CHAPTER 4

DELAY ESTIMATION AT UNDERSATURATED SIGNALIZED APPROACHES

Delay is one of the key parameters that are utilized in the optimization of traffic signal timings. Furthermore, delay is a key parameter in computing the level of service provided to motorists at signalized intersections. Delay, however, is a parameter that is difficult to estimate because it includes the delay associated with decelerating to a stop, the stopped delay and the delay associated with accelerating from a stop. While many methods and software are currently used to estimate the delay incurred by motorists as they approach a signalized intersection, very little research has been conducted to assess the consistency of delay estimates among the various analytical and simulation approaches. In an attempt to systematically evaluate and demonstrate the assumptions and limitations of different delay estimation approaches, this chapter compares the delay estimates from numerous models for an undersaturated signalized intersection considering uniform and random arrivals. Specifically, the chapter compares a theoretical vertical queuing analysis model, the queue-based models used in the 1994 and 2000 versions of the Highway Capacity Manual, the queue-based model in the 1995 Canadian Capacity Guide for Signalized Intersections, a theoretical horizontal queuing model derived from shock wave analysis, and the delay estimates produced by the INTEGRATION microscopic traffic simulation software.

The results of the comparisons when considering uniform arrivals indicate that all delay models produce identical results under such traffic conditions, except for the estimates produced by the INTEGRATION software, which tends to estimate slightly higher delays than the other approaches. These differences are attributed to two factors. First, the INTEGRATION model computes the delay associated with a constrained vehicle deceleration and acceleration, while the analytical approaches assume instantaneous unconstrained vehicle deceleration and acceleration levels. Second, the INTEGRATION model, unlike the analytical queuing and shock wave analysis models, captures the discrete nature of traffic flow, causing the delay estimates computed by model to be sensitive to the departure times of vehicles. The results of the comparisons for scenarios assuming random arrivals also indicate that the delay estimates obtained by a micro-simulation model like INTEGRATION are consistent with the delay estimates computed by the analytical approaches.

4.1 INTRODUCTION

4.1.1 OVERVIEW OF QUEUING THEORY

A primary objective in operational problems involving flow is to ensure that the average capacity can handle the average flow, so that persistent traffic jams do not occur. However, because of fluctuations in demand and service times, merely guaranteeing that highway capacity can handle traffic demand on the average does not preclude the formation of transient or even permanent bottlenecks.

Queuing theory was developed in order to describe the behavior of a system providing services for randomly arising demands. It originated in a paper written by A. K. Erlang in 1909 on the problem of congestion in telephone traffic. In later works Erlang observed that telephone systems were characterized by Poisson inputs, exponential holding times, and multiple service channels. Over the past four decades, much research into queuing theory has been conducted, particularly in the field of operations research. In order to characterize the performance of a queuing system, a number of input parameters are required including:

(a) the distribution of arrivals;
(b) the input source, whether finite or infinite;
(c) the queue discipline, whether first-in-first-out, priority, or random selection;
(d) the channel configuration, that is, the number of channels and whether the channels are in series or in parallel; and
(e) the distribution of service times for each channel.

Input and service-time distributions may be of any form, however the science of queuing theory has been developed around three special types of distributions: deterministic *D*, random *M*, and Erlang E_a. Using this terminology, an M/D/1 model denotes a single-channel queue (parameter 1) with a random (Poisson) distribution of arrivals (parameter M) and a deterministic distribution of service times (parameter D).

A queuing system is said to be in state *n* if it contains exactly *n* items ($n > 0$), including those items in queue and those in service. If the arrival rate *q* is less than the service rate *C*, stability exists in the system and there is a finite time-independent probability of the queue being in any state. However, if the ratio of flow rate to service rate is greater than unity, the state of the waiting line increases in length and is no longer independent of time.

Using standard queuing theory, the average number of vehicles in a system assuming an infinite source can be computed using Equation 4.1. The relationship results in an infinite number of vehicles at a volume-to-capacity (v/c) ratio (denoted as ρ) of 1.0. The average delay within the system is computed using Equation 4.2 by dividing the expected number of vehicles by the channel capacity (*C*), which again tends to infinity at a v/c ratio of 1.0.

$$E(n) = \frac{\rho}{1-\rho} \tag{4.1}$$

$$E(w) = \frac{\rho}{1-\rho} \times \frac{1}{C} \tag{4.2}$$

4.1.2 Significance of Delay as a Measure of Effectiveness

Vehicle delay is perhaps the most important parameter used by transportation professionals to measure the performance of signalized intersections. The importance of this parameter is evident in the fact that it is utilized in both the design and the evaluation of traffic signalized intersections. As an example, delay minimization is frequently used as a primary optimization criterion when determining the operating parameters of traffic signals at both isolated and coordinated signalized intersections. In another example, the Highway Capacity Manual (HCM) uses the average delay incurred by vehicles on approaches to signalized intersections as a criterion for determining the level of service provided to motorists by the traffic signals (TRB, 1998). Delay is frequently used as an optimization and evaluation criterion because it is a measure of performance that a driver can directly relate to. Moreover, delay is a criterion whose meaning is easily comprehended by both traffic professionals and the general public. However, delay is also a parameter that is not easily determined. On this subject, Teply (1989) indicated that a perfect match between delay measured in the field and analytical formulas could not be expected. In addition, while various models have been proposed for estimating delays at signalized intersections (Hurdle, 1984; Akcelik, 1988; Teply 1989; Akcelik and Rouphail, 1993), very little research has been concerned with the issue of consistency of delay estimates from one model to the other. In one study, McShane and Roess (1990) suggested that the delay models derived from queuing analysis and shock wave analysis yield different results when both applied to the analysis of a bottleneck. Using a simple numerical example, they demonstrated that the use of a model derived from queuing analysis may underestimate the overall magnitude of delays when compared to the estimates produced by a model derived from shock wave analysis. Nam (1998) and Chin (1996) have demonstrated similar findings. The analysis presented in this chapter, however, will demonstrate that the queuing theory and shock wave analysis do indeed provide identical results in terms of vehicle delay. Furthermore, the chapter demonstrates that the computation of delay using car-following without an explicit delay formulation results in similar delay estimates compared to analytical formulations. Furthermore, the chapter presents the assumptions, limitations, and strengths of each of the delay estimation approaches.

4.1.3 Objectives and Layout of the Chapter

The objectives of the chapter are twofold. First, the chapter describes the assumptions, limitations, and strengths of the different delay computation methods. Second, the chapter demonstrates the consistency and/or inconsistency between the different methods in terms of the approach and the computed delays for undersaturated fixed-time signalized intersections assuming uniform and random arrivals.

In order to achieve this goal, the chapter first presents some background material on vehicle delay attributed to traffic signal operations at signalized intersections. This presentation is then followed by a description of the various delay models that are compared in the chapter. These models include a vertical theoretical delay model based on queuing analysis, the delay models used in the HCM 1994 (TRB, 1994), 1995 Canadian Capacity Guide for Signalized Intersections (ITE, 1995), and the HCM 2000 (TRB, 1998), a horizontal theoretical delay model based on shock wave analysis, and six delay estimates produced by the INTEGRATION microscopic traffic simulation software (Van Aerde, 1998). Following these descriptions, comments are made regarding the consistency of delay estimates from one model to the other for the case of an undersaturated signalized intersection. The comparison of delay estimates for the case of a congested intersection is the subject of a forthcoming chapter.

4.2 DELAYS AT SIGNALIZED INTERSECTIONS

The delay experienced by a vehicle approaching a signalized intersection approach is defined as the difference between the travel time experienced by the vehicle in order to traverse the intersection and the travel time that would have been experienced by the same vehicle in the absence of the traffic signal. To illustrate this definition, Figure 4.1 shows the speed profile of a vehicle stopping at a traffic signal before accelerating to its desired speed after the traffic signal turns green. The profile, which was generated using the INTEGRATION microscopic simulation software, indicates that the vehicle enters the intersection approach link after 783 seconds of simulation and starts to decelerate after 900 seconds of simulation to come to a complete stop at 916 seconds. Once the traffic signal turns green, the vehicle accelerates and attains its desired speed of 60 km/h at 947 seconds. Knowing that 32 seconds is normally required by a vehicle to travel across the simulated intersection and that it took the simulated vehicle a total of 212 seconds to do so, it can be determined that the vehicle incurred in this case a 180-second delay. As indicated in the figure, this delay graphically corresponds to the area between the line indicating the vehicle free-flow speed and the actual speed profile of the vehicle. Mathematically, this delay can be computed from instantaneous speed measurements using Equation 4.3.

$$D = \int_{t=0}^{T} \left[\frac{u_f - u(t)}{u_f} \right] dt \tag{4.3}$$

where:

T = duration of the vehicle's trip,

u_f = vehicle speed under free-flow conditions,

$u(t)$ = instantaneous speed at time t.

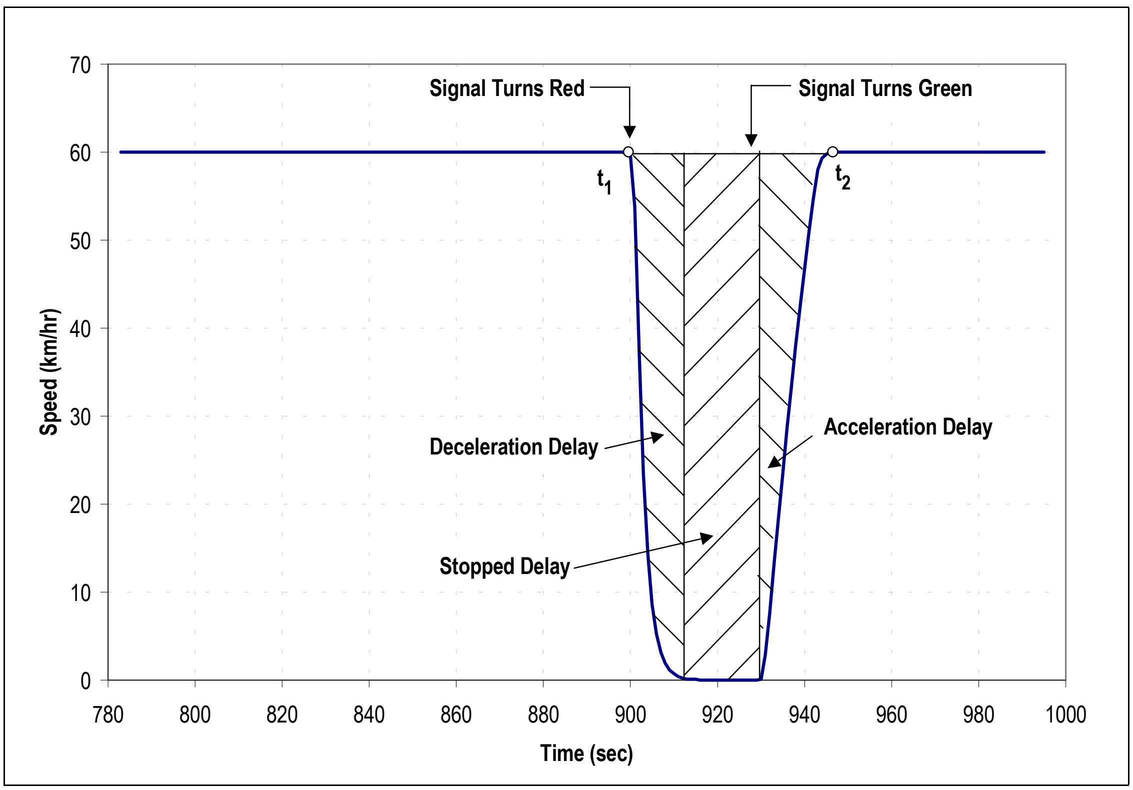

Figure 4.1: Simulated Speed Profile of a Vehicle Crossing a Signalized Intersection

Figure 4.1 also illustrates the notion of deceleration delay, stopped delay and acceleration delay. Typically, transportation professionals define stopped delay as the delay incurred by a vehicle when fully immobilized, while deceleration and acceleration delay are the delay incurred by a moving vehicle when it is either decelerating or accelerating. In some studies, stopped delay also includes any delay incurred by the vehicle while it is moving at an extremely low speed (e.g. less than 5 km/h). Specifically, the 1995 Canadian Capacity Guide for Signalized Intersections (CCG) defines stopped delay as any delay incurred by a vehicle traveling at a speed lower than the average pedestrian speed (ITE, 1995).

