TRAFFIC FLOW AT SIGNALIZED INTERSECTIONS

BY NAGUI ROUPHAIL[15]
ANDRZEJ TARKO[16]
JING LI[17]

[15] Professor, Civil Engineering Department, North Carolina State University, Box 7908, Raleigh, NC 276-7908

[16] Assistant Professor, Purdue University, West LaFayette, IN 47907

[17] Principal, TransSmart Technologies, Inc., Madison, WI 53705

Chapter 9 - Frequently used Symbols

$$I = \frac{variance\ of\ the\ number\ of\ arrivals\ per\ cycle}{mean\ number\ of\ arrivals\ per\ cycle}$$

I_i	=	cumulative lost time for phase *i* (sec)
L	=	total lost time in cycle (sec)
q	=	$A(t)$ = cumulative number of arrivals from beginning of cycle starts until *t*,
B	=	index of dispersion for the departure process,

$$B = \frac{variance\ of\ number\ of\ departures\ during\ cycle}{mean\ number\ of\ departures\ during\ cycle}$$

c	=	cycle length (sec)
C	=	capacity rate (veh/sec, or veh/cycle, or veh/h)
d	=	average delay (sec)
d_1	=	average uniform delay (sec)
d_2	=	average overflow delay (sec)
$D(t)$	=	number of departures after the cycle starts until time *t* (veh)
e_g	=	green extension time beyond the time to clear a queue (sec)
g	=	effective green time (sec)
G	=	displayed green time (sec)
h	=	time headway (sec)
i	=	index of dispersion for the arrival process
q	=	arrival flow rate (veh/sec)
Q_0	=	expected overflow queue length (veh)
$Q(t)$	=	queue length at time *t* (veh)
r	=	effective red time (sec)
R	=	displayed red time (sec)
S	=	departure (saturation) flow rate from queue during effective green (veh/sec)
t	=	time
T	=	duration of analysis period in time dependent delay models
U	=	actuated controller unit extension time (sec)
$Var(.)$	=	variance of (.)
W_i	=	total waiting time of all vehicles during some period of time *i*
x	=	degree of saturation, $x = (q/S) / (g/c)$, or $x = q/C$
y	=	flow ratio, $y = q/S$
Y	=	yellow (or clearance) time (sec)
Δ	=	minimum headway

9. TRAFFIC FLOW AT SIGNALIZED INTERSECTIONS

9.1 Introduction

The theory of traffic signals focuses on the estimation of delays and queue lengths that result from the adoption of a signal control strategy at individual intersections, as well as on a sequence of intersections. Traffic delays and queues are principal performance measures that enter into the determination of intersection level of service (LOS), in the evaluation of the adequacy of lane lengths, and in the estimation of fuel consumption and emissions. The following material emphasizes the theory of *descriptive models of traffic flow,* as opposed to prescriptive (i.e. signal timing) models. The rationale for concentrating on descriptive models is that a better understanding of the interaction between demand (i.e. arrival pattern) and supply (i.e. signal indications and types) at traffic signals is a prerequisite to the formulation of optimal signal control strategies. Performance estimation is based on assumptions regarding the characterization of the traffic arrival and service processes. In general, currently used delay models at intersections are described in terms of a deterministic and stochastic component to reflect both the fluid and random properties of traffic flow.

The deterministic component of traffic is founded on the fluid theory of traffic in which demand and service are treated as continuous variables described by flow rates which vary over the time and space domain. A complete treatment of the fluid theory application to traffic signals has been presented in Chapter 5 of the monograph.

The stochastic component of delays is founded on steady-state queuing theory which defines the traffic arrival and service time distributions. Appropriate queuing models are then used to express the resulting distribution of the performance measures. The theory of unsignalized intersections, discussed in Chapter 8 of this monograph, is representative of a purely stochastic approach to determining traffic performance.

Models which incorporate both deterministic (often called uniform) and stochastic (random or overflow) components of traffic performance are very appealing in the area of traffic signals since they can be applied to a wide range of traffic intensities, as well as to various types of signal control. They are approximations of the more theoretically rigorous models, in which delay terms that are numerically inconsequential to the final result have been dropped. Because of their simplicity, they have received greater attention since the pioneering work by Webster (1958) and have been incorporated in many intersection control and analysis tools throughout the world.

This chapter traces the evolution of delay and queue length models for traffic signals. Chronologically speaking, early modeling efforts in this area focused on the adaptation of steady-state queuing theory to estimate the random component of delays and queues at intersections. This approach was valid so long as the average flow rate did not exceed the average capacity rate. In this case, stochastic equilibrium is achieved and expectations of queues and delays are finite and therefore can be estimated by the theory. Depending on the assumptions regarding the distribution of traffic arrivals and departures, a plethora of steady-state queuing models were developed in the literature. These are described in Section 9.3 of this chapter.

As traffic flow rate approaches or exceeds the capacity rate, at least for a finite period of time, the steady-state models assumptions are violated since a state of stochastic equilibrium cannot be achieved. In response to the need for improved estimation of traffic performance in both under and oversaturated conditions, and the lack of a theoretically rigorous approach to the problem, other methods were pursued. A prime example is the time-dependent approach originally conceived by Whiting (unpublished) and further developed by Kimber and Hollis (1979). The time-dependent approach has been adopted in many capacity guides in the U.S., Europe and Australia. Because it is currently in wide use, it is discussed in some detail in Section 9.4 of this chapter.

Another limitation of the steady-state queuing approach is the assumption of certain types of arrival processes (e.g Binomial, Poisson, Compound Poisson) at the signal. While valid in the case of an isolated signal, this assumption does not reflect the impact of adjacent signals and control which may alter the pattern and number of arrivals at a downstream signal. Therefore performance in a system of signals will differ considerably from that at an isolated signal. For example, signal coordination will tend to reduce delays and stops since the arrival process will be different in the red and green portions of the phase. The benefits of coordination are somewhat subdued due to the dispersion of platoons between signals. Further, critical signals in a system could have a metering effect on traffic which proceeds downstream. This metering reflects the finite capacity of the critical intersection which tends to truncate the arrival distribution at the next signal. Obviously, this phenomenon has profound implications on signal performance as well, particularly if the critical signal is oversaturated. The impact of upstream signals is treated in Section 9.5 of this chapter.

With the proliferation of traffic-responsive signal control technology, a treatise on signal theory would not be complete without reference to their impact on signal performance. The manner in which these controls affect performance is quite diverse and therefore difficult to model in a generalized fashion. In this chapter, basic methodological approaches and concepts are introduced and discussed in Section 9.6. A complete survey of adaptive signal theory is beyond the scope of this document.

9.2 Basic Concepts of Delay Models at Isolated Signals

As stated earlier, delay models contain both deterministic and stochastic components of traffic performance. The deterministic component is estimated according to the following assumptions: a) a zero initial queue at the start of the green phase, b) a uniform arrival pattern at the arrival flow rate (q) throughout the cycle c) a uniform departure pattern at the saturation flow rate (S) while a queue is present, and at the arrival rate when the queue vanishes, and d) arrivals do not exceed the signal capacity, defined as the product of the approach saturation flow rate (S) and its effective green to cycle ratio (g/c). The effective green time is that portion of green where flows are sustained at the saturation flow rate level. It is typically calculated at the displayed green time minus an initial start-up lost time (2-3 seconds) plus an end gain during the clearance interval (2-4 seconds depending on the length of the clearance phase).

A simple diagram describing the delay process in shown in Figure 9.1. The queue profile resulting from this application is shown in Figure 9.2. The area under the queue profile diagram represents the total (deterministic) cyclic delay. Several

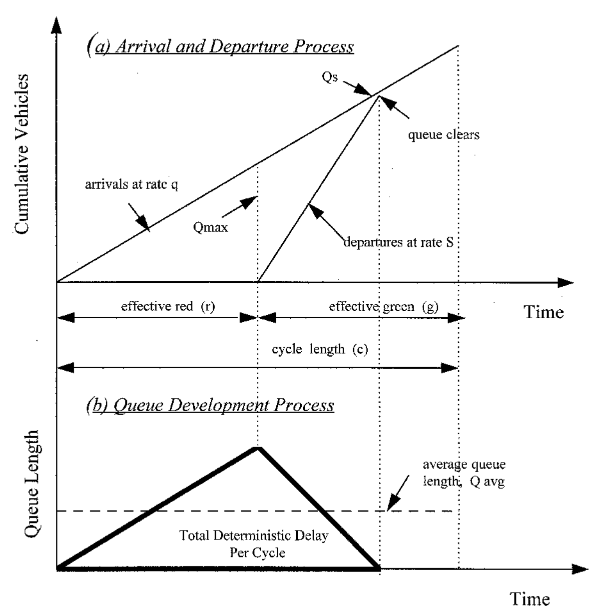

Figure 9.1
Deterministic Component of Delay Models.

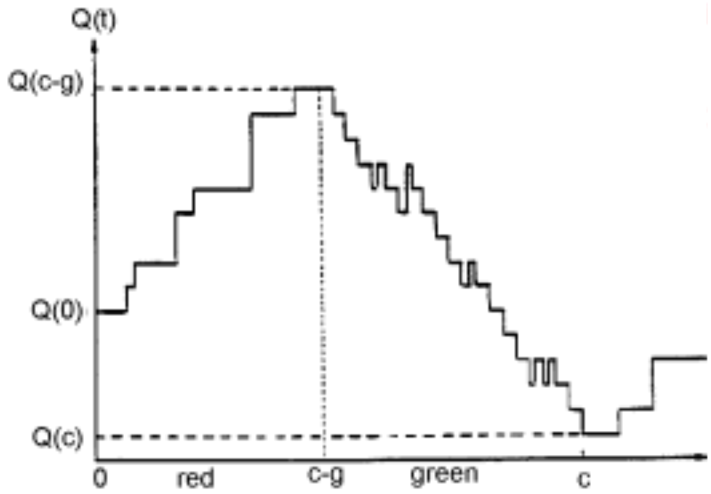

Figure 9.2
Queuing Process During One Signal Cycle
(Adapted from McNeil 1968).

performance measures can be derive including the average delay per vehicle (total delay divided by total cyclic arrivals) the number of vehicle stopped (Q_s), the maximum number of vehicles in the queue (Q_{max}) , and the average queue length (Q_{avg}). Performance models of this type are applicable to low flow to capacity ratios (up to about 0.50), since the assumption of zero initial and end queues is not violated in most cases.

As traffic intensity increases, however, there is a increased likelihood of "cycle failures". That is, some cycles will begin to experience an overflow queue of vehicles that could not discharge from a previous cycle. This phenomenon occurs at random, depending on which cycle happens to experience higher-than-capacity flow rates. The presence of an initial queue (Q_o) causes an additional delay which must be considered in the estimation of traffic performance. Delay models based on queue theory (e.g. M/D/n/FIFO) have been applied to account for this effect.

Interestingly, at extremely congested conditions, the stochastic queuing effect are minimal in comparison with the size of oversaturation queues. Therefore, a fluid theory approach may be appropriate to use for highly oversaturated intersections. This leaves a gap in delay models that are applicable to the range of traffic flows that are numerically close to the signal capacity. Considering that most real-world signals are timed to operate within that domain, the value of time-dependent models are of particular relevance for this range of conditions.

In the case of vehicle actuated control, neither the cycle length nor green times are known in advance. Rather, the length of the green is determined partly by controller-coded parameters such as minimum and maximum green times, and partly by the pattern of traffic arrivals. In the simplest case of a basic actuated controller, the green time is extended beyond its minimum so long as a) the time headway between vehicle arrivals does not exceed the controller's unit extension (U), and b) the maximum green has not been reached. Actuated control models are discussed further in Section 9.6.

9.3 Steady-State Delay Models

9.3.1 Exact Expressions

This category of models attempts to characterize traffic delays based on statistical distributions of the arrival and departure processes. Because of the purely theoretical foundation of the models, they require very strong assumptions to be considered valid. The following section describes how delays are estimated for this class of models, including the necessary data requirements.

The expected delay at fixed-time signals was first derived by Beckman (1956) with the assumption of the binomial arrival process and deterministic service:

$$d = \frac{c-g}{c(1-q/S)}\left[\frac{Q_o}{q}+\frac{c-g+1}{2}\right] \tag{9.1}$$

where,

- c = signal cycle,
- g = effective green signal time,
- q = traffic arrival flow rate,
- S = departure flow rate from queue during green,
- Q_o = expected overflow queue from previous cycles.

The expected overflow queue used in the formula and the restrictive assumption of the binomial arrival process reduce the practical usefulness of Equation 9.1. Little (1961) analyzed the expected delay at or near traffic signals to a turning vehicle crossing a Poisson traffic stream. The analysis, however, did not include the effect of turners on delay to other vehicles. Darroch (1964a) studied a single stream of vehicles arriving at a fixed-time signal. The arrival process is the generalized Poisson process with the Index of Dispersion:

$$I = \frac{var(A)}{qh} \tag{9.10}$$

where,

$var(.)$= variance of (.)
q = arrival flow rate,
h = interval length,
A = number of arrivals during interval $h = qh$.

