MODELING TRAFFIC FLOW ON OVERSATURATED ARTERIALS

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

XINKAI WU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

FACULTY ADVISOR: HENRY X. LIU

OCTOBER 2010

© XINKAI WU 2010

ALL RIGHTS RESERVED

To my dear grandmother

Acknowledgements

I would like to express the deepest appreciation to my advisor, Dr. Henry X. Liu. His infectious enthusiasm has encouraged me to carry out successful research. Without his continuous guidance and support, this dissertation would not have been possible. I am also greatly thankful to Dr. Panos G. Michalopoulos. He offered me the opportunity to join the transportation program at the university and introduced me to traffic flow theory.

Many thanks go to Dr. David M. Levinson. I had wonderful experiences of working with him. His energy, ambition, and sense of novel research ideas have inspired me. I also want to thank Dr. Nikolas Geroliminis. As a good friend, I admire his intelligence and thank him for sharing his successful experiences with me. I also appreciate the help from Dr. Gary A. Davis. I have benefited much from his profound knowledge in statistics, and I respect his intelligence and integrity. Finally, a special thanks goes to Dr. Shashi Shekhar for his service on my committee. His insightful opinions have improved this dissertation.

This research could not have been done without the financial support from the Minnesota Department of Transportation and the National Cooperative Highway Research Program. Specially, I would like to acknowledge Mr. Steve Misgen and Mr. Ron Christopherson from the Minnesota Department of Transportation.

I would also like to thank my fellow colleagues and friends, especially Mr. Heng Hu, Dr. Guizhen Yu, and Mr. Sundeep Bhimireddy for their help in data collection and algorithm programming. I also much appreciate Mr. Hui Xiong and Mr. Xiaozheng He for their help in my life.

Last but not least, my wholehearted gratitude goes to my family, especially to Ms. Fengqin Lian, the lovely young lady I am willing to spend my life with and who is willing to do the same with me.

Abstract

Traffic congestion is a national issue in the United States and has gotten worse in regions of all sizes. Now, more and more intersections are operated in oversaturated situations where the traffic demand exceeds the capacity of the system. Although a significant amount of literature has been devoted to how to manage oversaturated traffic signal systems, our understanding of the characteristics of oversaturation remains limited, particularly with regard to identification of oversaturation and the transition process from under-saturated condition to oversaturation. It has become increasingly obvious that successful traffic management requires efficient methods to identify and model oversaturated conditions.

This research moves towards a better understanding of oversaturation, by 1) providing coherent methodologies to quantify oversaturation and 2) developing a simplified model to describe oversaturation at signalized intersections based on high-resolution traffic signal data collected by the SMART-SIGNAL (Systematic Monitoring of Arterial Road Traffic Signals) system. In particular, the research focuses on the following four areas:

1) Quantification of oversaturation: Traditional definitions of oversaturation are not applicable for existing detection systems. This research circumvents this issue by quantifying the detrimental effects of oversaturation on signal operations, both temporally and spatially. In the temporal dimension, the detrimental effect is characterized by a residual queue at the end of a cycle, which occupies a portion of green time in the next cycle. In the spatial dimension, the detrimental effect is characterized by a downstream spillover, which blocks the traffic and reduces usable green time. From these observations, we derive two types of an oversaturation severity index (OSI): one temporally-based (T-OSI) and one spatially-based (S-OSI). Both T-OSI and S-OSI are designed to yield a ratio between the unusable green time due to detrimental effects and the total available green time in a cycle, using high resolution traffic signal data. T-OSI is quantified by estimating the residual queue length; and S-OSI is quantified by measuring

the time period of spillover. Since different types of OSI (T-OSI or S-OSI) point to different underlying causes of oversaturation, this research has the potential to provide guidance for the mitigation strategies of signal oversaturation.

2) Real-time queue length estimation for congested intersections: To quantify T-OSI, this research proposes a novel shockwave-based algorithm to estimate time-dependent queue length even when the signal links are congested with long queues, a situation that the traditional input-output approach for queue length estimation cannot handle. Using high-resolution "event-based" traffic signal data, the new algorithm first identifies traffic state changes; and then applies Lighthill-Whitham-Richards (LWR) shockwave theory to estimate maximum and minimum (*i.e.* residual) queue length. This algorithm is also applicable for other aspects of arterial performance such as travel time, delay, and level of service.

3) Queue-Over-Detector (QOD): To quantify S-OSI, we study a phenomenon we call Queue-Over-Detector (QOD). QOD occurs when a vehicle stops and rests on a detector for a period of time creating a large occupancy value. This research demonstrates that a main cause of QOD is spillover from downstream intersections. Thus QOD identification can be used to quantify oversaturation in the spatial dimension, *i.e.* S-OSI. This research also briefly studies the relationship between QOD and the cycle-based arterial fundamental diagram (AFD) by microscopically investigating individual vehicle trajectories derived from event-based data. Results show that proper treatment of QOD results in a stable form of the AFD which clearly identifies three different regimes, under-saturation, saturation, and over-saturation with queue spillovers. Achieving a stable form of the AFD is of great importance for traffic signal control because of its ability to identify traffic states on a signal link.

4). Traffic flow modeling for oversaturated arterials: The culmination of this research project is a simplified traffic flow model for congested arterial networks, which we call the shockwave profile model (SPM). Unlike conventional macroscopic models, in which space is often discretized into small cells for numerical solution, SPM treats each

homogeneous road segment with constant capacity as a section; then categorizes the traffic within each section simply as free-flow, saturated, or jammed. Traffic dynamics are analytically described by tracing the shockwave fronts which explicitly separate these three traffic states. SPM is particularly suitable for simulating traffic flow on congested signalized arterials, especially with queue spillover problems. In SPM, queue spillover can be treated as either extending a red light or creating new smaller cycles. Since only the essential features of arterial traffic flow, *i.e.*, queue build-up and dissipation, are considered, SPM significantly simplifies arterial network design and improves numerical efficiency. For these reasons, we fully expect this model to be adopted in real-time applications such as arterial performance prediction and signal optimization.

Table of Contents

List of Tables

List of Figures

Acronyms

AFD:	Arterial Fundamental Diagram
CTM:	Cell Transmission Model
LWR:	Lighthill-Whitham-Richards
MAPE:	Mean Absolute Percentage Error
MPE:	Mean Percentage Error
OSI:	Oversaturation Severity Index
QOD:	Queue-Over-Detector
QOD-I:	The first type of QOD: caused by red phases
QOD-II:	The second type of QOD: caused by a spillover
S-OSI:	OSI in the spatial dimension
SMART-SIGNAL:	Systematic Monitoring of Arterial Road Traffic Signals
SPM:	Shockwave Profile Model
T-OSI:	OSI in the temporal dimension

Notation

d_{jam}:	Space headway at jammed traffic conditions
$g_n(t)$:	Effective green interval in the current cycle for the through movement at intersection n when time t belongs to the current cycle
h_s:	Saturation discharge time headway
i:	Index of cycle i
k:	Density
k_a^i:	Average density of arrival flow during the i^{th} cycle
k_s^n:	Saturation density (when traffic flow is maximum) at link L_n
k_{jam}^n:	Jam density at link L_n
$l_n^{w_j}(t)$:	Distance from the front of shockwave w_j ($j = 1, 2, 3$, or 4) to the stop line at intersection n at time t
$l_n^{\hat{w}}(t)$:	Queue length, defined as the distance from the front of shockwave w_1 or w_3 to the stop line at intersection n at time t
n:	Index of Intersection n
q:	Flow
q_a^i:	Average arrival flow rate during the i^{th} cycle
q_s^n:	Maximum traffic flow rate at link L_n
$q_{n-1}(t)$:	Arrival flow rate at the entrance of link L_n at time t

$\bar{q}_n(t)$:	Departure flow rate at intersection n at time t for the major approaches
$\bar{q}_n^m(t)$:	Departure flow rate at intersection n at time t for the minor approaches
$q_n^{V1}(t)$:	Arrival flow rate at the entrance of section L_n^{V1}
$q_n^{V2}(t)$:	Arrival flow rate at the entrance of section L_n^{V2}
$r_n(t)$:	Effective red interval in the current cycle for the through movement at intersection n when time t belongs to the current cycle
t:	Time t
v:	Speed
v_f:	Free-flow speed
w:	The velocity of a shockwave
w_1:	Queuing wave
w_2:	Discharge wave
w_3:	Departure wave
w_4:	Compression wave
w^*:	The absolute value of w_2 or w_4
β_n:	Turning percentage for the through movement at link L_n
Δt	A small time interval

D_e:	Effective vehicle length
L_d:	Distance between advance detector and stop-bar
$L_{max}^{\ i}$:	Maximum queue length during the i^{th} cycle
$L_{min}^{\ i}$:	Minimum queue length during the i^{th} cycle
L_n:	Link L_n (with length L_n)
$L_n^{\ A}$:	Turning bay at link L_n
$L_n^{\ B}$:	Downstream through section at link L_n (starting from separation point U to link end)
$L_n^{\ U}$:	Upstream through section at link L_n (starting from link entrance to separation point U)
$L_n^{\ V1}$:	Blocked lane(s) in section $L_n^{\ U}$
$L_n^{\ V2}$:	Unblocked lane(s) in section $L_n^{\ U}$
$O_n(t)$:	Start time of the effective red (or the end of effective green) in the current cycle for the through movement at link L_n
T:	End time of simulation or the whole simulation time period
$T_g^{\ i}$:	End time of the effective green during the i^{th} cycle
$T_{end}^{\ QOD}$:	Time instant when QOD ends
$T_{max}^{\ i}$:	Time instant when the approach has the maximum queue length during the i^{th} cycle
$T_{min}^{\ i}$:	Time instant when the approach has the minimum queue length during the i^{th} cycle

T_r^i: End time of the effective red during the i^{th} cycle

T_{start}^{QOD}: Time instant when QOD starts

T_A: Time instant when break point A happens

T_B: Time instant when break point B happens

T_C: Time instant when break point B happens

1 Introduction

1.1 Problem Statement

Traffic congestion is a national issue in the United States and has gotten worse in regions of all sizes (Texas Transportation Institute, 2009). Now, more and more urban and suburban intersections are operated in congested conditions and situations where the traffic demand exceeds the capacity of the system. Under this condition of oversaturation, typical traffic control strategies do not work as efficiently as necessary. As indicated by the results of the 2007 Traffic Signal Operation Self Assessment surveys (Institute of Transportation Engineers, 2007), the majority of agencies involved in the operation and maintenance of traffic signal systems are already stretched thin and challenged to provide adequate service to drivers in their jurisdictions. Oversaturated conditions present an additional burden for practitioners that lack adequate tools for addressing such situations.

Although a significant amount of literature has been devoted to how to manage oversaturated traffic signal systems, our understanding of the characteristics of oversaturation, particularly identification of oversaturation and the transition process from under-saturated condition to oversaturation, is limited. As a result, there is a critical need for new methods to identify and model oversaturation.

This research moves towards a better understanding of oversaturation by providing coherent methodologies to quantify and model oversaturation at signalized intersections. Two major themes are the focus of this dissertation:

1). Quantification of oversaturation. Previous research studies using traffic data from signal systems to diagnose and identify oversaturation are mostly qualitative and incomplete. Conceptual definitions of oversaturation, which define an oversaturated traffic intersection as one where traffic demand exceeds the capacity, cannot be applied directly to identify oversaturated intersections because traffic demand under congested

conditions is not measurable, particularly with fixed-location sensors. Furthermore, successful mitigation strategies require the ability both to detect the onset of oversaturation and to quantify its severity. Thus it is imperative to have an implementable and quantifiable metric of oversaturation as well as a coherent methodology to identify such situation.

2). Modeling of oversaturation. Compared to the continuous and intensive research efforts on modeling of freeway traffic, there have been far fewer reported efforts to study arterial traffic flow. Existing traffic flow models are designed to model continuum freeway traffic, and thus have deficiencies when applied to congested arterials where traffic is interrupted by traffic lights. This is particularly true for large arterial networks during oversaturated conditions. Since traffic flow modeling serves as the foundation for all signal optimization schemes, there is a need for a quick and approximate approach, with sufficient descriptive power, to model oversaturated traffic flow dynamics in a signalized network for real-time applications.

The research to address these issues is built upon the high-resolution traffic signal data collected and archived by the SMART-SIGNAL (Systematic Monitoring of Arterial Road Traffic Signals) system developed by the University of Minnesota (Liu & Ma, 2009; Liu *et al.*, 2010; Wu *et al.*, 2010a; Wu *et al.*, 2010b; Wu *et al.*, 2010c). Without the real-time data collected by this system, our project would not have been possible.

1.2 Research Contributions

This thesis makes contributions in the areas of arterial performance measurement, arterial traffic flow theory, and arterial signal control. Specifically:

1) It proposes a systematic approach to quantify oversaturation. Traditional definitions of oversaturation are not applicable for existing detection systems. This research circumvents this issue by quantifying the detrimental effects of oversaturation

on signal operations, both temporally and spatially. The detrimental effect is characterized either temporally by a residual queue at the end of a cycle, which occupies a portion of green time in the next cycle; or spatially by the spillover from downstream traffic which reduces usable green time due to the downstream blockage. An oversaturation severity index (OSI), based either on the temporal dimension (T-OSI) or spatial dimension (S-OSI), can then be calculated as the ratio between the unusable green time due to detrimental effects and the total available green time in a cycle. T-OSI is quantified by estimating the residual queue length; and S-OSI is quantified by measuring the time period of spillover. Since different types of OSI (T-OSI or S-OSI) point to different underlying causes of oversaturation, this research is expected to provide guidance for strategies to mitigate signal oversaturation.

2) It develops a shockwave-based intersection queue length estimation algorithm. The traditional input-output approach for queue length estimation can only handle queues that are shorter than the distance between the vehicle detector and intersection stop line, because cumulative vehicle counts for arrival traffic are not available once the detector is occupied by the queue. To estimate queue length under congested conditions, a novel shockwave-based queue estimation algorithm is developed in this research. Using high-resolution "event-based" traffic signal data, the new algorithm first identifies traffic state changes that distinguish queue discharge flow from upstream arrival traffic; it then applies Lighthill-Whitham-Richards (LWR) shockwave theory to estimate maximum and minimum (*i.e.* residual) queue length. Since this approach can estimate time-dependent queue length even when signal links are congested with long queues, we use it to quantify T-OSI. This algorithm can also be used to measure other aspects of arterial performance such as travel time, delay, and level of service.

3) It identifies the Queue-Over-Detector (QOD) phenomenon and its significance. This work is the first to study the QOD phenomenon. We demonstrate the relationship between QOD and spillover, and then propose an algorithm to quantify S-OSI by identifying QOD. This research also studies the impact of QOD on the cycle-based arterial fundamental diagram (AFD) by microscopically investigating individual vehicle

trajectories derived from event-based data; analysis reveals that proper treatment of QOD yields a stable form of AFD which clearly identifies three distinct regimes: under-saturation, saturation, and over-saturation with queue spillovers. Having a stable form of AFD is of great importance for traffic signal control because of its ability to identify traffic states on a signal link.

4) It builds a shockwave profile traffic flow model for congested arterial networks. The culmination of this work is a new traffic flow model for congested arterial networks, the shockwave profile model (SPM). Unlike conventional macroscopic models, in which space is often discretized into small cells for numerical solution, SPM treats each homogeneous road segment with constant capacity as a section, and then categorizes traffic in a section into three distinct states: free-flow, saturated, and jammed. Traffic dynamics are analytically described by tracing the shockwave fronts which explicitly separate the three traffic states. SPM is particularly suitable for simulating traffic flow on congested signalized arterials, especially those with queue spillover problems. In SPM, queue spillover is treated as either extending a red light or creating new smaller cycles. Since it only considers the essential features of arterial traffic flow, *i.e.*, queue build-up and dissipation, SPM greatly simplifies the process of network design and improves numerical efficiency. For these reasons, adoption of this model can reasonably be expected in a number of real-time applications such as arterial performance prediction and signal optimization.

1.3 Dissertation Organization

The rest of the dissertation is organized as follows: Chapter 2 briefly reviews the SMART-SIGNAL System. Chapter 3 introduces a quantifiable measurement of oversaturation - oversaturation severity index (OSI). In Chapter 4, a novel algorithm is proposed for queue length estimation at congested intersections. The algorithms to quantify OSI in both temporal and spatial dimensions are described in Chapter 5. Chapter 6 proposes our shockwave profile traffic flow model for oversaturated arterials. Chapter 7 concludes this dissertation and lays out the direction of future research.

2 Background – SMART-SIGNAL

The SMART-SIGNAL (Systematic Monitoring of Arterial Road Traffic and SIGNAL) system, developed by the University of Minnesota, offers an easily implementable approach to collect and archive continuous and high-resolution traffic data on signalized arterials (Liu *et al.*, 2008). In this system, a complete history of traffic signal control, including all vehicle actuation events and signal phase change events, is archived and stored. Based on the event-based data, arterial performance measures, such as arterial travel time, intersection queue length, and level of service, are produced. The SMART-SIGNAL system has been installed on 11 intersections along France Avenue in Hennepin County, Minnesota, and 6 intersections along Trunk Highway 55, Minnesota, since February 2007. Event-based traffic data are being collected in a 24/7 mode and then archived in a database system, thus yielding a tremendous amount of field data available for research. This chapter will briefly introduce the system architecture, data collection hardware, and data processing procedure of the SMART-SIGNAL.

2.1 System Architecture

As illustrated in Figure 2-1, the SMART-SIGNAL system has three major components, data collection, performance measurements, and performance presentation through user interfaces. The data collection component collects high resolution raw data directly from the field on an event-by-event basis. Signal phase change events and vehicle-detector actuation events are acquired separately from data collection units located in traffic signal cabinets. The data are recorded in daily log files and sent to a data server at the master cabinet by serial port communication. The daily log files are finally transmitted to a database located at the Minnesota Transportation Observatory (MTO) lab at the University of Minnesota through the Digital Subscriber Line (DSL) or wireless communication. The second component of the SMART-SIGNAL system is performance measure calculation using the field-collected data. Analysis of the stored data log files

yields a set of performance measures, one covers intersection level measures (*e.g.* queue length) and a second related to arterial level measures (*e.g.* travel time). Once all performance measures have been derived from the raw data, the results are made accessible to a variety of users.

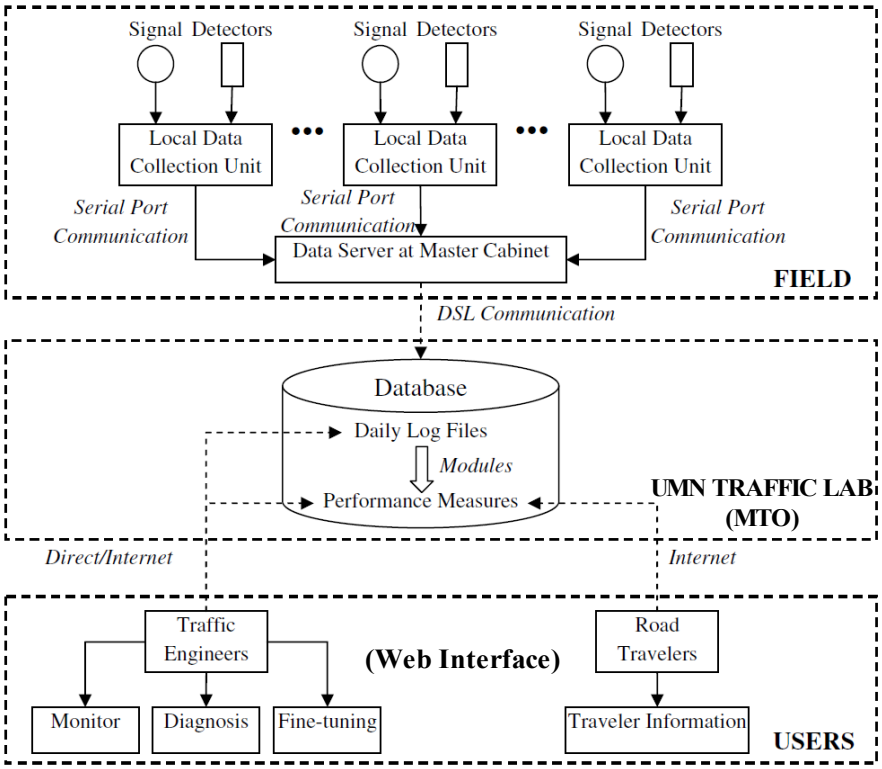

Figure 2-1 SMART-SIGNAL System Architecture (Source: Liu & Ma, 2009)

2.2 Data Collection Unit

The key element of the SMART-SIGNAL is the data collection unit, which consists of an industrial PC and a data acquisition card. At each intersection, an industrial PC with a data acquisition card is installed, and event data collected at each intersection is transmitted to the data server in the master controller cabinet through the existing communication line (in this case, spare twisted pair) between signalized intersections. The data acquisition cards (PCI-6511 from National Instruments (2006)) used in the

SMART-SIGNAL system, as shown in Figure 2-2a, have 64 input channels. If the total number of detector inputs and signal phases for one intersection exceed 64, an additional data acquisition card needs to be installed. A terminal box is used in order to limit the input direct current (DC) to a safe range and establish the connection between the data acquisition card and back panel of the traffic cabinet, as shown in Figure 2-2b. The terminal box allows digital voltage changes on the back panel, which indicate different traffic events in the field, to be captured by the data acquisition card installed in the industrial computer. A traffic event recorder software program, developed using the Microsoft Visual C# program, runs on the industrial computer in the field to record the events (for example, a phase 1 green change from "ON" to "OFF") into a log file.

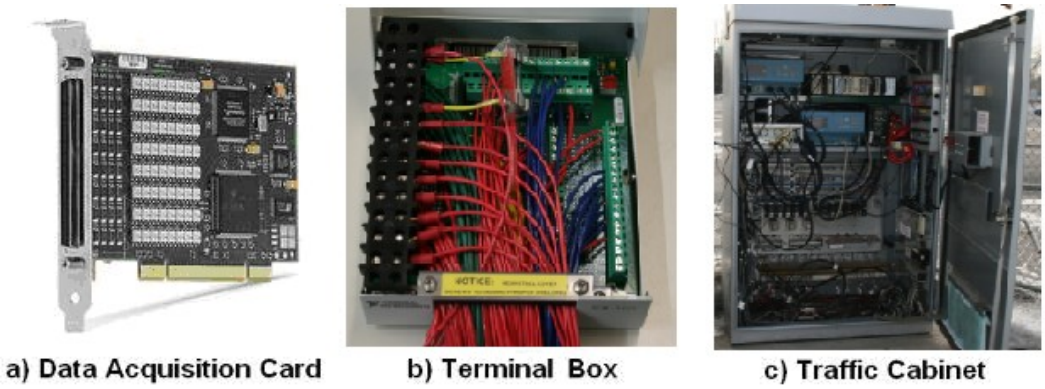

Figure 2-2 Demonstration of the Traffic Data Collection Components

Data communication between two controller cabinets is done using the existing twisted pair communication lines. A protocol of RS-485 is used to transmit data and synchronize time between cabinets (B&B Electronics, 2007). After the data in the local cabinets is transferred to the master cabinet, DSL or a wireless unit installed in the master cabinet is used to send the data back to the database located in the MTO.

A sample of data is shown in Figure 2-3. Each logged event starts with a time stamp that includes the date, hour, minute, second and millisecond based on the computer system time, followed by different types of event data including phase changes, detector actuation and pedestrian calls. A complete history of traffic signal events is thus recorded.

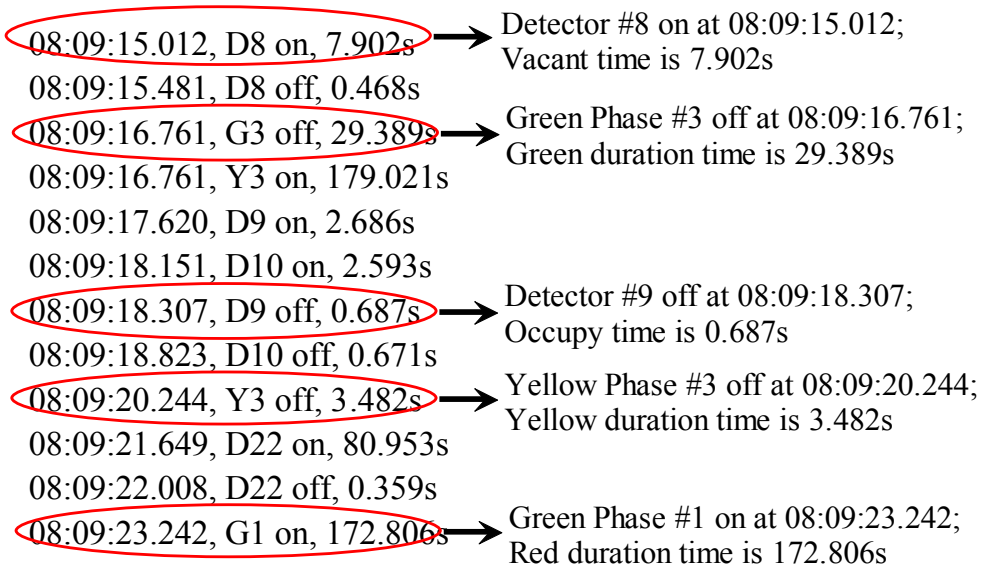

Figure 2-3 Sample Data (Source: Liu *et al*., 2010)

2.3 Data Processing Procedure

The raw data collected from the field needs to be preprocessed and converted to an easy-read format before the performance measures can be derived. Based on event data, the signal phase duration can be calculated from the time difference between the start and end of a signal event. The time interval between the start and end of a vehicle actuation event is the detector occupancy time, and the time interval between the end of a vehicle actuation event and the start of a following vehicle actuation event (from the same detector) is the time gap between two consecutive vehicles crossing the detector. Further processing can be done to determine second-by-second volume, occupancy, and cycle-by-cycle signal timing plan. The processed data can now be used to generate performance measures, such as queue length and average vehicle delay for intersections and travel time for arterial links. Figure 2-4 demonstrates the data flow of the SMART-SIGNAL system.

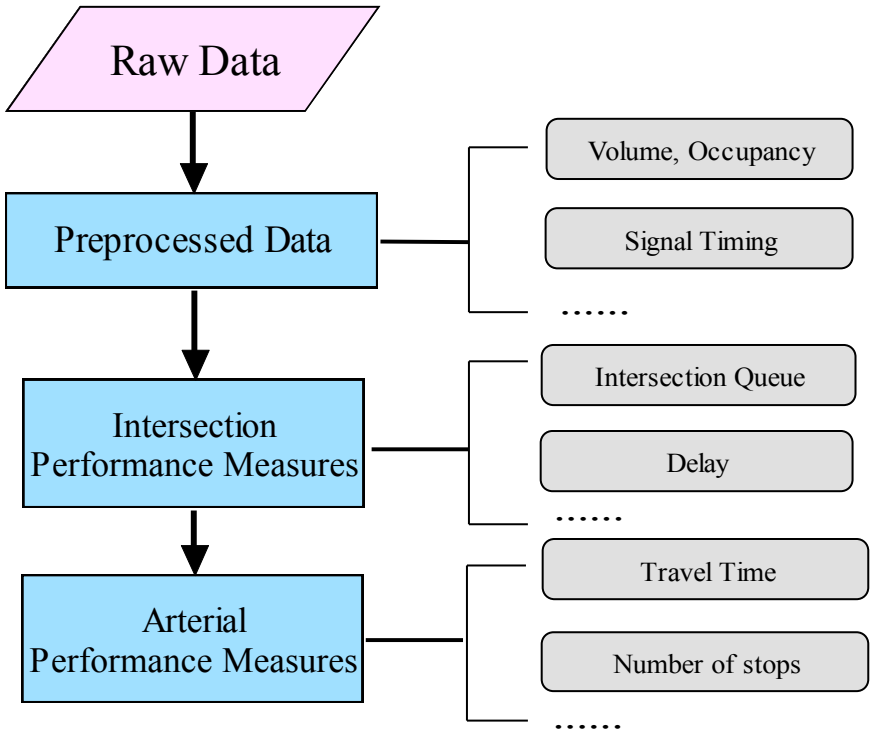

Figure 2-4 Data Flow of SMART-SIGNAL System (Source: Liu *et al.*, 2008)

Although many existing signal control systems are capable of generating data to support performance assessment, most do not make it "easy" for the managing agencies to prioritize improvements and plan for future needs. The SMART-SIGNAL System fills this gap. The high-resolution event data collected by the system is extremely valuable. As will be explained in the following chapters, the data is used to estimate real-time intersection queue length, identify queue spillover, and quantify the severity of oversaturation.

3 A QUANTIFIABLE MEASURE OF OVERSATURATION

3.1 The Conventional Definition of Oversaturation

Conventionally, an oversaturated intersection is defined as one where the demand exceeds the capacity (Gazis, 1964). The degree of saturation, *i.e*. the volume/capacity ratio, is defined as:

$$X_j = \frac{VOL_j}{CAP_j} \tag{3.1}$$

where X_j is the degree of saturation for lane group j and VOL_j and CAP_j are the demand flow rate and capacity for lane group j, respectively. A lane group is oversaturated when $X_j > 1$.

For a single intersection with two competing streams, Gazis (1964) expanded this concept by proposing the following inequality:

$$q_1/S_1 + q_2/S_2 > 1 - (LT/C) \tag{3.2}$$

where q_1 and q_2 are the arrival rates for two conflicting directions, S_1 and S_2 are the saturation flow rates for the two directions, LT is the total lost time, and C is the cycle length.

Direct application of these two definitions to detect the onset of oversaturation and quantify its duration and extent is difficult, however. This is partly due to the uncertainty of the capacity and saturation flow, and partly due to the difficulty of measuring the arrival flow using current data collection systems during oversaturated situations (the very condition that we are trying to identify). Most existing detection systems,

particularly those with inductance loop detectors, provide observations of traffic flows at a fixed point on a link when they are not fully occupied. Traffic demand is simply not measurable when a fixed-location detector is occupied with a vehicular queue.

Alternatively, some researchers have characterized oversaturation as "a stopped queue that cannot be completely dissipated during a green cycle" (Gazis, 1964), or as "traffic queues that persist from cycle to cycle either due to insufficient green splits or because of blockage" (Abu-Lebdeh & Benekohal, 2003). However, these authors provided no methodology for measuring queue length or identifying the existence of such situations.