To better illustrate the distinction between deceleration, stopped and acceleration delay, consider Figures 4.2 and 4.3, which illustrate the trajectory and selected speed profiles of a series of vehicles arriving at a signalized intersection within a 60-second cycle at a rate of 720 veh/h. This rate corresponds to an arrival rate of 15 vehicles per 60 seconds, or 5 vehicles per second. As was the case in Figure 4.1, these figures were created using output from the INTEGRATION model. In Figure 4.2, it can be observed that vehicles 1 through 8 come to a complete stop, either as a direct consequence of the display of a red signal indication or because at least one vehicle is queued at the intersection stop line. All these vehicles thus incur deceleration, stopped and acceleration delay. In the figure, it can also be observed that vehicles 9 through 11, which reach the intersection near the end of the queue dissipation, only experience deceleration and acceleration delay as they only need to slow down to maintain a safe distance with the vehicles ahead. Finally, the last vehicle (vehicle 12) to reach the intersection before the return of the red interval experiences no delay, as this vehicle joins the platoon of previously queued vehicles after they have started to move at free-flow speed.

While most of the delay experienced by motorists at signalized intersections is directly caused by the signal operation, a fraction of the delay is attributable to the time required by individual drivers to react to changes in the signal display and to accelerate from a stop to their desired speed. This delay in vehicle departures is termed *start-up* lost time. As an example, Figure 4.4 illustrates the variation in the time interval between successive stop line cross times for each of the 12 simulated vehicle arrivals of Figure 4.2. The figure demonstrates that the first vehicle started moving across the stop line 4.3 seconds after the green initiation. The second, third, and fourth vehicles then followed with respective headways of 3.0, 2.7 and 2.4 seconds. Following these vehicles, the headway then tends to the 2-second saturation flow rate headway, which corresponds to a saturation flow rate of 1800 veh/h. In the case of Figures 4.2 and 4.3, it should be noted that the longer headways that are simulated for the first four simulated vehicles are not caused by delayed reactions from drivers, but are the result of constrained vehicle accelerations. This is due to the fact that the INTEGRATION model assumes instantaneous reaction times. As with many simulation models, true delayed reaction time is take into consideration by simulating effective signal timings.

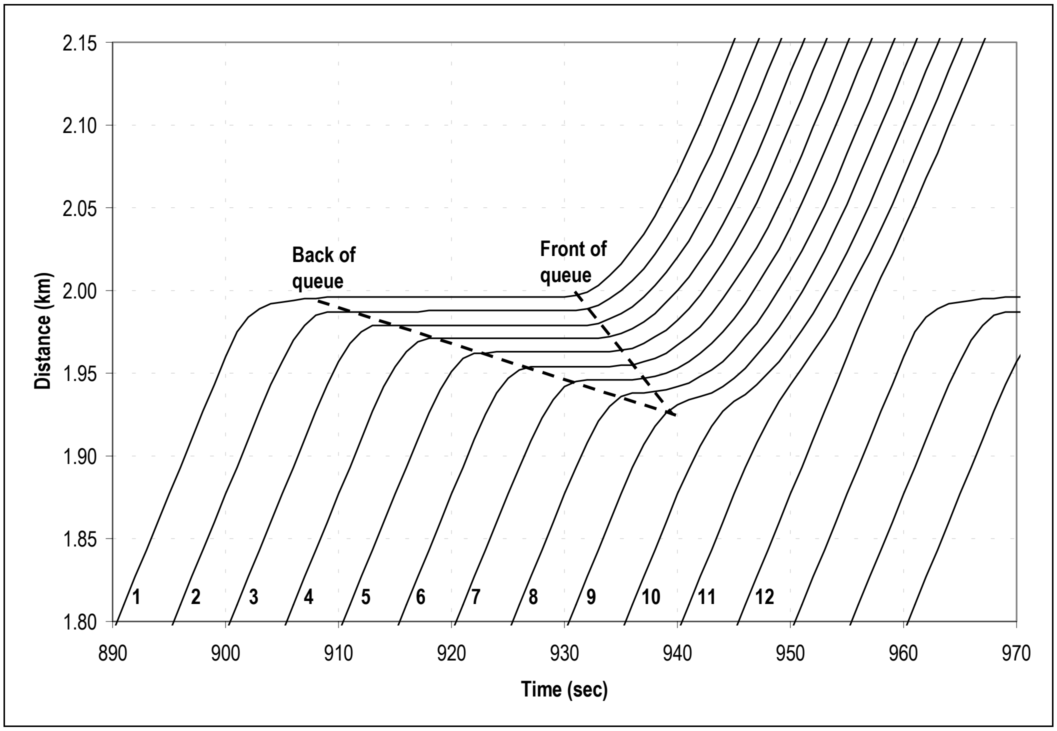

Figure 4.2: Simulated Time-Space Diagram for Typical Traffic Signal Cycle

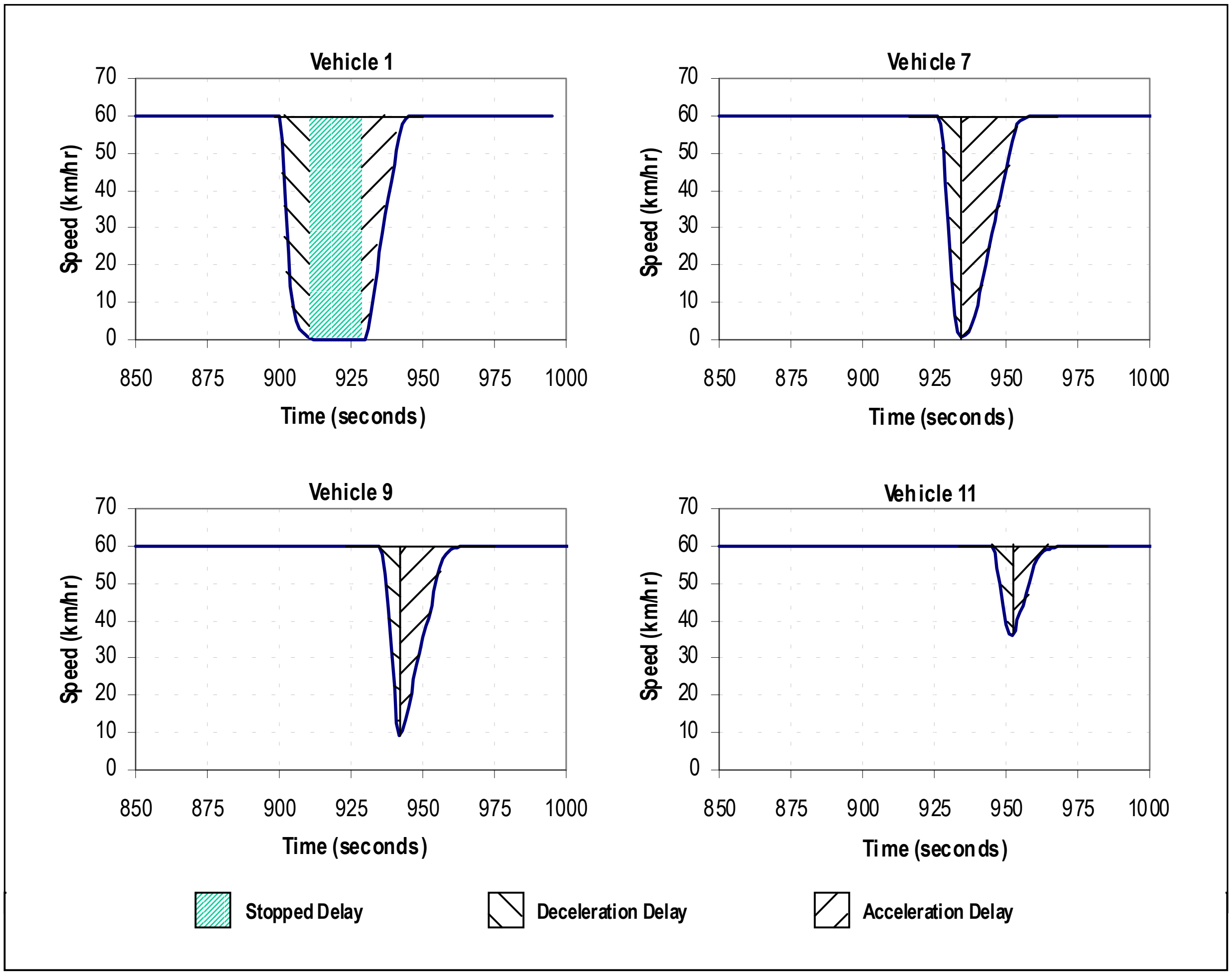

Figure 4.3: Speed Profile of Selected Vehicles in Typical Signal Cycle

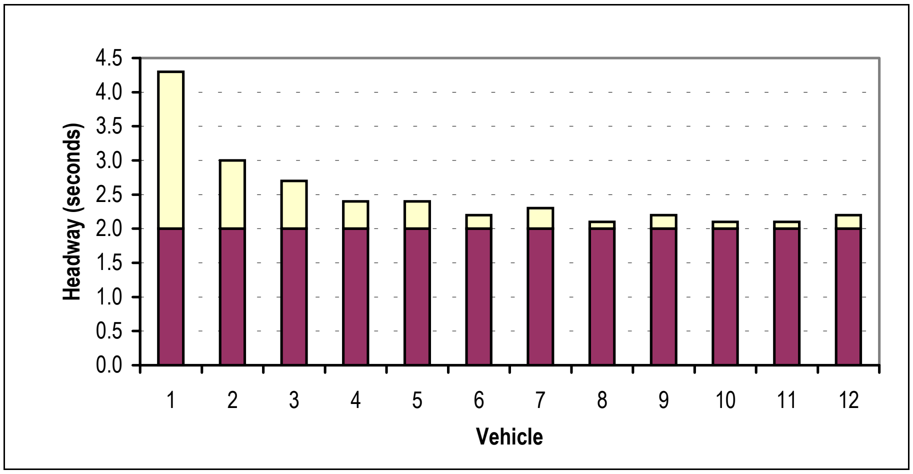

Figure 4.4: Simulated Headway Distribution at Beginning of Green Interval

The start up lost time is conceptually illustrated in Figure 4.5 by the delayed increase in the vehicle departure rate at the beginning of the green interval. Assuming that the demand is high and that vehicles are queued at the end of the green interval, field observations also indicate that vehicles continue to depart at the saturation flow rate during a portion of the amber interval. This utilization of the green interval is termed the *end-gain*. The portion of the cycle length that involves flow at saturation flow rate is termed the effective green, which is computed using Equation 4.4.

$$g = G - T_s + T_e \tag{4.4}$$

where:

g = effective green time (seconds),

G = displayed green time (seconds),

T_s = start-up lost time (seconds),

T_e = end-gain (seconds).