The departure process is described by a flexible service mechanism and may include the effect of an opposing stream by defining an additional queue length distribution caused by this factor. Although this approach leads to expressions for the expected queue length and expected delay, the resulting models are complex and they include elements requiring further modeling such as the overflow queue or the additional queue component mentioned earlier. From this perspective, the formula is not of practical importance. McNeil (1968) derived a formula for the expected signal delay with the assumption of a general arrival process, and constant departure time. Following his work, we express the total vehicle delay during one signal cycle as a sum of two components

$$W = W_1 + W_2, \tag{9.3}$$

where

W_1 = total delay experienced in the red phase and
W_2 = total delay experienced in the green phase.

$$W_1 = \int_0^{(c-g)}[Q(0)+A(t)]\,dt \tag{9.4}$$

and

$$W_2 = \int_{(c-g)}^{c} Q(t)dt \tag{9.5}$$

where,

$Q(t)$ = vehicle queue at time t,
$A(t)$ = cumulative arrivals at t,

Taking expectations in Equation 9.4 it is found that:

$$E(W_1) = (c-g)\,Q_o + \frac{1}{2}q\,(c-g)^2. \tag{9.6}$$

Let us define a random variable Z_2 as the total vehicle delay experienced during green when the signal cycle is infinite. The variable Z_2 is considered as the total waiting time in a busy period for a queuing process $Q(t)$ with compound Poisson arrivals of intensity q, constant service time $1/S$ and an initial system state $Q(t=t_0)$. McNeil showed that provided $q/S<1$:

$$E(Z_2) = \frac{(1 + I q/S - q/S)\, E[Q(t_0)]}{2S(1-q/S)^2} + \frac{E[Q^2(t_0)]}{2S(1-q/S)}. \qquad (9.7)$$

Now W_2 can be expressed using the variable Z_2:

$$E(W_2) = E[Z_2 \mid Q(t=c-g)] - E[Z_2 \mid Q(t=c)] \qquad (9.8)$$

and

$$E[W_2] = \frac{(1 + I g/S + q/S)\{E[Q(c-g) - Q(c)]\}}{2S(1-q/S)^2} + \frac{E[Q^2(c-g)] - E[Q^2(c)]}{2S(1-q/S)}. \qquad (9.9)$$

The queue is in statistical equilibrium, only if the degree of saturation x is below 1,

$$x = \frac{q/S}{g/c} < 1. \qquad (9.10)$$

For the above condition, the average number of arrivals per cycle can discharge in a single green period. In this case $E[\,Q(0)\,] = E[\,Q(c)\,]$ and $E[\,Q^2(0)\,] = E[Q^2(c)\,]$. Also $Q(c-g) = Q(0) + A(c)$, so that:

$$E[Q(c-g) - Q(c)] = E[A(c-g)] = q(c-g) \qquad (9.11)$$

and

$$\begin{aligned} E[Q^2(c-g) - Q^2(c)] &= 2E[A(c-g)]E[Q(0)] + E[A^2(c-g)] \\ &= 2q(c-g)Q_o + q^2(c-g)^2 + q(c-g)I \end{aligned} \qquad (9.12)$$

Equations 9.9, 9.11, and 9.12 yield:

$$\begin{aligned} E(W_2) = \frac{1}{2S(1-q/S)^2}\,[(1 + I q/S - q/S)\, g(c-g) \\ + (1-q/S)(2q(c-g)Q_o + q^2(c-g)^2 + q(c-g)I)] \end{aligned} \qquad (9.13)$$

and using Equations 9.3, 9.4 and 9.13, the following is obtained:

$$E(W) = \frac{(c-g)}{c(1-q/S)}\left[\frac{Q_0}{q} + \left(\frac{c-g}{2}\right)qc + \frac{1}{S}\left(1 + \frac{I}{1-q/S}\right)\right] \qquad (9.14)$$

The average vehicle delay d is obtained by dividing $E(W)$ by the average number of vehicles in the cycle (qc):

$$d = \frac{c-g}{2c(1-q/S)}\,[(c-g) + \frac{2}{q}Q_o + \frac{1}{S}(1 + \frac{I}{1-q/S})] \qquad (9.15)$$

which is in essence the formula obtained by Darroch when the departure process is deterministic. For a binomial arrival process I=1-q/S, and Equation 9.15 becomes identical to that obtained by Beckmann (1956) for binomial arrivals. McNeil and Weiss (in Gazis 1974) considered the case of the compound Poisson arrival process and general departure process obtaining the following model:

$$d = \frac{(c-g)}{2c(1-q/S)}\left\{(c-g) + \frac{2}{q}\left[1 + \frac{(1-q/S)(1-B^2)}{2S}\right]Q_o + \frac{1}{S}(1 + \frac{I+B^2 q/S}{1-q/S})\right\} \qquad (9.16)$$

An examination of the above equation indicates that in the case of no overflow (Q_o= 0), and no randomness in the traffic process (I=0), the resultant delay becomes the uniform delay component. This component can be derived from a simple input-output model of uniform arrivals throughout the cycle and departures as described in Section 9.2. The more general case in Equation 9.16 requires knowledge of the size of the average overflow queue (or queue at the beginning of green), a major limitation on the practical usefulness of the derived formulae, since these are usually unknown.

A substantial research effort followed to obtain a closed-form analytical estimate of the overflow queue. For example, Haight (1959) specified the conditional probability of the overflow queue at the end of the cycle when the queue at the beginning of the cycle is known, assuming a homogeneous Poisson arrival process at fixed traffic signals. The obtained results were then modified to the case of semi-actuated signals. Shortly thereafter, Newell (1960) utilized a bulk service queuing model with an underlying binomial arrival process and constant departure time, using generating function technique. Explicit expressions for overflow queues were given for special cases of the signal split.

Other related work can be found in Darroch (1964a) who used a more general arrival distribution but did not produce a closed form expression of queue length, and Kleinecke (1964), whose work included a set of exact but complicated series expansion for Q_o, for the case of constant service time and Poisson arrival process.

9.3.2 Approximate Expressions

The difficulty in obtaining exact expressions for delay which are reasonably simple and can cover a variety of real world conditions, gave impetus to a broad effort for signal delay estimation using approximate models and bounds. The first, widely used approximate delay formula was developed by Webster (1961, reprint of 1958 work with minor amendments) from a combination of theoretical and numerical simulation approaches:

$$d = \frac{c(1-g/c)^2}{2[1-(g/c)x]} + \frac{x^2}{2q(1-x)} - 0.65\left(\frac{c}{q^2}\right)^{\frac{1}{3}} x^{2+5(g/c)} \qquad (9.17)$$

where,

d = average delay per vehicle (sec),
c = cycle length (sec),
g = effective green time (sec),
x = degree of saturation (flow to capacity ratio),
q = arrival rate (veh/sec).

The first term in Equation 9.17 represents delay when traffic can be considered arriving at a uniform rate, while the second term makes some allowance for the random nature of the arrivals. This is known as the "random delay", assuming a Poisson arrival process and departures at constant rate which corresponds to the signal capacity. The latter assumption does not reflect actual signal performance, since vehicles are served only during the effective green, obviously at a higher rate than the capacity rate. The third term, calibrated based on simulation experiments, is a corrective term to the estimate, typically in the range of 10 percent of the first two terms in Equation 9.17.

Delays were also estimated indirectly, through the estimation of Q_o, the average overflow queue. Miller (1963) for example obtained a approximate formulae for Q_o that are applicable to any arrival and departure distributions. He started with the general equality true for any general arrival and departure processes:

$$Q(c) = Q(0) + A - C + \Delta C \qquad (9.18)$$

where,

$Q(c)$ = vehicle queue at the end of cycle,
$Q(0)$ = vehicle queue at the beginning of cycle,
A = number of arrivals during cycle,
C = maximum possible number of departures during green,
ΔC = reserve capacity in cycle equal to (C-$Q(0)$-A) if $Q(0)+A < C$, zero otherwise.

Taking expectation of both sides of Equation 9.18, Miller obtained:

$$E(\Delta C) = E(C-A), \qquad (9.19)$$

since in equilibrium $Q(0) = Q(c)$.

Now Equation 9.18 can be rewritten as:

$$Q(c) - [\Delta C - E(\Delta C)] = Q(0) - [C - A - E(C-A)] \qquad (9.20)$$

Squaring both sides, taking expectations, the following is obtained:

$$E[Q(c)]^2 + 2E[Q(c)]E(\Delta C) + Var(\Delta C) = E[Q(0)]^2 + Var(C-A) \qquad (9.21)$$

For equilibrium conditions, Equation 9.21 can be rearranged as follows:

$$Q_o = \frac{Var(C-A) - Var(\Delta C)}{2E(C-A)} \tag{9.22}$$

where,

C = maximum possible number of departures in one cycle,
A = number of arrivals in one cycle,
ΔC = reserve capacity in one cycle.

The component $Var(\Delta C)$ is positive and approaches 0 when $E(C)$ approaches $E(A)$. Thus an upper bound on the expected overflow queue is obtained by deleting that term. Thus:

$$Q_o \le \frac{Var(C-A)}{2E(C-A)} \tag{9.23}$$

For example, using Darroch's arrival process (i.e. $E(A)=qc$, $Var(A)=Iqc$) and constant departure time during green ($E(C)=Sg$, $Var(C)=0$) the upper bound is shown to be:

$$Q_o \le \frac{Ix}{2(1-x)} \tag{9.24}$$

where $x=(qc)/(Sg)$.

Miller also considered an approximation of the excluded term $Var(\Delta C)$. He postulates that:

$$I \approx \frac{Var(\Delta C)}{E(C-A)} \tag{9.25}$$

and thus, an approximation of the overflow queue is

$$Q_o \approx \frac{(2x-1)I}{2(1-x)} \quad , \quad x \ge 0.50 \tag{9.26}$$

which can now be substituted in Equation 9.15. Further approximations of Equation 9.15 were aimed at simplifying it for practical purposes by neglecting the third and fourth terms which are typically of much lower order of magnitude than the first two terms. This approach is exemplified by Miller (1968a) who proposed the approximate formula:

$$d = \frac{(1-g/c)}{2(1-q/s)} \left[c(1-g/c) + \frac{2Q_0}{q} \right] \tag{9.27}$$

which can be obtained by deleting the second and third terms in McNeil's formula 9.15. Miller also gave an expression for the overflow queue formula under Poisson arrivals and fixed service time during the green:

$$Q_o = \frac{\exp\left[-1.33\sqrt{Sg(1-x)/x}\right]}{2(1-x)}. \tag{9.28}$$

Equations 9.15, 9.16, 9.17, 9.27, and 9.28 are limited to specific arrival and departure processes. Newell (1965) aimed at developing delay formulae for general arrival and departure distributions. First, he concluded from a heuristic graphical argument that for most reasonable arrival and departure processes, the total delay per cycle differs from that calculated with the assumption of uniform arrivals and fixed service times (Clayton, 1941), by a negligible amount if the traffic intensity is sufficiently small. Then, by assuming a queue discipline LIFO (Last In First Out) which does not effect the average delay estimate, he concluded that the expected delay when the traffic is sufficiently heavy can be approximated:

$$d = \frac{c(1-g/c)^2}{2(1-q/S)} + \frac{Q_o}{q}. \tag{9.29}$$

This formula gives identical results to formula (Equation 9.15) if one neglects components of $1/S$ order in (Equation 9.15) and when $1\text{-}q/S=1\text{-}g/c$. The last condition, however, is never met if equilibrium conditions apply. To estimate the overflow queue, Newell (1965) defines F_Q as the cumulative distribution of the overflow queue length, $F_{A\text{-}D}$ as the cumulative distribution of the overflow in the cycle, where the indices A and D represent cumulative arrivals and departures, respectively. He showed that under equilibrium conditions:

$$F_Q(x) = \int_0^{\infty} F_Q(z)\, dF_{A-D}(x-z) \qquad (9.30)$$

The integral in Equation 9.30 can be solved only under the restrictive assumption that the overflow in a cycle is normally distributed. The resultant Newell formula is as follows:

$$Q_o = \frac{qc(1-x)}{\Pi} \int_0^{\Pi/2} \frac{\tan^2\theta}{-1+\exp[Sg(1-x)^2/(2\cos^2\theta)]} d\theta. \qquad (9.31)$$

A more convenient expression has been proposed by Newell in the form:

$$Q_o = \frac{IH(\mu)x}{2(1-x)}. \qquad (9.32)$$

where,

$$\mu = \frac{Sg-qc}{(ISg)^{1/2}}. \qquad (9.33)$$

The function $H(\mu)$ has been provided in a graphical form.

Moreover, Newell compared the results given by expressions (Equation 9.29) and (Equation 9.31) with Webster's formula and added additional correction terms to improve the results for medium traffic intensity conditions. Newell's final formula is:

$$d = \frac{c(1-g/c)^2}{2(1-q/S)} + \frac{Q_o}{q} + \frac{(1-g/c)I}{2S(1-q/S)^2}. \qquad (9.34)$$

Table 9.1
Maximum Relative Discrepancy between the Approximate Expressions and Ohno's Algorithm (Ohno 1978).

Approximate Expressions (Equation #, Q_0 computed according to Equation #)	**Range of y = 0.0 ≈ 0.5**				**Range of *g/c* = 0.4 ~ 1.0**	
	s = 0.5 v/s		**s = 1.5 v/s**		**s = 0.5 v/s**	**s = 1.5 v/s**
	c = 90 s	**c = 30 s**	**c = 90 s**	**c = 30 s**	**c = 30 s**	**c = 90 s**
	g = 46 s	**g = 16 s**	**g = 45.33 s**	**g = 15.33 s**	**q = 0.2 s**	**q = 0.6 s**
Modified Miller's expression (9.15, 9.28)	0.22	2.60	-0.53	0.22	2.24	0.26
Modified Newell's expression (9.15, 9.31)	0.82	2.53	0.25	0.82	2.83	0.25-
McNeil's expression (9.15, Miller 1969)	0.49	1.79	0.12	0.49	1.51	0.08
Webster's full expression (9.17)	-8.04	-21.47	3.49	-7.75	119.24	1381.10
Newell's expression (9.34, 9.31)	-4.16	10.89	-1.45	-4.16	-15.37	-27.27

More recently, Cronje (1983b) proposed an analytical approximation of the function *H(μ)*:

$$H(\mu) = \exp[-\mu - (\mu^2/2)] \qquad (9.35)$$

where,

$$\mu = (1-x)(Sg)^{1/2}. \qquad (9.36)$$

He also proposed that the correction (third) component in Equation 9.34 could be neglected.

