerrito, City of Richmond, City of San Pablo, etc.) must be involved, in addition to Caltrans.
(18) Multiple agencies operate the arterial traffic signal, thus it is very difficult to have the site supported by centralized data system and control system.
(19) The unique road geometry makes the site less representative of typical freeway-arterial corridors.

A.2.2 Candidate Site #2: I-680 Northbound AM Peak

Segment of interest: Capitol Expy. to Berryessa Rd. (4 miles)

Parallel arterial: Capitol Ave. (5 major signalized intersections)

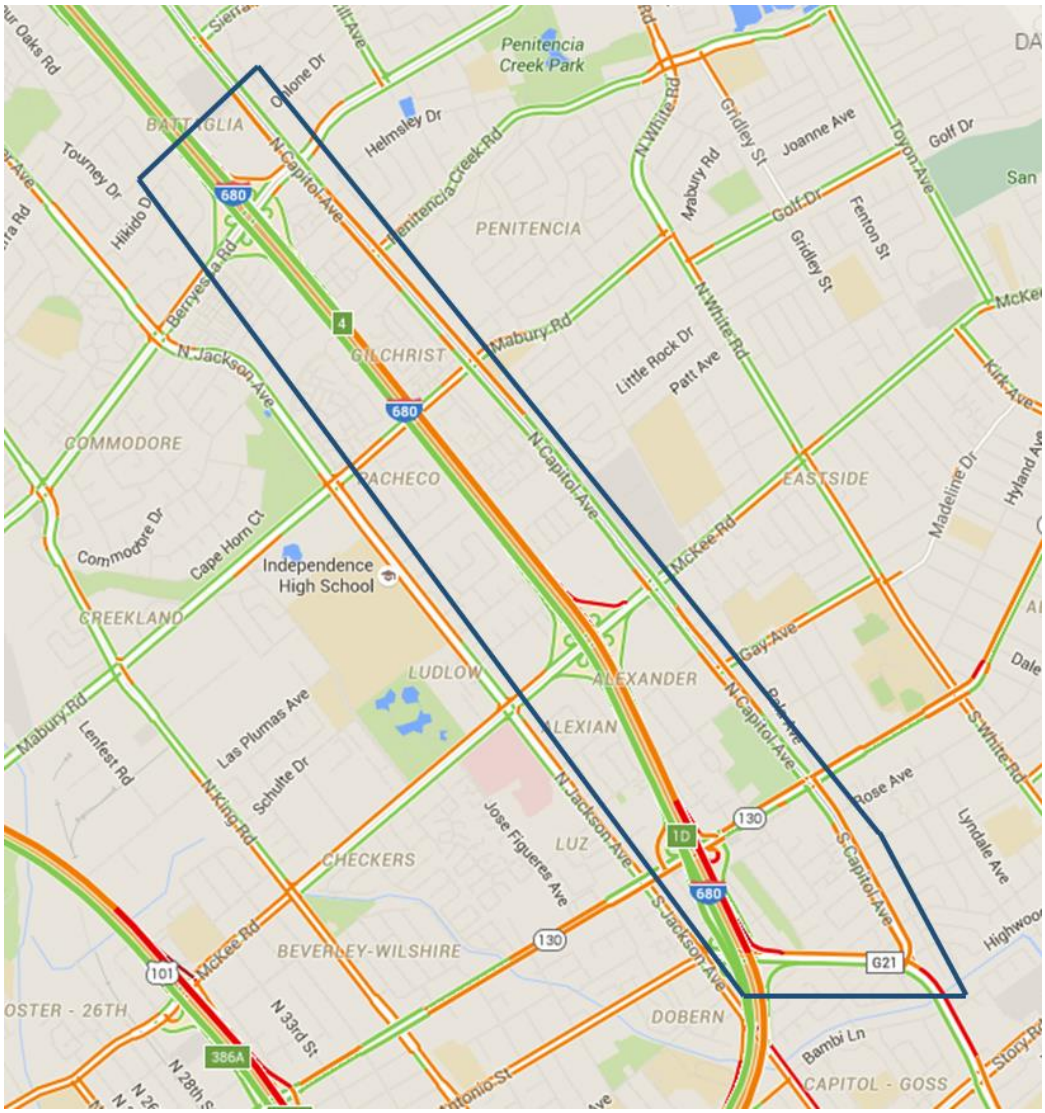


Figure A.2 Map of I-680 Northbound AM peak and Capitol Ave.

(1) There are 4 freeway-arterial interchanges.
(2) No freeway-to-freeway interchange.
(3) Multiple recurrent bottlenecks.
(4) Recurrent congestion is primarily caused by high on-ramp demand.
(5) Free-flow conditions can be observed upstream of Capitol Expy., and downstream of Berryessa Rd.
(6) The freeway capacity is fixed.
(7) Very low frequency of incidents and non-recurrent delay.
(8) The on-ramp lengths fit the requirement. There is sufficient storage on all on-ramps.
(9) Capitol Ave. is a major arterial immediately adjacent to the freeway.
(10) Capitol Ave. connects the perpendicular arterials with freeway access and the road geometry resemble that of a grid network.
(11) Capitol Ave. is primarily used to facilitate freeway access during the morning peak.
(12) High demand for freeway access is the main cause of oversaturation on the arterials.
(13) There are 5 major signalized intersections along Capitol Ave.
(14) There is limited pedestrian and bicyclist traffic due to the suburban setting.
(15) There were no active work zones at the time of site selection. Construction activities for the bus rapid transit project on Alum Rock Ave. (perpendicular arterial) were scheduled to take place after field data collection and will complete before field implementation.
(16) Good detector health due to newly installed detectors, and the traffic signals are properly functioning.
(17) Agreement with the City of San Jose has been reached. Santa Clara County has to be involved but it operates only one signalized intersection (Capitol Ave. at Capitol Expy.).
(18) Under the current agreement between City of San Jose and Caltrans, the site is supported by centralized data system and control system. However, cooperation with Santa Clara County is uncertain.
(19) The road geometry resembles the typical freeway-arterial corridors.

A.2.3 Candidate Site #3: I-680 Northbound AM Peak

Segment of interest: E. San Antonio St./Capitol Expy. to Berryessa Rd. (4 miles)

Parallel arterial: Jackson Ave. (5 major signalized intersections)

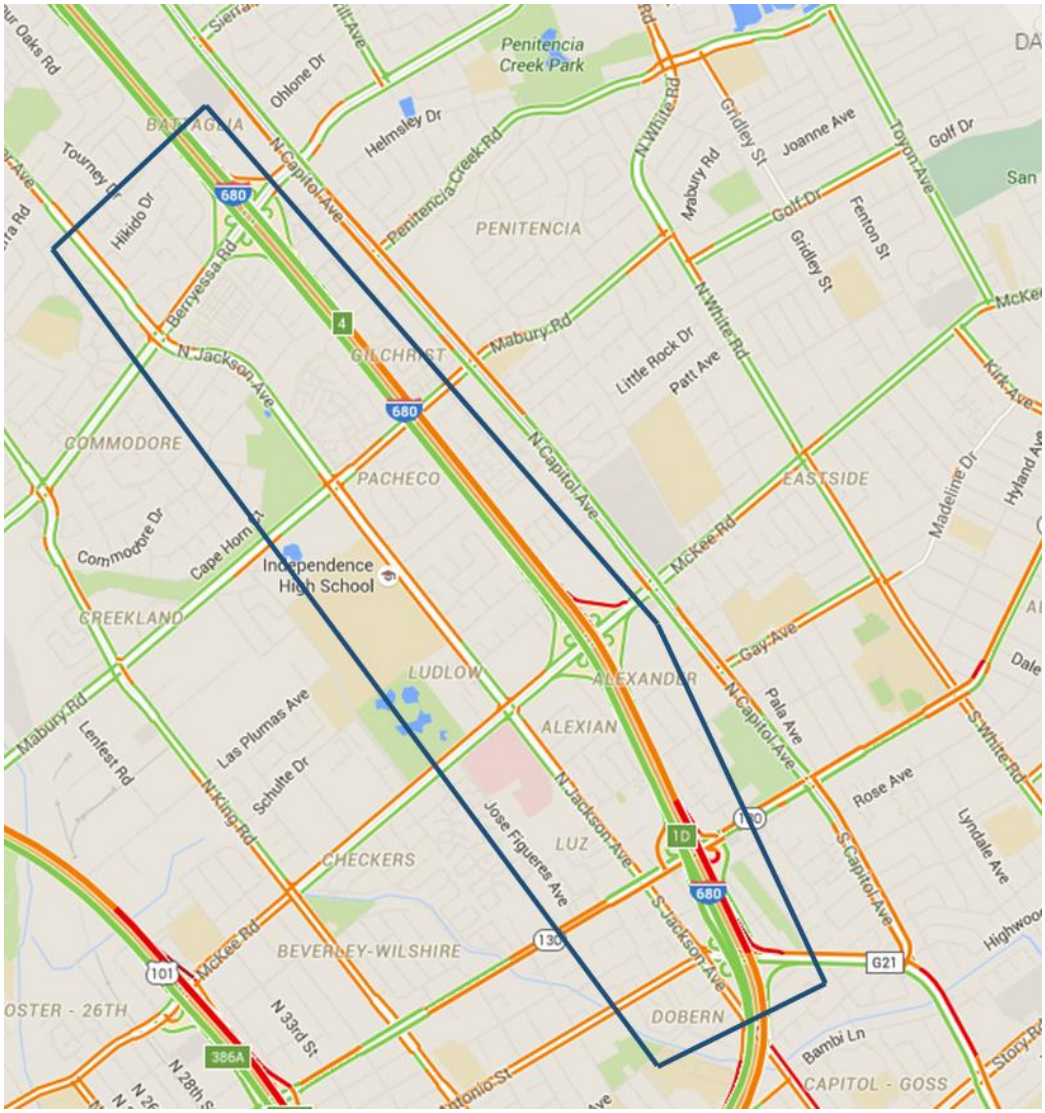


Figure A.3 Map of I-680 Northbound AM peak and Jackson Ave.

(1) There are 4 freeway-arterial interchanges.
(2) No freeway-to-freeway interchange.
(3) Multiple recurrent bottlenecks.
(4) Recurrent congestion is primarily caused by high on-ramp demand.
(5) Free-flow conditions can be observed upstream of Capitol Expy., and downstream of Berryessa Rd.
(6) The freeway capacity is fixed.
(7) Very low frequency of incidents and non-recurrent delay.
(8) The on-ramp lengths fit the requirement. There is sufficient storage on all on-ramps.
(9) Jackson Ave. is a major arterial immediately adjacent to the freeway.
(10) Jackson Ave. connects the perpendicular arterials with freeway access and the road geometry resembles that of a grid network.
(11) Jackson Ave. is used to facilitate freeway access during the morning peak but this is not the arterial's primary function.
(12) The demand for freeway access is not very high and the parallel arterial is not very saturated.
(13) There are 5 major signalized intersections along Jackson Ave.
(14) There is a fair amount of pedestrian traffic due to the presence of schools and hospitals nearby.
(15) There were no active work zones at the time of site selection. Construction activities for the bus rapid transit project on Alum Rock Ave. (perpendicular arterial) were scheduled to take place after field data collection and will complete before field implementation.
(16) Good detector health due to newly installed detectors, and the traffic signals are properly functioning.
(17) Agreement with the City of San Jose have been reached. All of the signalized intersections are operated by the City of San Jose.
(18) Under the current agreement between City of San Jose and Caltrans, the site is supported by centralized data system and control system.
(19) The road geometry resembles the typical freeway-arterial corridors.

A.2.4 Candidate Site #4: SR-87 Northbound AM Peak

Segment of interest: Branham Ln. to W. Alma Ave. (4 miles)

Parallel arterial: Almaden Expy. (5 major signalized intersections)

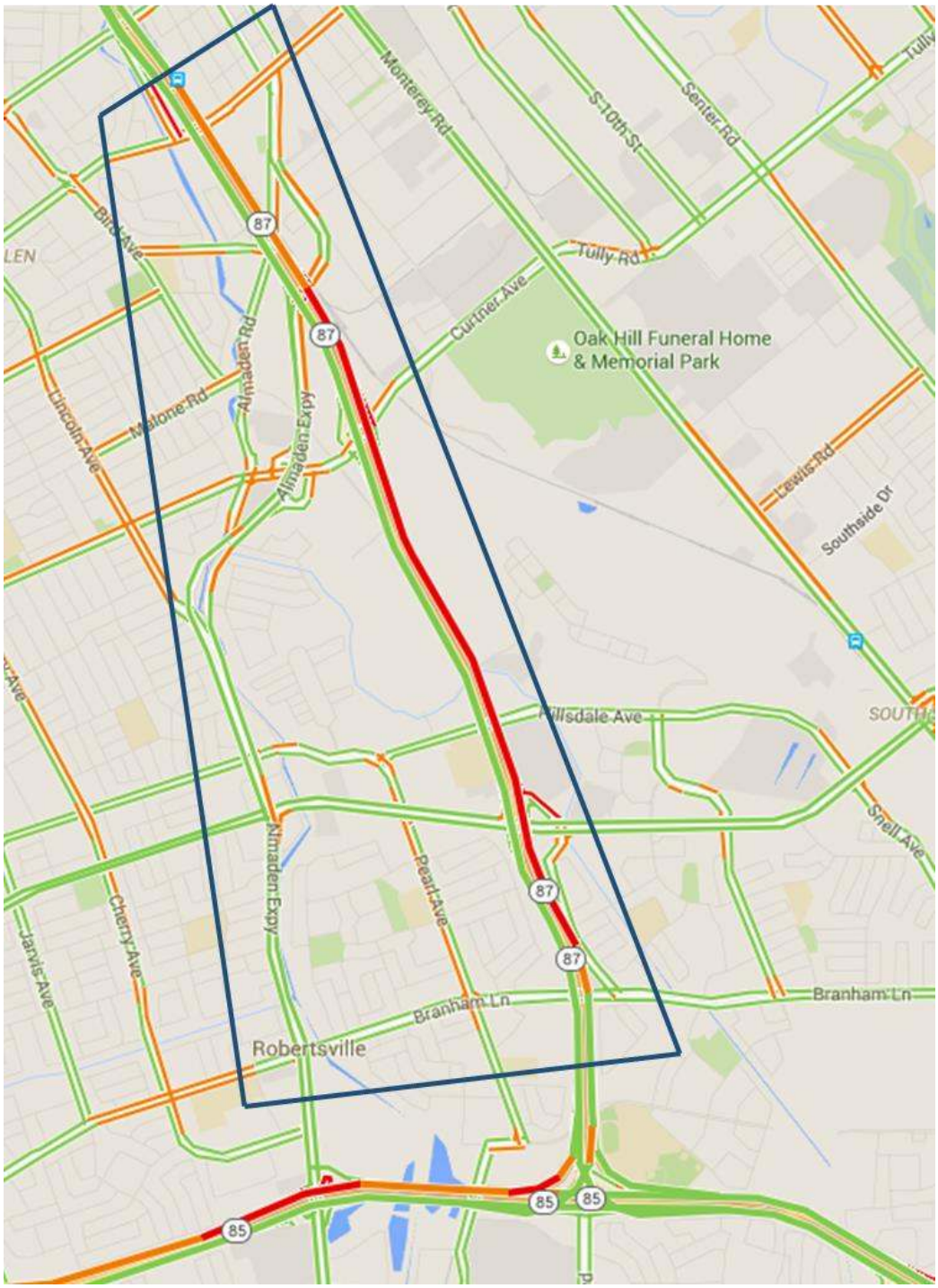


Figure A.4 Map of SR-87 Northbound AM peak and Almaden Expy.