Another measure that can potentially be used to detect oversaturation is green phase utilization. If a signal phase is oversaturated with a long queue, once the green light starts, vehicles will continue to discharge at the saturation flow rate until the end of the effective green. The ratio between the "used" green time and the green phase duration can be treated as an indicator of saturation for a particular approach. This concept has been incorporated into several adaptive traffic control systems, for example, SCATS (Sydney Coordinated Adaptive Traffic System) (Sims & Dobinson, 1980) and ACS-Lite (Adaptive Control Software Lite) (Luyanda, *et al.*, 2003; Gettman, *et al*., 2007). This method, however, cannot estimate the degree to which a certain traffic phase is oversaturated. The indicator simply identifies that green time is insufficient to serve the traffic demand, but the severity of oversaturation is unknown. In addition, it is necessary to point out that high green utilization of a given phase is not necessarily an indicator of oversaturation. In some cases, well-coordinated signalized intersections can also generate high values of green utilization.

To the best of our knowledge, previous research studies using traffic data from signal systems to diagnose and identify oversaturation are mostly qualitative and incomplete. Conceptual definitions discussed above are either not applicable in the real world or have other deficiencies. Since detection of the onset of oversaturation as well as quantifying the severity of oversaturation is a critical step before appropriate mitigation strategies can be applied, it is imperative to have an implementable and quantifiable measure of

oversaturation as well as a coherent methodology to identify such situations. We believe our research begins to fill this gap.

3.2 A Quantifiable Measure of Oversaturation

Since the general definition of oversaturation, *i.e.* traffic demand exceeding the capacity of a facility, cannot be applied directly to detect the occurrence of oversaturation, we propose a measure of oversaturation by quantifying its *detrimental effects*. The detrimental effects of oversaturation are any effects which lead to the reduction of usable green time in a cycle for a signalized approach. These effects can be described from either a temporal or a spatial perspective.

A detrimental effect of oversaturation in the temporal dimension is characterized by a residual queue at the end of a cycle. The residual vehicles, which are part of a discharging platoon and supposed to pass through the intersection in the current cycle, cannot be discharged due to insufficient green splits, thereby creating detrimental effects on the following cycle by occupying a portion of green time. The portion of green time utilized by the residual queue then becomes unusable for the traffic arrivals during that cycle.

A detrimental effect of oversaturation in the spatial dimension can be characterized by a spillover from downstream traffic. When spillover happens, a downstream link of the intersection is blocked and vehicles cannot be discharged from the intersection even in a green phase. Here again, a portion of the green time becomes unusable with detrimental effects. While spillover is common when vehicular queues fully occupy a downstream link, oversaturated conditions occur only if spillover from a downstream intersection leads to unusable green time.

Therefore, in our approach, the condition of oversaturation is characterized 1) by a residual queue at the end of a cycle that creates detrimental effects on the following cycle, and/or 2) by a downstream spillover within a cycle that creates detrimental effects on upstream traffic facilities. To quantify the detrimental effects in either temporal or

spatial dimensions, we introduce the ***oversaturation severity index (OSI)*** by using the ratio between unusable green time and total available green time in a cycle. OSI will be a non-negative percentage value between 0 and 100, with 0 indicating no detrimental effect for signal operation, and 100 denoting the worst case where all available green time becomes unusable.

We further differentiate OSI into ***T-OSI*** and ***S-OSI***. T-OSI describes the detrimental effects created by residual queue, *i.e.* the detrimental effect in the temporal dimension; and S-OSI describes the detrimental effects caused by spillover, *i.e.* the detrimental effect in the spatial dimension. Although both T-OSI and S-OSI can be calculated using the ratio between unusable green time and total available green time, the meanings of "unusable" are different. For T-OSI, the "unusable" green time is the equivalent green time to discharge the residual queue in the following cycle, in which case vehicles are discharged at saturation flow rate during that time period. By contrast, for S-OSI, the "unusable" green time is the time period during which a downstream link is blocked, in which case the discharge rate is zero.

Since T-OSI quantifies the detrimental effect of oversaturation on the following cycle, the persistence, duration, and frequency of T-OSI greater than zero becomes an important indicator of traffic congestion at the intersection level. On the other hand, S-OSI describes the detrimental effect of oversaturation caused by downstream queue spillover, indicating the spatial extent that traffic congestion has spread network-wide.

The differentiation of T-OSI and S-OSI may also help to identify the causal relationship of arterial traffic congestion. For example, for an intersection, a positive T-OSI indicates that the available green time is insufficient for queue discharge and a residual queue is generated at the end of a cycle. Subsequently, in the following cycles, the queue may grow and spill over to an upstream intersection, resulting in an S-OSI at the upstream intersection that is positive. Clearly in this case a positive S-OSI at the upstream intersection is caused by the downstream bottleneck. Note that the downstream bottleneck may lead to a situation in which both the S-OSI and T-OSI are greater than

zero, simply because a portion of the green time is wasted due to downstream blockage (*i.e.* S-OSI> 0), and a residual queue may be generated (*i.e.* T-OSI> 0) due to the reduction of green time. Here positive T-OSI and S-OSI indicate that traffic congestion may start to spread even further upstream.

With the proposed oversaturation severity index, the focus of the identification algorithms shifts from *measuring travel demand* to *quantifying detrimental effects* in both temporal and spatial dimensions. It is necessary to point out that the ability to classify and quantify the detrimental effects created by oversaturation is very important for traffic management, because different oversaturated situations may call for different strategies to mitigate congestion. For example, for an isolated intersection with positive T-OSI values, extension of green time to discharge residual queue may be sufficient (Quinn, 1992). By contrast for an arterial corridor with multiple overloaded intersections, simultaneous or even negative offsets (see Pignataro, *et al.* 1978) may be needed to prevent further deterioration of the oversaturated condition. Since the focus of this study is to identify and quantify oversaturated conditions, how to map OSI with different mitigation strategies is left for future research.

In the following two chapters, we propose two algorithms for the identification and quantification of oversaturated conditions, one that estimates residual queue length, and a second that detects spillover conditions. Our focus here is not only to identify oversaturation qualitatively, but also to quantify its severity.

4 Real-Time Queue Length Estimation for Congested Intersections

It has long been recognized that vehicular queue length is crucial both for measuring signal performance in terms of vehicle delay and stops (Webster & Cobbe, 1966; Cronje, 1983a and 1983b; Balke *et al.*, 2005), and for signal optimization (Webster, 1958; Gazis, 1964; Newell, 1965; Green, 1968; Michalopoulos & Stephanopolos, 1977a, b; Chang & Lin, 2000; Mirchandani & Zou, 2007). Over the years, many researchers have dedicated themselves to this topic, and two types of queue estimation models have been developed. The first one, which is based on the analysis of cumulative traffic input-output to a signal link, was proposed by Webster (1958) and later improved by a number of researchers (Newell, 1965; Robertson, 1969; Gazis, 1974; May, 1965; Catling, 1977; Akcelik, 1999; Strong, *et al.*, 2006; Sharma, *et al.*, 2007; Vigos *et al.*, 2008). This type of model is commonly used to describe traffic queuing processes, but, as noted by Michalopoulos & Stephanopolos (1981), "it is insufficient for obtaining the spatial distribution of queue lengths in time." In addition, cumulative input-output techniques can only be used for the estimation of queue length when the rear of the queue does not extend beyond the vehicle detector. Since the techniques can only handle relatively short queues, applications of the approach are limited.

The second model is based on traffic flow theory, which was first demonstrated by Lighthill and Whitham (1955) and Richards (1956) for uninterrupted flow; and later expanded by Stephanopolos & Michalopoulos (1979 & 1981) to signalized intersections. Essentially, this model estimates queue lengths by tracing the trajectory of shockwaves based on continuum traffic flow theory. Although shockwave theory-based models can successfully describe the complex queuing process in both temporal and spatial dimensions, these elegant theoretical models have had limited practical applications as

they require "perfect" input information. Specifically, these models assume known vehicle arrivals. This requirement is often impossible to satisfy, however, because vehicle arrivals cannot be measured once a detector is fully occupied, which is usually the case with congested arterials. Without arrival information, existing shockwave models cannot be used to estimate intersection queue lengths.

To estimate queue lengths longer than the distance from an intersection stop bar to an advance detector, Muck (2002) used a linear regression model that describes the relationship between queue size and "fill-up" time, *i.e.,* the time that "passes from the beginning of the red time of a signal until continuous occupancy of a detector." The evaluation results showed that this method can estimate queues up to 5-10 times further upstream from the actual detector location. However, Muck's heuristic approach is based on the premise that the arrival rate of traffic flow is constant within a cycle. This means that the measured fill-up time can be proportionally inversed to the queue build-up. Such an approach could be problematic when the arrival rate of traffic flow fluctuates greatly, which is not unusual with the "gating effect" of an upstream traffic signal.

More recently, Skabardonis and Geroliminis (2008) proposed a method to estimate intersection queue length using aggregated loop detector data in 30-second intervals. Their method for determining queue rear ends entails examining the data for flow and occupancy changes. The problem is that 30-second aggregation smoothes out variations for between-vehicle gaps, making it difficult to identify a queue rear unless the arrival traffic flow is significantly different with the queue discharge flow. In addition, since queue length estimation is only part of their travel time estimation model, no evaluation results are reported in their paper to verify the accuracy of their queue length estimation method.

The research presented here provides a method for estimating long queues (meaning queues "longer than the distance between an intersection stop line and the vehicle detectors"). We are particularly interested in the estimation of maximum queue length, *i.e.* the distance from the intersection stop line to the position of the last vehicle that has

to stop in a cycle. The proposed methodology relies on high resolution traffic signal data, which has become increasingly available in recent years. For example, second-by-second detector data has been used by ACS-Lite (**A**daptive **C**ontrol **S**oftware **Lite**) (Luyanda, *et al.*, 2003).

Even more significant for our work is the availability of event-based data (including both vehicle-detector actuation events and signal phase change events) can now be automatically collected and archived by the SMART-SIGNAL system. High-resolution traffic data is valuable because it potentially allows us to recover the event history of a traffic signal; it thus can help provide a basis to analyze the relationship between signal phase changes and traffic flow during queue formation and discharge. In particular, high resolution data reveals some "break points" that identify traffic flow pattern changes. These break points represent concrete positions of some major shockwave and contribute significantly to our approach of the long queue estimation. Indeed, we will demonstrate with such data that simple shockwave theory, originally developed by Lighthill and Whitham (1955) and Richards (1956), can be successfully applied to estimate intersection queue length.

It should be noted that our queue length estimation methods are designed to work with typical detector configurations for signal operation, *i.e.*, with either stop-bar detectors for vehicle presence detection or advance detectors (detectors placed a few hundred feet upstream from a stop-line) for green extension, or both. We make no provisions for link input detectors, *i.e.*, detectors placed sufficiently upstream so that input traffic flow to the traffic signal can be measured. With link input detectors, input-output approaches can be used for queue estimation. In addition, such detectors are usually not available in the real-world.

This chapter begins with a review of shockwave theory, followed by data description and "break points" identification. The analytical model is presented in Section 4.2, along with two expansion models that allow variations of high-resolution traffic signal data to be

used for specific real world scenarios. The testing results are given in Section 4.3 followed by discussion and conclusions.

4.1 Shockwave Analysis & Break Points Identification

4.1.1 Shockwave Analysis

Traffic shockwave theory is derived from the Lighthill-Whitham-Richards (LWR) traffic flow model. LWR hypothesizes that flow is a function of density at any point of the road. A shockwave is defined as "the motion (or propagation) of an abrupt change (discontinuity) in concentration" (Stephanopolos & Michalopoulos, 1979). Traffic shockwave theory is derived from LWR model by applying the method of characteristics to analytically solve the partial differential equation (PDE) in the LWR model. Basically, when characteristic curves (along which, the density is constant) interact, a shockwave is formed and whose velocity can be determined by the equation:

$$w = \frac{\Delta q}{\Delta k} = \frac{q_2 - q_1}{k_2 - k_1} = \frac{k_2 v_2 - k_1 v_1}{k_2 - k_1} \tag{4.1}$$

where q_1 (q_2), k_1 (k_2), v_1 (v_2) represent, respectively, the flow, density and velocity of the upstream (downstream) region. A traffic shockwave can also be illustrated with a fundamental diagram (flow-density curve). The tangent of the chord drawn between any two points on the curve defines the shockwave speed, as shown in Figure 4-1 (We will explain w_1, w_2, and w_3 in the following).

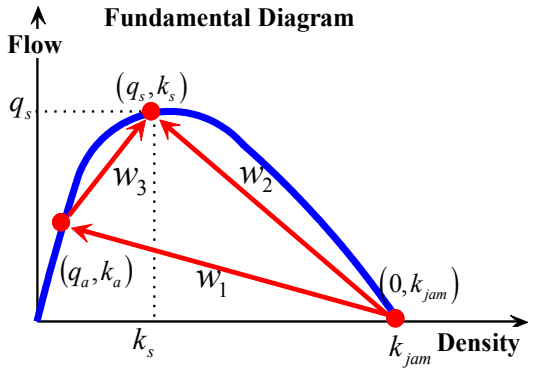

Figure 4-1 Representation of Shockwaves in the Fundamental Diagram

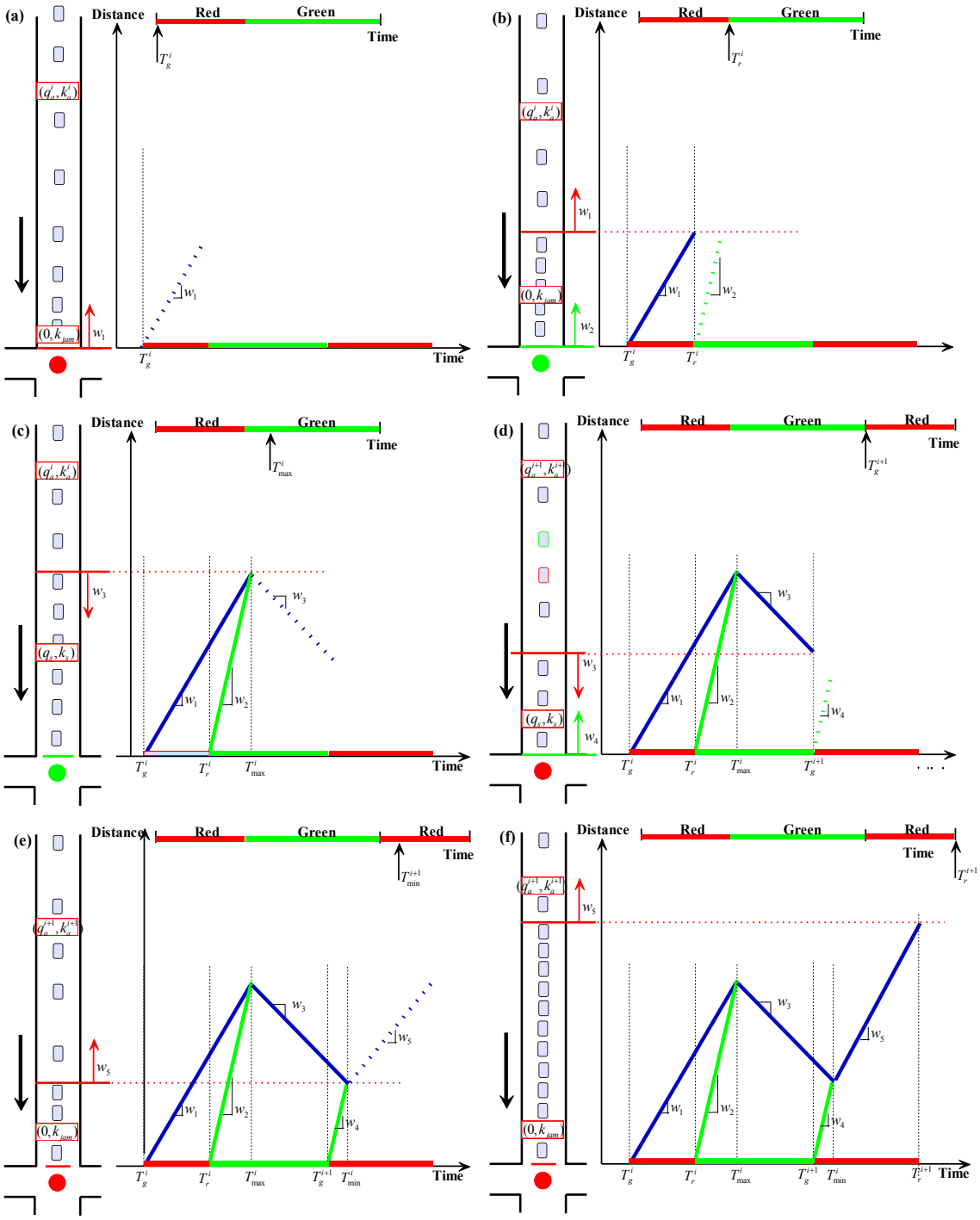

Figure 4-2 Shockwave Propagation

Signalized intersections generate multiple shockwaves due to the stop-and-go traffic caused by signal changes. For a more complete explanation, assume that a queue has

been fully discharged during the last green phase. In the following red interval, vehicles are forced to stop, which creates different flow and density conditions between the arrival and stopped traffic. Such interruption of traffic flow, as indicated in Figure 4-2a, forms a queuing shockwave (w_1 in Figure 4-1) moving upstream with velocity

$$w_1 = \frac{0 - q_a^i}{k_{jam} - k_a^i} \tag{4.2}$$

where *0* and k_{jam} represent the jammed flow and density; and $q_a{}^i$ and $k_a{}^i$ are the average arrival flow rate and density during the i^{th} cycle (subscript *a* means "arrival traffic"). In Figure 4-2a, $T_g{}^i$ and $T_r{}^i$ indicate the end time and the start time, respectively, of the effective green during the i^{th} cycle.

Queuing shockwave w_1 continues to propagate to upstream. At the beginning of the effective green ($T_r{}^i$ in Figure 4-2b), and assuming there is no congestion downstream, vehicles begin to discharge at the saturation flow rate, thereby, forming a second shock wave, or discharge shockwave (w_2 in Figure 4-1), at the stop line that travels upstream with speed

$$w_2 = \frac{q_s - 0}{k_s - k_{jam}} \tag{4.3}$$

where q_s and k_s are the saturation flow (capacity) and density.

Discharge shockwave w_2 usually has higher speed than w_1, so the two waves will meet at time $T_{max}{}^i$ (in Figure 4-2c), which is the time at which this approach has the maximum queue length. As soon as the two shockwaves meet, a third wave (known as a departure shockwave, w_3 in Figure 4-1) is generated propagating toward the stop line with speed

$$w_3 = \frac{q_s - q_a^i}{k_s - k_a^i} \tag{4.4}$$

At the end of this cycle, *i.e.,* the start of the red phase of the next cycle, if the queue cannot be fully discharged, a residual queue is formed, which builds a fourth wave w_4 (defined as a compression wave, Figure 4-2d) at the stop line moving upstream with speed

$$w_4 = \frac{0 - q_s}{k_{jam} - k_s} \tag{4.5}$$

Shockwave w_4 illustrates the queue compression process. Waves w_3 and w_4 have inverse directions therefore they meet at time $T_{min}^{\ i}$ (Figure 4-2e), which is the time the approach has minimum queue length, *i.e.,* the residual queue. As soon as the two waves meet a fifth wave w_5 forms a new queuing wave of the $(i+1)^{th}$ cycle, and moves upstream with similar speed as shock wave w_1.

$$w_5 = \frac{0 - q_a^{i+1}}{k_{jam} - k_a^{i+1}} \tag{4.6}$$

A similar process is repeated in the following cycles as indicated from Figure 4-2a to f.

It is clear that the tail end of the queue follows the trajectory of shockwaves w_1 and w_3. However, the queue dynamics can be analytically calculated only if accurate vehicle arrivals are known. As we mentioned before, in the real world, once a vehicular queue spills back to detectors, traffic arrival information at the detector location in the current cycle is unavailable until the queue starts to discharge.

The shockwave motion described above demonstrates the repetitive queuing and discharging process at a signalized intersection. The queues begin to accumulate at the start of a red interval and discharge at the beginning of a green phase. Maximum queue length is achieved shortly after the green start, while minimum queues (residual queues) form after a red start. The repeated nature of this process suggests that if some crucial "break points" can be identified utilizing detector data, the queuing and discharging process can be recovered for past cycles. As indicated in Figure 4-3, most waves propagate through the detector line (if there exist waves that do not cross the detector line, then this link either has short queues or is oversaturated, which will be discussed later). Points A, B, and C in Figure 4-3 represent the times that traffic flow changes at a detector location. If recognized, these break points can be used to recover the queuing process in this cycle. In particular, we are interested to estimate the maximum queue

length in a cycle, *i.e.*, point H in Figure 4-3. As will be explained in Section 4.1.2, if high-resolution event-based traffic signal data can be stored and archived, such data can be used to identify these points and estimate queue length in the immediately preceding cycle.

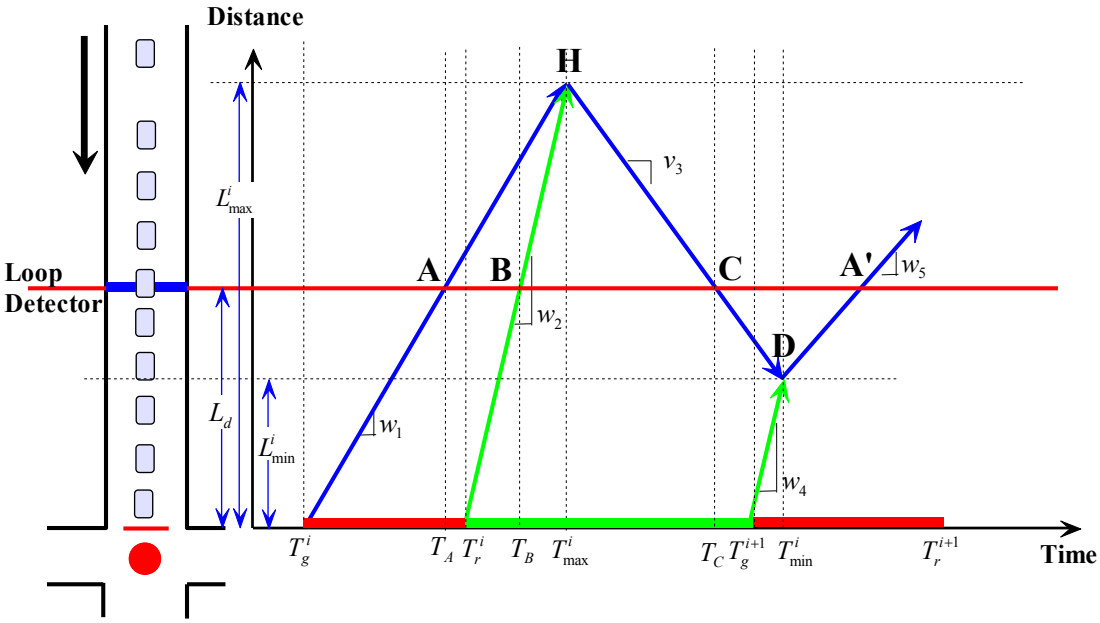

Figure 4-3 Break Points A, B, C, & Traffic Shockwaves at an Intersection

4.1.2 Identification of Break Points

The high-resolution data utilized in this research was collected by the SMART-SIGNAL system. As introduced in Chapter 2, SMART-SIGNAL can continuously collect and archive high resolution event-based data including vehicle and signal events (Figure 2-2). Event data provides the start and end times of every vehicle-detector actuation "event" and every signal phase change "event". Therefore, the time difference between the start and end of a signal event represents a phase interval. The time interval between the start and end of a vehicle actuation event, which is the time it takes for a vehicle to pass the detector, is the detector occupancy time. By assuming an effective vehicle length, the occupancy time can be used to calculate travel speed. Finally, a vehicle gap is the time interval between the end of a vehicle actuation event and the start of a following event (at

the same detector). In other words, it is the time interval between two consecutive vehicles crossing the detector.

As noted in Section 4.1.1, "break points" A, B, and C represent the time instants that a traffic condition changes within a cycle. (Here we define a cycle start as the effective red start and a cycle end as the effective green end). In detail, the time at which point A appears (T_A) is the moment that the queuing shock wave w_1 propagates backward to the loop detector line. Between T_g^i (the end of green in the i^{th} cycle) and T_A, the vehicles pass the loop detector with the arrival traffic state (q_a^i, k_a^i). By contrast, between T_A and T_{max}^i (the time that maximum queue is achieved), no vehicle can pass the loop detector because of the jam traffic condition (0, k_{jam}). Point A can be used to judge whether there is a long queue or not, because if point A does not exist, which means that the queuing shockwave does not propagate to the detector, then the queue length must be less than the distance between the stop line and the advance detector. Point A is not difficult to identify since after T_A, the detector is occupied for a relatively long time, making the value of the detector occupancy time is relatively large as well.

For practical application, a threshold value for point A identification is necessary. In this study, we observed that 3-second is long enough to check whether point A exists. If the detector occupancy time is longer than 3 seconds after T_g^i, then the intersection has a long queue; and vice versa.

Figure 4-4a demonstrates "event-based" detector occupancy time in a sample signal cycle. As indicated in the figure, after the queue spills back to the advance detector, the occupancy time is significantly greater than 3 seconds (about 45 seconds in this particular cycle). We should point out that second-by-second occupancy data can also be used to identify point A, *i.e.*, the occupancy remains 100% for more than 3 seconds.

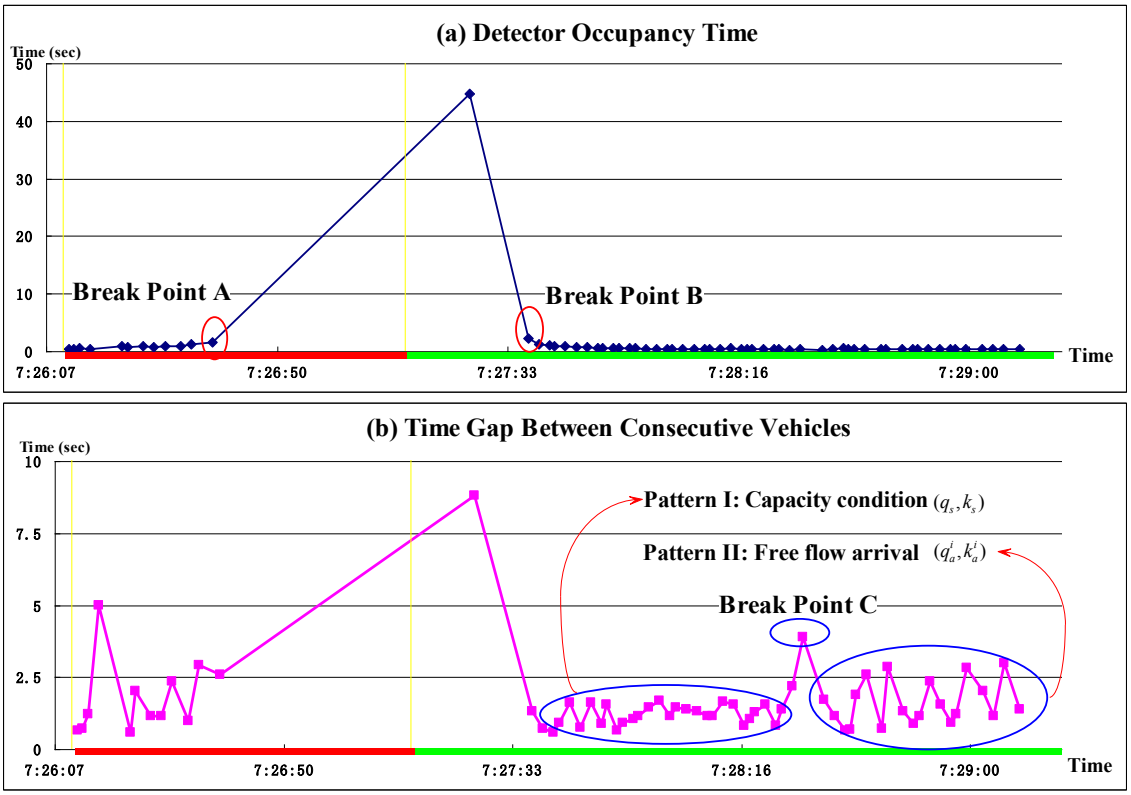

Figure 4-4 (a) Detector Occupancy Time Profile in a Cycle; (b) Time Gap between Two Consecutive Vehicles in a Cycle.

Point B indicates the time (T_B) that the discharge shockwave passes the detector. Between effective green start time (T_r^i) and T_B, the traffic state over the detector is (0, k_{jam}); after T_B, vehicles are discharged at saturation flow rate and the traffic state changes to (q_s, k_s). It is also not difficult to identify point B using high-resolution data. After the green starts and before T_B, traffic volume is zero, and detector occupancy time is high (larger than 3 seconds) or second-by-second occupancy continues to be 100% for more than 3 seconds. After T_B, queued vehicles begin to discharge over the detector, detector occupancy time significantly drops (as indicated in Figure 4-4a), making break point B easy to spot.

Most importantly, it is to identify break point C. As will be explained in the next section, we can use the time of point C (T_C), combined with the discharge shock wave w_2, to estimate the maximum queue length and re-construct the queue forming and discharging process. Point C indicates the time (T_C) when the rear end of the queue passes the

detector. The time duration between T_B and T_C is closely related to the estimation of maximum queue length. As noted earlier, wave w_3 is the interface between the saturation traffic state (q_s, k_s) and the arrival traffic state (q_a^i, k_a^i). Therefore, before point C appears, vehicles discharge at the saturation flow rate at the location of the loop detector, *i.e.,* the traffic state is (q_s, k_s). After the wave propagates to the detector location, the traffic condition becomes (q_a^i, k_a^i), *i.e.,* the arrival traffic states.

A threshold should be selected to identify the two different traffic states (q_s, k_s) and (q_a^i, k_a^i). Based on our observation, the time gap between two consecutive vehicles is sensitive to traffic state change. As indicated in Figure 4-4b, traffic separates two states by break point C. Before T_C, the time gaps between vehicles are small (less than 2.5 seconds) and the variance is small. This means that most of vehicles are discharged at the saturation flow rate. But after T_C, the vehicle gaps become much bigger and the variance is significantly increased. More importantly, there usually exists a time lag between the saturated queue discharge flow and newly arrived traffic, as shown in Figure 4-4b.

Ideally, statistical analysis of vehicle gap data (for example, comparing the means and variances of time gaps of two vehicle groups) would be used to recognize a traffic state change. However, for simplicity and practicality, we determined a threshold value for the time lag between saturated queue discharge flow and arrival traffic based on our observations. In this study, if the value of a time gap is greater than 2.5 seconds, we consider that the end of the queue has propagated forward and reached the detector line, therefore T_C is identified. Considering the variation of time gaps, using a single value to separate traffic states may result in large errors. In our implementation, if the time gap is between 2.5 seconds and 3 seconds, the system will continue searching the second and third points with time gaps over 2.5 seconds to make sure that the traffic state has really changed. It should also be noted that the method described above is still feasible if only second-by-second data is available. This is true because we know that a gap larger than 2.5 seconds means 0% occupancy for at least two consecutive seconds.