Earlier evaluations of delay models by Allsop (1972) and Hutchinson (1972) were based on the Webster model form. Later on, Ohno (1978) carried out a comparison of the existing delay formulae for a Poisson arrival process and constant departure time during green. He developed a computational procedure to provide the basis for evaluating the selected models, namely McNeil's expression, Equation 9.15 (with overflow queue calculated with the method described by Miller 1969), McNeil's formula with overflow queue according to Miller (Equation 9.28) (modified Miller's expression), McNeil's formula with overflow queue according to Newell (Equation 9.31) (modified Newell's expression), Webster expression (Equation 9.17) and the original Newell expression (Equation 9.34). Comparative results are depicted in Table 9.1 and Figures 9.3 and 9.4. Newell's expression appear to be more accurate than Webster, a conclusion shared by Hutchinson (1972) in his evaluation of three simplified models (Newell, Miller, and Webster). Figure 9.3 represents the percentage relative errors of the approximate delay models measured against Ohno's algorithm (Ohno 1978) for a range of flow ratios. The modified Miller's and Newell's expressions give almost exact average delay values, but they are not superior to the original McNeil formula. Figure 9.4 shows the same type of errors, categorized by the *g/c* ratio. Further efforts to improve on their estimates will not give any appreciable reduction in the errors. The modified Miller expression was recommended by Ohno because of its simpler form compared to McNeil's and Newell's.

9.4 Time-Dependent Delay Models

The stochastic equilibrium assumed in steady-state models requires an infinite time period of stable traffic conditions (arrival, service and control processes) to be achieved. At low flow to capacity ratios equilibrium is reached in a reasonable period of time, thus the equilibrium models are an acceptable approximation of the real-world process. When traffic flow approaches signal capacity, the time to reach statistical equilibrium usually exceeds the period over which demand is sustained. Further, in many cases the traffic flow exceeds capacity, a situation where steady-state models break down. Finally, traffic flows during the peak hours are seldom stationary, thus violating an important assumption of steady-state models. There has been many attempts at circumventing the limiting assumption of steady-state conditions. The first and simplest way is to deal with arrival and departure rates as a function of time in a deterministic fashion. Another view is to model traffic at signals, assuming stationary arrival and departure processes but not necessarily under stochastic equilibrium conditions, in order to estimate the average delay and queues over the modeled period of time. The latter approach approximates the time-dependent arrival profile by some mathematical function (step-function, parabolic, or triangular functions) and calculates the corresponding delay. In May and Keller (1967) delay and queues are calculated for an unsignalized bottleneck. Their work is nevertheless representative of the deterministic modeling approach and can be easily modified for signalized intersections. The general assumption in their research is that the random queue fluctuations can be neglected in delay calculations. The model defines a cumulative number of arrivals *A(t)*:

$$A(t) = \int_0^t q(\tau)\,d\tau \qquad (9.37)$$

and departures *D(t)* under continuous presence of vehicle queue over the period [0,*t*]:

$$D(t) = \int_0^t S(\tau)\,d\tau \qquad (9.38)$$

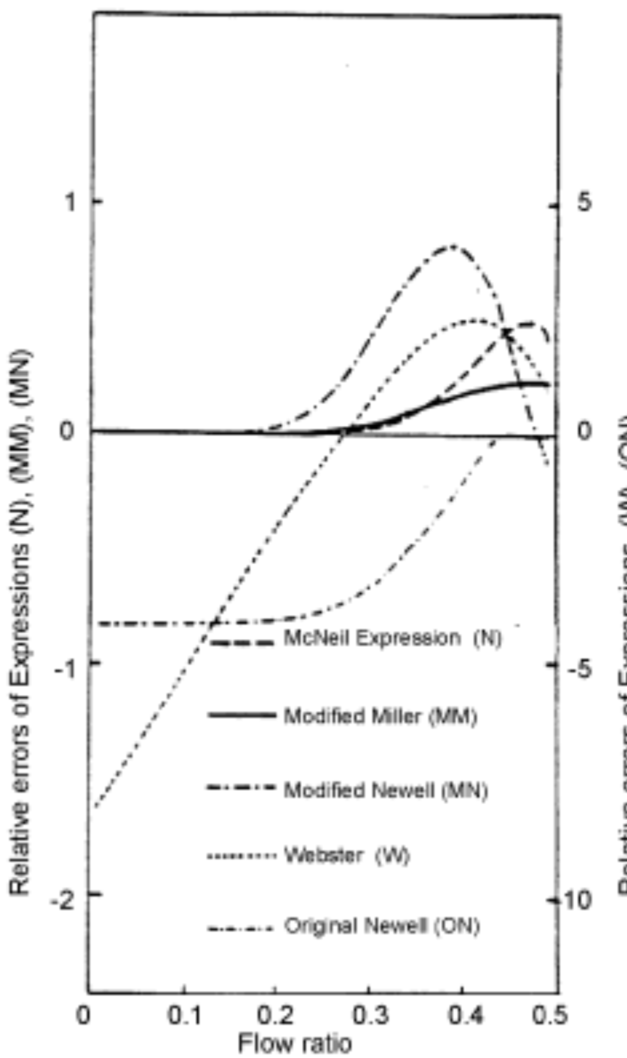

Figure 9.3
Percentage Relative Errors for Approximate Delay Models by Flow Ratios (Ohno 1978).

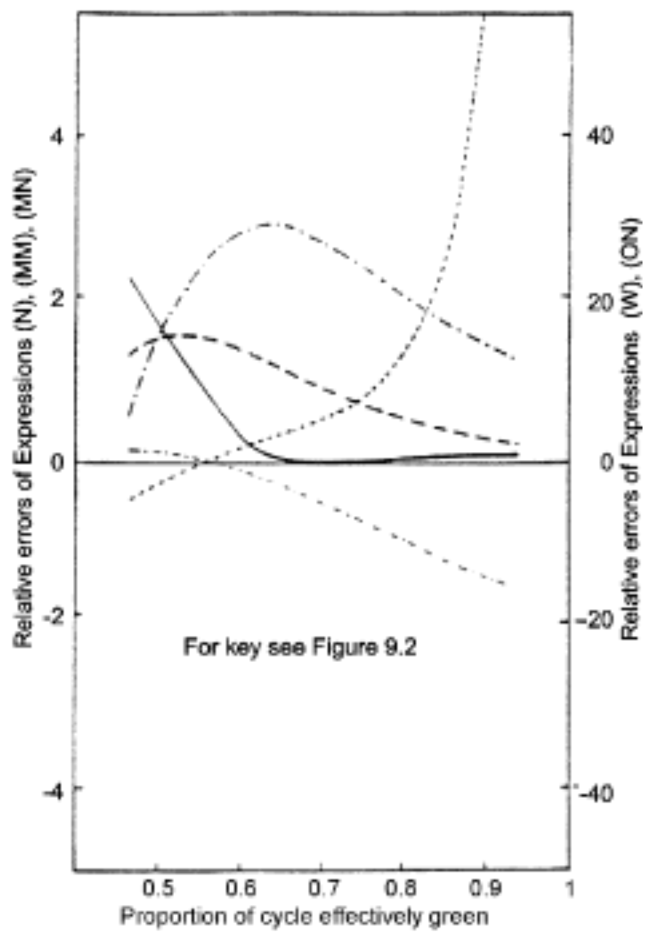

Figure 9.4
Relative Errors for Approximate Delay Models by Green to Cycle Ratios (Ohno 1978).

The current number of vehicles in the system (queue) is

$$Q(t) = Q(0) + A(t) - D(t) \tag{9.39}$$

and the average delay of vehicles queuing during the time period $[0,T]$ is

$$d = \frac{1}{A(T)} \int_0^T Q(t)dt \tag{9.40}$$

The above models have been applied by May and Keller to a trapezoidal-shaped arrival profile and constant departure rate. One can readily apply the above models to a signal with known signal states over the analysis period by substituting $C(\tau)$ for $S(\tau)$ in Equation 9.38:

$$\begin{aligned} C(\tau) &= 0 \text{ if signal is red,} \\ &= S(\tau) \text{ if signal is green and } Q(\tau) > 0, \\ &= q(\tau) \text{ if signal is green and } Q(\tau) = 0. \end{aligned}$$

Deterministic models of a single term like Equation 9.39 yield acceptable accuracy only when x<<1 or x>>1. Otherwise, they tend to underestimate queues and delays since the extra queues causes by random fluctuations in q and C are neglected.

According to Catling (1977), the now popular coordinate transformation technique was first proposed by Whiting, who did not publish it. The technique when applied to a steady-state curve derived from standard queuing theory, produces a time-dependent formula for delays. Delay estimates from the new models when flow approaches capacity are far more realistic than those obtained from the steady-state model. The following observations led to the development of this technique.

- At low degree of saturation (x<<1) delay is almost equal to that occurring when the traffic intensity is uniform (constant over time).
- At high degrees of saturation (x>>1) delay can be satisfactory described by the following deterministic model with a reasonable degree of accuracy:

$$d = d_1 + \frac{T}{2}(x-1) \tag{9.41}$$

where d_1 is the delay experienced at very low traffic intensity, (uniform delay) T = analysis period over which flows are sustained.

- steady-state delay models are asymptotic to the y-axis (i.e generate infinite delays) at unit traffic intensity (x=1). The coordinate transformation method shifts the original steady-state curve to become asymptotic to the deterministic oversaturation delay line--i.e.-- the second term in Equation 9.41--see Figure 9.5. The horizontal distance between the proposed delay curve and its asymptote is the same as that between the steady-state curve and the vertical line x=1.

There are two restrictions regarding the application of the formula: (1) no initial queue exists at the beginning of the interval $[0,T]$, (2) traffic intensity is constant over the interval $[0,T]$. The time-dependent model behaves reasonably within the period $[0,T]$ as indicated from simulation experiments. Thus, this technique is very useful in practice. Its principal drawback, in addition to the above stated restrictions (1) and (2) is the lack of a theoretical foundation. Catling overcame the latter difficulties by approximating the actual traffic intensity profile with a step-function. Using an example of the time-dependent version of the Pollaczek-Khintchine equation (Taha 1982), he illustrated the calculation of average queue and delay for each time interval starting from an initial, non-zero queue.

Kimber and Hollis (1979) presented a computational algorithm to calculate the expected queue length for a system with random arrivals, general service times and single channel service (M/G/1). The initial queue can be defined through its distribution. To speed up computation, the average initial queue is used unless it is substantially different from the queue at equilibrium. In this case, the full computational algorithm should be applied. The non-stationary arrival process is approximated with a step-function. The total delay in a time period is calculated by integrating the queue size over time. The coordinate transformation method is described next in some detail.

Suppose, at time T=0 there are $Q(0)$ waiting vehicles in queue and that the degree of saturation changes rapidly to x. In a deterministic model the vehicle queue changes as follows:

$$Q(T) = Q(0) + (x-1)CT. \tag{9.42}$$

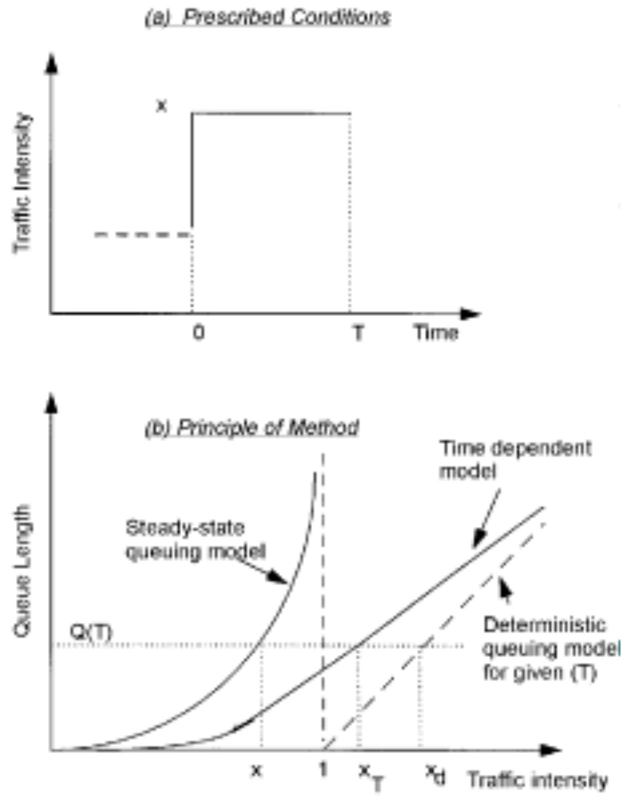

Figure 9.5
The Coordinate Transformation Method.

The steady-state expected queue length from the modified Pollaczek-Khintczine formula is:

$$Q = x + \frac{Bx^2}{1-x} \tag{9.43}$$

where B is a constant depending on the arrival and departure processes and is expressed by the following equation.

$$B = 0.5\left(1 + \frac{\sigma^2}{\mu^2}\right) \tag{9.44}$$

where σ^2 and μ are the variance and mean of the service time distribution, respectively.