(1) There are 4 freeway-arterial interchanges.
(2) No freeway-to-freeway interchange.
(3) Multiple recurrent bottlenecks.
(4) Recurrent congestion is primarily caused by high on-ramp demand.
(5) Free-flow conditions can be observed upstream of Capitol Expy., and downstream of Alma Ave.
(6) The freeway capacity is fixed.
(7) Very low frequency of incidents and non-recurrent delay.
(8) The on-ramp lengths fit the requirement. There is sufficient storage on all on-ramps.
(9) Almaden Expy. is the major arterial near the freeway but not very close to the freeway.
(10) Almaden Expy. connects the perpendicular arterials with freeway access but the road geometry does not resemble a grid network. In addition, Almaden Expy. has grade separated interchanges rather than signalized intersections with a few of its perpendicular arterials.
(11) Almaden Expy. is used to facilitate freeway access during the morning peak but this is not the arterial's primary function. Almaden Expy. is often used as an alternate route to SR-87.
(12) The demand for freeway access is very high and it is the primary cause of arterial congestion.
(13) There are 5 major signalized intersections along Almaden Expy. but there are also two grade-separated interchanges with the perpendicular arterials.
(14) Very low pedestrian traffic due to the suburban setting.
(15) There were no active work zones at the time of site selection. No construction planned for the duration of this project.
(16) There is extensive coverage of detectors but not all of them are properly functioning. Those at the most important locations are in good condition. The traffic signals are working properly.
(17) All of the signals on Almaden Expy. are operated by Santa Clara County. Cooperation from this agency is uncertain.
(18) Under the current agreement between City of San Jose and Caltrans, the site is supported by centralized data system and control system. However, cooperation from Santa Clara County is also required but is uncertain and challenging to obtain.
(19) The road geometry does not resemble the typical freeway-arterial corridors.

A.2.5 Candidate Site #5: US-101 Northbound AM Peak & PM Peak

Segment of interest: Wilfred Ave. to Baker Ave. (4 miles)

Parallel arterial: Santa Rosa Ave. (5 major signalized intersections)

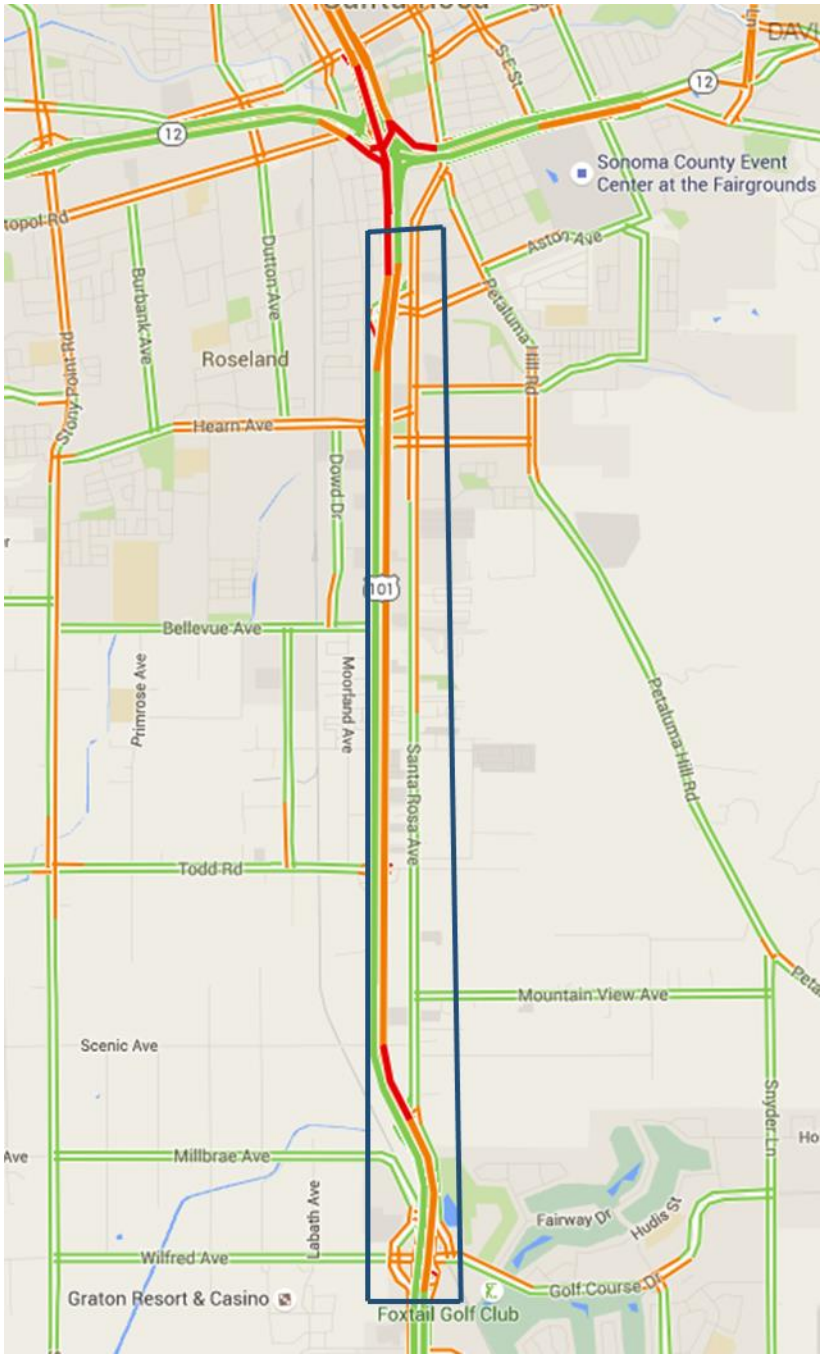


Figure A.5 Map of US-101 Northbound AM/PM peak and Santa Rosa Ave.

(1) There are 5 freeway-arterial interchanges.
(2) No freeway-to-freeway interchange.
(3) Multiple recurrent bottlenecks.
(4) Recurrent congestion is primarily caused by high on-ramp demand.
(5) Free-flow conditions can be observed upstream of Wilfred Ave., and downstream of Baker Ave.
(6) The freeway capacity is fixed.
(7) Relatively low frequency of incidents and non-recurrent delay.
(8) The on-ramp lengths fit the requirement. There is sufficient storage on all on-ramps.
(9) Santa Rosa Ave. is the major arterial immediately adjacent to the freeway.
(10) Santa Rosa Ave. connects the perpendicular arterials with freeway access and the road geometry somewhat resembles a grid network.
(11) Santa Rosa Ave. is used to facilitate freeway access during the peak hours but it is also used as an alternate route to US 101 for shorter trips, mostly due to the rural setting (in areas further away from Santa Rosa) and relatively high speed limit.
(12) The demand for freeway access is very high and it is the primary cause of arterial congestion.
(13) There are 5 major signalized intersections along Santa Rosa Ave.
(14) Very low pedestrian traffic due to the suburban setting.
(15) There were no active work zones at the time of site selection. No construction planned for the duration of this project.
(16) There is extensive coverage of detectors and video cameras on the freeway. The traffic signals are working properly.
(17) Some signals are operated by the City of Santa Rosa while others are operated by the City of Rohnert Park. Cooperation from these agencies is uncertain.
(18) Due to the uncertainty of whether the City of Santa Rosa and the City of Rohnert Park are willing to participate, whether the site will be supported by centralized data system and control system is unknown.
(19) The road geometry somewhat resembles the typical freeway-arterial corridors. However, the perpendicular arterials do not connect opposite sides of the freeway.

A.2.6 Candidate Site #6: SR-4 Westbound AM Peak

Segment of interest: Railroad Ave. to Willow Pass Rd. (6 miles)

Parallel arterial: Leland Rd. (5 major signalized intersections)

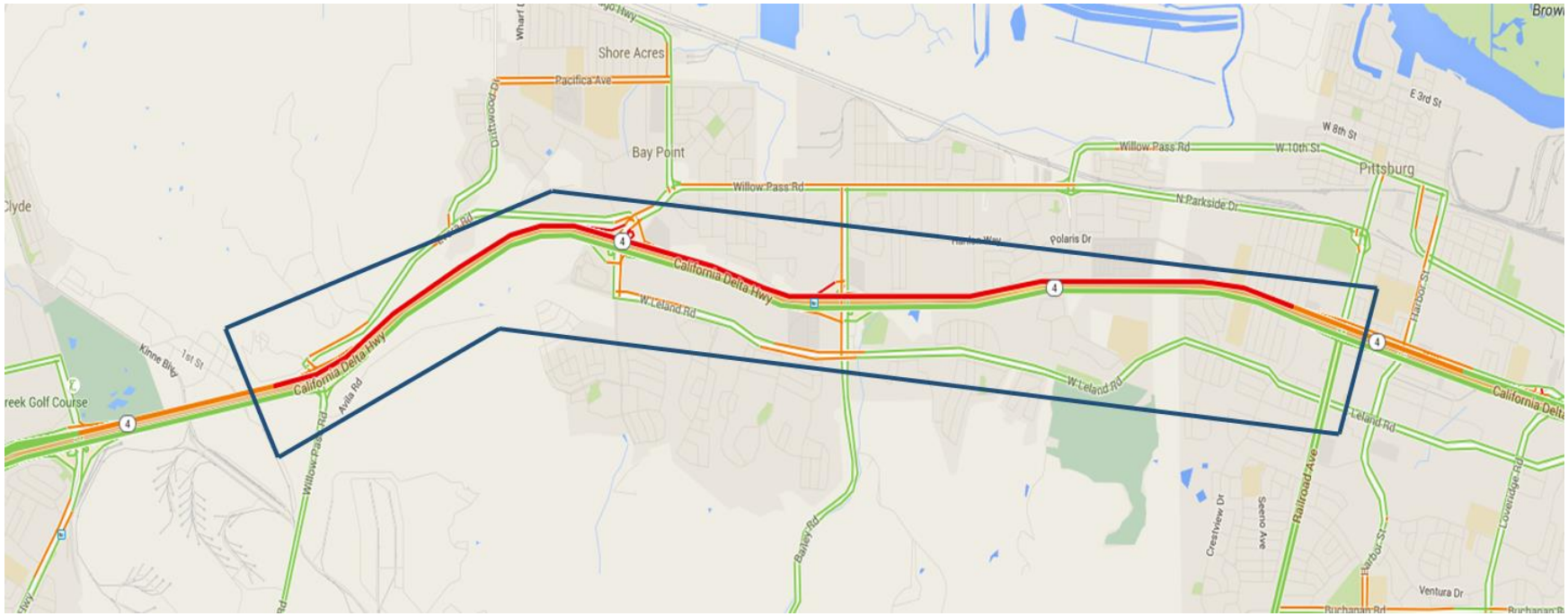


Figure A.6 Map of SR-4 Westbound AM peak and Leland Rd.

(1) There are 5 freeway-arterial interchanges.
(2) No freeway-to-freeway interchange.
(3) Multiple recurrent bottlenecks.
(4) Recurrent congestion is primarily caused by high on-ramp demand.
(5) Free-flow conditions can be observed upstream of Railroad Ave., and downstream of Willow Pass Rd.
(6) The freeway capacity is fixed.
(7) Relatively low frequency of incidents and non-recurrent delay.
(8) The on-ramp lengths fit the requirement. There is sufficient storage on all on-ramps.
(9) Leland Rd. is the major arterial immediately adjacent to the freeway.
(10) Leland Rd. connects the perpendicular arterials with freeway access and the road geometry somewhat resembles a grid network.
(11) Leland Rd. is used to facilitate freeway access during the morning peak hours.
(12) The demand for freeway access is very high and it is the primary cause of arterial congestion.
(13) There are 5 major signalized intersections along Leland Rd.
(14) Very low pedestrian traffic due to the suburban setting.
(15) There were no active work zones at the time of site selection. No construction planned for the duration of this project.
(16) The detector health is satisfactory. The traffic signals are working properly.
(17) Some signals are operated by the City of Bay Point while others are operated by the City of Pittsburg. Cooperation from these agencies is uncertain.
(18) Due to the uncertainty of whether the City of Bay Point and the City of Pittsburg are willing to participate, whether the site will be supported by centralized data system and control system is unknown.
(19) The road geometry somewhat resembles the typical freeway-arterial corridors.