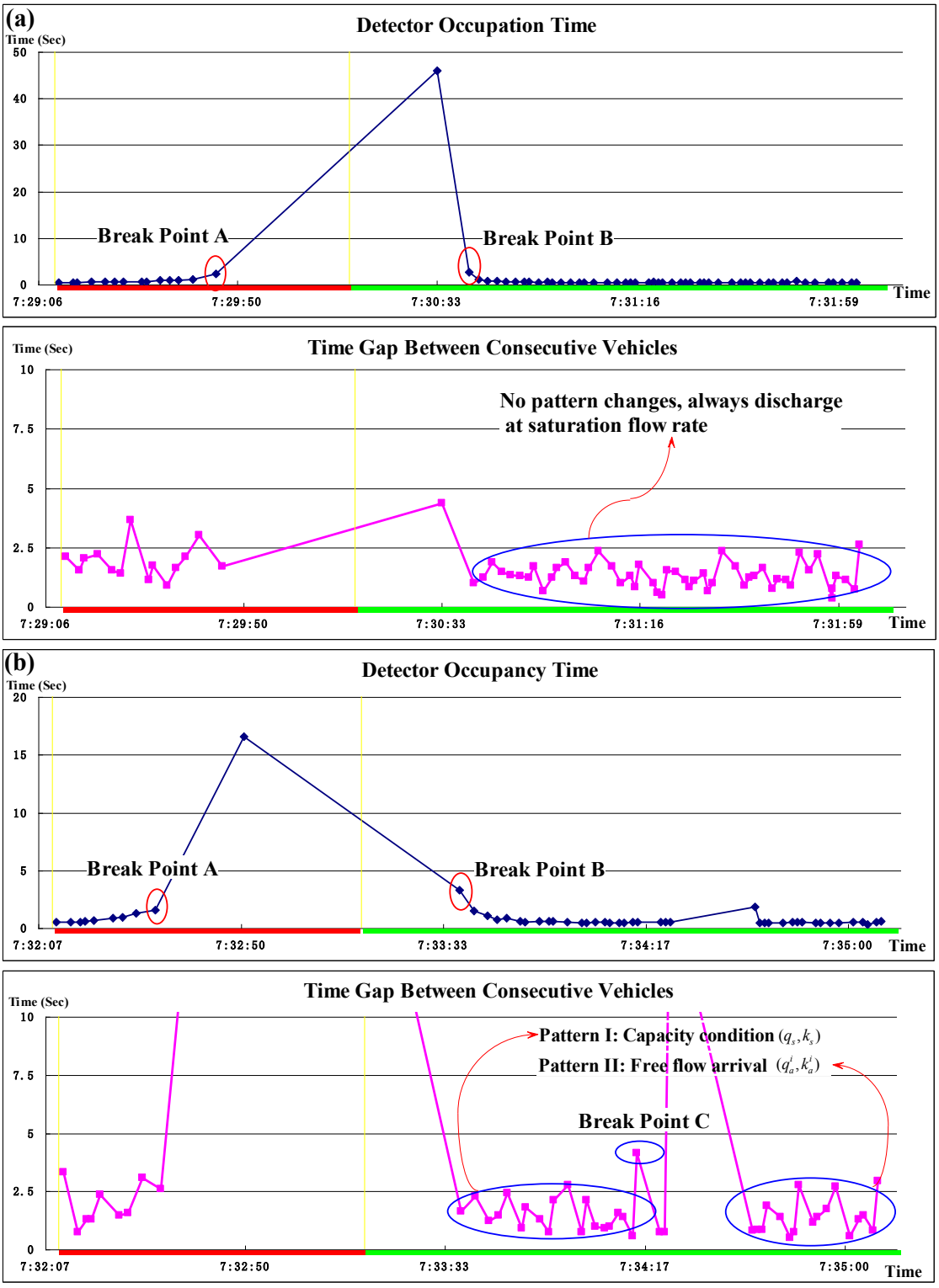

Figure 4-5 Two Other Examples of Occupancy time & Gap Data – (a) Oversaturation; (b) Platoon Arrival

Two other examples of occupancy time and gap data which indicate two interesting situations are demonstrated in Figure 4-5, one where the break point C cannot be found; and the other where vehicles arrive as a platoon. In Figure 4-5a, during the green phase, the traffic pattern does not change and vehicles keep discharging at the saturation flow rate. This could occur for two reasons: 1) the vehicle queue cannot be discharged completely and this approach at this cycle is possibly oversaturated; 2) a vehicle platoon arrives within a small time lag so that a large vehicle gap cannot be found. The second situation, as indicated in Figure 4-5b, is that although the break point is easy to find, the traffic pattern before and after the break point is nevertheless similar. The explanation is that a platoon discharged from an upstream intersection arrives after the queue has been discharged. This situation, sometimes, will lead to errors to our queue estimation model, an issue we discuss in later sections.

4.2 Queue Estimation Models

The queue estimation models proposed here are based on the identification of breakpoints A, B, and C. As indicated in Figure 4-3, the crucial task is figuring out how to estimate the coordinate of point H, the maximum queue, in both spatial (L_{max}^{i}) and temporal (T_{max}^{i}) dimensions. Point H is the intersection point of three waves and its coordinate can be determined by any two of them. As mentioned above, wave w_1 is a queuing wave, which highly depends on traffic arrival. Although wave w_3 is also related to arrival flow (see Eq.(4.3)), it represents the arrival traffic after queue discharge. Since discharging process is more stable compared with queuing process and the traffic state during discharging process can be estimated using detector data, this research uses waves w_2 and w_3, instead of w_1, to identify the coordinate of point H. Note that wave w_2, according to Eq.(4.4), has a constant velocity if we assume that saturation flow rate (q_s), saturation density (k_s), and jam density (k_{jam}) are known *a priori*. In this section, we present our basic model plus two expansions, especially designed for some practical applications.

4.2.1 Basic Model

To estimate shockwave speeds w_2 and w_3, we need to "mine" more information from the high resolution event-based data from one single loop detector. w_2 can be estimated using the distance between advance detector and stop-bar (L_d) and the time difference between the green start (T_r^i) and the time when discharge wave propagates back to advance detector (T_B), *i.e.* $w_2=L_d/(T_B - T_r^i)$. w_2 can also be estimated using Eq.(4.3) with assumed saturation flow rate (q_s), saturation density (k_s), and jam density (k_{jam}).

To estimate shockwave speed w_3, we will need to estimate two traffic states (q_s, k_s) and (q_a^i, k_a^i) and apply Eq.(4.4). We know we can estimate traffic states because event-based data contain both the occupancy time and gap between consecutive vehicles passing the loop detector. Occupancy time is used to estimate the speed of individual vehicles by assuming an effective vehicle length; and the sum of occupancy time and vehicle gap, *i.e.,* headway, is used to estimate the average flow. Density can then be estimated by average flow and space mean speed. The equation for estimating the space mean speed (v_s), flow (q) and density (k) is (4.7)

$$\begin{cases} v_j = D_e \Big/ t_{o,j} \\ v_s = 1 \Big/ \left(\dfrac{1}{N} \displaystyle\sum_{j=1}^{N} \dfrac{1}{v_j} \right) \\ q = 1 \Big/ \left(\dfrac{1}{N} \displaystyle\sum_{j=1}^{N} h_j \right) = 1 \Big/ \left(\dfrac{1}{N} \displaystyle\sum_{j=1}^{N} (t_{o,j} + t_{g,j}) \right) \\ k = q \Big/ v_s \end{cases} \tag{4.7}$$

where, $t_{o,j}$ and $t_{g,j}$ are the detector occupancy time and time gap of vehicle j, respectively; v_j, h_j are the speed and time headway of vehicle j, respectively; D_e is the effective vehicle length, *i.e.,* the sum of average vehicle length and detector length, which may require calibration; and N is the number of vehicles identified as having the same traffic state.

As we discussed before, the traffic state is (q_s, k_s) between T_B and T_C, and (q_a^i, k_a^i) after T_C (and before T_g^{i+1}). We estimate values of q_s, k_s, q_a^i, and k_a^i by applying Eq.(4.7). Note that we use observed data to estimate traffic state (q_s, k_s) instead of assuming the constant values; this is because the capacity or saturation flow rate will decrease if the traffic is affected by downstream intersections. The velocity of wave w_3 is calculated based on the two estimated traffic conditions (Eq.(4.4)). Using estimated w_3 and w_2, the maximum queue length L_{max}^i and time T_{max}^i during i^{th} cycle can be calculated as:

$$\begin{cases} L_{\max}^i = L_d + (T_C - T_B) \Big/ \left(\dfrac{1}{w_2} + \dfrac{1}{w_3} \right) \\ T_{\max}^i = T_B + (L_{\max}^i - L_d) \Big/ w_2 \end{cases} \tag{4.8}$$

where L_d is the distance from the stop line to the loop detector.

The minimum queue length L_{min}^i and time T_{min}^i (if a residual queue exists) during i^{th} cycle can be also calculated:

$$\begin{cases} L_{\min}^i = \left(\dfrac{L_{\max}^i}{w_3} + T_{\max}^i - T_g^{i+1} \right) \Big/ \left(\dfrac{1}{w_3} + \dfrac{1}{w_4} \right) \\ T_{\min}^i = T_g^{i+1} + L_{\min}^i \Big/ w_4 \end{cases} \tag{4.9}$$

where w_4 is the velocity of shock wave w_4, which has same value as velocity of the shock wave w_2 (Eq.(4.3)).

The dynamic queue discharging process after T_{max}^i (before T_{min}^i) is easy to formulate because we know the wave speed w_3; and the queuing process before T_A has been recorded by loop detectors (as we can see in Figure 4-6, the arrival may not be constant). The problem now is to estimate the queue length between points A and H. Without any other information, we can assume a constant velocity for wave w_1, which can be estimated by

$$w_1 = (L^i_{\max} - L_d) \Big/ (T^i_{\max} - T_A) \tag{4.10}$$

With w_1, w_2, w_3, and w_4, the entire queue accumulation and discharge processes can be fully described.

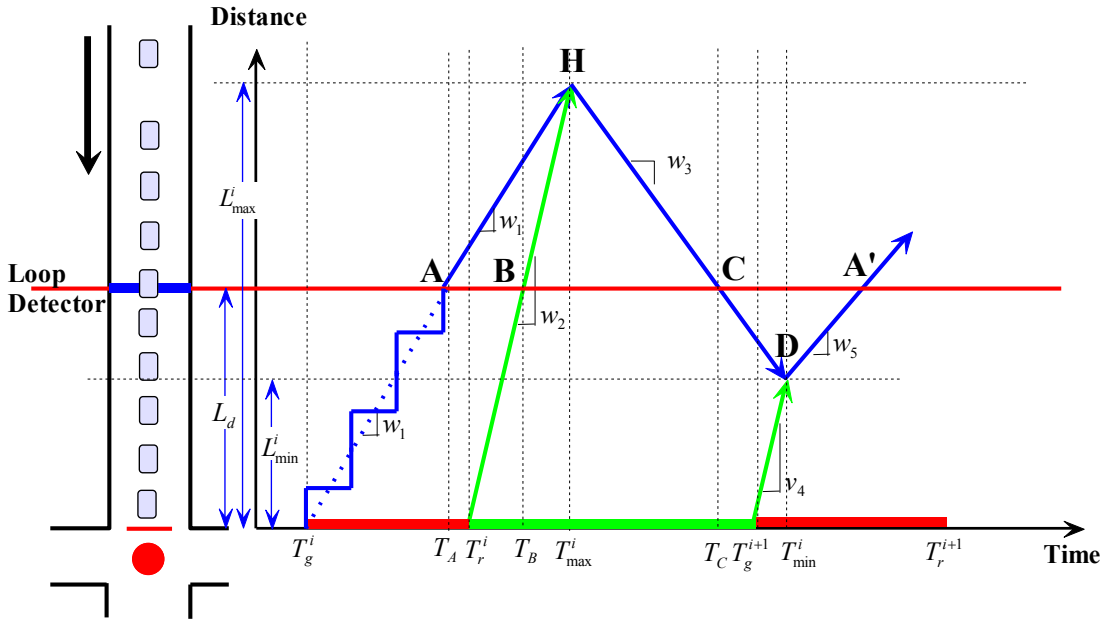

Figure 4-6 Basic Model for Intersection Queue Length Estimation

4.2.2 Expansion I – Using Second-by-second Detector Data

In the basic model, event-based data is required to identify the shockwave speeds; such requirement may not be satisfied in the real world. With second-by-second detector data and signal phase data available (such as the ACS-Lite system), we can use Model Expansion I, which provides a simple alternative to calculate maximum and minimum queue length, without the need to estimate traffic states using Eq.(4.7), as explained below.

We know that most of the vehicles passing loop detectors before T_C, belong to maximum queue $L_{max}{}^i$, while a small portion of vehicles joined the queue after the tail end of the

original maximum queue began to move. This small portion of vehicles theoretically makes no contribution to the maximum queue length, but these cars are affected by the queue. Therefore, an approximation can be made to treat all vehicles passing the detector between the green start and T_C as queued vehicles. Because the number of vehicles can be easily obtained using second-by-second detector data, by assuming a constant jam density, we can directly estimate maximum queue length ($L_{max}^{i'}$) and time ($T_{max}^{i'}$) as well as wave speed w_3' under such an approximation.

$$\begin{cases} L_{\max}^{i}{}' = N / k_{jam} + L_d \\ T_{\max}^{i}{}' = T_r^i + L_{\max}^{i}{}' / w_2 \\ w_3' = \left(L_{\max}^{i}{}' - L_d\right) / \left(T_C - T_{\max}^{i}{}'\right) \end{cases} \tag{4.11}$$

where N is the number of vehicles passing a detector between T_g^i and T_C.

By using a constant shockwave speed w_4, Eq.(4.9) can then be applied to estimate minimum queue length ($L_{min}^{i'}$) and time ($T_{min}^{i'}$) as well as wave speed w_1'.

Essentially, Expansion I is a simplified queue estimation model. As demonstrated in Figure 4-7, point H represents the true maximum queue, while H*'* is an approximation. Theoretically, the expanded model overestimates the maximum queue length since it includes some portion of vehicles which do not belong to the stopped vehicle queue.

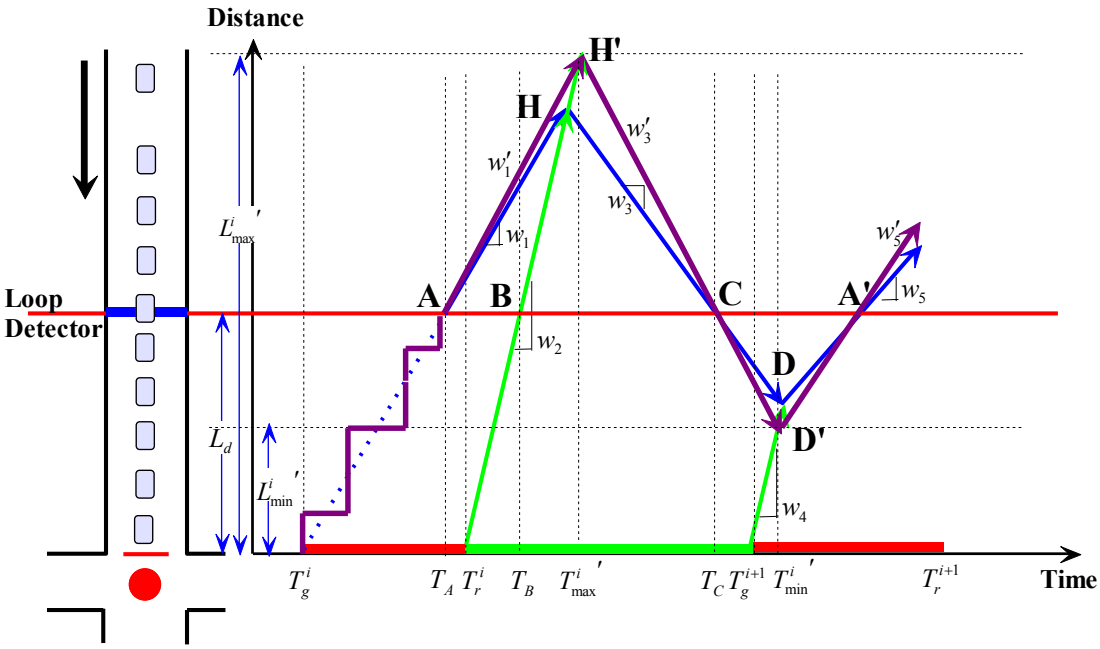

Figure 4-7 Expansion I for Intersection Queue Length Estimation

4.2.3 Expansion II – Dealing with Wired-together Detectors

The expanded model II is specifically designed to deal with cases where detectors in different lanes (but the same approach) are wired together, or where a single detector covers multiple lanes. In this case, vehicles traveling in different lanes are counted only once if they pass the detector at the same time. This detector configuration is common for many signalized intersections in Minnesota. Under such a design, break points can still be successfully identified using second-by-second or event-based detector occupancy, but the vehicle counts are no longer accurate; rendering the basic Model and Expansion I unworkable. An approximation is made here to deal with this problem. This approximation is based on the assumption that after the tail end of a maximum queue begins to move, no additional vehicles can join the queue. In this case, the trajectory of the last stopped vehicle is actually shockwave w_3''. We can analytically derive this trajectory by assuming *a priori* known constant acceleration speed (3.5 ft/sec, for example).

Depending on whether the last vehicle in the queue reaches maximum speed at point C, the time interval between $T_{max}^{i''}$ and T_C (t_x in Figure 4-8) can be estimated using the following equation:

$$\begin{cases} \frac{1}{2}a\left(\frac{v_f}{a}\right)^2 + v_f\left(t_x - \frac{v_f}{a}\right) = w_2(T_C - T_B - t_x) & \text{if } t_x \geq \frac{v_f}{a} \\ \frac{1}{2}at_x^{\,2} = w_2(T_C - t_x) & \text{if } t_x < \frac{v_f}{a} \end{cases} \tag{4.12}$$

where v_f and a are free flow speed and acceleration speed, respectively.

Then the maximum queue length ($L_{max}^{i''}$) and the time ($T_{max}^{i''}$) can be estimated by:

$$\begin{cases} L_{\max}^{i\,\prime\prime} = L_d + w_2(T_C - T_B - t_x) \\ T_{\max}^{i\,\prime\prime} = T_C - t_x \end{cases} \tag{4.13}$$

Similar formulations are used to estimate the minimum queue length ($L_{min}^{i''}$) and the time ($T_{min}^{i''}$) as well as wave speed w_1''.

Clearly, the assumption made in this model may not be true in reality. As with Expansion I, Expansion II will sometimes overestimates the maximum queue. As indicated in Figure 4-8, point H is the real maximum queue length, and estimated maximum queue length (H'') is actually an overestimation.

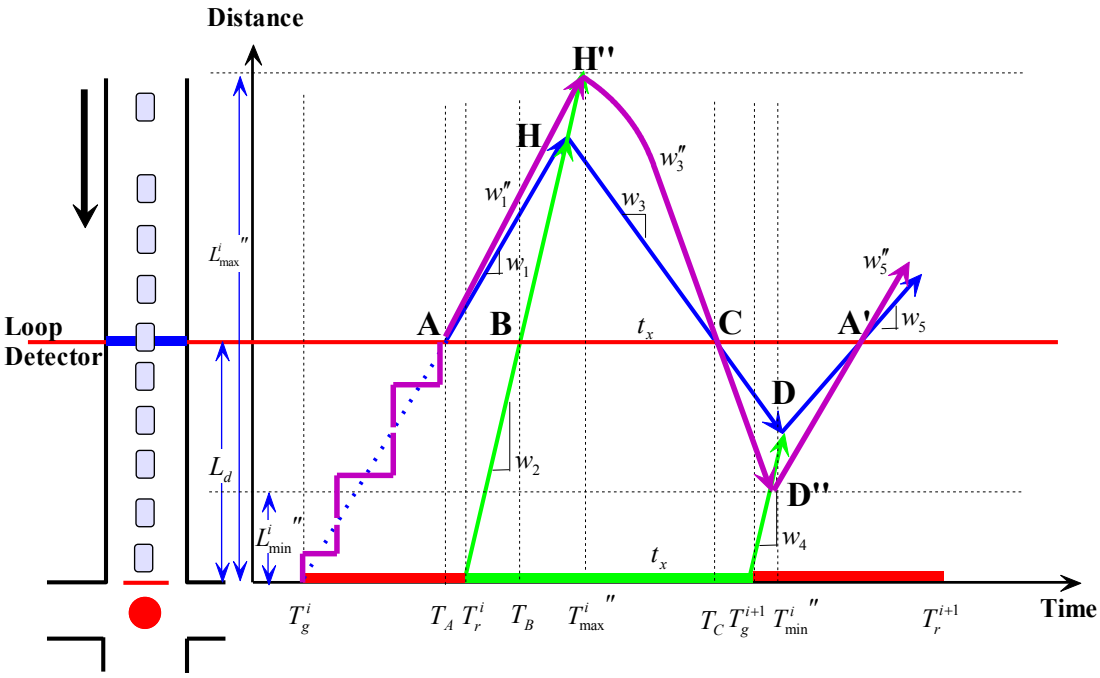

Figure 4-8 Expansion II for Intersection Queue Length Estimation

4.3 Implementation

4.3.1 Implementation Procedure

Figure 4-9 illustrates the implementation procedure for the queue length estimation algorithms. The first step is to check whether break point A exists. If point A cannot be found, it is a short queue case and a simple input-output method can be applied to estimate queue size. This simple method counts number of vehicles passing loop detector, assumes saturation flow rate discharging when signal turns green, and calculates the residual number of vehicles within the area between stop line and the location of the advance detectors. Queue length can then be estimated by simply assuming an effective vehicle length in the jam traffic state. If point A exists, then points B and C must be identified. Actually, it may not be necessary to identify point B because the wave speed v_2 for most situations is a constant value. If point C can be identified, one of the models is chosen to estimate the long queue, depending on the resolution of the detector

data. If point C cannot be identified, as we discussed in the last section, the approach is considered to be under oversaturated. The queue length under oversaturated conditions is difficult to estimate, and requires a modified model that will be introduced in the next chapter.

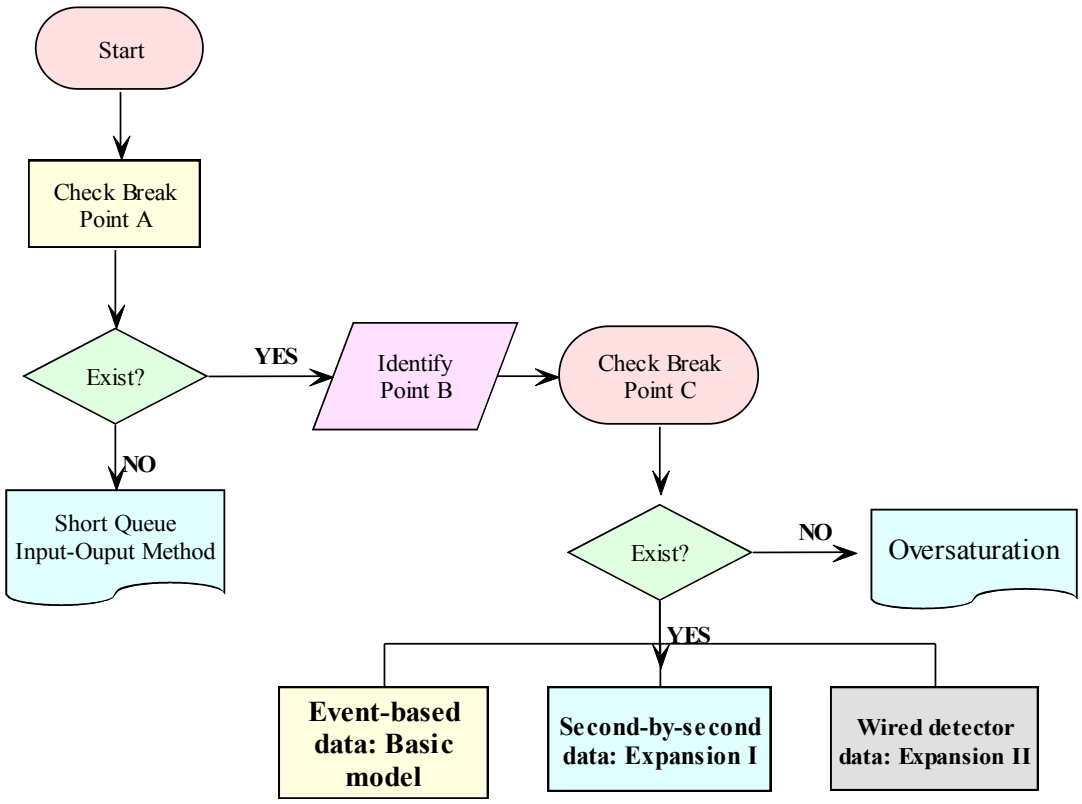

Figure 4-9 Flow Chart of Implementation Procedure for Intersection Queue Length Estimation

4.3.2 Field Evaluation & Results by the Research Team

We tested our models to see how well they estimated intersection queue lengths. The intersection of Trunk Highway 55 and Rhode Ave. in Minnesota was selected as the testing site because of the frequent occurrence of extended queues in the west bound direction (from Glenwood Ave. to Rhode Ave.) during morning peak hours. Figure 4-10a is a map of the intersection (including six coordinated intersections on Trunk Highway

55) and Figure 4-10b shows the detector layout of the intersection. Data was collected from detectors #9 and #10 during two morning peak hours with fixed cycle-length (180sec) on June 11th, 2008 (Wednesday). The data was used to estimate the intersection queue length. The estimated values were then compared with the ground truth data, which was recorded by a camera installed at the intersection. The maximum queue length for each cycle was manually extracted from the video. Since the data collected by the detectors was event-based, all three models (basic, expansion I, and expansion II) were applied. The results of maximum queue length are presented in Figure 4-11. Table 4-1shows the Mean Absolute Percentage Error (MAPE) which is calculated by

$$MAPE = \frac{1}{m}\sum_{m}\left|\frac{Observation - Estimation}{Observation}\right| \times 100\% \tag{4.14}$$

where m is the total sample size.

As showed in Figure 4-11, the basic model successfully estimates the maximum queue length for both left and right lane; and the average MAPE is around 7% (Table 4-1). However, as predicted, Expansion I and Expansion II overestimate the maximum queue; and the MAPEs are about 14% and 20%. Regarding their ability to capture queue dynamics, we can clearly see in Figure 4-12 that the proposed models successfully describe queue formation and discharge. In addition, Table 4-2 presents the time differences when the maximum queue length is reached, using the three proposed models and real-world observations. The absolute errors for the three models are around 5 sec, which is quite respectable.