The delay function for vehicle arrivals at a signalized approach is non-linear. This is illustrated in Figure 4.6. Specifically, the figure illustrates the analytical delay relationship as a function of the v/c ratio. As mentioned earlier, standard queuing theory would indicate that delay tends to infinity as the v/c ratio tends to 1.0 because it assumes an infinite analysis period. Instead, field observation indicates that delay is finite at v/c ratios in excess of 1.0 because oversaturation delay is not the product of an infinite demand, but the product of a high demand followed, at some point in time, by a lower demand that allows the congestion to dissipate.

The randomness of vehicle arrivals results in a delay function that tends to a uniform delay model at low v/c ratios and a deterministic oversaturation delay model at high v/c ratios (in excess of 1.3). At v/c ratios in the range of 0.8 to 1.2, the stochastic nature of traffic arrivals results in significantly higher delays than estimated by standard deterministic queuing models. In this range of v/c ratios, the non-linear relationship between delay and the v/c ratio means that the marginal delay associated with an increase in demand is higher than the one associated with a decrease in demand. This causes the delay associated with random arrivals to be higher than the delay associated with uniform arrivals.

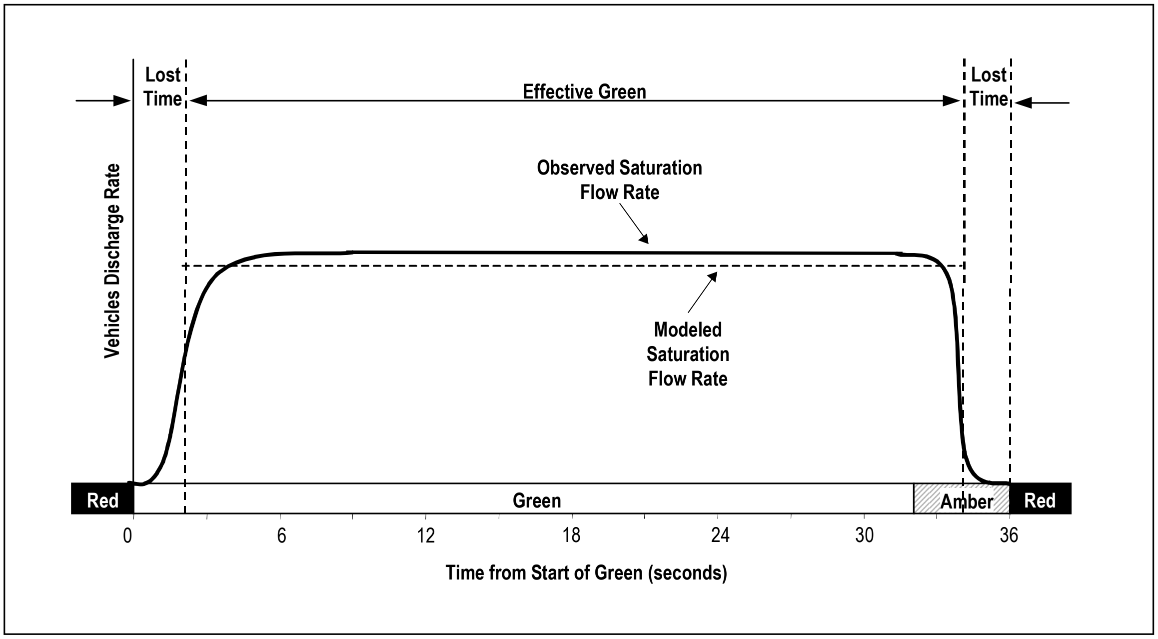

Figure 4.5: Typical Stop Line Departure Profile under Over-Saturated Conditions

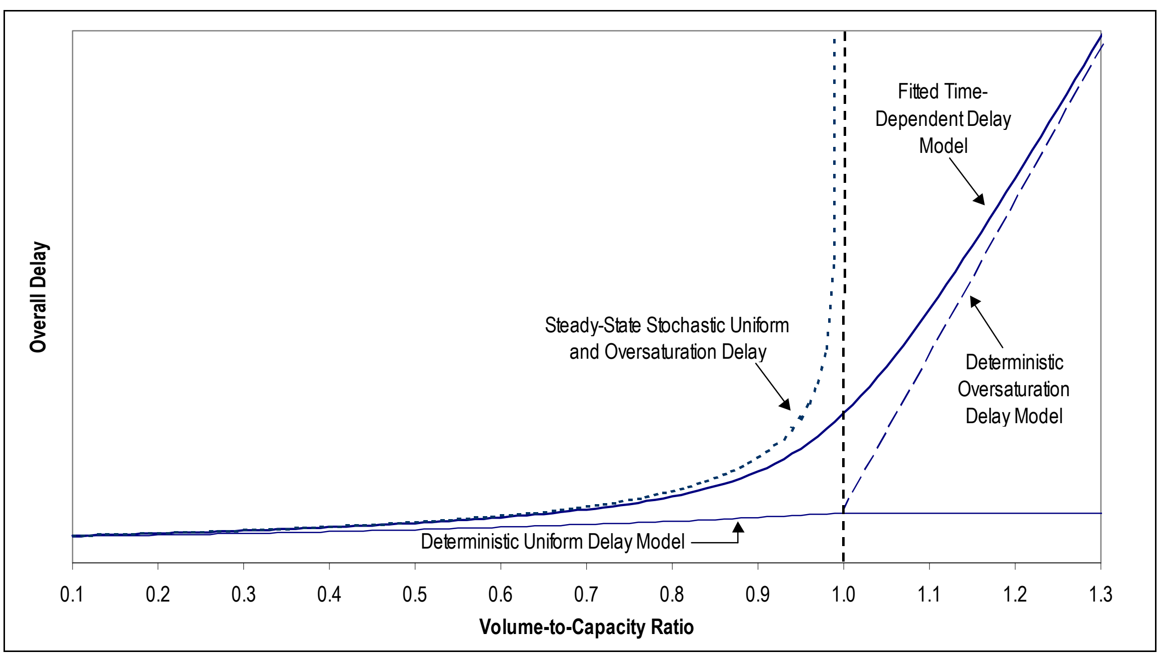

Figure 4.6: Deterministic Queuing Theory Delay Functions

4.3 MICROSCOPIC SIMULATION DELAY MODELS

Microscopic simulation models are commonly used to evaluate alternative traffic-improvement projects prior to their field implementation. A key factor in the use of a simulation tool is its validity, or consistency with standard traffic flow theory. Unfortunately, the analytical approach to traffic flow theory is typically limited in scope and, furthermore, requires simplifying assumptions. Consequently, an analysis of these simplifying assumptions and limitations together with their impacts on the accuracy of the analytical approaches should be accounted for while validating simulation models.

Microscopic traffic simulation models have the ability to track individual vehicle movements within a simulation environment. Vehicle behavior is usually modeled utilizing car-following, lane-changing and gap-acceptance logic. As a result, these models can compute the delay incurred by vehicles without the need for analytical delay formulas. In addition, by constraining the vehicle deceleration and acceleration capabilities, micro-simulation models can capture the deceleration and acceleration components of the delay, a factor that is beyond the scope of the current state-of-the-practice analytical approaches.

This section describes how delay is computed within the INTEGRATION software. Furthermore, this section describes the different delay outputs that are provided by the model. Subsequent sections will compare the delay estimates from the INTEGRATION model to standard traffic flow theory. The objective of the comparison is twofold. First, the comparison will demonstrate the validity of the simulation software within the scope of the state-of-the-art analytical approaches. Second, the comparison will demonstrate the assumptions and limitations of the state-of-the-art analytical approaches.

In the analysis presented in the chapter, the INTEGRATION software is used to estimate delays at an approach to a fixed-time signalized intersection. This model features an integrated dynamic traffic simulation and traffic assignment model. As indicated in Table 4.1, the current version of the model allows delay estimates to be produced based on second-by-second vehicle trackings using Equation 4.3. Depending on the output file analyzed, the model offers the opportunity to directly compare delay estimates at the individual vehicle, link, and network levels.

Table 4.1: INTEGRATION Delay Estimation Outputs

Output File	Delay Estimate	Delay Formula
File 10	Total travel on individual links	Total link travel time – Total link free flow travel time
File 11	Average delay on individual links	Sum of travel times experienced by vehicles on link – Sum of free flow link travel times
File 12	Average delay on individual links in time interval t	Sum of travel times experienced by vehicles on link in interval t – Sum of free flow link travel times in interval t
File 15	Network delay experience by individual probe vehicles	Time taken to complete a network trip – Free flow travel time along network path
File 16	Link delay experienced by individual probe vehicles	Time taken to travel along a link – Link free flow travel time
Summary File	Average vehicle network delay	Sum of individual network travel times – Sum of network free flow travel times

4.4 VERTICAL QUEUING ANALYSIS DELAY MODELS

Current state-of-the-practice queuing models rely on two critical simplifying assumptions in computing the delay incurred by vehicles arriving at signalized intersections. The first assumption establishes the hypothesis that vehicles can decelerate and accelerate instantaneously. Figure 4.7 illustrates the effect of this assumption on projected vehicle trajectories and speed profiles. When analyzing the two diagrams, it can first be observed that the assumption of instantaneous acceleration and deceleration attempts to estimate the overall delay incurred by a vehicle by converting a portion of the deceleration and acceleration delay to stopped delay. Consequently, the resulting delay estimates cannot directly be categorized as stopped, deceleration or acceleration delay. The second assumption assumes that vehicles queue vertically, which means that vehicles travel the full length of an approach link before stopping. These two assumptions result in an underestimation of the delay and the maximum queue size as vehicles are assumed to arrive at the end of the queue later than they would in reality.

Deterministic queuing models assume a basic cumulative arrival and departure pattern, as illustrated in Figure 4.8. In the diagram, the arrival curve represents the number of vehicles that would have reached the intersection if traffic were not stopped by the signal operation. The departure curve, on the other hand, represents the number of vehicles that actually depart from the intersection. As a result, the vertical distance between the arrival and departure curves represent the number of vehicles that have not been able to cross the intersection and that have therefore joined the stop line queue, while the horizontal distance represents the time that a particular vehicle spends waiting in queue.

In Figure 4.8, three distinct periods can be identified regarding queue size evolution within a cycle length. The first period corresponds to the interval during which the departure curve is horizontal. This period corresponds to the portion of the cycle length during which the traffic signal is red and no traffic can cross the stop line, which results in a growth in the queue size. The second period corresponds to the first portion of the green phase during which the queue is being served and traffic leaves the intersection at the saturation flow rate. For undersaturated conditions the queue is always served before the traffic signal turns red again. In the last period, which only occurs in undersaturated operations, both the cumulative arrival and departure curves overlap, indicating that all arriving traffic is able to cross the stop line without incurring any delay, i.e., that the queue that formed during the previous red interval has completely dissipated.