A.3 Selected Site

After careful consideration and discussion with the project panel, the 6 candidate sites were reduced to two candidate sites: I-680 Northbound AM Peak with Capitol Ave. as the parallel arterial and I-680 Northbound AM Peak with Jackson Ave. as the parallel arterial. The other four candidate sites were not selected by the project team due to lack of cooperation from the city-level jurisdictions in charge of the arterial traffic signal operations. Finally, the project team decided to select I-680 Northbound AM Peak with Capitol Ave. as the parallel arterial. Although the two final candidate sites share the same segment of freeway and both arterials have desirable road geometries, the site with Capitol Ave. as the parallel arterial is more desirable due to the following reasons:

- Majority of the traffic heading onto the congested northbound direction of the freeway come from the east side of the freeway, which is the area surrounding Capitol Ave.
- City of San Jose strongly objected the inclusion of Jackson Ave as part of the corridor because various changes such as pedestrian signals and road diet have been planned in order to slow down traffic along the arterial and enhance pedestrian safety (Jackson Ave has relatively high volumes of pedestrian crossing and rates of pedestrian fatality).
    - Future field tests may jeopardize pedestrian safety.
    - Improvements are unlikely due to pedestrian crossings rather than the interaction between freeway and arterial traffic becoming the major source of congestion on the arterial.
- Although there is concern about the transit signal priority (TSP) granted to the light rail vehicles that operate in the median of Capitol Ave, the relatively low frequency (every 15 minutes) and the field observations help us confirm that it is not problematic.

Lastly, the project team was required to adjust the scope of the project in order to account for the lack of cooperation from Santa Clara County. The most upstream signalized intersection was removed from the study site due to the institutional barrier, and the most upstream freeway bottleneck (at Capitol Expy, on-ramp) was excluded. The finalized study site is shown in figure 3.1. This includes a three mile section (post mile 1.65 to 4.48) of Northbound I-680 and a parallel arterial (Capitol Ave.) with 5 major signalized intersections.

The site selection process was a valuable experience in how to compromise between desired operating conditions (i.e. presence of bottlenecks) and practical constraints such as infrastructure, data availability, and institutional barriers.

Additional information relevant to the selected site is shown in the figures and tables below.

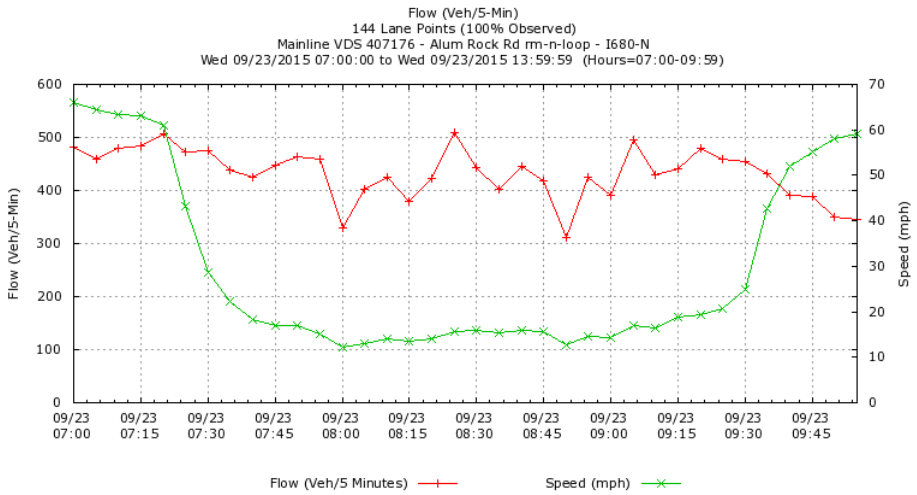


Figure A.7 Flow and speed of I-680 Northbound near Alum Rock Ave. on-ramp.

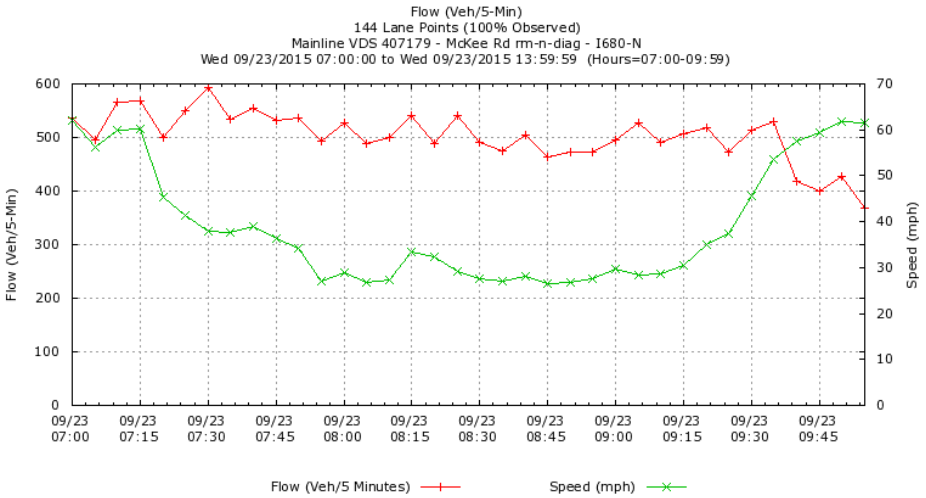


Figure A.8: Flow and speed of I-680 Northbound near McKee Rd. on-ramp.

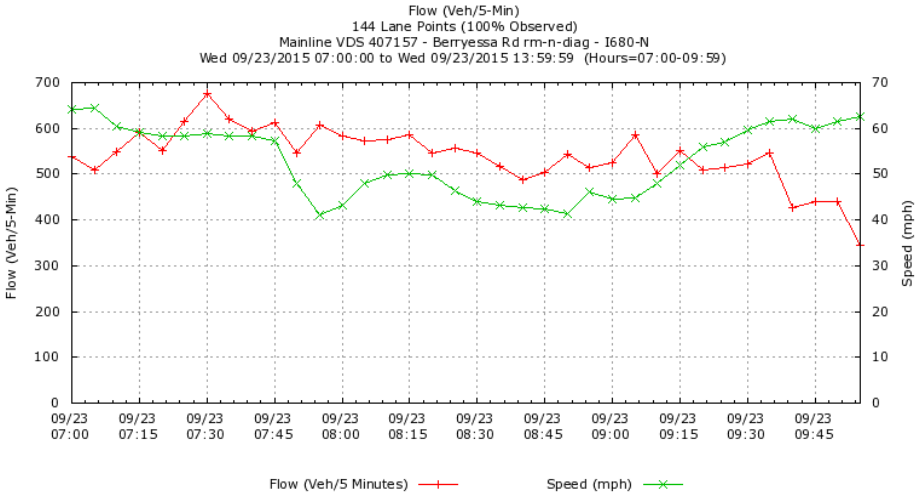


Figure A.9: Flow and speed of I-680 Northbound near Berryessa Rd. on-ramp.

Table A.1: Alum Rock Ave. AM peak ramp metering rates.

Time of Day	Alum Rock Ave. (loop)		Alum Rock Ave. (diagonal)	
	Mainline Occupancy	Meter Rate	Mainline Occupancy	Meter Rate
6:00 – 7:00 AM	≤ 3%	No metering	≤ 4%	No metering
	3% to 12%	900 vph/lane	4% to 14%	900 vph/lane
	≥ 12%	300 vph/lane	≥ 14%	560 vph/lane
7:00 – 7:30 AM	≤ 3%	No metering	≤ 4%	No metering
	3% to 5%	900 vph/lane	4% to 7%	900 vph/lane
	≥ 5%	300 vph/lane	≥ 7%	480 vph/lane
7:30 – 9:00 AM	≤ 3%	No metering	≤ 4%	No metering
	3% to 5%	900 vph/lane	4% to 7%	900 vph/lane
	≥ 5%	400 vph/lane	≥ 7%	480 vph/lane
9:00 – 10:00 AM	≤ 10%	No metering	≤ 12%	No metering
	10% to 12%	900 vph/lane	12% to 14%	900 vph/lane
	≥ 12%	360 vph/lane	≥ 14%	560 vph/lane

Table A.2: McKee Rd. and Berryessa Rd. AM peak ramp metering rates.

Time of Day	McKee Rd.		Berryessa Rd.	
	Mainline Occupancy	Meter Rate	Mainline Occupancy	Meter Rate
6:00 – 7:00 AM	≤ 4%	No metering	≤ 3%	No metering
	4% to 14%	900 vph/lane	3% to 14%	900 vph/lane
	≥ 14%	420 vph/lane	≥ 14%	420 vph/lane
7:00 – 7:15 AM	≤ 4%	No metering	≤ 3%	No metering
	4% to 6%	900 vph/lane	3% to 5%	900 vph/lane
	≥ 6%	400 vph/lane	≥ 5%	560 vph/lane
7:15 – 7:30 AM	≤ 4%	No metering	≤ 3%	No metering
	4% to 6%	900 vph/lane	3% to 5%	900 vph/lane
	≥ 6%	560 vph/lane	≥ 5%	560 vph/lane
7:30 – 7:45 AM	≤ 4%	No metering	≤ 3%	No metering
	4% to 6%	900 vph/lane	3% to 5%	900 vph/lane
	≥ 6%	700 vph/lane	≥ 5%	600 vph/lane
7:45 – 8:00 AM	≤ 4%	No metering	≤ 3%	No metering
	4% to 6%	900 vph/lane	3% to 5%	900 vph/lane
	≥ 6%	700 vph/lane	≥ 5%	650 vph/lane
8:00 – 8:15 AM	≤ 4%	No metering	≤ 3%	No metering
	4% to 6%	900 vph/lane	3% to 5%	900 vph/lane
	≥ 6%	600 vph/lane	≥ 5%	600 vph/lane
8:15 – 8:30 AM	≤ 4%	No metering	≤ 3%	No metering
	4% to 6%	900 vph/lane	3% to 5%	900 vph/lane
	≥ 6%	400 vph/lane	≥ 5%	600 vph/lane
8:30 – 9:00 AM	≤ 4%	No metering	≤ 3%	No metering
	4% to 6%	900 vph/lane	3% to 5%	900 vph/lane
	≥ 6%	400 vph/lane	≥ 5%	510 vph/lane
9:00 – 10:00 AM	≤ 12%	No metering	≤ 12%	No metering
	12% to 14%	900 vph/lane	12% to 14%	900 vph/lane
	≥ 14%	420 vph/lane	≥ 14%	450 vph/lane

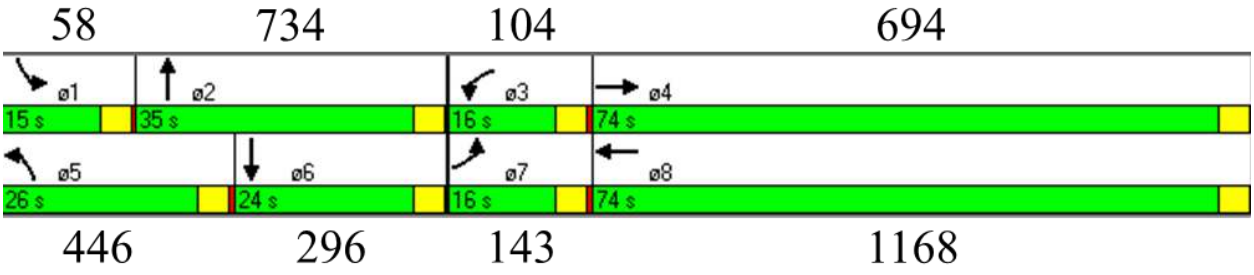


Figure A.10 Signal timing plan and typical volumes at Capitol Ave. and Alum Rock Ave.

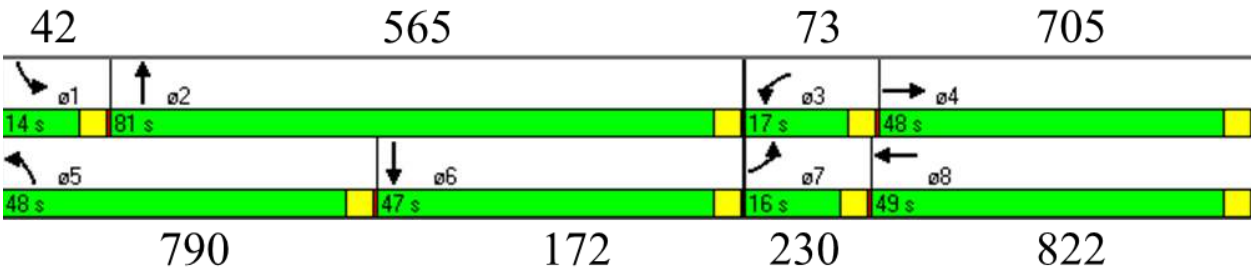


Figure A.11 Signal timing plan and typical volumes at Capitol Ave. and Berryessa Rd.