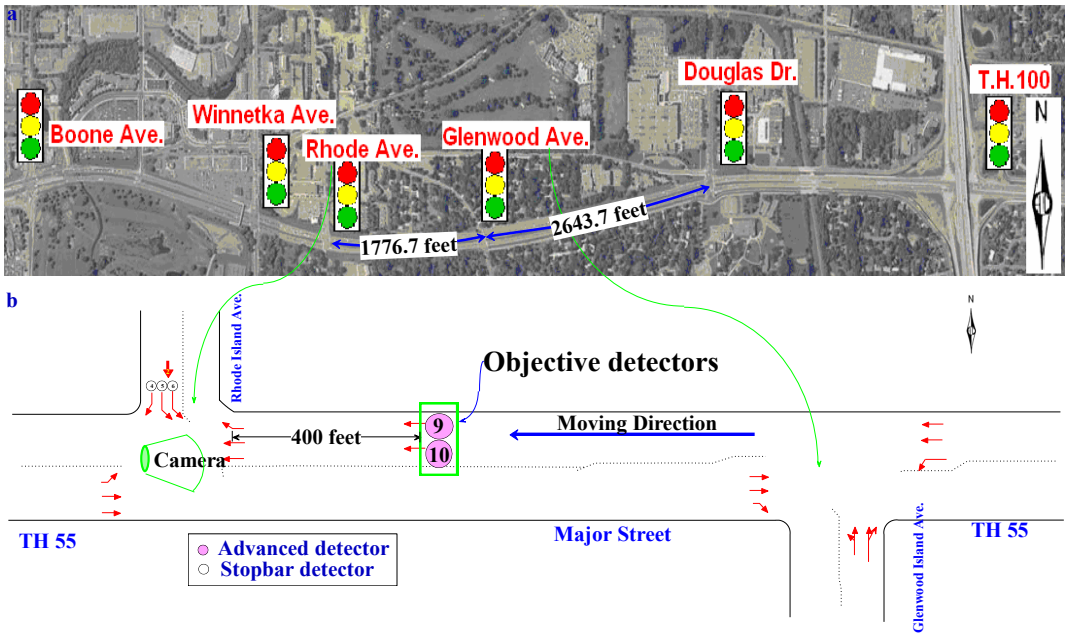

Figure 4-10 (a) Data Collection Site; (b) Detector Layout

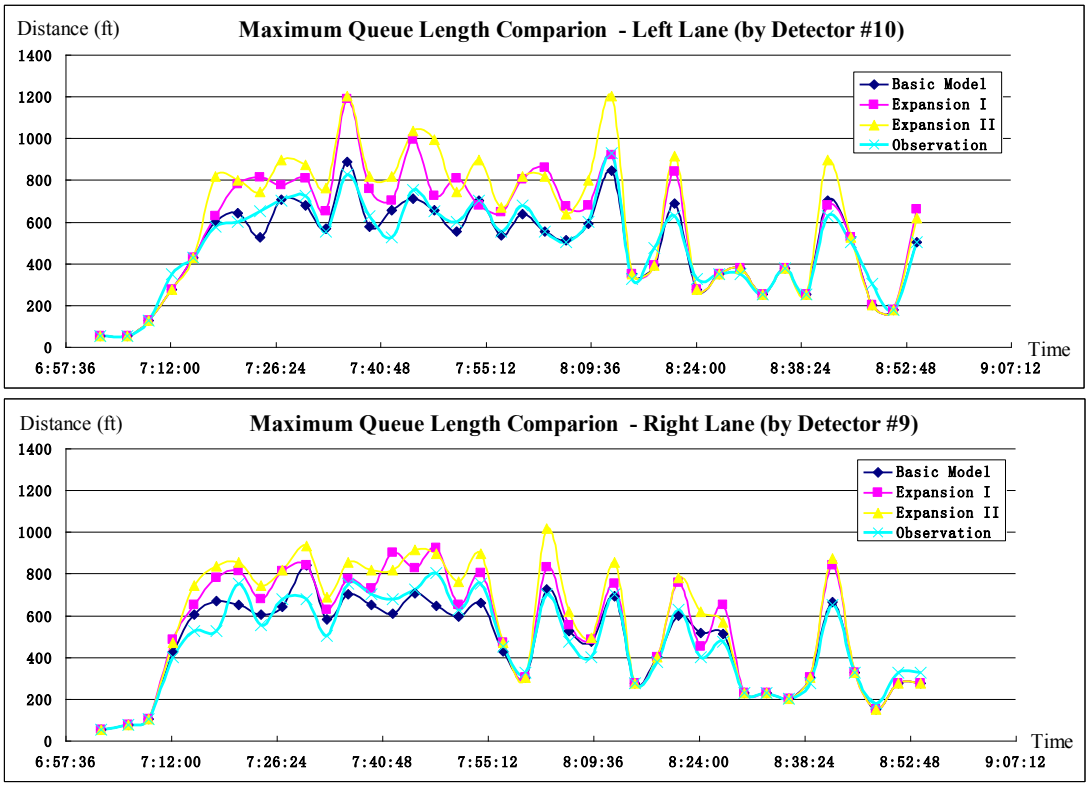

Figure 4-11 Testing Results – Maximum Queue Length Comparison

Table 4-1 Mean Absolute Percentage Errors of Three Models

MAPE of Maximum Queue Length	Basic Model	Expansion I	Expansion II
Left Lane (Detector #10)	**6.5%**	**15.9%**	**22.3%**
Right Lane (Detector #9)	**8.7%**	**13.4%**	**18.8%**

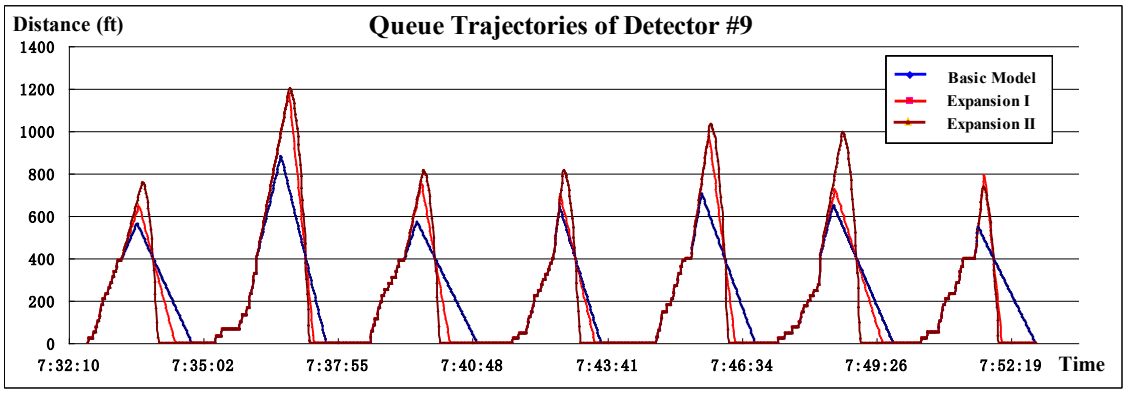

Figure 4-12 Queue Trajectories Estimated by Three Models

Table 4-2 Absolute Errors of Time of Maximum Queue Length

Absolute Errors of Time of Maximum Queue Length	Basic Model	Expansion I	Expansion II
Left Lane (Detector #10)	**6 sec**	**6 sec**	**5 sec**
Right Lane (Detector #9)	**5 sec**	**6 sec**	**4 sec**

4.3.3 Independent Evaluation Results by Alliant Engineering, Inc.

To verify our results, the Minnesota Department of Transportation (Mn/DOT) hired a Minneapolis-based Transportation Consulting firm, Alliant Engineering, Inc. to conduct an independent evaluation of the queue length estimation algorithm. Alliant sent observers to the field (the Rhode Island intersection) during morning peak (7:00am-9:00am) on three randomly selected days in 2008: Jul. 23rd, Oct. 29th, and Dec. 10th. These observers manually counted each vehicle as it entered the queue and recorded the time when the queue was maximum. They were instructed to count a stopped vehicle as one that was traveling at less than 5 mph. Figure 4-13 compares the time and length of

the maximum queues estimated by the basic model and the independently conducted observations (only right lane results are presented here). As can be seen, the proposed model tracks the trend of cycle-based queue dynamics successfully. The MAPE is relatively high (14.93%, on average, Table 4-3) compared with our own evaluation due to two possible reasons: 1) The collected queue size data (*i.e.*, the number of queued vehicles) was converted to queue length by simply multiplying 25 ft (the assumed constant effective vehicle length). Variation of actual vehicle length in the field may bring in conversion errors. For example, it was not unusual that when a detector was occupied (equivalent to a queue length of 400 ft), there were only 10 or 12 vehicles within the area between the stop bar and the detector line (equivalent to a queue length of 250 or 300 ft). 2) Engineering judgment was required during data collection to identify whether vehicles had joined the queue. This may also involve some errors. Overall, the proposed model accurately tracked the changes of cycle-based queue length for all three testing scenarios.

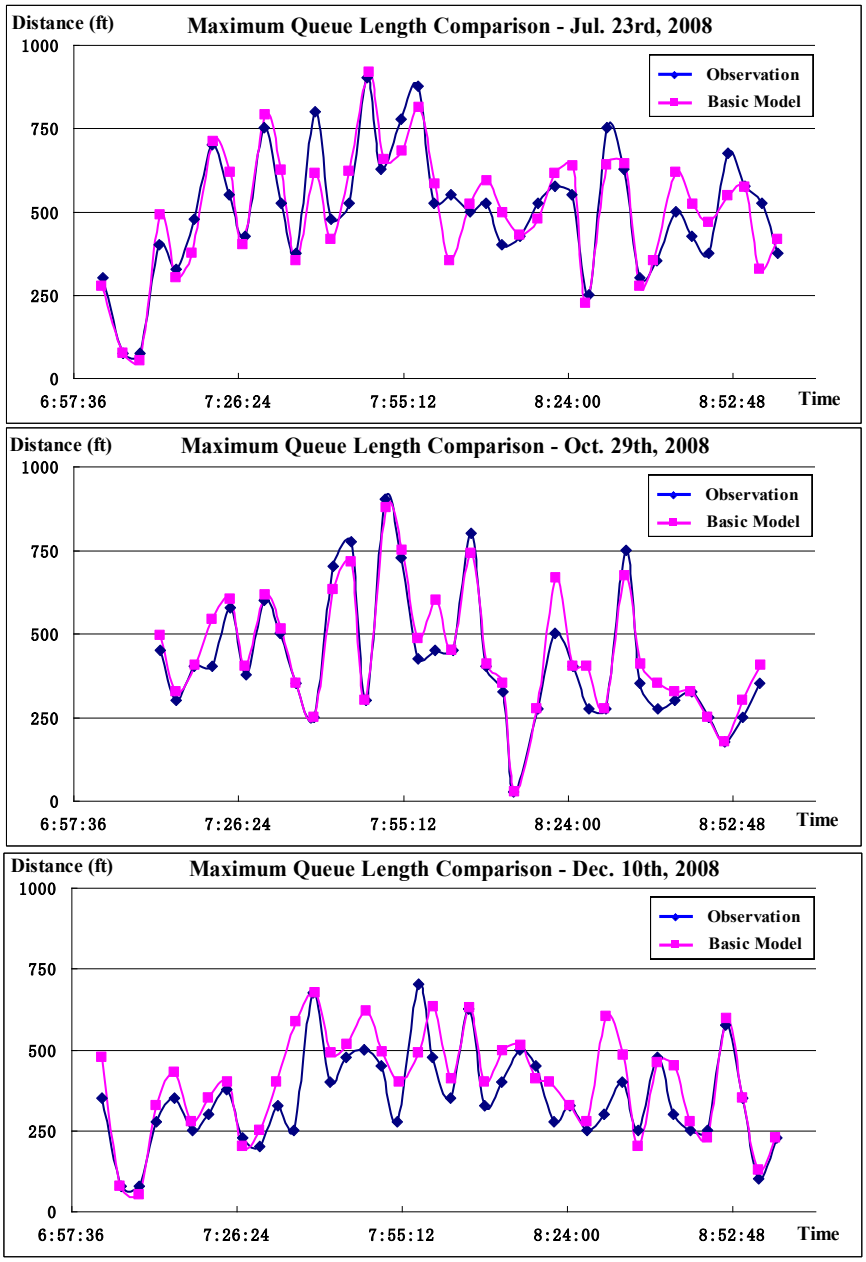

Figure 4-13 Comparison of Maximum Queue Length Obtained from the Basic Model & Independent Observers

Table 4-3 Mean Absolute Percentage Errors of Maximum Queue Length

	Jul.23rd, 2008	Oct.29th, 2008	Dec.10th, 2008	Average
MAPE	**12.89%**	**9.34%**	**22.03%**	**14.93%**

4.4 Discussion

Our queue length estimation models exhibit some limitations which need to be discussed.

1) Breakpoint C identification. When arrival traffic is at saturation flow rate and the time lag between the arrival flow and queue discharge is equal to saturation headway, an identification error can occur. Fox example, as indicated in Figure 4-14, a platoon discharged from an upstream intersection arrives right after the tail of a queue starts to discharge. In such situation, break point C cannot be identified, potentially leading to estimation errors. However, the likelihood of this occurrence is very small.

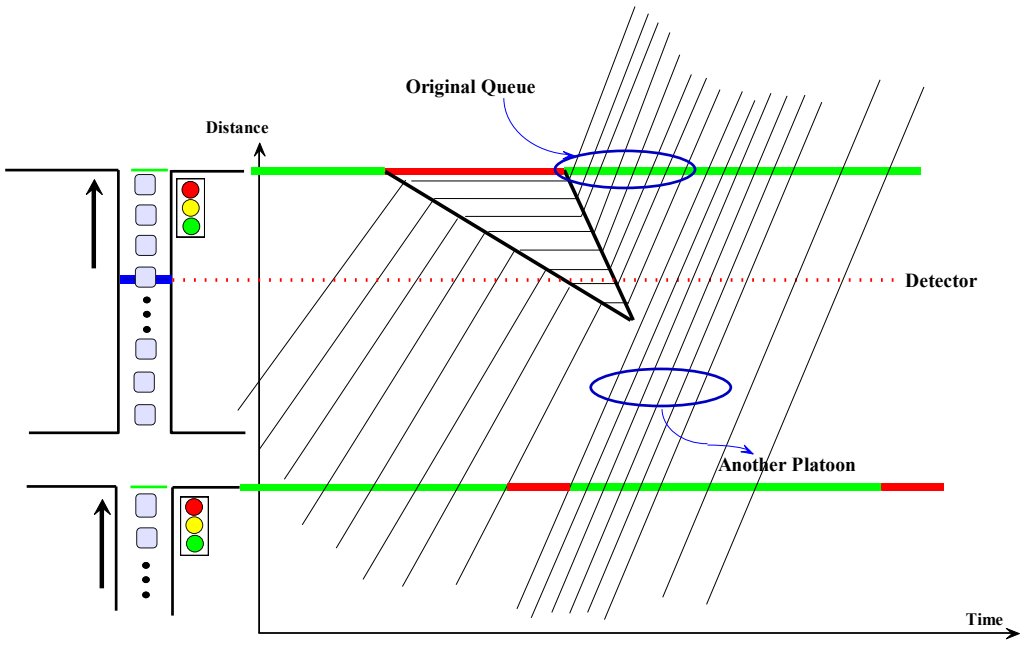

Figure 4-14 Immediate Platoon Arrival after Queue Discharge

2) Oversaturation. A failure to identify point C suggests that the intersection is potentially oversaturated, a situation our current model has limitation to deal with. A modified model will be presented in the next chapter.

3) Accuracy comparison among models. Theoretically, the basic model is most accurate of the three proposed. This is also supported by the testing results. However, since the basic model uses pre-determined effective vehicle length to estimate individual vehicle speed, queue estimation may be incorrect if the effective vehicle length is not accurate. By contrast, since Expansion I does not need the parameter of effective vehicle length, it may be more accurate. For the situations of frequent platoon arrivals, since the trajectory of the last queued vehicle is identical to that of shockwave trajectory, Expansion II is more accurate.

4) Detector errors. We need to point out that the proposed models do not consider the impact of detector errors, such as miscounting or over-counting. An error filtering method, such as the Kalman Filter, may improve the reliability of the proposed model. Such research is left for future study.

4.5 Summary

In this chapter, we proposed an innovative approach to estimate intersection queue length using existing detectors. A key methodological contribution of our approach is that it can estimate time-dependent queue length even when the signal links are congested with long queues. By applying LWR shockwave theory to high-resolution traffic signal data, we are able to distinguish different traffic states at the intersection, so that queue length estimation under congested conditions becomes possible. Two expanded models are also provided for practical applications to deal with cases where only second-by-second data is available or that if the detectors are wired together. The three models were evaluated by comparing the estimated maximum queue length with the ground truth data recorded by camera and human observers. The results indicate that all three models can accurately estimate intersection long queues, but the basic model is more accurate. Limitations of the proposed models are also discussed.

In the next chapter, this proposed model will be applied to quantify the severity level of oversaturation in temporal dimension, *i.e.* T-OSI.

5 Quantification of Oversaturation

Chapter 3 introduced T-OSI (oversaturation severity index in the temporal dimension) and S-OSI (oversaturation severity index in the spatial dimension). This chapter will propose two algorithms to measure T-OSI and S-OSI. Note that the identification algorithms discussed in this chapter will work with typical size 6 ft x 6 ft detector configurations for a vehicle-actuated signalized intersection, *i.e.*, with either stop-line detectors for vehicle presence detection or advance detectors (a few hundred ft upstream from the stop line) for green extension, or both. We assume advance detectors are available and we note the necessary changes to be made if the only available detector is located at the stop-bar. Although the algorithms presented in this chapter are demonstrated using event-based data from the SMART-Signal system, they are also applicable to second-by-second signal data coming from any other traffic signal management system.

This chapter is organized as follows. The next section adopts the queue length estimation method discussed in the last chapter to quantify the severity of oversaturation in the temporal dimension (T-OSI). In Section 5.2, we first experimentally examine Queue-Over-Detector (QOD) phenomenon, which is crucial for oversaturation identification; and then introduce an approach to quantify the severity of oversaturation in the spatial dimension (S-OSI). Section 5.3 presents the results from a field test, followed by a summary in Section 5.4.

5.1 Algorithm for Residual Queue Length Estimation

A residual queue at an intersection refers to those vehicles that are part of the discharging platoon that cannot pass the intersection during the green time. Residual queue also represents the minimum queue length at the end of a cycle. Vehicles in the residual queue will then occupy a portion of green time in the next cycle for discharging. The ratio

between the residual queue discharging time and the total available green time is thus denoted as T-OSI, as discussed in Chapter 3.

Estimation of residual queue length requires the reconstruction of the queue length profile within a cycle, including the maximum queue length that was reached. As the traditional input-output approach for queue length estimation can only handle queues shorter than the distance between vehicle detector and intersection stop line, we adopt the queue length estimation method described in the last chapter. We should note that the queue estimation method needs to be modified for the case of queue spillover from downstream, as will be discussed in Section 5.2.

The queue estimation method is based on the identification of traffic state changes and the associated shockwaves presented in a cycle. As described in the last chapter, if break point C can be identified, the maximum queue (both length $L_{max}^{\ i}$ and time $T_{max}^{\ i}$) and the minimum queue (both length $L_{min}^{\ i}$ and time $T_{min}^{\ i}$) during i^{th} cycle can be calculated by the following equations:

$$\begin{cases} L_{\max}^{i} = L_d + (T_C - T_B) \Big/ \left(\dfrac{1}{w_2} + \dfrac{1}{w_3} \right) \\ T_{\max}^{i} = T_B + (L_{\max}^{i} - L_d) \Big/ w_2 \end{cases} \tag{5.1}$$

$$\begin{cases} L_{\min}^{i} = \left(\dfrac{L_{\max}^{i}}{w_3} + T_{\max}^{i} - T_g^{i+1} \right) \Big/ \left(\dfrac{1}{w_3} + \dfrac{1}{w_4} \right) \\ T_{\min}^{i} = T_g^{i+1} + L_{\min}^{i} \Big/ w_4 \end{cases} \tag{5.2}$$

where L_d is the distance from the stop line to the loop detector; and $T_g^{\ i+1}$ is the green end of the $(i+1)^{th}$ cycle.

The presence of a residual queue at the end of a cycle must be determined before the calculation of the minimum queue length:

$$\begin{cases} \dfrac{L^i_{\max}}{w_3} + T^i_{\max} < T_g^{i+1} & \text{Without Residual Queue} \\ \dfrac{L^i_{\max}}{w_3} + T^i_{\max} \geq T_g^{i+1} & \text{With Residual Queue} \end{cases} \tag{5.3}$$

For severely congested traffic conditions, break point *C* may not be found during the green phase. In such cases, the traffic pattern does not change during the green phase and vehicles keep discharging at the saturation flow rate. This means that a residual queue must exist, at least, between the detector location and the stop line. Eq.(5.1) does not work since we are not able to estimate shockwave speed v_3. Under such conditions, the complete queue profile cannot be recovered from the detector data. However, since the entire green time has been used for queue discharge, the number of vehicles passing detector locations during the green time can be counted (between T_B and T_g^{i+1}), so that a minimum of the maximum queue length, *i.e.* $\min(L_{max}^i)$, can be estimated by simply taking the end of cycle (T_g^{i+1}) as T_C (see Figure 5-1). Since the space headway at jammed traffic conditions (d_{jam}) and the velocity of the discharge wave (w_2) are assumed constant, the following equation holds:

$$\text{If Point } C \text{ cannot be identified: } \begin{cases} \min(L^i_{\max}) = d_{jam} \cdot N + L_d \\ \min(T^i_{\max}) = T^i_r + \dfrac{\min\left(L^i_{\max}\right)}{w_2} \end{cases} \tag{5.4}$$

where N is the traffic count between T_B and T_g^{i+1}; d_{jam} is the space headway at jammed traffic conditions (assumed as a known constant); and T_r^i is the end of the red phase of an i^{th} cycle.

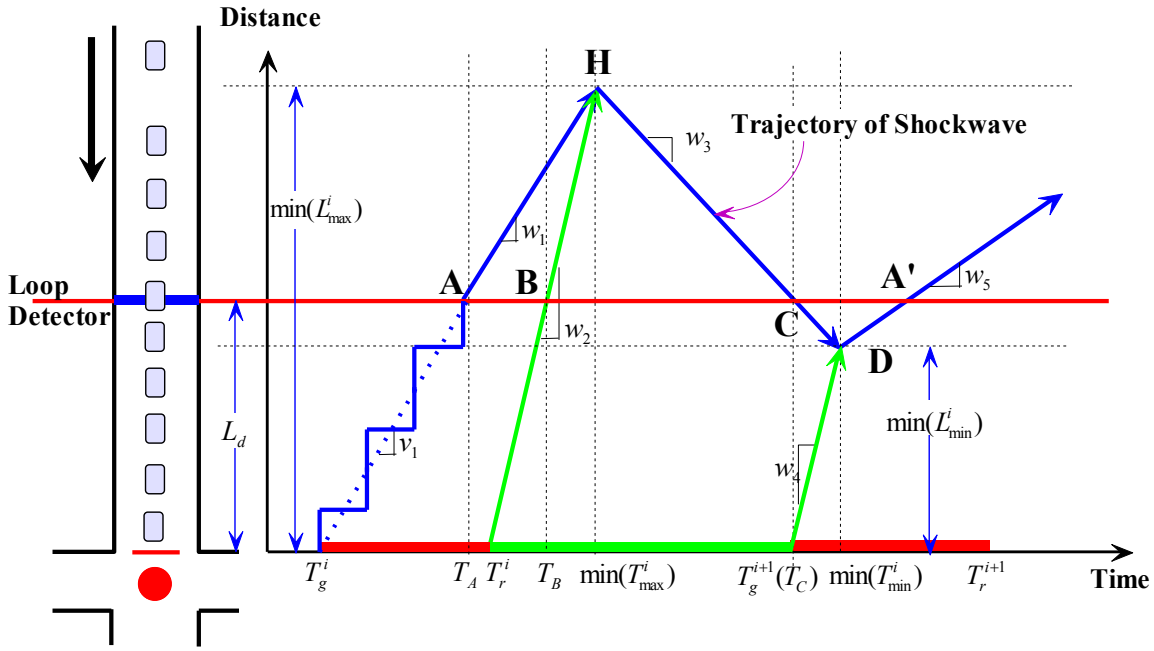

Figure 5-1 Calculation of Residual Queue Length when Point C cannot be Identified

Then v_3 can be calculated by Eq.(5.5):

$$\text{If point C cannot be identified: } w_3 = \frac{\min\left(L^i_{\max}\right) - L_d}{T^{i+1}_g - \min\left(T^i_{\max}\right)} \tag{5.5}$$

The coordinate of the minimum of the residual queue length, *i.e.* $\min(L_{min}{}^i)$ and $\min(T_{min}{}^i)$ for Point *D*, can then be estimated using Eq.(5.2).

If a residual queue exists at the end of a signal cycle, some portion of the green time in the following cycle will be used to discharge the residual vehicles, thereby becoming "unusable" green time for that cycle. The unusable green time can be calculated by taking the number of vehicles in the residual queue multiplied by the saturation discharge headway (around 2 seconds). The detrimental effect caused by the residual queue can therefore be quantified by the oversaturation severity index in temporal dimension (T-OSI):

$$\text{T-OSI} = \frac{\textit{unusable green time}}{\textit{total available green time}} \times 100\% = \frac{L^{i}_{\min} / d_{jam} \cdot h_s}{G} \times 100\% \qquad (5.6)$$

where G is the effective green time, and h_s is the saturation discharge time headway.

We should note that when only stop-line detection is available, a residual queue will exist as long as break point C cannot be identified within the green time. Although the length of the residual cannot be measured with the stop-line detection only, it is sufficient to say that oversaturation may have occurred at this intersection for this cycle, *i.e.* T-OSI > 0.

5.2 Algorithm for Identification of Spillover

Spillover creates detrimental effects for the operation of upstream traffic signals. Identification of spillover is particularly important because it indicates that traffic congestion has started to spread out in the network involving multiple intersections. To identify spillover using traffic signal data, we first need to illustrate the concept of Queue-Over-Detector (QOD), *i.e.* the complete occupation of a detector for a relatively long time due to a vehicular queue.

5.2.1 Queue-Over-Detector (QOD)

When a vehicular queue spills back to the location of a detector at a signalized link, one of the vehicles may stop and "sit" on the detector for a period of time. This relatively prolonged time period can affect detector occupancy values significantly. We call this phenomenon "Queue-Over-Detector" (QOD). Identifying the causes of QOD gives us a means to evaluate the severity of a traffic spillback.

Generally, there are two types of QOD. QOD-I is caused by cyclic red signal phases, which cause vehicles to slow down and stop before resuming their travel when the light turns green and the queuing wave propagates back. If one vehicle in the queue "sits" on the detector because of a red light, occupancy time recorded by detector could increase continuously. The second type QOD (QOD-II) is caused by spillover. When a queue

spills back upstream from a downstream intersection, the upstream intersection may be blocked and vehicles cannot discharge even when the signal is green. Some vehicles will remain stopped on the detector for a relatively long time, creating prolonged detector occupancy time. QOD-II represents an arterial link that is truly congested, while QOD-I only reflects the presence of transient queues.

5.2.2 Relationship between QOD & Cycle-based Arterial Fundamental Diagram (AFD)

To better understand the effects of QOD, we experimentally examined the relationship between QOD and the cycle-based AFD, which describes cycle-based flow-occupancy relationship (Wu *et al.*, 2010b). We studied the impacts of QOD on AFD, because we wanted to understand the relationship between QOD and congestion, especially queue spillover; and congestion can be clearly identified in the fundamental diagram.

Due to frequent stop-and-go, cycle-based flow-occupancy diagrams are greatly scattered. To illustrate, we extracted one-day (Nov. 17th, 2008) loop detector data from a stop-bar detector (Detector #1) and an advance detector (Detector #2) at the intersection of Rhode Island Ave (see Figure 5-2) and then plotted the data, as shown in Figure 5-3a and b. Both diagrams are greatly scattered (AM data (7:00am ~ 9:00am) and PM data (3:30pm ~ 6:30pm) are colored in purple and yellow respectively for comparison).

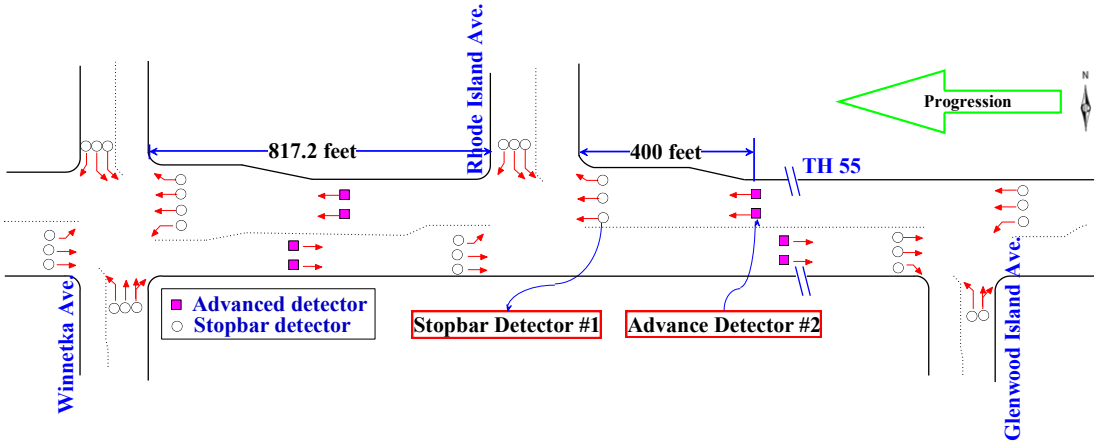

Figure 5-2 Data Collection Site on TH55 in Minneapolis, Minnesota

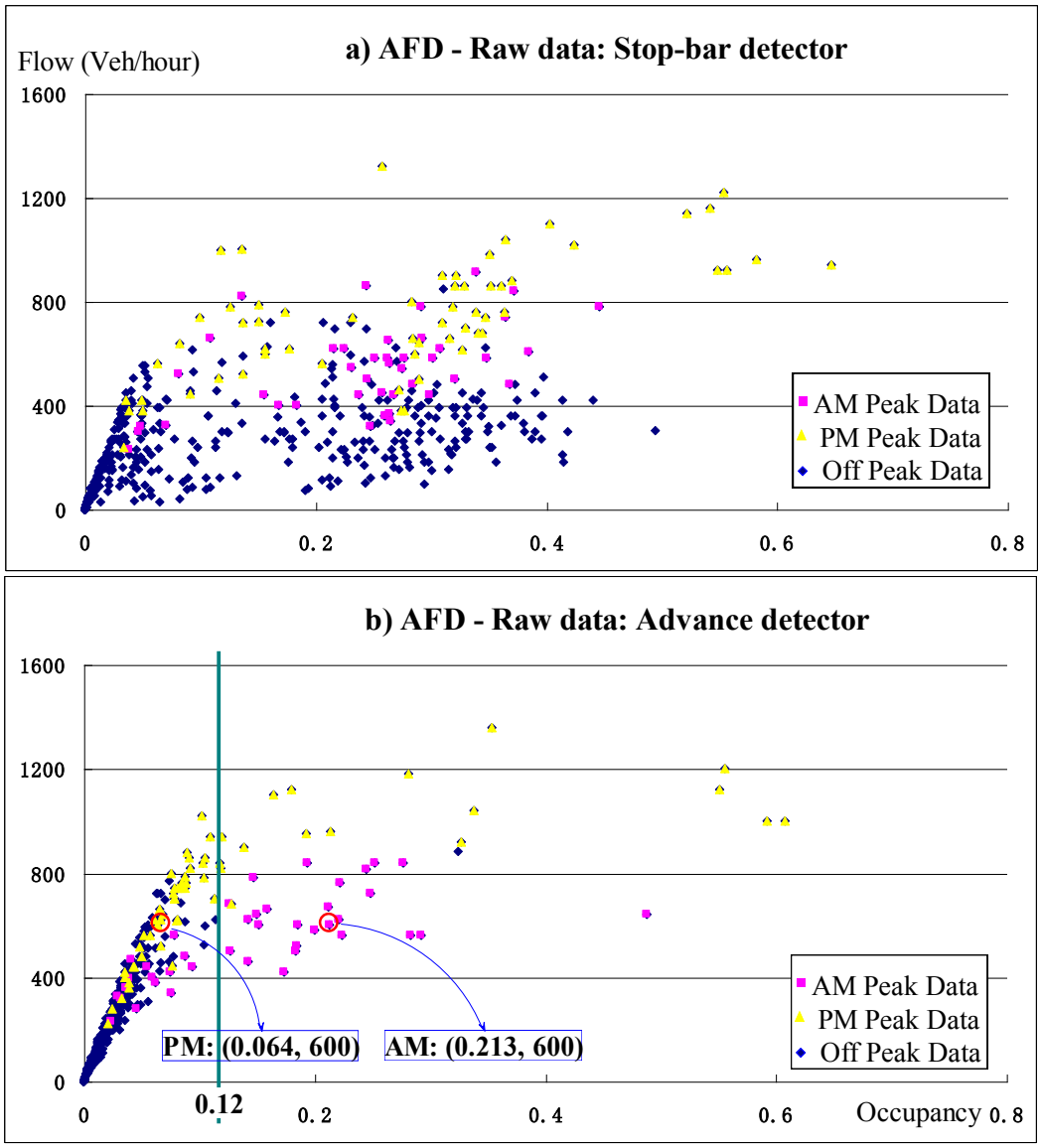

Figure 5-3 Cycle-based AFDs: (a) Data Collected by Stop-bar Detector; (b) Data Collected by Advance Detector

The scatter is presented by some data points with low flow rates staying in the middle region of the diagram. To understand why these points stay in the middle region, we can explain the reasons for the scatter in the diagram; and we also can find whether these points with relatively high occupancy values indeed represent that the arterial link is congested. So we carefully compared two data points in the flow-occupancy diagram

with the same flow (600 veh/h) but different occupancy values (0.213 vs. 0.064) as highlighted in Figure 5-3b. Apparently, aggregated macroscopic traffic flow data was not enough. More detailed microscopic vehicle trajectories were required. We thereby derived detailed microscopic vehicle trajectories. The vehicle trajectories were estimated using event-based data and application of a simplified car-following model, similar to the one proposed by Newell (2001). In the simplified model, lane-changing behaviors were not taken into account because we were only estimating the trajectories for a relatively short distance (for example, from the location of an advance detector to a stop line). Vehicles were assumed to accelerate if their speeds are less than free flow speed (55mph, in this study), and decelerate if speeds were higher than free-flow. The distance between any two vehicles had to satisfy a safety distance constraint, which was determined by current speeds of two consecutive vehicles, reaction time (about 1.0 sec), and stop distance (assumed 25 ft). Finally, a vehicle's decision whether to pass an intersection or not during yellow time was modeled as a function of its current state including its speed, yellow time left, distance to the front vehicle, *etc.*

We now derived all vehicle trajectories for the two cycles with flow-occupancy data (600, 0.213) and (600, 0.064) based on the model described above. As shown in Figure 5-4, during the cycle with flow-occupancy data (600, 0.064) (Figure 5-4b), only a small number of vehicles arrived during the red phase and the queue is not long enough to occupy the advance detector, thereby creating 0% occupancy until the next vehicle arrives during green. By contrast, in the cycle with flow-occupancy data (600, 0.213) (Figure 5-4a), more vehicles arrive during red creating a long queue; and one vehicle stops at the advance detector (QOD-I) generating an occupancy value of 100% until the discharge shockwave propagates back to the detector. Note that these two cycles have same flow, but the dramatic change in occupancy values shifts some data points with relatively small flows from the low occupancy area in the AFD to the high occupancy region, creating occupancy fluctuations.

