PERFORMANCE ANALYSIS OF ISOLATED INTERSECTION TRAFFIC SIGNALS

A Dissertation

by

KAI YIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Xiubin Wang
Committee Members,	Yunlong Zhang
	Luca Quadrifoglio
	Swaroop Darbha
Head of Department,	Robin Autenrieth

August 2013

Major Subject: Civil Engineering

ABSTRACT

This dissertation analyzes two unsolved problems to fulfill the gap in the literature: (1). What is the vehicle delay and intersection capacity considering left-turn traffic at a pre-timed signal? (2). What are the mean and variance of delay to vehicles at a vehicle-actuated signal?

The first part of this research evaluates the intersection performance in terms of capacity and delay at an isolated pre-timed signal intersection. Despite of a large body of literature on pre-timed signals, few work has examined the interactions between left-turn and through vehicles. Usually a protected left-turn signal phase, before (leading) or after (lagging) through signal, is applied to a signalized intersection when the traffic demand is relatively high. A common problem for leading left-turn operation is the blockage to left-turn vehicles by through traffic, particularly at an intersection with a short left-turn bay. During the peak hour, some vehicles on the through lane might not be able to depart at the end of a cycle, resulting in an increased probability of left-turn blockage. In turn, the blocked left-turn vehicles may also delay the through traffic to enter the intersection during the following cycle. Those problems may not exist for a lagging left-turn operation, since left-turn vehicles intend to spill out of the bay under heavy traffic. In this case, the through capacity is reduced, leading to an increase of total delay. All of these factors contribute to the difficulties of estimating the delay and capacity for an isolated intersection. In order to examine this missing part of study on the signalized intersection, two probabilistic models are proposed to deal with the left-turn bay blockage and queue spillback in a heuristic manner. Numerical case studies are also provided to test the proposed models.

The second part of this research studies an isolated intersection with vehicle-actuated signal. Typically an advanced detector is located at a distance prior to the intersection such that an arriving vehicle triggers a green time extension in order to pass through without any stop. This extended time period actuated by the vehicle is called unit extension in this study. If no vehicle actuation occurs during a unit extension, the green phase would terminate in order to clear queues in other approaches. In this way, the actuated system dynamically allocates the green time among multiple approaches according to vehicle arrivals. And the unit extension is the only control parameter in this case. We develop a model to study the vehicle delay under a general arrival distribution with a given unit extension. Our model allows to optimize the intersection performance over the unit extension.

The third part of this research applies graphical methods and diffusion approximations to the traffic signal problems. We reinterpret a graphical method which is originally proposed by Newell in order to directly measure the variance of the time for the queue clearance at a signalized intersection, which remains yet to be carefully examined in practice and would be rather challenging if only using the conventional queuing techniques. Our results demonstrate that graphical method explicitly presents both the deterministic and stochastic delay. We also illustrate that the theoretical background for the graphical methods in this particular application is inherently the diffusion approximation. Furthermore, we investigate the problems of disruptions occurred during a pre-timed traffic signal cycle. By diffusion approximation, we provide quantitative estimation on the duration that the effects of disruptions would dissipate.

DEDICATION

To my parents and my wife

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Bruce Wang, for giving me freedom in exploring research topics and for insightful guidance to me.

I would like to thank Dr. Yunlong Zhang, for the research opportunities that he has provided me and for his insightful suggestions on the research topics about traffic signal control. I would thank Dr. Luca Quadrifoglio, for the pleasant collaboration during the past. My thanks also go to Dr. Swaroop Darbha for the chances to talk with him and for his advice. All my committee members are very nice to me. I am grateful for such a convenient environment for the research at Texas A&M University.

I would thank all my past and current colleagues in the transportation engineering division, for the happiness they brought to me. I have greatly enjoyed being with them at Texas A&M University. In particular, I thank David Zeng for many agreeable discussions, and Yao Xing for the enjoyable collaboration. I also thank my roommate Lantao Liu, from whom I have learned very much and from whom I realized the attitude towards research.

I am grateful for all funding agencies that have supported me during my graduate studies. Specifically, I wish to thank the Southwest Region University Transportation Center (SWUTC).

I am deeply indebted to my parents for their love and unconditional support throughout years. Special thanks go to my wife Minle Xu, who greatly encouraged me during a time I was very depressed, and who must endure the lonely life as we live in different cities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE Page

1. INTRODUCTION

1.1 Background and Motivation

Transportation is both an art, providing practical solutions to fulfill the demand of the public with various advanced modern technologies and a science, founded on measurable experimental observations and rigorous reasoning. As a core in transportation, road intersection traffic signal control serves as a perfect subject featuring both art and science.

Traffic signal control remains a dominant approach to the urban mobility and a vital treatment to ease never-ending traffic congestion. According to the 2011 Urban Mobility Report, operational treatment saved $330 million from congestion in 2010 for fifteen large urban areas in the United States [100]. Realizing the potential utmost benefit, the goal of practice engineers is to design the most efficient operational strategy for traffic signals. The first challenge to achieve this goal is the well understand the fundamental properties of traffic signal operations.

Although there is a tremendous amount of research about traffic signal operations, there are still some important issues that demand answers. This study will analyze the following two questions at an isolated intersection (see Figure 1.1 for a physical illustration) to fulfill the gap in the literature: (1). What is the delay to vehicles considering left-turn traffic at a pre-timed signal? (2). What is the mean delay to vehicles and variance of cycle times at a vehicle-actuated signal?

1.2 Outline

The rest of this dissertation is organized as follows: Chapter 2 is the overview of the models for queues and delay at traffic signals. It provides background for discrete-time queueing model, fluid and diffusion approximations. Chapter 3 describes the

probabilistic models for left-turn bay blockage and spillback. The capacity and delay are estimated from the proposed models. Then Chapter 4 provides the approaches to characterize the delay, the mean and variance of green times at one-way intersection with actuated signals. We first propose an analytical model based on queueing models under the condition of heavy traffic. Then we present graphical methods in the next Chapter, in many ways similar to the Newell's original proposals, to illustrate the stochastic effects of arrivals in the context of diffusion approximation.

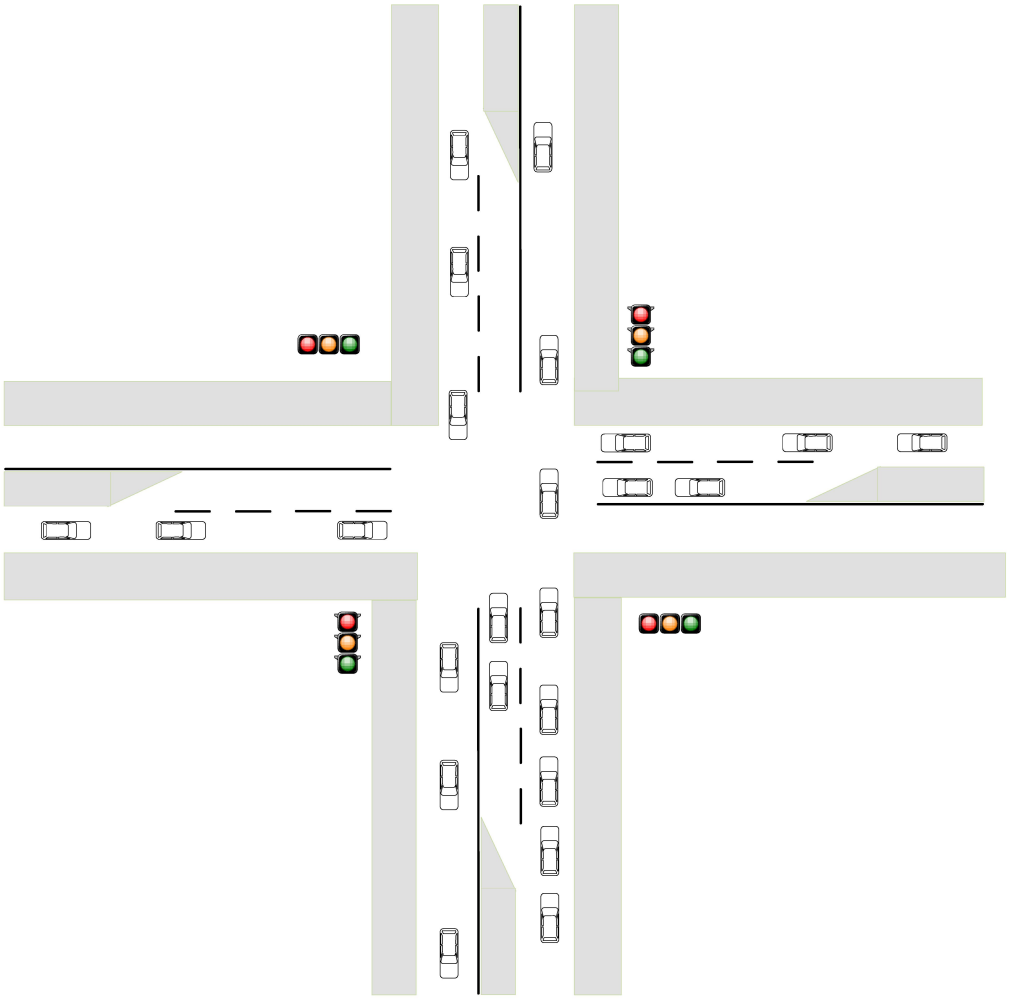

Figure 1.1: A signalized intersection.

2. REVIEW OF CLASSIC MODELS FOR TRAFFIC SIGNAL

2.1 Introduction

There is an appreciable amount of research on the performance evaluation of traffic signals, including the remarkable monograph by Newell [84]. Among those studies, the vehicular queueing at signalized intersections is central, because almost all the measures depend on the good understanding of this matter. In this section, we will overview theoretical models with a focus on the conventional traffic signals. Even though most of them were developed in 1960s, studies on traffic signal find those research relevant.

2.2 Models for Pre-Timed Traffic Signals

Pre-timed traffic signal alternates between red and green phases with fixed time periods in each cycle. Vehicles approach the intersection to either form a queue or pass without delay. The basic models on residual queues (overflow queue at the end of a green phase[1]) at pre-timed traffic signal can be classified as discrete-time and continuous approximation models. We will present them respectively.

2.2.1 Discrete-Time Models: Darroch's Approach

Darroch [33] presented a generalized discrete-time model, in line with Beckmann, McGuire and Winsten [16] and Newell [74]. Kleinecke[50] also obtained some similar results with Poisson arrivals. All the authors assumed that the departing vehicles were separated by constant time intervals of unit length when they crossed the stop line at the intersection. This assumption can be summarized as the following (see also in van Leeuwaarden [106]; Van den Broek et al. [19]; Bruneel and Kim [20]):

[1]Throughout this dissertation, we refer to the overflow queue as the residual queue.

Assumption 1 (DISCRETE-TIME ASSUMPTION). *The time axis is comprised of constant time* slots *of unit length, where each queued vehicle going straight or turning right needs a slot to pass the stop line. A traffic signal cycle, denoted by c, consists of r consecutive time slots designated as a red phase, and g consecutive slots designated as a green phase, where $c = r + g$, and r and g are integers. Further, if vehicles arrive in a slot and are delayed by a queue, then they join the queue at the end of the same slot.*

This assumption implies that the amber phase is equivalently comprised of effective green and red signals. Observations show that the first few vehicles in a discharging queue usually have larger departing headways (See, for example, Jin *et al.* [46] and Luttinen [61]). After the third vehicle departs the discharge headway remains constant [88]. These small fluctuations in headways contribute little to the vehicular delay. In addition, although the assumption that the signal phases are of multiple time slots is slightly strict, it proves to be a good approximation. The last statement in Assumption 1 allows the model to easily handle with the evolution of queues.

Let random variable $Y_{k,n}$ denote the number of vehicles arriving at the intersection during slot k in the nth cycle, where $k = 0, 1, 2, ..., c-1$. Throughout this section, by time $t = k$ we mean the time at the beginning of the slot k. Additionally, the time $t = c$ represents the end of one cycle or the time $t = 0$ at the beginning of a cycle. The following assumption is made for arrivals:

Assumption 2 (INDEPENDENT ARRIVAL ASSUMPTION [33, 106]). *The random variables $Y_{k,n}$, for all k and n, are assumed to be independent and identically distributed (i.i.d.). $Y_{k,n}$ therefore have the same probability generating function (p.g.f.) denoted by $Y(z) = \mathbb{E}(z^{Y_{k,n}})$, where z is a complex variable with $|z| \leq 1$.*

Clearly the above assumption for arrivals has limitations, the most obvious one being that the arriving vehicles may not be independent. However, Assumption 2 allows to describe a broad family of arrival processes, including the Poisson process and bulk arrivals, and it allows to derive the expected delay formula. Figure 2.1 represents a realization of discrete-time queues during a cycle according to both Assumptions 1 and 2. Note that queue may grow up during the green time because of bulk arrivals. In order to derive explicitly expression for delay of vehicles, we need the following assumption:

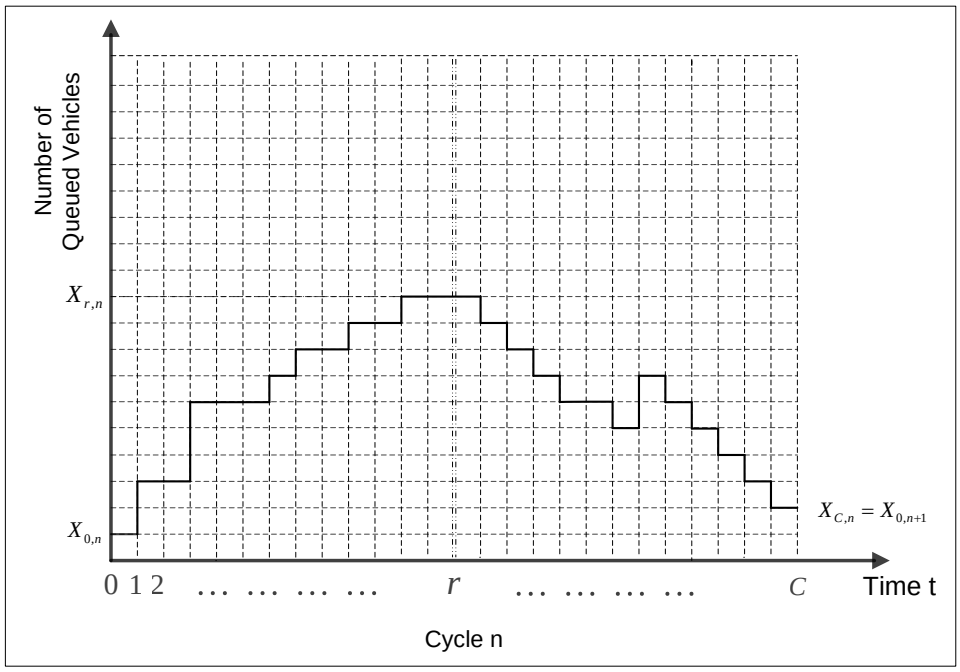

Figure 2.1: Illustration of discrete-time queues during a signal cycle.

Assumption 3 (DELAY ASSUMPTION [64, 106]). *There is no delay for vehicles that arrive during the green phase after the queue is cleared.*

In van Leeuwaarden [106], Assumption 3 is called Fixed-Cycle Traffic-Light (FCTL) assumption. This assumption is also described in McNeil [64]. Note that Assumption 3 implies that vehicles cross the intersection without slowing down, even they making right right, during the green phase without any queue. That is to say, if a bunch of them arrive during one slot after the slot of queue clearance, according to the assumption 3, they would not be delayed. In real-world situation, it might not be the case if these vehicles have to be separated by any reasonable spacing for safety reasons [106]. Nevertheless, the amount of delay shall be neglected when comparing to the counterpart by vehicle queues.

Let $Q_{k,n}$ denote the length of queue at the time $t = k$ in cycle n. Since we assume that each cycle begins with a red phase, $Q_{0,n}$ can be thought of as the residual queue (or overflow) at the end of green period of the cycle $n-1$ or as the initial queue during the cycle n. In green phase, $Q_{k,n}$ shall be equal to the number of arrivals during slot k plus $Q_{k-1,n}$ from the preceding slot less one departure vehicle if $Q_{k-1,n} > 0$. We therefore have the following recursive relations in green phase, i.e., for $k = r, r+1, ..., c-1$:

$$Q_{k+1,n} = \begin{cases} Q_{k,n} + Y_{k,n} - 1 & \text{if } Q_{k,n} \geq 0, \\ 0, & \text{if } Q_{k,n} = 0, \end{cases} \tag{2.1}$$

and in red phase, i.e., $k = 0, 1, 2, ..., r-1$,

$$Q_{k+1,n} = Q_{k,n} + Y_{k,n}. \tag{2.2}$$

In his original settings, Darroch [33] attempted to incorporate the left-turn vehicles on the shared lane with through traffic. Hence, he inserted a binomial distributed random variable $U_{k,n}$ to the first recursive relation, i.e., $Q_{k+1,n} = Q_{k,n} + Y_{k,n} + U_{k,n} - 1$

if $Q_{k,n} > 0$, by assuming it is complete random that whether or not a vehicle at the front of the queue turns left so he/she may have to wait. This section restricts the attention to the through and right-turn vehicles only.

According to Assumptions 1 to 3, the queue length $Q_{k,n}$ is a discrete-time Markov chain. The first of our concern is whether or not this Markov chain is able to converge to stationary distribution. Assuming that the mean and variance of the arrival $Y_{k,n}$ to be μ_Y and σ_Y^2, respectively, independent of k and n, we have the following results.

Theorem 1 (Darroch [33]) *The sufficient and necessary condition to guarantee the convergence of $Q_{k,n}$ to the stationary distribution is*

$$(g + r)\mu_Y < g. \tag{2.3}$$

Recall that g and r are integers that represent the number of green and red time slots, respectively. The proof of Theorem 1 can be found in Appendix B. In order to analyze equilibrium of recursive relation (2.1), let $\chi_{k,n}(z) = \mathbb{E}(z^{Q_{k,n}})$, and we have for $k = r, r+1, ..., c-1$,

$$\mathbb{P}(Q_{k+1,n} = i) \;=\; \sum_{j=1}^{i} \mathbb{P}(Y_{k,n} = j)\mathbb{P}(Q_{k,n} = i + 1 - j),\ i \in \mathbb{Z}_+, \tag{2.4}$$

and when the queue length is zero,

$$\mathbb{P}(Q_{k+1,n} = 0) \;=\; \mathbb{P}(Q_{k,n} = 0) + \mathbb{P}(Y_{k,n} = 0)\mathbb{P}(Q_{k,n} = 1). \tag{2.5}$$

Hence, it yields in green phase

$$\mathbb{E}(z^{Q_{k+1,n}}) \;=\; \mathbb{P}(Q_{k,n} = 0) + z^{-1}\mathbb{E}(z^{Y_{k,n}})(\mathbb{E}(z^{Q_{k,n}}) - \mathbb{P}(Q_{k,n} = 0)). \tag{2.6}$$

Note that we have $\chi_{r,n}(z) = \chi_{0,n}(z)(Y(z))^r$ in red phase. Moreover, since the stationary distribution does not depend on any specific cycle, we shall have $\chi_{c,n}(z) = \chi_{0,n+1}(z)$ in stationary state. If we use Q_k instead of $Q_{k,n}$ in stationary state with the p.g.f. $\chi_k(z)$, by recursively using Equation (2.6), we have

$$\chi_c(z) = \frac{Y(z)^g \zeta(z)^{-r}(\zeta(z)-1)\sum_{k=r}^{c-1} q_k \zeta(z)^k}{z^g - Y(z)^c}. \tag{2.7}$$

where $\zeta(z) = z/Y(z)$ and $q_k = \mathbb{P}(Q_k = 0)$ for $k = r, r+1, ..., c-1$.

To obtain the explicit formulas for q_k in Equation (2.7), one can investigate the roots of $z^g - Y(z)^c = 0$ within and on the unit cycle in the complex plane (see, for example, Takács [102], Newell [74] and van Leeuwaarden [106]). If there were g distinct roots, there would be $g-1$ equations for q_k except the one associated with root 1. This is because $\chi_c(z)$ shall be analytic within the unit cycle, the denominator and the numerator should both vanish at each root. In addition, $\lim_{z\uparrow 1} \chi_c(z) = 1$ gives another equation [106].

Hence, it is first of all to validate that $z^g = Y(z)^c$ has g distinct roots. It is worth noting that those roots are complex numbers and any theoretical attempt to obtain the number of distinct roots has to resort to Rouché's theorem in complex analysis [22, 6]. Adan, van Leeuwaarden and Winands[4] proved the existence theorem for a general class of distributions, which is discussed in Appendix B.

Then, by the normalization condition $\lim_{z\uparrow 1} \chi_c(z) = 1$, one obtains

$$\sum_{k=r}^{c-1} q_k = \frac{g - c\mu_Y}{1-\mu_Y} \triangleq K. \tag{2.8}$$

An intuitive interpretation of the above equation in the sense of probability is given by van Leeuwaarden [106] and van den Broek *et al.* [19]: If we rewrite as $g - \sum_{k=r}^{c-1} q_k =$

$(c - \sum_{k=r}^{c-1} q_k)\mu_Y$, it indicates that the mean number of green slots for discharging queues is equal to the mean number of delayed vehicles per cycle.

With Equation (2.8), one is able to express $\sum_{k=r}^{c-1} q_k \zeta(z)^{k-r}$ by the roots $z_1, ..., z_{g-1}$, i.e.,

$$\sum_{k=r}^{c-1} q_k \zeta(z)^{k-r} = K \prod_{k=1}^{g-1} \left(\frac{\zeta(z) - \zeta(z_k)}{1 - \zeta(z_k)} \right). \tag{2.9}$$

where $\zeta(z) = z/Y(z)$. Then Equation 2.7 becomes

$$\chi_c(z) \;=\; \frac{Y(z)^g(\zeta(z) - 1)}{z^g - Y(z)^c} \cdot K \prod_{k=1}^{g-1} \left(\frac{\zeta(z) - \zeta(z_k)}{1 - \zeta(z_k)} \right). \tag{2.10}$$

Equation (2.10) features the basic properties of residual queues at the intersection. For example, the mean delay of vehicles can be done by using this equation, and the distribution of length of queue is calculated by taking inversion of this equation.

The above model depends on the root-finding of $z^g = Y(z)^c$. This problem is of long historical interest. For example, Pollaczek [93, 94] and Crommelin [29] explored the roots of the equation $z^s = \exp(\lambda(z-1))$ (Poisson case) within the unit circle in the complex plane. McNeil [64] applied a similar method to the case of compound Poisson distribution. However, explicit formulas for the roots, which usually have contour integration involved, are difficult to obtain. Under some conditions, the expanded series expression for the roots may be found by the Lagrange inversion theorem (see pp. 132–133 in [117]). But the series usually converges slowly and thereby may not be readily applicable. In spite of analytic methods for special cases, there are well-documented techniques in queueing theory to address this issue numerically. One direct estimation is to solve a fixed-point equation. With the aid of the roots for $z^g = 1$ and an appropriate starting point, successive substitutions as

$z_k^{(n+1)} = \exp(2\pi ik/g)Y(z_k^{(n)})^{c/g}$, $k = 1, ..., g-1$, can be applied to find roots, where $i = \sqrt{-1}$ and $n \in \mathbb{Z}_+$ (see Janssen and van Leeuwaarden [44, 45], and Adan and Zhao [5]). More general methods can be found in Chaudhry, Harris and Marchal [23], Adan and van Leeuwaarden [4] and van Leeuwaarden [44], for example. Note that the inversion of Equation (2.10) gives the distribution of residual queues, though it demands numerical methods and relies on the Fourier transformation. We refer readers to the work by Abate and Whitt [2, 3] and Abate, Choudhury and Whitt [1].

Next, we will discuss one of the important results directly derived from Equation (2.10) and the bulk service queue model, respectively.

Average Length of Residual Queue

Equation (2.10) gives rise to many results. One of which is the mean length of residual queue important to the mean vehicular delay. It follows that the mean length of residual queue is $\frac{d\chi_c(z)}{dz}|_{z=1}$, i.e.,

$$\begin{aligned}\mathbb{E}(Q_0) &= \frac{(1-\mu_Y)^2}{g-c\mu_Y}\sum_{k=r}^{c-1}(k-r)q_k - \frac{\sigma_Y^2}{2(1-\mu_Y)} + \frac{1-\mu_Y}{2} \\ &+ \frac{c\sigma_Y^2 + (r\mu_Y)^2 - g^2(1-\mu_Y)^2}{2(g-c\mu_Y)},\end{aligned} \tag{2.11}$$

where $\mathbb{E}(Q_0) = \mathbb{E}(Q_c)$ is applied. Equation has been first derived in Darroach [33] (in a more general setting) and quoted in McNeil and Weiss [65], van Leeuwaarden [106]) and van den Broek et al. [19]. However, one can further simplify this equation into the following form by noting that $(r\mu_Y)^2 - g^2(1-\mu_Y)^2 = (r\mu_Y - g(1-\mu_Y))(r\mu_Y + g(1-\mu_Y))$ and $\frac{\sigma_Y^2}{2(1-\mu_Y)} - \frac{c\sigma_Y^2}{2(g-c\mu_Y)} = \frac{r\sigma_Y^2}{2(g-c\mu_Y)(1-\mu_Y)}$:

$$\mathbb{E}(Q_0) = \frac{(1-\mu_Y)^2}{g-c\mu_Y}\sum_{k=0}^{g-1} kq_{k+r} + \frac{r\sigma_Y^2}{2(g-c\mu_Y)(1-\mu_Y)} - \frac{(g-1)(1-\mu_Y)+r\mu_Y}{2}. \tag{2.12}$$

Taking a derivative with respect to $\zeta(z)$ in Equation (2.9) to get $\sum_{k=0}^{g-1} kq_{k+r}$, Equation (2.12) then turns to

$$\mathbb{E}(Q_0) = (1-\mu_Y)\sum_{k=1}^{g-1}\frac{1}{1-\zeta(z_k)} + \frac{r\sigma_Y^2}{2(g-c\mu_Y)(1-\mu_Y)} - \frac{(g-1)(1-\mu_Y)+r\mu_Y}{2}. \tag{2.13}$$

McNeil [64], in a slightly different setting from the discrete-time queueing model, generalized the method used by Crommelin [29] to obtain the expression for $\mathbb{E}(Q_0)$ in power series and then he discusses some special cases. Kleinecke[50] also provided a formula for $\mathbb{E}(Q_0)$ in power series for Poisson arrivals.

Perhaps the complexity of the root formulas motivated Darroach [33] and van den Broek et al. [19] to work directly with Equation (2.12) and to provide bounds for $\mathbb{E}(Q_0)$. By noting that $q_r \leq q_{r+1} \leq ... \leq q_{c-1}$, it can be shown :

$$\begin{aligned} & \frac{r\sigma_Y^2}{2(g-c\mu_Y)(1-\mu_Y)} - \frac{r\mu_Y}{2} \leq \mathbb{E}(Q_0) \\ \leq\ & \frac{r\sigma_Y^2}{2(g-c\mu_Y)(1-\mu_Y)} + \frac{r\mu_Y}{2}\left(\frac{(g-1)(1-\mu_Y)}{(g-c\mu_Y)} - 1\right). \end{aligned} \tag{2.14}$$

where the bounds of $\sum kq_{k+r}$ in van den Broek et al. [19] are applied. A new approximation to $\mathbb{E}(Q_0)$ is given by van den Broek et al. [19] based on heavy traffic limit and scaling argument; that is $\frac{(c\mu_Y)^2}{g^2}\frac{r\sigma_Y^2}{2(g-c\mu_Y)(1-\mu_Y)}$ (by plugging Equation (4.13) into (4.11) in [19]). It is very interesting to compare this approximation with the lower bound in Equation (2.14): they are close to each other when $c\mu_Y \to g$.

Bulk Service Queue

There is an alternative way to avoid the complex analysis of iterated relationship as in Equation (2.1). If one approximates Equation (2.1) by

$$Q_{g,n} = \max\{Q_{g,n-1} + \sum_{k=0}^{c-1} Y_{k,n-1} - g, 0\}, \tag{2.15}$$

then one regards the signalized intersection as serving queues with bulk service of g vehicles at a time. This actually is the classic *bulk service queue*, first studied by Bailey [14] and Downton [35] with Poisson arrivals.

Newell [74] used this approach and made a more simpler assumption which leads to a binomial arrival per slot, i.e., $Y(z) = 1 - \mu_Y + \mu_Y z$ with a mean μ_Y. Under this assumption, the solution for Q_0 can be derived exactly in the equilibrium condition. Newell further identified a key parameter $\mu = [g - \mu_Y(r+g)][rg/(r+g)]^{-\frac{1}{2}}$ to distinguish the traffic conditions between light and heavy. Interestingly, there is a meaning of the factor $[rg/(r+g)]^{-\frac{1}{2}}$ in μ. If we consider the heavy traffic case, i.e., $\mu_Y \to g/(r+g)$, then $r/(r+g)$ would be the approximation of variance-to-mean ratio (denoted by I) of the total arrivals during a cycle, i.e., $\frac{\mu_Y(1-\mu_Y)(r+g)}{(r+g)\mu_Y}$. Hence, the factor can be approximately considered as Ig, the same scaling factor Newell used in diffusion approximation to the queues [76]. In Appendix B, we will discuss how this factor arises.