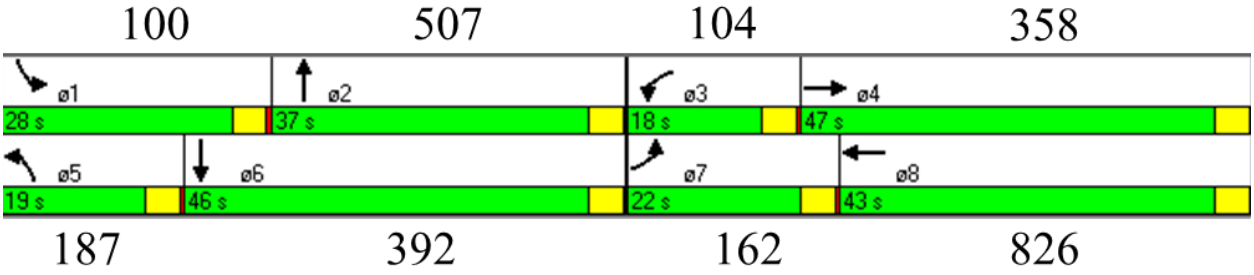


Figure A.12 Signal timing plan and typical volumes at Capitol Ave. and Mabury Rd.

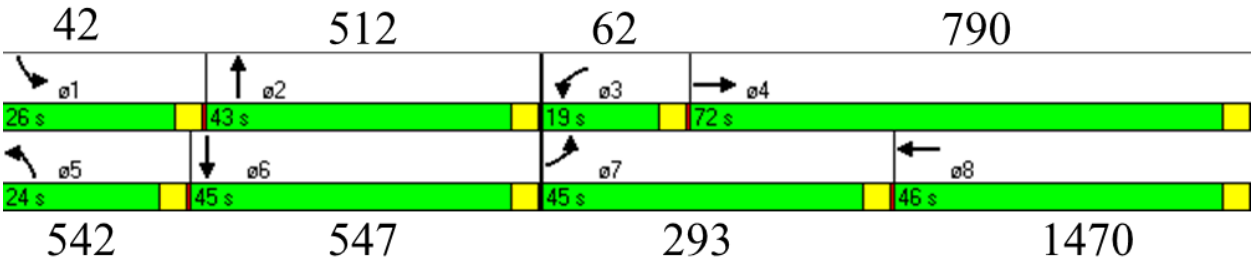


Figure A.13 Signal timing plan and typical volumes at Capitol Ave. and McKee Rd.

APPENDIX B: ADDITIONAL QUEUE OVERRIDE FIELD STUDY DATA

Additional details for data collected from Tuesday May 17, 2016 to Thursday May 19, 2016 are shown in Figures B.1 through B.3. Each figure provides a detailed view of the cumulative vehicle count curves of the mainline lanes, obtained at all three cameras locations shown in figure 3.2, vs time, t, in an oblique scale.

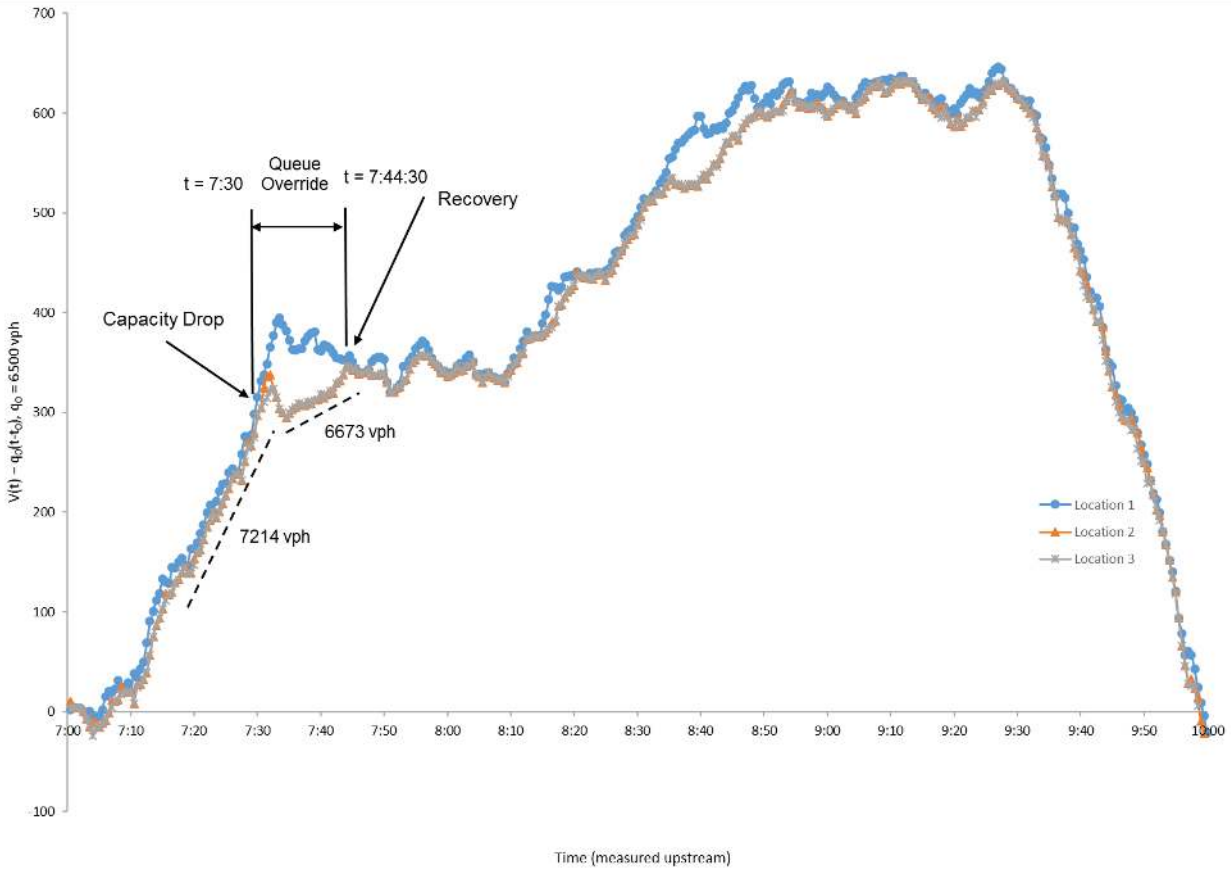


Figure B.1 May 17, 2016: $O(t)$ curves for locations 1 through 3.

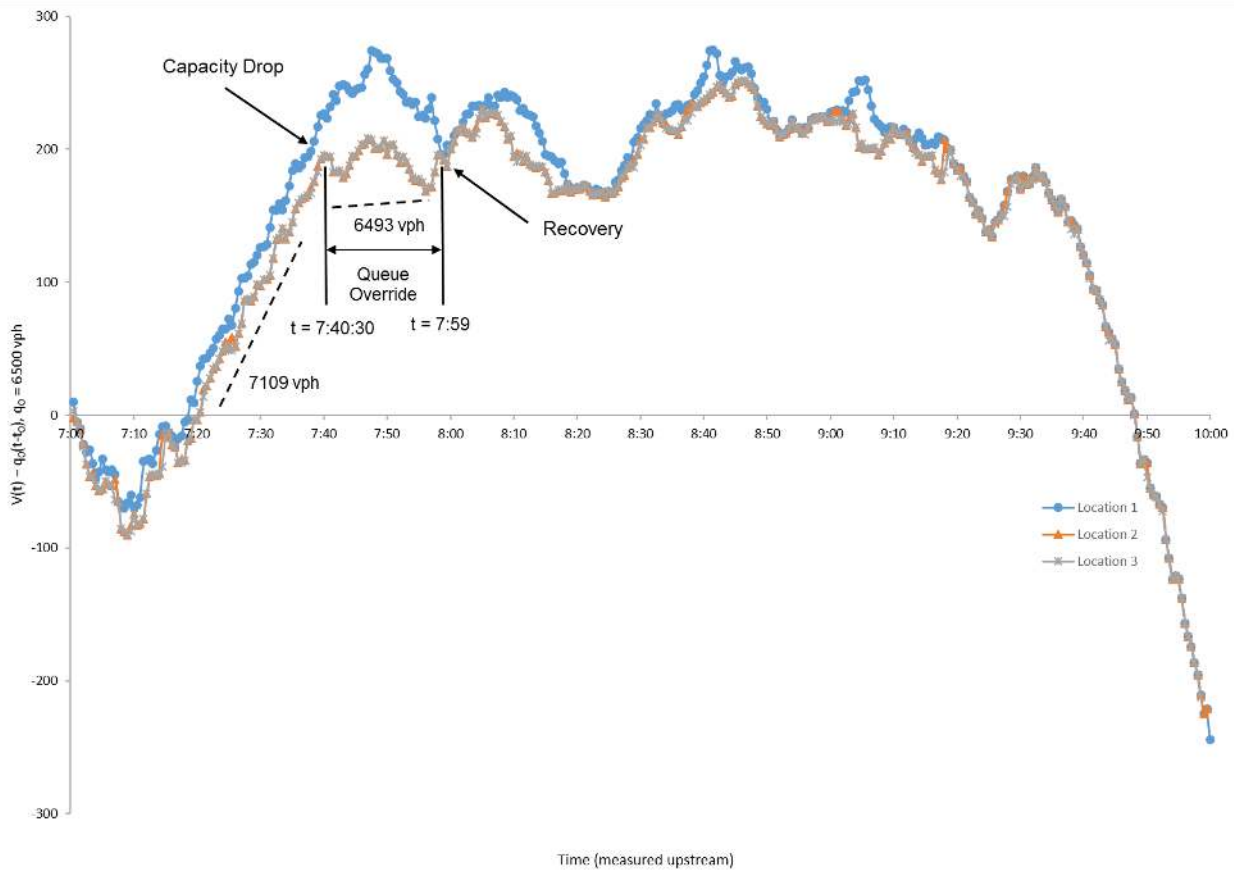


Figure B.2 May 18, 2016: $O(t)$ curves for locations 1 through 3.

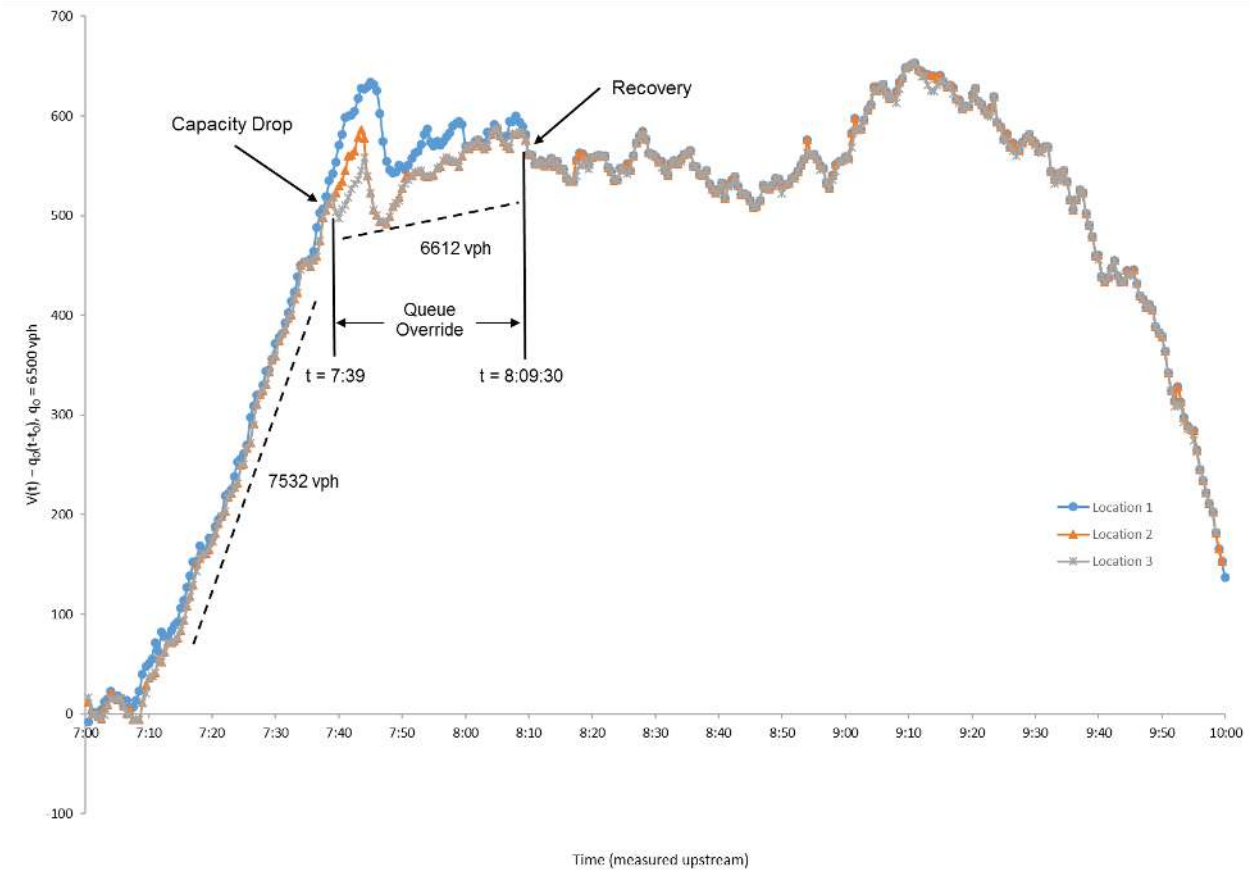


Figure B.3 May 19, 2016: $O(t)$ curves for locations 1 through 3.