In Figure 4.8 the total delay incurred by traffic within a cycle length can be estimated by calculating the area between the arrival and departure curves. To better understand how delay is computed from the queuing diagram, Figure 4.8 can be reconstructed to illustrate the queue size evolution within a single cycle length, as illustrated in Figure 4.9. In this diagram, it can be observed that the maximum queue size is assumed to occur immediately before the start of the effective green time. It is also observed that the time required to clear the queue is taken as a function of the difference between the rate at which vehicles arrive at the back of the queue and the rate at which they discharge across the stop line. Based on these observations, Equations 4.5

and 4.6 can be derived to calculate the total and average delays incurred by vehicles at an undersaturated fixed-time signalized approach.

$$D = \frac{r^2 \cdot q}{2 \cdot 3600} \cdot (1 + \frac{q}{S - q}) \tag{4.5}$$

$$d = \frac{r^2}{2 \cdot C} \cdot (1 + \frac{q}{S - q}) = 0.5\mathrm{C} \frac{\left(1 - \frac{\mathrm{g}}{\mathrm{C}}\right)^2}{\left(1 - X\frac{g}{C}\right)} \tag{4.6}$$

where:

D = total delay at signalized approach (seconds),

d = average delay per vehicle (seconds),

r = effective red interval duration (seconds),

g = effective green interval duration (seconds),

C = traffic signal cycle time (seconds),

S = saturation flow rate (vehicles/hour of green),

c = capacity (S×g/C) (vehicles/hour),

X = arrival to capacity ratio (unitless), and

q = arrival flow rate (vehicles/hour).

Since Equations 4.5 and 4.6 assume uniform vehicle arrivals, both equations only estimate uniform delay. It is also observed that both equations define the signal operation in terms of an effective red interval duration instead of an actual red duration. In both equations, the effective red interval constitutes the remainder of the cycle length.

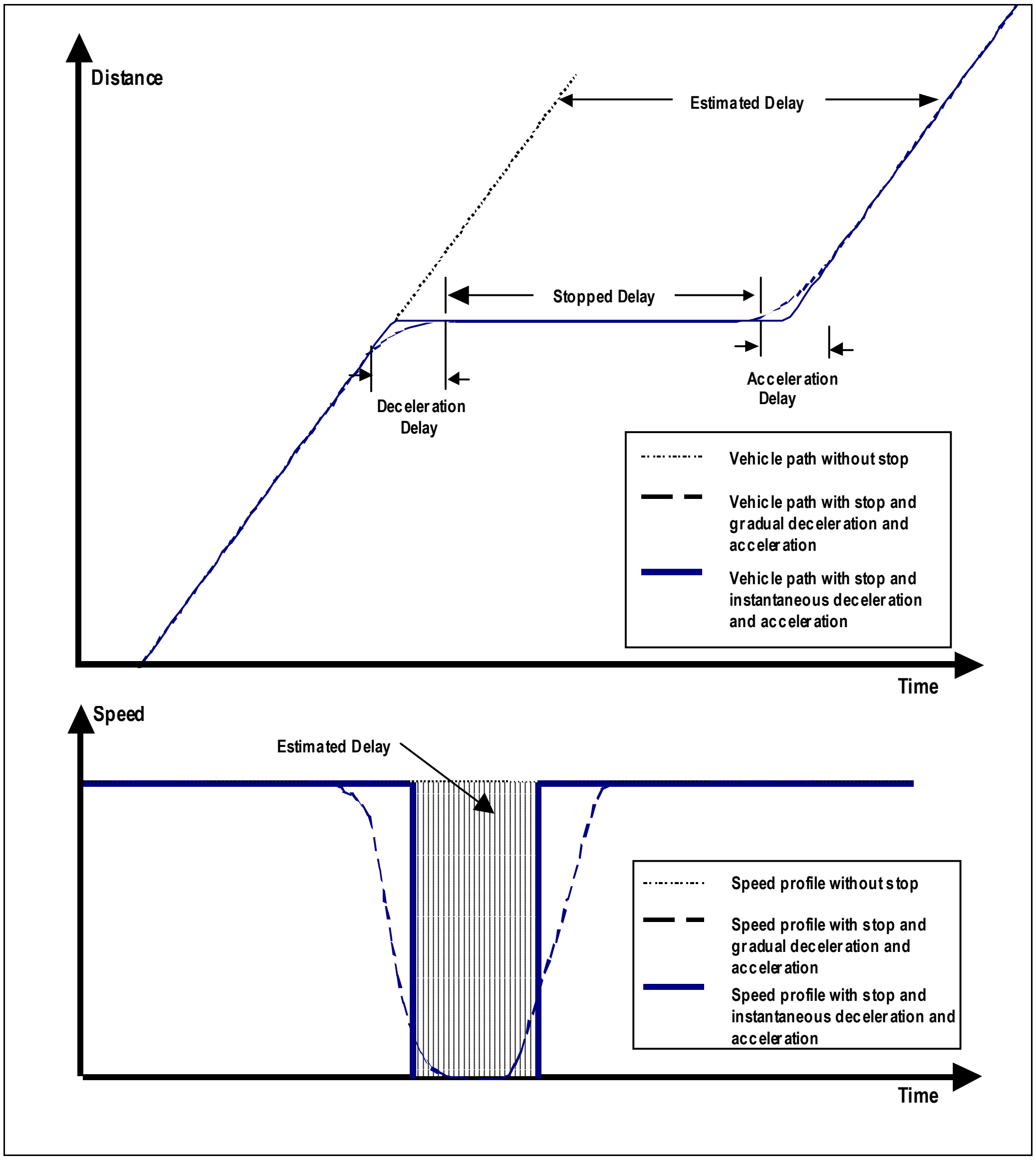

Figure 4.7: Instantaneous acceleration and deceleration speed-time profile

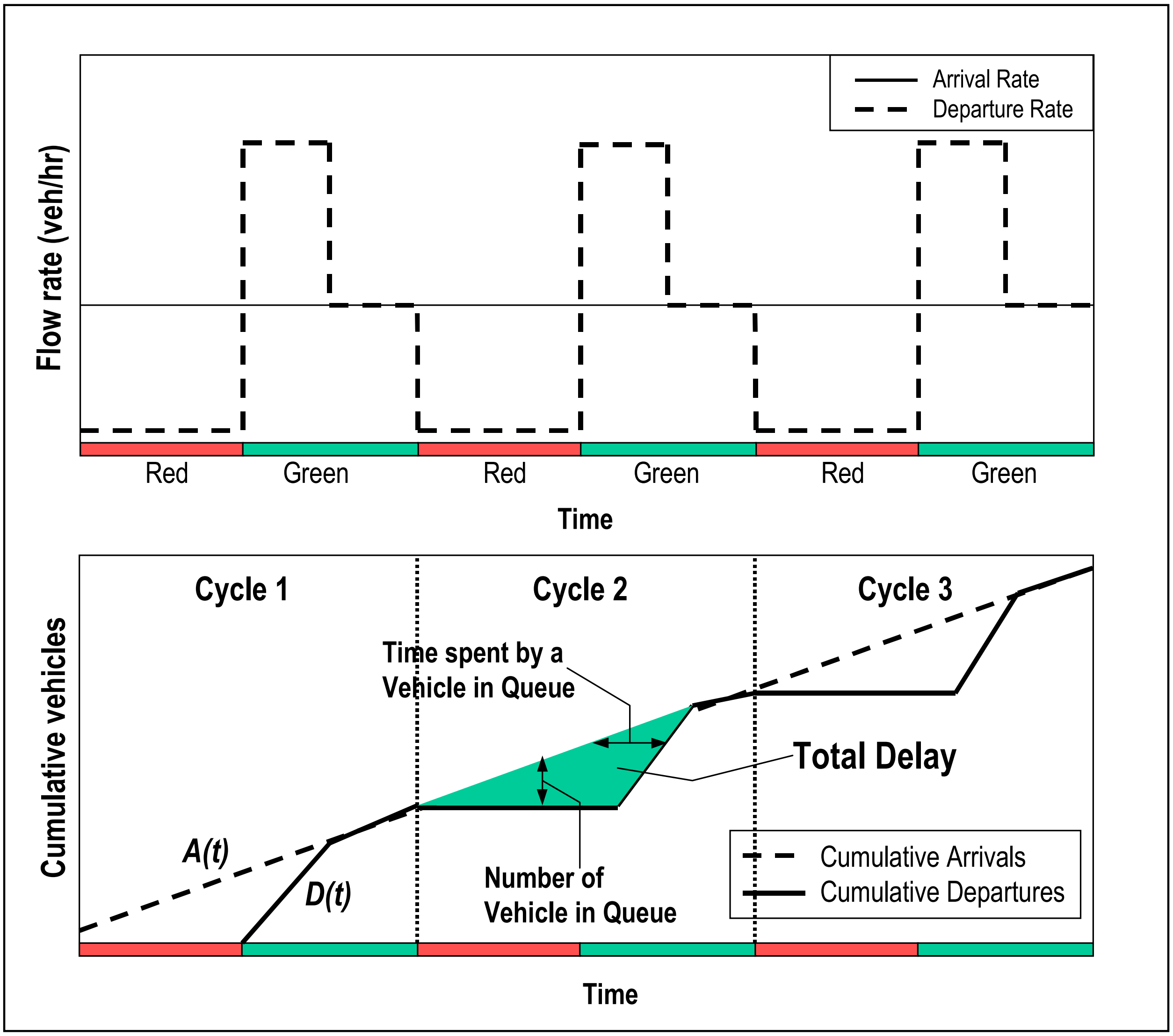

Figure 4.8: Idealized Cumulative Arrivals and Departures

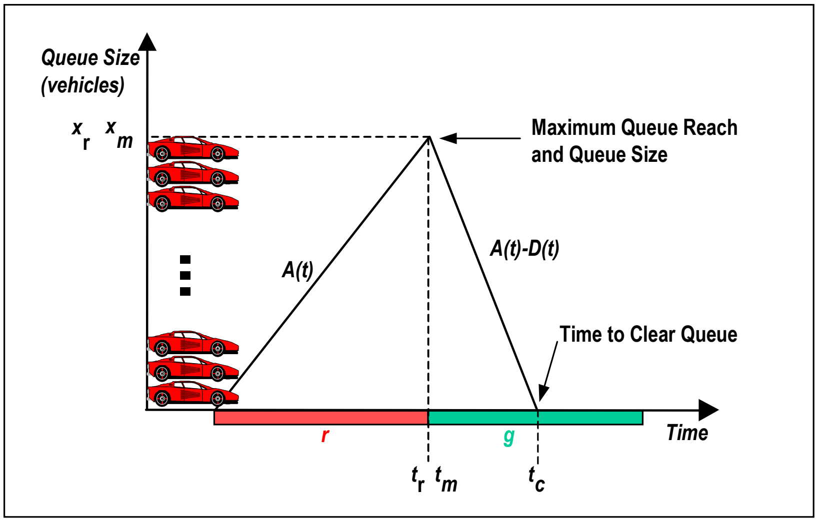

Figure 4.9: Queue Modeling under Deterministic Queuing Analysis

4.5 CAPACITY GUIDE MODELS

Numerous efforts have been devoted to the development of delay estimation models that would take into account both the deterministic and random aspects of traffic behavior. Following the work of Webster and Cobbe (1966), a number of stochastic delay models that attempt to account for these components have been developed using queuing analysis principles (Hurdle, 1984; Akcelik, 1988; Teply, 1989; Akcelik and Rouphail, 1993 among numerous publications). These models all share the same basic assumptions. First, they implicitly consider that vehicles can decelerate and accelerate instantaneously. Second, they all assume that vehicles queue vertically. Finally, it is usually assumed that the relation of delay to the arrival pattern is deterministic and that a Poisson distribution can describe the process by which vehicles arrive at an intersection. The HCM 1994, the Canadian Capacity Guide (CCG) 1995, and the HCM 2000, as these models are among the most widely used in North America, were used for comparison to assess the consistency. Those capacity guide models already described in Chapter 2

4.6 HORIZONTAL SHOCK WAVE DELAY MODELS

The flow of traffic is similar to the flow of a compressible fluid. The first successful attempts at demonstrating the consistency between the flow of traffic and the flow of a compressible fluid were made by Lighthill and Whitham (1955) and Richards (1956), who demonstrated the existence of traffic shock waves and proposed a first theory of one-dimensional waves that could be applied to the prediction of highway traffic flow behavior. The main postulate of their theory was that there exists some functional relation between traffic volume and traffic density and that this relation could be used to describe the speed at which a change in traffic flow characteristics propagates either upstream or downstream.