If $\mu > 1$, then the traffic demand is light and the queue length can be calculated directly under the assumption that there is no queue at the beginning of the cycle. The parameter μ actually arises when analyzing this case. If $\mu < 1$, the previous methods of probability generating function can be employed. For binomial arrivals, $z^g - Y(z)^c$ is a polynomial with degree c. This fact allowed Newell to find the

approximation to the roots of $z^g = Y(z)^c$ under the heavy traffic limit as the nearly critical arrival rate, one of the interesting results being [74]

$$\mathbb{E}(Q_0) = \frac{rg}{2(r+g)(g-\mu_Y c)} + O(1), \text{ for } \mu_Y \to g/(r+g). \tag{2.16}$$

This expression yields almost the same magnitude of the lower bound in Equation (2.14) for $\mu_Y \to g/(r+g)$.

When the number of arrivals is larger than one during one slot, the solution by the bulk service queue would become an approximation. As noticed in McNeil [64], this is simply due to the Assumption 3 that some of vehicles may pass intersection without delay. Nevertheless, if we assume such effect on the residual queues is negligible, we will obtain the upper bound of $\mathbb{E}(Q_0)$[106].

2.2.2 Diffusion Approximation: Newell's Approach

Exact solutions to the average residual queues due to overflow are often cumbersome to be readily applicable. The difficulty of obtaining simple expressions for residual queues has motivated many scholars to look for approximation methods. The possibility of approximation was probably based on an observation that in most applications the average delay appears insensitive to the detailed stochastic structure of arrival processes, more so when traffic becomes heavy. It indicates that there might be a way that does not depend on the particular discrete distribution.

Newell [76] was the first to use the technique of diffusion approximation to obtain the average residual queue length and the average delay. As remarked by Newell in [75], if the number of arrivals is fairly large, it is convenient to consider the queue length as a continuous random variable normally distributed about its mean. So we can use Brownian motion, diffusion or central limit theorem type approximations. The critical argument here is that the relative changes in the residual queue length

from one cycle to the next are small. Although the residual queue length would drift over a wide time in order to produce a certain value of variance, it only does so in small displacement [80]. An illustration of such approximation is seen from Figure 2.2, where the cycle length and green time are considered large enough.

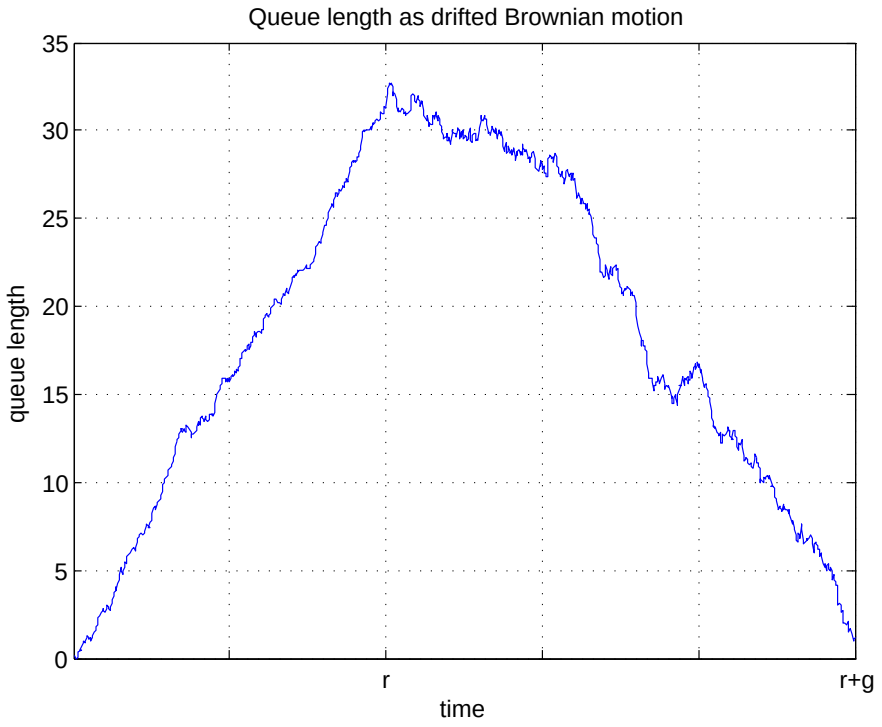

Figure 2.2: Illustration of queue length as Brownian motion in one cycle assuming r and g large enough. Brownian approximation allows the queue length occasionally to decrease a little during the red signal.

We denote Q_t as the queue length at time t, arrival rate as q and the departing headway from the queue as $1/s$. We also use g and r to denote the length of green and red times (here they are continuous variables). Consider the traffic demand close to saturation. Under equilibrium condition, the variance-to-mean ratio of arrivals and

departures might be pretty stable, especially for the high demand. Hence, if we let the cumulative number of arrivals be $A(t)$ and number of departures be $D(t)$ at time t, it is reasonable to assume $Var(A(r+g) - D(g)) = Iq(r+g)$ for some proper variance-to-mean ratio I. Note that the standard deviation of $A(r+g) - D(g)$ is also in the order of $(sg)^{1/2}$.

The diffusion approximation can be made if $\mathbb{E}(Q_0)$ is large compared with $(sg)^{1/2}$ and Q_{r+g} has a variation around Q_0 at most in the order of $(sg)^{1/2}$ (with probability one). These conditions are expected to be met in heavy traffic situation. Then $Q_t/\mathbb{E}(Q_0) - 1$ will have the expectation almost 0 and the variance in the order $O(1)$. That is to say, $Q_t/\mathbb{E}(Q_0) - 1$ behaves like a standard Brownian motion on a time scale large compared with the cycle length. The above analysis mainly follows Newell [76]. Miller [67] employed the similar idea when developing the delay and average residual queue length formulas.

To make the above analysis mathematically tractable, let $F_Q(z) = \mathbb{P}\{Q_0 \leq z\}$ and $F_{A-D}(x) = \mathbb{P}\{A(r+g) - D(g) \leq z\}$. Then Newell [76] deduced the Wiener-Hopf type integral equation (see also pp. 118–119 in Cox and Smith [28] and Noble [89])

$$F_Q(z) = \int_0^\infty F_Q(x) dF_{A-D}(z-x). \tag{2.17}$$

By this equation, the implication is that we allow the case of $A(r+g) - D(g) < 0$ to occur and think of such case as an imaginary part. The integral between zero and infinity guarantees that the queue length larger than or equal to zero. In the current literature, this phenomenon can be captured by modeling the queue as a reflected Brownian motion. Although using Fourier transform can solve Equation (2.17), we can obtain the approximation based on the above recognition on Brownian motion. By expanding $F_Q(x)$ in the righthand of the above equation to second order, we can

obtain [76]

$$\frac{dF_Q(z)}{dz} \cdot \mathbb{E}(A(r+g) - D(g)) = \frac{1}{2}\frac{d^2F_Q(z)}{dz^2} \cdot \mathbb{E}(A(r+g) - D(g))^2. \tag{2.18}$$

This is the actually the forward equation for diffusion process. The average residual queue length is then deduced from the solution of the Equation (2.18), yielding [76]

$$\mathbb{E}(Q_0) \approx \frac{Iq}{2s}\left[\frac{g}{r+g} - \frac{q}{s}\right]^{-1}. \tag{2.19}$$

This approximation is accurate near saturation since the original formula for the distribution of residual queue has zero probability for the residual queue of length zero. When analyzing residual queue delay by this approximation, Newell [76] found that the error between this formula and Webster's might not be small when q decreases. One can remedy the problem by finding another asymptotic solution that rapidly decreases as q decreases. To obtain the second approximation, Newell [76] turned back to Equation (2.17). By assuming the scaled variable $[A(r+g) - D(g)]/(IsG)^{1/2}$ to be Gaussian distributed, the final average residual queue length is [76]

$$\mathbb{E}(Q_0) \approx \frac{sg - q(r+g)}{\pi}\int_0^{\pi/2} \frac{\tan^2\theta}{\exp[(sg - q(r+g))^2/(2Isg\cos^2\theta)] - 1} d\theta. \tag{2.20}$$

The virtue of the above methods developed in [76] is the systematic treatment of approximation to the queueing problem at the pre-timed traffic signal, though some steps rely on intuition. When applying this approximation to the delay, the formula has much agreement with Webster's. It demonstrated that the power and accuracy of the asymptotic approximation. The above approximation techniques are further emphasized and developed in the book by Newell [83].

2.2.3 Vehicular Delay

In most applications, the average vehicular delay is the central concern and serves as the important performance measure for traffic signal systems. In order to estimate the delay, one may start with a simple situation, i.e., no overflow queue at the commencement of red time. When vehicles arrived, the delay will be due to the red times and the wait in the queue discharging process during green intervals. If the traffic demand is moderately large and can be treated as continuous fluid, then the so-called deterministic delay will be very easy to obtain. However, if the traffic flow occasionally exceeds the capacity or if the demand becomes high, the delay is directly related to the residual queues from the previous cycle. In this case, one can hypothetically consider the residual queue to be postponed to discharge after all arrivals during one cycle depart the cycle. Then it is relatively easy to get the average delay by combining the previous two components. In fact, this approach was essentially the same as that in the original work by Beckmann, McGuire and Winsten [16].

Certainly there is an issue about the effective green or red times and loss time. In real situations, it might be appropriate to consider the loss time as deterministic, whose value could be statistically determined by real data. Hence, the effective signal times are treated as deterministic ones. However, some observations in modern days indicate that the loss time might be random and such effects may have adverse consequences. Indeed, people are frequently observed to talk to their cell phones during waiting for green time. It results in an unexpected longer loss time and longer residual queues in the next consecutive cycles during peak hours. While this phenomenon needs to be further studied in practice, it may not be a problem in most situations where the phenomenon rarely takes place. Therefore, it is fairly reasonable

to ignore the random effects of loss time.

For discrete-time models, the expected amount of total waiting time per cycle in terms of time slot is given by simply summarizing up all expected queue length in each time slot, as indicated in Darroach [33]. Then the average delay per vehicle can be obtained easily. An alternative perspective of obtaining the average delay per vehicle is to use Little's law by calculating first the average queue length in an arbitrary slot, as in van Leeuwaarden [106] (There is a typo in [106], i.e., $\mathbb{E}D$ shall be given by $\mathbb{E}\overline{X}/\mu_Y$ for Equation (18) in [106]). Nevertheless, the above two approaches are essentially the same. Let w be the delay for an arbitrary vehicle, then $\mathbb{E}(w)$ is expressible as [106]

$$\mathbb{E}(w) = \frac{r}{2c\mu_Y(1-\mu_Y)}\left[2\mathbb{E}(Q_0) + \frac{\sigma_Y^2}{1-\mu_Y} + r\mu_Y\right]. \tag{2.21}$$

Miller [67] and McNeil [64] applied the same reasoning to derive the average delay as that in Beckmann, McGuire and Winsten [16]. They generalized the distribution of arrivals by considering the variance-to-mean ratio and made some approximations. Their results are very useful in many practical cases, and we will discuss Miller's work in Appendix .

Perhaps the most appropriate model readily for applications is to regard the arrivals as stochastic fluid. Newell [76] concluded that the average total wait per cycle for all cars is given approximately by

$$\mathbb{E}(w) = \frac{1}{q(r+g)}\left\{\frac{qr^2}{2(1-q/s)} + (r+g)\mathbb{E}(Q_0) + \frac{qrI}{2s(1-q/s)^2}\right\}. \tag{2.22}$$

where I is the variance-to-mean ratio for the difference between arrivals and departures as in the previous section. In the bracket, the first two terms account for

deterministic and residual queue delay. While the third term is a very crude approximation as it is not accurate for extremely light flows, it is accurate for heavy flows when central-limit theorem can be applied and it is fairly good for the Poisson distributed arrivals. This formula along with the expression for $\mathbb{E}(Q_0)$ is accurate and is comparable with Webster's formula [30].

Although most of efforts have been made to estimate the mean delay, understanding the distribution of delay plays an important role in practice as well. Newell [71] proposed perhaps the first distribution model of the delay to the vehicle coming to the signalized intersection. He explored some special cases and showed to a first approximation one can obtain the mean delay disregarding the detailed statistical assumptions. This method was considered too much elaborate by Newell himself if one merely wishes to get the mean delay [76]. In view of the aforementioned methods used in the mean delay models, it seems that there is no essential difficulty to obtain the description of distribution of delay. However, it seems impossible to get the closed-form expression for delay distribution. Any attempts in this subject must involve numerical techniques to solve the dynamical equations which characterizes the process. See, for instance, Heidemann [42] and Van Zuylen and Viti [107, 109]. Among others, van Leeuwaarden [106] proposed a technique based on the property of Fourier series to make the calculation of distribution possible in terms of probability generating function, though more numerical efforts have involved.

The above delay analysis is largely concerned with the isolated intersection with traffic signals. If the intersections on an arterial are the concern, one cannot treat all of them as isolated due to some interdependency among them. However, the critical intersection, where the crossing traffic is with the largest degree of saturation among all intersections, is reasonably to be treated as isolated [85]. To analyze the performance along the arterial, one can decompose the delay into deterministic and

stochastic parts as that for isolated signals. Newell [85] provided some arguments that whenever no turning traffic on the arterial the stochastic delay may be largely influenced by critical intersections and should be small compared with deterministic delay. Although the explicit expression for the total delay is very hard to obtain, some qualitative descriptions are at least to be done based on the understanding of the above two components of delay.

2.2.4 Effects of Left-Turn Vehicles

All the discussions so far have been made on the signalized intersections without left-turn traffic. If there is enough space to accommodate the left-turn vehicles along one approach, then one can consider through and turning traffic independently. Otherwise, the situations can be complicated as there might be some interdependence between through and turning traffic. Nevertheless, one may expect them to cause the delay more or less proportional to the fraction of turning vehicles.

The influence of unprotected left-turn traffic on the shared lane is the first subject to be investigated in the literature. If a lane shared by through and turning traffic, the turning vehicles may block the traffic on the shared lane whenever they need to wait for filtering through the opposing traffic [9]. Newell [72] considered an unsignalized intersection with only one lane for each approach. He estimated the capacity through a Markov chain model by assuming that either a left-turn vehicle at the intersection must yield to the through vehicles from another approach or both left-turn vehicles can depart simultaneously. Darroach [33] approached this problem by assigning each vehicle a probability independent of other factors to determine whether or not it could make turn. Although this approach is designated for the shared lane traffic of the discrete model, the independence of probability for capability of turning may result in the limited use for practice.

It seems that the protected left-turn bay may relieve the pressure of reduction in capacity. However, the effects of left-turn movements on the capacity and delay would be severe if there is a protected left-turn bay not long enough. The entry of left-turn bay may be blocked during the red phase or even on a portion of the left-turn green phase [66], leading to a significant reduction in left-turn capacity. Obviously, such reduction may vary with the fraction of left-turn vehicles and the green splits in a complex manner. Moreover, sometimes there are left-turn bay overflow spills over onto the adjacent through lane. The spillback (or spillover) is a difficult phenomenon to deal with in analytical model. Apart from a substantial variation in the intersections between through and turning movements, the realistic issue is how left-turn drivers behave in response to the overflow (See Chapter 31 in HCM2010 [105]).

Messer and Fambro [66] conducted a simulation study to examine the capacity in terms of left-turn bay length and signal phasing. They found that the lagging green phase slightly better than leading green and provided empirical formula for the length of left-turn bay. Zhang and Tong [122] proposed a probabilistic model for protected left-turn capacity at a signalized intersection with a short left-turn bay. They estimated the probability of the left-turn bay blockage and left-turn spillback by considering the order of through and left-turn arrivals during one cycle. Yin, Zhang and Wang [120, 121] performed some follow-up studies and used Newell's results on residual queue to investigate the capacity and delay. No matter what kind of probability approaches have been made, the basic structure in the above studies is simply to find a coefficient to smoothly combine two components of delay, i.e., the delay with the occurrence of blockage or spillback and the delay without. This treatment is somewhat crude but perhaps practical. Another difficulty in [120, 121] is the choice of probability of arrivals. With the increase of traffic intensity, the arrival

pattern is not Poisson any more. On the other hand, the blockage and spillback depend on some detail discrete structure around the short left-turn bay. In this respect, perhaps the discrete time model is more appropriate to study this subject.

Recently, Haddad and Geroliminis [41] used the negative binomial distribution as in Zhang and Tong [122] to quantify the uncertainty of queue spillback from the left bay. However, this work still does not overcome the difficulties mentioned earlier. Among others, Akçelik [9] provided a treatment of shared lane capacity from the practice perspective. Wang and Benekohal [110] considered the effects of platoon arrival rate to the left-turn operation. Qi et al. [95] considered the overflow of left-turn traffic and applied discrete-time Markov chain model to analyze left-turn queue lengths at signalized intersections. Liu and Chan [60] made an optimization model to maximize the capacity by considering the queue spillback and left-turn bay blockage. However, due to the excessive scenarios and the discrete nature of the problem, it is unlikely to attain the original objective.

Having discussed various existing approaches, we have to conclude that so far no overall mathematical model included all the identified variables has been developed for the spillback and left-turn bay blockage. A clear understanding of the major process during these phenomena taking place is needed the first. To do so one has to identify which factors are more important than others and then one can apply some mathematical methods. Perhaps a similar graphical method in Newell [86] is worth an attempt to this subject.

2.3 Models for Vehicular-Actuated Signals

Vehicle-actuated signals, on the contrary to pre-timed signals, is capable to respond to fluctuations of arrival patterns. For the basic full actuated control strategy, the green time for each approach traffic consists of two components, i.e., queue clear-

ance time and green extension. Queue clearance time refers to the interval that takes for the vehicular queue to vanish. The green extension, immediately followed the queue clearance time, is controlled by a vehicle detector which usually places upstream of the intersection. Once a vehicle passes the detector, it extends the green time to a fixed interval. Such fixed interval is usually called unit extension (also called passage time or gap time in the literature). During the unit extension, if there is any vehicle passing the detector, the green will be extended to another unit extension further. Otherwise, the green will terminate and switch to red time. Usually the signal controller has set a minimum and a maximum value for green time. Needless to say, the green time is forced to be longer than the minimum and no longer than the maximum no matter whether there is a triggered unit extension or not. In the basic controller setting, once the unit extension is set, it cannot be changed. On the contrary, modern controllers allow it to change according to the traffic demand. Through this dissertation, we will not discuss the nonbasic settings. Figure 2.3 shows a typical cycle for one approach.

The investigation of vehicle-actuated signals is much more complicated than pre-timed signals. It seems impossible for a model to involve all the relevant parameters in the real setting. This is due to not only the interdependence of signal times for different approaches, but also the order of queue clearance times of opposing approaches.

Apart from the efforts of exploring models to describe the behaviors of the signals, the studies generally indicate that a unconstrained vehicle-actuated signal at the intersection of two one-way streets behaves better than an optimal pre-timed signal. However, it may not be so at an intersection of two-way streets [87].

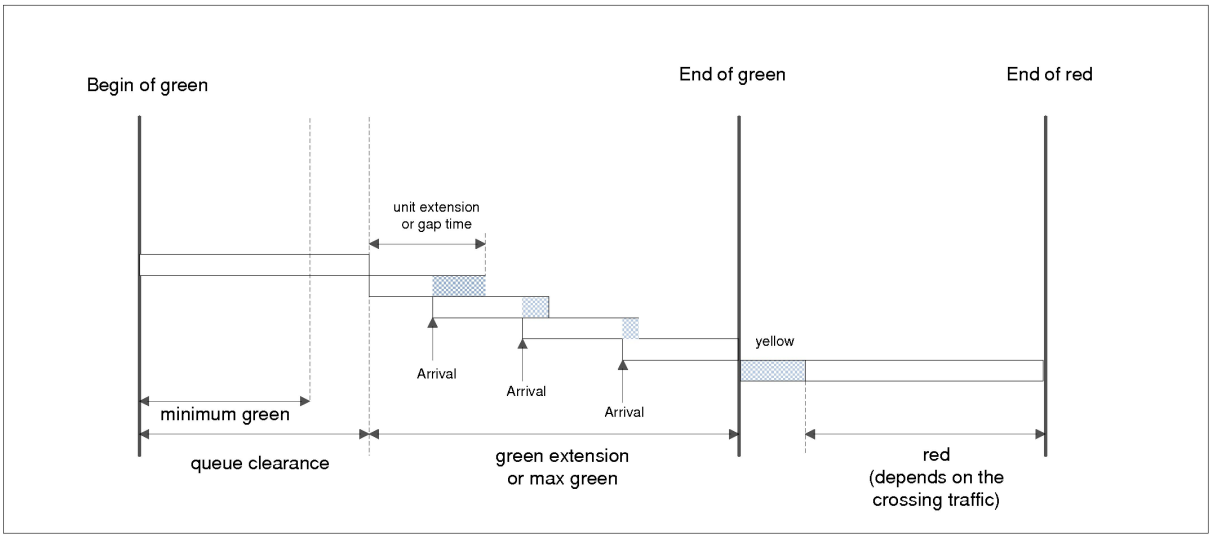

Figure 2.3: A typical cycle for vehicle-actuated signal (adapted from the Figure 4 in [10]).

2.3.1 General Approaches

Research on vehicle-actuated control by the tools of probability starts in the 1940s. Garwood [39] directly used Poisson distribution to describe the arrivals and find the probability that a vehicle has to wait for the whole maximum green time. Clayton [24] studied the average delay at vehicle-actuated signals on a road with light traffic. Tanner [103] estimated the delays of two opposing streams of vehicles trying to cross a length of road wide enough for only one vehicle at a time. Based on his assumption, he found explicit results of waiting times for some special cases.

Darroch, Newell and Morris [32] gave a comprehensive analysis based on assumptions of general distribution of departure vehicle headways and lost time, and of exponentially distributed arrival headways. The authors used results of busy period of $M/G/1$ queues to investigate the influence of green extension to signal cycle and

average delay. This work might be the first to obtain some general results for a one-way intersection with actuated signals. However, the limitation of this study lies on the assumption of Poisson arrivals. As remarked in [84], since their model did not assume minimum green time, in the extreme case with no traffic flows, it will allow signal to switchback and forth many times during a short period.

Based on a binomial vehicle arrival, Dunne [36] presented a discrete model following Darroch, Newell and Morris [32]. He gave an analytic expression of the probability generating function for vehicle delay and showed a connection to the discrete cases of [32]. Lehoczky [53] also considered the discrete type of signal times and compound Poisson arrivals. However, the setting fell short of realistic meaning. In Lehoczky [52], he regarded two one-way streets as a server with two input channels and the control mechanism as alternating priorities queues. Little [59] did a similar study and obtained probability distribution of queue length in the steady state.

Taking bunching arrivals into consideration, Cowan [26] considered vehicle-actuated traffic control. The intersection under his consideration was also the one of two one-way streets without maximum and minimum green times. He basically assumed the compound Poisson arrivals, i.e., the bunch of arrivals, separated by inter-bunch headway with shifted exponential distribution and the intra-bunch gap with a constant one unit time. Since his model was essentially of the discrete time type, the treatment was similar to that in Lehoczky [53, 52] to some extent.

There are also many other proposed models similar to the ones above. Examples include Morris and Pak-Poy [69]. Recently, Viti and Zuylen [108] developed a computational model based on an assumption of the time dependent distribution of arrivals. and attempted to incorporate many real parameters such as maximum and minimum green times. The temporal evolution of queue length, traffic signal sequence probabilities, delay, and waiting time could be computed based on their

proposed equations. Their model served suitably as a probabilistic evaluation for microscopic simulations. However, it seems that the computational model did not give direct description of what would happen had some parameters changed.

On the practical front, Kruger, May and Newell [51] studied the real implementation of vehicle-actuated control at an intersection. They investigated the location of detectors for a specific strategy: queue control. Queue control strategy aims to detect the end of the queue and switches the signal at the same time that the end of the queue is expected to reach the intersection. Based on numerical simulation, they showed that, in general, queue control at an isolated intersection is better than other control strategies.

The above discussed the attempts which have been made for finding exact methods of vehicle-actuated signals. One may wonder why we have to find the general solution if there is only a little improvement for understanding the properties of actuated signals. In fact, any efforts looking for some models should aim at making inference about the real world. The accurate solutions for the problem may be too complicated to obtain the useful information. In this respect, perhaps some appropriate approximations made in the mathematical analysis can be greatly of help for making predictions about the real applications. Newell [80, 87] made the progress and applied diffusion approximation to the analysis of vehicle-actuated signals.

2.3.2 Diffusion Approximation

Deterministic fluid model and diffusion approximation are two useful and appropriate tools discovered by Newell to analyze the control strategies of traffic signals, first applied to pre-timed signals [76]. They are also proper to analyze the performance of vehicle-actuated signals. While the deterministic fluid model can reveal the basic dependence relationship of phase time between different traffic approaches,

the diffusion approximation easily captures the stochastic effects of traffic demand on the phase time and vehicular queues.

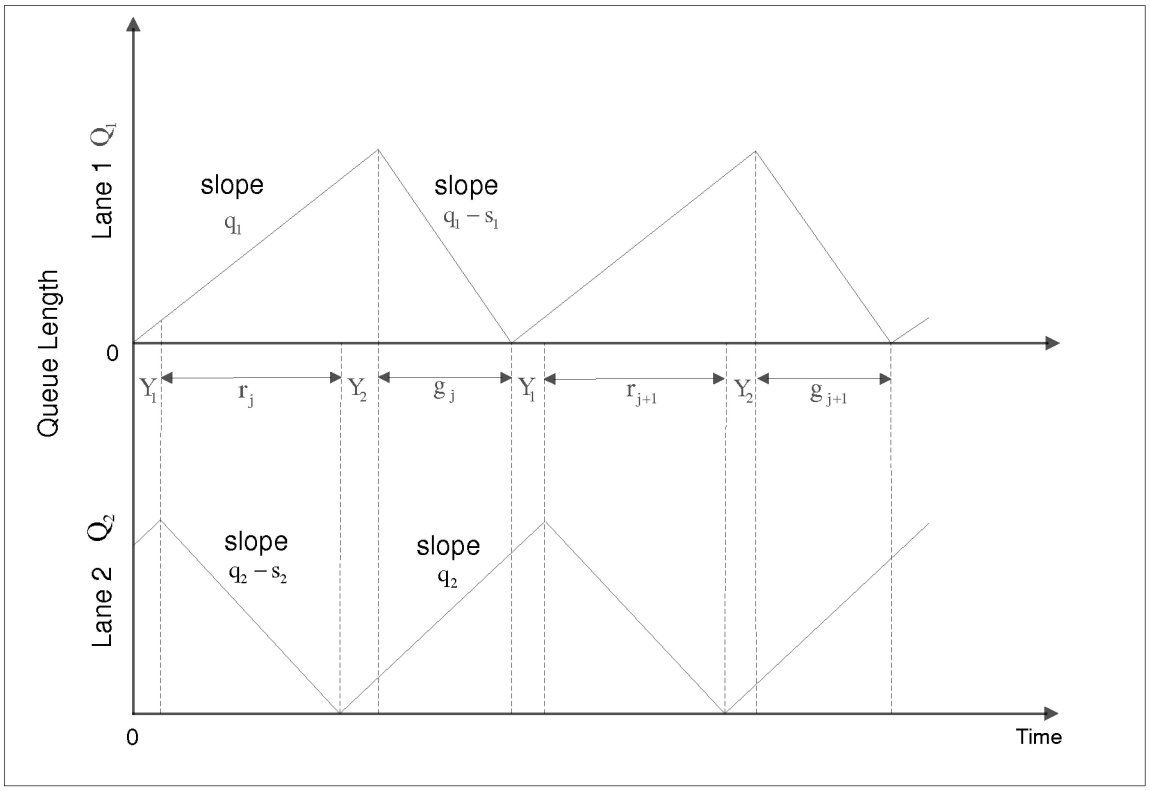

Figure 2.4: Continuous representation for queue length v.s. time at one-way intersection (reproduced from Figure 1 in [80]).