APPENDIX C: PRELIMINARY MODEL CALIBRATION

The microscopic simulation model was calibrated to existing conditions prior to the evaluation of the proposed control strategy. Twenty replications of the simulation model runs with different random number seeds were made using the existing demand, turning percentages, ramp metering algorithm, and signal timing plans. The predicted flows and speeds at selected locations on the freeway mainline were compared with real traffic measurements in every 5 minutes to assess the accuracy of the simulation model in representing observed conditions. For flows, we need at least 85% of the flows to be acceptable and GEH<5 (Dowling et al, 2004). According to this criterion, the predicted flow is acceptable if it satisfies the requirement below.

Link flow quantity

- If 700 vph < real flow < 2700 vph, simulated flow has an error within 15%;
- If real flow < 700 vph, simulated flow has an error within 100 vph;
- If real flow > 2700 vph, simulated flow has an error within 400 vph.

The GEH statistic is computed as

$$GEH(k) = \sqrt{\frac{2[M(k) - C(k)]^2}{M(k) + C(k)}} \quad \text{(C.1)}$$

Where:

$M(k)$: simulated flow during the k-th time interval (veh/hour)
$C(k)$: flow measured in the field during the k-th time interval (veh/hour)

For speed, the relative root mean squared error (RRMSE) of simulated speed values are required to be 15% or lower, on average of all detectors. For arterial flows, GEH < 5 must be satisfied for at least 85% of all 5-minute time intervals, for each turning movement of the major intersections.

Tables C.1 and C.2 summarize the calibration results for the three detectors along the four mile stretch of Northbound I-680, as well as the 5-minute turning movement flows of the major arterial intersections. Figures C.1 through C.3 compare the speed data collected from the field with the simulated data. It can be seen that on average, for the freeway, the simulated flows and speeds satisfy the calibration criteria. Similarly, the simulated flows at major arterial intersections satisfy the calibration criterion. Calibration was not performed in the southbound freeway direction because of the low volume during the morning peak analysis periods.

Table C.1 Calibration of freeway and arterial flows.

Freeway: 5-min flows of I-680 Northbound					
Detector Location	Target	Cases	Cases Met	% Met	Target Met?
Alum Rock Ave. on-ramp (loop)	GEH < 5 for > 85% of *k*	24	24	100.00%	Yes
McKee Rd. on-ramp	GEH < 5 for > 85% of *k*	24	24	100.00%	Yes
Berryessa Rd. on-ramp	GEH < 5 for > 85% of *k*	24	23	95.83%	Yes
Overall	GEH < 5 for > 85% of *k*	72	71	98.61%	Yes
Arterial: 5-min flows of major intersections					
Capitol Ave. & Alum Rock Ave.					
	Target	Cases	Cases Met	% Met	Target Met?
Northbound	GEH < 5 for > 85% of *k*	48	47	97.92%	Yes
Southbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Westbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Eastbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Capitol Ave. & Berryessa Rd.					
	Target	Cases	Cases Met	% Met	Target Met?
Northbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Southbound	GEH < 5 for > 85% of *k*	72	72	100.00%	Yes
Westbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Eastbound	GEH < 5 for > 85% of *k*	72	72	100.00%	Yes
Capitol Ave. & Mabury Rd.					
	Target	Cases	Cases Met	% Met	Target Met?
Northbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Southbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Westbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Eastbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Capitol Ave. & McKee Rd.					
	Target	Cases	Cases Met	% Met	Target Met?
Northbound	GEH < 5 for > 85% of *k*	48	46	95.83%	Yes
Southbound	GEH < 5 for > 85% of *k*	72	71	98.61%	Yes
Westbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Eastbound	GEH < 5 for > 85% of *k*	72	70	97.22%	Yes
Alum Rock Ave. & I-680 Northbound Off-ramp					
	Target	Cases	Cases Met	% Met	Target Met?
Northbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Westbound	GEH < 5 for > 85% of *k*	48	48	100.00%	Yes
Eastbound	GEH < 5 for > 85% of *k*	48	47	97.92%	Yes
Arterial: Overall	GEH < 5 for > 85% of *k*	1008	1001	99.30%	Yes

Table C.2 Calibration of freeway speeds.

	Detector Location		
	Alum Rock Ave. on-ramp (loop)	McKee Rd. on-ramp	Berryessa Rd. on-ramp
RMSSE	2.21%	16.14%	13.80%
Target	<15%	<15%	<15%
Target Met?	Yes	No	Yes
Overall	**10.72%**		
Target	**RMSSE<15%**		
Target Met?	**Yes**		

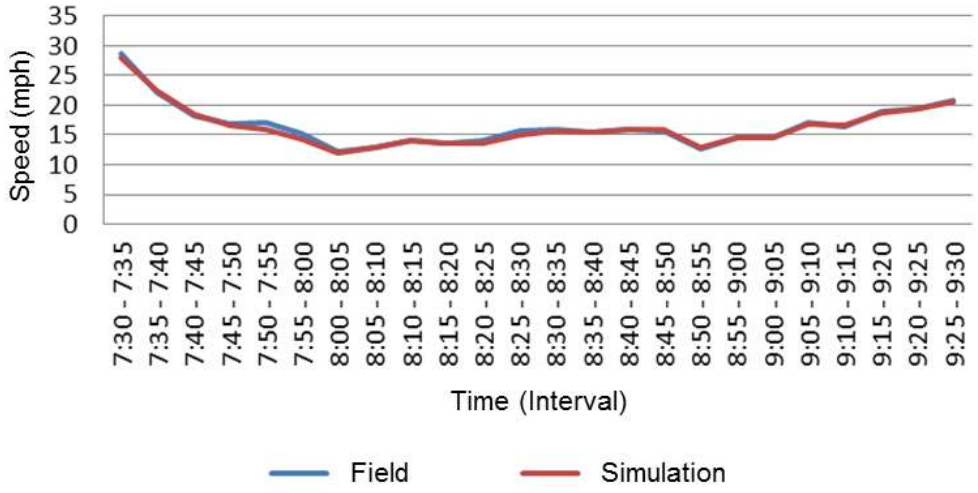


Figure C.1 Observed and simulated speeds near Alum Rock Ave. on-ramp (loop).

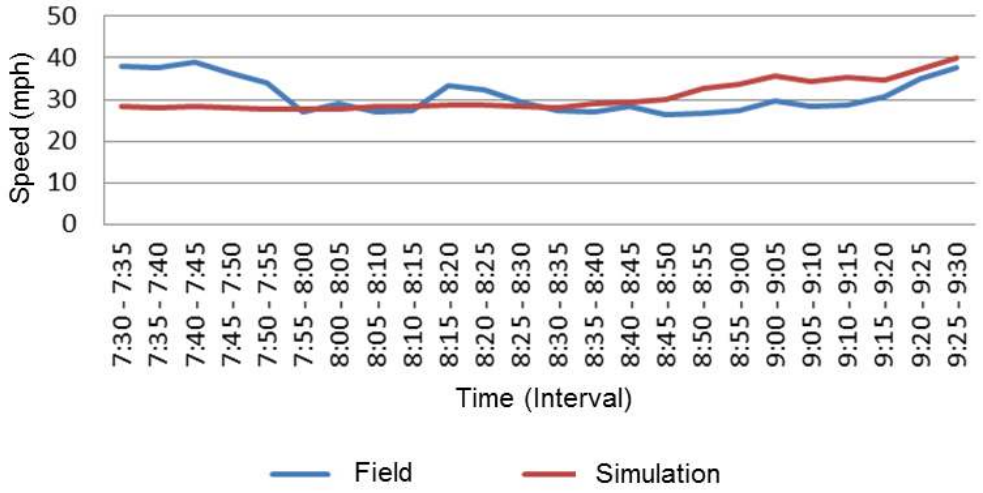


Figure C.2 Observed and simulated speeds near McKee Rd. on-ramp.

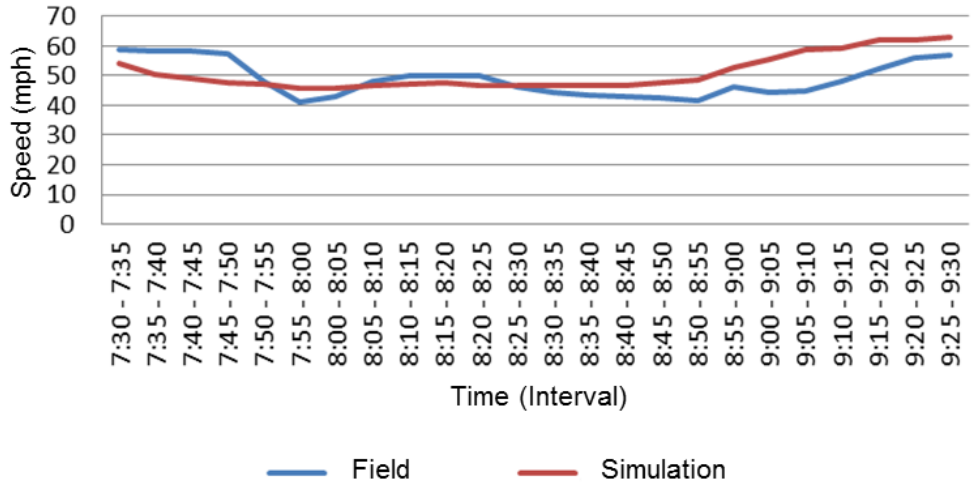


Figure C.3 Observed and simulated speeds near Berryessa Rd. on-ramp.

C.1 Reference

Dowling, R., A. Skabardonis, and V. Alexiadis, 2004. Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software. Federal Highway Administration FHWA-HRT-04-038.

APPENDIX D: ENHANCED SIMULATION MODEL

The microscopic simulation model was enhanced according to the work in Lu et al. (2017), which states that microscopic vehicle behavior and interaction with the nearby vehicles determine overall traffic behavior at the macroscopic level based upon the following factors: maximum acceleration/deceleration and driver behaviors such as preferred headway and response time, gap acceptance threshold for lane changing, and perception advance time-period or distance for lane changing. Those parameters directly affect density and delays in the simulation, and thus the overall traffic pattern.

The proposed driving behavior model is built upon the basic framework of the NGSIM oversaturated flow model proposed by Yeo et al. (2008). Some important extensions and modifications were made in order to depict detailed car following and lane changing behavior that were not represented in the original model.

To determine the trajectory of a vehicle at a microscopic level, it is necessary and sufficient to iteratively determine its location at each time step, which can be realized through a discrete kinematic model if the desired acceleration and current speed are known. The latter is known from the last step calculation. The former is determined by the dynamic interactions with the adjacent vehicles, geometric constraints, and the overall traffic conditions. The dynamic interactions include time/clearance gaps for safety and mobility, and possible scenarios associated with lane changes. Those scenarios are further partitioned into fundamental scenarios (or movement phases) and transitions between them for continuous/smooth speed trajectories:

CF: Regular car following mode
YCF: Yielding (cooperative) car following mode
LC: Lane change mode, which includes discretionary lane change (DLC) and mandatory lane change (MLC)
ACF: After lane changing car following mode (a driver temporarily adopts a short gap following a lane change maneuver)
BCF: Before lane changing car following mode (a driver speeds up or slows down to align with an acceptable gap in the target lane)
RCF: Receiving car following mode (a driver temporarily adopts a short gap after a vehicle from the adjacent lane merges in front)

D.1 Discrete Kinematic Model

The discretized kinematic model is used at the microscopic level to determine vehicle position x for the next simulation time step $t + \Delta t$ based on all of the information at the current time step t. The following are the first and second order Taylor series expansion approximations of $x_{des}(t + \Delta t)$, the desired location of the subject vehicle (SV) for the next time step:

$$x_{des}(t + \Delta t) \approx x(t) + \Delta t \cdot v_{des}(t + \Delta t) \approx x(t) + \Delta t \cdot v(t) + \frac{\Delta t^2}{2} \cdot a_{des}(t) \qquad \text{(D.1)}$$

Equation D.1 states that the expected (or desired) location of the subject vehicle can be determined as follows:

(a) First order approximation: if one knows the desired speed v_{des} and current location $x(t)$
(b) Second order approximation: if one knows the desired acceleration a_{des} and the current location $x(t)$ and speed $v(t)$

We use the term *desired* (or expected) because the actual speed is subject to constraints imposed by adjacent traffic conditions, and the maneuvers of the subject vehicle (SV), which will be discussed in the following. Similarly, one could easily deduce the expression for the relative distance with respect to the leading vehicle.

D.2 Car Following Model

The car following model used in this study is Newell's simplified car following model with constraints for safety and acceleration (Newell, 2002). The Gipps' deceleration component (Gipps, 1986; Ciuffo et al., 2012) is used here to place a safety margin on Newell's simplified equation. The headway parameter in Newell's model and the reaction time in Gipps' model need to be carefully selected. The free flow acceleration a_F is described by the following equation:

$$a_F = a_M \left[1 - \left(\frac{v(t)}{V_0}\right)^{\alpha}\right] \tag{D.2}$$

where,

V_0: free flow speed [m/s]
a_M: maximum acceleration [m/s^2]
α: acceleration exponent

The final desired acceleration a_{des} equation reads:

$$a_{des} = \min(a_F, a_N, a_G) \tag{D.3}$$

where,

a_F: free-flow acceleration
a_N: acceleration from Newell's model
a_G: acceleration from Gipps model.