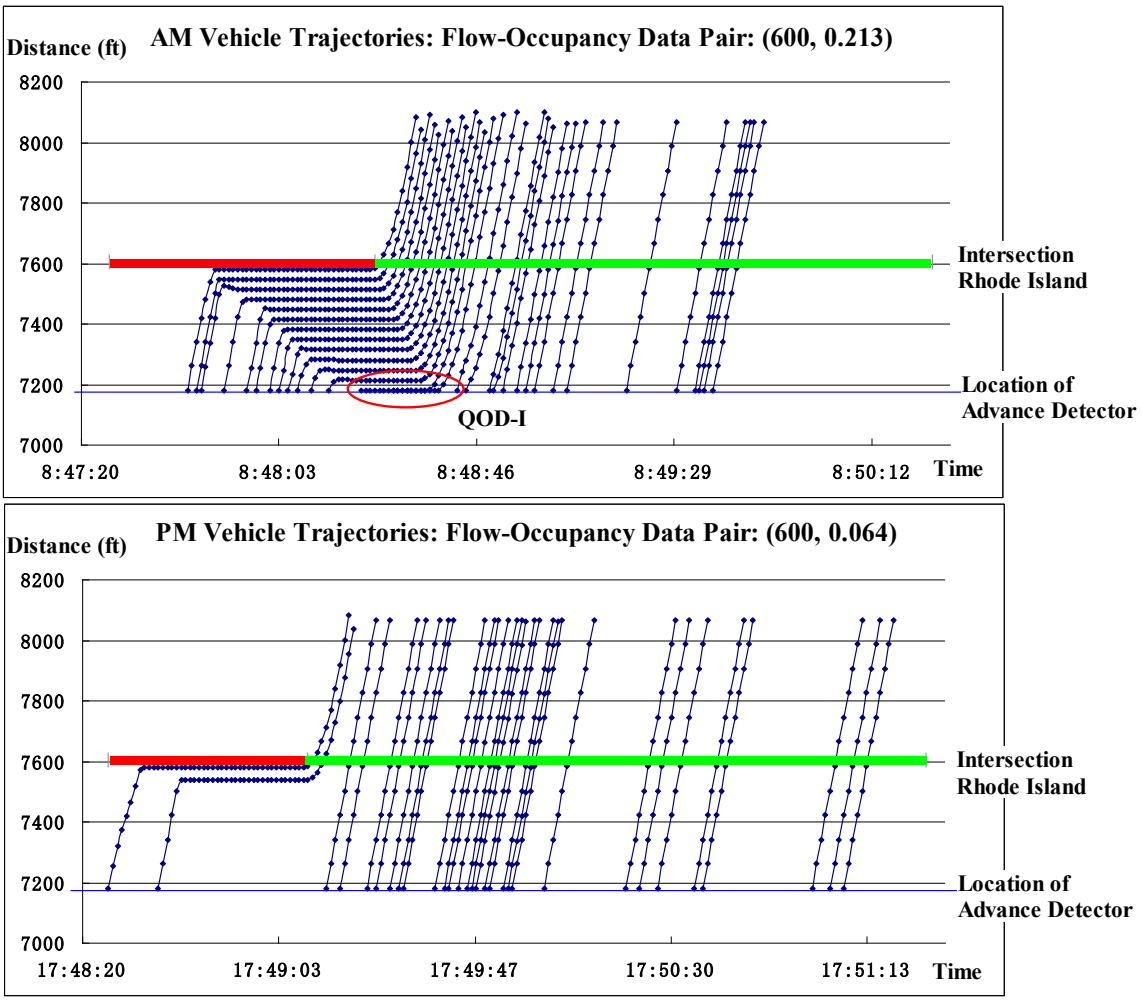

Figure 5-4 Vehicle Trajectories for Two Cycles: (a) Data Point at AM - (600, 0.213); (b) Data Point at PM - (600, 0.064)

The above experiment indicates that QOD-I is the main cause of the occupancy expansion in the AFD. This finding is further verified by a simple experiment that removes QOD-I's impact. In this experiment, we simply move the vehicle stopped on the detector a little bit forward or backward for those cycles with a long queue (*i.e.* queue length longer than the distance from the stop line to the advance detector). The overall flow and queue length do not change, but since the detector is not occupied by any vehicle now, occupancy suddenly changes from 100% to a smaller value near 0% for a period of time, reflecting the removal of the QOD-I factor.

We need to emphasize here that this experiment only removes the occupancy expansion caused by QOD-I. If QOD is caused by spillover, its impact should not be removed. In detail, the occupancy time created by the queues because of red signal is replaced by a normal occupancy time equivalent to the time required by a free-flowing vehicle that passes detector. The removed time period starts when a vehicle arrives at the detector and ends when the discharge shockwave propagates to the detector (*i.e.* the time interval between point B and C in Figure 5-5). Note that the maximum time interval which can be removed is the length between two shockwaves (*i.e.* between points A and B in Figure 5-5, which is equivalent to one red interval). If vehicles arrive before time A or stay on the detector after time B, spillover from the downstream intersection occurs. In such situations, only the portion within A and B can be replaced by a regular passing time.

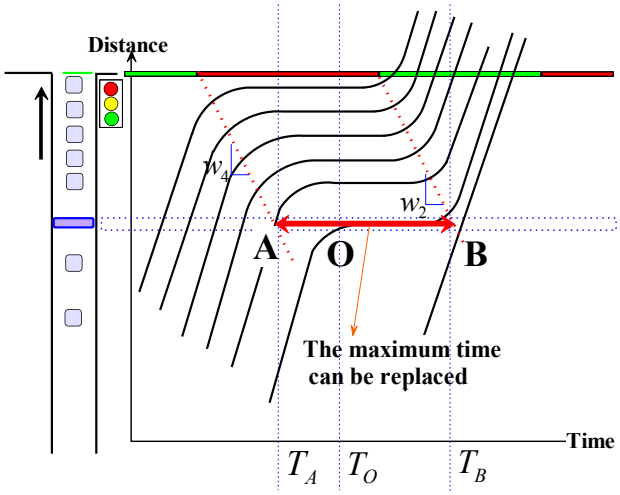

Figure 5-5 Maximum Replaceable QOD-I Time

With our high-resolution event-based data, we can determine the occupancy expansion for each cycle by Eq.(5.7). Essentially, the expanded occupancy time is the interval between time T_O and T_B in Figure 5-5.

$$Occ_{i,n}^{\exp} = \frac{T_B - \max(T_A, T_O)}{C_{i,n}} \tag{5.7}$$

where $Occ_{i,n}^{\exp}$ is the expanded occupancy within the i^{th} cycle at the n^{th} intersection; $C_{i,n}$ is the cycle length for the i^{th} cycle at the n^{th} intersection; and T_A, T_O, and T_B are time instants as presented in Figure 5-5.

Assuming the velocity of discharge shockwave (w_2) and compression shockwave (w_4) is a known constant, then:

$$Occ_{i,n}^{\exp} = \frac{\left(T_{r,n}^{i} + \frac{L_d}{w_2}\right) - \max\left(\left(T_{g,n}^{i} + \frac{L_d}{w_4}\right), T_O\right)}{C_{i,n}}, \tag{5.8}$$

where $T_{r,n}{}^{i}$ and $T_{g,n}{}^{i}$ are the time instants when the effective red and green phases end for the i^{th} cycle at the n^{th} intersection; L_d is the distance between the stop-bar and the location of the advance detector; and w is the velocity of discharge shockwave (see Figure 5-5).

We then applied the QOD-I removal procedure to the raw flow-occupancy diagram previously shown in Figure 5-3. The revised diagrams for both the advance and stop-bar detectors are presented in Figure 5-6. Compared with the diagrams drawn from raw data (Figure 5-3), the AFDs after removing QOD-I are significantly changed. In particular, the highly scattered cloud in the flow-occupancy diagram for both detectors disappears, and the two revised flow-occupancy diagrams now have a similar shape.

Our analysis above shows that the plotting of some data points in the raw AFD in the congested regime with high occupancy values does not necessarily indicate that the signal link is saturated or oversaturated due to QOD-I. It may only mean local congestion in the proximity of the detector, but no occurrence of downstream queue spillback. Only the right hand side of the revised diagrams, in which the QOD-I's impact has been removed, can be said to adequately represent a congested situation caused by blockage on a downstream link.

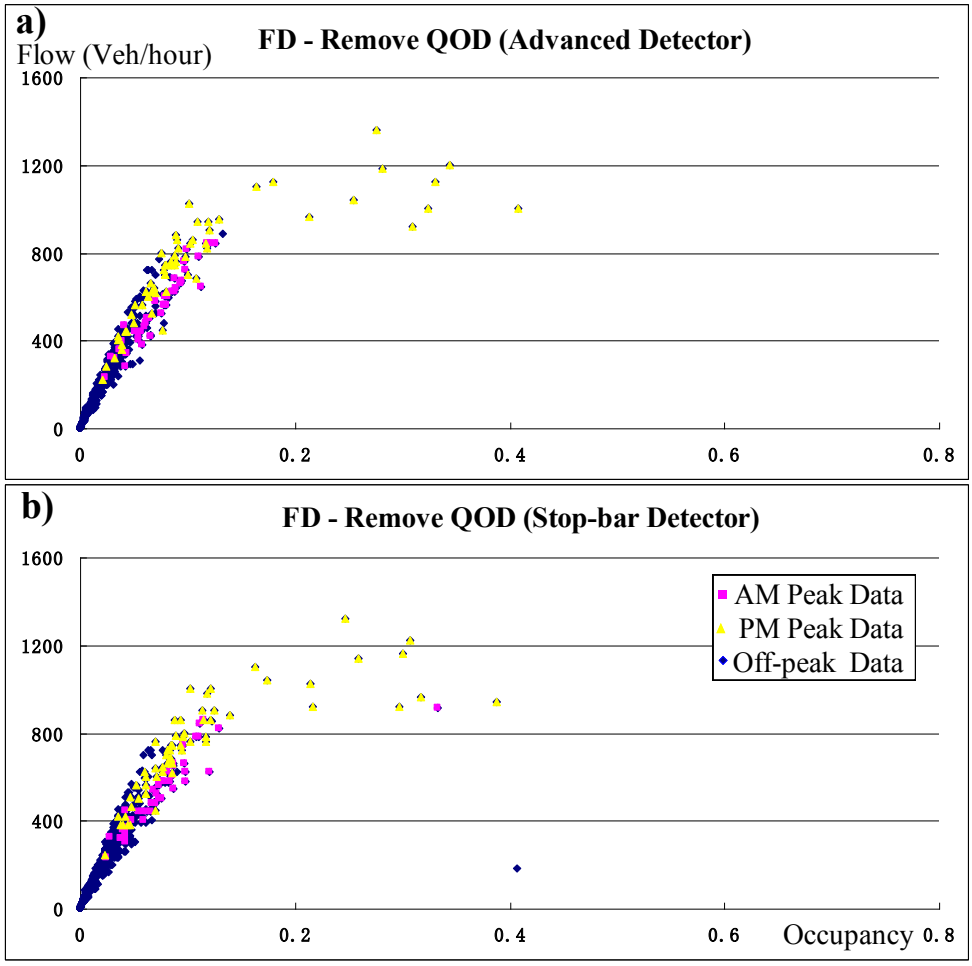

Figure 5-6 AFDs after Removing QOD-I: (a) Data Collected by Advance Detector; (b) Data Collected by Stop-bar Detector.

Another important benefit of QOD-I removal is that it unveils an orderly and stable form of AFD. The left-hand side of the AFD indicates under-saturated traffic and the right-hand side implies oversaturation, specifically, spillover from downstream intersections. Such an AFD is also observed using two-week's historical data collected from Rhode Island intersection (Figure 5-7). The appearance of the diagram seems to suggest a trapezoidal form, which is consistent with traditional understanding of the AFD. However, due to limited data source in congested regimes, we are not able to clearly identify the shape of the right hand side in the AFD. This deserves more effort and is left for future research.

Figure 5-7 also indicates that the data collected from advance and stop-bar detectors generate an almost identical form of the AFD (after removing QOD-I). This is very significant because, as traffic in a signalized arterial link is inhomogeneous, the form of the AFD usually depends on where the data is collected, and how the data is aggregated. That is why the raw AFDs generated by advance and stop-bar detectors differ so significantly (Figure 5-7a & b). But by applying QOD-I removal, we uncover a stable form for the AFD. The diagram now essentially represents the traffic characteristics of an arterial link, rather than just that of the street nearest the detector. In other words, the data can be collected from a set of different detectors (stop-bar, advance or entrance), possibly widely spaced, and generate the similar "answers" about the conditions of the traffic, as long as the impact of QOD-I is removed.

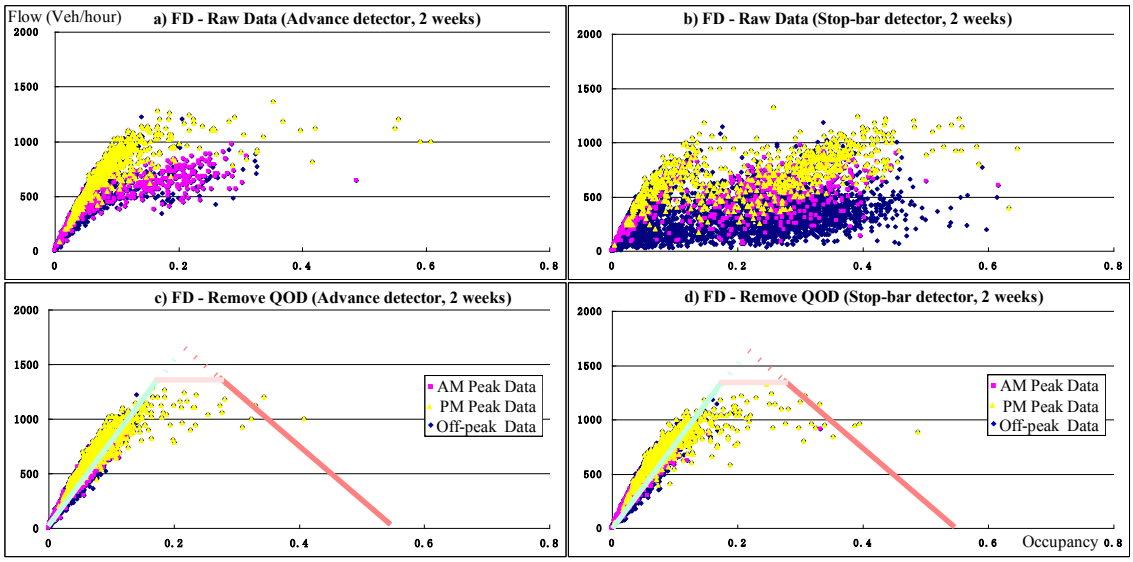

Figure 5-7 A Stable AFD (Based on Two-Week's Data: From Nov. 10th, 2008 to Nov. 21st, 2008, weekdays' data only)

Essentially, what our study of the relationship between QOD and AFD reveals is that true oversaturation states are characterized by spillover, or QOD-II. This observation

suggests a potentially important approach to quantify the severity of oversaturation as will described in the following section.

5.2.3 Quantifying Spillover using High-resolution Data

As found in the last section, QOD-II represents spillover. Therefore the detrimental effect of a spillover can be quantified by measuring the duration of QOD-II.

It is not difficult to identify QOD using high resolution data, since it is evident by a relatively large occupancy time (or percentage occupancy value remaining at 100% for some time). In our implementation, we identify QOD using a threshold value of 3 seconds (roughly equivalent to 5mph of speed assuming a 22ft effective vehicle length).

We need to first distinguish two types of QOD so that we can eliminate the one that does not indicate spillover. Figure 5-8 demonstrates both types of QOD with drawings of each vehicle trajectory starting from upstream to downstream. Since QOD-I is caused by a red signal, the maximum occupancy time is the red signal interval. When we consider the residual queue from the last cycle and the queue propagation at the green start, QOD-I can only occur within the range of time between two parallel shockwaves (compression shockwave v_4 and discharge shockwave v_2, which have the same velocity, see Figure 5-1). Therefore if QOD occurs between ($T_g^i+L_d/w_4$) and ($T_r^i+L_d/w_2$), it is of the first type.

QOD-II occurs outside of the time interval [$T_g^i+L_d/w_4$, $T_r^i+L_d/w_2$]. Therefore, if the QOD starting time T_{start}^{QOD} or ending time T_{end}^{QOD} falls outside of the time interval [$T_g^i+L_d/w_4$, $T_r^i+L_d/w_2$], it is a QOD-II event. QOD-II indicates that a spillover has happened at a downstream location. This creates unusable green time, meaning that vehicles cannot be discharged during the green time because of downstream congestion. When a QOD-II event is identified, the event is classified as spillover. An example of QOD-II is demonstrated in Figure 5-8.

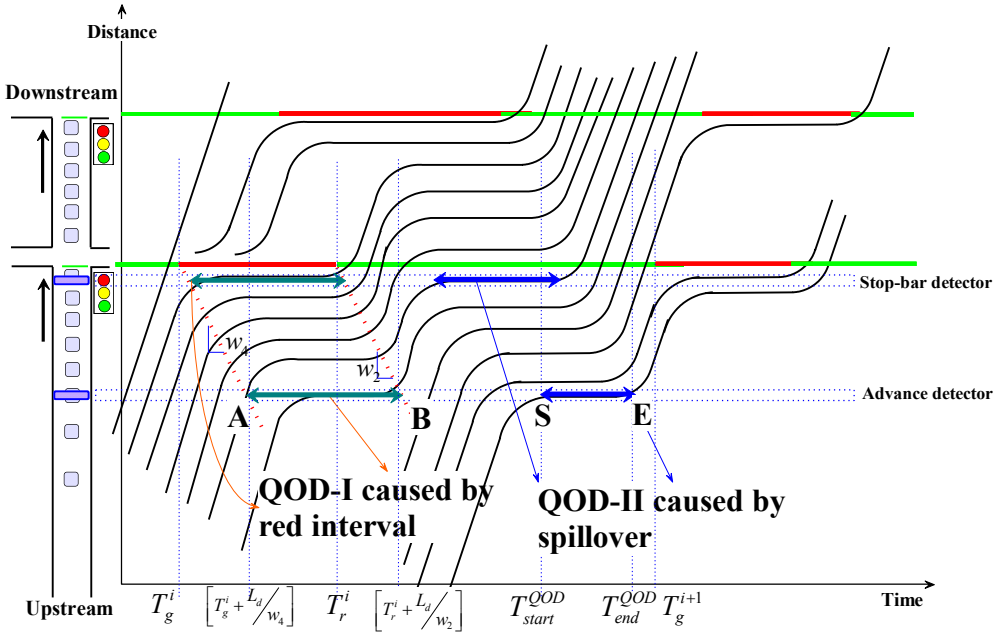

Figure 5-8 Two Types of Queue-Over-Detector Events

Once QOD-II is identified, we can measure the severity of a spillover using the oversaturation severity index in space dimension (S-OSI):

$$\text{S-OSI} = \frac{\textit{unusable green time}}{\textit{total available green time}} \times 100\% = \frac{\sum\left(T_{end,j}^{QOD} - T_{start,j}^{QOD}\right)}{G} \times 100\% \qquad (5.9)$$

where $T_{start,j}^{QOD}$ and $T_{end,j}^{QOD}$ are the starting and ending times of the j^{th} occurrence of QOD-II.

In order to improve the robustness of spillover identification, we should also estimate the maximum queue length of the downstream intersection. If the estimated maximum queue length is longer than or equal to the link length, then oversaturation is confirmed. This verification process can be used to avoid some diagnosis errors caused by "incidents" (for example, a detector occupied by a broken vehicle for a relatively long time). These "incidents" may generate QOD-II, but do not necessarily indicate an oversaturated condition.

5.2.4 Queue Length Estimation with Downstream Spillover

We should note that the queue length estimation method discussed in Section 5.1 cannot be applied directly to an intersection with queue spillover originating downstream. With spillover, queued vehicles can only be discharged when the downstream blockage is cleared and while the signal light remains green. For the example shown in Figure 5-9, when the traffic light turns green at intersection n ($T_r^{i,n}$), queued vehicles start to discharge, but the discharging process is disturbed because the queue at downstream intersection ($n+1$) is growing and eventually spills over to the upstream intersection n. The advance detector will identify a QOD-II, starting at time T_{start}^{QOD} and ending when the spillover is cleared, *i.e.* at time T_{end}^{QOD}. Under such oversaturated conditions, the queue estimation method needs to be modified because there is a new shockwave profile generated. Before spillover happens, we first indentify break points A, B, and C and then estimate the coordinates of point H and D. But after spillover happens, we need to re-identify break points A', B', and C' (as shown in Figure 5-9), then recalculate the new maximum queue (point H') and minimum queue (point D'). Original equations (Eq.(5.1)-(5.5)) will remain valid for spillover cases.

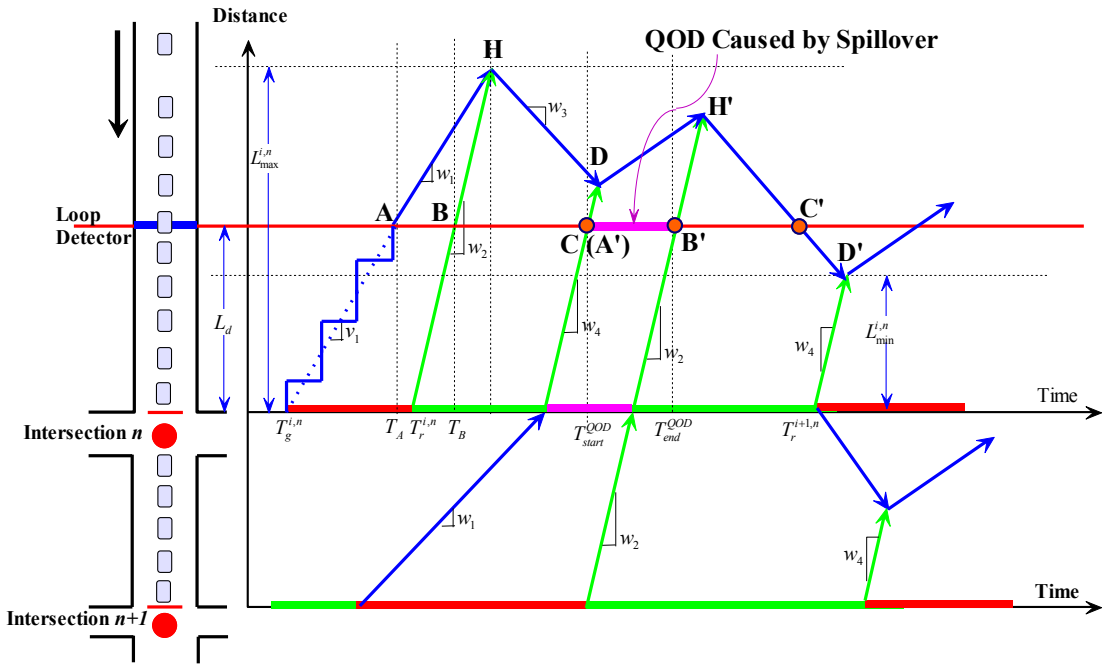

Figure 5-9 Shockwave Profile with Downstream Spillover

5.3 Field-Testing & Results

5.3.1 Data Collection

We conducted field tests to verify our approaches of the quantification of OSI. The test site was six intersections along Trunk Highway 55 (TH55) in Minneapolis, Minnesota. Figure 5-10 shows the detector layout of four of the intersections where oversaturation has been known to recur. The six intersections were equipped with vehicle-actuated traffic controllers, with advance detectors on the major approach and stop-bar detectors on the minor approach, for the purpose of traffic signal operations. Stop-bar detectors were used to detect the presence of vehicles and advance detectors were located about 400 ft upstream from the stop line to detect vehicles for green extension on the coordinated phases. For the purpose of verification of the estimated queue length, we have installed stop-bar and link entry detectors along TH55 at the six intersections (see Figure 5-10 for the detector configurations at the four intersections of interest). These additional detectors were not used for regular traffic signal operations, but for data collection to provide further evidence for the results presented below.

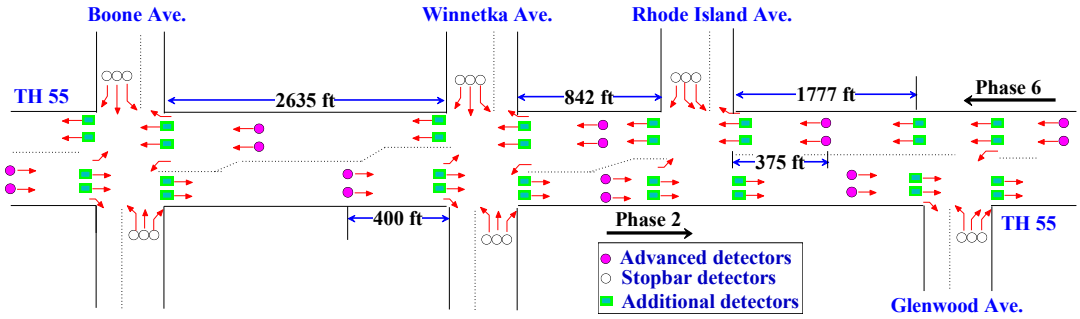

Figure 5-10 Detector Layout in Data Collection Site

High-resolution event data including signal phase changes and vehicle-detector actuations were continuously collected from the six intersections and archived by the SMART-Signal system and then transmitted back to the University of Minnesota lab in real-time. The event data were used to derive signal timings, occupancy times, and time gaps between two consecutive vehicles crossing the detector.

5.3.2 Results of Residual Queue Length Estimation

As introduced in Section 5.1, we apply a residual queue length estimation method to quantify T-OSI. This section presents some oversaturation cases to demonstrate how to calculate T-OSI.

Figure 5-11 presents an oversaturation case based on the data collected by an advance detector in the eastbound direction at the intersection of Glenwood Ave. on Feb. 28th, 2008. As can be seen, residual queues appeared at the end of the first two cycles, meaning that these two cycles were oversaturated. In this particular case, oversaturation was due to signal preemption, which created a shorter cycle length in the second cycle (the cycle length was 132 seconds during the preemption, 48 seconds less than the normal cycle length). Due to the insufficient green time, some queued vehicles could not be discharged until the next cycle, creating detrimental effects on the following cycles.

Using the residual queue length estimated by our method, we calculate OSIs for these two cycles of 7.5% and 7.0%, meaning that at least 7.5% and 7.0% of green time in these cycles was needed for the discharge of the residual queue. In the calculation of these severity indices, we assumed a space headway of 25ft for the jammed condition and a saturation time headway of two seconds.

It should be noted that the estimated maximum queue lengths (500 - 600ft) during the three cycles were not long when compared with the link length (1777ft from Glenwood to Rhode Island). However, the fact that residual queuing occurred at the end of the first two cycles despite these relatively short queue lengths tells us that the traffic volume joining the discharge platoon after the last stopped vehicle had started to move was rather high. A portion of those newly arriving vehicles joined the discharge platoon but could not pass the intersection during the green phase, thus forming the residual queue.

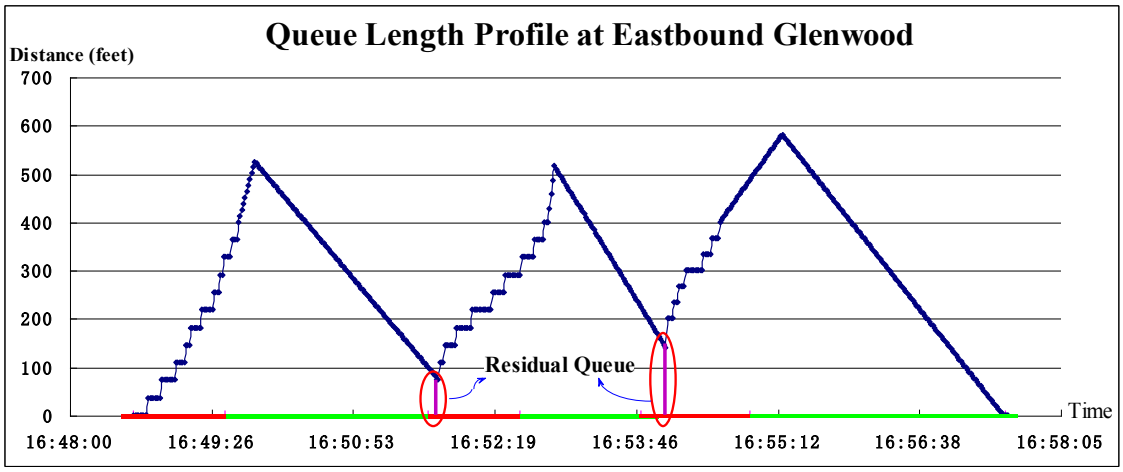

Figure 5-11 Estimation Results of Residual Queue for Eastbound Approach at Glenwood

As we discussed previously, if the entire green time is used for queue discharge, the departure shockwave cannot be identified. In this case we can only measure the minimum of the maximum queue length within a cycle. We present 5 cases in Figure 5-12. The data was collected from an advance detector in the eastbound direction at the intersection of Boone Ave. on Feb. 28th, 2008. During the first five cycles, the departure shockwaves could not be identified, so we can only estimate the minimal values of the maximum queue length. The queue lengths in this case are quite long, averaging around 1500ft in the first five cycles. The minimum values of the residual queue length are also estimated, as shown in Figure 5-12. The minimal oversaturation severity indices (T-OSIs) are estimated at 9.8%, 19.4%, 10.5%, 11.3%, and 10.3% for these five cycles. The oversaturated condition persists until the sixth cycle.