Newell [80] was primarily concerned with vehicle-actuated control strategies under the arrival probability invariant to time translations at the intersection of two one-way streets. In his setting, the signal switches the green as soon as the queue vanishes. As in Figure 2.4 for deterministic fluid model, the two lanes, indexed by 1 and 2, have the arrival rate q_i and the discharge rate s_i, $i = 1, 2$. Y_1 and Y_2 are loss times on two approaches, respectively, and g_j and r_j are green and red times in jth cycle,

respectively. Obviously Figure 2.4 illustrates a linear recursive relations of two signal phases, whose condition of stability was $\alpha = 1 - q_1/s_1 - q_2/s_2 > 0$ shown by Newell [80]. In fact, the relationship between r_j and g_j defines a Markov chain. If one searches for two-step dependence, one will find a relationship between r_j and r_{j+1} (or between two consecutive green times). As long as the traffic intensity is large, it is appropriate to think of r_j as some random value varying about the mean. One can also justify such variants during two consecutive cycles should be small if one proceeds to scale the time r_j as αr_j. It implies that one is able to apply the same trick of diffusion approximation to the distribution of red times. Let $f_j(x) = \frac{d}{dx}\mathbb{P}\{\alpha r_j < x\}$, then under certain conditions the processes of r_j are approximately described by the following Fokker-Planck type equation [80]

$$f_{j+1}(x) - f_j(x) \cong -\frac{d}{dx}[f_j(x)a(x)] + \frac{1}{2}\frac{d^2}{dx^2}[f_j(x)b(x)]. \tag{2.23}$$

where $a(x) = \mathbb{E}(\alpha(r_{j+1} - r_j)|\alpha r_j = x)$ and $b(x) = \mathbb{E}(\alpha^2(r_{j+1} - r_j)^2|\alpha r_j = x)$. The values of $a(x)$ and $b(x)$ are given in [80], where it did not assume any specific probability function for the arrivals but a linear relationship between the mean and variance of arrivals within any given time period. The equilibrium solution of the above equation is obtained by setting $f_{j+1}(x) = f_j(x)$ and is shown to be a gamma-type distribution. This conclusion has some practical meaning. First, it is rare to observe a heavy-tailed distribution for cycle time. In fact, it actually decays exponentially, meaning that under equilibrium condition the cycle times should fluctuate around the mean by a small factor. Second, if the distribution of cycle time is not equilibrium, the actual time to reach equilibrium is proportional to α^{-2} [80].

As for delay, the important feature is that the total delay on approach 1 is proportional to the variance of red time $Var(r_j)$ and $(\mathbb{E}(r_j) + Y_1 + Y_2)^2$. Based on this

fact, one can compare the performance of actuated control with that of pre-timed counterpart. The conclusion is drawn for the symmetric two one-way streets that a vehicle-actuated signal generates only one third the delay of an optimal pre-timed signal.

Newell [87] further explored the two-way intersection with unconstrained vehicle-actuated signals, i.e., no maximum and minimum green times. The investigated control policy is to switch the green once all queues on opposing directions have vanished. For unbalanced flows on the opposing approaches, Newell noted that the approach with lower traffic intensity preferred a pre-timed control to an actuated control as to eliminate the effect of $Var(r_j)$ on delay, but it was not so to the approach with higher traffic intensity. This tension may exert the influence on the performance of signal. Nevertheless, the actuated signal gives less delay than the pre-timed signal in a large range of arrival and saturation flow rates. For almost balanced flows, the problem lies in the fact that the queue on either approach can vanish first due to the stochastic effect. In the diffusion approximation, this leads to the expectation of green time given red time $\mathbb{E}(g_j|r_j)$ that is proportional to r_j and $(r_j + Y_1 + Y_2)^{1/2}$, not a single linear relationship with red time any more. Moreover, the uncertainty in the red time becomes small compared with the mean [87]. The main consequence from the above facts is that the investigated actuated control is inefficient for nearly saturated steady flows compared with pre-timed control [87].

Recently, Boon *et al.* [17] also used diffusion limit based on the studies about the pooling system [18] and provided the results for both heavy and light traffic demands. They interpolated the results for the moderate demand case. However, their results might have limited practical meaning.

In sum, the deterministic fluid model and diffusion approximation can be regarded as the basic tools to analyze most of queueing related traffic problems. Because the

fluid model is the first order approximation and the diffusion is the second order approximation, they are suitable to most static realistic applications. In order to tackle with time-dependent arrivals, the diffusion approximation can be generalized [82, 77, 78, 79]. The recent development involves the application of second order approximation to traffic flow (See Jabari and Liu [43]).

2.4 Learning-Based Traffic Signal Control

In addition to the pre-timed and vehicle-actuated signals, there are emerging technologies that enable other types of signal control based on the past information and the prediction of traffic flow. The adaptive signal control systems, including Urban Traffic Control System (UTCS), Split, Cycle and Offset Optimization Technique (SCOOT), and Sydney Coordinated Adaptive Traffic System (SCATS), have been more or less employed advanced optimization or learning techniques to implement various control strategies. More recently, the genetic approach [92], enforcement learning [119] and scheduled based [118] signalized intersection control have been investigated. However, the architectures and control policies in above systems are too complex to be explicitly analyzed. The quality and the optimality of the control performance are not guaranteed as well. Besides, these signal control systems heavily rely on the advanced surveillance devices such as digital cameras, wireless communication devices, and associated computing systems, which may be too expensive for public agencies to afford. Hence, they are not widely adopted in practice.

3. PRE-TIMED SIGNAL: MODELING SPILLBACK AND BLOCKAGE

3.1 Introduction

One of the most important issues of signal timing plan relies on the left-turn treatment. [1] Usually a protected left-turn signal phase, before (leading) or after (lagging) through signal, is applied to a signalized intersection with high traffic demand. The traffic delay, as a criterion, is used to evaluate the signal operations. For a light traffic demand, the average delay can be easily determined by treating left-turn and through traffic independently. However, when the demand is very heavy, say, close to the capacity, the situations become much more complicated since not only residual queues become a problem but also some interactions would exist between left turns and through vehicles. Such interactions affect the delay to arriving vehicles and the capacity of the signalized intersection as well.

A common problem for leading left-turn operation is the blockage to left-turning vehicles by through traffic, particularly so at an intersection with a short left-turn bay. During the peak hour, some vehicles in the through lane might not be able to depart at the end of one cycle, resulting in an increased probability of left-turn blockage. In turn, the blocked left-turning vehicles may also delay the through traffic to enter the intersection. Those problems may not exist during a lagging left-turn operation since left-turning vehicles intend to spill out of the bay under heavy traffic. In this case, the through capacity is reduced, leading to an increase of total delay as

[1] Part of this chapter is reprinted with permission from K. Yin, Y. Zhang, and B. Wang, Modeling Delay During Heavy Traffic for Signalized Intersections with Short Left-Turn Bay, Transportation Research Record: Journal of the Transportation Research Board, No.2257, pp.103-110, 2011, and K. Yin, Y. Zhang, and B. Wang, Analytical models for protected plus permitted left turn capacity at signalized intersection with heavy traffic, Transportation Research Record: Journal of the Transportation Research Board, No.2192, pp.177-184, 2010. Copyright, National Academy of Sciences, Washington, D.C., 2011 and 2010, respectively.

well. All of these factors contribute to the difficulties for estimating average delay and the total capacity. Furthermore, these issues are not included in the considerations of the current Highway Capacity Manual (HCM) [104] methods.

Therefore, it is necessary to investigate the traffic capacity and delay during heavy traffic at pre-timed intersections with a short left-turn bay. Due to the complex interactions between left-turn and through traffic, the objective of this chapter is to propose heuristic models for practice purposes. The methods described below are not rigorous in essence. They simply serve as a framework for the future study.

3.2 Basic Models for Signals with Protected Left-Turn Phase

3.2.1 Leading Protected Left-Turn Phase

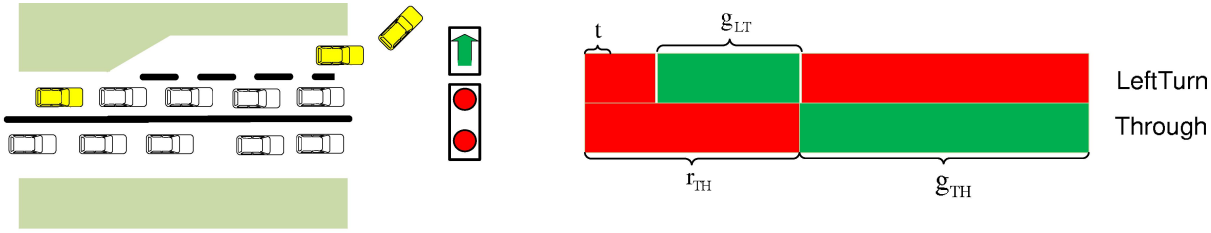

Figure 3.1: Left-turn blockages and signal timing.

Blockage to a short left-turn bay often occurs when traffic signal is operating with a leading left-turn phase during the peak hours. Although the spillback of left-turn vehicles can take place occasionally, it is less severer than the blockage for the traffic signal with leading phase in terms of total capacity and delay at the intersection. Hence, we focus on the blockage to left-turn bay in this section. To model the blockage to a left-turn bay from an adjacent through lane, illustrated in

Figure 3.1, the influence of through residual queues should be taken into account. During the period with high traffic demand, it suffices to consider the queue as a random continuous flow. Denote by $Q_{TH}(t)$ the length of queue for through traffic at time t after the commencement of a red period and denote the distribution of residual queue $Q_{TH}(0)$ by $F_Q^{TH}(x)$, i.e., $F_Q^{TH}(x) = Pr\{Q_{TH(0)} \leq y\}$. The closed form for $F_Q^{TH}(x)$ can be obtained by diffusion approximation discussed in Chapter 2.

We assume the length of the left-turn bay to be N normal vehicles and $N+2$ queued through vehicles to block the left-turn bay. To count the influence by residual queues, the probability of blockage takes place as long as there are $N+2-Q_{TH}(0)$ through vehicles arriving in the adjacent lane and no spillback occurs. If the residual queue length is greater than $N+2$, the blockage will remain at least for the whole red time. Considering that $Q_{TH}(0)$ varies cycle by cycle, the blockage probability can be estimated on average as:

$$P_{block} = \sum_{n=0}^{N+2} Pr\left(\{X_{TH} \geq N+2-n\} \cap \{X_{LT} \leq N+2\}\right) Pr(Q_{TH}(0) = n), \tag{3.1}$$

where X_T and X_{LT} represent the number of vehicles on the adjacent trough lane and in the bay, respectively. In general, the calculation of probability in Equation (3.1) is not trivial. Although Poisson arrival assumption can help easily calculate this probability, and has been used in several left-turn traffic studies [48, 95], it is known that the heavy traffic does not follow Poisson distribution. Without any knowledge of the construction of traffic, perhaps one reasonable approach is to use negative binomial distribution for general cases, by considering the arrivals on the adjacent through lane as failures and those in the left-turn bay as successes during the through red interval. The blockage probability can be obtained accordingly by

cumulating each discrete probability for the possible number of vehicles in the left-turn bay. The errors that may arise in this method are due to the fluctuations in the arrival number and the distribution of vehicles on the multiple through lanes. If, however, the qualitative description is the only concern, the Poisson distribution can be applied as an approximation.

Now we analyze the process of blockage and its effects. When blockage takes place, no new vehicles can join the queue in the left-turn bay. However, the blocked vehicles still wait at the intersection during the red interval. It is more convenient to imagine a hypothetical queue which accumulates according to the left-turning arrivals after blockage, as described as dashed curve around the shaded area in Figure 3.2. We can avoid estimating the beginning time of blockage, which leads to extra tedious calculation. Before the protected green time, the analysis of left-turn queue is the same with the usual analysis. The blocked vehicles that lead to the shaded area would contribute to the residual queue during this cycle. In order to obtain residual queue length, one may want to calculate the probability of the number of vehicles in the bay during blockage. Besides, if necessary, one has to take care of the starting time of left-turn arrivals (denoted as t_L in Figure 3.2), since the new left-turning can enter the bay only after the $N+2$ adjacent through arrivals departure. However, it would be too tedious to do so.

The probability of x left-turning vehicles that arrive in the bay before blockage, denoted by $P(x)$, follows a negative binomial distribution with formula:

$$P(x) = C_1 \sum_n \begin{pmatrix} x+N+1-n \\ N+1-n \end{pmatrix} (1-p_t)^x p_t^{N+2-n} Pr(Q_{TH}(0)=n), \qquad (3.2)$$

where p_t denotes the proportion of through traffic and C_1 is a constant such that

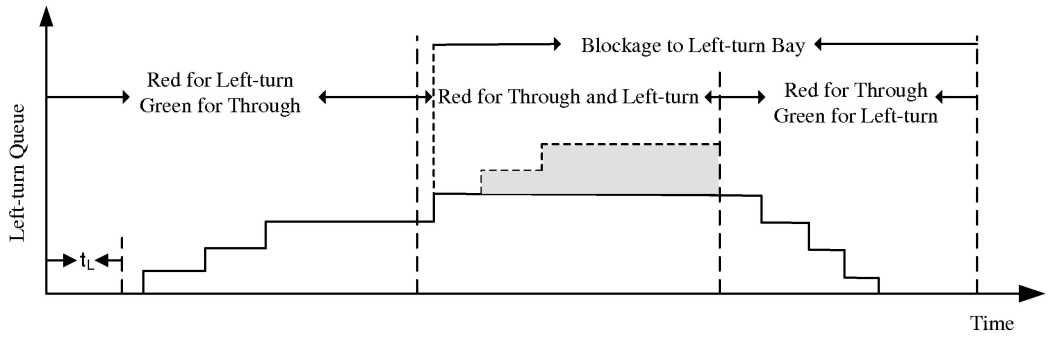

Figure 3.2: Illustration of left-turn queues during left-turn blockage.

$\sum_{x=0}^{N+2} P(x) = 1$. One can see that we adopt a heuristic approach here since the probability of blockage actually relates to the order of arrivals from left-turn and through traffic streams and the intersection state (whether blockage occurs or not) of the previous cycle.

The queues of through traffic contribute to the most complex part of analysis in our proposed approach. As we mentioned in Chapter 2, the expression for residual queues can be complicated, but it receives simplification if diffusion approximation is employed. While diffusion approximation is accurate for heavy demand situation, for light traffic demand, its errors should be of order of the reciprocal for the number of arrivals during one cycle. Here we use Newell's results for $F_Q^{TH}(y)$ and $\mathbb{E}(Q_{TH})$ [76]:

$$F_Q^{TH}(y) = 1 - e^{\mathbb{E}(Q_{TH})^{-1}y}, \tag{3.3}$$

and

$$\mathbb{E}(Q_{TH}) \simeq \frac{Iv_{TH}}{2s_{TH}} \left[\frac{s_{TH}C}{s_{TH}g_{TH} - v_{TH}C} \right], \tag{3.4}$$

where

$v_{TH} =$ average through arrival rate during the entire cycle;

$C =$ signal cycle length;

$I =$ an appropriate variance-to-mean ratio for arrivals and departures;

$s_{TH} =$ saturation flow rate;

$g_{TH} =$ green time for the through traffic.

$r_{TH} =$ red time for the through traffic.

Obviously, with the technique of diffusion approximation, $F_Q^{TH}(y)$ is a continuous function of y. However, one of our interests is the discrete distribution of $\mathbb{E}(Q_{TH})$, the residual queue on the adjacent through lane of a left-turn bay. Since $\mathbb{E}(Q_{TH})$ can be obtained through dividing Q_{TH} by the number of through lanes, one can convert $F_Q^{TH}(y)$ into discrete type as:

$$Pr(Q_{TH} = n) = F_Q^{TH}((n+1)l) - F_Q^{TH}(nl), \tag{3.5}$$

where l is the number of through lanes. Since our interests focus on the length of queues with relative high probability in practice, one can use the truncated distribution of the Equation (3.5).

Here, we see a tension between discrete models and continuous approximation. On one hand, the short left-bay related problems require a discrete model. On the other hand, the expression for distributions of residual queues has a simplified expression by continuous approximations. The approach used in this section is not the best way to achieve the balance between the requirement of models and simplified expressions for related quantities. Moreover, the proposed model for left-turn bay blockage is very crude since it overlooks many important aspects during the interactions between through and left-turn traffic. It still needs more efforts in the

future.

3.2.2 Lagging Protected Left-Turn Phase

One can observe that the effects of left-turn spillback are much severer than that of left-turn bay blockage in terms of capacity and delay for traffic signals with lagging protected left-turn phase. Hence, we concentrate on the left-turn spillback and its probability in this section.

Denote the left-turn residual queue by Q_{LT}. The probability of left-turn spillback from the bay is a joint of two events: at least $N + 3 - Q_{LT}$ left-turn vehicles arrive at the intersection and the adjacent through vehicles are not able to block the bay. Here, we assume the transitional area between left-turn bay and through lane can contain 2 vehicles. Thus, the spillback probability is estimated as:

$$P_{spill} = \sum_{m,n=0}^{N+1} Pr\left(\{X_{TH} \geq N + 3 - m\} \cap \{X_{LT} \leq N + 1 - n\}\right) Pr(Q_{TH}(0) = n) Pr(Q_{LT} = m). \tag{3.6}$$

To estimate the probability in the Equation (3.6), one encounters the same problem for the Equation (3.1). The method suggested in the discussion for the Equation (3.1) can be applied to this equation.

3.3 Problems of Capacity with Protected Left-Turn Phase

3.3.1 Leading Protected Left-Turn Phase

Based on the modeling results, the protected left-turn capacity can then be calculated by combining the expected number of vehicles during blockage and the number of left-turn vehicles in non-blockage condition:

$$C_{protected} = nP_{block}\mathbb{E}(X_{LT}) + (1 - P_{block})\frac{s_{LT}g_{LT}}{C}, \tag{3.7}$$

where n represents the number of cycles in a peak hour, s_{LT} the saturation flow rate for protected left-turn movement, g_{LT} the effective green interval, C the cycle length and $\mathbb{E}(X_{LT})$ the expected number of vehicles in the left-turn bay when blockage occurs. To estimate $\mathbb{E}(X_{LT})$, one needs to consider the distribution of left-turn residual queue $F_Q^{LT}(x)$.

Here one needs to be careful about the definition of capacity. The usual definition of capacity should be independent of any arrival rates. However, Equation (3.7) involves P_{block} which depends on the left-turn and through traffic rates. From this perspective, Equation (3.7) should be considered as a sort of "conditional" capacity. The capacity should be the maximum of value of Equation (3.7) by varying different arrival rates.

3.4 Delay Problems with Protected Left-Turn Phase

3.4.1 Leading Protected Left-Turn Phase

The delay of through traffic contributes to the most part of total traffic delay at a signalized intersection. During peak hours when there is a high demand of through traffic, the leading protected left-turn signal operation would lead to a situation that quickly queued adjacent through vehicles block the left-turn bay. Such phenomenon would occur especially for some cycles during which the fluctuations in the number of arrivals are large. When blockage to the left-turn bay takes place, only the vehicles in the left-turn bay can depart during protected left-turn phase. And the blocked left-turning vehicles are delayed to the next cycle. Obviously, those blocked vehicles increase the total delay at a signalized intersection. In this section, we present the delay estimation for both left-turn and through vehicles according to the results of blockage probability discussed in the previous section.

The total uniform delay for all left-turn vehicles per cycle, denoted by d_{LT}^T, can

be calculated:

$$d_{LT}^{T} = P_{block} d_{block}^{T} + (1 - P_{block}) d_{unblock}^{T}, \tag{3.8}$$

$$d_{block}^{T} = \frac{1}{2} r_{LT}^{2} v_{LT} + r_{LT} v_{LT} g_{LT} + \sum_{x=0}^{N+2} P(x)(\frac{x^2}{2 s_{LT}} - g_{LT} x) + \frac{1}{2} g_{LT}^{2} v_{LT} \tag{3.9}$$

$$d_{unblock}^{T} = \frac{1}{2} \frac{v_{LT} r_{LT}^{2}}{(1 - v_{LT}/s_{LT})}, \tag{3.10}$$

where

$s_{LH} =$ saturation flow rate for protected left-turn movement, in vehicle per second unit,

$x =$ the number of left-turning vehicles arrive in the bay before blockage,

$r_{LT} =$ red signal time for left-turning vehicles,

$g_{LT} =$ protected left-turn green time,

$v_{LT} =$ average left-turning arrival rate (veh/sec).

Equation (3.9) is derived from estimating the shaded area (the imaged queue) and the area between the queue curve and time axis in Figure 3.2 from the previous section. Equation (3.10) is essentially the same with the uniform delay in HCM methods when the traffic condition is not oversaturated. It is apparent that the proposed average uniform delay model, Equation (3.8), is essentially a weighted combination among the blocked left-turning vehicles and others. To obtain the average uniform delay d_{LT} for per vehicle, one can divide d_{LT}^{T} by the average number of arrivals per cycle:

$$d_{LT} = \frac{d_{LT}^{T}}{v_{LT} C}, \tag{3.11}$$

where C is one signal cycle length. This equation can be used to replace the uniform delay term in the HCM methods [104] to estimate the control delay for left-turn traffic.

One may be concerned with the influence of blocked left-turning vehicles to the residual queue and the through traffic delay. In fact, such influence cannot cause any large errors since during heavy traffic the through delay caused by blocked left-turning vehicles should be small compared to total delay. To see this, let us consider the number of blocked vehicles as the fluctuations in the number of through traffic and investigate its influence to delay. Note that during heavy traffic, at most cycles, the processes of departures and arrivals are uncorrelated. By mimicking the arguments in Newell [76], one can show the fraction of the average invoked errors d_{ε}^{T} to the average total through delay d_{TH}^{T} as follows.

$$\frac{d_{\varepsilon}^{T}}{d_{TH}^{T}} = \frac{I}{r_{TH}(s_{TH} - v_{TH})} = O\left(\frac{1}{s_{TH} g_{TH}}\right), \tag{3.12}$$

where r_{TH} is red time for through traffic and $O(\cdot)$ is big-O notation meaning the same order. In the situation of heavy traffic, this term is very small and therefore, we can disregard the error terms contributed by blocked vehicles. Based on the above analysis, the total uniform delay for through traffic per cycle can be estimated as

$$d_{TH}^{T} = \frac{s_{TH} v_{TH} v_{TH}^{2}}{2(s_{TH} - v_{TH})} + C\mathbb{E}(Q_{TH}). \tag{3.13}$$

To obtain the average uniform delay per vehicle per cycle for an entire intersection, we have

$$d = \frac{d_{TH}^{T} + d_{LT}^{T}}{(v_{TH} + v_{LT})C}. \tag{3.14}$$

Again, this equation can be used to replace the uniform delay in the HCM methods

to estimate the control delay.

3.4.2 Lagging Protected Left-Turn Phase

For the lagging protected left-turn signal during the period of high traffic demand, the left-turn vehicles are most likely to spill out of the bay rather than be blocked by through vehicles. Such phenomenon would occur particularly for some cycles when left-turn flow is close to capacity. If the left-turning vehicles spill out of the bay, there would be some vehicles leftover to next cycle and the left-turn spillover queue would block adjacent through traffic. Hence, the through capacity is reduced and total delay increases during through green phase.

When the adjacent through lane is blocked during the time that the spillback of left-turn vehicles, the through vehicles should wait behind the spilled left turns if they cannot seek a chance to move to another unblocked through lane. In this study, we only consider the case that the through volume is less than the reduced capacity and leave the other case to future work. In addition, we assume the blocked through vehicles on the adjacent lane wait until the spillback dissipates. Therefore, the delay for through vehicles in this situation can be estimated from two parts, one accounting the queue departure delay due to the reduced capacity, and the other accounting the waiting time for the blocked through vehicles during the left-turn red time. Since the number of through lanes (denoted by l) reduces to $l-1$ during the condition of spillback, the total delay for through traffic can be calculated as follows:

$$d_{TH}^{T} = P_{spill} d_{spill}^{T} + (1 - P_{spill}) d_{nonspill}^{T}, \tag{3.15}$$

$$d_{spill}^{T} = \frac{s_{TH} v_{TH} r_{TH}^{2} (l-1)/l}{2(s_{TH}(l-1)/l - v_{TH})} + \frac{v_{TH}}{2l} r_{LT}^{2}, \tag{3.16}$$

where $d_{nonspill}$ can be calculated similarly to Equation (3.10). The total uniform

delay d_{spill}^{T} can be also converted to the average delay per vehicle to replace the uniform model in the HCM methods.

3.5 Simulation Results

3.5.1 Results for Capacity Models

A two-lane isolated signalized intersection was set up in CORSIM and generated simulation data for evaluation of the developed probabilistic model. The intersection operates with a leading left-turn strategy and has two 12-ft lanes with one left-turn bay, all passenger cars, no parking and no pedestrians. In this simulation setup, the left-turn signals run with a protected phase following by a permissive phase. The length of the left-turn bay was selected as a variable in the capacity calculation. The main input data for the capacity calculation are as follows:

- Number of through lanes: 2
- Protected Left-turn (LT) green: 13 s
- Through vehicle (TH) green: 50 s
- TH red: 63 s
- Total cycle length: 117 s
- Change interval for each phase: 4 s
- Through vehicle volume: 1,550 veh/h
- LT volume: 388 veh/h
- Mean discharge headway: 1.9 s
- Opposing TH volume: 850 veh/h

- Opposing TH arrival type: 3

For each of the eight length scenarios, i.e., the left-turn bay length from 5 to 12 vehicles, 5 simulation runs of 15 minutes were done by changing the random number seeds in CORSIM. During each run, we managed to obtain the left-turn results from CORSIM by increasing the total demand until the output reached its maximum.

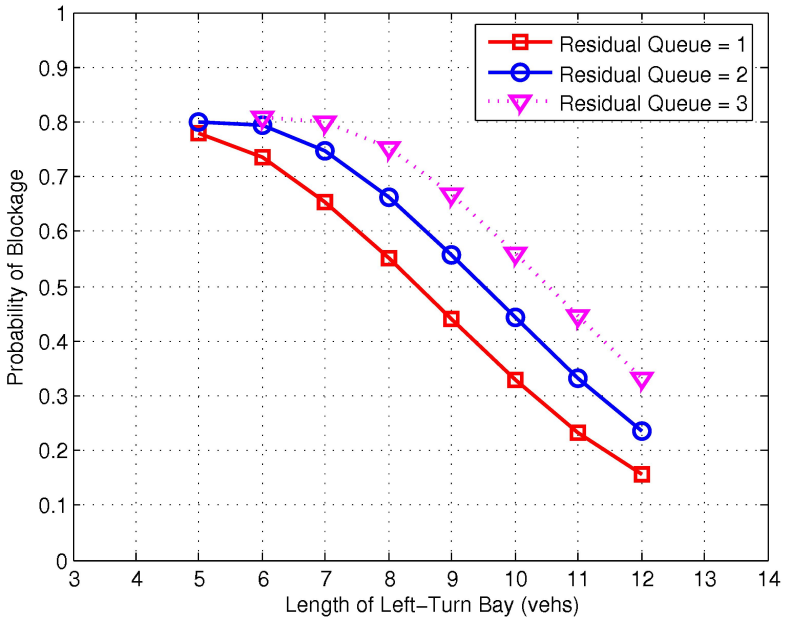

Figure 3.3: Influence of residual queue length on probability of blockage.

The proposed probabilistic model was used for protected left-turn capacity. In regard to protected left-turn capacity, the first is to estimate the length of residual queue. It is appropriate to set up 51 seconds for green time because drivers may use the first second of yellow time to pass the intersection. Obviously the variance-to-mean ratio for arrivals and departure is bounded by 1 because of under dispersed distributed traffic flow rate. This value can be also estimated from the real-world or

simulation data. Suppose the vehicles arrive at intersection with equal distribution for each through lane. Accordingly the expected residual queue length was estimated to be 1 vehicle for each lane and was bounded by 3 vehicles. It is worth noting that the calculation of residual queue length would be sensitive to the arrival rate and the effective green time. In the real-world the residual queue lengths may vary greatly cycle by cycle, especially when the traffic is heavy. Thus knowing the bound of residual queue length is also important for an application. Based on this residual queue estimation, it is easy to calculate protected left-turn capacity after obtaining the average number of left-turn vehicles blocked in the left-turn bay by the adjacent through traffic and the probability of blockage. Figure 3.3 shows the relationship between the length of left-turn bay and the probability of blockage with residual queue lengths of 1, 2 and 3 vehicles in the adjacent through lane.