For smooth transition between different car following modes, the following transition treatment is adopted:

$$a(t) = a(t - \Delta t) + \frac{a_{des}(t) - a(t - \Delta t)}{\omega(t)} \tag{D.4}$$

where $\omega(t)$ is a smoothing factor and $\omega(t)$: $1 \rightarrow 0$ monotonically, which is a time based interpolation. In this model, $\omega(t)$ is defined as the following:

$$\omega(t) = \frac{t}{T}, 0 \le t \le T$$

D.2.1 Permissive Speed

The permissible speed of the subject vehicle considers the real-world phenomenon that speeds are typically lower on the right lanes of roads in U.S. and other similar settings, thus a lane-based speed limit distribution is applied (Lane 1 is the leftmost lane).

$$V_L = \begin{cases} V_{limit}, & lane\ 1 \\ 0.95V_{limit}, & lane\ 2 \\ 0.9V_{limit}, & lane\ 3,4,\cdots \end{cases} \quad \text{(D.5)}$$

where,

V_L: redistributed speed limit [m/s]
V_{limit}: the section posted speed limit [m/s]

Incorporating this lane-based speed limit, the free-flow speed $V_{lane\ limit}$ becomes:

$$V_{lane\ limit} = min\{V_L, V_{des}\}$$

where,

V_{des}: individual driver's desired speed [m/s]

D.2.2 Speed Friction across Lanes

It is generally recognized that most drivers do not drive significantly faster than those in the adjacent lane due to safety concerns, and significantly reduce their speeds if planning a lane change into the slower adjacent lane. A model for the *friction effect* is proposed to account for this real-world scenario:

$$V_0 = \begin{cases} min\{v_r, v_l\} + \dfrac{V_{lane\ limit} - min\{v_r, v_l\}}{c_f}, & V_{lane\ limit} > min\{v_r, v_l\} \\ V_{lane\ limit}, & Otherwise \end{cases} \quad \text{(D.6)}$$

where,

V_0: desired speed adjusted for lane friction [m/s]
v_l, v_r: speeds ahead on the left/right lane, respectively [m/s]
c_f: coefficient of lane friction, a constant tunable parameter adjusted in model calibration

D.3 Lane Change Models

The fundamental scenarios associated with lane changing (LC) are described in the following sections and illustrated in Figure D.1.

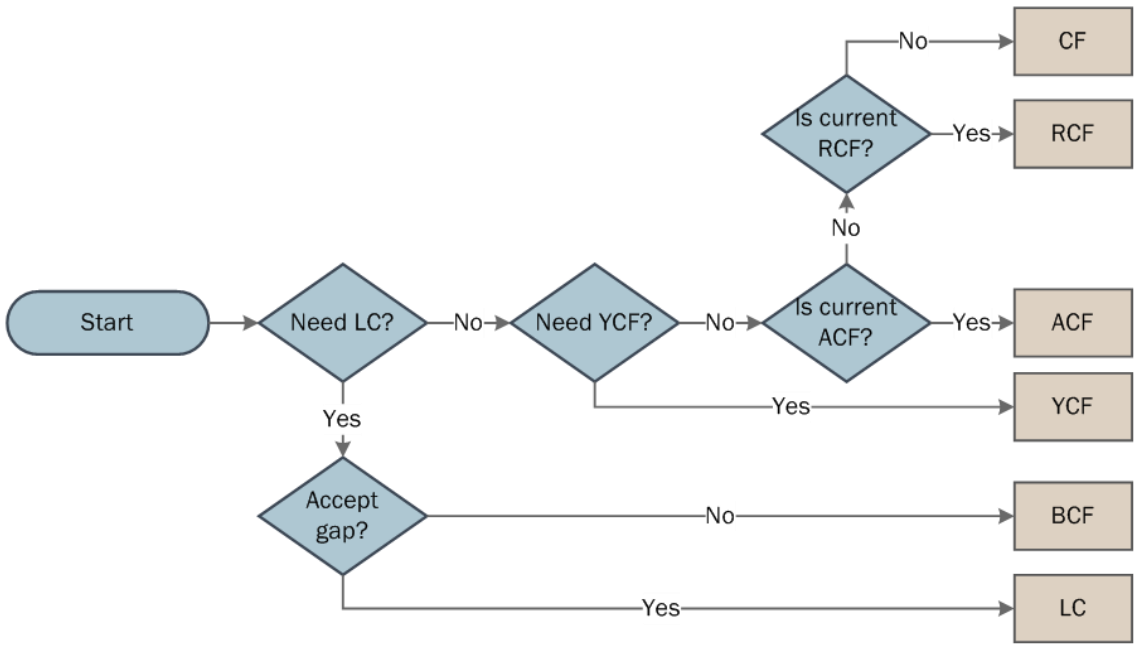


Figure D.1 Lane change behavior structure.

D.3.1 Lane Change Motivation Generation

The lane changing motivation is measured by a desire index γ between zero and one. Zero means the driver has no intention to change lane, while one indicates the driver has the highest intention.

(a) Mandatory Lane Change (MLC) Motivation: If the driver must merge onto the freeway or to the lane for exit, a mandatory lane change desire $\gamma_{m,\{l,r\}}$ is generated by using the following equation:

$$\gamma_{m,\{l,r\}} = \begin{cases} 0 & d_e \geq N_{lc}E_{max}, t_e \geq N_{lc}T_{max} \\ 1 - min\left[\frac{d_e - E_{min}}{N_{lc}(E_{max} - E_{min})}, \frac{t_e - T_{min}}{N_{lc}(T_{max} - T_{min})}\right] & otherwise \\ 1 & d_e \leq N_{lc}E_{min}, t_e \leq N_{lc}T_{min} \end{cases} \quad \text{(D.7)}$$

where,

d_e: distance to the end of the acceleration lane/turning point [m]
t_e: $\frac{d_e}{v_f}$, time to the end of the acceleration lane/turning point [s]
E_{max}, E_{min}: distance parameters [m]
T_{max}, T_{min}: time parameters [s]
N_{lc}: number of lane changes required

The parameters E_{max}, E_{min}, T_{max} and T_{min} are different for on-ramp and off-ramp mandatory lane changes.

(b) Discretionary Lane Change (DLC) Motivation: The discretionary lane change motivation is generated based on the anticipatory speed ahead on the current lane and the adjacent lanes. The anticipatory speed is determined as:

$$v_{ant} = \min\left(\tilde{v}_{\{l,r\}}, v_{l'}\right) \tag{D.8}$$

where,

v_{ant}: anticipatory speed on the target lane [m/s]
$\tilde{v}_{\{l,r\}}$: average speed of the target lane (left or right) ahead [m/s]
$v_{l'}$: speed of the leader in the target lane [m/s]

The discretionary lane change incentive $\gamma_{d\{l,r\}}$ is determined as:

$$\gamma_{d\{l,r\}} = \min\left(\max\left(0, \frac{v_{ant} - \tilde{v}_0}{\max(\tilde{v}_0, V_{dlc})}\eta_{\{l,r\}}\right), 1\right) \tag{D.9}$$

where,

$\tilde{v}_0$: average speed of the current lane ahead [m/s]
$\eta_{\{l,r\}}$: parameter (=1 for left lane change and <1 for right lane change)
V_{dlc}: minimum speed parameter [m/s]

(c) DLC Motivation for High Occupancy Vehicles (HOVs): To account for carpoolers' extra incentive to access the HOV lane, $v_{ant,HOV}$, the anticipated speed on the HOV lane is increased by a pre-defined value as:

$$v_{ant,HOV} = v_{ant} + \Delta V_{HOV} \tag{D.10}$$

where,

ΔV_{HOV}: Bias value toward HOV lane for HOVs (> 0) [m/s]

D.3.2 Gap Acceptance Models

The lane change gap acceptance model was defined separately for mandatory and discretionary lane changes. For mandatory lane changes, safety is the primary concern. For discretionary lane changes, both comfort and safety are taken into account. If both the forward and backward gaps are accepted (see Figure D.2), the vehicle will make an LC maneuver immediately. We separate those two factors since, in practice, the distance to the tentative leader can be shorter than that with respect to the follower for driver's safety and comfort.

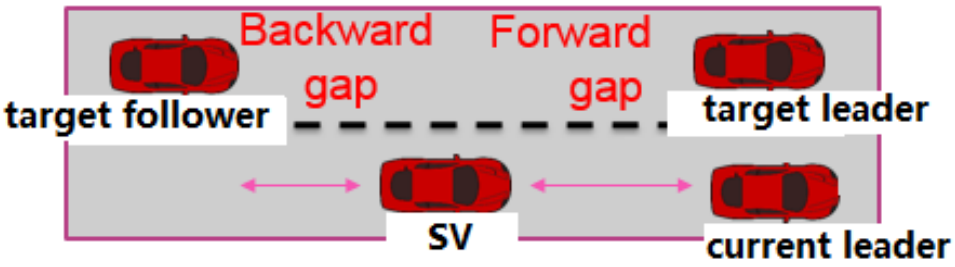


Figure D.2 Illustration of forward and backward gaps.

(a) MLC Gap Acceptance: Once the driver decides to change lane, the driver will search for gaps in the target lane. The gap acceptance is based on the minimum gap that the driver anticipates to be available after the lane change maneuver. The forward and backward gaps have the following relationships:

$$g_f \geq g_r \geq g_j$$

g_f: anticipated minimum forward gap after lane change
g_r: anticipated minimum backward gap after lane change
g_j: jam gap

The forward and backward gap are accepted if the following condition is met:

$$g_{f,min} \geq d_{jam} \text{ and } g_{b,min} \geq d_{jam} \quad \text{(D.11)}$$

Calculation of g_f and g_r follows the standard kinematics (Eq. D.1) given the anticipated acceleration/deceleration of the leader/follower and is calculated following the ACF rule:

$$\hat{b}_g = \phi_f b_f \quad \text{(D.12)}$$

where,

ϕ_f: relaxation parameter for anticipated deceleration of the leader/follower in the target lane with respect to the maximum comfortable deceleration (different for on-ramp and off-ramp mandatory lane changes) and ϕ_f is smaller for the leader and larger for the follower
b_f: maximum deceleration that the lane-changer is comfortable applying [m/s^2]
$\hat{b}_g$: anticipated deceleration of the leader/follower in the target lane [m/s^2]

(b) DLC Gap Acceptance: Besides safety, gap acceptance for discretionary lane changing also considers the anticipatory acceleration.

- Forward gap:

$$g_{f,min} \geq d_{jam} \text{ and } a_{ant} \geq 0 \quad \text{(D.13)}$$

- Backward gap:

$$g_{b,min} \geq d_{jam} \quad \text{(D.14)}$$

Because the forward gap is only accepted when the lane-changer anticipates an acceleration, the disturbance caused to the follower in the target lane is also reduced. Therefore, the discretionary lane change is likely to cause smaller disturbances to the following traffic flow.

D.3.3 Combining DLC and MLC Motivations

At each time step, the following lane change motivations are generated for each driver: left-lane mandatory lane change motivation $\gamma_{m,l}$, right-lane mandatory lane change motivation $\gamma_{m,r}$, left-lane discretionary lane-change motivation $\gamma_{d,l}$, and right-lane discretionary lane-change motivation $\gamma_{d,r}$. All four are combined, with priority given to mandatory lane changes.

- Case 1: Mandatory lane change desire is larger than zero ($\gamma_{m,r} > 0$ or $\gamma_{m,l} > 0$): If the driver has a mandatory lane change motivation for a target lane (left/right), then the discretionary lane change desire for the opposite target lane (right/left) is set to zero. That is, if $\gamma_{m,r} > 0$ $(\gamma_{m,l} > 0)$, then $\gamma_{d,l} = 0$ ($\gamma_{d,r} = 0$). The final desire is then determined by:
$$\gamma = \gamma_m + \kappa\gamma_d$$
where κ is weighting parameter for discretionary lane change.

- Case 2: Mandatory lane change desire is zero ($\gamma_{m,r} = 0$ and $\gamma_{m,l} = 0$): In this case, the desire is determined by:
$$\gamma = \max(\gamma_{d,l}, \gamma_{d,r})$$
A random variable ϵ, that follows a normal distribution with the driver's average lane changing desire threshold as the mean, is generated at the beginning of the simulation. If $\gamma > \epsilon$, the driver decides to change lane at the current time step and starts searching for gaps in the target lane, otherwise, the driver remains in the current lane.

D.3.4 Yielding Car Following (YCF)

At each simulation step, drivers monitor leaders in the adjacent lanes. If a leader has an intent for a lane change to the current lane of the SV, the driver will decide if YCF should be applied based on a cooperative factor ζ. A value between 0 and 1 is randomly generated for each driver at the beginning of the simulation, from a normal distribution with user specified mean $\bar{\zeta}$ and variance $\bar{\sigma}$. Any value less than the cooperative factor ζ will require the driver to apply YCF. Note that drivers only yield to *mandatory lane changes* (i.e. lane change is the only feasible scenario), which will be discussed later.

However, in the yielding process, the "deadlock" phenomena may appear when both the yielding SV and the lane changer are stationary while the gap acceptance criterion is not satisfied; the yielding vehicle is not moving because the driver decides to yield; the lane changer cannot move because insufficient backward gap will cause the driver to slow down and skip the current gap according to the lane changing algorithm presented above. To avoid deadlock, a constraint on the yielding intent is added. Under this constraint, the driver will not yield if:

1) The current speed is less than a minimum threshold;
2) The yielding car following mode has been active over a certain period of time for the same lane changer;

3) The current spacing is negative (the front-bumper of the subject vehicle has already passed the rear-bumper of the potential lane changer).

In YCF mode, smaller jam gap, reaction time, and headway are adopted as follows:

$$\begin{aligned} \tau_h' &= \varphi_h \tau_h, \quad & 0 < \varphi_h < 1 \\ d_j' &= \varphi_j d_j, \quad & 0 < \varphi_j < 1 \\ \tau_r' &= \varphi_r \tau_r, \quad & 0 < \varphi_r < 1 \end{aligned} \tag{D.15}$$

where,

τ_h, τ_h': headway and reduced headway, respectively
d_j, d_j': jam gap and reduced jam gap, respectively
τ_r, τ_r': reaction time and reduced reaction time, respectively
$\varphi_h, \varphi_j, \varphi_r$: reduction factor used to adjust the headway, jam gap and reaction time, respectively. The values of φ need to be determined in model calibration.