The shock wave model proposed by Lighthill, Whitham and Richards is computed using Equations 4.7 and 4.8. The first equation defines the relation between volume, density and speed that was developed using fluid dynamics theory and reflects the hypothesis that traffic can be considered as a compressible fluid, as illustrated in Figure 4.10. Using Equation 4.8, the second equation then describes the speed at which a change in traffic characteristics, or shock wave, propagates along a roadway.

$$q_i = k_i \cdot u_i \tag{4.7}$$

$$SW_{ij} = \frac{q_j - q_i}{k_j - k_i} \tag{4.8}$$

where:

SW_{ij} = speed of shock wave between zones i and j (meters/second),
q_j = traffic flow in zone i (vehicles/second),
k_i = traffic density in zone i (vehicles/meter),
u_i = traffic speed in zone i (meters/second).

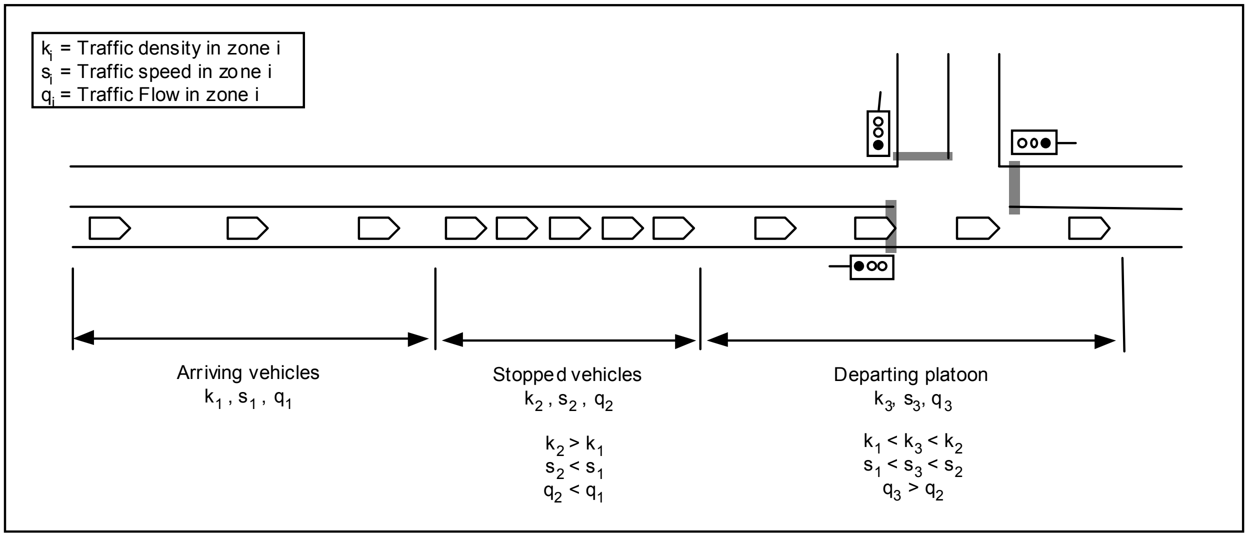

Figure 4.10: Traffic Flow Characteristics Upstream a Traffic Signal

Shock wave theory can be used to analyze flow at signalized intersections. For instance, Rorbech (1968) applied the shock wave theory of Lighthill and Whitham to investigate queue formation on intersection approaches at the beginning of red intervals. Stephanopoulos *et al.* (1979) further investigated the dynamics of queue formation and dissipation at isolated intersections using the flow conversation principle of Equation 2.13. In another example, Michalopoulos *et al.* (1980) analyzed traffic dynamics between signalized intersections and demonstrated the existence of shock waves caused by the traffic signal operation that periodically propagates downstream of an intersection. Michaloploulos *et al.* (1981a, 1981b) further developed a real-time control algorithm based on shock wave theory that minimizes total intersection delay at isolated intersections subject to constraints regarding maximum queue lengths on individual approaches.

The main difference between shock wave and queuing analysis models is in the way vehicles are assumed to queue at the intersection stop line. While queuing analysis assumes vertical queuing, shock wave analysis considers that vehicles are queued horizontally one behind each other, i.e., that each vehicle occupies a physical space. This treatment allows shock wave delay models to capture more realistic queuing behavior. Consider for example the diagram of Figure 4.11, which illustrates the formation and dissipation of queue of a vehicles through shock wave analysis. In the figure, it can first be observed that the shock wave theory still assumes instantaneous decelerations and accelerations. This is illustrated by the sharp angles along the vehicle paths. More importantly, it can be observed that while the maximum number of queued vehicles still occurs at the end of the red interval, the maximum reach of the back of queue is now correctly modeled to occur later, as it usually happens in reality.

In addition to queue formation and dissipation, Figure 4.11 illustrates the shock waves that are created by the traffic signal operation. The fist wave that is created, SW_I, is associated with the back of queue. This wave defines the boundary between incoming traffic and queued vehicles. It can further be observed that this wave corresponds to the line identifying the back of queue in the time-space diagram of Figure 4.2. Two other waves generated by the signal operation are produced when the signal turns green. The first wave, SW_R, moves downstream from the stop line and is associated with the front of the surge of vehicles that leave the intersection at saturation flow rate at the beginning of the green interval. The second wave, SW_R, moves upstream of the stop line and divides the vehicles stopped in the queue from those that have started to accelerate forward. The last wave, SW_N, defines the end of the platoon of vehicles leaving the intersection at saturation flow rate.

In Figure 4.11, the total travel time spent by all vehicles going through the intersection can be estimated using Equation 4.9. Since delay represents the added travel time caused by the traffic signal operation, the total delay incurred by traffic within one signal cycle at the intersection of Figure 4.11 can be estimated by comparing the total travel time in scenarios with and without traffic signals. The resulting calculations are summarized in Equation 4.10.

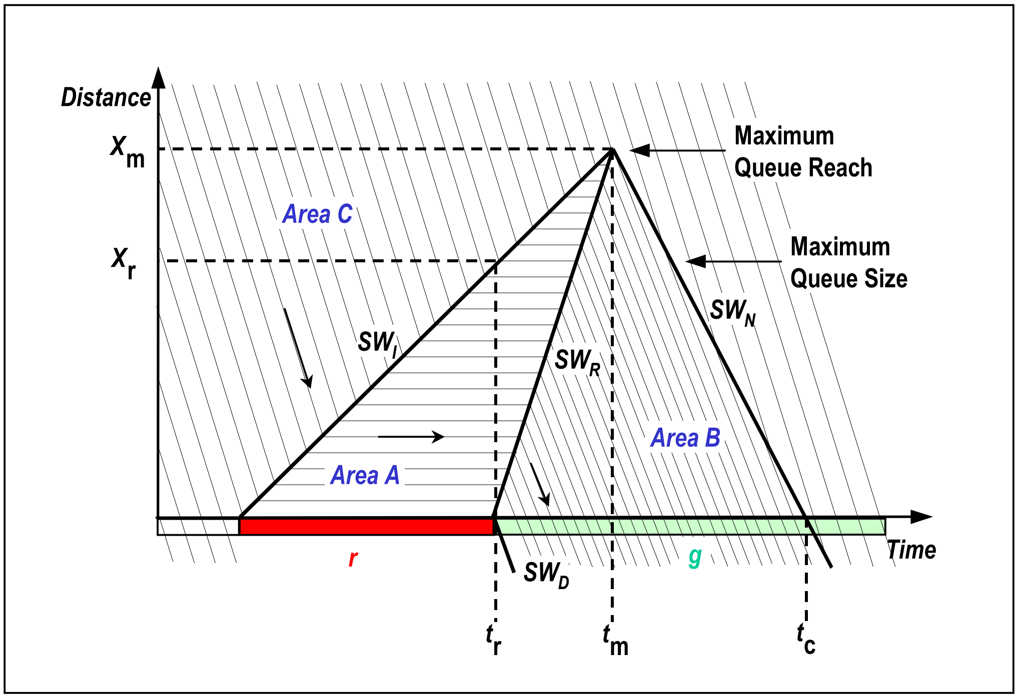

Figure 4.11: Graphical Illustration of Shock Wave Analysis

$$TT = \sum_{i=A,B,C} A_i k_i \quad (4.9)$$

$$D = TT_{with\ signals} - TT_{no\ signals} = \sum_{i=A,B,C} \left[A_i (k_i - k_C) \right] \quad (4.10)$$

where:

TT = travel time,

A_i = surface area of area i,

k_i = traffic density in area i.

Based on Equation 4.10, Equations 4.11 and 4.12 can be derived for determining the total delay caused by the signal operation and the average delay experienced by vehicles going through the intersection.

$$D = \frac{x_m}{2} \cdot [r \cdot (k_j - k_a) + (t_m + t_c) \cdot (k_d - k_a)] \quad (4.11)$$

$$d = \frac{x_m \cdot 3600}{2 \cdot q \cdot C} \cdot [r \cdot (k_j - k_a) + (t_m + t_c) \cdot (k_d - k_a)] \quad (4.12)$$

where:

x_m = maximum spatial extent of queue (kilometers),

= $t_m \cdot SW_R$

= $\dfrac{q \cdot r \cdot S}{3600 \cdot [q \cdot (k_d - k_j) + S \cdot (k_j - k_a)]}$,

t_m = time for maximum extent of queue (seconds),

= $\dfrac{-q \cdot r \cdot (k_d - k_j)}{3600 \cdot [q \cdot (k_d - k_j) + S \cdot (k_j - k_a)]}$,

t_c = time to clear queue (seconds),

= x_m / SW_N,

r = red interval (seconds),

k_j = jam density (vehicles/kilometer),

k_a = approach density (vehicles/kilometer),

k_d = discharge density (vehicles/kilometers),

q = arrival rate (vehicles/hour),

C = cycle length (seconds).

Similar to the vertical queue analysis model, the horizontal shock wave analysis model defined by Equations 4.11 and 4.12 only estimates uniform delay. There is no account for the incremental delay caused by the randomness of traffic flow.

4.7 TEST SCENARIOS

In order to evaluate the consistency of delay estimates among the various analytical and INTEGRATION simulation delay models presented in the chapter, delay evaluations were carried out for the example of Figure 4.12 using the INTEGRATION traffic simulation software and the analytical models of Sections 4.4, 4.5 and 4.6. For ease of comparison, the example features a single intersection approach leading traffic to a signalized intersection operated in fixed-time with 30-second effective green and red intervals. The example further includes a one-kilometer exit link to allow INTEGRATION to capture the delay incurred by vehicles while accelerating away from the intersection. In order to ensure consistency with the analytical procedures a constant demand was loaded a time 13 minutes in order to ensure that vehicles arrived at the intersection after 15 minutes of simulation. The demand was loaded for 15 minutes followed by a 15 minutes period of no demand, in order to clear any remaining queues. Figure 4.13 illustrates the demand that was simulated. This demand scenario is consistent with the assumptions of the Canadian Capacity Guide (1995) and the Highway Capacity Manual (1994, 2000).

Using the above settings, two sets of ten test scenarios were developed to estimate delays in a range of traffic conditions. The first set considers uniform arrivals, while the second set adds randomness to the arrival process. Within each set, all scenarios are identical, except for the vehicle arrival rate. Arrival rates are varied to produce v/c ratio on the intersection approach varying from 0.1 to 1.0. No congested scenarios were considered in this study, as oversaturation is the subject of a forthcoming chapter.