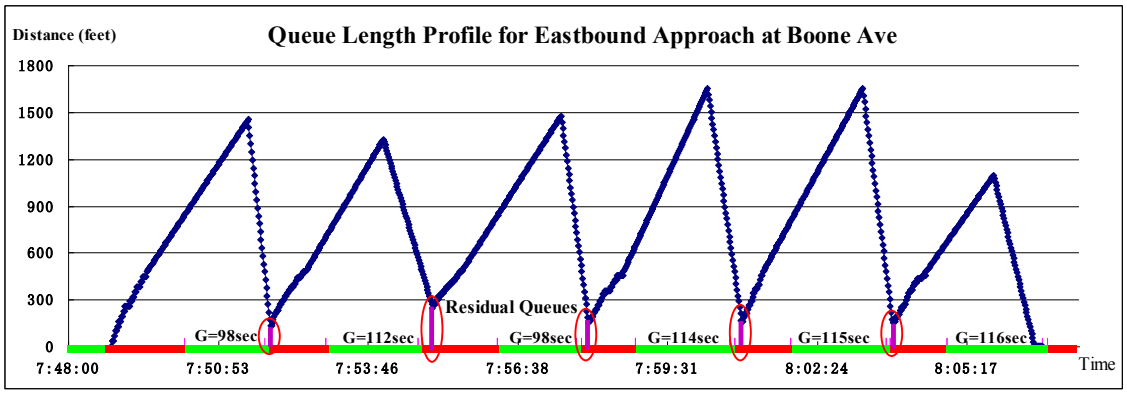

Figure 5-12 Estimated Residual Queue Length for Eastbound Approach at Boone Ave.

5.3.3 Results of Spillover Detection

As introduced in Section 5.2, we need to identify QOD-II in order to quantify S-OSI. This section presents some cases to demonstrate how to identify QOD-II and how to calculate S-OSI.

We use occupancy time recorded by detectors to identify QOD-II, as discussed in Section 5.2. In Figure 5-13, we present the detector occupancy time within a cycle for an afternoon peak hour on Nov. 17th, 2008 for westbound TH55 at Rhode Island Ave. As shown in the figure, QOD-II is found. This means that vehicles cannot be discharged from the intersection although the traffic light is green, due to a spillover happening in the downstream link. Oversaturation is therefore identified at this intersection for this cycle.

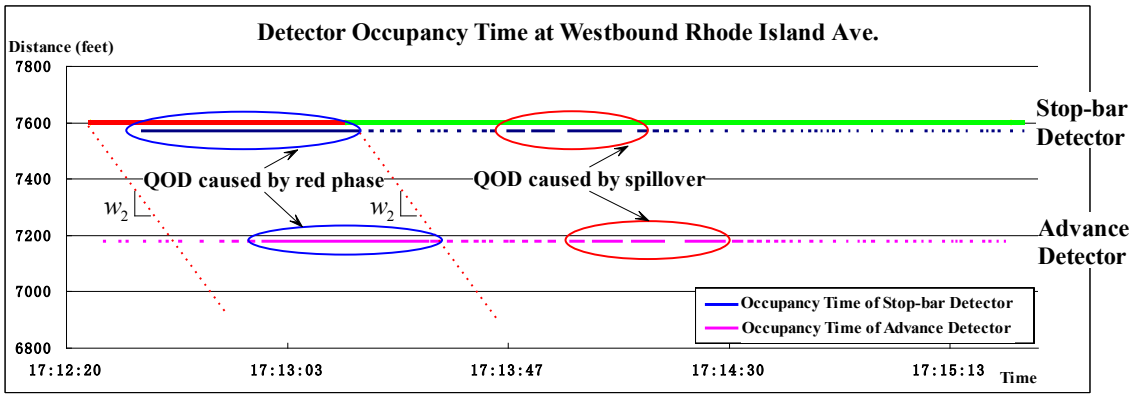

Figure 5-13 Identification of QOD-II using Detector Occupancy Time Data

To further verify that there is a spillover happening in the downstream link, vehicle trajectories were derived based on the vehicle events collected by the advance detector at the same intersection of Rhode Island Ave. The estimated vehicle trajectories starting from the advance detector line at the intersection of Rhode Island Ave. and ending at 500 ft downstream from the intersection of Winnetka Ave. are presented in Figure 5-14. As clearly indicated in the figure, the queue spills back from Winnetka to Rhode Island and blocks the Rhode Island intersection during green time, confirming a QOD-II event.

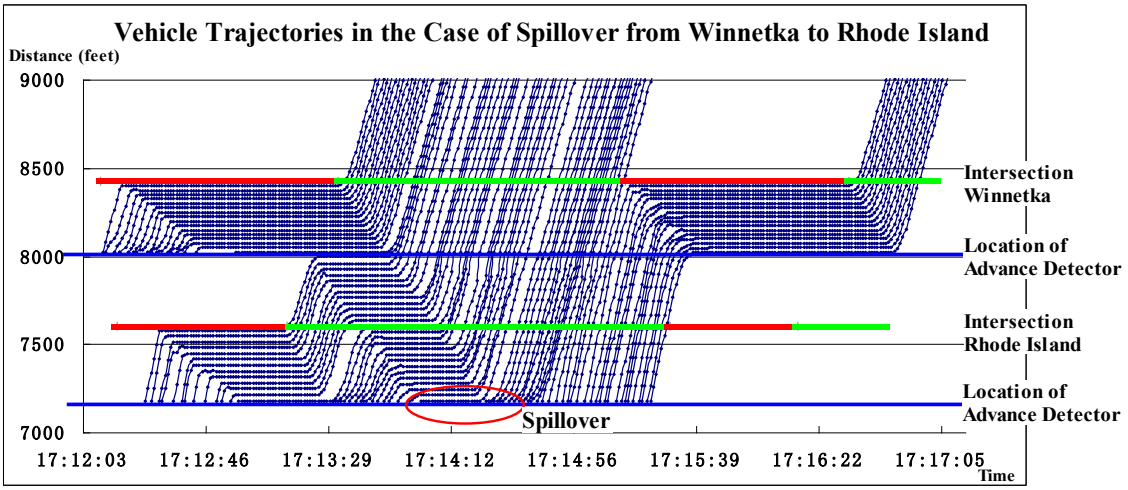

Figure 5-14 Vehicle Trajectories in the Case of Spillover from Winnetka to Rhode Island

We also applied AFD to identify spillover cycles. As presented in Figure 5-15, after QOD-I removal, there are still 9 cycles (indicated by letters A – I) staying in the congested regime. We further investigated the event data for these 9 cycles and found QOD-II consecutively occurred in these cycles. From the figure, we can see that the spillover started at 17:06:31, stayed about 30 minutes, and ended at 17:36:31. Spillover was further confirmed by looking at the queue length profiles at the downstream intersection at Winnetka Ave. during these cycles. As shown in Figure 5-16, the maximum queue lengths for these 9 cycles are around 1200 – 1500ft, which is significantly longer than the link length (842ft). This indicates that, during these cycles, the Rhode Island intersection must have been blocked for a portion of green time (S-OSI > 0). Interestingly, because of the reduction of usable green time, residual queues were also generated at the Rhode Island Intersection for some cycles (*i.e.* T-OSI > 0). This suggests that an oversaturated traffic condition at Winnetka (T-OSI > 0 and S-OSI = 0) had spread upstream, leading to insufficient green at Rhode Island Intersection to discharge the queue (T-OSI > 0 and S-OSI > 0). Please see Table 5-1 and Table 5-2 for oversaturation severity indices at Winnetka and Rhode Island. Note that for Winnetka, since there was no downstream blockage, S-OSI was always zero during that time period.

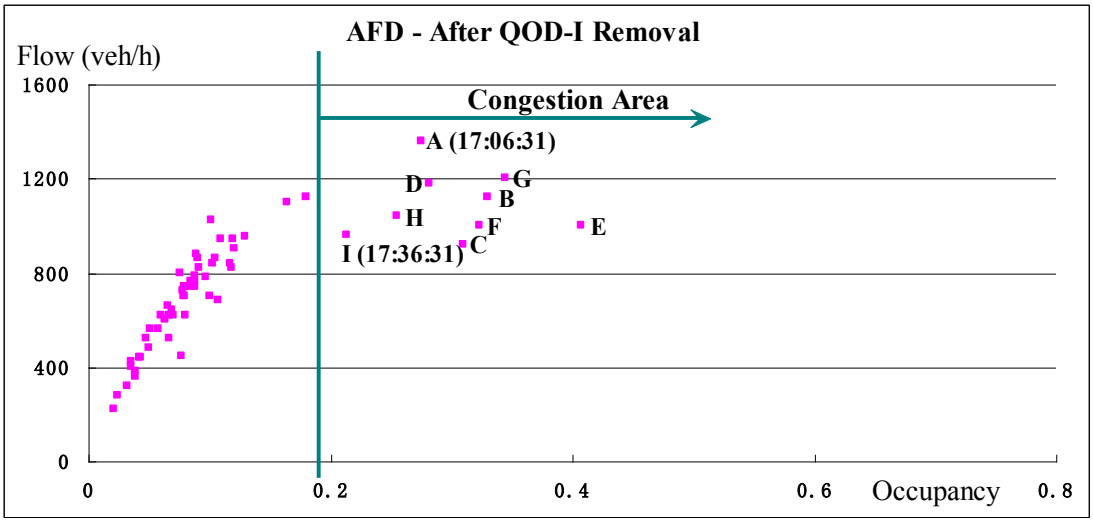

Figure 5-15 AFD – After QOD-I Removal

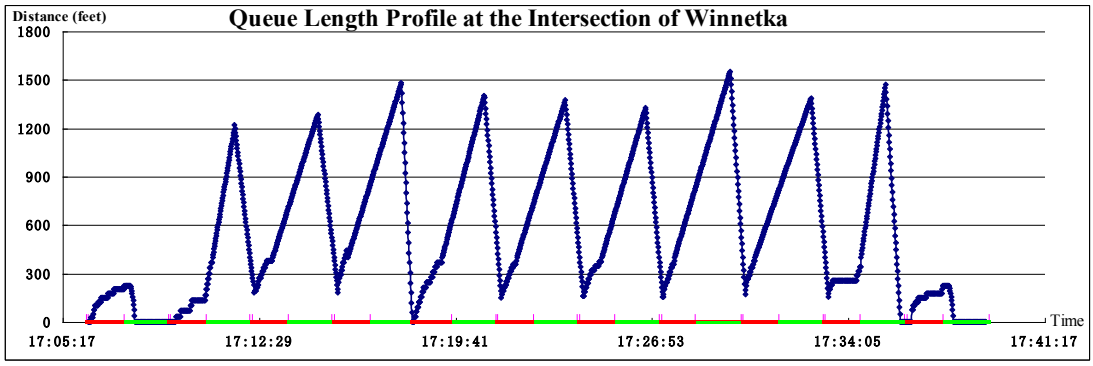

Figure 5-16 Queue Length Profile at the Intersection of Winnetka

Table 5-1 Oversaturation Severity Indices (OSI) for Winnetka Ave. Intersection

Winnetka Ave						
Cycle Start	Available Green (sec)	OSI: Created by Residual Queue			OSI: Created by Spillover	
		Residual Queue (ft)	Unusable Green (sec)	T-OSI (%)	Unusable Green (sec)	S-OSI (%)
17:06:14	101	0.0	0.0	0.0	0.0	0.0
17:09:14	101	180.3	0.0	0.0	0.0	0.0
17:12:14	101	178.8	14.4	14.28	0.0	0.0
17:15:14	101	0.0	14.3	14.16	0.0	0.0
17:18:14	101	149.1	0.0	0.00	0.0	0.0
17:21:14	101	157.6	11.9	11.81	0.0	0.0
17:24:14	102	156.4	12.6	12.36	0.0	0.0
17:27:14	106	130.1	12.5	11.81	0.0	0.0
17:30:14	101	153.4	10.4	10.31	0.0	0.0
17:33:14	105	0.0	12.3	11.69	0.0	0.0
17:36:14	102	0.0	0.0	0.00	0.0	0.0

Table 5-2 Oversaturation Severity Indices (OSI) for Rhode Island Intersection

Rhode Island						
Cycle Start	Available Green (sec)	OSI: Created by Residual Queue			OSI: Created by Spillover	
		Residual Queue (ft)	Unusable Green (sec)	T-OSI (%)	Unusable Green (sec)	S-OSI (%)
17:06:31	136	0.0	0.0	0.0	0.0	0.0
17:09:31	136	0.0	0.0	0.0	3.0	2.2
17:12:31	136	89.6	0.0	0.0	28.0	20.6
17:15:31	136	164.3	7.2	5.3	28.8	21.2
17:18:31	136	0.0	13.1	9.7	15.0	11.1
17:21:31	136	180.4	0.0	0.0	41.7	30.6
17:24:31	135	165.3	14.4	10.7	34.1	25.2
17:27:31	139	138.2	13.2	9.5	25.2	18.1
17:30:31	120	125.3	11.1	9.2	16.3	13.6
17:33:31	141	0.0	10.0	7.1	8.6	6.1
17:36:31	135	0.0	0.0	0.0	0.0	0.0

5.4 Summary

In this chapter, we proposed two algorithms to identify oversaturated signalized intersections, one for the estimation of residual queue length using a shockwave based method and the second for detection of spillover by identifying long detector occupancy time during green phase. We defined an oversaturation severity index using the unusable green time caused by residual queue or spillover. Our field-test results from a major arterial in the Twin Cities area demonstrate that the developed algorithms are very effective in identifying oversaturated conditions. We should note that, although the identification algorithms focus on a single approach of a signalized intersection, the proposed methodology can be easily expanded to identify oversaturation of an intersection with multiple approaches, an arterial, or a network of intersections.

This chapter also briefly studies the relationship between QOD and the cycle-based AFD by microscopically investigating individual vehicle trajectories derived from event-based data. We show how a stable form of AFD can be produced which clearly identifies three different regimes: under-saturation, saturation, and over-saturation with queue spillovers. The ability to represent traffic states on a signal link is of great importance for traffic signal control.

6 Traffic Flow Modeling for Oversaturated Arterials

In the previous chapters, we talked about how to quantify the severity of oversaturation using high-resolution data. This chapter aims to develop a traffic flow model, which will serve as a foundation of all signal optimization schemes. This model should be able to provide a quick and approximate approach, with sufficient descriptive power, to simulate oversaturated traffic flow dynamics in a signalized network for real-time applications.

6.1 Background

Over the last fifty years, following the seminal works of Lighthill & Whitham (1955) and Richards (1956), continuum traffic flow theories have been studied by many researchers. For a recent review of current continuum traffic flow models, we refer to Zhang (2001). Although LWR-type models have been criticized for their inability to account for the acceleration and deceleration process of traffic flow, it has commonly been argued that simple continuum models are sufficient to describe traffic behavior in signalized networks, because, as one author puts it, "traffic flow dynamics are dominated by external events (red traffic lights) rather than by the inherent traffic flow dynamics" (Papageorgiou, 1998). An extensive literature search also indicates that compared to the continuous and intensive efforts to model freeway traffic flow, far less work has been done to understand dynamics of arterial traffic flow. A notable exception is Michalopoulos *et al.* (1980), in which an analytical solution was derived for describing the evolution of the queue length on a signalized link, based on the continuum principle and method of characteristics. Despite the theoretical appeal of the method of characteristics, there has been almost no field testing to evaluate the potential

effectiveness of this approach. As has been repeatedly stressed in the literature (Beskos *et al.*, 1984; Yi *et al.*, 2001; Liu *et al.*, 2009), the main obstacle is the lack of comprehensive, large scale, and detailed traffic data on signalized arterials.

Other obstacles that have confronted researchers are numerical errors and computational efforts. For numerical reasons, the solutions of LWR-based continuum traffic flow models usually involve the discretization of space and time to describe the spatio-temporal variations of traffic flow and density. The well-known cell transmission model (CTM), indicated by its very name, applies the finite difference method to simulate the evolution of traffic density in each cell, whose length is equal to the distance traveled by a free-flowing vehicle in one time interval (Daganzo, 1994, 1995). When CTM is applied to model arterial flow, however, numerical errors often occur because it is not uncommon that a signalized link cannot be decomposed into an integer number of cells. Although this may not be a problem if the cell size is small, using small sizes increases the number of cells for an arterial link, thereby compromising computational efficiency.

In this study, we propose a section-based approach to model arterial traffic flow dynamics, based on the LWR shockwave theory (as introduced in Chapter 4). Instead of using the usual differential approach, however, we integrate traffic over finite road sections. Therefore a signalized road section will no longer be decomposed into uniform cells; rather, each homogeneous road segment with constant capacity is treated as a section and the major shockwaves generated within a section or between two adjacent sections are traced explicitly. We call our approach the shockwave profile model (SPM). We should note that the concept of a "section-based" traffic flow model was initially proposed by Helbing (2003), but it was applied to study freeway traffic flow for travel time analysis.

The purpose of SPM is to provide a quick and approximate approach, with sufficient descriptive power, to model traffic flow dynamics in a signalized network for a number of real-time applications. As indicated in Dell'Olmo and Mirchandani (1996), for real-time applications, signal control strategies need to be evaluated quickly. Therefore, we do

not aim to track traffic densities in each small cell, which, we argue, it is not necessary; rather, we are more interested in the position of the tail of an intersection queue, because queue length is a key indicator for traffic performance at an intersection, from which vehicle delay and level of service can be calculated accordingly. As we observed, for signalized arterials, because of traffic lights, repetitive shockwaves for queue build-up and dissipation clearly separate a road section into three different traffic states: free-flow, saturated, and jammed. The proposed model takes advantage of these simplified traffic states to describe traffic dynamics. Such simplification allows us to derive an analytical solution for queuing dynamics, while reducing computational costs significantly.

SPM is particularly suitable for simulating traffic flow on congested signalized arterials especially with queue spillover problems, where the steady-state periodic pattern of queue build-up and dissipation process may break down. However, depending on when and where the spillover happens along a signalized arterial, a large number of queuing patterns may be possible. Therefore the conventional approach tracking of shockwave fronts, as shown in Michalopoulos *et al.* (1980), becomes tedious and time consuming. To overcome the difficulty of spillover, a novel approach is proposed as part of SPM, in which queue spillover is treated as either extending a red light or creating new smaller cycles, so that the analytical solutions for tracing the shockwave fronts can be easily applied.

SPM has been empirically validated using real-world traffic signal data, which were collected and archived using the SMART-SIGNAL system. We compared estimation results from SPM against the field data for a PM peak during which a nine-cycle queue spillover occurred (Wu, *et al.*, 2010b). The results clearly demonstrate the effectiveness and accuracy of the model.

Section 6.2 of this chapter introduces the SPM model for a signalized approach with or without downstream spillover. A discussion on the issues of model implementation follows in Section 6.3. In Section 6.4, we present a simple numerical example that demonstrates the difference between CTM and SPM in dealing with congested arterial

traffic. Field validation results are given in Section 6.5. Finally, Section 6.6 concludes this chapter with some remarks.

6.2 The Shockwave Profile Model (SPM)

In this section, we present our shockwave profile model to describe the traffic flow dynamics within a signalized arterial link. Without loss of generality, Figure 6-1 shows three one-way signalized intersections, *n-1*, *n*, and *n+1*, connected by two links with lengths L_n and L_{n+1}. These three intersections could be part of a large urban traffic network. For simplicity, we use a one-way network as an example; but the proposed model can handle two-way networks as long as the boundary conditions have been given.

We assume that traffic signal control parameters for all three intersections are known in advance for the time duration [*0, T*], where *T* is the ending time of simulation. For any time $t \in [0,T]$, we let $O_n(t)$ be the start time of the effective red (or the end of effective green) in the current cycle for the through movement of link L_n at intersection *n*, and let $r_n(t)$ and $g_n(t)$ be the effective red and green intervals for the same approach in the same cycle. In Figure 6-1, $q_{n-1}(t)$ represents the flow rate at the entrance of link L_n (*i.e.*, inflow rate) at time *t*, and $\bar{q}_n(t)$ and $\bar{q}_n^m(t)$ the departure flow rates at intersection *n* (*i.e.* outflow rates) at time *t* for the major and minor approaches, respectively.

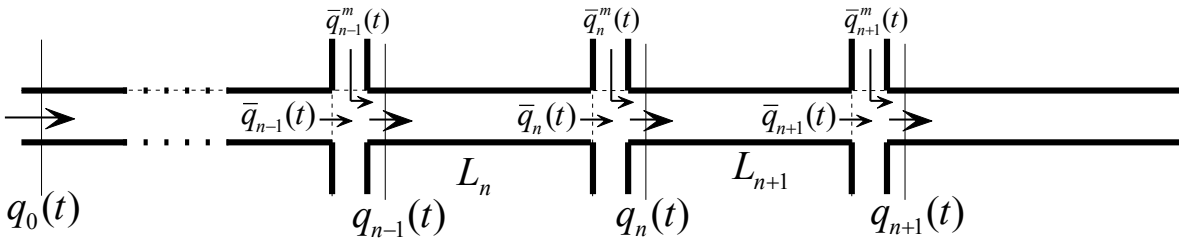

Figure 6-1 Layout of a Signalized Arterial with Three Intersections

To describe traffic flow dynamics, *i.e.*, the trajectories of queue formation and dissipation on signalized links, we assume that the inflow rates at network boundaries are given. For the three-intersection network shown in Figure 6-1, inflow rates at network boundaries including $q_{n-2}(t), \bar{q}_{n-1}^{m}(t), \bar{q}_{n}^{m}(t)$, and $\bar{q}_{n+1}^{m}(t)$ are assumed known. The initial condition of the network (*i.e.* traffic states within each link at the beginning) should also be given *a priori*. We now need to derive the outflow and inflow rates at all the intersections (*i.e.*, $\bar{q}_{n-1}(t), q_{n-1}(t), \bar{q}_{n}(t), q_{n}(t), \bar{q}_{n+1}(t)$ and $q_{n+1}(t)$) in addition to queue trajectories of each link. We should note that here every signalized link is a section, within which the road segment is homogeneous with constant capacity. A link with turning bays will be discussed in Section 6.3.

Similar to previous macroscopic traffic flow models, the foundation of SPM is the flow conservation law. For a signalized link L_n between time t_1 and t_2, assuming no sinks and sources between the link entry and exit, the following flow conservation equation holds:

$$\int_{t_1}^{t_2} q_{n-1}(t)dt + N_n(t_1) = \int_{t_1}^{t_2} \bar{q}_n(t)dt + N_n(t_2) \tag{6.1}$$

where $N_n(t)$ is the number of vehicles within link L_n at time t. Note that, instead of using the differential form of the conservation law as most previous studies do, we use the integral form.

However, traffic dynamics cannot be fully described using solely the flow conservation equation. It must be supplemented by additional relations, such as the flow-density relation (*i.e.* the fundamental diagram) or a momentum equation describing the evolution of traffic speed. Our model, however, is built upon the observation that traffic dynamics at signalized intersections can be represented by a shockwave profile which describes queue build-up and dissipation. Due to the cyclic disruption introduced by traffic lights, there exists a clear pattern of the first-order shockwaves at signalized intersections. These shockwaves clearly divide traffic states into free-flow, saturated, and jammed conditions. SPM therefore takes advantage of the simplified traffic states to describe traffic dynamics based on a shockwave profile.

Consequently, the following assumptions are made: **(1)** vehicles travel at free-flow speed before reaching the tail of a queue; **(2)** vehicles in a queue discharge at the saturation flow rate (when they are not constrained by downstream congestion); and **(3)** the velocity for a queue discharge shockwave is assumed to be known.

It is necessary to differentiate the above assumptions from those underlying traditional fundamental diagram (FD), which hypothesizes a flow-density relationship. Our assumption (1) actually indicates that the left hand side (uncongested area) in the FD is a straight line. This assumption is consistent with many empirical observations, *i.e.*, vehicles are free-flow traveling when traffic is uncongested. The second and third assumptions imply known capacity and jam density, only two points in the FD. Except for these two points, the precise function of the congested regime for the FD does not need to be known. Indeed, till now, the exact shape of the right hand side of the FD has been unknown, especially for signalized arterials (Wu *et al.*, 2010b). Although much research has been devoted to this topic, whether the shape is linear (for example, the piecewise linear function applied in CTM by Daganzo (1994 & 1995)), concave (for example, the parabolic function proposed by Greenshields (1935)), or some other form, is still an open question. This research presented here avoids this question by simplifying congested traffic conditions into two states: saturated or jammed. These assumptions for traffic flow on a signalized link significantly simplify the model design and improve computational efficiency.

6.2.1 SPM for Intersections without Spillover from Downstream Links

The model for an individual intersection without downstream spillover is introduced in this section. SPM is built upon the cyclic shockwave profile on a signalized link. As introduced in Chapter 4, a shockwave is derived when applying the method of characteristics to analytically solve the partial differential equation (PDE) in the LWR model. Basically, when characteristic curves (along which the density is constant) interact, a shockwave is formed and wave velocity can be determined using Eq.(6.2):

$$w = \frac{q_2 - q_1}{k_2 - k_1} \tag{6.2}$$

where q_1 (q_2) and k_1 (k_2) are the traffic flow rate and density of upstream (downstream) respectively.

Due to the cyclic nature of signal phase changes, the shockwave profile at a signalized intersection also has a cyclic pattern. As indicated in Figure 6-2a, at the beginning of red (for better explanation, we assume there is no residual queue at the beginning of a cycle), a **queuing shockwave (w_1)** is generated and propagates backward. The queue reaches its maximum length when the queuing shockwave meets a **discharge shockwave (w_2)**, which also propagates backward from stop-bar after a green light starts. As soon as these two waves meet, a third shockwave called a **departure wave (w_3)** is generated and propagates forward to the stop line. If a queue does not fully discharge by the end of a cycle, a residual queue is formed (Figure 6-2b). The minimum queue will be achieved some time after the start of the red light for the next cycle when the departure wave meets a **compression wave (w_4)**. The compression wave has the same speed as the discharge wave, as both waves are generated from the discontinuity between the saturated and jammed traffic conditions. The velocities of all waves can be estimated using Eq.(6.2). The fact the shockwave patterns in this queuing process repeat from cycle to cycle makes them potentially very useful for simulating traffic dynamics in signalized settings.

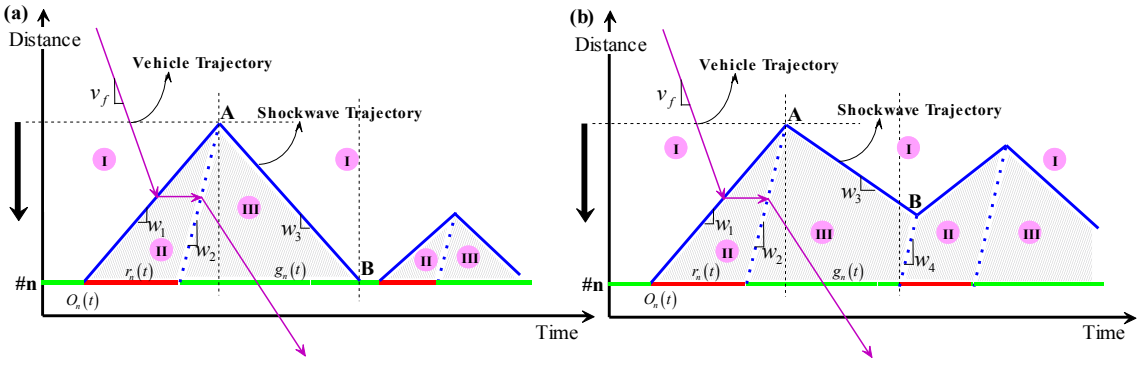

Figure 6-2 Shockwave Profile of Single Intersection: (a) Without Residual Queues; (b) With a Residual Queue

Specifically, a shockwave profile can be used to identify traffic states on a signalized link. Figure 6-2 shows the shockwave profiles of one intersection with and without residual queues. In region I of both diagrams, vehicles are free-flow traveling before reaching the front of queuing wave w_1 or departure wave w_3. Since a linear flow-density relationship is assumed for uncongested traffic, the density in region I can be approximated by integrating traffic flow entering this area. Between the front of queuing wave w_1 (or compression wave w_4, if there is a residual queue at the end of the previous cycle) and the stop line (or the front of discharge wave w_2), *i.e.*, the shadowed region II in the figure, vehicles cannot move and traffic density reaches the maximum (jam density). By contrast, in the shadowed region III, which is from the front of discharge wave w_2 and departure wave w_3 to the stop line (or the front of compression wave w_4, if there is a residual queue at the end of the current cycle), vehicles are discharged at the saturation flow rate.

Before discussing the SPM model, some important notation in this section is introduced below:

w: the velocity of a shockwave; there are four major shockwaves: queuing wave (w_1), discharge wave (w_2), departure wave (w_3), and compression wave (w_4);

$l_n^{w_j}(t)$: the distance from the front of shockwave w_j (j = *1*, *2*, *3*, or *4*) to the stop line of intersection n at time t;

$l_n^{\hat{w}}(t)$: the queue length, defined as the distance from the front of shockwave w_1 or w_3 to the stop line of intersection n at time t, *i.e.*, $l_n^{\hat{w}}(t) = l_n^{w_1}(t)$ or $l_n^{w_3}(t)$;

q_s^n, k_{jam}^n, and k_s^n: saturation flow rate, jam density, and saturation density at link L_n;

v_f: free-flow speed.

By integrating vehicles over a shockwave profile using Eq.(6.1) and applying the shockwave theory (Eq.(6.2)), the velocities of the four shockwaves w_1, w_2, w_3, and w_4 at time t can be derived based on the current signal status and locations of the shockwave fronts. Note that these waves only exist within certain time intervals, which are determined not only by signal status but also by other shockwaves.

$$w_1(t)=\begin{cases}\dfrac{-q_{n-1}\left(t-\dfrac{L_n-l_n^{\hat{w}}(t)}{v_f}\right)}{k_{jam}^n-q_{n-1}\left(t-\dfrac{L_n-l_n^{\hat{w}}(t)}{v_f}\right)/v_f}, & \text{if } O_n(t)+\dfrac{l_n^{\hat{w}}(t)}{w^*}\le t<O_n(t)+r_n(t)+\dfrac{l_n^{\hat{w}}(t)}{w^*}\\ 0 & \text{otherwise}\end{cases} \quad (6.3)$$

$$w_2(t)=\begin{cases}-w^*, & \text{if } O_n(t)+r_n(t)\le t<O_n(t)+r_n(t)+\dfrac{l_n^{\hat{w}}(t)}{w^*}\\ 0, & \text{otherwise}\end{cases} \quad (6.4)$$

$$w_3(t)=\begin{cases}\dfrac{q_s^n-q_{n-1}\left(t-\dfrac{L_n-l_n^{\hat{w}}(t)}{v_f}\right)}{k_s^n-q_{n-1}\left(t-\dfrac{L_n-l_n^{\hat{w}}(t)}{v_f}\right)/v_f}, & \text{if } O_n(t)+r_n(t)+\dfrac{l_n^{\hat{w}}(t)}{w^*}\le t<O_n(t)+r_n(t)+g_n(t)\\ & \text{or } O_n(t)\le t<O_n(t)+\dfrac{l_n^{\hat{w}}(t)}{w^*}\\ 0, & \text{otherwise}\end{cases} \quad (6.5)$$

$$w_4(t)=\begin{cases}-w^*, & \text{if } O_n(t)\le t<O_n(t)+\dfrac{l_n^{\hat{w}}(t)}{w^*}\\ 0, & \text{otherwise}\end{cases} \quad (6.6)$$

where w^* is a constant and can be calculated by Eq.(6.7):

$$w^*=\frac{q_s^n}{k_{jam}^n-k_s^n} \quad (6.7)$$

Notice that the velocity of w_3 automatically becomes v_f if the maximum flow rate (*i.e.* the capacity) keeps the same before and after traffic breaks down due to congestion. But if the capacity drops after traffic breaks down, the velocity of w_3 will be smaller than v_f.