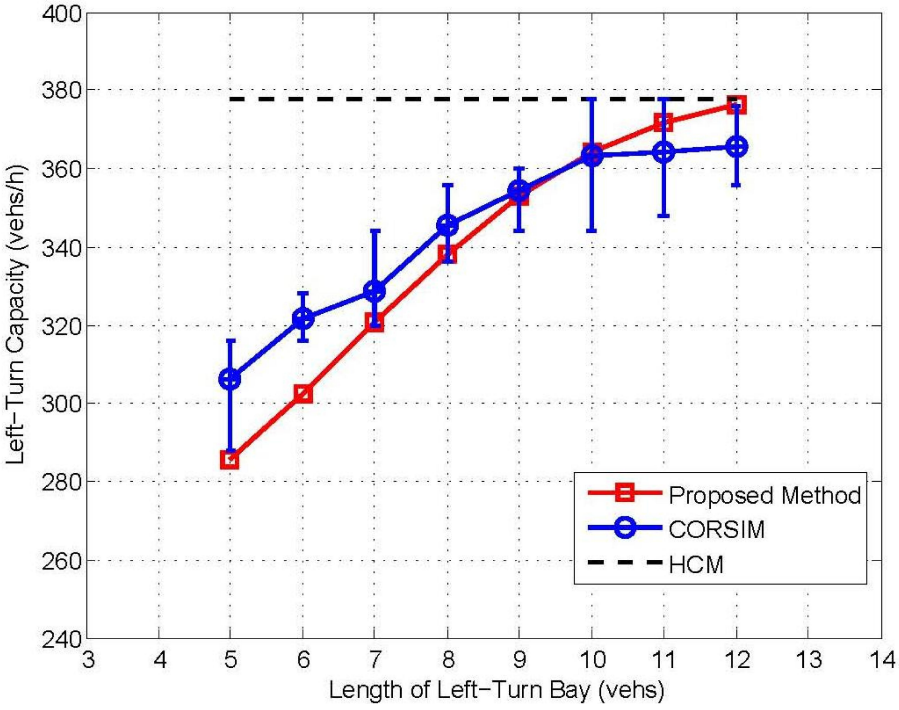

Figure 3.4: Results for the proposed left-turn capacity model.

In regard to the permitted left-turn capacity, the HCM method was used due to the light opposing through vehicles. The critical gap and the follow-up headway used are 4.5 s and 2.5 s, respectively, which have been also used in the current version of HCM. Finally combining the results of protected and permitted left-turn capacity calculation, the total left-turn capacity can be estimated from 285 vehicles when the length of left-turn bay is 5 vehicles to 376 vehicles when the length of left-turn bay is 12 vehicles. Figure 3.4 illustrates the left-turn capacity comparing to the simulated results with respect to the length of the left-turn bay. To illustrate the range of simulation results, the minimum and maximum values for each scenario are plotted as error bars associated with mean value of simulated results. In this figure, the left-turn capacity results that are obtained by the HCM methods only are also shown and remain a constant 378 vph for different scenarios. Obviously, the HCM methods overestimate the left-turn capacity, especially when the length of the left-turn bay is not sufficiently long. Due to the uncertainty of left-turn spillback, the residual queues may not be equally distributed in each through lanes and this phenomenon occurs frequently at short length left-turn bays, i.e., the bays with 5 to 6 vehicles length. The interaction between left-turn and through vehicles is thus complicated at the very short length left-turn bays and the proposed model may under-estimate the left-turn capacity. It is also worth noting that the proposed model overestimates the left-turn capacity when the left-turn bay is long. This is because the residual queue would occur when the left-turn bay gets longer. Considering the stochastic nature of simulation, from the results, the proposed probabilistic model well reflects the left-turn capacity due to the residual queue under heavy traffic.

3.5.2 Results for Delay Models

A two-lane isolated signalized intersection is set up in VISSIM and the simulation data is generated with different scenarios for the evaluation of proposed model. The intersection operates with protected left-turn strategy only (no permitted left-turn phase) and has ideal conditions: two 12-ft through lanes with one left-turn bay, all passenger cars, no parking and no pedestrians. The length of the left-turn bay was selected as a variable in the capacity calculation. Since in VISSIM all passenger cars have the length varying from 13.48 ft to 15.62 ft, the length of the bay was adjusted by observing the number of vehicles during a simulation. The basic calibrated data for the delay calculation follow:

- Protected Left-turn (LT) green: 13 s
- Through vehicle (TH) green: 50 s
- TH Red and change time: 56 s
- Total cycle length: 106 s
- TH saturation flow rate: 1,800 vph
- Protected LT saturation flow rate:1,700 vph

Different through volumes and left-turn volumes are set for the leading and lagging left-turn operations, aiming to enhance the left-turn phenomena for different strategies. In addition, since there is no choice to directly control the saturation flow rate in VISSIM, we managed to change some environment setups to calibrate these values. The saturation flow rate was obtained by averaging the outcome of lane throughput from 15 multiple runs, each of which lasted 100 seconds for discharging queued vehicles under a fully congested situation.

In order to manage the stochastic nature of VISSIM, fifteen simulation runs for each of seven length scenarios of left-turn bay, one hundred and five in total, were conducted for each of the leading and lagging left-turn operations by changing the random number seeds in VISSIM. Each run lasted one hour period with an increment of 15 minutes and the highest 15 min delay was chosen to compute the average control delay for each left-bay scenario. The reason of doing so is that the control delay model in the HCM is developed based on the highest flow level among different 15 min time periods. In the simulation for the leading left-turn operation,

Table 3.1: Standard deviation of left-turn delay in simulation.

Bay length	Std of delay
5	3.83
6	5.04
7	3.46
8	7.45
9	6.09
10	4.58
11	3.98

the through volume is set to 1650 vph and the left-turn volume is set to 100 vph. In this case, the through demand is very high and hence, the residual queue problem becomes a concern, resulting in the high possibility of blockage to the left-turn bay. The authors in the reference [122] observed the blockage in field, especially when the left-turn bay is short under the heavy through traffic. For a longer left-turn bay, the left-turn delay is expected to be smaller since the chance of blockage should become lower. In simulation, the delay for left-turn traffic varies from 69 second per vehicle for a left-turn bay of five vehicles length to 55 second per vehicle for a bay of eleven

vehicles length, as shown in Figure 3.5. It is generally consistent with the expectation as well. However, it is noticed that the left-turn delay for the bay of length eight is a little larger than the ones for the bays of length seven and night. Table 3.1 shows the standard deviation of delay in the sample simulation data. Within the fifteen simulation runs, the standard deviation of left-turn delay for the eight length bay is larger than the others. Nevertheless, it does not mean that the former delay is actually longer. Such results are in fact due to the stochastic nature of arrivals. For this case, a reasonable explanation is that there is no significant difference among the delays for the bay of length seven to nine. Moreover, observations from the field and the simulations show that not only blockage occurs when the bay is very short but also the left-turn spillback can take place. The blocked adjacent through vehicles increase the chance of the longer residual queues and in turn, the blockage to left-turn bay would be most likely to occur. Such complex phenomena cause a larger left-turn delay in the case of shorter bay. As shown in Figure 3.5, the simulation result of a larger delay for the bay of length five than that of length six is due to this reason.

The proposed models for left-turn delay, Equations (3.8) to (3.11), are used to calculate the control delay by replacing the uniform delay term in the HCM delay model (Chapter 16 in the HCM [104]). Regarding the increment term in HCM model, the recommended values for isolated intersections are used to set the parameters [104]. And the progression adjustment factor was set to 1 because of the isolated intersection. No initial queue delay term in the HCM methods was added. As Figure 3.6 shows, the comparison of all results demonstrates that the proposed left-turn delay model well estimates the increase of left-turn delay due to the blockage of through traffic. It is obvious that the HCM methods significantly underestimate the left-turn control delay in this case. However, since the proposed model does not

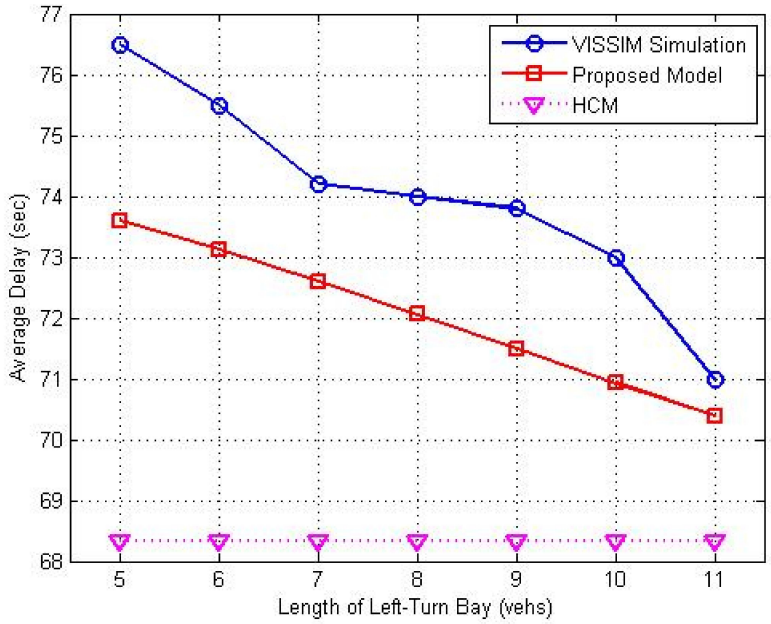

Figure 3.5: Comparison of left-turn delays under leading left-turn operation.

consider the complex phenomena discussed previously for very short left-turn bay, it has a gap between the simulation result and proposed model for the bay of length five. Such inaccuracy will be considered in the future work. Another case with a left-turn volume 160 vph and a through volume 1500 vph was used to validate the proposed left-turn delay model. Note that in this case the v/c ratio for left-turn traffic is as high as 0.76. Again, as shown in Figure 3.6, the results demonstrate the merits of the proposed model.

In the simulation for the lagging left-turn operation, the through volume is set to 1300 vph and the left-turn volume is set to 205 vph under a leading left-turn operation. It is important to note that the left-turn volume is quite close to the saturation flow rate. Therefore, overflows of left-turn traffic are expected for some cycles, resulting in the left-turn spillback from the bay with insufficient length. Such phenomenon was observed from the field as well [122]. The delay for through traffic

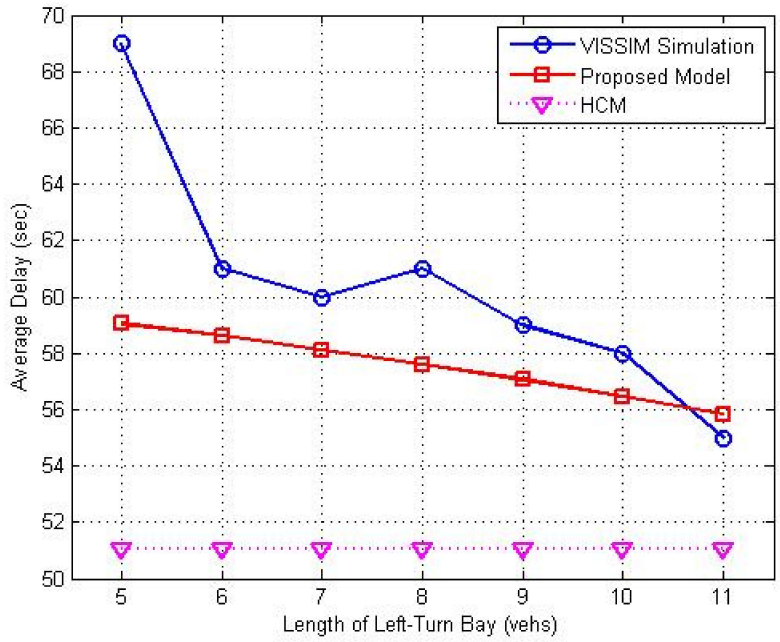

Figure 3.6: Comparison of left-turn delays with left-turn volume 160 vph and through 1500 vph.

varies from 48 second per vehicle for a left-turn bay of length five to 34 second per vehicle for a bay of length eleven, as shown in Figure 3.7. Generally speaking, it is consistent with the intuition as well, saying that the delay decreases with respect to the increase of bay length. However, it is noted that the through delay for a seven length bay is a little larger than the one for a six length bay. This phenomenon occurs similar to the situation of leading left-turn operation: the standard deviation of through delay with the seven length bay is larger than the one with the six length bay within the fifteen simulation runs. It does not mean that the former delay is actually longer but attributes to the samples one could randomly obtain. It is most likely to occur when one deals with the field data as well. For this case, it indicates these two delays are almost same. The proposed models for through delay, Equations (3.15) to (3.16), are used to calculate the control delay by replacing the

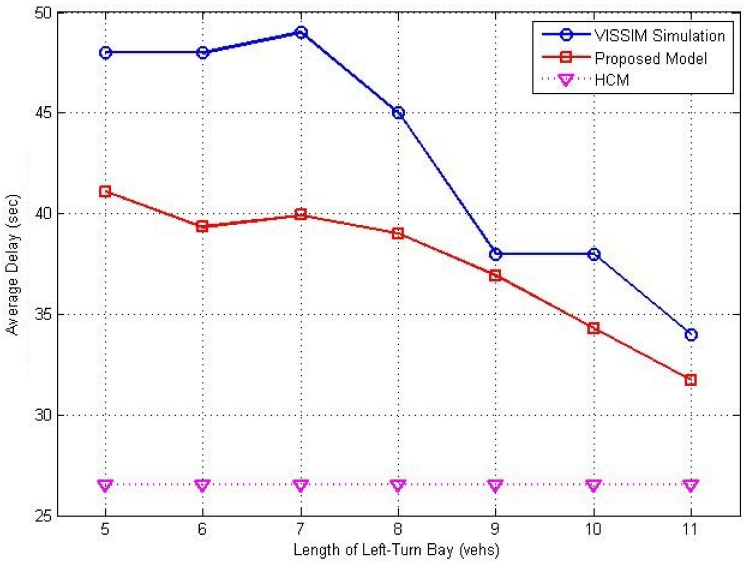

Figure 3.7: Comparison of through delays under lagging left-turn operation.

uniform delay term in the HCM delay model. Figure 3.7 shows the results from the proposed models and the HCM methods. The comparison of all results demonstrates that the proposed through delay model well reflects the increase of delay due to the left-turn spillback. Obviously, the HCM methods significantly underestimate the through control delay in the case of left-turn spillback. However, there is still a gap between VISSIM simulation and the proposed model when the left-turn bay is not very long. It is partially due to the longer queue of left-turn spillback for a shorter bay. Consequently, the drivers on the blocked through lane are much harder to get out of the bottleneck even they get the chance to change to the right lane. The proposed model does not count this issue. It is also noticed that the proposed through delay with the bay of six vehicles length is slightly lower than that with the bay of seven length. The reason is due to the methods of estimating spillback probability. When the left-turn bay is relative short, the independent treatment of

through and left-turn arrivals is not accurate. These issues will be studied in the future.

3.6 Summary

This chapter proposes the models for left-turn bay blockage and left-turn spillback. The models are derived in a quite heuristic manner. They only serve as a starting point for the future research. The difficulties in modeling arise when one attempts to estimate the probability of blockage or spillback. As these phenomena can take place during anytime within the cycle, the exact estimation involves the investigation of the precise structures of arrivals of two traffic streams. The order of arrivals from different traffic streams also plays an important role. Because of the interactions between left-turn and through traffic, the probability of blockage or spillback during one cycle heavily depends on the state of previous cycle. Furthermore, the residual queues for two traffic streams are not independent any more. All of the above factors contribute to the difficulties. It seems that the similar techniques illustrated in this chapter are not suitable for a rigorous study. Appropriate approximations are needed to be developed in the future research.

4. VEHICLE-ACTUATED SIGNAL: DELAY AND QUEUE ANALYSIS

In this chapter, we will study the delay problem at a vehicle-actuated signal intersection under two scenarios, Poisson processes of vehicle arrivals and general stationary arrivals, respectively, both of which trace back to their origins of literature as early as in the 1950s. The first scenario addresses a similar problem as in Darroch *et al.* [32], but with a different modeling methodology and extensive numerical tests. A primary contribution of this section is made in the second scenario with stationary and heavy traffic, which has rarely been studied before. The first scenario sets the framework for the second to build on. In the second scenario, we develop models in general traffic taking green extension as endogenous.

4.1 Problem Statement

Consider a fully actuated signal system at an isolated intersection between two one-way streets without turning vehicles, denoted by major and minor approaches respectively. We assume that there are only two green phases, each dedicated to traffic in one approach. Vehicle arrivals from the two approaches follow two Poisson processes with *i.i.d.* headways in each incoming direction. We consider a special yet popular control scheme that first ensures queue clearance during a green phase. The green extension is controled by a vehicle detector placed upstream of each incoming approach to activate green phase extension. During a green phase, a vehicle passage by the detector at time t triggers the green phase extension until time $t + \Delta$ or until queued vehicles are cleared, whichever is later. Δ is called *unit extension* to allow an arriving vehicle to pass through the intersection without stop. If no vehicle arrival triggers further green extension, the green signal phases out and switches to its conflict approach automatically, resulting in a loss of effective green time. The

unit extensions are denoted by Δ_s and Δ_L for the minor and major approaches, respectively. In each approach, there is a constant discharge rate to clear waiting vehicles. Obviously, the unit extensions Δ_s and Δ_L are the only control variables. The objective is to decide the values of Δ_s and Δ_L such that the average vehicle delay at the intersection over a long period is minimized.

Figure 4.1 illustrates such an intersection and Figure 4.2 shows an example for the green extension in one cycle. In the case of multi-lanes, an equivalent one lane case is obtained by projecting vehicles onto one lane. Therefore, vehicle headway can be any non-negative value, which justifies the point assumption of vehicles. This is in line with the models of many early research such as Darroch *et al.* [33].

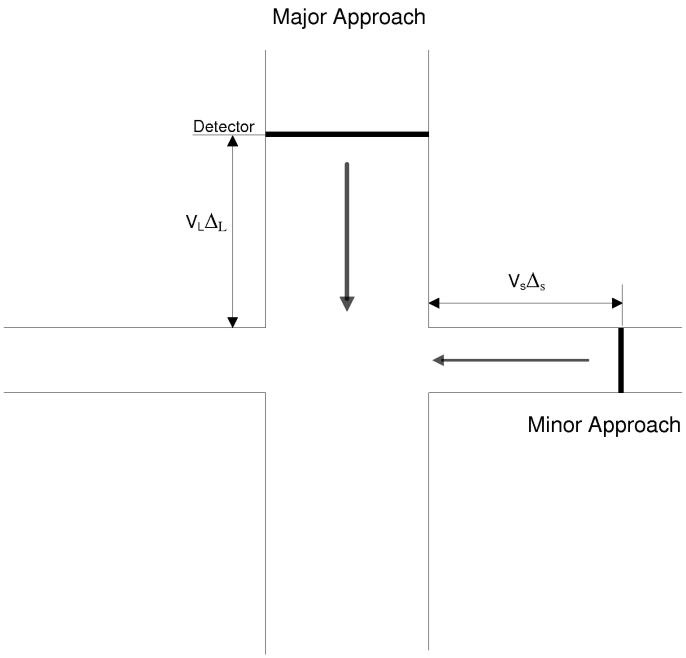

Figure 4.1: A major-minor intersection

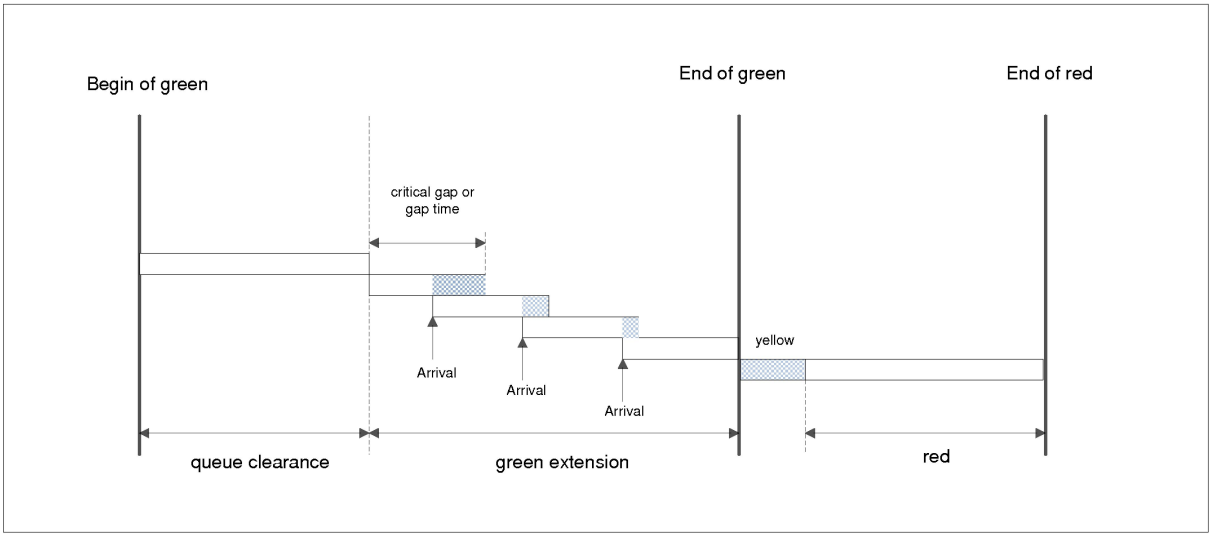

Figure 4.2: A typical process of green extension for one phase (adapted from the Figure 4 in [10]).

In the next sub section, we develop models addressing traffic of Poisson processes. Similar problem setups are also seen in Tanner [103], Darroch *et al.* [33], Newell [84], Cowan [26], and Viti and van Zuylen [108] in which the unit extension is the only control variable.

We assume that loop detectors are at a distance, $v_s\Delta_s$ and $v_L\Delta_L$, before the intersection for both incoming approaches respectively, where v_s and v_L are the average speeds for both roads. We do not consider minimum green time and maximum green time, the same as in Darroch *et al.* [33]. This implies a reasonable assumption that technologies are available to detect the vehicle queue presence. This setup allows us to explore for the full potential of efficient signal control. Note again that this paper does not concern practical implementations, but the maximum potential of intersection control. However, we believe the findings directly shed light on practices.

For simplicity, we assume a *strategy* in which the green phase always switches to its conflict approach when there is no vehicle arrival during the period of the most recently extended unit extension. This assumption is critical to modeling as argued in Darroch *et al.* [33]. In contrary, Tanner [103] considers a strategy not switching signal if no vehicle arrives in the conflict direction, meeting insurmountable technical difficulties.

We refer to our strategy as the *always-switch* strategy. Although the *always-switch* strategy could give rise to a 'peculiar' situation under light traffic in which the green phase switches to the minor approach without any waiting vehicles, and then switches back to the major approach, there is a high probability of vehicle presence at the time of switch in heavy traffic. Therefore, Darroch *et al.* [33] argue that models under this strategy make good approximation to the actual performance.

As will be seen, vehicle delay and green times are functions of the total switching loss of green time in a cycle, irrespective of the loss time split between switches. For simplicity, we denote with δ the total time loss for the two switches during one signal cycle.

We present the notation next. Here, the subscripts s and L correspond to minor and major approaches, respectively.

Notation

λ_s	vehicle arrival rate along the minor approach
λ_L	vehicle arrival rate along the major approach
Δ_s	unit extension in the minor approach
Δ_L	unit extension in the major approach
t_s	random green time in the minor approach
t_L	random green time in the major approach

f_s	discharge rate of vehicular queue along the minor approach
f_L	discharge rate of vehicular queue along the major approach
δ	total green time loss in a signal cycle due to phase switches
$E[\cdot]$	expected value function
$Var(\cdot)$	variance function
$X(\cdot)$	random number of arrivals with the parameter being time period
D_s, D_L	constant discharge headway for the minor and major directions respectively

The parameters of the intersection, such as the green time loss and queue discharge rates, are all deterministic here. We believe that assuming randomness for them would only increase technical complexity slightly and that it would not change the nature of the findings. In addition, the unit extensions, once set up, do not change during a control process.

We first examine the expected duration of green phases at the intersection.

4.2 Expected Green Times

In each approach, the green time consists of two random components: queue clearance time and free flow time. The queue clearance time represents the period in which the discharge rate of vehicles equals the saturation rate, denoted by t_{sa} and t_{La} for both approaches, respectively. The free flow time, denoted by t_{sb} and t_{Lb} for both approaches respectively, corresponds to the period of time, the length of which is a function of the unit extension and vehicle arrivals. In the free flow time, there is no presence of vehicular queue. The unit extension is the endogenous variable to be studied. As a special case, if the unit extension is set to be zero, which means the signal switches immediately after queue clearance, the free flow time becomes zero.

As studied in Wang [111], t_{sb} and t_{Lb} are determined by an information relay process: $E[t_{sb}] = \frac{1}{\lambda_s} e^{\lambda_s \Delta_s} - \frac{1}{\lambda_s} - \Delta_s$, and $E[t_{Lb}] = \frac{1}{\lambda_L} e^{\lambda_L \Delta_L} - \frac{1}{\lambda_L} - \Delta_L$. In addition,

we have $t_s = t_{sa} + t_{sb}$ and $t_L = t_{La} + t_{Lb}$. Due to the Markov property, t_{sa} and t_{sb} are independent of each other. The same is true for t_{La} and t_{Lb}. Figure 4.3 is an illustrative queuing process in the minor direction during a cycle.

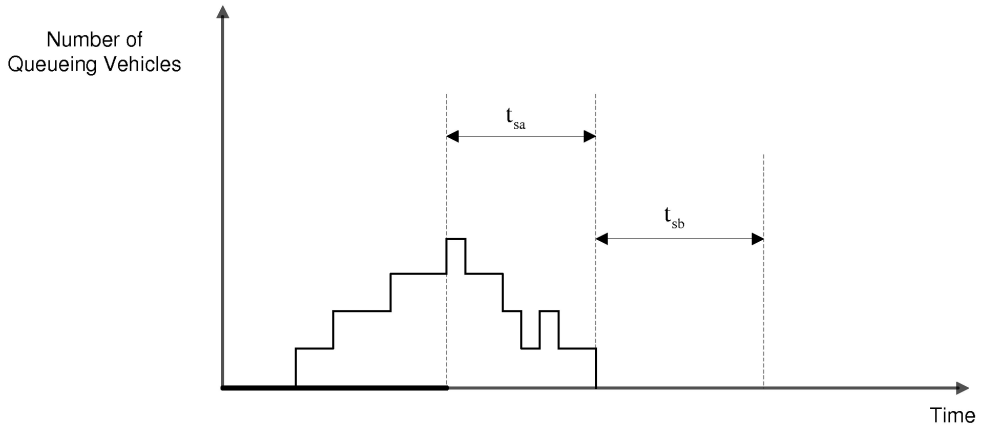

Figure 4.3: An illustrative queuing process in the minor direction

Next, we show how to evaluate t_{sa}. t_{sa} satisfies the following relationship for traffic flow conservation.

$$X(t_L + \delta - \Delta_s) + X(t_{sa}) - f_s t_{sa} = 0, \tag{4.1}$$

where $X(\cdot)$ represents the number of arrivals in a Poisson process along the minor approach. Equation (4.1) states that the total number of vehicles discharged at the saturation rate in time t_{sa} equals to the total arrivals during $t_L + \delta + t_{sa} - \Delta_s$. Here Δ_s is a proven time interval in which no vehicles arrive at the intersection owing to the upstream vehicle detectors. Equation (4.1) ensures a full discharge rate during t_{sa}.

Taking expectation at both sides of Equation (4.1) gives the following expected queue clearance time.

$$E[t_{sa}] = \frac{\lambda_s E[t_L] + \lambda_s \delta - \lambda_s \Delta_s}{f_s - \lambda_s}.$$

Similarly,

$$E[t_{La}] = \frac{\lambda_L E[t_s] + \lambda_L \delta - \lambda_L \Delta_L}{f_L - \lambda_L}.$$

We have

$$\begin{aligned} E[t_s] &= E[t_{sa}] + E[t_{sb}] \\ &= \frac{\lambda_s E[t_L] + \lambda_s \delta}{f_s - \lambda_s} + \frac{1}{\lambda_s} e^{\lambda_s \Delta_s} - \frac{1}{\lambda_s} - \frac{f_s}{f_s - \lambda_s} \Delta_s. \end{aligned} \tag{4.2}$$

In the same way,

$$E[t_L] = \frac{\lambda_L E[t_s] + \lambda_L \delta}{f_L - \lambda_L} + \frac{1}{\lambda_L} e^{\lambda_L \Delta_L} - \frac{1}{\lambda_L} - \frac{f_L}{f_L - \lambda_L} \Delta_L. \tag{4.3}$$

Solving (4.2) and (4.3) gives $E[t_s]$ and $E[t_L]$ as follows.