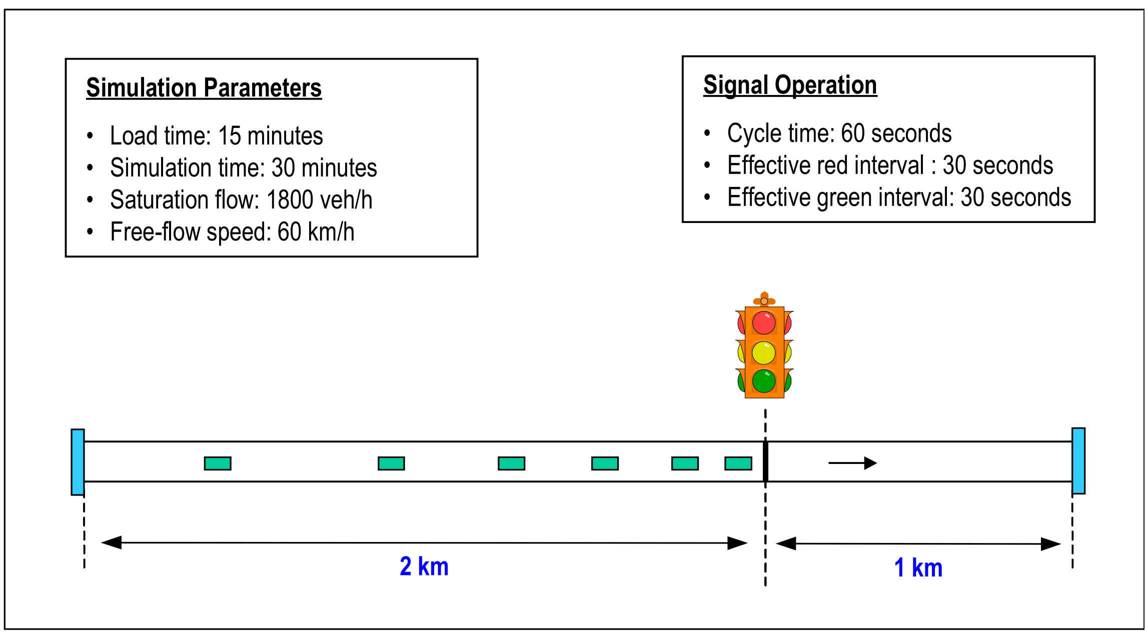

Figure 4.12: Delay Evaluation Scenario

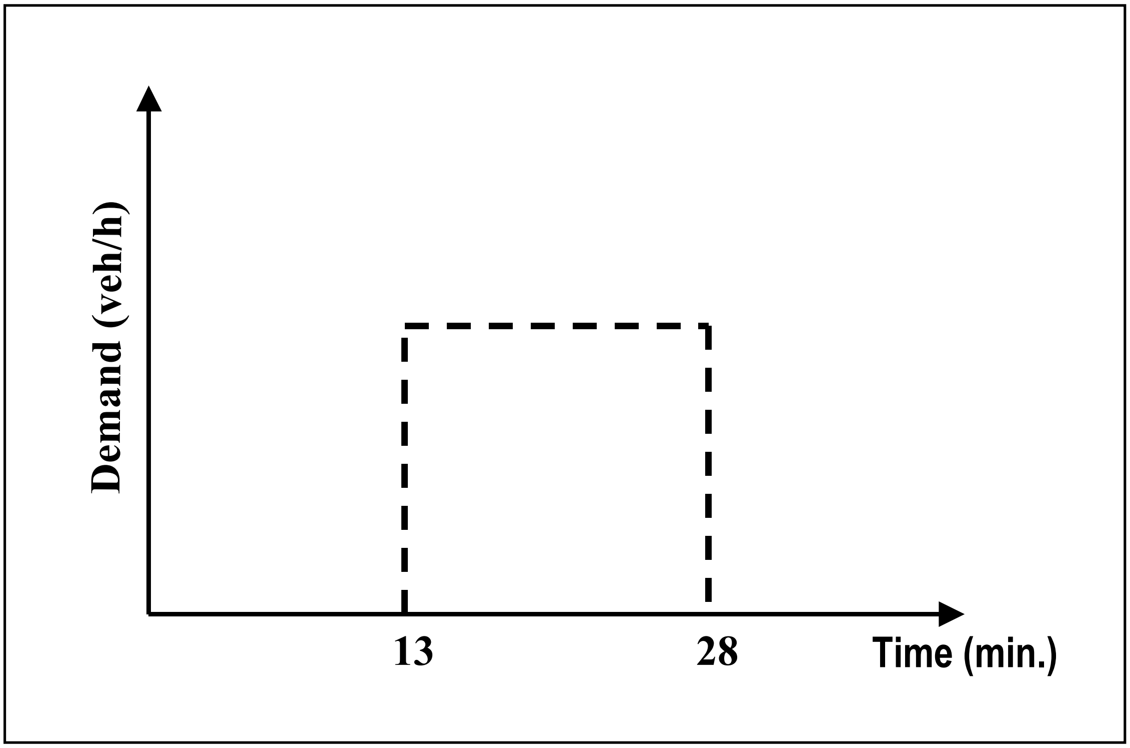

Figure 4.13: Demand Graph for the Simulation Model

4.8 TEST RESULTS

4.8.1 CONSISTENCY OF OVERALL DELAYS ESTIMATES UNDER UNIFORM ARRIVALS

Table 4.2 and Figure 4.14 provides the results of the delay estimations that were carried out for the scenarios considering uniform arrivals only. For this set of scenarios, Equations 4.6 and 4.12 were used to calculate the overall delay with the vertical queue analysis and horizontal shock wave models, respectively. For the three capacity guide delay models, only the first term of Equations 2.4, 2.8 and 2.12 in Chapter 2 was used, as the remaining portions of these equations were developed to specifically take into account the randomness of vehicle arrivals and the probability of temporary signal cycle oversaturation due to this randomness. The results predicted by Equation 2.4 were also multiplied by 1.3 to convert the estimated stopped delay into overall delay.

The results of Table 4.2 indicate that all the delay models produce similar delay estimates when applied to analyze the delay caused to uniform traffic flows at undersaturated pretimed signalized intersections. In this case, the similarly of the results between the theoretical queuing analysis and the three capacity guide models was expected since the later models were derived from queuing analysis. For the INTEGRATION delay models, it is observed that there is a general agreement with the other models despite some added variability and the fact that the simulation models produces slightly higher delay estimates.

Table 4.2: Overall Delay Estimates under Uniform Arrivals

Delay Model	v/c Ratio									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Vertical Queuing Analysis	7.89	8.33	8.82	9.38	10.00	10.71	11.54	12.50	13.64	15.00
Horizontal Shock Wave Analysis	7.89	8.33	8.82	9.38	10.00	10.71	11.54	12.50	13.64	15.00
HCM 1994	7.89	8.33	8.82	9.38	10.00	10.71	11.54	12.50	13.64	15.00
HCM 2000	7.89	8.33	8.82	9.38	10.00	10.71	11.54	12.50	13.64	15.00
CCG 1995	7.89	8.33	8.82	9.38	10.00	10.71	11.54	12.50	13.64	15.00
INTEGRATION File 10	9.51	8.09	10.07	9.41	10.98	11.08	12.93	12.58	15.19	15.12
INTEGRATION File 11	9.40	7.90	10.00	9.30	10.90	10.90	12.80	12.50	15.00	15.00
INTEGRATION File 12	9.40	7.90	10.00	9.30	10.90	10.90	12.80	12.50	15.00	15.00
INTEGRATION File 15	9.60	8.21	10.19	9.51	11.09	11.17	13.04	12.72	15.32	15.27
INTEGRATION File 16	9.60	8.21	10.18	9.49	11.09	11.17	13.05	12.73	15.33	15.27
INTEGRATION Summary File	9.50	8.04	10.10	9.38	11.00	11.10	12.95	12.58	15.22	15.11

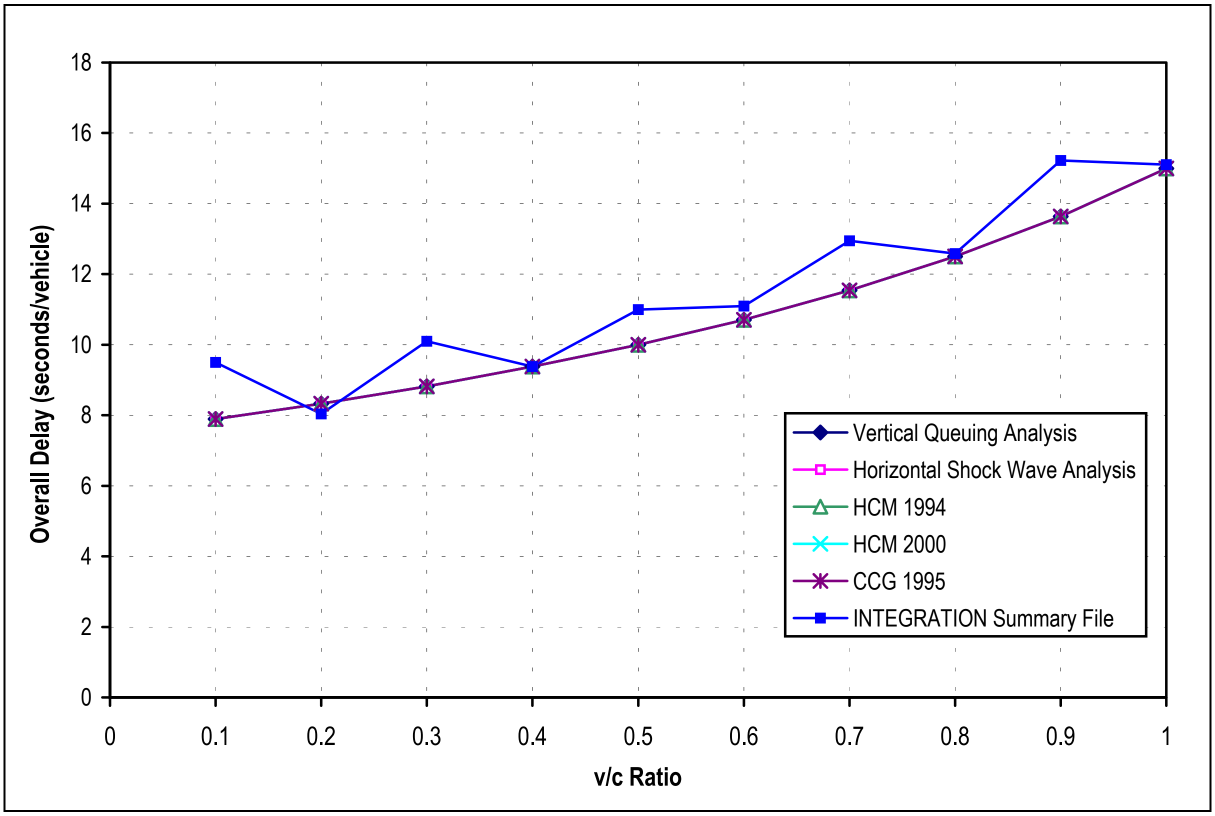

Figure 4.14: Overall Delay Estimates under Uniform Arrivals

The variability of delay estimates with INTEGRATION is attributed to the fact that the simulation model only allows integer numbers of vehicles to go across an intersection while the analytical delay models consider average hourly flow rates that often yield fractional average vehicle arrivals within a single signal cycle. For example, an average vehicle arrival rate of 810 vehicles per hour translates into one arrival every 4.44 seconds and in 13.5 arrivals in every 60-second cycle. In this case, an assumed average arrival rate of 13.5 vehicles per cycle would thus be used by the analytical models to calculate delays over a 60-second cycle, while the INTEGRATION model would average delays from simulated cycles considering 13 arrivals and cycles considering 14 arrivals.