The following derivation considers the case of exponential service times, for which $\sigma^2 = \mu^2$, $B = 1$. Let x_d be the degree of saturation in the deterministic model (Equation 9.42), x refers to the steady-state conditions in model (Equation 9.44), while x_T refers to the time-dependent model such that $Q(x,T)=Q(x_T,T)$. To meet the postulate of equal distances between the curves and the appropriate asymptotes, the following is true from Figure 9.5:

$$1 - x = x_d - x_T \tag{9.45}$$

and hence

$$x = x_T - (x_d - 1) \tag{9.46}$$

and from Equation 9.42:

$$x_d = \frac{Q(T) - Q(0)}{CT} + 1, \qquad (9.47)$$

the transformation is equivalent to setting:

$$x = x_T - \frac{Q(T) - Q(0)}{CT}. \qquad (9.48)$$

From Figure 9.5, it is evident that the queue length at time T, $Q(T)$ is the same at x, x_T, and x_d. By substituting for $Q(T)$ in Equation 9.44, and rewriting Equation 9.48 gives:

$$\frac{Q(T)}{1+Q(T)} = x_T - \frac{Q(T)-Q(0)}{CT} \qquad (9.49)$$

By eliminating the index T in x_T and solving the second degree polynomial in Equation 9.49 for $Q(T)$, it can be shown that:

$$Q(T)=\frac{1}{2}[(a^2+b)^{1/2}-a] \qquad (9.50)$$

where

$$a=(1-x)CT+1-Q(0) \qquad (9.51)$$

and

$$b = 4[Q(0) + xCT]. \qquad (9.52)$$

If the more general steady state Equation 9.43 is used, the result for Equation 9.51 and 9.52 is:

$$a = \frac{(1-x)(CT)^2+[1-Q(0)]CT-2(1-B)[Q(0)+xCT]}{CT+(1-B)} \qquad (9.53)$$

and

$$b = \frac{4[Q(0) + xCT][CT - (1-B)(Q(0) + xCT)]}{CT + (1-B)}. \qquad (9.54)$$

The equation for the average delay for vehicles arriving during the period of analysis is also derived starting from the average delay per arriving vehicle d_d over the period [0,T],

$$d_d=\frac{[Q(0)+1]+\frac{1}{2}(x-1)CT}{C} \qquad (9.55)$$

and the steady-state delay d_s,

$$d_s = \frac{1}{C}(1+\frac{Bx}{1-x}). \qquad (9.56)$$

The transformed time dependent equation is

$$d=\frac{1}{2}[(a^2+b)^{1/2}-a] \qquad (9.57)$$

with the corresponding parameters:

$$a=\frac{T}{2}(1-x)-\frac{1}{C}[Q(0)-B+2] \qquad (9.58)$$

and

$$b = \frac{4}{C}[\frac{T}{2}(1-x) + \frac{1}{2}xTB - \frac{Q(0)+1}{C}(1-B)]. \qquad (9.59)$$

The derivation of the coordinate transformation technique has been presented. The steady-state formula (Equation 9.43) does not appear to adequately reflect traffic signal performance, since a) the first term (queue for uniform traffic) needs further elaboration and b) the constant B must be calibrated for cases that do not exactly fit the assumptions of the theoretical queuing models.

Akçelik (1980) utilized the coordinate transformation technique to obtain a time-dependent formula which is intended to be more applicable to signalized intersection performance than Kimber-Hollis's. In order to facilitate the derivation of a time-dependent function for the average overflow queue Q_o, Akçelik used the following expression for undersaturated signals as a simple approximation to Miller's second formula for steady-state queue length (Equation 9.28):

$$Q_o = \begin{cases} \dfrac{1.5(x-x_o)}{1-x} & when\ x>x_o, \\ 0 & otherwise \end{cases} \qquad (9.60)$$

where

$$x_o = 0.67 + \frac{Sg}{600} \qquad (9.61)$$

Akçelik's time-dependent function for the average overflow queue is

$$Q_o = \begin{cases} \dfrac{CT}{4}[(x-1)+\sqrt{(x-1)^2+\dfrac{12(x-x_o)}{CT}}] & when\ x>x_o, \\ 0 & otherwise. \end{cases} \qquad (9.62)$$

The formula for the average uniform delay during the interval $[0,T]$ for vehicles which arrive in that interval is

$$d = \left\{ \begin{array}{ll} \dfrac{c(1-g/c)^2}{2(1-q/S)} & when\ x<1 \\ (c-g)/2 & when\ x\geq 1 \end{array} \right\} + \frac{Q_o}{C}. \qquad (9.63)$$

Generalizations of Equations 9.60 and 9.61 were discussed by Akçelik (1988) and Akçelik and Rouphail (1994). It should be noted that the average overflow queue, Q_0 is an approximation of the McNeil (Equation 9.15) and Miller (Equation 9.28) formulae applied to the time-dependent conditions, and differs from Newell's approximations Equation 9.29 and Equation 9.34 of the steady-state conditions. According to Akçelik (1980), this approximation is relevant to high degrees of saturation x and its effect is negligible for most practical purposes.

Following certain aspects of earlier works by Haight (1963), Cronje (1983a), and Miller (1968a); Olszewski (1990) used non-homogeneous Markov chain techniques to calculate the stochastic queue distribution using the arrival distribution $P(t,A)$ and capacity distribution $P(C)$. Probabilities of transition from a queue of i to j vehicles during one cycle are expressed by the following equation:

$$P_{i,j}(t) = \sum_{C=0}^{\infty} P_{i,j}(t,C)P(C) \qquad (9.64)$$

and

$$P_{i,0}(t,C) = \begin{cases} \displaystyle\sum_{k=0}^{C-i} P(t,A=k) & when\ i\leq C, \\ 0 & otherwise \end{cases} \qquad (9.65)$$

and

$$P_{i,j}(t,C) = \begin{cases} P(t,A=j-i+C) & when\ j\geq i-C, \\ 0 & otherwise. \end{cases} \qquad (9.66)$$

The probabilities of queue states transitions at time t form the transition matrix $P(t)$. The system state at time t is defined with the overflow queue distribution in the form of a row vector $P_Q(t)$. The initial system state variable distribution at time t =0 is assumed to be known: $P_Q(0)=[P_1(0), P_2(0),...P_m(0)]$, where $P_i(0)$ is the probability of queue of length i at time zero. The vector of state probabilities in any cycle t can now be found by matrix multiplication:

$$P_Q(t) = P_Q(t-1)\,P(t). \qquad (9.67)$$

Equation 9.67, when applied sequentially, allows for the calculation of queue probability evolution from any initial state.

In their recent work, Brilon and Wu (1990) used a similar computational technique to Olszewski's (1990a) in order to evaluate existing time-dependent formulae by Catling (1977), Kimber-Hollis (1979), and Akçelik (1980). A comparison of the models results is given in Figures 9.6 and 9.7 for a parabolic arrival rate profile in the analysis period T_o. They found that the Catling method gives the best approximation of the average delay. The underestimation of delays observed in the Akçelik's model is interpreted as a consequence of the authors' using an average arrival rate over the analyzed time period instead of the step function, as in the Catling's method. When the peak flow rate derived from a step function approximation of the parabolic profile is used in Akçelik's formula, the results were virtually indistinguishable from Brilon and Wu's (Akçelik and Rouphail 1993).

Using numeric results obtained from the Markov Chains approach, Brilon and Wu developed analytical approximate (and rather complicated) delay formulae of a form similar to Akçelik's which incorporate the impact of the arrival profile shape (e.g. the peaking intensity) on delay. In this examination of delay models in the time dependent mode, delay is defined according to the path trace method of measurement (Rouphail and Akçelik 1992a). This method keeps track of the departure time of each vehicle, even if this time occurs beyond the analysis period T. The path trace method will tend to generate delays that are typically longer than the queue sampling method, in which stopped vehicles are sampled every 15-20 seconds for the duration of the analysis period. In oversaturated conditions, the measurement of delay may yield vastly different results as vehicles may discharge 15 or 30 minutes beyond the analysis period. Thus it is important to maintain consistency between delay measurements and estimation methods. For a detailed discussion of the delay measurement methods and their impact on oversaturation delay estimation, the reader is referred to Rouphail and Akçelik (1992a).

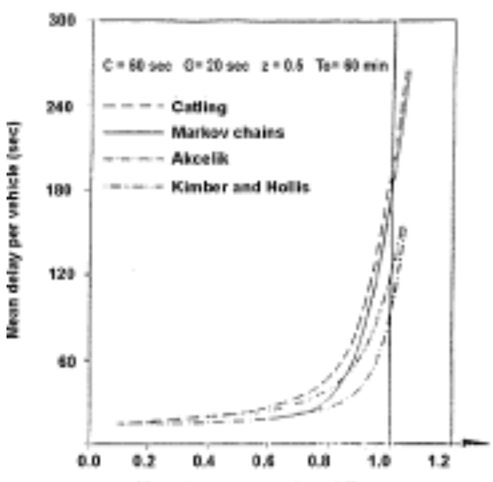

Figure 9.6
Comparison of Delay Models Evaluated by Brilon and Wu (1990) with Moderate Peaking (z=0.50).

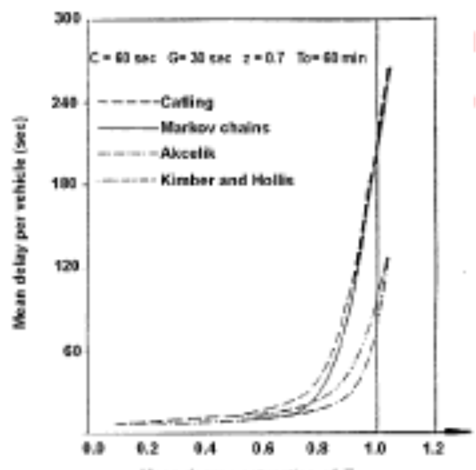

Figure 9.7
Comparison of Delay Models Evaluated by Brilon and Wu (1990) with High Peaking (z=0.70).

9.5 Effect of Upstream Signals

The arrival process observed at a point located downstream of some traffic signal is expected to differ from that observed upstream of the same signal. Two principal observations are made: a) vehicles pass the signal in "bunches" that are separated by a time equivalent to the red signal (platooning effect), and b) the number of vehicles passing the signal during one cycle does not exceed some maximum value corresponding to the signal throughput (filtering effect).

9.5.1 Platooning Effect On Signal Performance

The effect of vehicle bunching weakens as the platoon moves downstream, since vehicles in it travel at various speeds, spreading over the downstream road section. This phenomenon, known as platoon diffusion or dispersion, was modeled by Pacey (1956). He derived the travel time distribution $f(\tau)$ along a road section assuming normally distributed speeds and unrestricted overtaking:

$$f(\tau) = \frac{D}{\tau^2 \sigma \sqrt{2\pi}} \exp\left[- \frac{(\frac{D}{\tau} - \frac{D}{\bar{\tau}})^2}{2\sigma^2} \right] \qquad (9.68)$$

where,

D = distance from the signal to the point where arrivals are observed,
τ = individual vehicle travel time along distance D,
$\bar{\tau}$ = mean travel time, and
σ = standard deviation of speed.

The travel time distribution is then used to transform a traffic flow profile along the road section of distance D:

$$q_2(t_2)\, dt_2 = \int_{t_1} q_1(t_1)\, f(t_2 - t_1)\, dt_1\, dt_2 \qquad (9.69)$$

where,

$q_2(t_2)dt_2$ = total number of vehicles passing some point downstream of the signal in the interval $(t, t+dt)$,
$q_1(t_1)dt_1$ = total number of vehicles passing the signal in the interval $(t, t+dt)$, and
$f(t_2\text{-}t_1)$ = probability density of travel time $(t_2 - t_1)$ according to Equation 9.68.

The discrete version of the diffusion model in Equation 9.69 is

$$q_2(j) = \sum_i q_1(i)\, g(j-i) \qquad (9.70)$$

where i and j are discrete intervals of the arrival histograms.

Platoon diffusion effects were observed by Hillier and Rothery (1967) at several consecutive points located downstream of signals (Figure 9.8). They analyzed vehicle delays at pretimed signals using the observed traffic profiles and drew the following conclusions:

- the deterministic delay (first term in approximate delay formulae) strongly depends on the time lag between the start of the upstream and downstream green signals (offset effect);
- the minimum delay, observed at the optimal offset, increases substantially as the distance between signals increases; and
- the signal offset does not appear to influence the overflow delay component.

The TRANSYT model (Robertson 1969) is a well-known example of a platoon diffusion model used in the estimation of deterministic delays in a signalized network. It incorporates the Robertson's diffusion model, similar to the discrete version of the Pacey's model in Equation 9.70, but derived with the assumption of the binomial distribution of vehicle travel time:

$$q_2(j) = \frac{1}{1 + a\tau}\, q_1(j) + \left(1 - \frac{1}{1 + a\tau}\right) q_2(j-1) \qquad (9.71)$$

where τ is the average travel time and a is a parameter which must be calibrated from field observations. The Robertson model of dispersion gives results which are satisfactory for the

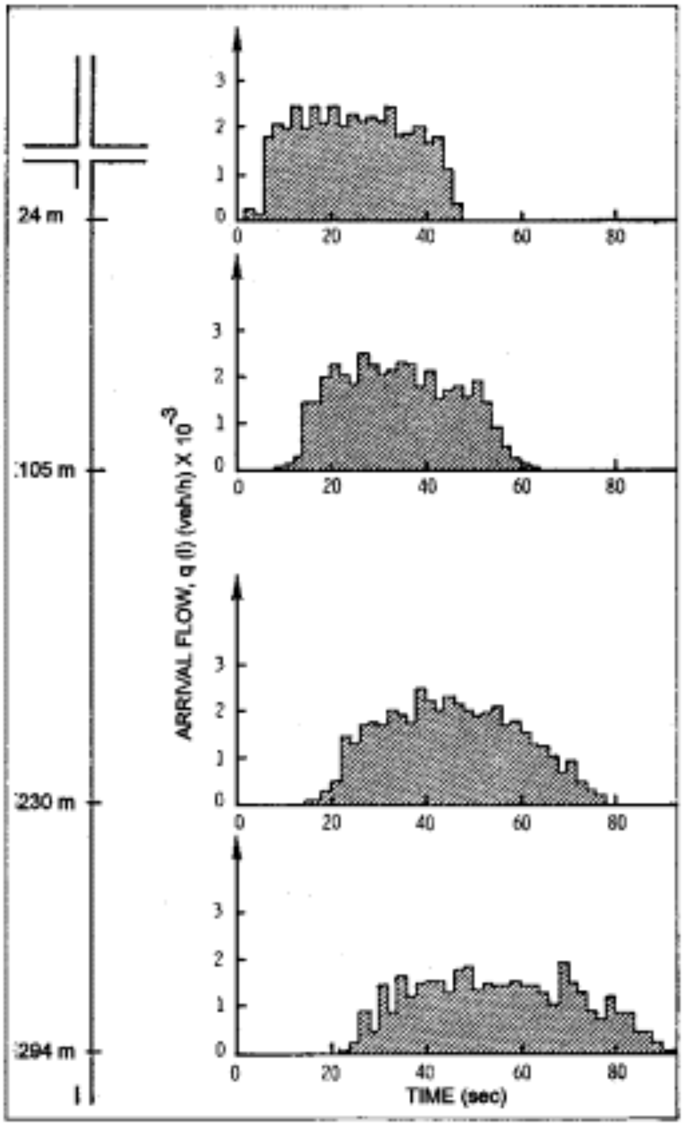

Figure 9.8
Observations of Platoon Diffusion
by Hillier and Rothery (1967).

purpose of signal optimization and traffic performance analysis in signalized networks. The main advantage of this model over the former one is much lower computational demand which is a critical issue in the traffic control optimization for a large size network.