Proposition 1 *The expected lengths of green phases are given as follows.*

$$\begin{aligned} E[t_s] &= \frac{(f_s-\lambda_s)(f_L-\lambda_L)}{f_sf_L-f_s\lambda_L-f_L\lambda_s} \\ &\times \left\{\frac{\lambda_s\delta}{f_s-\lambda_s}+\frac{1}{\lambda_s}e^{\lambda_s\Delta_s}-\frac{1}{\lambda_s}-\frac{f_s}{f_s-\lambda_s}\Delta_s\right. \\ &+ \left.\frac{\lambda_s}{f_s-\lambda_s}\left(\frac{\lambda_L\delta}{f_L-\lambda_L}+\frac{1}{\lambda_L}e^{\lambda_L\Delta_L}-\frac{1}{\lambda_L}-\frac{f_L}{f_L-\lambda_L}\Delta_L\right)\right\}, \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} E[t_L] &= \frac{(f_s-\lambda_s)(f_L-\lambda_L)}{f_sf_L-f_s\lambda_L-f_L\lambda_s} \\ &\times \left\{\frac{\lambda_L\delta}{f_L-\lambda_L}+\frac{1}{\lambda_L}e^{\lambda_L\Delta_L}-\frac{1}{\lambda_L}-\frac{f_L}{f_L-\lambda_L}\Delta_L\right. \\ &+ \left.\frac{\lambda_L}{f_L-\lambda_L}\left(\frac{\lambda_s\delta}{f_s-\lambda_s}+\frac{1}{\lambda_s}e^{\lambda_s\Delta_s}-\frac{1}{\lambda_s}-\frac{f_s}{f_s-\lambda_s}\Delta_s\right)\right\}. \end{aligned} \tag{4.5}$$

From Proposition 1, it is clear that green duration in each approach is a function of the unit extensions and all other characteristics (such as discharge rate, arrival rate, and green loss) in both approaches. This demonstrates the dynamics between both approaches. When $\frac{\lambda_s}{f_s}+\frac{\lambda_L}{f_L}\to 1$, the expected green times in both approaches become infinite. This is clear by making a change to the denominator of the first term: $f_sf_L-f_s\lambda_L-f_L\lambda_s=f_sf_L(1-\frac{\lambda_L}{f_L}-\frac{\lambda_s}{f_s})$.

In the case of general renewal processes for vehicle arrivals, $E[t_{sa}]$ and $E[t_{La}]$ can be both obtained similarly.

The effect of the green loss δ becomes clearer when we set the unit extensions to zero (*e.g.* a queue control policy). In this case, Equation (4.4) and (4.5) become the following.

$$E[t_s] = \frac{\lambda_s f_L \delta}{f_s f_L - f_s \lambda_L - f_L \lambda_s} = \frac{\lambda_s \delta}{f_s(1 - \frac{\lambda_s}{f_s} - \frac{\lambda_L}{f_L})}, \tag{4.6}$$

and

$$E[t_L] = \frac{\lambda_L f_s \delta}{f_s f_L - f_s \lambda_L - f_L \lambda_s} = \frac{\lambda_L \delta}{f_L(1 - \frac{\lambda_L}{f_L} - \frac{\lambda_s}{f_s})}. \tag{4.7}$$

We can easily have the following result,

$$\frac{E[t_s]}{E[t_L]} = \frac{\lambda_s}{\lambda_L} \times \frac{f_L}{f_s}.$$

Specially, the ratio of green times is proportionate to the ratio of vehicle arrival intensities if the discharge rates are equal in both approaches.

Then in this case, the expected duration of a full signal cycle becomes as follows.

$$\begin{aligned} C &= E[t_s] + E[t_L] + \delta \\ &= \frac{\delta}{1 - \frac{\lambda_L}{f_L} - \frac{\lambda_s}{f_s}}. \end{aligned}$$

Clearly, the above property mimics that from the fixed cycle system with uniform continuum vehicle arrivals.

When the unit extension is set to zero, the control policy simply becomes a queue control (Kruger *et al.* [51]) one ensuring queue clearance in each direction, which implies the following relationship.

$$\frac{Q_s}{Q_L} = \frac{(1 - \frac{\lambda_s}{f_s})\lambda_s}{(1 - \frac{\lambda_L}{f_L})\lambda_L}, \tag{4.8}$$

where Q_s and Q_L are the expected queue lengths at the beginning of the green signal for the minor and major directions respectively.

When $\frac{\lambda_s}{f_s} + \frac{\lambda_L}{f_L} \to 1.0$, Equation (4.8) becomes $\frac{Q_s}{Q_L} \to \frac{f_s}{f_L}$.

When the discharge rates are equal in both directions at saturated intersection, a queue clearance policy essentially ensures $Q_s = Q_L$. The queue clearance policy, i.e., the unit extensions are set to zero, is not optimal as will be seen at later tests where the optimal unit extensions are not zero.

4.3 Variances of Green Times

Variances of green phases will be needed in the calculation of the average vehicle delay.

Variances in both approaches are inter-related. If the duration of the green phase in one approach becomes longer, expectedly the duration of the green phase in its conflict approach becomes longer as well. As a result, if the variance of the green duration in one approach becomes larger, expectedly the variance in its conflict approach becomes larger as well. The derivation in this section exploits just this observation to establish recursive equations. Note that the following derivation implies that there exist stationary variances, which is generally true at reasonable intersections.

In our problem setting, the saturation flow headway of departure at the intersection is assumed constant.

Consider one vehicle at the stop line to be immediately discharged in the minor approach. Discharging this vehicle takes a constant time, D_s, where $D_s = \frac{1}{f_s}$. During this interval there might be several vehicles arriving to join the queue. We denote

the time for discharging this vehicle and the newly arrived ones by χ_1, and denote its probability distribution function by $P(x)$, where $P(x) = P\{\chi_1 \leq x\}$. If there is a queue containing i vehicles at the beginning of the green time, we denote the total time for discharging this queue by χ_i and its related distribution by $P_i(x)$ where $P_i(x) = P\{\chi_i \leq x\}$. Here $\chi_i = \chi_1^1 + \chi_1^2 + ... + \chi_1^i$, each χ_1^k is i.i.d with χ_1. Clearly, $P_i(x)$ is an i-fold convolution of $P(x)$, which is equivalent to completely cleaning i queues of only one vehicle each at the beginning. By the theorem of total probability, the following is obvious.

$$P\{\chi_1 \leq x\} = \sum_{i=0}^{\infty} \frac{(\lambda_s D_s)^i}{i!} e^{-\lambda_s D_s} P\{\chi_i \leq x - D_s\}.$$

If we let $\Gamma(s) = E(e^{-s\chi_1})$, the Laplace-Stieltjes transform of distribution $P(x)$, then from the above analysis we have the following relationship:

$$\begin{aligned} \Gamma(s) &= \sum_{i=0}^{\infty} \frac{(\lambda_s D_s)^i}{i!} e^{-\lambda_s D_s} E(e^{-s(\chi_i + D_s)}) \\ &= \exp\{D_s(\lambda_s \Gamma(s) - \lambda_s - s)\}. \end{aligned} \tag{4.9}$$

The second equation uses the fact that Laplace-Stieltjes transform of $P_i(x)$ is i-th product of $\Gamma(s)$. Using the derivatives of Equation (4.9) at the point where $s = 0$, we can get any moment of χ_1.

Now we can evaluate the random variable t_{sa}. Suppose that the red time t_L for the minor approach and the lost time δ, are known. During the time interval $t_L + \delta - \Delta_s$, the number of arrivals follows Poisson distribution. Therefore,

$$P\{t_{sa} \leq x\} = \sum_{i=0}^{\infty} \frac{(\lambda_s(t_L + \delta - \Delta_s))^i}{i!} e^{-\lambda_s(t_L + \delta - \Delta_s)} P\{\chi_i \leq x\}.$$

Forming the Laplace-Stieltjes transform of $P\{t_{sa} \leq x\}$, denoted by $F(s)$, we have the following equation:

$$\begin{aligned} F(s) &= \int_0^{+\infty} e^{-sx} dP\{t_{sa} \leq x\} \\ &= \exp\{\lambda_s(t_L + \delta - \Delta_s)(\Gamma(s) - 1)\}. \end{aligned} \tag{4.10}$$

From Equation (4.9) it is easy to evaluate the first and second derivatives of $\Gamma(s)$ at $s = 0$:

$$\Gamma'(0) = -\frac{1}{f_s - \lambda_s}, \tag{4.11}$$

$$\Gamma''(0) = \frac{f_s}{(f_s - \lambda_s)^3}. \tag{4.12}$$

Hence, the expected t_{sa} conditional on $t_L + \delta - \Delta_s$ is:

$$\begin{aligned} E[t_{sa}|t_L + \delta - \Delta_s] &= -F'(0) \\ &= \frac{\lambda_s(t_L + \delta - \Delta_s)}{f_s - \lambda_s}. \end{aligned} \tag{4.13}$$

Equation (4.13) agrees with Equation (4.1) in the previous section if we take expectation of it conditional on $t_L + \delta - \Delta_s$. Equation (4.13) implies a necessary condition for derivation: $\delta \geq \Delta_s$. A similar condition is implied later, $\delta \geq \Delta_L$.

The variance of t_{sa} conditional on $t_L + \delta - \Delta_s$ can be estimated as:

$$\begin{aligned} Var(t_{sa}|t_L + \delta - \Delta_s) &= F''(0) - (F'(0))^2 \\ &= \frac{\lambda_s f_s(t_L + \delta - \Delta_s)}{(f_s - \lambda_s)^3}. \end{aligned} \tag{4.14}$$

To obtain unconditional variances of t_{sa} and t_{La}, we need to use the relation:

$$Var(t_{sa}) = E[Var(t_{sa}|t_L + \delta - \Delta_s)] + Var(E[t_{sa}|t_L + \delta - \Delta_s]).$$

By using Equations (4.13) and (4.14) we have:

$$Var(t_{sa}) = \frac{\lambda_s f_s(E[t_L] + \delta - \Delta_s)}{(f_s - \lambda_s)^3} + \frac{\lambda_s^2}{(f_s - \lambda_s)^2} Var(t_L). \tag{4.15}$$

In previous analyses, the green time contains two periods: queue clearance time and free flow time. Because of the Markov property, these two periods are independent of each other. As a result, we have:

$$Var(t_s) = Var(t_{sa}) + Var(t_{sb}). \tag{4.16}$$

According to Wang [111], $Var(t_{sb}) = -\frac{1}{\lambda_s^2} - \frac{2\triangle_s e^{\lambda_s \triangle_s}}{\lambda_s} + \frac{e^{2\lambda_s \triangle_s}}{\lambda_s^2}$. Therefore, we have the equations for variances:

$$\begin{aligned} Var(t_s) &= \frac{\lambda_s f_s(E[t_L] + \delta - \Delta_s)}{(f_s - \lambda_s)^3} + \frac{\lambda_s^2}{(f_s - \lambda_s)^2} Var(t_L) \\ &- \frac{1}{\lambda_s^2} - \frac{2\triangle_s e^{\lambda_s \triangle_s}}{\lambda_s} + \frac{e^{2\lambda_s \triangle_s}}{\lambda_s^2}, \end{aligned} \tag{4.17}$$

and similarly,

$$\begin{aligned} Var(t_L) &= \frac{\lambda_L f_L(E[t_s]+\delta-\Delta_L)}{(f_L-\lambda_L)^3}+\frac{\lambda_L^2}{(f_L-\lambda_L)^2}Var(t_s) \\ &- \frac{1}{\lambda_L^2}-\frac{2\triangle_L e^{\lambda_L\triangle_L}}{\lambda_L}+\frac{e^{2\lambda_L\triangle_L}}{\lambda_L^2}. \end{aligned} \tag{4.18}$$

It is clear now that $Var(t_s)$ and $Var(t_L)$ are linear functions of each other. Solving Equations (4.17) and (4.18) gives values of $Var(t_s)$ and $Var(t_L)$ as follows.

Proposition 2 *The equilibrium solutions for variances of green phases are given as follows.*

$$\begin{aligned} Var(t_s) &= \left(1-\frac{\lambda_s^2\lambda_L^2}{(f_L-\lambda_L)^2(f_s-\lambda_s)^2}\right)^{-1}\left(\frac{\lambda_s f_s(E[t_L]+\delta-\Delta_s)}{(f_s-\lambda_s)^3}+\frac{\lambda_s^2}{(f_s-\lambda_s)^2}\right. \\ &\times \left.\left[\frac{\lambda_L f_L(E[t_s]+\delta-\Delta_L)}{(f_L-\lambda_L)^3}+Var(t_{Lb})\right]+Var(t_{sb})\right), \end{aligned} \tag{4.19}$$

$$\begin{aligned} Var(t_L) &= \left(1-\frac{\lambda_s^2\lambda_L^2}{(f_L-\lambda_L)^2(f_s-\lambda_s)^2}\right)^{-1}\left(\frac{\lambda_L f_L(E[t_s]+\delta-\Delta_L)}{(f_L-\lambda_L)^3}+\frac{\lambda_L^2}{(f_L-\lambda_L)^2}\right. \\ &\times \left.\left[\frac{\lambda_s f_s(E[t_L]+\delta-\Delta_s)}{(f_s-\lambda_s)^3}+Var(t_{sb})\right]+Var(t_{Lb})\right), \end{aligned} \tag{4.20}$$

where $Var(t_{sb})=-\frac{1}{\lambda_s^2}-\frac{2\triangle_s e^{\lambda_s\triangle_s}}{\lambda_s}+\frac{e^{2\lambda_s\triangle_s}}{\lambda_s^2}$ *and* $Var(t_{Lb})=-\frac{1}{\lambda_L^2}-\frac{2\triangle_L e^{\lambda_L\triangle_L}}{\lambda_L}+\frac{e^{2\lambda_L\triangle_L}}{\lambda_L^2}$. $E[t_s]$ *and* $E[t_L]$ *are given by Equation (4.4) and (4.5).*

It is interesting to evaluate the green phases when the unit extension is set to zero. In this case, both $Var(t_{sb})$ and $Var(t_{Lb})$ become zero. Equations (4.19) and (4.20) become:

$$Var(t_s) = C_1\left(\frac{(f_L-\lambda_L)^2\lambda_s f_s(E[t_L]+\delta)}{(f_s-\lambda_s)}+\frac{\lambda_s^2\lambda_L f_L(E[t_s]+\delta)}{(f_L-\lambda_L)}\right), \tag{4.21}$$

and

$$Var(t_L) = C_1 \left(\frac{(f_s - \lambda_s)^2 \lambda_L f_L (E[t_s] + \delta)}{(f_L - \lambda_L)} + \frac{\lambda_L^2 \lambda_s f_s (E[t_L] + \delta)}{(f_s - \lambda_s)} \right), \quad (4.22)$$

where

$$C_1 = \frac{1}{(1 - \frac{\lambda_s}{f_s} - \frac{\lambda_L}{f_L}) \left(f_s^2 f_L^2 (1 - \frac{\lambda_s}{f_s} - \frac{\lambda_L}{f_L}) + 2 f_s \lambda_s f_L \lambda_L \right)}. \quad (4.23)$$

With Equations (4.21) and (4.22), we can examine the impact of saturation rate on variances and intersection stability. Now we take a look at the constant C_1, represented as Equation (4.23) at the right-hand sides of both Equations (4.21) and (4.22).

It is easily seen by Equations (4.6), (4.7), (4.21) and (4.22) that the expected green times and their variances increase at a rate of $O\left(\left[1 - \frac{\lambda_L}{f_L} - \frac{\lambda_s}{f_s}\right]^{-1}\right)$ and $O\left(\left[1 - \frac{\lambda_L}{f_L} - \frac{\lambda_s}{f_s}\right]^{-2}\right)$, respectively. These equations imply the following result which indicates the critical role of green time loss in the performance of an intersection.

Proposition 3 *Given an actuated intersection with zero unit extension, both the expected values and variances of the green times increase in a linear manner with the total green time loss δ.*

As one can easily verify that with $\Delta_s = \Delta_L = 0$, symmetric intersections have $E[t_s] = E[t_L] = \frac{\lambda\delta}{f - 2\lambda}$, $Var(t_s) = Var(t_L) = \frac{\lambda\delta}{(f-2\lambda)^2}$, and $\frac{\sqrt{Var(t_s)}}{E[t_s]} = \frac{\sqrt{Var(t_L)}}{E[t_L]} = \frac{1}{\sqrt{\lambda\delta}}$.

4.4 Vehicle Delay

4.4.1 Vehicle Delay During A Signal Cycle

We start with the minor approach, assuming that t_L and δ are given. The waiting time has two components. When the signal first turns green in the minor approach,

there has been a queue with an expected waiting time W_s^0 before the start of green time. The other component of the waiting time is incurred during the period in which the queue is being discharged. We denote this waiting time by W_{sa}. Be aware that there is no new queue during the free flow period t_{sb} and t_{Lb}, as explained before. Accordingly, W_L^0 and W_{La} are the corresponding notations for the major approach.

Proposition 4 *At the time when the green light first turns on in the minor approach, the total vehicle delay, denoted by W_s^0, of those arrivals N_s, during the red signal period is given as follows.*

$$E[W_s^0] = \frac{1}{2}\lambda_s \left(Var(t_L) + E^2[t_L] + 2(\delta - \Delta_s)E[t_L] + (\delta - \Delta_s)^2 \right), \tag{4.24}$$

and similarly for the major approach,

$$E[W_L^0] = \frac{1}{2}\lambda_L \left(Var(t_s) + E^2[t_s] + 2(\delta - \Delta_L)E[t_s] + (\delta - \Delta_L)^2 \right). \tag{4.25}$$

Proof.

With a given number of arrivals from a Poisson process during a time period, the distribution of each arrival is uniform within the given time interval. Therefore, the expected waiting time of the queue can be calculated by conditioning on the length of queue as follows.

$$
\begin{aligned}
E[W_s^0] &= E[E[E[E[W_s^0 \mid X(t_L+\delta-\Delta_s)]] \mid t_L]] \\
&= E[(t_L+\delta-\Delta_s)\times\lambda_s\times\frac{1}{2}\times(t_L+\delta-\Delta_s)] \\
&= \frac{1}{2}\lambda_s E[(t_L+\delta-\Delta_s)^2] \\
&= \frac{1}{2}\lambda_s\left(E[t_L^2]+2(\delta-\Delta_s)E[t_L]+(\delta-\Delta_s)^2\right) \\
&= \frac{1}{2}\lambda_s\left(Var(t_L)+E^2[t_L]+2(\delta-\Delta_s)E[t_L]+(\delta-\Delta_s)^2\right).
\end{aligned}
$$

Similarly, we can show the result for $E[W_L^0]$.

We further have the following results for W_{sa} and W_{La}, respectively.

Proposition 5 *The waiting times during queue discharge are given as follows.*

$$
E[W_{sa}] = \frac{1}{2}(f_s-\lambda_s)\left(Var(t_{sa})+E^2[t_{sa}]\right), \tag{4.26}
$$

and similarly,

$$
E[W_{La}] = \frac{1}{2}(f_L-\lambda_L)\left(Var(t_{La})+E^2[t_{La}]\right). \tag{4.27}
$$

Proof.

We only prove for the minor approach. The result for the major approach can be obtained similarly.

We condition on queue discharge time t_{sa}.

$$
\begin{aligned}
E[W_{sa}|t_{sa}] &= E\left[\int_0^{t_{sa}} [X(t_L+\delta-\Delta_s)+X(t)-f_s t]dt \mid t_{sa}\right] \\
&= E\left[\int_0^{t_{sa}} X(t_L+\delta-\Delta_s)dt \mid t_{sa}\right] + E\left[\int_0^{t_{sa}} (\lambda_s t - f_s t)dt \mid t_{sa}\right] \\
&= (f_s-\lambda_s-\Delta_s)t_{sa}^2 + \frac{1}{2}(\lambda_s-f_s)t_{sa}^2 \\
&= \frac{1}{2}(f_s-\lambda_s)t_{sa}^2.
\end{aligned}
$$

The third equality above uses the fact that $E[X(t_L+\delta-\Delta_s)|t_{sa}] = (f_s-\lambda_s)t_{sa}$ according to Equation (4.1).

We therefore have,

$$
\begin{aligned}
E[W_{sa}] &= E\left[E[W_{sa}|t_{sa}]\right] \\
&= \frac{1}{2}(f_s-\lambda_s)\left(Var(t_{sa})+E^2[t_{sa}]\right).
\end{aligned}
$$

To summarize, the total expected vehicle delay during the period of a signal cycle at the intersection can be expressed in a closed form by $E[W_s^0]+E[W_L^0]+E[W_{sa}]+E[W_{La}]$ explicitly. We can calculate the expected waiting time by substituting the equations for the expected values and variances.

4.4.2 Vehicle Delay Per Unit Time

The signal cycling can be considered as a renewal process, from the beginning of a green time to the beginning of the green time again. According to the renewal theory, the average vehicle delay per unit time over this renewal process is the ratio between the total expected vehicle delay in a cycle and the expected cycle length. This average vehicle delay can therefore be expressed in terms of Δ_s and Δ_L with

given parameters λ_s, λ_L, f_s and f_L. We denote by $F(\Delta_L, \Delta_s)$ the function of average vehicle delay per unit time. The function $F(\Delta_L, \Delta_s)$ can be expressed in the following way.

$$F(\Delta_L, \Delta_s) = \frac{E[W_s^0] + E[W_L^0] + E[W_{sa}] + E[W_{La}]}{E[t_s] + E[t_L] + \delta}. \tag{4.28}$$

Similarly, one can easily get directional average waiting time per vehicle or per time unit, which we do not pursue in details here.

The closed form (4.28) of $F(\Delta_L, \Delta_s)$ consists of a large number of terms, making it very inconvenient to analytically examine its properties such as concavity or convexity. However, our observation is that the delay function appears to be a coercive function in the unit extension. This means that as the unit extension increases, the delay increases and tends to infinity. We can easily study this function numerically with latest computing technologies. Figure 4.4 graphically demonstrates an example delay function when $\delta = 6.0$, $\lambda_s = 0.20$, $\lambda_L = 0.25$, $f_s = 0.6$ and $f_L = 0.6$.

In addition, we examined a case as illustrated in Figure 4.5, in which the discharge rates in both approaches are 0.6 and 0.4 vehicle per second, respectively. The arrival rates are 0.05 and 0.25 vehicles per second, respectively. The green loss during a cycle is 6 seconds. The average vehicle delay reaches its minimum when $\Delta_s = 5.2$ and $\Delta_L = 5.2$ seconds. Intuitively, the close proximity of the unit extensions is surprising as one would intuitively think the unit extension for the minor direction should be much smaller than that for the major direction. Now we are able to evaluate intersection performance in various settings. Table 4.1 is a sample test results where $f_s = f_L = 0.6$ vehicle per second and Table 4.2 is with $f_s = 0.4$ and $f_L = 0.6$ vehicle per second.

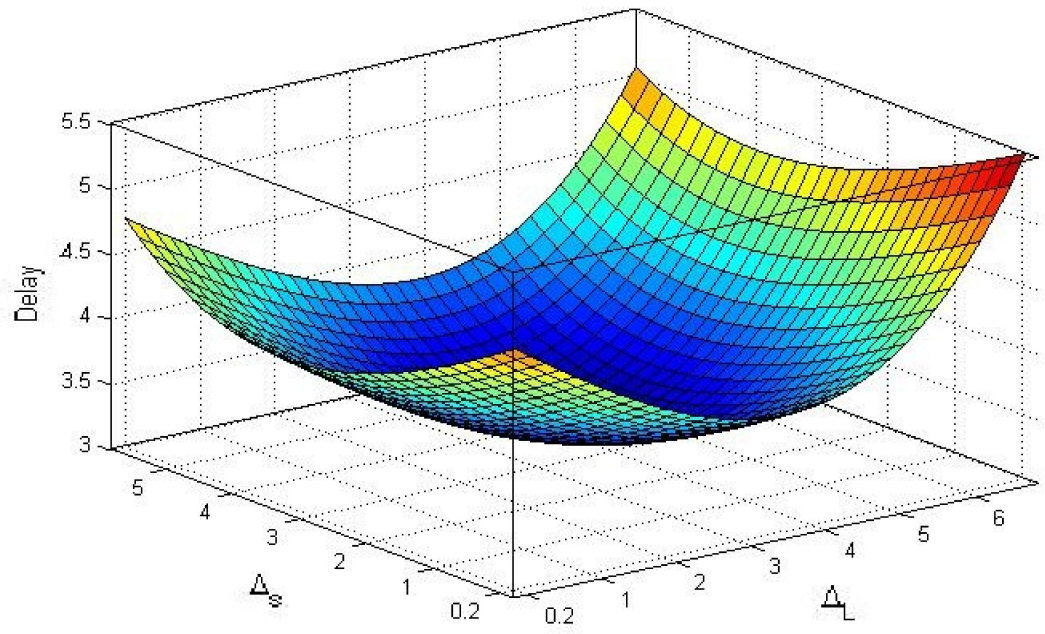

Figure 4.4: Example average delay per unit time with $\delta = 6.0$, $\lambda_s = 0.20$ and $\lambda_L = 0.25$

From Table 4.1, we have the following observations.

Observation 1 *When roadway capacities are comparable between the two approaches, the optimal green extensions in both directions are almost equivalent regardless of traffic intensities, although the major direction has a slightly larger extension than the minor direction.*

4.5 The Case of General Traffic

Our primary concern is intersection performance with general stationary arriving traffic under an actuated signal system. In this case, we assume that vehicle arrivals follow a general renewal process. In this renewal process, vehicle headways are independent and identically distributed (*i.i.d.*) random variables whose density function

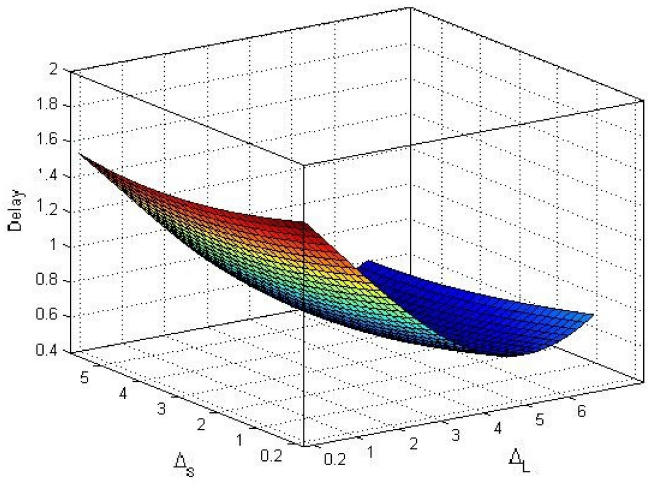

Figure 4.5: Example average delay per unit time $\delta = 6.0$, $\lambda_s = 0.05$ and $\lambda_L = 0.25$

Table 4.1: Optimal unit extension and intersection performance with $f_s = f_L = 0.6, \delta = 4.0$.

λ_s/λ_L	Δ_s/Δ_L	$Var(t_s)$	$Var(t_L)$	$F(\cdot,\cdot)$
0.05/0.25	3.8/4.4	2.1	24.9	0.234
0.08/0.25	3.6/4.2	3.7	24.2	0.441
0.10/0.25	3.6/4.0	5.3	23.6	0.603
0.15/0.25	3.4/3.6	11.3	26.7	1.142
0.20/0.25	3.2/3.4	25.9	38.9	2.029

is denoted by $f(\cdot)$. The arrival intensity is λ, and we use H for the stochastic arrival headway. Clearly, we have $\lambda E[H] = 1$. We use σ^2 for the variance of H, and the subscripts L and s denote the major and minor directions, respectively. In the following, notations without subscriptions indicate results applicable to both directions. We will develop analytical results following the framework earlier.

First, we characterize the free flow green time with general vehicle headway. The

Table 4.2: Optimal unit extension and intersection performance with $f_s = 0.4, f_L = 0.6, \delta = 6.0$.

λ_s/λ_L	Δ_s/Δ_L	$Var(t_s)$	$Var(t_L)$	$F(\cdot,\cdot)$
0.05/0.25	5.2/5.2	8.1	60.4	0.607
0.08/0.25	4.8/4.6	15.9	51.5	1.105
0.10/0.25	4.6/4.4	25.1	56.2	1.543
0.15/0.25	4.4/3.6	88.8	95.3	3.481
0.20/0.25	4.2/2.8	751.0	492.7	11.095

notation has the same meaning as in the previous sections. Derivation of the results is seen in Wang *et al.* (2010).

$$E[t_{sb}] = \frac{\int_0^{\Delta_s} t f(t)dt}{1 - F(\Delta_s)}, \tag{4.29}$$

$$E[t_{Lb}] = \frac{\int_0^{\Delta_L} t f(t)dt}{1 - F(\Delta_L)}, \tag{4.30}$$

$$Var(t_{sb}) = \frac{\int_0^{\Delta_s} t^2 f(t)dt}{1 - F(\Delta_s)} + E^2[t_{sb}], \tag{4.31}$$

$$Var(t_{Lb}) = \frac{\int_0^{\Delta_L} t^2 f(t)dt}{1 - F(\Delta_L)} + E^2[t_{Lb}]. \tag{4.32}$$

4.5.1 Expectations

Following similar steps as in the Poisson process using the results (4.29) to (4.32), we have the following accordingly.