4.8.2 Sensitivity of Delay Uniform Estimates to Arrival Patterns

At an intersection, vehicles do not always arrive at exactly the same time relative to the beginning of the green and red intervals. While the overall number of arriving vehicles may be the same from one cycle to another, there could be substantial variability in the cyclic arrival patterns. As an example, Table 4.3 computes the delay associated with three arriving flows for the example of Figure 4.12. In all three cases, the vehicles arrive at the intersection with a 5-second headway and leave at saturation during the effective green interval with a 2-second headway. The only difference between the arriving flows in the first two signal cycle is the offset in the time at which each vehicle is assumed to arrive relative to the start of the effective green interval. In the second cycle, all the vehicles reach the intersection stop line one second later than their corresponding vehicle in the first cycle. In the third cycle, the vehicle arrivals are offset by four seconds relative to arrivals in the first cycle. This offset causes the last arriving vehicle to cross the intersection stop line one second before the end of the effective green interval. As it can be observed a one-second offset in arrival times is sufficient in this case to cause a 6.1-percent reduction in the estimated delays, while a four-second offset further results in a 23.6-percent total delay reduction.

Table 4.3: Sensitivity of Delay Estimates to Various Arrival Patterns

Vehicle	Cycle 1			Cycle 2			Cycle 3		
	Arrival	Departure	Delay	Arrival	Departure	Delay	Arrival	Departure	Delay
1	0	30	30	1	30	29	4	30	26
2	5	32	27	6	32	26	9	32	23
3	10	34	24	11	34	23	14	34	20
4	15	36	21	16	36	20	19	36	17
5	20	38	18	21	38	17	24	38	14
6	25	40	15	26	40	14	29	40	11
7	30	42	12	31	42	11	34	42	8
8	35	44	9	36	44	8	39	44	5
9	40	46	6	41	46	5	44	46	2
10	45	48	3	46	48	2	49	49	0
11	50	50	0	51	51	0	54	54	0
12	55	55	0	56	56	0	59	59	0
	Total Delay		165	Total Delay		155	Total Delay		126
	Average Delay		13.75	Average Delay		12.92	Average Delay		10.50

If delay calculations were made using the vertical queue analysis model of Equation 4.6, the estimated average delay would be 12.5 seconds for all cycles. An identical estimate would also be obtained with the horizontal shock wave delay model. In both cases, there would be no sensitivity to the arrival pattern as there is no change in the variables used to compute the delays. While the time at which vehicles arrive at the intersection stop line is varying from one cycle to the other, the arrival rate, departure rate, traffic densities and signal timings remain the same.

The above results clearly demonstrate the sensitivity of delay estimates to the assumed arrival patterns. They also demonstrate the potential variability of delay estimates between delay models that explicitly consider vehicle arrival patterns in their calculation and models that do not. Such difference accounts for most of the differences in delay estimates observed in Figure 14 between the delay estimated by micro-simulation with INTEGRATION and the analytical delay models.

4.8.3 Comparison of Acceleration and Stopped Delay Estimates

The consistency of delay estimates between some of the analytical delays models and the micro-simulation delay estimates from INTEGRATION can be further assessed by comparing the ratio of stopped delay to overall delay obtained with each model. Among the models evaluated in this study, such comparison is only possible between INTEGRATION and the CCG 1995 model. Comparison is not possible with the HCM 2000 as there is no relation of stopped to overall delay provided in the manual. While a fixed ratio of stopped to overall delay is provided in HCM 1994, this ratio cannot be used for analyses with the HCM 2000 delay model as this model is totally different than the one used in the HCM 1994. Comparison are also not possible with the vertical queue analysis model and the horizontal shock wave analysis models as both models were developed assuming instantaneous accelerations and decelerations.

The ratios of stopped to overall delay assumed in the CCG 1995 delay model were given in Table 2.3 of Chapter 2. To estimate the same ratio for the simulated delays in INTEGRATION, manual calculations must be made. As an example, consider Figure 4.15, which illustrates the delay obtained under uniform arrivals for a v/c ratio of 0.80. By summing the delays incurred by vehicles while traveling at a speed lower than that of a pedestrian (5 km/h), which is the criterion defining stopped delay in the CCG 1995, it is found that 54.9 percent of the total simulated delay could be considered as stopped delay. The ratio that has been estimated compares very favorably with the 56-percent stopped to overall delay ratio assumed in the CCG 1995 delay model for similar traffic conditions.

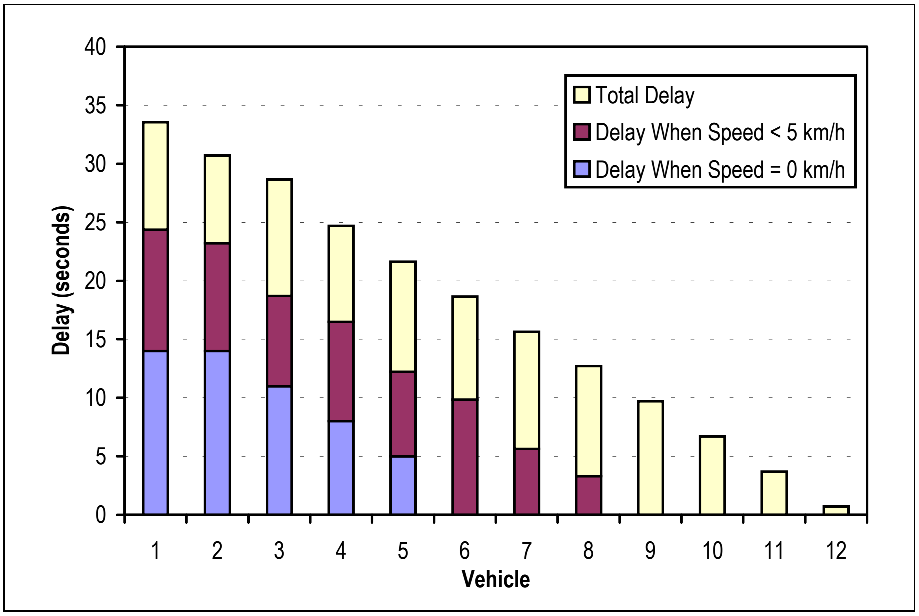

Figure 4.15: Stopped and Overall Delay for Arriving Vehicles in Typical Signal Cycle

4.8.4 Overall Delays Estimates with Random Arrivals

Table 4.4 and Figure 4.16 reports the results of the delay estimations that were carried out for the example of Figure 4.12 using the random arrival scenarios. For this set of scenarios, Equations 4.6 and 4.12 were again used to calculate the overall delay predicted by the vertical queue and horizontal shock wave analysis models. However, contrary to the previous analysis, all the terms of Equations 2.4, 2.8 and 2.12 were used in this case to calculate the overall delays predicted by the three capacity guide models. Similarly to the previous analysis, the stopped delay estimates produced by the HCM 1994 model were again multiplied by 1.3 to obtain the corresponding overall delay and allow a direct comparison of its delay estimates with other models. In addition, the delays reported for the INTEGRATION simulation model are in this case the average of the delays reported by ten replications of the test example. Replications were made for this set of scenarios to account for the stochastic variability of the simulation processes within INTEGRATION.

The results of Table 4.4 and Figure 4.16 indicate that there is a general agreement between the simulation delays from INTEGRATION and the delays predicted by the HCM 1994, CCG 1995 and HCM 2000 models when these models are applied to the analysis of undersaturated flows at signalized intersections. As it can be observed, all four models predict small increases in delay with increasing demands at low v/c ratios and significantly larger increases as traffic approaches saturation (v/c ratio of 1.0). Similarly, it can be observed that there is a general disagreement between the delays predicted by the four previously mentioned models and the delays predicted by the vertical queue analysis and the horizontal shock wave analysis models, as these two models predict much smaller delay increases as traffic approaches saturation. This disagreement was expected, as both the vertical queue analysis and horizontal shock wave analysis models only assume uniform vehicle arrivals. Since both models ignore the potential for additional delays that arises from the probability of having temporary oversaturation caused by surges of arriving vehicles, it is normal for these two theoretical models to predict lower delays, especially at high v/c ratios where the impact of arrival surges is more prominent.

Table 4.4: Overall Delay Estimates under Stochastic Arrivals

Delay Model	v/c Ratio									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Vertical Queuing Analysis	7.89	8.33	8.82	9.38	10.00	10.71	11.54	12.50	13.64	15.00
Horizontal Shock Wave Analysis	7.89	8.33	8.82	9.38	10.00	10.71	11.54	12.50	13.64	15.00
HCM 1994	7.89	8.35	8.90	9.59	10.50	11.78	13.75	17.23	24.79	44.99
HCM 2000	8.12	8.83	9.68	10.70	11.98	13.67	16.05	19.89	27.42	45.00
CCG 1995	8.12	8.83	9.68	10.70	11.98	13.67	16.05	19.89	27.42	45.00
INTEGRATION File 10	10.18	9.06	10.02	11.53	12.51	12.75	13.72	17.37	22.07	37.84
INTEGRATION File 11	10.10	9.00	9.90	11.40	12.40	12.70	13.60	17.20	21.00	37.70
INTEGRATION File 12	10.01	9.85	9.67	10.49	12.73	12.33	13.61	18.23	25.93	48.62
INTEGRATION File 15	10.30	9.17	10.13	11.63	12.62	12.86	13.84	17.49	22.21	37.99
INTEGRATION File 16	10.30	9.16	10.12	11.63	12.62	12.86	13.85	17.49	22.16	38.00
INTEGRATION Summary File	10.24	9.17	10.16	11.72	12.77	13.09	14.15	17.94	22.81	38.78