In the TRANSYT model, a flow histogram of traffic served (departure profile) at the stopline of the upstream signal is first constructed, then transformed between two signals using model (Equation 9.71) in order to obtain the arrival patterns at the stopline of the downstream signal. Deterministic delays at the downstream signal are computed using the transformed arrival and output histograms.

To incorporate the upstream signal effect on vehicle delays, the Highway Capacity Manual (TRB 1985) uses a progression factor (*PF*) applied to the delay computed assuming an isolated signal. A PF is selected out of the several values based on a platoon ratio f_p. The platoon ratio is estimated from field measurement and by applying the following formula:

$$f_p = \frac{PVG}{g/c} \tag{9.72}$$

where,

PVG = percentage of vehicles arriving during the effective green,
g = effective green time,
c = cycle length.

Courage et al. (1988) compared progression factor values obtained from Highway Capacity Manual (HCM) with those estimated based on the results given by the TRANSYT model. They indicated general agreement between the methods, although the HCM method is less precise (Figure 9.9). To avoid field measurements for selecting a progression factor, they suggested to compute the platoon ratio f_p from the ratios of bandwidths measured in the time-space diagram. They showed that the proposed method gives values of the progression factor comparable to the original method.

Rouphail (1989) developed a set of analytical models for direct estimation of the progression factor based on a time-space diagram and traffic flow rates. His method can be considered a simplified version of TRANSYT, where the arrival histogram consists of two uniform rates with in-platoon and out-of-platoon traffic intensities. In his method, platoon dispersion is also based on a simplified TRANSYT-like model. The model is thus sensitive to both the size and flow rate of platoons. More recently, empirical work by Fambro et al. (1991) and theoretical analyses by Olszewski (1990b) have independently confirmed the fact that signal progression does not influence overflow queues and delays. This finding is also reflected in the most recent update of the Signalized Intersections chapter of the Highway Capacity Manual (1994). More recently, Akçelik (1995a) applied the HCM progression factor concept to queue length, queue clearance time, and proportion queued at signals.

The remainder of this section briefly summarizes recent work pertaining to the filtering effect of upstream signals, and the resultant overflow delays and queues that can be anticipated at downstream traffic signals.

9.5.2 Filtering Effect on Signal Performance

The most general steady-state delay models have been derived by Darroch (1964a), Newell (1965), and McNeil (1968) for the binomial and compound Poisson arrival processes. Since these efforts did not deal directly with upstream signals effect, the question arises whether they are appropriate for estimating overflow delays in such conditions. Van As (1991) addressed this problem using the Markov chain technique to model delays and arrivals at two closely spaced signals. He concluded that the Miller's model (Equation 9.27) improves random delay estimation in comparison to the Webster model (Equation 9.17). Further, he developed an approximate formula to transform the dispersion index of arrivals, I, at some traffic signal into the dispersion index of departures, B, from that signal:

$$B = I \exp(-1.3 F^{0.627}) \tag{9.73}$$

with the factor F given by

$$F = \frac{Q_o}{\sqrt{I_a q c}} \tag{9.74}$$

This model (Equation 9.73) can be used for closely spaced signals, if one assumes the same value of the ratio I along a road section between signals.

Tarko et al. (1993) investigated the impact of an upstream signal on random delay using cycle-by-cycle macrosimulation. They found that in some cases the ratio I does not properly represent the non-Poisson arrival process, generally resulting in delay overestimation (Figure 9.10).

They proposed to replace the dispersion index I with an adjustment factor f which is a function of the difference between the maximum possible number of arrivals m_c observable during one cycle, and signal capacity Sg:

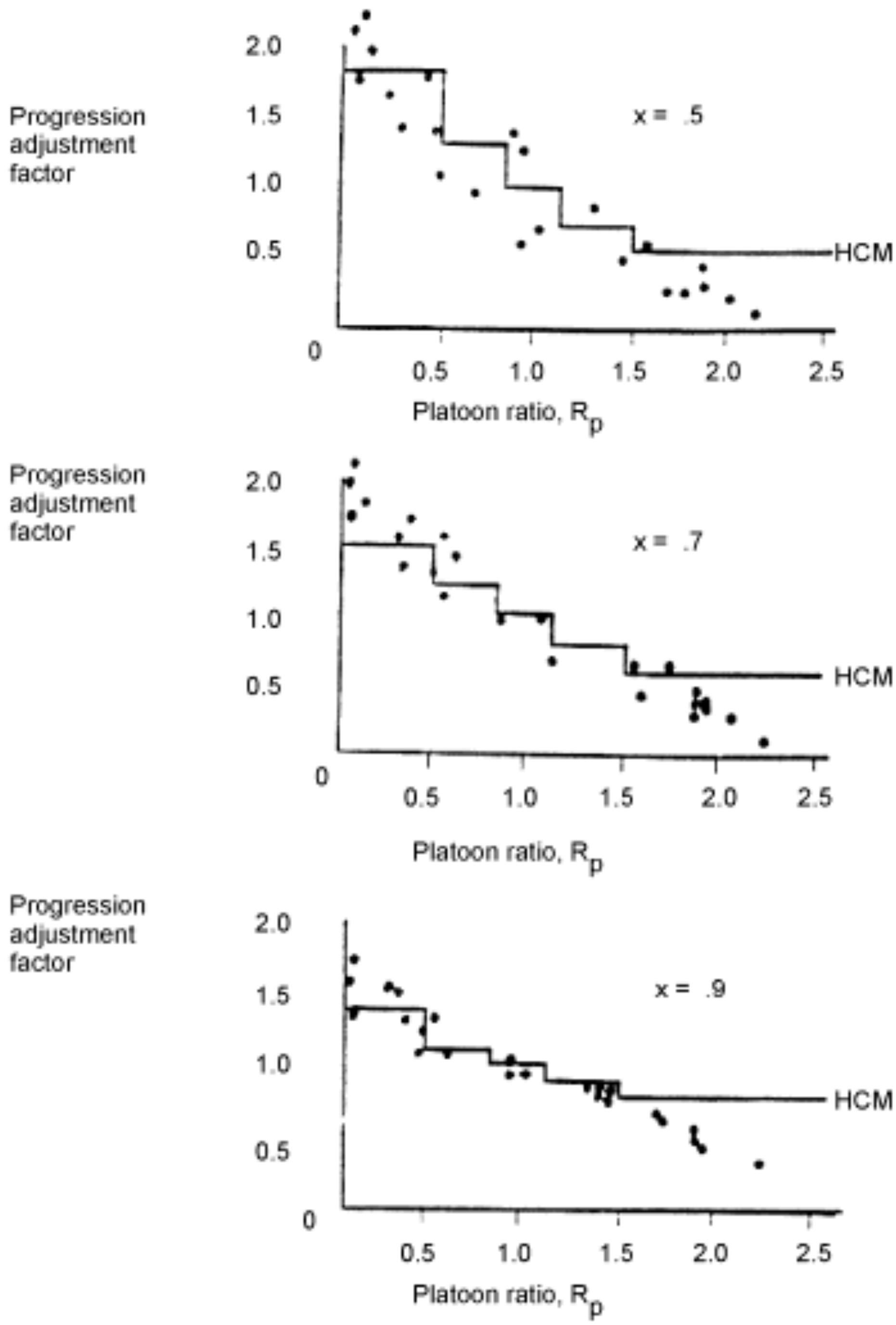

Figure 9.9
HCM Progression Adjustment Factor vs Platoon Ratio
Derived from TRANSYT-7F (Courage et al. 1988).

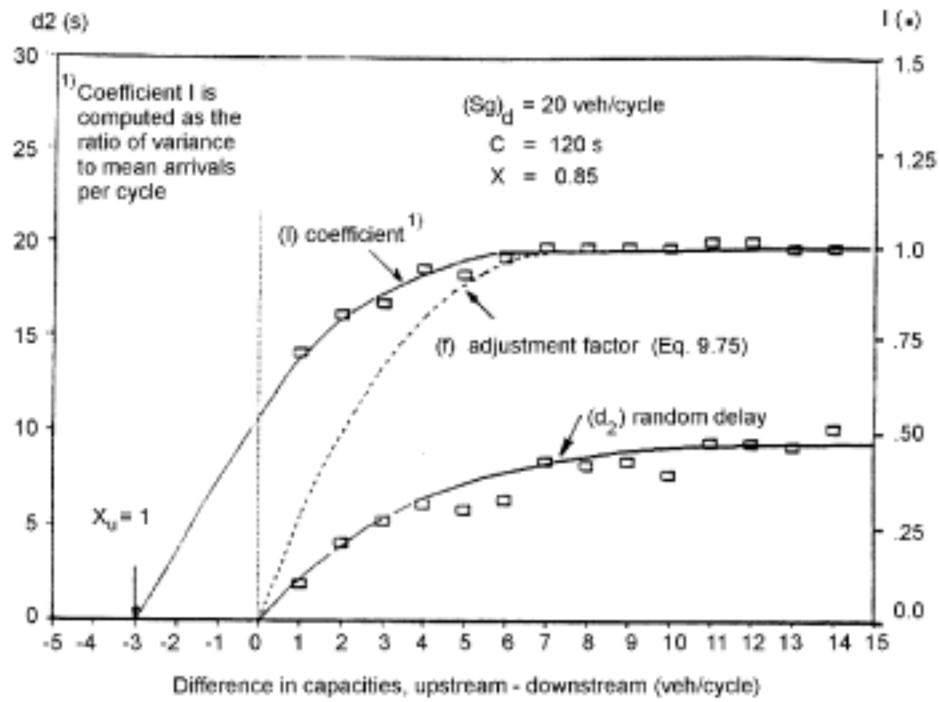

Figure 9.10
Analysis of Random Delay with Respect to the Differential Capacity Factor (f) and Var/Mean Ratio of Arrivals (I)- Steady State Queuing Conditions (Tarko et al. 1993).

$$f = 1 - e^{a(m_c - Sg)} \tag{9.75}$$

where a is a model parameter, $a < 0$.

A recent paper by Newell (1990) proposes an interesting hypothesis. The author questions the validity of using random delay expressions derived for isolated intersections at internal signals in an arterial system. He goes on to suggest that the sum of random delays at all intersections in an arterial system with no turning movements is equivalent to the random delay at the critical intersection, assuming that it is isolated. Tarko et al. (1993) tested the Newell hypothesis using a computational model which considers a bulk service queuing model and a set of arrival distribution transformations. They concluded that Newell's model estimates provide a close upper bound to the results from their model. The review of traffic delay models at fixed-timed traffic signals indicate that the state of the art has shifted over time from a purely theoretical approach grounded in queuing theory, to heuristic models that have deterministic and stochastic components in a time-dependent domain. This move was motivated by the need to incorporate additional factors such as non-stationarity of traffic demand, oversaturation, traffic platooning and filtering effect of upstream signals. It is anticipated that further work in that direction will continue, with a view towards using the performance-based models for signal design and route planning purposes.

9.6 Theory of Actuated and Adaptive Signals

The material presented in previous sections assumed fixed time signal control, i.e. a fixed signal capacity. The introduction of traffic-responsive control, either in the form of actuated or traffic-adaptive systems requires new delay formulations that are sensitive to this process. In this section, delay models for actuated signal control are presented in some detail, which incorporate controller settings such as minimum and maximum greens and unit extensions. A brief discussion of the state of the art in adaptive signal control follows, but no models are presented. For additional details on this topic, the reader is encouraged to consult the references listed at the end of the chapter.

9.6.1 Theoretically-Based Expressions

As stated by Newell (1989), the theory on vehicle actuated signals and related work on queues with alternating priorities is very large, however, little of it has direct practical value. For example, "exact" models of queuing theory are too idealized to be very realistic. In fact the issue of performance modeling of vehicle actuated signals is too complex to be described by a comprehensive theory which is simple enough to be useful. Actuated controllers are normally categorized into: fully-actuated, semi-actuated, and volume-density control. To date, the majority of the theoretical work related to vehicle actuated signals is limited to fully and semi-actuated controllers, but not to the more sophisticated volume-density controllers with features such as variable initial and extension intervals. Two types of detectors are used in practice: passage and presence. Passage detectors, also called point or small-area detectors, include a small loop and detect motion or passage when a vehicle crosses the detector zone. Presence detectors, also called area detectors, have a larger loop and detect presence of vehicles in the detector zone. This discussion focuses on traffic actuated intersection analysis with passage detectors only.

Delays at traffic actuated control intersections largely depend on the controller setting parameters, which include the following aspects: unit extension, minimum green, and maximum green. *Unit extension* (also called vehicle interval, vehicle extension, or gap time) is the extension green time for each vehicle as it arrives at the detector. *Minimum green*: summation of the initial interval and one unit extension. The initial interval is designed to clear vehicles between the detector and the stop line. *Maximum green*: the maximum green times allowed to a specific phase, beyond which, even if there are continuous calls for the current phase, green will be switched to the competing approach.

The relationship between delay and controller setting parameters for a simple vehicle actuated type was originally studied by Morris and Pak-Poy (1967). In this type of control, minimum and maximum greens are preset. Within the range of minimum and maximum greens, the phase will be extended for each arriving vehicle, as long as its headway does not exceed the value of unit extension. An intersection with two one-way streets was studied. It was found that, associated with each traffic flow condition, there is an optimal vehicle interval for which the average delay per vehicle is minimized. The value of the optimal vehicle interval decreases and becomes more critical, as the traffic flow increases. It was also found that by using the constraints of minimum and maximum greens, the efficiency and capacity of the signal are decreased. Darroch (1964b) also investigated a method to obtain optimal estimates of the unit extension which minimizes total vehicle delays.