Proposition 6 *For general traffic, the expected lengths of green phases are given as*

follows.

$$\begin{aligned} E[t_s] &= \frac{(f_s-\lambda_s)(f_L-\lambda_L)}{f_sf_L-f_s\lambda_L-f_L\lambda_s} \\ &\times \left\{ \frac{\lambda_s\delta}{f_s-\lambda_s} + \frac{\int_0^{\Delta_s} tf(t)dt}{1-F(\Delta_s)} - \frac{\lambda_s}{f_s-\lambda_s}\Delta_s \right. \\ &+ \left. \frac{\lambda_s}{f_s-\lambda_s}\left(\frac{\lambda_L\delta}{f_L-\lambda_L} + \frac{\int_0^{\Delta_L} tf(t)dt}{1-F(\Delta_L)} - \frac{\lambda_L}{\lambda_L-\lambda_L}\Delta_L \right) \right\}, \end{aligned} \tag{4.33}$$

$$\begin{aligned} E[t_L] &= \frac{(f_s-\lambda_s)(f_L-\lambda_L)}{f_sf_L-f_s\lambda_L-f_L\lambda_s} \\ &\times \left\{ \frac{\lambda_L\delta}{f_L-\lambda_L} + \frac{\int_0^{\Delta_L} tf(t)dt}{1-F(\Delta_L)} - \frac{\lambda_L}{f_L-\lambda_L}\Delta_L \right. \\ &+ \left. \frac{\lambda_L}{f_L-\lambda_L}\left(\frac{\lambda_s\delta}{f_s-\lambda_s} + \frac{\int_0^{\Delta_s} tf(t)dt}{1-F(\Delta_s)} - \frac{\lambda_s}{f_s-\lambda_s}\Delta_s \right) \right\}. \end{aligned} \tag{4.34}$$

Discretization of the integral easily gives rise to its numerical solution for any known probability density function $f(\cdot)$ above.

4.5.2 Some Prerequisite Results

The following results are needed for the subsequent derivations of variances and waiting time.

Theorem 2 *Given a constant t, the following holds for both directions under heavy traffic:*

$$E[X(t)] \approx \lambda t, \tag{4.35}$$

$$Var(X(t)) \approx kt, \tag{4.36}$$

$$where\ k = \frac{\sigma^2}{E^3[H]}.$$

In the special case of Poisson arrivals (e.g., headway with exponential distribution), we can easily verify that $k = \lambda$, in which case values in Theorem 2 are accurate. In general cases of heavy traffic where vehicle headway is small compared to the green time in each direction, Theorem 2 provides good approximations. One way of its proof is indicated in Grimmett and Stirzaker (2001), which states that for a renewal process, the mean and variance of number of arrivals follow the Central Limit Theorem during a large time period.

The following result about waiting time corresponds to Proposition 4.

Proposition 7 *Under heavy traffic, the expected vehicle waiting time $w(t)$ is approximately $\frac{\lambda t^2}{2}$, i.e., $w(t) \approx \frac{\lambda t^2}{2}$, where t is the time that the signal has been red in the direction of interest, assuming no queue present at the beginning of the red time.*

Proof.

We have

$$w(t) = \int_0^t (t - l + w(t - l))\, f(l) dl.$$

The above equation conditions on the first vehicle arrival at time l from the beginning of the red time. We only need to see if the assumption can approximately satisfy the above equation.

$$
\begin{aligned}
\text{Left handside} &= \frac{1}{2}\lambda t^2. \\
\text{Right handside} &= t - E[H] + \int_0^t \frac{1}{2}\lambda (t-l)^2 f(l) dl \\
&= t - E[H] + \frac{1}{2}\lambda \left(t^2 - 2tE[H] + \sigma^2 + E^2[H]\right) \\
&\approx \frac{1}{2}\lambda t^2 + \frac{1}{2}\lambda\sigma^2 - \frac{1}{2}E[H].
\end{aligned}
$$

Note that $\int_0^t (t-l)f(l)dl \approx t - E[H]$ when there are a few cars during the red time t each time, implying a big enough t relative to the average headway. We call the remnant in the right hand side $\frac{1}{2}\lambda\sigma^2 - \frac{1}{2}E[H]$ error term. The assumption can make both sides of the equation approximately equal, considering t outweighs $E[H]$ and t^2 outweighs σ^2. In the following, we show how t^2 outweighs σ^2.

$$
\begin{aligned}
\frac{1}{2}\lambda\sigma^2 &= \frac{1}{2}\lambda v E[H]\sigma \\
&= \frac{1}{2}v\sigma \\
&= \frac{1}{2}v^2 E[H],
\end{aligned}
$$

where v is the coefficient of variation (CV) between the standard deviation σ and the expected headway $E[H]$. If CV is a finite value, then the error compared with the major term $\frac{1}{2}\lambda t^2$ tends to zero in heavy traffic. In Poisson arrivals as a special case, the error term is zero, which can be verified by interested readers.

4.5.3 Variances

In previous analysis, we use χ_1 to denote the time needed for reducing the vehicle queue length by one vehicle. Because the vehicles having arrived during the red time

are $X(t_L+\delta)$ for the minor approach, the total time for discharging these vehicles is equal to $\sum_{i=1}^{X(t_L+\delta)} \chi_1^i$, where χ_1^i is i.i.d. with χ_1. Therefore, using $Var(t_{sa}) = E[Var(t_{sa}|X(t_L+\delta-\Delta_s))] + Var(E[(t_{sa}|X(t_L+\delta-\Delta_s)])$, we have:

$$\begin{aligned} Var(t_{sa}) &= E\Big[\sum_{i=1}^{X(t_L+\delta-\Delta_s)} Var(\chi_1^i)|X(t_L+\delta-\Delta_s)\Big] + Var\Big(\sum_{i=1}^{X(t_L+\delta-\Delta_s)} E[\chi_1^i]|X(t_L+\delta-\Delta_s)\Big) \\ &= Var(\chi_1)E[X(t_L+\delta-\Delta_s)] + (E[\chi_1])^2 Var(X(t_L+\delta-\Delta_s)) \\ &= \lambda_s Var(\chi_1)(E[t_L]+\delta-\Delta_s) + (E[\chi_1])^2\left(k(E[t_L]+\delta-\Delta_s) + \lambda_s^2 Var(t_L)\right). \end{aligned}$$

Similarly, we can evaluate $Var(t_{La})$ as follows

$$Var(t_L) = \lambda_L Var(\mu_1)(E[t_s]+\delta-\Delta_L) + (E[\mu_1])^2\left(k(E[t_s]+\delta-\Delta_L) + \lambda_L^2 Var(t_s)\right).$$

where μ_1 is the counterpart definition of χ_1 for the major road, which is the time for reducing the queue length by one vehicle. When the arrivals follow Poisson distribution, the equation for $Var(t_{sa})$ becomes Equation (4.14). Recall that $Var(t_s) = Var(t_{sa}) + Var(t_{sb})$. The following results become obvious.

Proposition 8 *The variances of green phases are given as follows for the minor and major roads respectively.*

$$\begin{aligned} Var(t_s) &= \lambda_s Var(\chi_1)(E[t_L]+\delta-\Delta_s) + (E[\chi_1])^2\left(k(E[t_L]+\delta-\Delta_s) + \lambda_s^2 Var(t_L)\right) \\ &+ Var(t_{sb}), \qquad (4.37) \\ Var(t_L) &= \lambda_L Var(\mu_1)(E[t_s]+\delta-\Delta_L) + (E[\mu_1])^2\left(k(E[t_s]+\delta-\Delta_L) + \lambda_L^2 Var(t_s)\right) \\ &+ Var(t_{Lb}), \qquad (4.38) \end{aligned}$$

where $E[\chi_1]$ and $Var(\chi_1)$ are given by the following equations as from the earlier results.

$$\begin{aligned} E[\chi_1] &= \frac{1}{(f_s-\lambda_s)^2}, \\ Var[\chi_1] &= \frac{f_s}{(f_s-\lambda_s)^3} - \frac{1}{(f_s-\lambda_s)^2}, \\ E[\mu_1] &= \frac{1}{(f_L-\lambda_L)^2}, \\ Var[\mu_1] &= \frac{f_L}{(f_L-\lambda_L)^3} - \frac{1}{(f_L-\lambda_L)^2}. \end{aligned}$$

The last four formulas in Proposition 8 are borrowed from those on Poisson assumptions. Note that Proposition 4 about waiting time provides asymptotic approximations in heavy traffic according to Proposition 7 and Theorem 2. Proposition 5 also holds in heavy traffic due to Theorem 2. For the according values in general traffic, we only need to update the variances and expected values of green times in Proposition 4 and 5 with their according values from Proposition 6 and 8. Using equation (4.28), we can estimate the average vehicle delay under heavy traffic.

Note that the results here about green time variance, waiting time and average vehicle waiting time are all asymptotically accurate under heavy traffic. The numerical tests next show that the formulas perform well in evaluating the intersection performance.

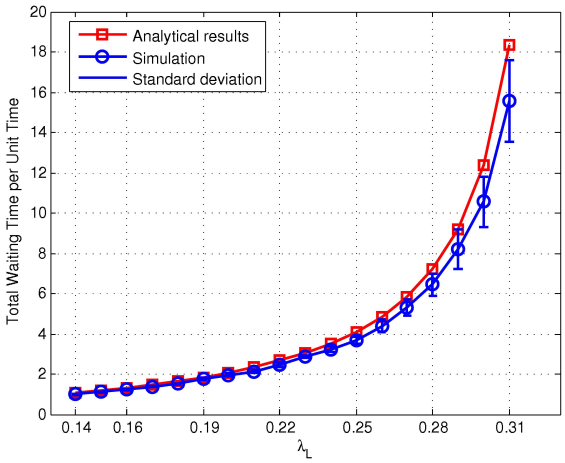

Figure 4.6: $f_s = 0.6, \lambda_s = 0.2, f_L = 0.5, CV = 0.5774$

4.5.4 Numerical Results in Heavy Traffic

In generating arriving traffic for the numerical tests, we assume the headway follows Gamma distribution.

$$f(x) = \frac{exp(-x/\theta)}{\Gamma(k)\theta^k} x^{k-1},$$

where the mean is $\theta k = \lambda_s$ (for minor direction) or λ_L (for major direction). The parameter k is given by the coefficient of variation (CV) in the following tests. The same CV value applies to both directions each time in the following tests. For each case, we simulate 120 times Monte Carlo simulation. In each time, we guarantee no less than 4000 arriving vehicles in each direction, which turns out to be at least 120 signal cycles. All other parameters in the simulation are set as in the analytical formulas such as discharge rates.

In Figure 4.6 and 4.7, $\lambda_s = 0.2, f_s = 0.6, f_L = 0.5, \delta = 6.0, \Delta_s = 3.5, \Delta_L = 3.2$. We vary λ_L from 0.14 to 0.31, i.e., $\frac{\lambda_L}{f_L} + \frac{\lambda_s}{f_s} \in [0.6133, 0.9533]$. The only difference

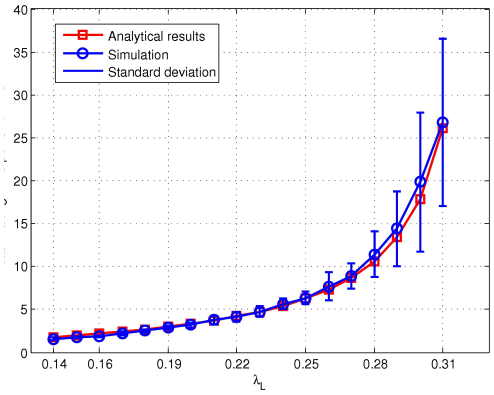

Figure 4.7: $f_s = 0.6, \lambda_s = 0.2, f_L = 0.5, CV = 1.4142$

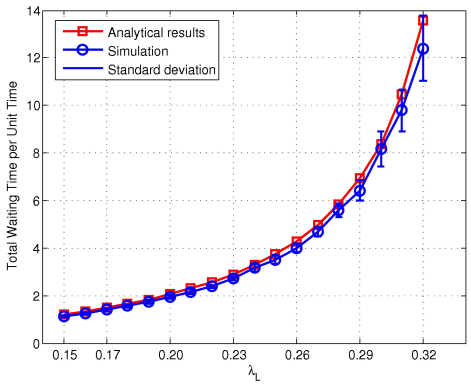

Figure 4.8: $f_s = 0.5, \lambda_s = 0.2, f_L = 0.6, \Delta_s = 3.5, \Delta_L = 3.5, CV = 0.5774$

between Figure 4.6 and 4.7 is the value of CV. In Figure 4.6, the CV is 0.5774 while in Figure 4.7, the CV is 1.4142. In these two figures, the blue error bar represents the standard deviation of simulated results.

We change the setting for f_s and f_L in Figure 4.8 and Figure 4.9, where $\lambda_s = 0.20, f_s = 0.5, f_L = 0.6, \delta = 6.0, \Delta_s = 3.5, \Delta_L = 3.5$. We vary λ_L from 0.15 to 0.32, which yields $\frac{\lambda_L}{f_L} + \frac{\lambda_s}{f_s} \in [0.6500, 0.9333]$. In Figure 4.8, the CV is 0.5744. In Figure 4.9, the CV is 1.4142. When the traffic intensity becomes heavy in the latter case, the green times changed cycle-by-cycle sometimes significantly. For example, when

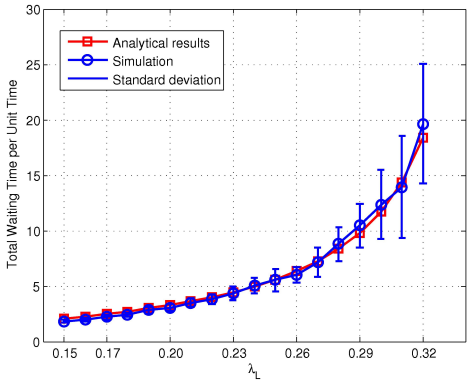

Figure 4.9: $f_s = 0.5, \lambda_s = 0.2, f_L = 0.6, \Delta_s = 3.5, \Delta_L = 3.5, CV = 1.4142$

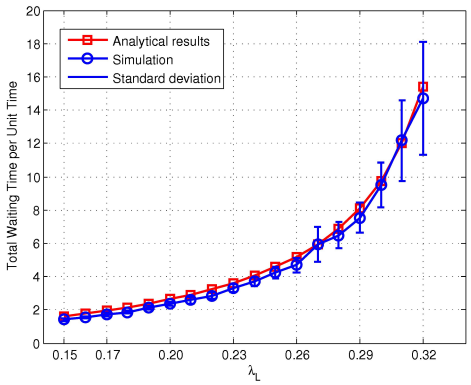

Figure 4.10: $f_s = 0.5, \lambda_s = 0.2, f_L = 0.6, \Delta_s = 3.5, \Delta_L = 3.5$, Poisson headway

λ_L increases to 0.31, the average green time is 30 and the standard deviation becomes 32.

The analytical results for heavy traffic are exact in the case of Poisson distributed headway. Figure 4.10 shows analytical results with Poisson arrivals compared with those from simulation, other parameters identical to those in Figure 4.8 and Figure 4.9. The slight disagreement between the analytical and simulation due to variations in simulation explains the quality of the performances in Figure 4.6 through 4.9. Overall, the analytical estimate of waiting time appears better compared with the

simulation when the CV of the arrival headway is larger. By information, we have examined the arriving headway using loop detector information at an intersection in Minneapolis. We found the CV ranged from 1.2 to 1.9 with time of day. It appears that our models for general traffic should work well practically.

Note that the figures also indicate very large variances of green times under very heavy traffic.

4.6 Summary

This chapter studies a vehicle-actuated signal control scheme in which the unit extension is the only control variable. We analytically model the intersection performance in two cases respectively: Poisson arrivals and general traffic. Equations for expected green times in both directions, variances, and the average waiting time are presented. The models in the case of Poisson processes are exact. The models in the case of heavy traffic are generally accurate, but appear sensitive to the variance of headway. The models can be used for studying setup of unit extensions and for example performance evaluation.

Our models illustrate some similarities between the actuated signal system and the pre-timed one. The latter usually assumes a uniform vehicle arrival. It has the same green time allocation as the expected green time allocation in an actuated signal system when the unit extension is set to zero.

One insight from our modeling is that the optimal unit extensions are generally not zero. This indicates that the queue clearance policy even in heavy traffic is not optimal. However, the findings here could change if the number of queuing vehicles is used as a condition for signal switching. In other words, the actuated signal control based solely on unit extensions is an adaptive but limited scheme.

Results in this chapter rely on an assumption of a stationary renewal arrival

process of vehicles, which does not agree well with the phenomenon of platoon arrivals. In addition, the assumption of no maximum/minimum green times ignores the overflow queues when there is a surge of arrival traffic.

5. GRAPHICAL METHODS AND DIFFUSION APPROXIMATIONS

This chapter will illustrate graphical methods and diffusion approximations to the traffic signal timing and delay problems. The practical importance and usefulness of graphical methods are advocated by Newell [81, 82]. What we will present next can be seen as an alternative interpretation of the graphical methods used in Newell [76]. From this interpretation, one can directly measure the variance of green time for the queue clearance, which remains yet to be carefully examined in practice and would be rather challenging if using conventional queueing techniques. Moreover, the graphical method explicitly presents both the deterministic and stochastic delay. We will also illustrate that the theoretical background for the graphical methods in this particular application is inherently diffusion approximation. Furthermore, we investigate the problems of disruptions occurred during a pre-timed traffic signal cycle. By diffusion approximation, we provide quantitative estimation on the duration that the effects of disruptions would dissipate.

5.1 Graphical Methods: Variance of Green Time, Deterministic and Stochastic Delay

We consider a two-phase traffic signal cycle where the green time G_c follows the red time R. Suppose the traffic demand is high and no residual queues at the beginning of the cycle. Without loss of generality, we assume the red time is known and the queue cumulated in red time can be cleared during the green time. Assume the arrival rate is λ and departure rate is f. We are interested in the variance of the portion of green time for the queue clearance and the delay to the vehicles in the queue. The above description is suitable for the queue problem at actuated signals without maximum green time. It is also convenient to the pre-timed signals.

If, however, there is residual queue for pre-timed signal, the average queue delay is simply equal to the delay due to the above process and the average residual queue delay.

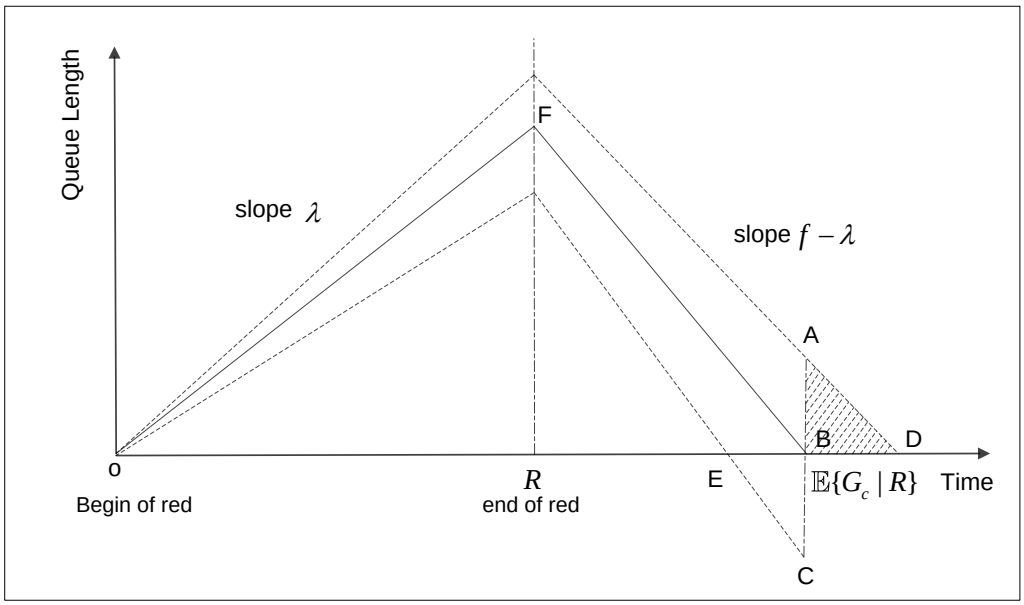

Figure 5.1: Illustration of the variance of green time for queue clearance by considering BD and EB as standard deviation. The value is directly estimated by the appropriate interpretation of the relationship between BD and AB. It also shows the deterministic delay (area OFB) and stochastic delay (area ABD).

To assess the variance of green time for queue clearance, it is suitable to adopt the cumulative queue curves as illustrated in Figure 5.1. The x-axis represents time from the beginning of red time, and the y-axis represents the length of cumulative queue. In Figure 5.1, the solid curve represents the expected cumulative queue length, which increases during the red and becomes zero at point B, the time $\mathbb{E}\{G_c|R\}$ (note we assume that the red time R is given). Two dashed lines represent two possible realizations due to the stochastic arrival and departure. For convenient analysis,

we can imagine that the lower curve is extended to the point C with x-coordinate $\mathbb{E}\{G_c|R\}$. Thus, at point B, the fluctuation of queue length is between point A and point C. Given red time R, such amplitude of queue fluctuation can be treated as the standard deviation of total arrivals and departures during the time $\mathbb{E}\{G_c|R\}$. In the same way, the deviation of queue clearance time from point B can be treated as the standard deviation. That is to say, we can approximately regard the length of BD and EB as the standard deviation of the green time for queue clearance. If traffic demand is high, the slope of any cumulative curve for queues near the time point can be approximated by $f - \lambda$. From the geometric illustration in Figure 5.1, we can use $f - \lambda$ to approximate the ratio of the following two standard deviations: (1) total arrivals and departures during one cycle and (2) the length of green time for queue clearance.

Therefore, if we denote the number of arrivals and departures by $A(t)$ and $D(t)$, respectively, we have the following approximate relation

$$(f - \lambda)^2 Var(G_c|R) = Var((A - D)(R + \mathbb{E}\{G_c|R\})|R). \tag{5.1}$$

In Equation (5.1), we square the standard deviation to become the variance. The above expression also implies that we regard the arrival and departure processes as independent. This is definitely an approximation when the traffic demand becomes high. Note that if in some cases we can assume the departure headway is constant, the righthand of Equation (5.1) reduces to $Var(A(R + \mathbb{E}\{G_c|R\})|R)$. If we further assume the variance to mean ratio for arrivals and departures as I_a and I_d, we will have $Var((A-D)(R+\mathbb{E}\{G_c|R\})|R) = (I_a+I_d)(R+\mathbb{E}\{G_c|R\})$. In fact, $\mathbb{E}\{R+G_c|R\}$ can be easily calculated as $(I_a + I_d)fR/(f - \lambda)$. Accordingly, we have from Equation

(5.1)

$$Var(G_c|R) = \frac{(I_a + I_d)fR}{(f-\lambda)^3}. \tag{5.2}$$

This is an exact formula in Newell [80], but here we use a different interpretation: We directly consider the fluctuations of certain quantity as the standard deviations. If one intends to obtain the variance of G_c, one needs to obtain $Var(\mathbb{E}\{G_c|R\})$ as well. In general, the term $Var(\mathbb{E}\{G_c|R\})$ is easy to calculate. The difficulty is actually to get Equation (5.2). When one obtains these two estimations, the variance of G_c can be calculated as $Var(G_c) = \mathbb{E}\{Var(G_c|R)\} + Var(\mathbb{E}\{G_c|R\})$.

It is interesting to explore the theoretical background of the above the treatment. In fact, the above treatment is to think of the queue as a Brownian motion (diffusion process) so that we can imagine a hypothetical queue fluctuates (AB and BC in Figure 5.1) according to the Gaussian distribution at the expected queue clearance time. Such approximation leads to the Gaussian distribution of the green time for the queue clearance. This is why we can reasonably consider the deviations in Figure 5.1 as standard deviations.

The above graphical method is useful particularly for the practical purpose since it avoids some formal calculations. If one wants a formal procedure, one may treat arrival and departure processes as independent during peak hours. Then one may model both the arrival and departure as drifted Brownian motions

$$A(R+G_c) = \sqrt{I_a\lambda}\mathbf{B}(R+G_c) + \lambda(R+G_c), \quad \text{conditional on } R, \tag{5.3}$$

$$D(G_c) = \sqrt{I_d \cdot f}\mathbf{B}(G_c) + f \cdot G_c, \quad \text{conditional on } R, \tag{5.4}$$

where $\mathbf{B}(\cdot)$ denotes the standard Brownian Motion. Since at the time G_c, we should

have $A(R+G_c) = D(G_c)$ [1]. After plugging this expression into the expressions (5.3) and (5.4), rearranging the terms and taking variance, we find

$$(\lambda - f)^2 Var\{G_c|R\} \quad = \quad Var\{\sqrt{I_a\lambda}\mathbf{B}(R+G_c) - \sqrt{I_d \cdot f}\mathbf{B}(G_c)|R\}. \qquad (5.5)$$

If we consider the arrivals and departures as independent, and note that $\mathbf{B}^2(t) - t$ is a Martingale, it is easy to obtain

$$Var\{\sqrt{I_a\lambda\mathbf{B}(R+G_c)}|R\} \quad = \quad I_a\lambda\mathbb{E}\{(R+G_c)\} \qquad (5.6)$$

$$Var\{\sqrt{I_d f\mathbf{B}(G_c)}|R\} \quad = \quad I_d f\mathbb{E}\{G_c|R\} \qquad (5.7)$$

Since the expected departures are equal to the expected arrivals, $\lambda\mathbb{E}\{(R+G_c)\} = f\mathbb{E}\{G_c|R\}$. Hence, we have the righthand of Equation (5.5) as

$$Var\{\sqrt{I_a\lambda}\mathbf{B}(R+G_c) - \sqrt{I_d \cdot f}\mathbf{B}(G_c)|R\} = (I_a + I_d)\lambda\mathbb{E}\{R+G_c|R\}. \qquad (5.8)$$

We obtain the same Equation (5.1) again. This result indicates that the above graphical methods has a background of diffusion approximation.

Furthermore, Figure 5.1 is of importance to understand the delay to the vehicles in the queue. From the description above, the area OFB under the solid curve, i.e., the delay due to the average queue length through the cycle, can be considered as the *deterministic delay*, as this part of delay can be calculated by considering the queue as deterministic. The shaded area, i.e., ABD in Figure 5.1, which represents the delay due to the random fluctuations of arrivals and departures, should be considered as the *stochastic delay*. This illustration leads to some useful application considerations. For example, if the ratio of area ABD over area OFB is small, we can discard the

[1] Rigorously speaking, $G_c = \arg\inf_T\{A(R+T) - D(T)|A - D \geq 0\}$.

random effects in practice. As indicated by Newell [83], usually the deterministic delay is the primary effect while the stochastic delay is the second.

5.2 Diffusion Approximation: Queues due to Disruptions

Disruptions are usually observed to affect the performance of the traffic signals. For example, a heavy truck arriving at the intersection during the red time may impose additional delay to the vehicles after its arrival due to its unusual long start-up time. Some individual drivers who are talking on cell phone and waiting during the red time are also observed to have a long response time to the green signal. If their response time is too long, there would be a longer residual queue at the end of this traffic signal cycle, particularly so during the rush hour. If at the beginning the traffic signal system is under equilibrium condition, the disruptions might take several cycles to get the system back to the equilibrium. We are interested here in how long the effects of disruption may dissipate.

In order to explore this problem, we will apply diffusion approximation techniques to the residual queues at a pre-timed traffic signal. To keep the presentation short, we use the same notions in Chapter 2 section 2.2.2 except that we add some superscripts to the notations to differentiate the cycles. We denote by $F^j_{A-D}(z)$ the distribution of $\{A(r+g) - D(g) \leq z\}$ during the j-th cycle, and by $F^j_Q(z)$ the distribution of $\{Q_0 \leq z\}$ during the j-th cycle. According to the Chapter 2 section 2.2.2, we have

$$F^{j+1}_Q(z) = \int_0^\infty F^j_Q(x) dF^j_{A-D}(z-x). \tag{5.9}$$

By diffusion approximation, we have

$$F^{j+1}_Q(z) - F^j_Q(z) = -\mathbb{E}(A-D) \cdot \frac{\partial F^j_Q(z)}{\partial z} + \frac{1}{2}\mathbb{E}(A-D)^2 \cdot \frac{\partial^2 F^j_Q(z)}{\partial z^2}. \tag{5.10}$$

If we observe the traffic signal performance at a large scale of time compared with a cycle time, we may approximate the discrete index j by a continuous variable t. Then the left hand of the above equation turns to $\frac{\partial F_Q^t(z)}{\partial t}$. If we denote $\mathbb{E}(A-D)$ by α and $\mathbb{E}(A-D)^2$ by β, we then have the following

$$\frac{\partial F_Q^t(z)}{\partial t} = -\alpha \cdot \frac{\partial F_Q^t(z)}{\partial z} + \frac{\beta}{2} \cdot \frac{\partial^2 F_Q^t(z)}{\partial z^2}. \tag{5.11}$$

The Equation (5.11) is under the boundary condition that $F_Q^t(z)$ should be a distribution as of z and vanish at $z = 0$. The disruptions to the traffic signal performance can be translated into the initial condition, since it is easy in practice to observe the effects on residual queue length during the same cycle. We denote the after the disruptions, the residual queue length is z_0. Thus the initial condition is $F_Q^0(z) = \delta(z - z_0)$ where $\delta(z - z_0)$ is the Direct function, i.e., $\delta(z - z_0) = 1$ if $z = z_0$ otherwise $\delta(z - z_0) = 0$.