D.3.5 Receiving Car Following (RCF)

Once a SV vehicle finishes a lane change maneuver, the new follower in the destination lane will apply RCF mode. The RCF is the same as the regular CF except that reduced headway, jam gap, and reaction time are adopted. Once the RCF mode is activated, it will apply the same set of parameters as described in Equation D.15, and then gradually increase the reduction factor, and finally return to the regular CF mode. Specifically, assuming the number of transition time steps is denoted by I_s, the reduction parameters during the transition process are determined as:

$$\begin{aligned} \varphi_{h_i} &= \frac{1-\varphi_h}{I_s} \cdot i + \varphi_h \\ \varphi_{j_i} &= \frac{1-\varphi_j}{I_s} \cdot i + \varphi_j \\ \varphi_{r_i} &= \frac{1-\varphi_r}{I_s} \cdot i + \varphi_r \end{aligned} \tag{D.16}$$

where,

I: an integer ranging from 0 to I_s;
φ_{h_i}, φ_{r_i}, φ_{j_i}: reduction factors at ith time step after RCF activates.

D.3.6 After Lane Change Car Following (ACF)

ACF follows the same rule as the RCF mode that adopts reduced headway, jam gap and reaction time parameters and gradually returns to the regular CF mode.

D.3.7 Before Lane Change Car Following (BCF)

BCF is defined separately for on-ramp, off-ramp and discretionary LCs.

(a) BCF for On-ramp Merging: For on-ramp mandatory lane changing, if either the forward or the backward gap is insufficient due to the safety constraints (i.e., the gap does not meet equation 11), the vehicle will adopt one of the two BCF modes: synchronizing or gap-skipping.

- On-ramp BCF synchronizing: If the forward gap is rejected and the backward gap is accepted, then the driver will start to reduce speed to synchronize with the leader in the target lane, which means the driver will follow both the leaders in the target lane and in the current lane. The synchronizing acceleration is determined as:

$$a_{sync} = \min(a_c, \max(a_s, a_{sm})) \quad \text{(D.17)}$$

where,

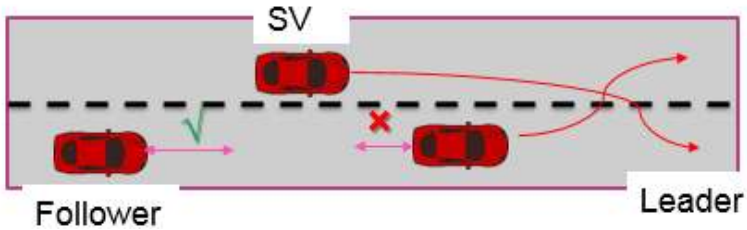
a_c: acceleration with respect to the current leader to keep a comfortable distance
a_s: acceleration with respect to the leader in the target lane
a_{sm}: acceleration with respect to both front and rear vehicles in the current lane to maintain a minimum coasting speed while synchronizing with the gap in the target lane
a_c and a_s are both calculated based on the basic car following model with shorter jam gap and reaction time and headway. a_{sm} is determined based on a minimum synchronizing coasting speed V_{sc}. If $v \leq V_{sc}$, $a_{sm} = 0$. Otherwise, $a_{sm} = \min(b_f, V_{sc} - v/\Delta t)$.

- On-ramp BCF gap skipping: If the backward gap is rejected, the driver will slow down with a comfortable deceleration, $b_c = \phi_c b_f$, to skip the current gap and seek the next gap behind, where ϕ_c is the comfortability factor ($0<\phi_c \leq 1$). The minimum speed during this process cannot be lower than V_{sk}.

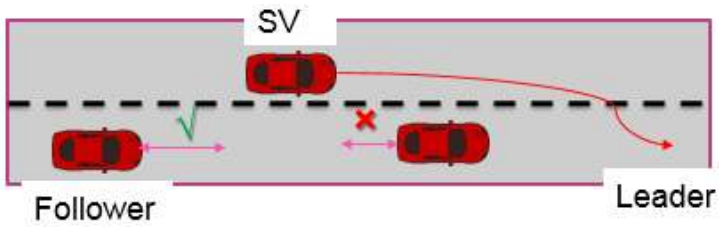
(b) BCF for off-ramp departure: The BCF for off-ramp departure includes four driving modes: synchronizing, overtaking, cruising, and advance LC.

- Synchronizing: The driver will follow the leaders in both the target lane and the current lane. Under either of the following conditions, the driver will adopt the synchronizing mode for BCF, as shown in Figure D.3.

1) The rear gap is rejected and
   a. The leader in the target lane has no intention for LC and the follower in the target lane intends to cooperate with the SV or;
   b. There is no leader in the target lane and the potential follower in the target lane intends to cooperate with the SV



(a)



(b)

Figure D.3 Lane changing under sufficient backward gap but insufficient forward gap.

2) Forward gap is rejected, rear gap is accepted and
    a. The leader has no intent to merge into the current lane; and passing is infeasible.

- Overtaking: In overtaking mode, the vehicle will apply a pre-defined acceleration to overtake the leader on the target lane. During overtaking, the speed cannot exceed a certain percentage of the speed limit. Under either of the following three conditions, the driver will adopt the overtaking mode for BCF.

1) Rear gap is rejected and
    a. the leader in the target lane intends to merge into the SV's lane; or
    b. both the potential leader and the potential follower in the target lane intend to merge into the SV's lane; or
    c. there is no leader in the target lane, and the follower in the target lane intends to merge into SV's lane; or
    d. there is no leader in the target lane, and it is feasible to pass the current gap and change lane.
2) Forward gap is rejected; rear gap is accepted
    a. the leader in the target lane does not intend to merge into the SV's lane: the SV accelerates until the rear gap is acceptable and completes the lane change if the gap is available;
    b. the leader in the target lane intends to merge into the SV's lane: wait until after the maneuver of the leader in adjacent lane

- Advance LC: Strictly speaking, this mode is not a typical gap-seeking driving mode. Instead, this mode happens if the current gap is acceptable and the vehicle is within a certain range to the turning/off-ramp. In this case, the driver will perform LC immediately.
- Cruising: In this mode, the driver will slow down to a certain cruising speed with a pre-defined deceleration.

D.4 Model Calibration

The driving behavior model was calibrated for a relatively complex freeway corridor in Sacramento, California. Details are discussed in the subsequent sub-sections.

D.4.1 Site Description

State Route (SR) 99 northbound was selected for model calibration. This section of freeway spans from the Elk Grove Blvd. interchange to the US-50 freeway interchange south of downtown Sacramento, CA. As indicated by the arrows in figure D.1, there are 9 interchanges with local arterial streets; 4 partial cloverleaf interchanges, 3 full cloverleaf interchanges, and 2 diamond interchanges with the local arterials. Furthermore, there are 3 lanes (one high occupancy vehicle or HOV lane and two general purpose lanes) in each direction upstream of the Calvine Road interchange while an additional general purpose lane is added downstream of that interchange. The on-ramp merging and weaving sections located downstream of the Elk Grove Blvd. interchange, as well as the off-ramp at the US-50 freeway interchange, contribute to the morning peak recurrent delay observed in this corridor. This peak period typically begins at 6:30 AM and ends around 9:00 AM, and the morning congestion pattern exhibits the typical peak period when there is high demand for suburb to downtown trips during the morning hours. Also shown in figure D.4, there is a wide coverage of detectors throughout the corridor. Detectors with good data quality are shown in blue, those with less acceptable data quality, shown in red, were not used to collect field data for calibration and validation. Currently, the on-ramps are metered using the local traffic responsive demand-capacity approach in order to control the flow of on-ramp traffic and mitigate the peak hour congestion.

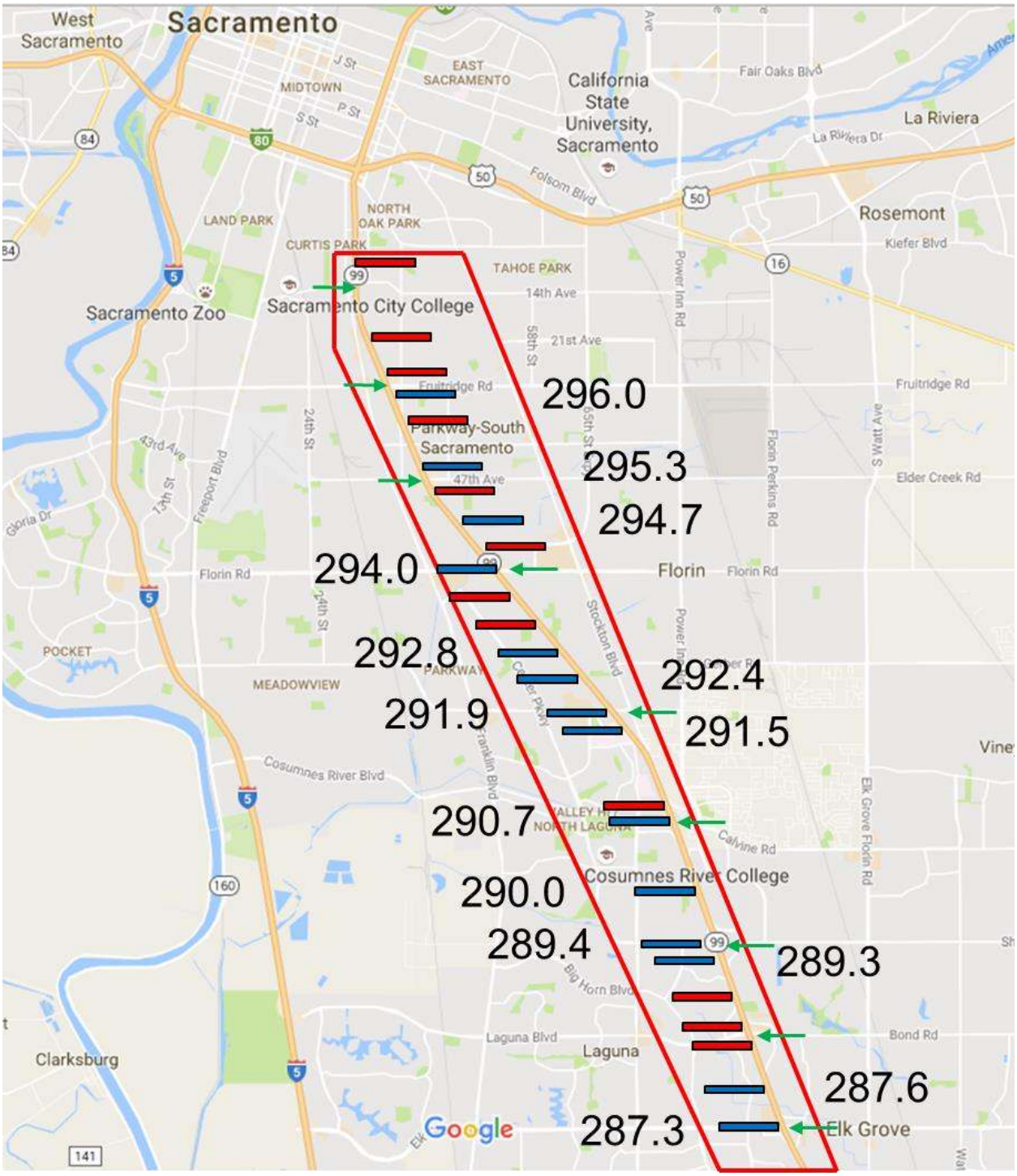


Figure D.4 Study site: SR-99 freeway south of downtown Sacramento, CA, with detector stations and post-mile identified.

D.4.2 Model Application and Calibration

A microscopic simulation model was built in the AIMSUN microscopic simulation framework (TSS, 2017) using the most up to date road geometry, lane configurations, and speed limits for the selected site. Freeway mainline and on-ramp data obtained from an 8-hour period (4:00 AM to 12:00 PM) on October 6, 2015 were used for the inputs in demand and turning percentages. This 8-hour period encompasses periods prior to, during, and after the peak. For this corridor, 5-minute interval loop detector data for flow were obtained from PeMS (Caltrans PeMS, 2016) and used as the demand input at the most upstream location of the simulation network and the entry points of the on-ramps, and as the turning percentages at any applicable mainline-off-ramp split. Ramp metering rates and algorithms were obtained from Caltrans District 3 and modeled in the microscopic simulation via the AIMSUN API (Application Programming Interface) (TSS, 2017).

Ten replications of the simulation model runs with different random number seeds were made in order to calibrate the model to the conditions of October 6, 2015. The predicted flows and speeds at selected locations on the freeway were compared with real traffic measurements at every 5 minute interval to assess the accuracy of the simulation model in representing the observed conditions. For flows, we need at least 85% of the flows to be acceptable and GEH<5 (Dowling et al., 2004). According to this criterion, the predicted flow is acceptable if it satisfies the requirement below.

Link flow quantity

- If 700 vph < real flow < 2700 vph, simulated flow has an error within 15%;
- If real flow < 700 vph, simulated flow has an error within 100 vph;
- If real flow > 2700 vph, simulated flow has an error within 400 vph.

The GEH statistic is computed as

$$GEH(k) = \sqrt{\frac{2[M(k) - C(k)]^2}{M(k) + C(k)}} \tag{D.18}$$

where,

$M(k)$: simulated flow during the k-th time interval (veh/hour)
$C(k)$: flow measured in the field during the k-th time interval (veh/hour)

In addition, the simulated flow-density relationship and queue propagation must be visually acceptable (WSDOT, 2002). This indicates that the fundamental diagrams of field observed and simulated flow vs. density should resemble similar patterns for key bottlenecks along the corridor, and the contour plots of the field observed and simulated speeds at all 5-minute intervals should exhibit similar trends over the length of the corridor as well as the duration of the study period.