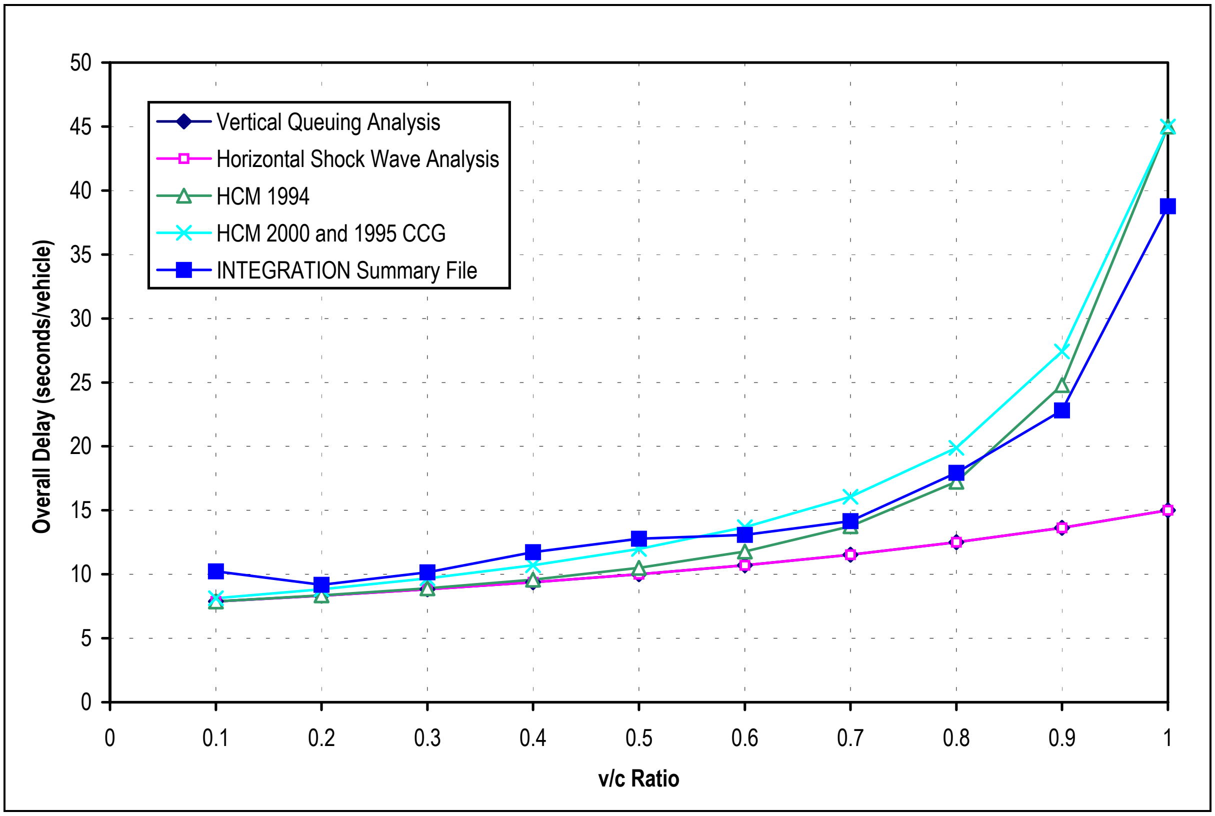

Figure 4.16: Overall Delay Estimates under Stochastic Arrivals

A more detailed analysis of the diagram of Figure 4.16 further reveals that the delays that are predicted by the HCM 2000 model are higher than those predicted by the 1994 model in the undersaturated domain. These higher delays reflect the changes that were made to the second term of Equation 2.12 in the HCM 2000 model to allow it to produce delay estimates that are asymptotic to the deterministic oversaturation delay model of Figure 4.6. The fact that the 1994 HCM model was not asymptotic to this model was considered as a weakness by many researchers (Akcelik, 1988; McShane and Roess, 1990; Fambro *et al.*, 1997) and was one of the major arguments that lead to its replacement. While the changes introduced in the model resulted in the prediction of more suitable lower delays in the oversaturation domain (Engelbrecht, 1997), they also caused the model to predict higher delays in the undersaturation domain. These changes were caused by the need to define a single continuous mathematical expression that could be used to estimate the added delay due to the randomness of vehicle arrivals in both undersaturated and oversaturated traffic conditions.

Figure 4.16 also indicates that the delays predicted by INTEGRATION are generally consistent to those predicted by the CCG 1995 and HCM 2000 models. This conclusion is further supported by the diagrams of Figure 4.17, which superimpose the delays predicted by the two capacity guide models to the those obtained with each of the ten replications that were conducted with the INTEGRATION model. As it can be observed in Figure 4.17, the both delays estimated by the CCG 1995 and the HCM 2000 models fall within the simulated results under all v/c ratios. This result thus clearly indicates that there is a consistency in this example between the simulated and analytical delays. Also, the diagram of Figure 4.18, which conducted the same procedure for two-lane delays, derives the same conclusion.

In Figures 4.16, 4.17 and 4.18, the non-uniform trend of increasing delays with increasing v/c ratios for the INTEGRATION simulation model is in part attributable to the stochastic nature of the simulation processes within traffic simulation model. Similar to the uniform scenario results, some variability is also attributable to the fact that the INTEGRATION results are sensitive to the actual vehicle arrival times at the test intersection while the delay estimates from the capacity guide models are not. As it was shown in Table 4.3, a small offset in vehicle arrival times could have a significant impact on the simulated delays. In the case of the capacity guide delays, this impact does not exists as the delays are calculated within these models on the basis of average hourly arrival and departure rates. Finally, similar to the uniform scenarios, another source of variability is the fact that the capacity guide delay models are not constrained to consider only integer number of arrivals and departures within a signal cycle while INTEGRATION has to respect such a constraint.

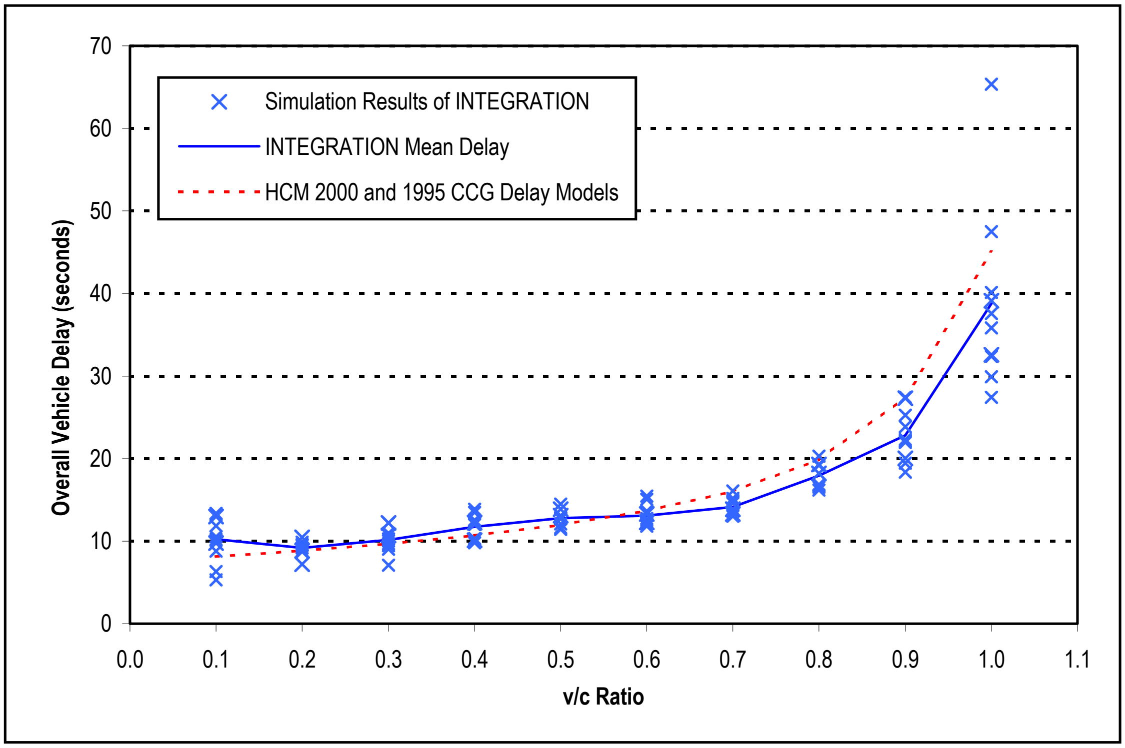

Figure 4.17: Simulation Results of INTEGRATION for Single Lane

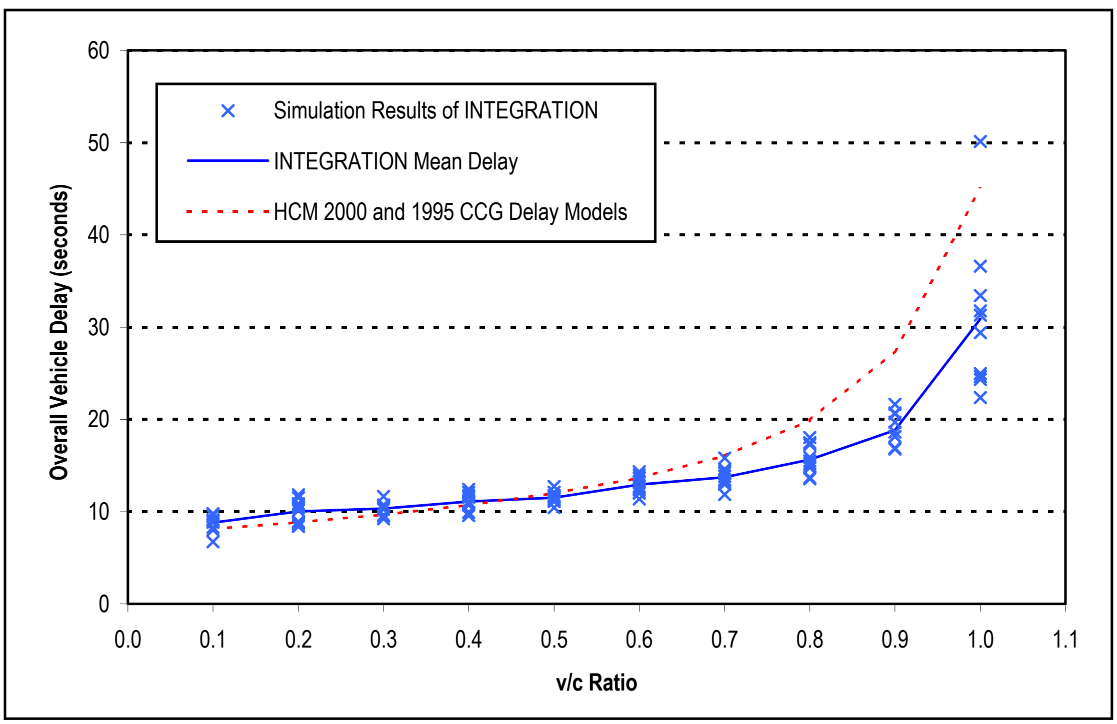

Figure 4.18: Simulation Results of INTEGRATION for Two-Lane

4.9 SUMMARY AND CONCLUSIONS

This chapter compared the delay estimates predicted by analytical and simulation models for the case of an undersaturated signalized intersection. The models that were compared are a theoretical vertical queue analysis model, the queue-based models used in the HCM 1994, the CCG 1995 and the HCM 2000, a theoretical shock wave analysis model, and six delay estimates from the INTEGRATION microscopic traffic simulation software.

The delay estimates predicted by each model under uniform and random arrival scenarios were compared to assess their consistency. For the uniform arrival scenarios, it is found that all the analytical models produce identical results and that the INTEGRATION model produces slightly higher results that follow the same general trend as the delay estimates from the analytical models. The difference in delay estimates between the analytical models and the simulation results is explained by the sensitivity of delay estimates from microscopic traffic simulation models to assumed traffic arrival patterns. Another source of variation was associated to the modeling of cyclic vehicle arrivals. While analytical delay models compute delays over one signal cycle using average arrival rates that may not correspond to an integer number of arrivals per cycle, simulation models can only simulate integer number of arrivals and departures per cycle and must therefore estimate delays by averaging traffic conditions over a certain number of cycles. Despite these differences, it was concluded that there is general consistency between the various delay models considered in this chapter for the uniform arrival scenarios.

A similar conclusion for the simulated delays predicted by the INTEGRATION model and delays predicted by the CCG 1995 and HCM 2000 was also reached for the scenarios considering random arrivals. In this case, the evaluation results demonstrate that the delay estimated by the CCG 1995 and the HCM 2000 models fall within the confidence interval of the delay obtained by simulation using INTEGRATION. The results also showed that the HCM 2000 model generally predicted higher delays than the HCM 1994 model when applied to the analysis of undersaturated traffic flows. At last, the results clearly indicated the inability of the two theoretical traditional delay models based on vertical queue analysis and horizontal shock wave analysis to accurately estimate delays when considering random arrivals.