The solution to the Equation (5.11) under the above initial condition and boundary condition is

$$F_Q^t(z) = \Phi\left(\frac{z - z_0 + \alpha t}{(\beta t)^{1/2}}\right) - e^{-2z\alpha/\beta}\Phi\left(\frac{-z - z_0 + \alpha t}{(\beta t)^{1/2}}\right), \tag{5.12}$$

where $\Phi(x)$ is the cumulative distribution function of standard Gaussian distribution. From this result, we see that the decay of the effects or the convergence to the equilibrium has the same rate as the function $\phi(t) = \Phi\left(\frac{z}{t^{1/2}}\right)$ goes to the infinity.

5.3 Summary

This chapter briefly presents the graphical methods with an application to the variance of time for queue clearance and the diffusion techniques to the effects of disruptions at the signalized intersections. We intend to emphasize here the importance

of graphical method because it can not only present the deterministic approximation but also reflect some results of the diffusion approximation. We also emphasize that the diffusion approximation is a suitable tool which can be applied to most queue related problems at traffic signals. What we have discussed is still one dimensional diffusion. For higher dimensional diffusion approximations, one can resort to the literature such as Newell's book on tandem queue [82].

6. CONCLUSION

In this research, the probabilistic models for left-turn bay blockage and spillback are presented to study the capacity and delay at the intersection with pre-timed signals. Then we provide the approaches to characterize the delay, the mean and variance of green times at one-way intersection with actuated signals. In all of the approaches, we choose to assume that only the mean and variance of arrivals are available. We also present graphical methods, in many ways similar to the Newell's original proposals, to illustrate the stochastic effects of arrivals in the context of diffusion approximation.

We do not intend to develop new techniques merely for their own sake. We attempt to choose appropriate techniques to solve the problems that arise in the practical applications. Although some methods we present here might be complicated to be readily used in practice, we provide heuristic or some rule-of-thumb methods to illustrate the results based on the obtained results.

In retrospect, the fluid and diffusion approximations are demonstrated their power in dealing with practical problems. In most cases, they allow us to describe some complicate phenomena in a fairly accurate way and to provide qualitative analysis for practical considerations. Sometimes the graphical method [81] is also useful in practice, and this approach should be illustrated more in the practical problems. in the future, we will consider the coordination between traffic signals [73, 15], and time-dependent diffusion methods [77, 78, 79].

REFERENCES

[1] J. Abate, G. L. Choudhury, and W. Whitt. Numerical inversion of multidimensional laplace transforms by the laguerre method. *Performance Evaluation*, 31(3-4):229–243, 1998.

[2] J. Abate and W. Whitt. The fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, 10(1-2):5–88, 1992.

[3] J. Abate and W. Whitt. Numerical inversion of probability generating functions. *Operations Research Letters*, 12(4):245–251, 1992.

[4] I. Adan, J. van Leeuwaarden, and E. Winands. On the application of rouché's theorem in queueing theory. *Operations Research Letters*, 34(3):355–360, 2006.

[5] I. J. Adan and Y. Zhao. Analyzing $GI/E_r/1$ queues. *Operations Research Letters*, 19(4):183– 190, 1996.

[6] L. V. Ahlfors. *Complex analysis*. McGraw-Hill, 1979.

[7] R. Akçelik. Time-dependent expressions for delay, stop rate and queue length at traffic signals. Technical report, Australian Road Research Board. Internal Report AIR 367-1, 1980.

[8] R. Akçelik. Traffic signals: Capacity and timing analysis. Technical report, ARR No.123. Australian Road Research Board, Kew, Victoria, Australia, 1981.

[9] R. Akçelik. Capacity of a shared lane. *Proceedings of the 14th ARRB Conference*, 14(2):228–241, 1988.

[10] R. Akçelik. Estimation of green times and cycle time for vehicle-actuated signals. *Transportation Research Record*, 1457:63–72, 1994.

[11] R. Akçelik and N. M. Rouphail. Estimation of delays at traffic signals for variable demand conditions. *Transportation Research B: Methodological*, 27(2):109–131, 1993.

[12] R. Akcelik, E. Chung, and M. Besley. Recent research on actuated signal timing and performance evaluation and its application in SIDRA5. Technical report, In the Compendium of Technical Papers of the 67th Annual Meeting of the Institution of Transportation Engineers, Boston, USA, 1997.

[13] R. Allsop. Delay at a fixed time traffic signal-i: Theoretical analysis. *Transportation Science*, 6(3):260–285, 1972.

[14] N. T. J. Bailey. On queueing processes with bulk service. *Journal of the Royal Statistical Society, Series B (Methodological)*, 16(1):80–87, 1954.

[15] E. Bavarez and G. F. Newell. Traffic signal synchronization on a one-way street. *Transportation Science*, 1:55–73, 1967.

[16] M. J. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics in Transportation*. Yale University Press, New Haven, 1956.

[17] M. A. A. Boon, I. J. B. F. Adan, E. M. M. Winands, and D. Down. Delays at signalised intersections with exhaustive traffic control. *Probability in the Engineering and Informational Sciences*, 26(3):337–373, 2012.

[18] M. A. A. Boon, R. D. van der Mei, and E. M. M. Winands. Applications of polling systems. *Surveys in Operations Research and Management Science*, 16:67–82, 2011.

[19] V. Broek, J. van Leeuwaarden, I. Adan, and O. Boxma. Bounds and approximations for the fixed-cycle traffic-light queue. *Transportation Science*, 40(4):484–496, 2006.

[20] H. Bruneel and B. G. Kim. *Discrete-Time Models for Communication Systems Including ATM*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1979.

[21] I. A. Burrow. Note on traffic delay formulas. *ITE Journal*, 59(10):9–32, 1989.

[22] H. Cartan. *Elementary Theory of Analytic Functions of One or Several Complex Variables*. Aiison-Wesley Publishing Company, Reading, Massachusetts, second edition, 1973.

[23] M. L. Chaudhry, C. M. Harris, and W. G. Marchal. Robustness of rootfinding in single-server queueing models. *INFORMS Journal on Computing*, 2(3):273–286, 1990.

[24] A. Clayton. Road traffic calculations. *Journal of Institute of Civil Engineering*, 16(7):247–264, June 1941.

[25] K. G. Courage, D. B. Fambro, R. Akcelik, P.-S. Lin, M. Anwar, and F. Viloria. Capacity analysis of traffic actuated intersections. Technical report, NCHRP Project 3-48 Final Report Prepared for National Cooperative Highway Research Program, Transportation Research Board, National Research Council, 1996.

[26] R. Cowan. An improved model for signalised intersections with vehicle-actuated control. *Journal of Applied Probability*, 15:384–396, 1978.

[27] R. J. Cowan. Useful headway models. *Transportation Research*, 9(6):371–375, 1975.

[28] D. R. Cox and W. L. Smith. *Queues*. London: Methuen & Co Ltd, 1961.

[29] C. D. Crommelin. Delay Probability Formulae When the Holding Times are Constant. *Post Office Electrical Engineers Journal*, 25:41–50, 1932.

[30] W. B. Cronjé. Analysis of existing formulas for delay, overflow and stops. *Transportation Research Record*, 905:89–93, 1983.

[31] J. Daniel, D. B. Fambro, and N. M. Rouphail. Accounting for nonrandom arrivals in estimate of delay at signalized intersections. *Transportation Research Record*, 1555:9–16, 1996.

[32] J. Darroch, G. Newell, and R. Morris. Queues for a vehicle actuated traffic light. *Operations Research*, 12(6):882–895, 1964.

[33] J. N. Darroch. On the traffic-light queue. *The Annals of Mathematical Statistics*, 35:380–388, 1964.

[34] F. Dion, H. Rekha, and Y.-S. Kang. Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections. *Transportation Research B: Methodological*, 38(2):99–122, 2004.

[35] F. Downton. Waiting time in bulk service queues. *Journal of the Royal Statistical Society, Series B (Methodological)*, 17(2):256–261, 1955.

[36] M. C. Dunne. Traffic delays at a signalized intersection with binomial arrivals. *Transportation Science*, 1:24–31, 1967.

[37] R. J. Engelbrecht, D. B. Fambro, and N. M. Rouphail. Validation of generalized delay model for oversaturated conditions. *Transportation Research Record*, 1572:122–130, 1997.

[38] D. Fambro and N. Rouphail. Generalized delay model for signalized intersections and arterial streets. *Transportation Research Record*, 1572:112–121, 1997.

[39] F. Garwood. An application of the theory of probability to the operation of vehicular-controlled traffic signals. *Journal of the Royal Statistical Societ: Series B (Methodological)*, 7(1):65–77, 1940.

[40] D. Gazis, R. Herman, and A. Maradudin. The problem of the amber signal light in traffic flow. *Transportation Research*, 8(1):112–132, 1960.

[41] J. Haddad and N. Geroliminis. The effect of left-turns in arterial capacity with queue spillbacks. In *The 92nd TRB Annual Meeting, Washington, D.C.*, 2013.

[42] D. Heidemann. Queue length and delay distributions at traffic signals. *Transportation Research Part B: Methodological*, 28(5):624–635, 1994.

[43] S. E. Jabari and H. X. Liu. A stochastic model of traffic flow: Gaussian approximation and estimation. *Transportation Research Part B: Methodological*, 47:15–41, 2013.

[44] A. J. E. M. Janssen and J. S. H. van Leeuwaarden. Analytic computation schemes for the discrete-time bulk service queue. *Queueing Systems*, 50(2-3):141–163, July 2005.

[45] A. J. E. M. Janssen and J. S. H. van Leeuwaarden. Back to the roots of the $M/D/1$ queue and the works of Erlang, Crommelin and Pollaczek. *Statistica Neerlandica*, 62(3):299–313, 2008.

[46] X. Jin, Y. Zhang, F. Wang, L. Li, D. Yao, Y. Su, and Z. Wei. Departure headways at signalized intersections: A log-normal distribution model approach. *Transportation Research Part C: Emerging Technologies*, 17(3):318–327, 2009.

[47] D. G. Kendall. Some problems in the theory of queues. *Journal of the Royal Statistical Society. Series B (Methodological)*, 13(2):151–185, 1951.

[48] S. Kikuchi, M. Kii, and P. Chakroborty. Lengths of double or dual left-turn lanes. *Transportation Research Record*, 1881:72–78, 2004.

[49] R. Kimber and E. Hollis. Traffic queues and delay at road junctions. Technical report, TRRL Laboratory Report No. 909. Transport and Road Research Laboratory, Berkshire, U.K., 1979.

[50] D. C. Kleinecke. Discrete time queues at a periodic traffic light. *Operations Research*, 12(6):809–814, 1964.

[51] P. Kruger, A. D. May, and G. Newell. The efficient control of an isolated vehicle-actuated controlled intersection: A comparison between queue control and volume-density control. Technical report, UCB-ITS-RR-90-1, Institute of Transportation Studies, University of California at Berkeley, 1990.

[52] J. P. Lehoczky. Traffic intersection control and zero-switch queues under conditions of Markov chain dependence input. *Journal of Applied Probability*, 9:382–395, 1972.

[53] J. P. Lehoczky. Stochastic models in traffic flow theory: intersection control. Technical report, Stanford University, California, August August, 1969.

[54] J. Li, N. M. Rouphail, and R. Akçelik. Overflow delay estimation for a simple intersection with fully actuated signal control. *Transportation Research Record*, 1457:73–81, 1994.

[55] J. Li, N. M. Rouphail, A. Tarko, and L. Velichansky. Overflow delay model for signalized arterials. *Transportation Research Record*, 1555:1–8, 1996.

[56] F.-B. Lin. Estimation of average phase durations for full-actuated signals. *Transportation Research Record*, 881:65–72, 1982.

[57] F.-B. Lin. Predictive models of traffic actuated cycle splits. *Transportation Research Part B: Methodological*, 16(5):361–372, 1982.

[58] P.-S. Lin and K. G. Courage. Phase time prediction for traffic-actuated intersections. *Transportation Research Record*, 1555:17–22, 1996.

[59] J. G. Little. Queueing of side-street traffic at a priority type vehicle-actuated signal. *Transportation Research*, 5:295–300, 1971.

[60] Y. Liu and G.-L. Chan. An arterial signal optimization model for intersections experiencing queue spillback and lane blockage. *Transportation Research Part C: Emerging Technologies*, 19(1):130–144, 2011.

[61] R. T. Luttinen. Statistical properties of vehicle time headway. *Transportation Research Record*, 1365:92–98, 1992.

[62] M. P. Malakapalli and C. J. Messer. Enhancements to the PASSER II-90 delay estimation procedures. *Transportation Research Record*, 1421:94–103, 1993.

[63] A. D. May and H. E. Keller. A deterministic queueing model. *Transportation Research*, 1(2):117 – 128, 1967.

[64] D. R. McNeil. A solution to the fixed-cycle traffic light problem for compound poisson arrivals. *Journal of Applied Probability*, 5(3):624–635, 1968.

[65] D. R. McNeil and G. H. Weiss. Chapter 2: Delay problems for isolated intersections. In D. C. Gazis, editor, *Traffic Science*, pages 109–174. John Wiley & Sons, Inc., 1974.

[66] C. J. Messer and D. B. Fambro. Effects of singnal phasing and length of left-turn bay on capacity. *Transportation Research Record*, 644:95–101, 1977.

[67] A. J. Miller. Settings for fixed-cycle traffic signals. *Operational Research Quarterly*, 14(4):373–386, December 1963.

[68] A. J. Miller. The capacity of signalized intersections in Australia. Technical report, ARRB Bulletin No.3. Australian Road Research Board, Nunawading, Australia, Martch 1968.

[69] R. W. T. Morris and P. G. Pak-Poy. Intersection control by vehicle actuated signals. *Traffic Engineering and Control*, 10:288–293, 1967.

[70] G. F. Newell. Mathematical models of freely-flowing highway traffic. *Operations Research*, 3:176–186, 1955.

[71] G. F. Newell. Statistical analysis of the flow of highway traffic through a signalized intersection. *Quarterly of Applied Mathematics*, 13:353–369, 1956.

[72] G. F. Newell. The effect of left turns on the capacity of a traffic intersection. *Quarterly of Applied Mathematics*, 17:67–76, 1959.

[73] G. F. Newell. The flow of highway traffic through a sequence of synchronized traffic signals. *Operations Research*, 8:390–405, 1960.

[74] G. F. Newell. Queues for a fixed-cycle traffic light. *The Annals of Mathematical Statistics*, 31:589–597, 1960.

[75] G. F. Newell. Discussion on Dr.Weiss's paper. In W. L. Smith and W. E. Wilkinson, editors, *Proceedings of the Symposium on Congestion Theory*, pages 281–282. The University of North Carolina Press, Chapel Hill, 1964.

[76] G. F. Newell. Approximation methods for queues with application to the fixed-cycle traffic light. *SIAM Review*, 7(2):223–240, 1965.

[77] G. F. Newell. Queues with time-dependent arrival rates: I. the transition through saturation. *Journal of Applied Probability*, 5(2):436–451, 1968.

[78] G. F. Newell. Queues with time-dependent arrival rates: II. the maximum queue and the return to equilibrium. *Journal of Applied Probability*, 5(3):579–590, 1968.

[79] G. F. Newell. Queues with time-dependent arrival rates: III. a mild rush hour. *Journal of Applied Probability*, 5(3):591–606, 1968.

[80] G. F. Newell. Properties of vehicle-actuated signals I: One-way streets. *Transportation Science*, 3:30–52, 1969.

[81] G. F. Newell. Graphical representations of queue evolution for multiple-server systems. In A. B. Clarke, editor, *Lecture notes in Economics and Mathematical Systems, Number 98, Proceedings of Conference on Mathematical Methods in Queueing Theory (1973)*, pages 63–79. Springer-Verlag, New York, 1974.

[82] G. F. Newell. *Approximate Behavior of Tandem Queues, Lecture Notes in Economic and Mathematical Systems number 171.* Springer-Verlag, 1st edition, 1979.

[83] G. F. Newell. *Applications of Queueing Theory.* Chapman and Hall, 2nd edition, 1982.

[84] G. F. Newell. Theory of highway traffic signals. Technical report, University of California at Berkeley, Institute of Transportation Studies, UCB-ITS-CN-89-1, 1989.

[85] G. F. Newell. Stochastic delays on signalized arterial highways. In M. Koshi, editor, *Transportation and Traffic Theory, Proceedings of the Eleventh International Symposium on Transportation and Traffic Theory, July 18-20,1990, Yokohama, Japan*, pages 589–598. Elsevier Science Publishing Co. Inc., New York, 1990.

[86] G. F. Newell. Delays caused by a queue at a feeway exit ramp. *Transportation Research Part B: Methodological*, 33(5):337–350, 1999.

[87] G. F. Newell and E. E. Osuna. Properties of vehicle-actuated signals II: Two-way streets. *Transportation Science*, 3:99–125, 1969.

[88] J. Niittymäki and M. Pursula. Saturation flows at signal-group-controlled traffic signals. *Transportation Research Record*, 1572:24–32, 1997.

[89] B. Noble. *Methods based on the Winer-Hopf Technique for the solution of partial differential equations*. Pergamon Press, INC, New York, 1958.

[90] K. Ohno. Computational algorithm for a fixed cycle traffic signal and new approximate expressions for average delay. *Transportation Science*, 12(1):29–47, 1978.

[91] P. L. Olson and R. W. Rothery. Drivers respoonse to the amber phase of traffic signals. *Transportation Research*, 9(5):650–663, 1961.

[92] B. Park, C. J. Messer, and T. Urbanik. Traffic signal optimization program for oversaturated conditions: Genetic algorithm approach. *Transportation Research Record*, 1683:133–142, 1999.

[93] F. Pollaczek. Über eine Aufgabe der Wahrscheinlichkeitstheorie. I. *Mathematische Zeitschrift*, 32:64–100, 1930.

[94] F. Pollaczek. Über eine Aufgabe der Wahrscheinlichkeitstheorie. II. *Mathematische Zeitschrift*, 32:729–750, 1930.

[95] Y. Qi, L. Yu, and M. Azimi. Determination of storage lengths of left-turn lanes at signalized intersections. *Transportation Research Record*, 2023:102–111, 2007.

[96] N. M. Rouphail and R. Akçelik. Oversaturation delay estimates with consideration of peaking. *Transportation Research Record*, 1365:71–81, 1992.

[97] N. M. Rouphail, M. Anwar, D. B. Fambro, P. Sloup, and C. E. Perez. Validations of generalized delay model for vehicle-actuated traffic signals. *Transportation Research Record*, 1572:105–111, 1997.

[98] N. M. Rouphail, K. G. Courage, and D. W. Strong. New calculation method for existing and extended Highway Capacity Manual delay estimation procedure. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006, 2006.

[99] N. M. Rouphail, A. Tarko, and J. Li. Chapter 9: Traffic flow at signalized intersections. In H. Lieu, editor, *Revised Monograph of Traffic Flow Theory*. Update and Expansion of Transportation Research Board (TRB) Special Report 165 "Traffic Flow Theory" published in 1975, 2000.

[100] D. Schrank and T. Lomax. The 2011 urban mobility report. Technical report, Texas Transportation Institute, College Station, Texas, 2011.

[101] D. W. Strong and N. M. Rouphail. Incorporating effects of traffic signal progression into proposed incremental queue accumulation method. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006, 2006.

[102] L. Takács. *Introduction to the Theory of Queues*. Oxford University Press, 1962.

[103] J. Tanner. A problem of interference between two queues. *Biometrika*, 40:58–69, 1953.

[104] TRB, National Research Council Washington, D.C. *Highway Capacity Manual 2000*, 2000.

[105] TRB, National Research Council Washington, D.C. *Highway Capacity Manual 2010*, 2010.

[106] J. van Leeuwaarden. Delay analysis for the fixed-cycle traffic-light queue. *Transportation Science*, 40:189–199, 2006.

[107] H. J. Van Zuylen and F. Viti. A probabilistic model for queues, delays, and waiting time at controlled intersections. *In the the 86th Transportation Research Board Annual Meeting CD-ROM, Washington, D. C.*, 2007.

[108] F. Viti and H. J. van Zuylen. A probabilistic model for traffic at actuated control signals. *Transportation Research Part C: Emerging Technologies*, 18(3):299–310, 2009.

[109] F. Viti and H. J. van Zuylen. Probabilistic models for queues at fixed control signals. *Transportation Research Part B: Methodological*, 44(1):120–135, 2010.

[110] M.-H. Wang and R. Benekohal. Arrival-based uniform delay model for exclusive protected-permitted left-turn lane at signalized intersections. *Transportation Research Record*, 2027:91–98, 2007.

[111] X. Wang. Modeling the process of information relay through intervehicle communication. *Transportation Research Part B: Methodological*, 41(6):684–700, July 2007.

[112] X. Wang, T. M. Adams, W.-L. Jin, and Q. Meng. The process of information propagation in a traffic stream with a general vehicle headway: A revisit. *Transportation Research Part C: Emerging Technologies*, 18(3):367–375, 2010.

[113] J. G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II*, 1:325–378, 1952.

[114] F. V. Webster. Traffic signal settings. Technical report, Road Research Technical Paper No. 39, Department of Scientific and Industrial Research, Road Research Laboratory, HMSO, London, U.K., 1958.

[115] F. V. Webster and B. M. Cobbe. Traffic signals. Technical report, Road Research Technical Paper No. 56, Department of Scientific and Industrial Research, Road Research Laboratory, HMSO, London, U.K., 1966.

[116] F. V. Webster and P. B. Ellson. Traffic signals for high-speed roads. Technical report, Road Research Technical Paper No. 74, Department of Scientific and Industrial Research, Road Research Laboratory, HMSO, London, U.K., 1965.

[117] E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis*. Cambridge University Press, 4th edition, 1963.

[118] X.-F. Xie, S. F. Smith, L. Lu, and G. J. Barlow. Schedule-driven intersection control. *Transportation Research Part C: Emerging Technologies*, 24:168–189, October 2012.

[119] Y. Xie. *Development and Evaluation of an Arterial Adaptive Traffic Signal Control System Using Reinforcement Learning*. PhD thesis, Texas A&M University, College Station, TX, December 2007.

[120] K. Yin, Y. Zhang, and B. X. Wang. Analytical models for protected plus permitted left turn capacity at signalized intersection with heavy traffic. *Transportation Research Record*, 2192:177–184, 2010.

[121] K. Yin, Y. Zhang, and B. X. Wang. Modeling delay during heavy traffic for signalized intersections with short left-turn bay. *Transportation Research*

Record, 2257:103–110, 2011.

[122] Y. Zhang and J. Tong. Modeling left-turn blockage and capacity at signalized intersection with short left-turn bay. *Transportation Research Record*, 2071:71–76, 2008.

APPENDIX A

LITERATURE REVIEW ON PRACTICAL RESEARCH OF TRAFFIC SIGNALS

> The mathematical approach, aimed at deeper understanding of traffic phenomena, has resulted in a seemingly unending number of theoretical solutions to traffic problems. Admittedly, this mathematical approach ..... is not always of immediate help to the practicing traffic and highway operations engineer. Yet, familiarity with the concepts brought out by traffic flow theorists is vitally needed in order to arrive at the practical solutions that must be found.
>
> —- from the Foreword in Highway Research Record No. 89 (1965).

A.1 Introduction

There is always a tension in modern days between traffic engineers, who are looking for the methods directly applicable to real-world situations, and those academic theorists who are seeking mathematical models on a solid ground. While the methods in practice may sacrifice rigorous reasoning to some degree, most of the theoretical models may lead theorists to focus on too much detail in precise mathematical structures and may be too complicated to be readily applicable. It seems that the gap between them results in the ever-lasting incompatibility. However, in retrospect, there are numerous noted researchers and engineers who made extensive contributions to bridge this gap. As Wardrop commented and warned, a theoretical background is required in the practical traffic engineering study, but the theory should not be divorced from experiment [113]. In addition to the discussion in Chapter 2, we will review the development on the practical side here, with a focus on how

the theoretical and practical work interact.

A.2 Isolated Pre-Timed Signal

A.2.1 Steady-State Condition

One of the main focuses of traffic signal studies is to obtain average delay under stationary conditions for both arrivals and departures. Systematic treatment of this subject from both theoretical and practical point of view can be traced back to as earlier as the work by Wardrop [113] and the references therein [24]. Except the statement of his two well-known equilibriums for network, Wardrop discussed the effects of random delay and its difference from uniform delay at a signalized intersection [113]. Some preliminary results of delay for pre-timed and vehicle-actuated signals were also provided. This work then inspired Newell to perform the first theoretical study on traffic delay in terms of distribution [70]. However, as the above scholars noticed, the average delay formula for pre-timed signals is in fact rather difficult to obtain.

Webster [114] conducted an extensively systematic study on the performance and control of pre-timed signals. He also commented that the developed methods might be applicable to the vehicle-actuated signals where the green times run to maximum due to heavy traffic demands, indicating a general use of his formula. The Webster's formula has the following expression:

$$d = \frac{c(1-\lambda)^2}{2(1-\lambda x)} + \frac{x^2}{2q(1-x)} - 0.65\left(\frac{c}{q^2}\right)^{(1/3)} x^{(2+5\lambda)}, \tag{A.1}$$

where d is the average delay per vehicle for a particular direction; c is the cycle time; s is the saturation flow; q is the flow; λ is the ratio of green time and cycle time, i.e., g/c; x is the degree of saturation, i.e., $q/\lambda s$. The first term accounts for uniform

delay, which according to Webster [114] is much agreement under light traffic flow condition. The expression of the second term is derived from the $M/D/1$ queues (Poisson arrival/constant service rate/one server) based on the pioneered theoretical work by Kendall [47]. The third term, on the other hand, is purely empirical with an aim of fitting the simulation results. By comparing with extensive simulation results, Webster noticed that the third term was generally below 15 percentage of average delay [114]. According to this fact, the third term was neglected when deriving the optimum green times and cycle time. Although these results were obtained in a more complicated formula, the expression for the optimal cycle time was approximated for practical purposes [114], yielding

$$c = \frac{1.5L + 5}{1 - Y}, \tag{A.2}$$

where L is the total lost time per cycle (sec) and Y is the sum of the highest flow to saturation ratio for one approach in each phase. Due to the simplicity of this expression and its root in theoretical reasoning, it has been widely used in engineering practice. Later, Webster and Cobbe [115] provided a more comprehensive treatment of the traffic signals.

In most of studies, the delay formulas depend on a particular arrival distribution. In order to derive the results insensitive to the detailed stochastic structures of arrivals, Miller [67] employed the variance-to-mean ratio, I, to capture the relationship between the first two moments of arrivals. He gave the expression of the average delay as

$$d = \frac{1 - \lambda}{2(1 - \lambda x)} \left\{ \frac{2\mathbb{E}(Q)}{q} + (c - g) + \frac{I - 1}{s} + \frac{\lambda x}{s} \right\}, \tag{A.3}$$

where $\mathbb{E}(Q)$ is the average length of residual queue (overflow) and the meaning of

other notations is kept the same with Equation (A.1). When dealing with the estimation of $\mathbb{E}(Q)$, Miller implicitly used the properties of Brownian motion (diffusion processes in general) to obtain an approximation when traffic intensity becomes heavy. He finally produced the following approximation [67]:

$$d = \frac{1-\lambda}{2(1-\lambda x)}\left\{\frac{I(2x-1)}{q(1-x)} + (c-g) + \frac{I-1}{s} + \frac{\lambda x}{s}\right\}. \tag{A.4}$$

This expression can be comparable to the Webster's for Poisson arrivals, and is useful in many practical cases. Perhaps this Miller's work, along with the Brownian motion theory in physics and the development of heavy traffic limits in queueing theory, motivated Newell to pursue much further along this line. Consequently, Newell [76] had established the powerful techniques of diffusion approximation and obtained the formula of average delay much agreed with Webster's in a systematic way, thereby successfully solving the problem. Newell's work has been discussed in Chapter 2.

The theoretical development of discrete-time model and bulk service queues (see Chapter 2) also have the influence on practice as well. In his another work, Miller [68] provided the following approximation expression instead of root-finding to estimate the $\mathbb{E}(Q)$:

$$\mathbb{E}(Q) = \frac{\exp[-1.33\sqrt{sg}(1-x)/x]}{2(1-x)}. \tag{A.5}$$

It seems that the parameters in the above equation were calibrated after evaluating the exact value of solutions under $sg = 10, 30, 90$, $0.4 \leq x \leq 0.96$, and Poisson arrivals (p.7 in [68]). The equation (A.5) with a simplified version of Equation (A.3) gives an accurate solution to various cases [90, 99], and leads to the further simplification for practice use [7].