Parameters in the model were adjusted by a few iterations of trial and error to find the best fit for the given field data. More sophisticated algorithms were avoided in order to ensure reasonable simulation and computation time for this complex corridor. Simulation experiments suggest that the following parameters provide a good fit:

- Reaction time (τ_r): 0.8 s
- Maximum acceleration (a_M): 2.4 m/s^2
- Maximum deceleration (b_f): 2.5 m/s^2
- Time headway (τ_h): 1.25 s
- Coefficient of lane friction (c_f): 0.5
- Speed difference threshold for discretionary lane changes (DLC) $\left(\frac{v_{ant}-\tilde{v}_0}{\max(\tilde{v}_0,V_{dlc})}\right)$: 0.15

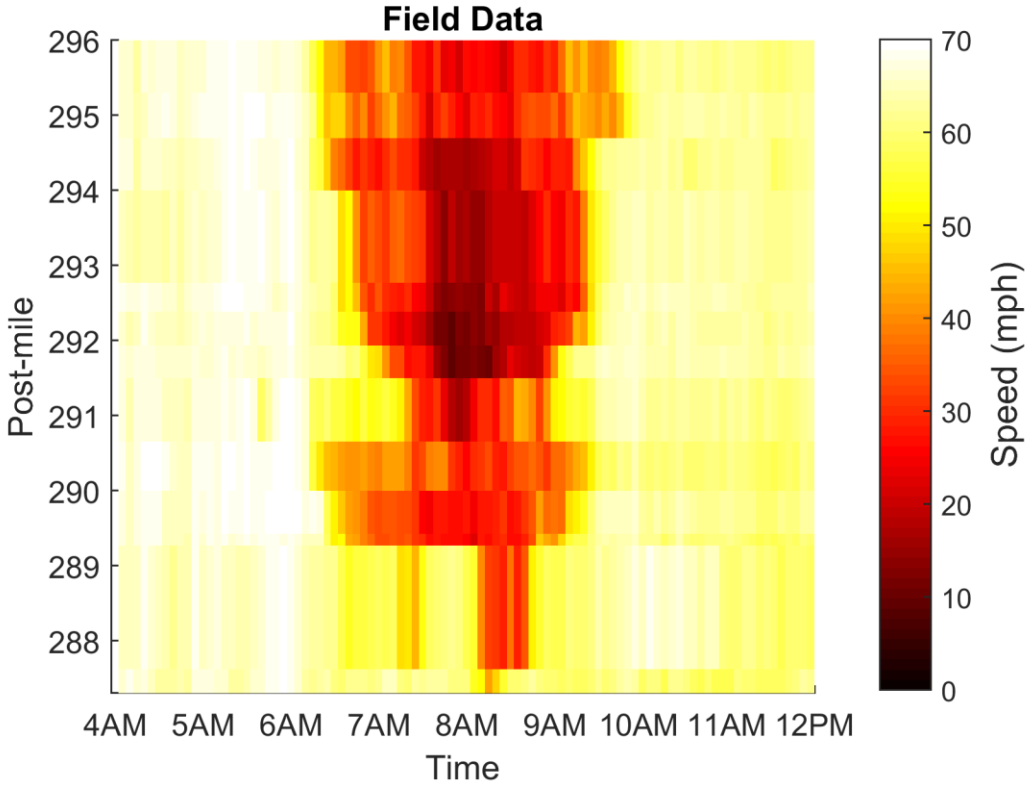
However, poor visibility near on-ramp merging areas on the upstream 2 mile section of the corridor required increasing the reaction time to 1.0 s to better reproduce the field observations.

Table D.1 summarizes the flow calibration results for the 14 detectors that provided good data along the 11-mile stretch of northbound SR-99. The HOV lane data were aggregated with those of the general purpose lanes because the HOV lane is equally congested and exhibits nearly the same congestion pattern (bottleneck location and duration). It can be seen that on average, the simulated flows satisfied the calibration criteria.

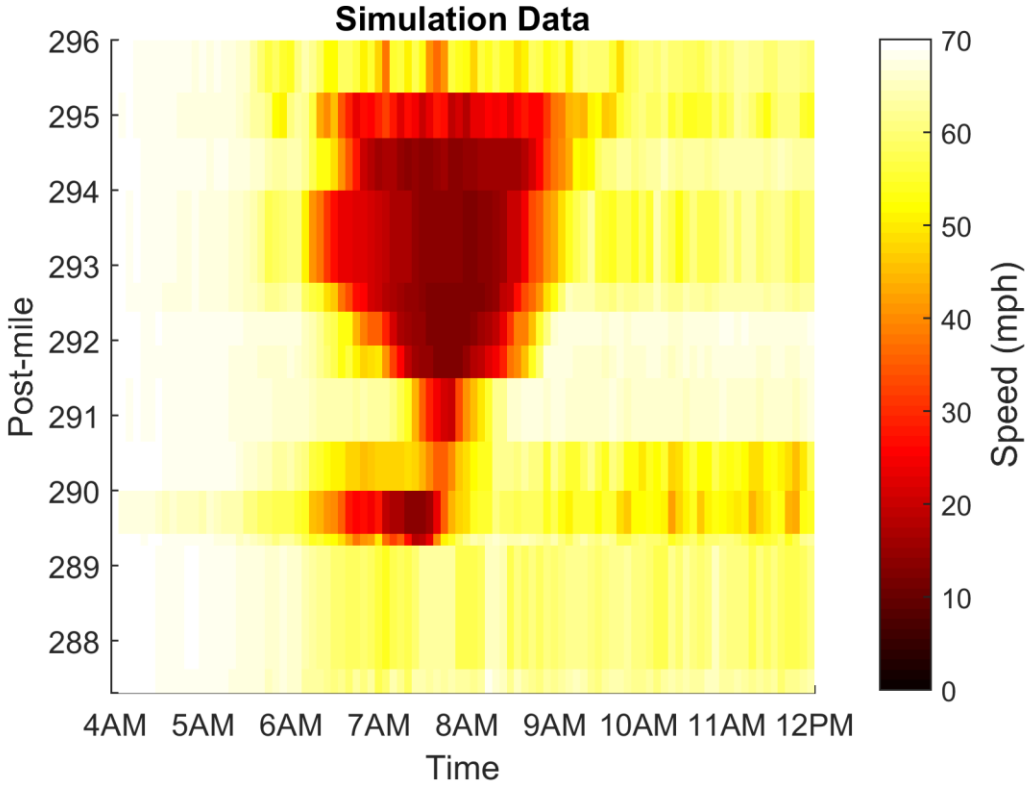
Table D.1 Calibration of freeway flows.

Freeway: 5-min flows of SR-99 Northbound					
Detector Location (post-mile)	Target	Cases	Cases Met	% Met	Target Met?
287.3	GEH < 5 for > 85% of *k*	930	921	99.0%	Yes
287.6	GEH < 5 for > 85% of *k*	930	904	97.2%	Yes
289.3	GEH < 5 for > 85% of *k*	930	915	98.4%	Yes
289.4	GEH < 5 for > 85% of *k*	930	914	98.3%	Yes
290.0	GEH < 5 for > 85% of *k*	930	921	99.0%	Yes
290.7	GEH < 5 for > 85% of *k*	930	920	98.9%	Yes
291.5	GEH < 5 for > 85% of *k*	930	870	93.5%	Yes
291.9	GEH < 5 for > 85% of *k*	930	852	91.6%	Yes
292.4	GEH < 5 for > 85% of *k*	930	875	94.1%	Yes
292.8	GEH < 5 for > 85% of *k*	930	887	95.4%	Yes
294.0	GEH < 5 for > 85% of *k*	930	889	95.6%	Yes
294.7	GEH < 5 for > 85% of *k*	930	915	98.4%	Yes
295.3	GEH < 5 for > 85% of *k*	930	909	97.7%	Yes
296.0	GEH < 5 for > 85% of *k*	930	850	91.4%	Yes
Overall	GEH < 5 for > 85% of *k*	13020	12542	96.3%	Yes

Figure D.5 summarizes the simulated vs. observed queue propagation. The contour plots show the 5-minute average speeds at the detectors throughout the selected peak period. The simulation reproduced the field observed peak duration and the length of queue fairly accurately, with the exception of the most upstream bottleneck at Sheldon Rd., which showed slightly shorter simulated duration and slightly less simulated queue propagation. Despite these minor discrepancies, the overall congestion pattern and the presence of severe speed reduction were accurately replicated.



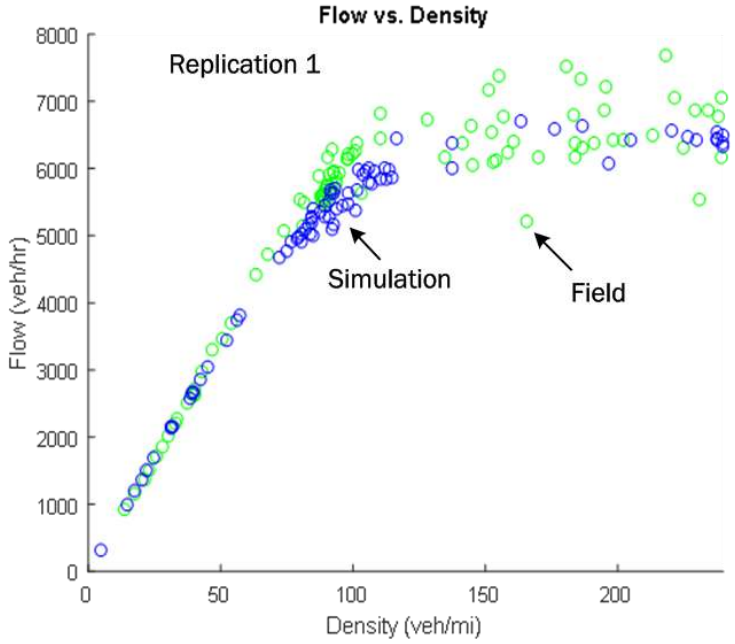
(a)



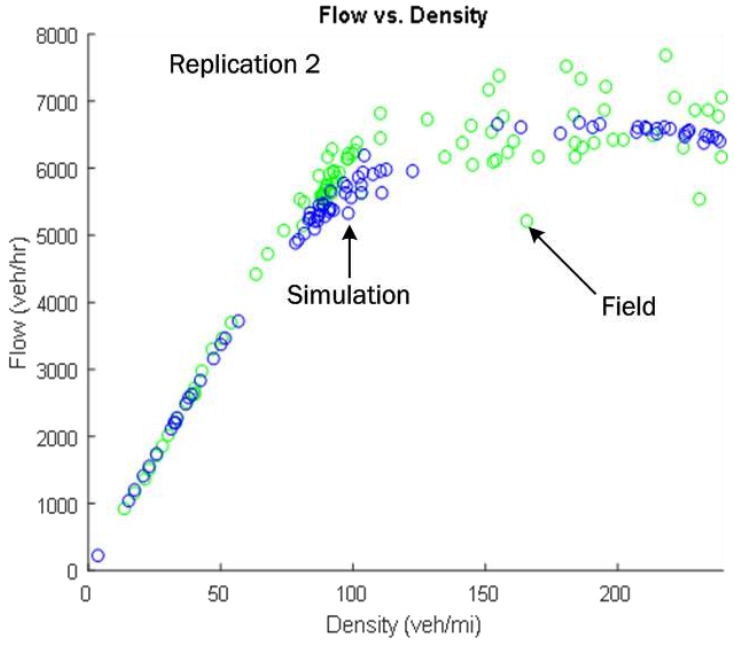
(b)

Figure D.5 Comparison of contour plots of (a) field observed and (b) simulated speeds.

Figure D.6 shows the field observed and simulated flow-density relationships of a four-lane mainline section at the most important bottleneck, the Florin Rd. on-ramps at mile 294. In both replications, the simulated data provided near-perfect match in the uncongested state, as well as good representation of the congested state. The fundamental diagram shows that the simulated maximum capacity is slightly lower than the observed maximum capacity, otherwise, the simulation provided good representation of the overall macroscopic traffic behavior.



(a)



(b)

Figure D.6 Comparison of (a) field observed and (b) simulated flow vs. density relationship at Florin Rd. bottleneck (mile 294).

D.5 References

Caltrans PeMS. http://pems.dot.ca.gov/. Accessed on October 24, 2016.

Ciuffo, B., et al, 2012. Thirty Years of Gipps' Car-Following Model: Applications, Developments, and New Features. Transportation Research Record 2315, 88-99.

Dowling, R., A. Skabardonis, and V. Alexiadis, 2004. Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software. Federal Highway Administration FHWA-HRT-04-038.

Gipps, P. G., 1986. A Model for the Structure of Lane-Changing Decisions. Transportation Research Part B: Methodological 20-B (5), 403-414.

Lu, X., Kan, X., Shladover, S.E., Wei, D., Ferlis, R.A., 2017. An enhanced microscopic traffic simulation model for application to connected automated vehicles. Transportation Research Board 96th Annual Meeting, Washington, DC.

Newell, G. F., 2002. A Simplified Car Following Theory: A Lower Order Model. Transportation Research Part B: Methodological 36, 195–205.

TSS|Aimsun. http://www.aimsun.com/. Accessed on January 5, 2017.

Yeo, H., et al., 2008. Oversaturated Freeway Flow Algorithm for Use in Next Generation Simulation. Transportation Research Record 2088, 68-79.

WSDOT, 2002. Freeway System Operational Assessment - Paramics Calibration and Validation Guidelines (Draft). Technical Report I-33, Wisconsin DOT, District 2.