The behavior of vehicle-actuated signals at the intersection of two one-way streets was investigated by Newell (1969). The arrival process was assumed to be stationary with a flow rate just slightly below the saturation rate, i.e. any probability distributions associated with the arrival pattern are time invariant. It is also assumed that the system is undersaturated but that traffic flows are sufficiently heavy, so that the queue lengths are considerably larger than one car. No turning movements were considered. The minimum green is disregarded since the study focused on moderate heavy traffic and the maximum green is assumed to be arbitrarily large. No specific arrival process is assumed, except that it is stationary.

Figure 9.11 shows the evolution of the queue length when the queues are large. Traffic arrives at a rate of q_1, on one approach, and q_2, on the other. r_j, g_j, and Y_j represent the effective red, green, and yellow times in cycle j. Here the signal timings are random variables, which may vary from cycle to cycle. For any specific cycle j, the total delay of all cars W_{ij} is the area of a triangular shaped curve and can be approximated by:

$$\mathrm{E}\{W_{1j}\} = \frac{q_1}{2(1-q_1/S_1)}\left[(E\{r_j\}+Y)^2 + Var(r_j) + \frac{I_1(E\{r_j\}+Y)}{S_1(1-q_1/S_1)} + \frac{V_1}{S_1 q_1}\right] \tag{9.76}$$

$$E\{W_{2j}\} = \frac{q_2}{2(1-q_2/S_2)}\left\{[(E\{g_j\}+Y)^2 + Var(g_j) + \frac{I_2[E(g_j)+Y]}{S_2(1-q_2/S_2)} + \frac{V_2}{S_2 q_2}\right\} \tag{9.77}$$

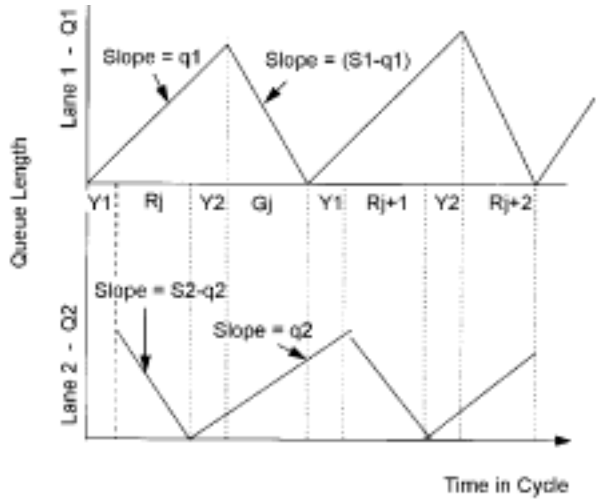

Figure 9.11
Queue Development Over Time Under
Fully-Actuated Intersection Control (Newell 1969).

where

$E\{W_{1j}\}, E\{W_{2j}\}$	=	the total wait of all cars during cycle j for approach 1 and 2;
S_1, S_2	=	saturation flow rate for approach 1 and 2;
$E\{r_j\}, E\{g_j\}$	=	expectation of the effective red and green times;
$\mathrm{Var}(r_j), \mathrm{Var}(g_j)$	=	variance of the effective red and green time;
I_1, I_2	=	variance to mean ratio of arrivals for approach 1 and 2; and
V_1, V_2	=	the constant part of the variance of departures for approach 1 and 2.

Since the arrival process is assumed to be stationary,

$$E\{r_j\} \rightarrow E\{r\}, \quad E\{g_j\} \rightarrow E\{g\} \qquad (9.78)$$

$$Var(r_j) \rightarrow Var(r), \quad Var(g_j) \rightarrow Var(g) \qquad (9.79)$$

$$E\{W_{kj}\} \rightarrow E\{W_k\}, \quad k=1,2 \qquad (9.80)$$

The first moments of r and g were also derived based on the properties of the Markov process:

$$E\{r\} = \frac{Yq_2/S_2}{1-q_1/S_1-q_2/S_2} \qquad (9.81)$$

$$E\{g\} = \frac{Yq_1/S_1}{1-q_1/S_1-q_2/S_2} \qquad (9.82)$$

Variances of r and g were also derived, they are not listed here for the sake of brevity. Extensions to the multiple lane case were investigated by Newell and Osuna (1969).

A delay model with vehicle actuated control was derived by Dunne (1967) by assuming that the arrival process follows a binomial distribution. The departure rates were assumed to be constant and the control strategy was to switch the signal when the queue vanishes. A single intersection with two one-lane one-way streets controlled by a two phase signal was considered.

For each of the intervals ($k\tau$, $k\tau+\tau$), k=0,1,2... the probability of one arrival in approach i = 1, 2 is denoted by q_i and the probability of no arrival by $p_i=1-q_i$. The time interval, τ, is taken as the time between vehicle departures. Saturation flow rate was assumed to be equal for both approaches. Denote

$W^{(2)}_r$ as the total delay for approach 2 for a cycle having effective red time of length r, then it can be shown that:

$$W^{(2)}_{r+1} = W^{(2)}_r + \mu[\delta_1 + c + \delta_2] \qquad (9.83)$$

where c is the cycle length, δ_1, and δ_2 are increases in delay at the beginning and at the end of the cycle, respectively, when one vehicle arrives in the extra time unit at the beginning of the phase and:

$$\begin{aligned} \mu = 0 \ \ & with\ \ probability\ \ p_2\ , \\ 1\ \ & with\ \ probability\ \ q_2\ . \end{aligned} \qquad (9.84)$$

Equation 9.83 means that if there is no arrivals in the extra time unit at the beginning of the phase, then $W^{(2)}_{r+1}=W^{(2)}_r$, otherwise $W^{(2)}_{r+1}=W^{(2)}_r + \delta_1 + c + \delta_2$.

Taking the expectation of Equation 9.83 and substituting for $E(\delta_1)$, $E(\delta_2)$:

$$E(W^{(2)}_{r+1}) = E(W^{(2)}_r) + q_2(r+1)/p_2 \qquad (9.85)$$

Solving the above difference equation for the initial condition $W^{(2)}_0=0$ gives,

$$E(W^{(2)}_r) = q_2(r^2+r)/2p_2 \qquad (9.86)$$

Finally, taking the expectation of Equation 9.86 with respect to r gives

$$E(W^{(2)}) = q_2\{var[r]E^2[r] + E[r]\}/(2p_2) \qquad (9.87)$$

Therefore, if the mean and variance of (r) are known, delay can be obtained from the above formula. $E(W^{(i)})$ for approach 2 is obtained by interchanging the subscripts.

Cowan (1978) studied an intersection with two single-lane one-way approaches controlled by a two-phase signal. The control policy is that the green is switched to the other approach at the earliest time, t, such that there is no departures in the interval $[t-\beta_i -1, t]$. In general $\beta_i \geq 0$. It was assumed that departure headways are 1 time unit, thus the arrival headways are at least 1 time unit. The arrival process on approach j is assumed to follow a bunched exponential distribution. It comprises random-sized bunches separated by inter-bunch headways. All bunched vehicles are assumed to have the same headway of 1 time unit. All inter-bunch headways follow the exponential distribution. Bunch size was assumed to have a general probability distribution with mean, μ_j, and variance, σ_j^2. The cumulative probability distribution of a headway less than t seconds, $F(t)$, is

$$F(t) = \begin{cases} 1 - \varphi e^{-\rho(t-\Delta)} & for\ \ t \geq \Delta \\ 0 & for\ \ t < \Delta \end{cases} \qquad (9.88)$$

where,

- Δ = minimum headway in the arrival stream, Δ=1 time unit;
- φ = proportion of free (unbunched) vehicles; and
- ρ = a delay parameter.

Formulae for average signal timings (r and g) and average delays for the cases of $\beta_j = 0$ and $\beta_j > 0$ are derived separately. $\beta_j = 0$ means that the green ends as soon as the queues for the approach clear while $\beta_j > 0$ means that after queues clear there will be a post green time assigned to the approach. By analyzing the property of Markov process, the following formula are derived for the case of $\beta_j = 0$.

$$E(g_1) = \frac{q_1 L}{1 - q_1 - q_2} \qquad (9.89)$$

$$E(g_2) = \frac{q_2 L}{1 - q_1 - q_2} \qquad (9.90)$$

$$E(r_1) = l_2 + \frac{q_2 L}{1 - q_1 - g_2} \qquad (9.91)$$

$$E(r_2) = l_1 + \frac{q_1 L}{1 - q_1 - q_2} \qquad (9.92)$$

where,

$E(g_1), E(g_2)$ = expected effective green for approach 1 and 2;
$E(r_1), E(r_2)$ = expected effective red for approach 1 and 2;
L = lost time in cycle;
l_1, l_2 = lost time for phase 1 and 2; and
q_1, q_2 = the stationary flow rate for approach 1 and 2.

The average delay for approach 1 is:

$$\frac{L(1-q_2)}{2(1-q_1-q_2)} + \frac{q_1^2(1-q_2)^2\rho_2(\sigma_2^2+\mu_2^2)+(1-q_1)^3(1-q_2)\rho_1(\sigma_1^2+\mu_1^2)}{2(1-q_1-q_2)(1-q_1-g_2+2q_1q_2)} \qquad (9.93)$$

Akçelik (1994, 1995b) developed an analytical method for estimating average green times and cycle time at a basic vehicle actuated controller that uses a fixed unit extension setting by assuming that the arrival headway follows the bunched exponential distribution proposed by Cowan (1978). In his model, the minimum headway in the arrival stream Δ is not equal to one. The delay parameter, ρ, is taken as $\varphi q_t/\theta$, where q_t is the total arrival flow rate and θ=1-Δq. In the model, the free (unbunched) vehicles are defined as those with headways greater than the minimum headway Δ. Further, all bunched vehicles are assumed to have the same headway Δ. Akçelik (1994) proposed two different models to estimate the proportion of free (unbunched) vehicles φ. The total time, g, allocated to a movement can be estimated as where g_{min} is the minimum green time and g_e, the green extension time. This green time, g, is subject to the following constraint

$$g \le g_{\max} \quad \vee \quad g_e < g_{emax} \qquad (9.94)$$

where g_{max} and g_{emax} are maximum green and extension time settings separately. If it is assumed that the unit extension is set so that a gap change does not occur during the saturated portion of green period, the green time can be estimated by:

$$g = g_s + e_g \qquad (9.95)$$

where g_s is the saturated portion of the green period and e_g is the extension time assuming that gap change occurs after the queue clearance period. This green time is subject to the boundaries:

$$g_{\min} \le g \le g_{\max} \qquad (9.96)$$

The saturated portion of green period can be estimated from the following formula:

$$g_s = \frac{f_q yr}{1-y} \qquad (9.97)$$

where,

f_q = queue length calibration factor to allow for variations in queue clearance time;
S = saturation flow;
r = red time; and
y = q/S, ratio of arrival to saturation flow rate.

The average extension time beyond the saturated portion can be estimated from:

$$e_g = n_g h_g + e_t \qquad (9.98)$$

where,

n_g = average number of arrivals before a gap change after queue clearance;
h_g = average headway of arrivals before a gap change after queue clearance; and
e_t = terminating time at gap change (in most case it is equal to the unit extension U).

For the case when $e_t = U$, Equation 9.98 becomes

$$e_g = -\frac{1}{q} + \left(\frac{\Delta}{\varphi} + \frac{1}{q}\right)e^{q(U-\Delta)} \qquad (9.99)$$

9.6.2 Approximate Delay Expressions

Courage and Papapanou (1977) refined Webster's (1958) delay model for pretimed control to estimate delay at vehicle-actuated signals. For clarity, Webster's simplified delay formula is restated below.

$$d = 0.9(d_1+d_2) = 0.9\left[\frac{c(1-g/c)^2}{2(1-q/s)} + \frac{x^2}{2q(1-x)}\right] \qquad (9.100)$$

Courage and Papapanou used two control strategies: (1) the available green time is distributed in proportion to demand on the critical approaches; and (2) wasted time is minimized by terminating each green interval as the queue has been properly serviced. They propose the use of the cycle lengths shown in Table 9.2 for delay estimation under pretimed and actuated signal control:

Table 9.2
Cycle Length Used For Delay Estimation for Fixed-Time and Actuated Signals Using Webster's Formula (Courage and Papapanou 1977).

Type of Signal	Cycle Length in 1st Term	Cycle Length in 2nd Term
Pretimed	Optimum	Optimum
Actuated	Average	Maximum

The optimal cycle length, c_0, is Webster's:

$$c_0 = \frac{1.5L + 5}{1 - \Sigma y_{ci}} \tag{9.101}$$

where L is total cycle lost time and y_{ci} is the volume to saturation flow ratio of critical movement i. The average cycle length, c_a is defined as:

$$c_a = \frac{1.5L}{1 - \Sigma y_{ci}} \tag{9.102}$$

and the maximum cycle length, c_{max}, is the controller maximum cycle setting. Note that the optimal cycle length under pretimed control will generally be longer than that under actuated control. The model was tested by simulation and satisfactory results obtained for a wide range of operations.

In the U. S. Highway Capacity Manual (1994), the average approach delay per vehicle is estimated for fully-actuated signalized lane groups according to the following:

$$d = d_1 * DF + d_2 \tag{9.103}$$

$$d_1 = \frac{c(1 - g/c)^2}{2(1 - xg/c)} \tag{9.104}$$

$$d_2 = 900Tx^2[(x-1) + \sqrt{(x-1)^2 + \frac{mx}{CT}}] \tag{9.105}$$

where, d, d_1, d_2, g, and c are as defined earlier and

- DF = delay factor to account for signal coordination and controller type;
- x = q/C, ratio of arrival flow rate to capacity;
- m = calibration parameter which depends on the arrival pattern;
- C = capacity in veh/hr; and
- T = flow period in hours (T=0.25 in 1994 HCM).