We want to close this brief discussion in this section by remarking two points

here. Although Ohno [90], Allsop [13] and others made comparisons by evaluating the above delay expressions through simulation, some of the formulas have their own advances especially in some particular situations. More important, the theoretical techniques and approximation methods developed among the efforts of obtaining these results are invaluable and should be noted in both theory-oriented and practice-oriented research.

A.2.2 Analysis Period Dependence

The above analysis of residual queues relies on the steady-state condition. This condition reflects in the tendency to infinity as the traffic demand approaches saturation. Furthermore, it may take a long period for a signalized intersection to reach the steady-state condition in real environment. The problem might become severe when the demand varies in time. However, establishing a theoretical solution for queues with a time-dependent demand is rather difficult, let along the subject of the varying demand around the capacity. To overcome this obstacle in practice, Kimber and Hollis [49] purposely employed a coordinate transformation technique to the residual queue length (overflow) and then the delay expression. But the very original idea maybe have the root in a study by Webster and Cobbe [115] when they investigated the effect of a parked vehicle on delay at an intersection (see Figure 27 in [115]). It was based on the results of designed simulations and the observation that long residual queues may remain for many cycles if demand fluctuates around the capacity. The approach is to smooth the delay under the steady-state condition into the oversaturated results obtained by deterministic approaches. As shown in Figure A.1 for residual queue length, the transformed curve maintains the relationship of $1.0 - X_1 = L_1 = L_2 = X_3 - X_2$ if all three curves have the same length of queue. Although the original formulas for both delay and residual queue length are

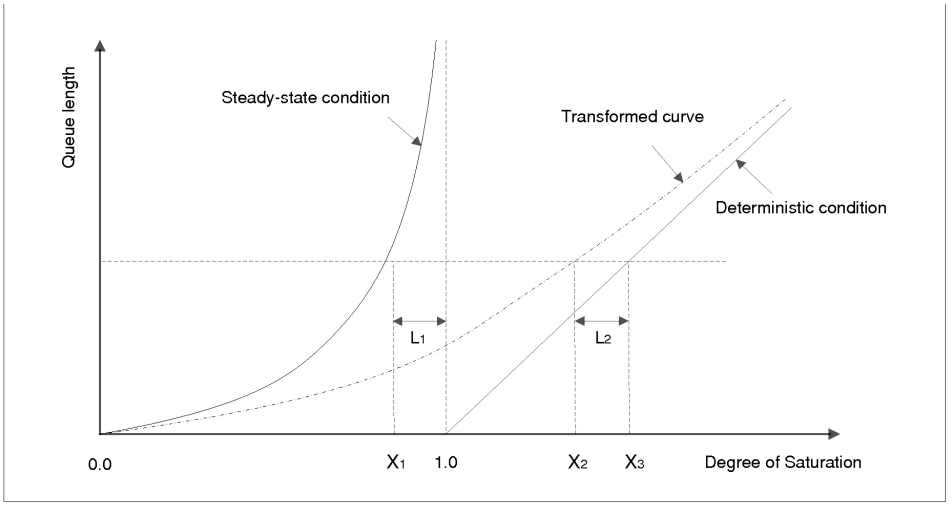

Figure A.1: Coordinate transformation (modified from Figure 6 in Kimber and Hollis[49]).

in a general form [21], they are considered as inconvenience in the sense of practice. Akçelik [8] simplified the formula and generated the following equation for residual queue length

$$\mathbb{E}(Q) = \frac{Q_{cp}T}{4}\left[(x-1) + \sqrt{(x-1)^2 + \frac{12(x-x_0)}{Q_{cp}T}}\right], \text{when } x > x_0, \tag{A.6}$$

where Q_{cp} is the capacity, T the length of analysis period and x_0 the degree of saturation below which the average residual queue has length of zero [8]. To make a convenient applicable formula, Akçelik and Rouphail [11] further proposed several modification rules for determining variable traffic flows in different analysis periods based on deterministic demand functions [63].

The coordinate transformation approach is essentially heuristic and does not represent theoretically validity [109]. The time-dependent analysis still needs future investigation. Nevertheless, given that rare progress has been made on the theo-

retical side, it is still widely adopted in the American, Canadian, and Australian capacity guides [109] in a variety of formulas. In fact, as Burrow [21] and Dion [34] noted, the delay formulas for pre-timed signals used in various countries are similar.

A.2.3 Other Issues

Since the establishment of the general form of delay model, numerical efforts have been made to improve the model and calibrate the parameters in a state-of-art manner. For example, In Li *et al.*[55], the effects of signal coordination were taken care of and were involved in the calibration of parameters. Roupail and Akçelik [96] described and compared the path-trace and queue sampling methods for estimating delay in practice. Engelbrecht *et al.* [37] applied these methods to oversaturated condition. In another study, Fambro and Rouphail [38] incorporated related parameters to count the effects of upstream signal metering, platoon dispersion and queue spillback. In addition, the effect of left-turn lane geometry on delay also received attention. Messer and Fambro [66] recognized that the left-turn bay length would significantly affect delay. This topic still needs some further investigation today.

Apart from the study on delay, the problem of dilemma zone when signal changing to amber also got attention from the early development. The adequacy of signal change interval was at first given from a theoretical perspective by Gazis *et al.* [40] and Olson and Rothery [91]. Then Webster and Ellson [116] carried out a thorough experiments and practical studies and proposed a method whereby the signal system would not change to amber when drivers fall into the dilemma zone [116]. The effects of types and locations of controllers and vehicle speed on such treatment were also discussed. Due to the limitation of the space, we do not elaborate all the efforts here.

A.3 Isolated Vehicle-Actuated Signal

It is much more difficult to study vehicle-actuated signal than pre-timed signal. It is partly because all the approaches can be treated separately at pre-timed signal, while the approaches at vehicle-actuated signal should be considered as a whole, as many researchers noted [114, 32, 26]. It is also partly because no "theory" can explicitly solve the problem of delay at vehicle-actuated signal and provide convenient formulas for practical advance.

It is common for a vehicle-actuated controller to set maximum and minimum green times. In a rigorous analysis, it means that one has to determine whether a queue has been cleared before or after the minimum green time and whether a queue could be cleared before the maximum green time. The consequence is a too much information involved equation from which a closed-form solution cannot be obtained. Even for a one-way street intersection with a basic control setting, i.e., fixed unit extension[1], the difficulties cannot be overcome. Perhaps due to this reason, most of theoretical studies make the assumption that there is neither maximum green nor minimum green time [32, 80, 26]. For a two-way street intersection, the difficulty arises when evaluating the maximum queue clearance time and the simultaneous satisfactory of unit extensions on two opposing approaches. Nevertheless, the problem can be tackled by focusing on a particular distribution or by approximate analysis for general cases. The noted work on theoretical side include Darroch *et al.* [32], Newell [80, 87] and Cowan [26]. Particularly, the fluid and diffusion approximation developed by Newell [80, 87] are very useful and powerful in both theoretical research and advanced practice.

Next section we will review some widely used approaches in practice for vehicle-

[1]In convention unit extension is also called passage time or gap time.

actuated signal. These approaches arise from empirical studies and have employed both heuristics and crude mathematical modeling. Admittedly, rigorous models that incorporate all practice-related factors and control parameters are unlikely to be developed and unlikely to be of help to the traffic engineers. Hence, the rule-of-thumb approach is necessary for the real-world complicated applications, as indicated in Appendix H in Akçelik [8]. However, it does not indicate that we can be unaware of the weakness of these approaches. Despite modern controllers able to adapt the gap time to the speed of actuated vehicles, our discussion will be limited to the operation of a basic type of actuated controller which uses a fixed unit-extension time.

A.3.1 Green Time and Cycle Length

The concept of degree of saturation for vehicle-actuated signals is of considerable concern, since the green times and cycle length vary cycle by cycle, leaving some challenges to determine the degree of saturation. It may be proper to define the degree of saturation by the average phase lengths. However, practicing operations of vehicle-actuated signals demands not only the average values but also the variance. Among the methods in practice, the approaches proposed in Lin[56, 57] and Akçelik[10] may represent the most useful methods.

A set of models have been proposed by Lin[56, 57] for the estimation of average cycle splits from semi-actuated and fully-actuated signal controls. For semi-actuated signals, the method is workable and of practical importance. For full-actuated signals, there may need some more discussions. The proposed models rely on the following

equations for estimation of average cycle splits:

$$\gamma_1 = 0.5Y_1 + Y_2 + I_1 + g_2, \tag{A.7}$$

$$\gamma_2 = 0.5Y_2 + Y_1 + I_2 + g_1, \tag{A.8}$$

$$g_i = \sum_{n=0}^{\infty}(I_i + D_n + E_n)P(n/\gamma_i),\ i = 1, 2, \tag{A.9}$$

where g_i represents green times of phase i, D_n represents the part of the green extended by a moving queue in the critical lane after the initial portion on that direction, E_n the additional green extension after queue clearance by successive arrivals whose headways are less than one unit extension, and $P(n/\gamma_i)$ the probability of number of arrivals during time period γ_i.

If γ_i is known, then, as the author claimed in [56, 57], the average green G_i can be estimated for phase i from the Equation(A.9). However, this method may be uncertain for the desired results due to the following reasons. First of all, it may need, in practice, to figure out how the n arrivals form a queue and how long it will take the queue to be clear. In reality, only a portion of the n arrivals form the queue, while such portion is totally random in different cycles. The practice engineers hence need to specify how many of these arrivals contribute the D_n and others contribute the E_n, which may lead to some difficulties. Or, one can use Poisson distribution for the probability $P(n/\gamma_i)$ as in [56, 57], in that the explicit expressions are readily available for $P(n/\gamma_i)$ and deduced exponential headway. In this case, the headway distribution cannot be shifted exponential any more as assumed in [56, 57].

In practice, it is of importance to gain some knowledge of the variance of green times in that such quantity attributes to the delay [80]. However, the above models are essentially deterministic so that they cannot provide an estimation of the variance of green times. If the methods would improve to take the randomness of evolution

processes of queues into consideration, or if the estimation for D_n would capture the variance of queueing processes, the practice would benefit from the knowledge of moments of green times.

Akçelik[10] presented alternative methods to estimate the average green times and cycle length at vehicle-actuated signals. In order to involve the platoon arrivals, the methods employed the Cowan's headway distributions [27] for estimation of the extension green time beyond the queue clearance. The proposed methods show the advantage in that they are direct and convenient enough for practical use.

The proposed method used the following equation for green time estimation:

$$g = g_{min} + g_e = g_{min} + g_s + e_g, \tag{A.10}$$

where g_{min} is minimum green, g_s queue clearance time or saturated portion of green time, and e_g the green extension time after g_s. The calculation of g_s was suggested as the result of the equation for the case of a single green period per cycle:

$$g_s = f_q(yr)/(1-y), \tag{A.11}$$

where f_q is a calibration factor, y flow ratio (arrival flow rate over saturation flow rate), s saturation flow rate and r red time. The Equation (A.11) depends on the red time thereby it is not a real closed form formula. As indicated by the author, it requires iterative computation to obtain the numerical solution. The implication of this equation is that the arrival traffic flow can be viewed as fluid or deterministic flow.

Another approximation is also proposed by the following equations:

$$c = \frac{L + G_m + E}{1 - Y'}, \tag{A.12}$$

$$g = f_q y c + (1 - y) e_g, \tag{A.13}$$

where L is lost time per cycle, G_m the sum of green times for critical movements whose green times are set to minimum or maximum, $E = \sum(1 - y_i)e_{gi}$ the adjusted extension time for all critical movements, and Y' adjusted flow ratio. This method was also used in Rouphail *et al.* [97] in an alternative version. The advantage of this set of equations rests on no requirement of iterative computation. It is based on heuristic and serves as an approximation to the average value of green time and cycle time. It should be noted that the calculation of E must be based on the arrival distribution.

As mentioned previously, the above analysis for green times and cycle length provides a way to approximate the average value but the variance, which is a missing part in practice. Moreover, it is vital to note that the green time for one approach partially depends on the queues (randomly) accumulated during the green time for the other approach. Hence, the random interdependence of signal phases for two conflict traffic flows may be large in some cases to affect the performance of signals. This effect may be taken into consideration in practice as well.

A.3.1.1 On Green Extension for Cowan's M3 Model

The green extension mentioned in previous section depends on the arrival distribution. Akçelik[10] provided an expression (Equation 14) and was cited in Lin and Courage [58], and Rouphail *et al.* [99]. However, it is incorrect in the general form, though it is correct for exponential case.

The following equation was provided in [10] for green extension:

$$e_g = \frac{\exp[\lambda(e_0 - \Delta)]}{\phi q} - \frac{1}{\lambda}, \tag{A.14}$$

We denote the gap time (unit extension or critical gap in our paper) by e_0 as the same notation in [10]. The underlining arrival headway for the equation (A.14) obeys the cumulative distribution function (Cowan's M3 model [27]):

$$F(t) = \begin{cases} 1 - \phi \exp[-\lambda(t - \Delta)], & \text{for } t \geq \Delta, \\ 0, & \text{for } t < \Delta, \end{cases} \tag{A.15}$$

where $\lambda = \phi q/(1 - \Delta q)$, q total arrival per second, Δ minimum headway and ϕ proportion of unbanched vehicles. For exponential distribution, we have $\Delta = 0$, $\phi = 1.0$ and $\lambda = q$. In this case, the green extension formula agrees with the correct expression (see, for example, Lin[56]).

For the M3 distribution in the general form, however, Equation (A.14) is incorrect. In fact, the green extension for any general distribution $F(t)$ in a renewal process can be calculated as (see, for example, Wang *et al.*[112] for the principle):

$$e_g = \frac{\int_0^{e_0} t dF(t)}{1 - F(e_0)}. \tag{A.16}$$

It is critical to note that for M3 distribution, $F(\Delta) = 1 - \phi$ and $F(t) = 0$ if $t < \Delta$. Hence, the integral $\int_0^{e_0} t dF(t)$ should be considered as the Stieltjes integral (not the

normal Riemann integral), i.e.,

$$\begin{aligned}
\int_0^{e_0} t dF(t) &= \int_{\Delta-}^{e_0} t dF(t) \\
&= \int_{\Delta}^{e_0} t dF(t) - (1-\phi)\Delta \\
&= \lambda\phi \exp[\lambda\Delta] \int_{\Delta}^{e_0} t \exp[-\lambda t] dt - (1-\phi)\Delta \\
&= \phi(-e_0 - \frac{1}{\lambda}) \exp[-\lambda(e_0 - \Delta)] \\
&+ \phi(\Delta + \frac{1}{\lambda}) - (1-\phi)\Delta. \qquad \text{(A.17)}
\end{aligned}$$

Since $1 - F(e_0) = \phi \exp[-\lambda(e_0 - \Delta)]$ and $\frac{1}{\lambda} = \frac{1}{\phi q} - \frac{\Delta}{\phi}$, then the Equation (A.16) yields:

$$e_g = \left(2\Delta\left(1 - \frac{1}{\phi}\right) + \frac{1}{\phi q}\right) \exp[\lambda(e_0 - \Delta)] - \frac{1}{\lambda} - e_0. \qquad \text{(A.18)}$$

The last term $-e_0$ can be dropped due to the different setting in [10]. However, the coefficient of $\exp[\lambda(e_0 - \Delta)]$ is different from the one in Equation (A.14). The Equation 14 in [10] incorrectly regards $2\Delta(1 - \phi^{-1})$ as 0.

A.3.2 On Delay Models

The delay models for vehicle-actuated signals are basically adopted from the Webster's formula for fixed-time signals. Such adoption is based on the observation that during the peak hours the vehicle-actuated signals are frequently running to maximum. This observation also has been made as early as in Webster [114] and the related adoption has been suggested by Akçelik in his report [8]. In practice, it contains the uniform delay d_1 and incremental delay d_2 [97], though there is another term called initial queue delay in the HCM 2010 [105]. The uniform delay is essentially the same as the first term in the Webster's formula, whereas the incremental

delay is obtained by the coordinate transformation described in the previous section [49].

The uniform delay requires the average cycle length of actuated signals as input. The transformation for the incremental delay allows to handle the case of degree of saturation larger than 1 in a heuristic way. Specifically, the incremental delay has the following expression in general:

$$d_2 = aT\left[(x-1)+\sqrt{(x-1)^2+\frac{bkI(x-x_0)}{Q_{cp}T}}\right], \tag{A.19}$$

where k is the incremental delay factor, Q_{cp} capacity, I ratio of variance to mean arrivals per cycle, and a, b and x_0 parameters. Most of the past literature focuses on how to modify these parameters to accommodate the issues arising from contemporary applications. For example, Li *et al.*[54] drop the progression factor and X^2 used as a part of the parameter a in the 1985 HCM. Daniel, Fambro and Rouphail [31] provide an estimate of the kI in Equation (A.19) for pre-timed, semiactuated and fully actuated signals, and compare with the model in Malakapalli and Messer [62]. Rouphail *et al.* [97] made the efforts to validate the delay model by extensive simulation studies. Similar models have been also applied in the series of software SIDRA [12] and in the NCHRP report [25].

More recently, another method, known as incremental queue accumulation (IQA), was proposed by Rouphail *et al.* ([98, 101]) for estimating the performance of a signalized intersection. The IQA method explicitly considers the cumulative arrivals and departures during each time interval. In practice, this method is presented in graphs, from which the results are easy to estimate. The original ideas have their roots in Newell [83], in which the graphical method is greatly emphasized. These contributions finally resulted in the delay formulas for actuated signals in HCM2010

[105].

Despite widely usage in real-world applications, the adoption from delay formulas for pre-timed signals lacks of rigorous examination. Although the current techniques such as queueing theory are not able to handle the transition between near saturated and oversaturated situations, it indeed demands the further research to better our understanding on this subject.

A.4 Conclusion

There are numerical achievements for practical solutions when theorists and traffic engineers have well communication. Needless to say, the existing problems in operating traffic signals today highly encourage such collaboration to go further. Overall, all the efforts made by the above practicing pioneers have demonstrated as excellent examples that adequately bring theoretical analysis to real-world applications and make useful results for the purpose of practice. This principle has to be carried on by all engineers and researchers.

APPENDIX B

SUPPLEMENT TO CHAPTER II

B.1 Proofs of Theorems

Theorem 3 (Darroch [33]) *The sufficient and necessary condition to guarantee the existence of the stationary distribution of $X_{k,n}$ is*

$$(g+r)\mu_Y < g. \tag{B.1}$$

PROOF. We first prove the sufficiency, mainly following the proof in Darroch [33]. Notice that the states $X_c = i, i \in \mathbb{Z}_+$ (here we ignore the index n for cycle) in consecutive cycles forms a Markov Chain. Obviously it is irreducible and aperiodic, it is sufficient to show this Markov Chain has a positive finite mean recurrence time. Let N denote the number of cycles between two events $\{X_c = 0\}$. If we denote by

$$S_m = \sum_{n=1}^{m} \left(\sum_{k=0}^{c-1} Y_{k,n} - g \right), \tag{B.2}$$

then $\{N > p\} = \cap_{m=1}^{p} \{S_m > 0\}$. Note that $\mathbb{P}(S_m > 0) = \mathbb{P}(\frac{1}{m} S_m - \mathbb{E}(S_1) > -\mathbb{E}(S_1))$ and $\mathbb{E}(S_1) = (g+r)\mu_Y - g < 0$. Suppose m is large enough, then $m^{-1/2}\sigma(S_1)^{-1}(S_m - m\mathbb{E}(S_1)) \sim N(0,1)$ where $N(0,1)$ is normal distribution. Hence, using Chebychev's inequality,

$$\begin{aligned} \mathbb{P}(\frac{1}{m} S_m - \mathbb{E}(S_1) > -\mathbb{E}(S_1)) &\leq \frac{\mathbb{E}(m^{-1/2}\sigma(S_1)^{-1}(\mathbb{E}(S_m) - m\mathbb{E}(S_1)))^4}{m^2\sigma(S_1)^2\mathbb{E}(S_1)^4} \\ &= \frac{1}{m^2}(6\sigma(S_1)^4(E)(S_1)^4) \\ &= \frac{K}{m^2}, \text{ as } m \text{ large enough.} \end{aligned} \tag{B.3}$$

Therefore, $\mathbb{P}(N > p) < \mathbb{P}(S_p) < K/p^2$, as p large enough, and then

$$\begin{aligned}
\mathbb{E}(N) &= \sum_{p=0}^{\infty} \mathbb{P}(N > p) \\
&\leq 1 + \sum_{p=1}^{M} \mathbb{P}(N > p) + \sum_{p=M}^{\infty} K/p^2 \\
&\leq 1 + M + \sum_{p=M}^{\infty} K/p^2 \\
&< \infty, \text{ for some } M \text{ large enough.}
\end{aligned} \tag{B.4}$$

Hence, $\{X_c = 0\}$ has finite mean recurrence time. Based on the previous discussion, the Markov Chain formed by X_c is ergodic. Obviously the same result is applied to the queue length distribution during all slots of the cycle. The sufficiency is clear since the ergodic property guarantees the existence of the stationary distribution.

The necessary condition is relatively easy to prove. From the Equation (2.8) in Chapter 2, we know $\sum_{k=r}^{c-1} q_k = \frac{g - c\mu_Y}{1-\mu_Y}$. Since $q_k \leq 1$, we should have $g - c\mu_Y > 0$. □

One of the important issues in discrete queues is about the number of zeros of equation $z^g = Y(z)^c$. Adan, van Leeuwaarden and Winands[4] proved the following theorem:

Theorem 4 *Suppose the following conditions are satisfied*

(i) $|Y(z)| = 1$ *is differentiable at* $z = 1$ *and* $Y'(1) < g$*;*

(ii) $|Y(z)| = 1$ *on the unit cycle only at* $z = 1$ *and* $Y(0) > 0$*;*

Then $z^g = Y(z)^c$ *has* g *distinct roots* $z_0, z_1, ..., z_{g-1}$ *within the unit circle, where* $z_0 = 1$ *and* $|z_k| < 1$ *for* $k \neq 0$.

In order to ease the presentation of the following theorem, we do not use the concept of the period of the series as in Adan, van Leeuwaarden and Winands [4] since the period will be exactly 1 in most cases. We also remark here that because $Y(z)$ is the probability generating function, $Y(z)$ is analytic in $|z| < 1$ and continuous up to the unit cycle.

The proof of existence can trace back to the work by Pollaczek [93, 94] and Crommelin [29] when they dealt with the case of Poisson arrivals. In these early literature including Darroch [33], a stronger condition is assumed, i.e., $Y(z)$ is analytic in $|z| < 1 + \delta$ for some $\delta > 0$. Adan, van Leeuwaarden and Winands[4] used the classical argument and truncation of the $Y(z)$ to prove the above the theorem. The advantage of the proof is that it does not exclude many heavy tailed distributions whose radius of convergence is exactly 1. These heavy tailed distributions include, for instance, discrete Pareto and discrete lognormal distributions [4].

B.2 Approximate Average Queue Length in Newell [74] for $\mu > 1$

As mentioned in the section Bulk Service Queue in Chapter 2, we will investigate how $\mu = [g - \mu_Y(r+g)][rg/(r+g)]^{-\frac{1}{2}}$ arises and what it means to the approximation. The discussion is the same as in Newell [74]. Since the approximation methods, which involve asymptotic analysis and error estimation, are useful in many applications, we will provide some details here.

During the light traffic demand under equilibrium, it is unlikely the residual queues would appear in several consecutive signal cycles. Suppose no queue at the beginning of the cycle (with probability one), we can evaluate the distribution of residual queues cycle by cycle until it converges. However, whereas in applications, a qualitative description can be obtained through the evaluation for only one cycle. As long as such result decays rapidly with the decrease of arrival rate, the result can

serve as an approximation to the average queue length.

In Newell [74], the results are obtained in the sense of asymptote, i.e., r and $g \to \infty$ with r/g fixed(or comparable with O(1)). The asymptotic methods can avoid some difficulties in analysis and obtain the results often approximately accuracy. In the case of binomial arrival per slot for the discrete time model, we have for $0 < j \ll r$

$$\begin{aligned}
&\mathbb{P}(X_c = j) = \binom{r+g}{g+j}(1-\mu_Y)^{r-j}\mu_Y^{g+j} \\
= \;& \frac{(r+g)!(1-\mu_Y)^r\mu_Y^g}{g!r!}\left\{\frac{r\mu_Y}{g(1-\mu_Y)}\right\}^j \exp\left(-\sum_{i=1}^{j}\ln(1+\frac{i}{g}) + \sum_{i=1}^{j-1}\ln(1-\frac{i}{r})\right) \\
= \;& \frac{(r+g)!(1-\mu_Y)^r\mu_Y^g}{g!r!}\left\{\frac{r\mu_Y}{g(1-\mu_Y)}\right\}^j \exp\left(\frac{-j^2(r+g)}{2rg} + O\left(\frac{i^3}{r}\right)\right), \qquad \text{(B.5)}
\end{aligned}$$

where $\ln(1-x) = -x + O(x^2)$ is applied.

Note in light traffic demand, the large values of probability would concentrate on the cases with small values of j. Hence, we can carry the following calculation

$$\begin{aligned}
&\mathbb{E}(X_c) = \sum_{j=1}^{\infty} j\mathbb{P}(X_c = j) \\
= \;& \frac{(r+g)!(1-\mu_Y)^r\mu_Y^g}{g!r!}\sum_{j=1}^{\infty} j\left\{\frac{r\mu_Y}{g(1-\mu_Y)}\right\}^j \left\{1 - \frac{j^2(r+g)}{2rg} + O\left(\frac{j^4}{r^2}\right)\right\}, \\
= \;& \frac{(r+g)!(1-\mu_Y)^r\mu_Y^g}{g!r!}\left\{\frac{rg\mu_Y(1-\mu_Y)}{[g-\mu_Y(g+r)]^2}\right\}\left(1 - \frac{r+g}{2rg} - 3\frac{(r+g)\mu_Y(1-\mu_Y)}{[g-\mu_Y(g+r)]^2} + ...\right) \\
= \;& \frac{(r+g)!(r+g)(1-\mu_Y)^{r+1}\mu_Y^{g+1}}{g!r!\mu^2}\left(1 + O\left(\mu^{-2}\right)\right), \qquad \text{(B.6)}
\end{aligned}$$

where $\sum jx^3 = 6x^2(1-x)^{-4} + x(1-x^{-2})$ is applied. Though we cannot see how μ arises from this equation, we use this notation to simply the expression. Note that there is a little difference between this equation and the original one in Newell [74].

Suppose r and g sufficiently large, then applying Stirling's formula $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)$ yields:

$$\begin{aligned}\mathbb{E}(X_c) &= \left(\frac{gr}{2\pi(r+g)}\right)^{1/2}\frac{1}{\mu^2}\exp\left\{(r+1)\ln\frac{(r+g)(1-\mu_Y)}{r}+(g+1)\ln\frac{(r+g)\mu_Y}{g}\right\}\\ &\quad\cdot\left(1+O\left(\mu^{-2}\right)\right).\end{aligned} \tag{B.7}$$

The parameter μ arises exactly if we expand the function of logarithm:

$$\begin{aligned}&(r+1)\ln\frac{(r+g)(1-\mu_Y)}{r} = (r+1)\ln\left(1+\frac{g-(r+g)\mu_Y}{r}\right)\\ =\ &(r+1)\left(\frac{g-(r+g)\mu_Y}{r}\right)-\frac{1}{2}(r+1)\left(\frac{g-(r+g)\mu_Y}{r}\right)^2+O\left(r^{-2}(g-(r+g)\mu_Y)^3\right)\end{aligned} \tag{B.8}$$

and

$$\begin{aligned}&(g+1)\ln\frac{(r+g)\mu_Y}{g} = (g+1)\ln\left(1-\frac{g-(r+g)\mu_Y}{g}\right)\\ =\ &-(g+1)\left(\frac{g-(r+g)\mu_Y}{g}\right)-\frac{1}{2}(g+1)\left(\frac{g-(r+g)\mu_Y}{g}\right)^2+O\left(g^{-2}(g-(r+g)\mu_Y)^3\right).\end{aligned} \tag{B.9}$$

If we sum up the above two equations and neglect the terms of order r^{-2} and $r-g$(since we assume r and g are comparable), then we see the main term is exactly $-\mu^2/2$. The error term will be small if we further assume $\mu \ll r^{1/6}$. Therefore, the asymptotic expression for $\mathbb{E}(X_c)$ under this condition is

$$\mathbb{E}(X_c) = \left(\frac{gr}{2\pi(r+g)}\right)^{1/2}\frac{1}{\mu^2}\exp\left\{-\frac{\mu^2}{2}\right\}\cdot\left(1+O\left(\frac{\mu^3}{r^{1/2}}\right)+O\left(\mu^{-2}\right)\right). \tag{B.10}$$

For $\mu > 1$ and r large enough, as μ_Y decreases, μ increases faster so that $\mathbb{E}(X_c)$ decreases faster. It indicates the above asymptotic expression (B.6), particularly the expression (B.10), is indeed a good approximation during the range $\mu > 1$.