Based on the values of $w_1(t)$ and $w_3(t)$, and the current queue length ($l_n^{\hat{w}}(t)$), the queue length at the next time step ($t+\Delta t$), *i.e.*, $l_n^{\hat{w}}(t+\Delta t)$, can be updated using Eq.(6.8). Note that the distance should be neither negative nor longer than the link length.

$$l_n^{\hat{w}}(t+\Delta t)=\begin{cases}\min\left\{L_n,\max\left\{0,l_n^{\hat{w}}(t)-w_1(t)\cdot\Delta t\right\}\right\}, & \text{if } w_1(t)\neq 0\\ \min\left\{L_n,\max\left\{0,l_n^{\hat{w}}(t)-w_3(t)\cdot\Delta t\right\}\right\}, & \text{if } w_3(t)\neq 0\\ \min\left\{L_n,\max\left\{0,l_n^{\hat{w}}(t)\right\}\right\}, & \text{otherwise}\end{cases} \tag{6.8}$$

Similarly, based on the values of $w_2(t)$ and $w_4(t)$, and the current wave front positions of w_2 and w_4 ($l_n^{w_2}(t)$ and $l_n^{w_4}(t)$), the wave front positions at the next time step ($t+\Delta t$), *i.e.* $l_n^{w_2}(t+\Delta t)$ and $l_n^{w_4}(t+\Delta t)$, can be updated using Eq.(6.9):

$$l_n^{w_j}(t+\Delta t)=\begin{cases}\min\left\{L_n,\max\left\{0,l_n^{w_j}(t)-w_j(t)\cdot\Delta t\right\}\right\}, & \text{if } w_j(t)\neq 0\\ 0, & \text{otherwise}\end{cases},\ j=2 \text{ or } 4 \tag{6.9}$$

Once the wave positions are updated, the shockwave profiles can be constructed. The shockwave profiles are then used to determine the outflow from the intersection: if the signal is red, the outflow rate is zero; when the signal turns green, the departure rate is equivalent to the saturation rate if there is a queue, or determined by the arrival flow with a time lag (L_n/v_f) after the queue has been fully discharged. Considering that the intersection output may also be constrained by the downstream capacity, the following equation summarizes the results:

$$\bar{q}_n(t)=\begin{cases}0, & \text{if } O_n(t)\le t<O_n(t)+r_n(t)\\ \min\left\{q_s^n,q_s^{n+1}\right\}, & \text{if } O_n(t)+r_n(t)\le t<O_n(t)+r_n(t)+g_n(t)\ \&\ l_n^{\hat{w}}(t)>0\\ \min\left\{q_{n-1}\left(t-\dfrac{L_n}{v_f}\right),q_s^{n+1}\right\}, & \text{if } O_n(t)+r_n(t)\le t<O_n(t)+r_n(t)+g_n(t)\ \&\ l_n^{\hat{w}}(t)=0\end{cases} \tag{6.10}$$

Given a known turning percentage (β_n) for the through movement at link L_n, the input for the downstream intersections can be easily derived based on the output from the upstream intersections:

$$q_n(t) = \begin{cases} \bar{q}_n^m(t), & \text{if } O_n(t) \le t \le O_n(t) + r_n(t) \\ \beta_n \cdot \bar{q}_n(t), & \text{if } O_n(t) + r_n(t) < t \le O_n(t) + r_n(t) + g_n(t) \end{cases} \quad (6.11)$$

where $\bar{q}_n^m(t)$ is the departure flow rate of minor streets at intersection *n*.

6.2.2 SPM for Intersections with Spillover from Downstream Links

For multiple intersections, the main issue is how to deal with spillover from downstream traffic. When a spillover from downstream happens, the cyclic process of queue build-up and dissipation is disrupted. The shockwave profile for an urban network with more than one intersection could be very complicated. Depending on when spillover happens and how long spillover lasts, it can be treated as either extending the original red time or creating new small cycles, as described in the following:

Case I: Extending the red phase: The first case, as presented in Figure 6-3a, describes a situation in which the queue at downstream intersection (*n+1*) spills back to upstream intersection (*n*), the signal at intersection *n* is red, and the spillover persists after the signal turns green. In this case, vehicles cannot be discharged when the signal turns green until a discharge wave generated at the downstream intersection propagates back to the upstream intersection. In other words, the red phase has to be extended to the time when the discharge wave arrives.

Case II: Creating new small cycles: Unlike Case I, the second case represents a situation where the queue from the downstream intersection (*n+1*) spills back to the upstream intersection (*n*), and the signal at the upstream intersection is green (Figure 6-3b). Vehicles are forced to stop until the discharge wave generated at the downstream intersection propagates back to the upstream intersection, rendering a portion of the green time unusable. As shown in Figure 6-3b, once spillover happens, a compression wave (w_4) is generated at the n^{th} intersection; this wave propagates backward and will meet the original departure wave (w_3), creating a new queuing wave (w_1). Note that if the queue at the n^{th} intersection has been fully discharged before the spillover occurs, a new queuing wave (w_1), not a compression wave (w_4), will be generated. When the spillover dissipates,

a new discharge wave (w_2) will be generated, propagating backward and eventually meeting up with the new queuing wave (w_1). The structure of this shockwave profile becomes much more complex compared with those without spillover. However, as clearly indicated in Figure 6-3b, spillover essentially creates a new cycle, in which the new red phase starts when spillover occurs and ends when the discharge wave from downstream intersection arrives at the upstream intersection. Within the new cycle, the shockwave profile presents the same pattern as described in Figure 6-2 for the situation without spillover.

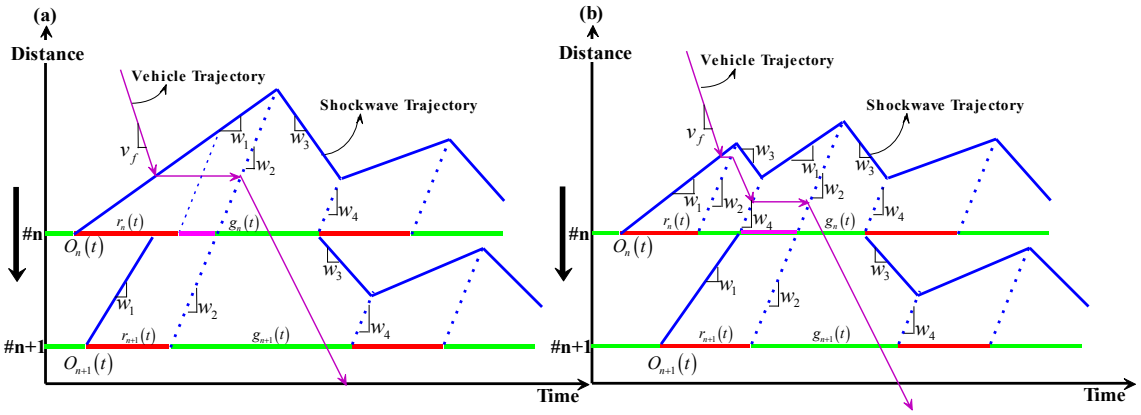

Figure 6-3 Two Interpretive Cases for Spillover: (a) I: Extending the Red Phase; (b) II: Creating New Cycles

The basic idea to deal with these two spillover cases is to update signal timing at the time when queue spills back to upstream intersection (*i.e.* $l_{n+1}^{\hat{w}}(t) < L_{n+1} \le l_{n+1}^{\hat{w}}(t+\Delta t)$ or $l_{n+1}^{w_4}(t) < L_{n+1} \le l_{n+1}^{w_4}(t+\Delta t)$) and when discharge wave propagates back to upstream ($i.e.\ l_{n+1}^{w_2}(t) < L_{n+1} \le l_{n+1}^{w_2}(t+\Delta t)$). However, since SPM updates traffic states at each time instant *t*, when spillover happens, the duration of spillover is not known. For modeling convenience, we can treat the rest of time in the cycle as red when spillover happens, and let the rest of time in the cycle as green when spillover dissipates. Note it is possible to have a situation in which the signal is red when the discharge wave (w_2) propagates back to the upstream intersection. For this situation, we cannot set all the rest time in the cycle

as green; instead, the left time for the red signal is considered as the new red period and the green remains the same. Eq.(6.12) summarizes all situations.

$$\begin{cases} \text{if } l_{n+1}^{\hat{w}}(t) < L_{n+1} \le l_{n+1}^{\hat{w}}(t+\Delta t) \\ \text{or } l_{n+1}^{w_4}(t) < L_{n+1} \le l_{n+1}^{w_4}(t+\Delta t) \end{cases} \text{ then } \begin{cases} O_n(t+\Delta t) = t \\ r_n(t+\Delta t) = O_n(t) + r_n(t) + g_n(t) - t \\ g_n(t+\Delta t) = 0 \end{cases}$$

$$\text{if } l_{n+1}^{w_2}(t) < L_{n+1} \le l_{n+1}^{w_2}(t+\Delta t) \quad \text{then } \begin{cases} O_n(t+\Delta t) = t \\ r_n(t+\Delta t) = \max\{O_n(t) + r_n(t) - t, 0\} \\ g_n(t+\Delta t) = O_n(t) + r_n(t) + g_n(t) - r_n(t+\Delta t) - t \end{cases} \tag{6.12}$$

Note that by comparing ($O_n(t) + r_n(t) - t$) and *0*, we can know whether there is red time left when discharge wave (w_2) propagates back upstream.

Once the signal status is updated, the model introduced in Section 6.2.1 can be applied to describe traffic dynamics with spillover. We should point out that the above module specifically designed for spillover is crucial for SPM. As we noted earlier, the shockwave profiles with spillover for multiple intersections are extremely complicated. Figure 6-4 presents two potential shockwave profiles for an arterial with three intersections, but there are many other possible profiles depending on when spillover happens, how long spillover lasts, and whether there is a residual queue at the end of the cycle. It is infeasible to enumerate all the possible profiles and apply the model described in Section 6.2.1 to estimate the traffic dynamics. However, all spillovers can be categorized as either an extension of red phase (Case I) or a creation of new cycles (Case II). The complicated cases with spillover can then be easily converted to cases without spillover by updating signal timings using Eq.(6.12). This is tremendously beneficial for large network applications since the simple model introduced for non-spillover situations can be directly applied no matter how complicated the shockwave profiles are.

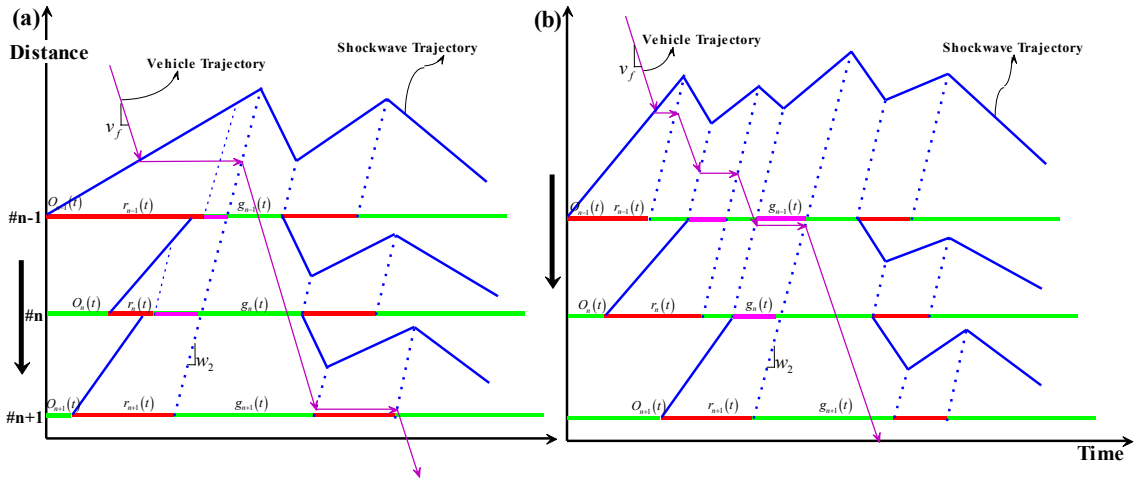

Figure 6-4 Shockwave Profiles for Three Intersections with Spillovers

6.3 Network Representation

6.3.1 Nodes & Arcs

To complete the model, we need to choose an appropriate method to represent an arterial network. Since SPM treats each homogeneous road segment as a section, it is convenient to use nodes to represent road sections with arcs between two adjacent nodes to indicate the direction of vehicle transfer. Here nodes carry most of the physical data including arterial link length, parameter values (such as jam density, saturation rate, *etc.*), and signal timing plans. The data is then used to construct shockwave profiles, to derive potential departure rates, and to determine arrivals for downstream nodes. Arcs, by contrast, play only a minor role, as they simply indicate travel directions to ensure vehicles are properly transferred among nodes. Figure 6-5 gives an example of a single intersection represented by nodes and arcs.

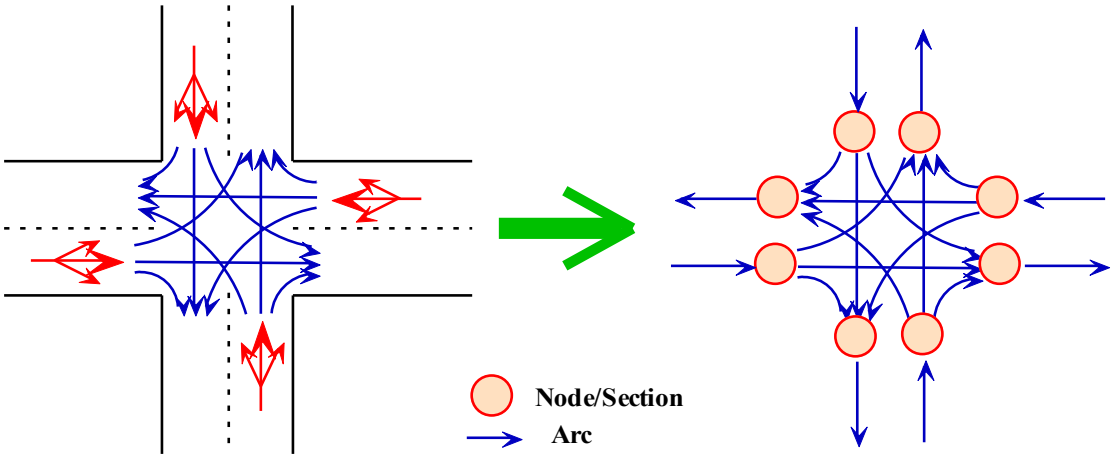

Figure 6-5 Nodes & Arcs

Note that a signal link should be subdivided into sections if the number of lanes changes; since new shockwaves may be generated at the location where traffic heterogeneity exists. The connecting point between two consecutive sections is modeled as a virtual intersection with an all-green phase. The model introduced in Section 6.2 can then be applied.

6.3.2 Network Representations for Intersections with Turning Bays

One case we have not addressed yet in our model is intersections with turning bays. A signal link with turning bays cannot be treated simply as two sub-sections because turning movements may have different signal timing with through movement; and more importantly, spillover from turning bays may lead to the blockage of through movement. In order to apply the SPM model described in Section 6.2, we divide the link into three sections as shown in Figure 6-6a: L_n^U, L_n^B, and L_n^A, representing upstream through movement (starting from the link entrance to the separation point U), downstream through movement (starting from the separation point U to the stop line), and turning bays, respectively (note $L_n^B = L_n^A$). These three sections are represented by three nodes as described in Figure 6-6b. Separation point U is treated as a virtual intersection with all-green, as shown in Eq.(6.13). The equations presented in Section 6.2 can then be applied.

$$\begin{cases} O_n(t) = 0 \\ r_n(t) = 0 \\ g_n(t) = T \end{cases} \tag{6.13}$$

where T is the whole simulation period.

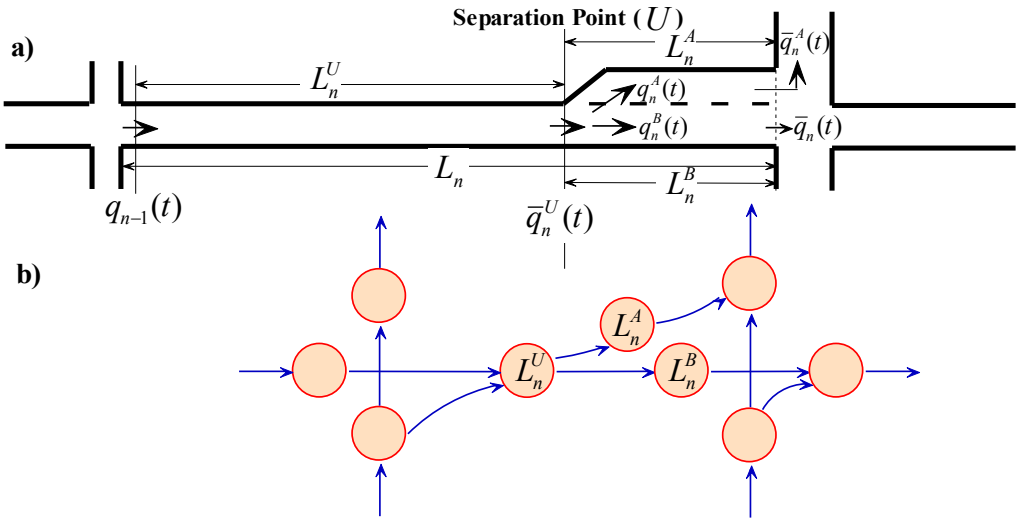

Figure 6-6 (a) Layout of an Intersection with Turning Bay; (b) Network Representation

An example of a shockwave profile for a signal link with turning bay is shown in Figure 6-7. When the queuing wave on the turning section $L_n{}^A$ (represented by the purple line) propagates back to the separation point U (see Figure 6-6a), a shockwave profile on section $L_n{}^U$ (represented by the dark blue line) is generated. Meanwhile the shockwave profile on section $L_n{}^B$ (represented by the light green line) has also been impacted as there are no more vehicles coming from upstream.

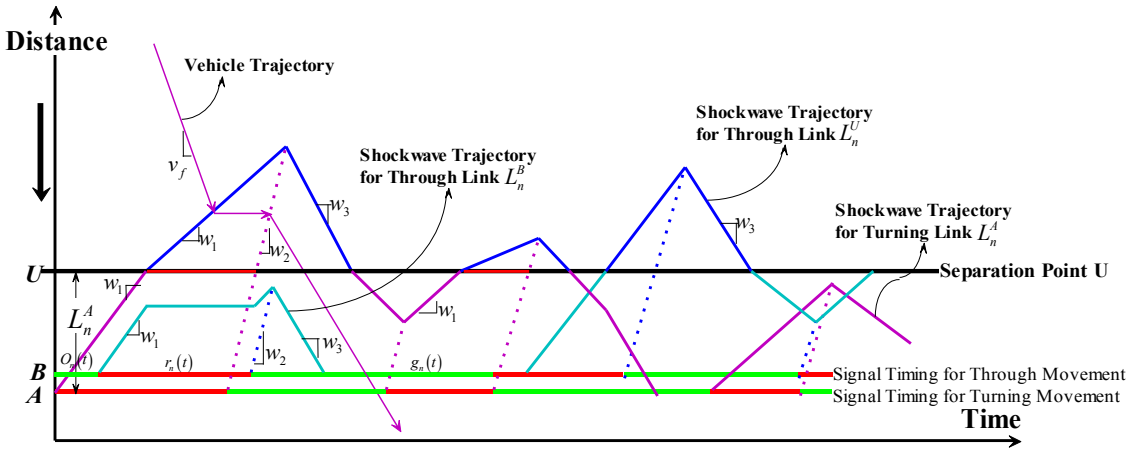

Figure 6-7 Shockwave Profiles for an Intersection with Spillover at Turning Bays

However, when a queue spills back from turning bay L_n^A to the separation point U and the section of L_n^U has multiple lanes, usually only the leftmost lane on L_n^U is blocked and the other lanes remain open (see Figure 6-8a). Under such circumstances, section L_n^U needs to be split into two parallel sub-sections, L_n^{V1} and L_n^{V2}, representing the blocked and unblocked lane(s) respectively (see Figure 6-8b). The inputs for the two sub-sections ($q_n^{V1}(t)$ and $q_n^{V2}(t)$) are assigned directly at the entrance of section L_n^U based on known turning percentages; and the output is determined by the signal timings at two virtual intersections (*i.e.* points V_1 and V_2) and shockwave profiles in two sub-sections L_n^{V1} and L_n^{V2}. Note we need to introduce a dummy node (with zero section length) to represent traffic flow coming from upstream intersections before we split section L_n^U into two sub-sections based on turning percentage (see Eq.(6.14)).

$$\begin{cases} q_n^{V1}(t) = (1-\beta_n)\cdot q_{n-1}(t) \\ q_n^{V2}(t) = \beta_n \cdot q_{n-1}(t) \end{cases} \tag{6.14}$$

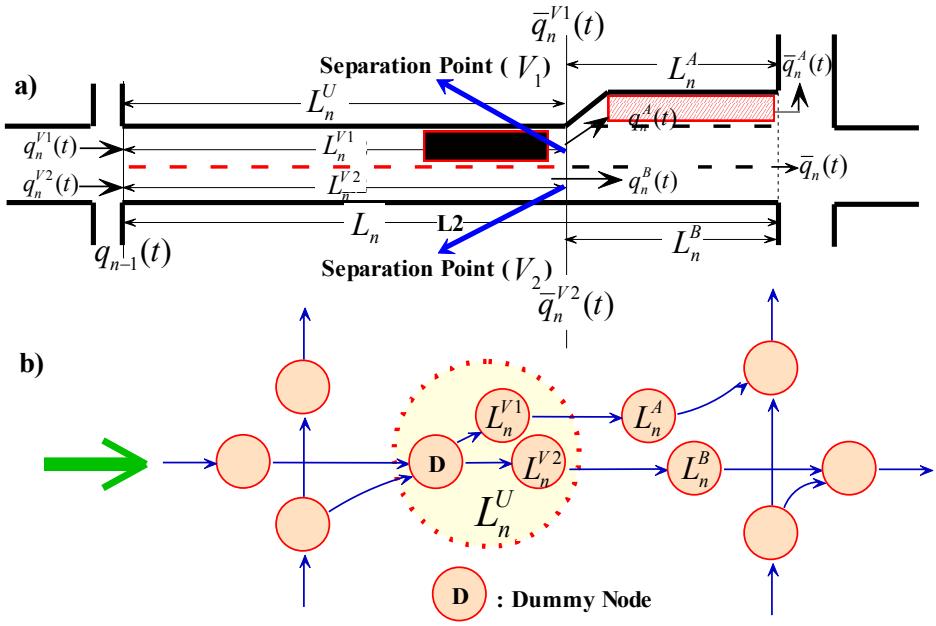

Figure 6-8 Layouts of Intersections with Turning Bays – Through Movement are Partially Blocked

It must be noted that the lane-based section design is only an approximation of the real world. We assume that vehicles have been assigned to different lanes (or lane groups) at the entrance of the link in order to avoid dealing with complex lane-changing behaviors. Although this assumption is not exactly true, it is still a reasonable one as when one lane is blocked, it is very likely that vehicles will make a lane choice earlier. A more sophisticated lane-changing model could increase accuracy, but at a cost. The model we propose is more numerically efficient and robust, and thus more appropriate for large arterial network applications.

6.4 Numerical Example Comparing SPM & CTM

To demonstrate the capability of SPM, we provide a small numerical example, in which, we compared results generated by SPM and CTM. For the experiment, we used a simple arterial with two intersections and two links but no turning movements, as shown in Figure 6-9a. The second link was designed relatively short in order to create spillover. The model inputs, including arrival flow at the entrance of the first intersection, signal

timings, free-flow speed, jam density, discharge wave velocity, saturation flow, and flow-density relation for CTM, are presented in Figure 6-9b-d.

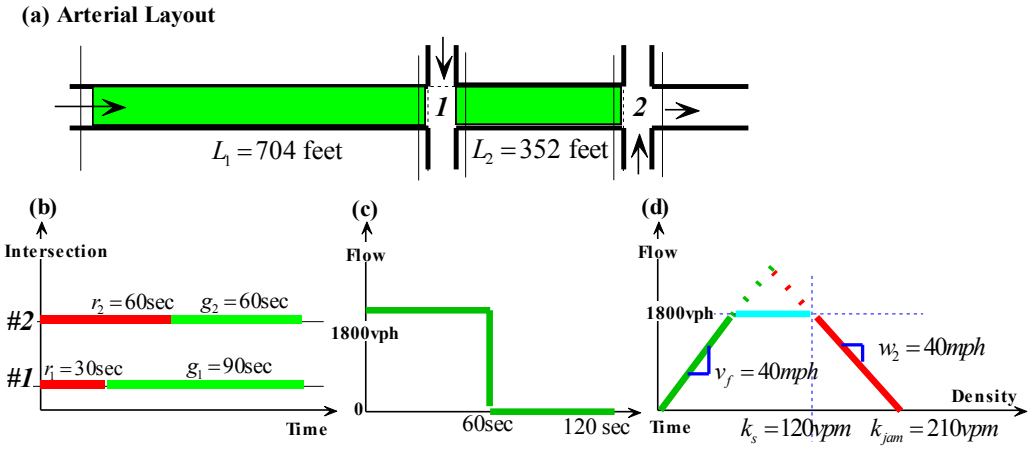

Figure 6-9 (a) Layout of the Two-Intersection Example; (b) Signal Timing Parameters; (c) Input Flow Rate; (d) The FD for CTM

6.4.1 CTM

We first briefly introduce CTM and its cell design. Essentially, CTM applies the following equations for the simulation of traffic dynamics.

$$\begin{cases} x_j(t+1) = x_j(t) + f_j(t) - f_{j+1}(t) \\ f_j(t) = \min\left\{ x_{j-1}(t), Q_j(t), (\frac{w_2}{v_f}) \cdot \left[Z_j(t) - x_j(t) \right] \right\} \end{cases} \tag{6.15}$$

where $x_j(t)$, $f_j(t)$, $Q_j(t)$, and $Z_j(t)$ denote the number of vehicles, the actual inflow, the inflow capacity (the maximum allowable inflow), and the holding capacity (the maximum allowable number of vehicles) in cell j at time t. Essentially, the first equation describes the flow conservation for each cell; and the second represents the possible number of vehicles entering a cell based on a hypothesized piecewise linear fundamental diagram.

One of the most important steps in CTM is to discretize links into homogenous cells such that the cell length is equal to the distance traveled by a free-flowing vehicle in one time

interval. Generally, CTM will generate more accurate results when cell length is relatively short. In this example, we use a one-second time interval and a cell length of 58.67 ft under the assumption that the free-flow speed is 40 mph. Figure 6-10a presents the cell design for the two-link arterial. Note that the last cell (the blue one) has infinity holding capacity so that the departure from the upstream cell is not restricted. Compared with the section design for SPM (Figure 6-11), the cell design for CTM (Figure 6-10) is much more complicated.

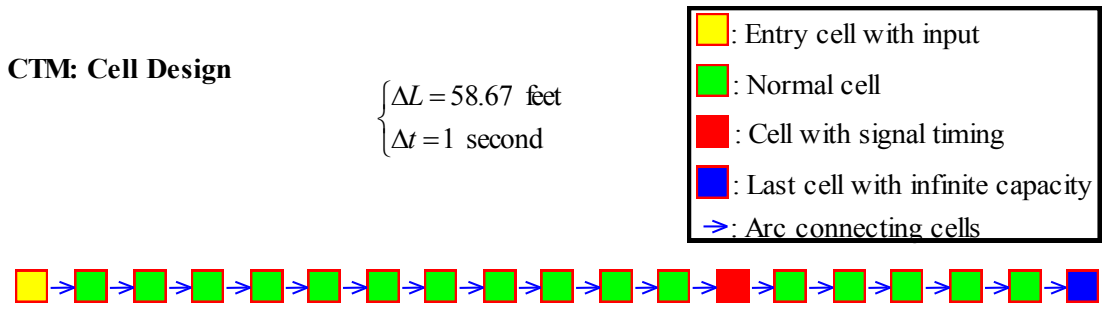

Figure 6-10 Cell Design for CTM

Figure 6-11 Section Design for SPM

6.4.2 Results

The results presented here are only for the first link (L_1 = 704ft) from one of the signal cycles. Since CTM provides the density for each cell at each time interval, we show the density contour in Figure 6-12a. Figure 6-12b displays the corresponding shockwave profile. As can be seen, the shockwave profile closely matches the density contour from CTM. Figure 6-13 presents the output flow from the stop-bar at the first intersection for both models. These results are also very consistent. Note that a downstream spillover occurs in this cycle and the purple bar represents the updated red phase due to spillover.

All these comparisons demonstrate that SPM can accurately describe traffic dynamics but with much simpler network design.