The delay factor DF=0.85, reduces the queuing delay to account for the more efficient operation with fully-actuated operation when compared to isolated, pretimed control. In an upcoming revision to the signalized intersection chapter in the HCM, the delay factor will continue to be applied to the uniform delay term only.

As delay estimation requires knowledge of signal timings in the average cycle, the HCM provides a simplified estimation method. The average signal cycle length is computed from:

$$c_a = \frac{Lx_c}{x_c - \Sigma y_{ci}} \tag{9.106}$$

where x_c = critical q/C ratio under fully-actuated control (x_c=0.95 in HCM). For the critical lane group i, the effective green:

$$g_i = \frac{y_{ci}}{x_c} c_a \tag{9.107}$$

This signal timing parameter estimation method has been the subject of criticism in the literature. Lin (1989), among others, compared the predicted cycle length from Equation 9.106 with field observations in New York state. In all cases, the observed cycle lengths were higher than predicted, while the observed x_c ratios were lower.

Lin and Mazdeysa(1983) proposed a general delay model of the following form consistent with Webster's approximate delay formula:

$$d = 0.9\{\frac{c(1-K_1\frac{g}{c})^2}{2(1-K_1\frac{g}{c}K_2x)} + \frac{3600(K_2x)^2}{2q(1-K_2x)}\} \quad (9.10)$$

where g, c, q, x are as defined earlier and K_1 and K_2 are two coefficients of sensitivity which reflect different sensitivities of traffic actuated and pretimed delay to both g/c and x ratios. In this study, K_1 and K_2 are calibrated from the simulation model for semi-actuated and fully-actuated control separately. More importantly, the above delay model has to be used in conjunction with the method for estimating effective green and cycle length. In earlier work, Lin (1982a, 1982b) described a model to estimate the average green duration for a two phase fully-actuated signal control. The model formulation is based on the following assumption: (1) the detector in use is small area passage detector; (2) right-turn-on-red is either prohibited or its effect can be ignored; and (3) left turns are made only from exclusive left turn lanes. The arrival pattern for each lane was assumed to follow a Poisson distribution. Thus, the headway distribution follows a shifted negative exponential distribution.

Figure 9.12 shows the timing sequence for a two phase fully actuated controller. For phase *i*, beyond the initial green interval, g_{mini}, green extends for F_i based on the control logic and the settings of the control parameters. F_i can be further divided into two components: (1) e_{ni} — the additional green extended by n vehicles that form moving queues upstream of the detectors after the initial interval G_{mini}; (2) E_{ni} — the additional green extended by n vehicles with headways of no more than one unit extension, U, after G_{mini} or e_{ni}. Note that e_{ni} and E_{ni} are random variables that vary from cycle to cycle. Lin (1982a, 1982b) developed the procedures to estimate e_i and E_i, the expected value of e_{ni} and E_{ni}, as follows. A moving queue upstream of a detector may exist when G_{mini} is timed out in case the flow rate of the critical lane q_c is high. If there are n vehicles arriving in the critical lane during time T_i, then the time required for the nth vehicle to reach the detector after G_{mini} is timed out can be estimated by the following equation:

$$t_n = nw + \sqrt{\frac{2(nL-s_i)}{a}} - G_{mini} \quad (9.109)$$

where w is the average time required for each queuing vehicle to start moving after the green phase starts, L is the average vehicle length, a is the vehicle acceleration rate from a standing position., and s is the detector setback. If $t_n \leq 0$, there is no moving queue exists and thus e_i=0; otherwise the green will be extended by the moving queue. Let s be the rate at which the queuing vehicle move across the detector. Considering that additional vehicles may join the queue during the time interval t_n, if t_n>0 and s>0, then:

$$e_{ni} = \frac{st_n}{s-q} \quad (9.110)$$

To account for the probability that no moving queues exist upstream of the detector at the end of the initial interval, the expected value of e_{ni}, e_i is expressed as:

$$e_i = \sum_{n=n_{\min}}^{\infty} \frac{P_j(n/T_i)\, e_{ni}}{1-p_j(n<n_{\min})} \quad (9.111)$$

where n_{min} is the minimum number of vehicles required to form a moving queue.

To estimate E_i, let us suppose that after the initial G_{mini} and additional green e_{ni} have elapsed, there is a sequence of k consecutive headways that are shorter than U followed by a headway longer than U. In this case the green will be extended k times and the resultant green extension time is $kJ+U$ with probability $[F(h \leq U)]^k F(h \geq U)$, where J is the average length of each extension and $F(h)$ is the cumulative headway distribution function.

$$J = \frac{\int_{\Delta}^{U_i} tf(t)\,dt}{F(h<U_i)} \quad (9.112)$$

and therefore

$$E_i = \sum_{k=0}^{\infty}(kJ+U)[F(h\leq U)]^k F(h>U) = -\frac{1}{q} + [\Delta + \frac{1}{q}]e^{q(U-\Delta)} \quad (9.113)$$

where Δ is the minimum headway in the traffic stream.

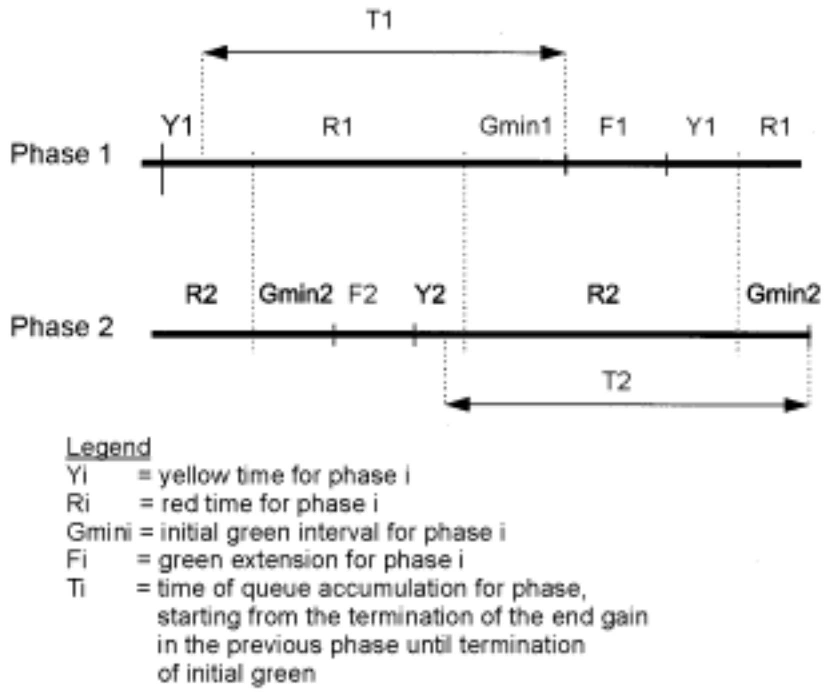

Figure 9.12
Example of a Fully-Actuated Two-Phase Timing Sequence (Lin 1982a).

Referring to Figure 9.12, after the values of T_1 and T_2 are obtained, G_i can be estimated as:

$$G_i = \sum_{n=0}^{\infty}(G_{mini}+e_i+E_i)P(n/T_i) \qquad (9.114)$$

subject to

$$G_{mini}+e_i+E_i \le (G_{\max})_i \qquad (9.115)$$

where $P(n/T_i)$ is the probability of n arrivals in the critical lane of the ith phase during time interval T_i. Since both T_1 and T_2 are unknown, an iterative procedure was used to determine G_1 and G_2.

Li et al. (1994) proposed an approach for estimating overflow delays for a simple intersection with fully-actuated signal control. The proposed approach uses the delay format in the 1994 HCM (Equations 9.104 and 9.105) with some variations, namely a) the delay factor, *DF*, is taken out of the formulation of delay model and b) the multiplier x^2 is omitted from the formulation of the overflow delay term to ensure convergence to the deterministic oversaturated delay model. Thus, the overflow delay term is expressed as:

$$d_2 = 900T[(x-1)+\sqrt{(x-1)^2+\frac{8kx}{CT}}] \qquad (9.116)$$

where the parameter (k) is derived from a numerical calibration of the steady-state for of Equation 9.105 as shown below.

$$d_2 = \frac{kx}{C(1-x)} \qquad (9.117)$$

This expression is based on a more general formula by Akçelik (1988) and discussed by Akçelik and Rouphail (1994). The calibration results for the parameter *k* along with the overall statistical model evaluation criteria (standard error and R^2) are depicted in Table 9.3. The parameter *k* which corresponds to pretimed control, calibrated by Tarko (1993) is also listed. It is noted that the pretimed steady-state model was also calibrated using the same approach, but with fixed signal settings. The first and most obvious observation is that the pretimed model produced the highest *k* (delay) value compared to the actuated models. Secondly, the parameter was found to increase with the size of the controller's unit extension (*U*).

Procedures for estimating the average cycle length and green intervals for semi-actuated signal operations have been developed by Lin (1982b, 1990) and Akçelik (1993b). Recently, Lin (1992) proposed a model for estimating average cycle length and green intervals under semi-actuated signal control operations with exclusive pedestrian-actuated phase. Luh (1991) studied the probability distribution of and delay estimation for semi-actuated signal controllers.

In summary, delay models for vehicle-actuated controllers are derived from assumptions related to the traffic arrival process, and are constrained by the actuated controller parameters. The distribution of vehicle headways directly impact the amount of green time allocated to an actuated phase, while controller parameters bound the green times within specified minimums and maximums. In contrast to fixed-time models, performance models for actuated have the additional requirement of estimating the expected signal phase lengths. Further research is needed to incorporate additional aspects of actuated operations such as phase skipping, gap reduction and variable maximum greens. Further, there is a need to develop generalized models that are applicable to both fixed time and actuated control. Such models would satisfy the requirement that both controls yield identical performance under very light and very heavy traffic demands. Recent work along these lines has been reported by Akçelik and Chung (1994, 1995).

9.6.3 Adaptive Signal Control

Only a very brief discussion of the topic is presented here. Adaptive signal control systems are generally considered superior to actuated control because of their true demand responsiveness. With recent advances in microprocessor technology, the gap-based strategies discussed in the previous section are becoming increasingly outmoded and demonstrably inefficient. In the past decade, control algorithms that rely on explicit intersection/network delay minimization in a time-variant environment, have emerged and been successfully tested. While the algorithms have matured both in Europe and the U.S., evident by the development of the MOVA controller in the U.K. (Vincent et al. 1988), PRODYN in France (Henry et al. 1983), and OPAC in the U.S. (Gartner et al. 1982-1983), theoretical work on traffic performance estimation under adaptive control is somewhat limited. An example of such efforts is the work by Brookes and Bell (1991), who investigated the use of Markov Chains and three heuristic approaches in an attempt to calculate the expected delays and stops for discrete time adaptive signal control. Delays are computed by tracing the queue evolution process over time using a `rolling horizon' approach. The main problem lies in the estimation (or prediction) of the initial queue in the current interval. While the Markov Chain approach yields theoretically correct answers, it is of limited value in practice due to its extensive computational and storage requirements. Heuristics that were investigated include the use of the mean queue length, in the last interval as the starting queue in this interval; the `two-spike' approach, in which the queue length

Table 9.3
Calibration Results of the Steady-State Overflow Delay Parameter (*k*) (Li et al. 1994).

Control	Pretimed	U=2.5	U=3.5	U=4.0	U=5.0
k (m=8k)	0.427	0.084	0.119	0.125	0.231
s.e.	NA	0.003	0.002	0.002	0.006
R^2	0.903	0.834	0.909	0.993	0.861

distribution has non-zero probabilities at zero and at an integer value closest to the mean; and finally a technique that propagates the first and second moment of the queue length distribution from period to period.

Overall, the latter method was recommended because it not only produces estimates that are sufficiently close to the theoretical estimate, but more importantly it is independent of the traffic arrival distribution.

9.7 Concluding Remarks

In this chapter, a summary and evolution of traffic theory pertaining to the performance of intersections controlled by traffic signals has been presented. The focus of the discussion was on the development of stochastic delay models.

Early models focused on the performance of a single intersection experiencing random arrivals and deterministic service times emulating fixed-time control. The thrust of these models has been to produce point estimates--i.e. expectations of-- delay and queue length that can be used for timing design and quality of service evaluation. The model form typically include a deterministic component to account for the red-time delay and a stochastic component to account for queue delays. The latter term is derived from a queue theory approach.

While theoretically appealing, the steady-state queue theory approach breaks down at high degrees of saturation. The problem lies in the steady-state assumption of sustained arrival flows needed to reach stochastic equilibrium (i.e the probability of observing a queue length of size Q is time-independent) . In reality, flows are seldom sustained for long periods of time and therefore, stochastic equilibrium is not achieved in the field at high degrees of saturation.

A compromise approach, using the coordinate transformation method was presented which overcomes some of these difficulties. While not theoretically rigorous, it provides a means for traffic performance estimation across all degrees of saturation which is also dependent on the time interval in which arrival flows are sustained.

Further extensions of the models were presented to take into account the impact of platooning, which obviously alter the arrival process at the intersection, and of traffic metering which may causes a truncation in the departure distribution from a highly saturated intersection. Next, an overview of delay models which are applicable to intersections operating under vehicle actuated control was presented. They include stochastic models which characterize the randomness in the arrival *and* departure process-- capacity itself is a random variable which can vary from cycle to cycle, and fixed-time equivalent models which treat actuated control as equivalent pretimed models operating at the average cycle and average splits.

Finally, there is a short discussion of concepts related to adaptive signal control schemes such as the MOVA systems in the United Kingdom and OPAC in the U.S. Because these approaches focus primarily on optimal signal control rather than performance modeling, they are somewhat beyond the scope of this document.

There are many areas in traffic signal performance that deserve further attention and require additional research. To begin with, the assumption of uncorrelated arrivals found in most models is not appropriate to describe platooned flow--where arrivals are highly correlated. Secondly, the estimation of the initial overflow queue at a signal is an area that is not well understood and documented. There is also a need to develop queuing/delay models that are constrained by the physical space available for queuing. Michalopoulos (1988) presented such an application using a continuous flow model approach. Finally, models that describe the interaction between downstream queue lengths and upstream departures are needed. Initial efforts in this direction have been documented by Prosser and Dunne (1994) and Rouphail and Akçelik (1992b).

References

Akçelik, R. (1980). *Time-Dependent Expressions for Delay, Stop Rate and Queue Length at Traffic Signals*. Australian Road Research Board, Internal Report, AIR 367-1.

Akçelik, R. (1988). *The Highway Capacity Manual Delay Formula for Signalized Intersections*. ITE Journal 58(3), pp. 23-27.

Akçelik, R. and N. Rouphail (1993). *Estimation of Delays at Traffic Signals for Variable Demand Conditions*. Transportation Research-B, Vol. 27B, No. 2, pp. 109-131.

Akçelik, R. (1994). *Estimation of Green Times and Cycle Time for Vehicle-Actuated Signals.* Transportation Research Board 1457, pp.63-72.

Akçelik, R. and E. Chung (1994). *Traffic Performance Models for Unsignalized Intersections and Fixed-Time Signals.* Proceedings, Second International Symposium on Highway Capacity, Sydney, Australia, ARRB, Volume I, pp. 21-50.

Akçelik, R. and N. Rouphail (1994). *Overflow Queues and Delays with Random and Platoon Arrivals at Signalized Intersections,* Journal of Advanced Transportation, Volume 28(3), pp. 227-251.

Akçelik, R. (1995a). *Extension of the Highway Capacity Manual Progression Factor Method for Platooned Arrivals,* Research Report ARR No. 276, ARRB Transport Research Ltd., Vermont South, Australia.

Akçelik, R. (1995b). *Signal Timing Analysis for Vehicle Actuated Control.* Working Paper WD TE95/007, ARRB Transport Research Ltd., Vermont South, Australia.

Akçelik, R. and E. Chung (1995). *Calibration of Performance Models for Traditional Vehicle-Actuated and Fixed-Time Signals.* Working Paper WD TO 95/103, ARRB Transport Research Ltd., Vermont South, Australia.

Allsop, R. (1972). *Delay at a Fixed Time Traffic Signal-I: Theoretical Analysis*. Transportation Science, 6, pp. 260-285.

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

Bell, M. G. H. and D. Brookes (1993). *Discrete Time-Adaptive Traffic Signal Control*. Transportation Research, Vol. 1C, No. 1, pp. 43-55.

Brilon, W. and N. Wu (1990). *Delays At Fixed-time Traffic Signals Under Time Dependent Traffic Conditions*. Traffic Engineering and Control, 31(12), pp. 623-63.

Brookes, D. and M. G. Bell (1991). *Expected Delays and Stop Calculation for Discrete Adaptive Traffic Signal Control*. Proceedings, First International Symposium on Highway Capacity, Karlsruhe, Germany, U. Brannolte, Ed.

Catling, I. (1977). *A Time-Dependent Approach To Junction Delays*. Traffic Engineering and Control, 18(11), pp. 520-523,526.

Clayton, A. (1941). *Road Traffic Calculations*. J. Inst. Civil. Engrs, 16, No.7, pp.247-284, No. 8, pp. 588-594.

Courage, K. G. and P. P. Papapanou (1977). *Estimation of Delay at Traffic-Actuated Signals*. Transportation Research Record, 630, pp. 17-21.

Courage, K. G., C. E. Wallace, and R. Alqasem (1988). *Modeling the Effect of Traffic Signal Progression on Delay*. Transportation Research Record, 1194, TRB, National Research Council, Washington, DC, pp. 139-146.

Cowan, R. (1978). *An Improved Model for Signalized Intersections with Vehicle-Actuated Control*. J. Appl. Prob. 15, pp. 384-396.

Cronje, W. B. (1983a). *Derivation of Equations for Queue Length, Stops, and Delays for Fixed Time Traffic Signals*. Transportation Research Record, 905, pp. 93-95.

Cronje, W. B. (1983b). *Analysis of Existing Formulas for Delay, Overflow, and Stops*. Transportation Research Record, 905, pp. 89-93.

Darroch, J. N. (1964a). *On the Traffic-Light Queue*. Ann. Math. Statist., 35, pp. 380-388.

Darroch, J. N., G. F. Newell, and R. W. J. Morris (1964b). *Queues for a Vehicle-Actuated Traffic Light*. Operational Research, 12, pp. 882-895.

Dunne, M. C. (1967). *Traffic Delay at a Signalized Intersection with Binomial Arrivals*. Transportation Science, 1, pp. 24-31.

Fambro, D. B., E. C. P. Chang, and C. J. Messer (1991). *Effects of the Quality of Traffic Signal Progression on Delay*. National Cooperative Highway Research Program Report 339, Transportation Research Board, National Research Council, Washington, DC.

Gartner, N. (1982-1983). *Demand Responsive Decentralized Urban Traffic Control*. Part I: Single Intersection Policies; Part II: Network Extensions, Office of University Research, U.S.D.O.T.

Gazis, D. C. (1974). *Traffic Science*. A Wiley-Intersection Publication, pp. 148-151.

Gerlough, D. L. and M. J. Huber (1975). *Traffic Flow Theory*. Transportation Research Board Special Report, 165, National Research Council, Washington, DC.

Haight, F. A. (1959). *Overflow At A Traffic Flow*. Biometrika. Vol. 46, Nos. 3 and 4, pp. 420-424.

Haight, F. A. (1963). *Mathematical Theories of Traffic Flow*. Academic Press, New York.

Henry, J., J. Farges, and J. Tuffal (1983). *The PRODYN Real-Time Traffic Algorithm*. 4th IFAC-IFIC--IFORS on Control in Transportation Systems, Baden-Baden.

Hillier, J. A. and R. Rothery (1967). *The Synchronization of Traffic Signals for Minimum Delays*. Transportation Science, 1(2), pp. 81-94.

Hutchinson, T. P. (1972). *Delay at a Fixed Time Traffic Signal-II. Numerical Comparisons at Some Theoretical Expressions*. Transportation Science, 6(3), pp. 286-305.

Kimber, R. M. and E. M. Hollis (1978). *Peak Period Traffic Delay at Road Junctions and Other Bottlenecks*. Traffic Engineering and Control, Vol. 19, No. 10, pp. 442-446.

Kimber, R. and E. Hollis (1979). *Traffic Queues and Delays at Road Junctions*. TRRL Laboratory Report, 909, U.K.

Li, J., N. Rouphail, and R. Akçelik (1994). *Overflow Delay Estimation for Intersections with Fully-Actuated Signal Control*. Presented at the 73rd Annual Meeting of TRB, Washington, DC.

Lighthill, M. H. and G. B. Whitham (1957). *On Kinematic Waves: II. A Theory of Traffic Flow on Long Crowded Roads*. Proceedings of the Royal Society, London Series A229, No. 1178, pp. 317-345.

Little, J. D. C. (1961). *Approximate Expected Delays for Several Maneuvers by Driver in a Poisson Traffic*. Operations Research, 9, pp. 39-52.

Lin, F. B. (1982a). *Estimation of Average Phase Duration of Full-Actuated Signals*. Transportation Research Record, 881, pp. 65-72.

Lin, F. B. (1982b). *Predictive Models of Traffic-Actuated Cycle Splits*. Transportation Research - B, Vol. 16B, No. 5, pp. 361-372.

Lin, F. (1989). *Application of 1985 Highway Capacity Manual for Estimating Delays at Signalized Intersection*. Transportation Research Record, 1225, TRB, National Research Council, Washington DC, pp. 18-23.

Lin, F. (1990). *Estimating Average Cycle Lengths and Green Intervals of Semiactuated Signal Operations for Level-of-Service Analysis*. Transportation Research Record, 1287, TRB, National Research Council, Washington DC, pp. 119-128.

Lin, F. (1992). *Modeling Average Cycle Lengths and Green Intervals of Semi-actuated Signal Operations with Exclusive Pedestrian-actuated Phase*. Transportation Research, Vol. 26B, No. 3, pp. 221-240.

Lin, F. B. and F. Mazdeyasa (1983). *Delay Models of Traffic Actuated Signal Controls.* Transportation Research Record, 905, pp. 33-38.

Luh, J. Z. and Chung Yee Lee (1991). *Stop Probability and Delay Estimations at Low Volumes for Semi-Actuated Traffic Signals*. Transportation Science, 25(1), pp. 65-82.

May, A. D. and H. M. Keller (1967). *A Deterministic Queuing Model.* Transportation Research, 1(2), pp. 117-128.

May, A. D. (1990). *Traffic Flow Fundamentals*. Prentice Hall, Englewood Cliffs, New Jersey, pp. 338-375.

McNeil, D. R. (1968). *A Solution to the Fixed-Cycle Traffic Light Problem for Compound Poisson Arrivals*. J. Appl. Prob. 5, pp. 624-635.

Miller, A. J. (1963). *Settings for Fixed-Cycle Traffic Signals*. Operational Research Quarterly, Vol. 14, pp. 373-386.

Miller, A. J. (1968a). *Australian Road Capacity Guide - Provisional Introduction and Signalized Intersections*. Australian Road Research Board Bulletin No.4, (Superseded by ARRB report ARR No. 123, 1981).

Miller, A. J. (1968b). *The Capacity of Signalized Intersections in Australia*. Australian Road Research Board, ARRB Bulletin No.3.

Morris, R. W. T. and P. G. Pak-Boy (1967). *Intersection Control by Vehicle Actuated Signals*. Traffic Engineering and Control, No.10, pp. 288-293.

Newell, G. F. (1960). *Queues for a Fixed-Cycle Traffic Light*. The Annuals of Mathematical Statistics, Vol.31, No.3, pp. 589-597.

Newell, G. F. (1965). A*pproximation Methods for Queues with Application to the Fixed-Cycle Traffic Light*. SIAM Review, Vol.7.

Newell, G. F. (1969). *Properties of Vehicle Actuated Signals: I. One-Way Streets*. Transportation Science, 3, pp. 30-52.

Newell, G. F. (1989). *Theory of Highway Traffic Signals*. UCB-ITS-CN-89-1, Institute of Transportation Studies, University of California.

Newell, G. F. (1990). *Stochastic Delays on Signalized Arterial Highways*. Transportation and Traffic Theory, Elsevier Science Publishing Co.,Inc., M. Koshi, Ed., pp. 589-598.

Back to the main publication page.

Newell, G. F. and E. E. Osuna (1969). *Properties of Vehicle-Actuated Signals: II. Two-Way Streets*. Transportation Science, 3, pp. 99-125.

Ohno, K. (1978). *Computational Algorithm for a Fixed Cycle Traffic Signal and New Approximate Expressions for Average Delay*. Transportation Science, 12(1), pp. 29-47.

Olszewski, P. S. (1988). *Efficiency of Arterial Signal Coordination*. Proceedings 14th ARRB Conference, 14(2), pp. 249-257.

Olszewski, P. (1990a). *Modelling of Queue Probability Distribution at Traffic Signals*. Transportation and Traffic Flow Theory, Elsevier Science Publishing Co., Inc., M. Koshi, Ed., pp. 569-588.

Olszewski, P. (1990b). *Traffic Signal Delay Model for Non-Uniform Arrivals.* Transportation Research Record, 1287, pp. 42-53.

Pacey, G. M. (1956). *The Progress of a Bunch of Vehicles Released from a Traffic Signal*. Research Note No. Rn/2665/GMP. Road Research Laboratory, London (mimeo).

Potts, R. B. (1967). *Traffic Delay at a Signalized Intersection with Binomial Arrivals*. pp. 126-128.

Robertson, D. I. (1969). *TRANSYT: A Traffic Network Study Tool*. Road Research Laboratory Report LR 253, Crowthorne.

Rorbech, J. (1968). *Determining The Length of The Approach Lanes Required at Signal-Controlled Intersections on Through Highways*. Transportation Research Record, 1225, TRB, National Research Council, Washington, DC, pp. 18-23.

Rouphail, N. (1989). *Progression Adjustment Factors at Signalized Intersections*. Transportation Research Record, 1225, pp. 8-17.

Rouphail, N. and R. Akçelik (1992a). *Oversaturation Delay Estimates with Consideration of Peaking*. Transportation Research Record, 1365, pp. 71-81.

Rouphail, N. and R. Akçelik(1992b). *A Preliminary Model of Queue Interaction at Signalized Paired Intersections*, Proceedings, 16th ARRB Conference 16(5), Perth, Australia, pp. 325-345.

Stephanopoulos, G., G. Michalopoulos, and G. Stephanopoulos (1979). *Modelling and Analysis of Traffic Queue Dynamics at Signalized Intersections*. Transportation Research, Vol. 13A, pp. 295-307.

Tarko, A., N. Rouphail, and R. Akçelik (1993b). *Overflow Delay at a Signalized Intersection Approach Influenced by an Upstream Signal: An Analytical Investigation*. Transportation Research Record, No. 1398, pp. 82-89.

Transportation Research Board (1985). Highway Capacity Manual, Special Report 209, National Research Council, Washington, DC.

Transportation Research Board (1993). Highway Capacity Manual, Chapter 9 (Signalized Intersections), National Research Council, Washington, DC.

Van As, S. C (1991). *Overflow Delay in Signalized Networks*. Transportation Research - A, Vol. 25A, No. 1, pp. 1-7.

Vincent, R. A. and J. R. Pierce (1988). *MOVA: Traffic Responsive, Self Optimizing Signal Control for Isolated Intersections*. TRRL Lab Research Report, 170.

Webster, F. V. (1958). *Traffic Signal Settings*. Road Research Laboratory Technical Paper No. 39, HMSO, London.

Wirasinghe, S. C. (1978). *Determination of Traffic Delays from Shock-Wave Analysis*. Transportation Research, Vol. 12, pp. 343-348.