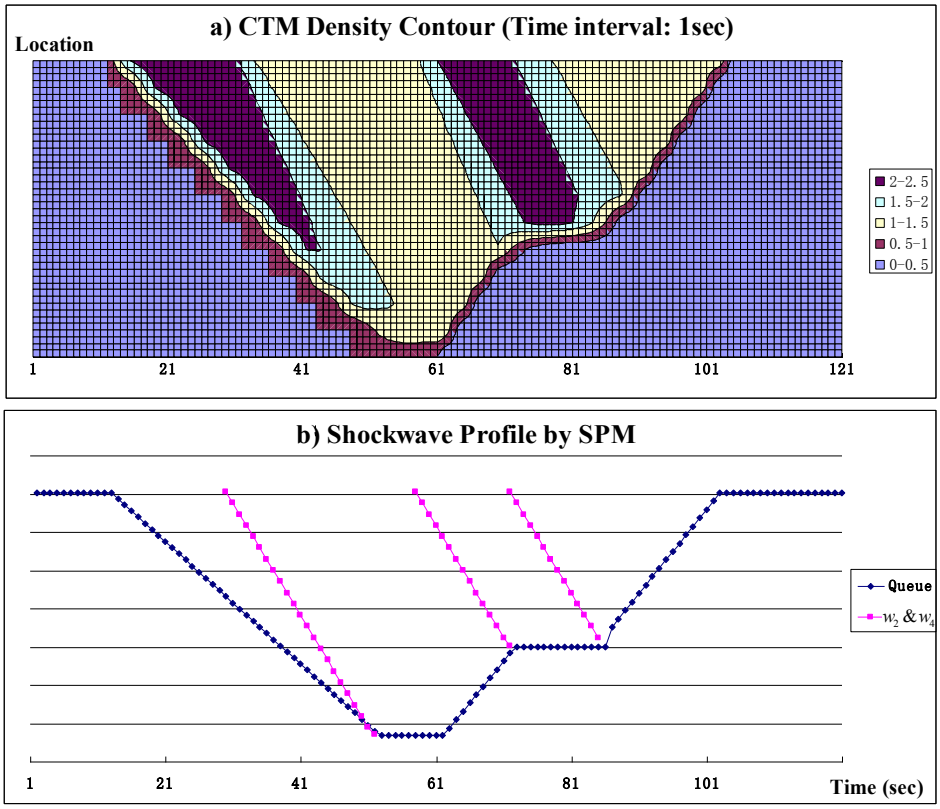

Figure 6-12 (a) CTM Density Contour (Time Interval: 1sec); (b) Shockwave Profile Generated by SPM.

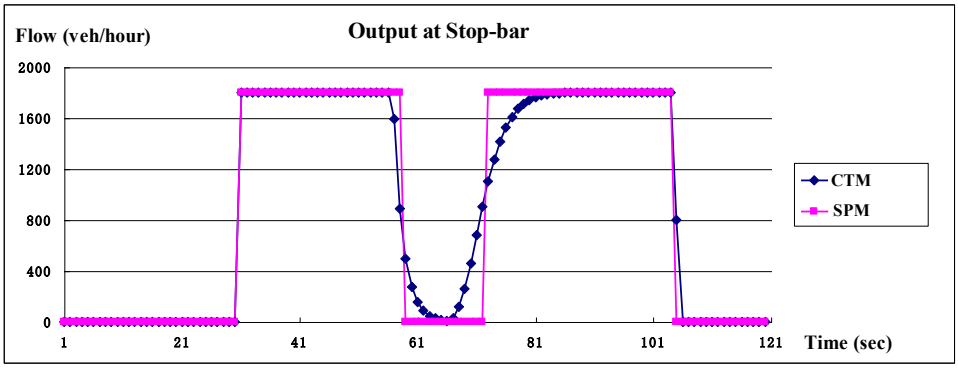

Figure 6-13 Output at the Stop-bar for the First Intersection

6.5 Field Test Results

6.5.1 Field Data Collection

We further validated the SPM model using the field data collected from Trunk Highway 55 (TH55). To review, the installation on TH55 included advance detectors on the major approaches (about 400 ft upstream from the stop line) to detect vehicles for green extension, and stop-bar detectors installed on minor streets (right behind the stop line) to detect the presence of vehicles. For research purposes, we also installed stop-bar and link entrance detectors on major approaches. Figure 6-14 shows the detector configurations for three intersections of interest (Winnetka Ave., Rhode Island Ave., and Glenwood Ave.). We do not include the advance detectors in the figure since no information was used from them in this testing.

All three intersections are installed with the SMART-Signal system, which continuously collects and archives high-resolution event-based traffic signal data in real time. The event data provide start and end times of each vehicle-detector actuation and signal phase change, making it possible to extract time-dependent traffic volume, turning percentage, and signal phase information at each intersection from the raw data.

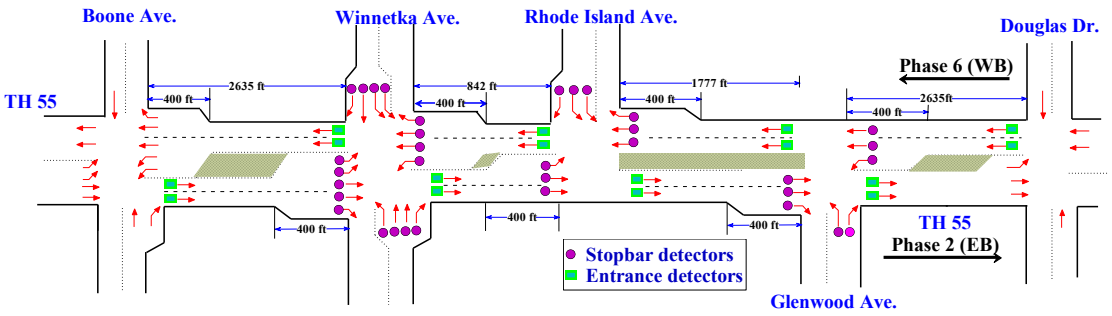

Figure 6-14 Layout of TH55 Test Site

The data we selected to test the model was collected during an afternoon peak on Nov. 17th, 2008. This day was specifically selected because a nine-cycle spillover was known

to have occurred that day between 17:06:31 and17:36:31 in the westbound of the intersection TH55/Rhode Island Ave, as shown in Chapter 5 (also see Wu *et al.*, 2010a for a detailed discussion of spillover and oversaturation). The input data to the SPM model includes the time-dependent traffic volume at the arterial boundaries, the turning percentage at each intersection, and the signal phase information for each approach. The inflow at boundaries was collected by the entrance detectors at the eastbound link of the intersection of Winnetka Ave. and the westbound link of the intersection of Glenwood Ave., and by the stop-bar detectors at all minor streets (see Figure 6-14).

6.5.2 Network Representation

The three intersections of interest on TH55 are represented as a network in Figure 6-15. Since there are turning bays, each link between two intersections is divided into two sections represented by two nodes and two arcs. The start or end nodes at boundaries are treated as dummy nodes, which have infinite length to store vehicles.

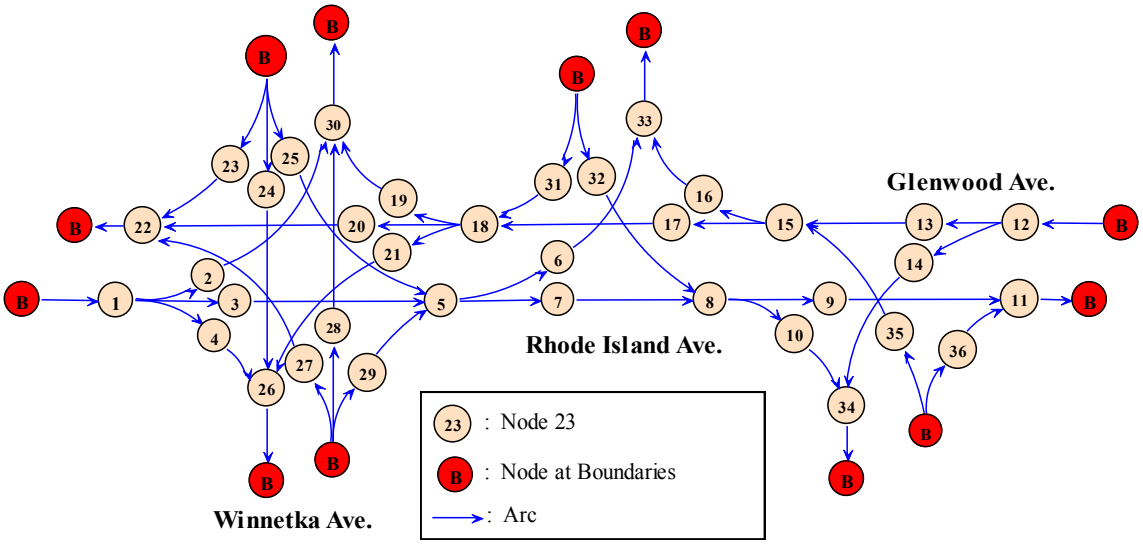

Figure 6-15 Network Representation of the Three Intersections on TH55

6.5.3 Results

The model simulated traffic dynamics for three hours (16:00:00 -19:00:00), but only two hours of data (16:30:00 -18:30:00) for the westbound of TH55 are presented here for demonstration purposes. In this study, the saturation flow rate, free flow speed, and jam density were set to 2100 veh/hr/ln, 45 m/hr, and 176 veh/m/ln, respectively. Consequently, the discharge shockwave speed used in this study is 16.2 m/hr. All of these parameters can be verified by the high-resolution data collected from the SMART-Signal system. The time interval we used for SPM was one second.

Figure 6-16 compares the simulated traffic volumes with the ground truth collected by the entrance and stop-bar detectors for the three intersections for every signal cycle between 16:30:00 and18:30:00. The comparisons for both detectors indicate that SPM generates consistent results with the ground truth. Table 6-1 also shows the mean percentage error (MPE) and mean absolute percentage error (MAPE) (Eq.(6.16)). The results validate that SPM can accurately describe traffic flow propagation.

$$\begin{cases} MPE = \dfrac{1}{m}\sum_{m}\left(\dfrac{Observation - Estimation}{Observation}\right)\times 100\% \\ MAPE = \dfrac{1}{m}\sum_{m}\left|\dfrac{Observation - Estimation}{Observation}\right|\times 100\% \end{cases} \tag{6.16}$$

where m is the sample size.

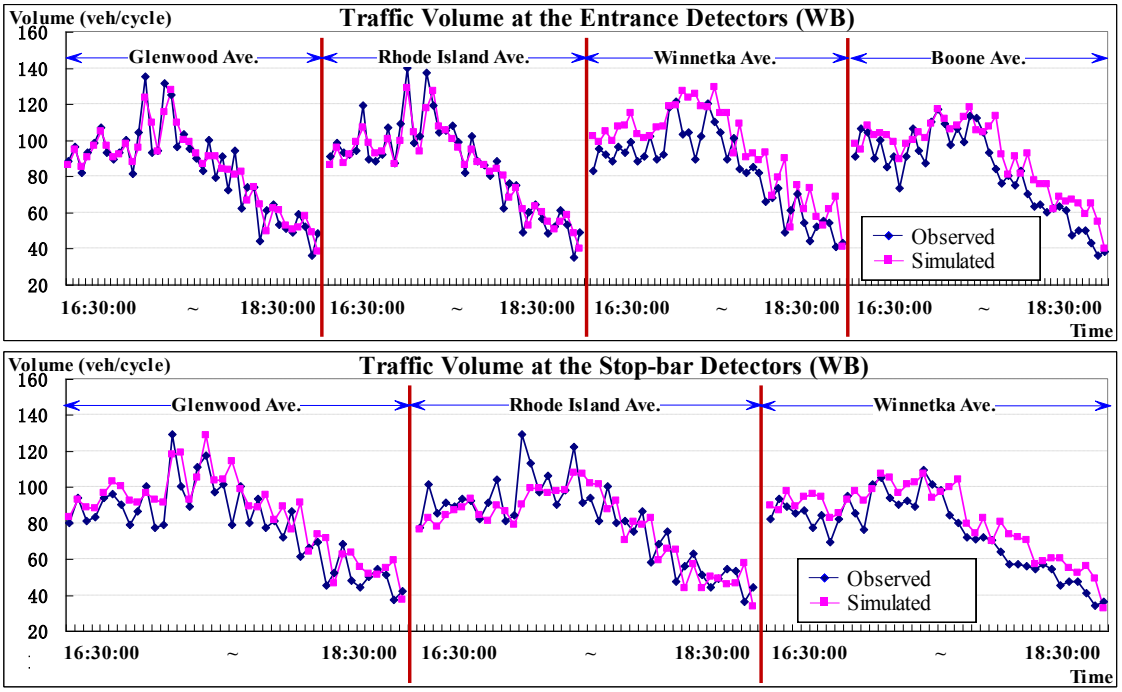

Figure 6-16 Comparison of Observed & Simulated Traffic Volumes

Table 6-1 Comparison of Error Rates of Traffic Volumes for Two Detector

	MPE	MAPE
Entrance Detectors	**5.51%**	**12.3%**
Stop-bar Detectors	**4.25%**	**13.8%**

This test also demonstrates that SPM can accurately simulate traffic dynamics with spillovers. As clearly presented in Figure 6-17, SPM creates small new red phases (marked as the pink bars in the figure) for spillover cycles. Figure 6-18 compares the time duration of the spillovers. The results indicate that SPM estimates a longer oversaturation period as spillover starts one cycle early and ends one cycle late. The reason could be the numerical errors generated at some time instants. For example, vehicles cannot be transferred from one section to another if the free-flow travel distance within a time interval is longer than the distance between the tail of the downstream queue and the section entrance. But in reality, vehicles can move slower (less than free-flow speed) and join the downstream queue. So SPM requires a little bit longer time to discharge queues at the upstream, particularly when the downstream section is congested.

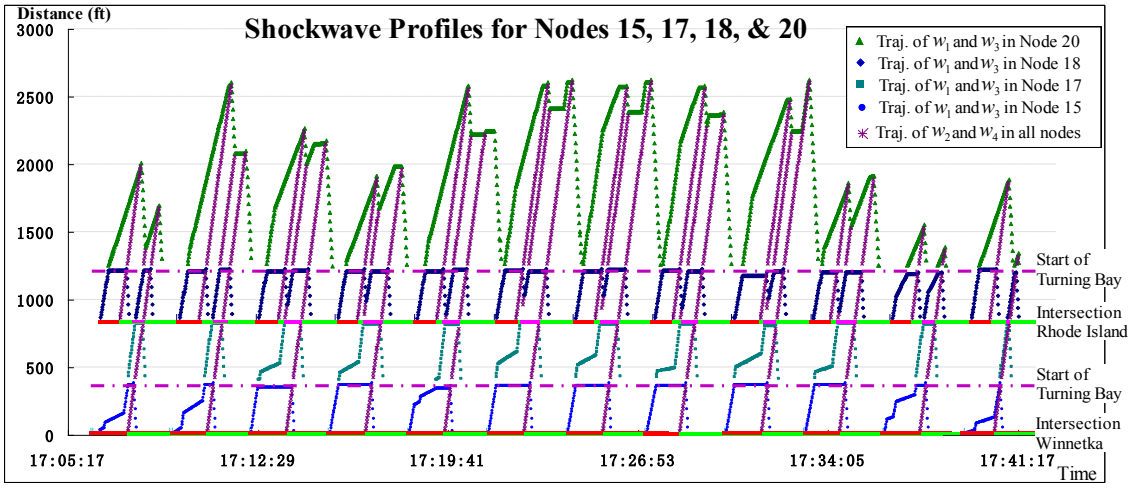

Figure 6-17 Shockwave Profiles for Nodes 15, 17, 18, & 20

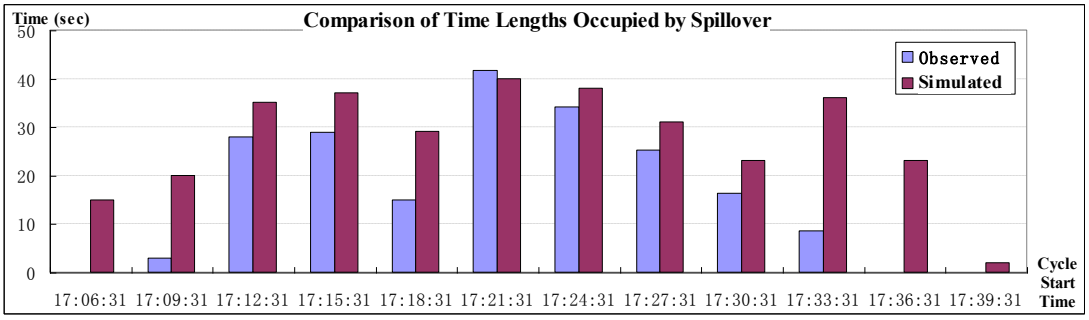

Figure 6-18 Comparison of Time Lengths Occupied by Spillovers

6.6 Summary

In this chapter, we propose a simplified shockwave profile model (SPM), which simulates the traffic dynamics in a congested urban network. Taking advantage of the fact that traffic states within a congested link can be simplified as free-flow, saturated, and/or jammed conditions, SPM simulates the traffic dynamics by deriving the trajectories of shockwaves analytically. Each link with the same number of lanes is treated as a section in this model. The queuing dynamics within each section are described by tracing the shockwave fronts which explicitly separate the three traffic states. SPM is specifically

designed to deal with saturated or oversaturated arterials. In this model, a novel design is developed to treat spillover as either extending a red time or creating new smaller cycles. This model is promising for analytical investigations of traffic in congested signalized arterials especially with queue spillover. The field test results verify that SPM can successfully and accurately replicate spillover situations. Therefore, SPM is highly appropriate for large network real-time applications, especially when traffic is oversaturated.

We expect that SPM can be applied in both arterial performance prediction and traffic signal optimization. Since SPM traces the shockwave fronts on a signalized link, the intersection queue length can be calculated easily. Given queue length information, other performance measures such as travel time, delay, and stops can be derived accordingly. As a traffic flow model, SPM also provides a way to bridge the relationship between signal timing and intersection performance, so that a signal optimization program can be formulated by maximizing benefits (travel time savings, intersection throughput, *etc.*) or minimizing costs (queue, travel time, delay, *etc.*).

7 Conclusions & Future Research

7.1 Conclusions

This dissertation serves as a critical first step in mitigating oversaturation. Building upon the high-resolution traffic signal data recently collected by the University of Minnesota, this research first proposes a systematic approach to identify and quantify oversaturation, and then develops a simplified traffic flow model to describe traffic dynamics under oversaturated conditions. In particular, this research makes contributions in four main areas:

1) Oversaturation quantification. This research first proposes a quantifiable measure of oversaturation by focusing on its detrimental effects in both temporal and spatial dimensions. The temporal detrimental effect is characterized by the occurrence of residual queue and the spatial detrimental effect is referred to the spillover from downstream intersection to upstream. The oversaturation severity index (OSI), in either temporal dimension (T-OSI) or spatial dimension (S-OSI) can then be measured by the ratio between the unusable green time due to detrimental effects and the total available green time in a cycle. T-OSI is quantified by estimating the residual queue length; and S-OSI is quantified by measuring the time period of spillover. Two algorithms are applied to identify oversaturated signalized intersection. Field-test results from a major arterial in the Twin Cities area demonstrate that these two algorithms are very effective in identifying oversaturated conditions.

2) Real-time queue length estimation for congested intersections. To quantify T-OSI, an innovative approach is proposed to estimate intersection queue length using existing detectors. A key methodological contribution of this approach is that it can estimate time-dependent queue length even when signal links are congested with long queues. By

applying LWR shockwave theory with high-resolution traffic signal data, this approach distinguishes different traffic states at the intersection, and then estimates queue length by constructing a shockwave profile. We evaluated the model by comparing the estimated maximum queue length with the ground truth data recorded by camera and human observers. The results that the proposed model can accurately estimate intersection queue length for congested intersections.

3) Queue-Over-Detector (QOD): To quantify S-OSI, we studied the QOD phenomenon. We demonstrated the relationship between QOD and spillover. From that we proposed an algorithm to quantify S-OSI by identifying QOD. We also briefly studied the impact of QOD on the cycle-based arterial fundamental diagram (AFD); and found a stable form of AFD that clearly identifies three distinct regimes: under-saturation, saturation, and over-saturation with queue spillovers. The stable form of AFD is of great importance for traffic signal control because of its ability to identify traffic states on a signal link.

4) Traffic flow modeling for congested arterials. Based on the better understanding of oversaturation gained from above three points, this dissertation further develops a simplified shockwave profile model (SPM), which can simulate the traffic dynamics in an oversaturated urban network. To construct the model, we take advantage of the fact that traffic states within a congested link can be simplified as free-flow, saturated, and jammed conditions. SPM simulates these three dynamic states by analytically deriving the trajectories of shockwaves. Each link with the same number of lanes is treated as a section in this model. The queuing dynamics within each section are described by tracing the distinctive shockwave fronts which explicitly mark each traffic state. SPM is specifically designed to deal with saturated or oversaturated arterials. The model uses a novel design that treats spillover as either extending red time or creating new smaller cycles. This model is promising for analytical investigations of traffic in congested signalized arterials especially with queue spillover. The field test results verify that SPM can successfully and accurately replicate spillover situations. Therefore, SPM is highly appropriate for large network real-time applications, especially when traffic is oversaturated.

7.2 Future Research

This dissertation lays a foundation for the development of computationally based strategies to mitigate arterial traffic congestion. Future research will need to address:

1) Mapping OSI with mitigation strategies. The current research demonstrates that different types of OSI (T-OSI or S-OSI) and different levels of OSI may require different mitigation strategies. Therefore, one future research task is to map identified OSI values with various mitigation strategies for better management of traffic signals under oversaturated conditions.

2) Understanding the transition process from cycle to cycle in the AFD. In the cycle-based AFD, each point represents the traffic state of a cycle. Our research reveals a stable form of the AFD; but the transition process from cycle to cycle is unclear. Research efforts are needed first, to discover characteristics of the congestion regime of the AFD, and second, to build a model that describes the transition process from cycle to cycle.

3) Optimizing traffic signals for oversaturated arterials. The arterial traffic flow model (SPM) developed in this research provides a bridging relationship between signal timing and intersection performance. It will be important to study how to apply SPM to optimize traffic signals, especially for oversaturated arterials. This requires deriving additional arterial performance measures including intersection delay, stops, and arterial travel time; and then optimizing signal control by maximizing benefits (travel time savings, intersection throughput, *etc.*) or minimizing costs (queue, travel time, delay, *etc.*). For real-time applications, we will need to continue looking for ways to mine the riches of real-time traffic data, such as using the information of OSI.

References

Abu-Lebdeh, G., & Benekohal, R. (2003). Design and evaluation of dynamic traffic management strategies for congested conditions. *Transportation Research Part A*, 37(2), 109-127.

Akcelik, R. (1999). *A queue model for HCM 2000*. *ARRB Transportation Research Ltd.*, Vermont South, Australia.

B&B Electronics (2007). *Model 485SD9TB Port-Powered RS-485 Converter*. Retrieved March 30, 2007, from B&B Electronics: http://www.bb-elec.com/bb-elec/literature/485sd9tb-3803ds.pdf.

Balke, K., Charara, H., & Parker, R. (2005). *Development of a traffic signal performance measurement system (TSPMS)*. Texas Transportation Institute, Report 0-4422-2.

Beskos, D.E., Okutani, I., & Michalopoulos, P. (1984). Testing of dynamic models for signal controller intersections. *Transportation Research Part B*, 18(4/5), 397-408.

Catling, I. (1977). A time dependent approach to junction delays. *Traffic Engineering and Control*, 18, 520-526.

Chang, T. H. & Lin, J. T. (2000). Optimal signal timing for an oversaturated intersection. *Transportation Research Part B*, 34(6), 471–491.

Cronje, W. B. (1983a). Analysis of existing formulas for delay, overflow. and stops. *Transportation Research Record* 905, 89-93.

Cronje, W. B. (1983b). Derivation of equations for queue length, stops, and delay for fixed-time traffic signals. *Transportation Research Record*, 905, 93-95.

Daganzo, C. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B*, 28(4), 269-287.

Daganzo, C. (1995). The cell transmission model: Part II: Network traffic. *Transportation Research Part B*, 29(2), 79-93.

Dell'Olmo, P., & Mirchandani, P. (1996). A model for real-time traffic coordination using simulation based optimization. In Bianco and Toth (Eds.), *Advanced Methods in Transportation Analysis*. Springer, Germany, 525-546

Gazis, D.C. (1964). Optimal control of a system of oversaturated intersections. *Operations Research*, 12, 815-831.

Gazis D. C. (1974). *Traffic Science*. Wiley Interscience, New York

Gettman, D., Shelby, S. G., Head, L., Bullock, D. M. & Soyke, N. (2007). Data-driven algorithms for real-time adaptive tuning of offsets in coordinated traffic signal systems. *Transportation Research Record*, 2035, 1–9.

Green, D.H. (1968). Control of oversaturated intersections. *Operational Research Quarterly*, 18(2), 161–173.

Greenshields, B. D. (1935). A study of traffic capacity. *Highway Research Board Proceedings*, 14, 448-477.

Helbing, D. (2003). A section-based queuing-theoretical traffic model for congestion and travel time analysis in networks. *Journal of Physics A: Mathematical and General*, 36, 593-598.

Highway Capacity Manual. (2000). Transportation Research Board, Washington D.C.

Institute of Transportation Engineers (2007), *2007 National Traffic Signal Report Card*, available: http://www.ite.org/reportcard/ 1/13/08 (visited on August 27, 2010)

Lighthill, M.J. & Whitham, G.B. (1955). On kinematic waves I: Flood movement in long rivers. II: A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 229, 281–345, London.

Liu, H., & Ma, W. (2009). A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials. *Transportation Research Part C*, 17(1), 11-26.

Liu, H., Ma, W., Hu, H., Wu, X., & Yu, G. (2008). SMART-SIGNAL: Systematic Monitoring of Arterial Road Traffic Signals. *Presented at the 11th International IEEE Conference on Intelligent Transportation System (ITSC)*, Beijing, China.

Liu, H., Ma, W., Wu, X., & Hu, H. (2010). Real-time estimation of arterial travel time under congested conditions. *Transportmetrica*, in press.

Liu, H., Wu, X., Ma, W., & Hu, H. (2009). Real-time queue length estimation for congested signalized intersections. *Transportation Research Part C*, 17(4), 412-427.

Luyanda, F., Gettman, D., Head, L., Shelby, S., Bullock, D., & Mirchandani, P. (2003). ACS-Lite algorithmic architecture applying adaptive control system technology to closed-loop traffic signal control systems. *Transportation Research Record*, 1856, 175-184.

May A. D. Jr. (1965). Traffic flow theory-the traffic engineers challenge. *Proceedings of the Annual Meeting of the Institute of Transportation Engineers*, 290-303

Michalopoulos, P., & Stephanopoulos, G. (1977a). Oversaturated signal system with queue length constraints: single intersection. *Transportation Research*, 11, 13-422.

Michalopoulos, P., & Stephanopoulos, G. (1977b). Oversaturated signal system with queue length constraints: systems of intersections. *Transportation Research*, 11, 423-428.

Michalopoulos, P. G., Stephanopoulos, G., & Pisharody, V. B. (1980). Modelling of traffic flow at signalized links. *Transportation Science*, 14(1), 9-41.

Michalopoulos P.G., & Stephanopoulos, G. (1981). An application of shock wave theory to traffic signal control. *Transportation Research B*, 15(1), 31–51.

Mirchandani, P. B., & Zou, N. (2007), Queuing models for analysis of traffic adaptive signal control. *IEEE Transactions on Intelligent Transportation Systems*, 8(1), 55-59

Muck, J. (2002). Using detectors near the stop-line to estimate traffic flows. *Traffic Engineering and Control*, 43, 429-434.

National Instruments (2006). *NI-6528 User Guide and Specifications*. Retrieved January 30, 2007, from National Instruments: http://www.ni.com/pdf/manuals/372124a.pdf

Newell, G. F. (1965). Approximation methods for queues with application to the fixed-cycle traffic light. *SIAM Review*, 7(2), 223-340

Newell, G.F. (1982). *Applications of Queuing Theory*, 2nd edition. Chapman & Hall, London, England, 287–300

Newell, G. F. (2001). A simplified car-following theory: a lower order model. *Transportation Research Part B*, 36(3), 195-205.

Papageorgiou, M. (1998). Some remarks on macroscopic traffic flow modeling. *Transportation Research Part A*, 32(5), 323-329

Pignataro, L. J., McShane, W. R., & Crowley, K. W. (1978). *Traffic control in oversaturated street networks*. National Cooperative Highway Research Program Report 194. Transportation Research Board, Washington, D.C.

Quinn, D. J. (1992). A review of queue management strategies. *Traffic Engineering and Control*, 3311, 600-605.

Richards, P.J. (1956). Shock waves on the highway. *Operations Research*, 4, 42–51

Robertson, D.I. (1969). TRANSYT: A traffic network study tool. *Road Research Laboratory, Laboratory Report LR* 253, Crowthorne.

Sharma, A., Bullock, D.M., & Bonneson, J. (2007). Input-output and hybrid techniques for real-time prediction of delay and maximum queue length at a signalized intersection. *Transportation Research Record*, 2035, 88-96.

Sims, A. G., & Dobinson, K. W. (1980). The Sydney Coordinated Adaptive Traffic (SCAT) system philosophy and benefits. *IEEE Transactions on Vehicular Technology*, 29 (2), 130-137

Skabardonis, A., & Geroliminis, N. (2008). Real-time monitoring and control on signalized arterials. *Journal of Intelligent Transportation Systems*, 12(2), 64-74.

Smaglik E.J., Sharma, A., Bullock, D.M., Sturdevant, J.R., & Duncan, G., (2007). Event-based data collection for generating actuated controller performance measures. *Transportation Research Record*, 2035, 97-106.

Stephanopoulos, G., & Michalopoulos, P.G. (1979). Modelling and analysis of traffic queue dynamics at signalized intersections. *Transportation Research Part A* 13(5), 295-307.

Strong, D. W., Nagui M. R, & Ken C. (2006). New calculation method for existing and extended HCM delay estimation procedure. *Proceedings of the 85th annual meeting of the Transportation Research Board*, Washington, D.C.

Texas Transportation Institute (2009). *2009 Urban Mobility Report*. College Station, Texas. Available: http://tti.tamu.edu/documents/mobility_report_2009_wappx.pdf (visited on August 27, 2010).

Vigos, G., Papageorgiou, M., & Wang, Y. (2008). Real-time estimation of vehicle-count within signalized links. *Transportation Research Part C*, 16(1), 18-35

Webster, F. (1958). *Traffic signal settings*. Road Research Technical Paper 39. Road Research Laboratory, Her Majesty's Stationery Office, London.

Webster, F. V., & Cobbe, B. M. (1966). *Traffic signals*. Road Research Technical Paper 56, Road Research Laboratory, Her Majesty's Stationery Office, London.

Wu, X., Liu, H., & Getttman, D. (2010a). Identification of oversaturated intersections using high-resolution traffic signal data. *Transportation Research Part C*, 18(4), 626-638.

Wu, X., Liu, H., & Geroliminis, N. (2010b). Empirical analysis on the arterial fundamental diagram. *Transportation Research Part B*, Accepted, DOI: 10.1016/j.trb.2010.06.003.

Wu, X. & Liu, H. (2010c). A Shockwave Profile Model for Traffic Flow on Congested Urban Arterials. Submitted to *Transportation Research Part B*.

Yi, P., Xin, C., & Zhao, Q. (2001). Implementation and field testing of characteristics-based intersection queue estimation model. *Network and Spatial Economics*, 1, 205-222.

Zhang, H.M. (2001). *Continuum flow models, traffic flow theory, a state-of-the-art report*, Chapter 5. Transportation Research Board. Original text by Reinhart Kuhne and Panos Michalopoulos.