TRANSPORTATION RESEARCH
# RECORD

No. 1484

*Highway Operations, Capacity, and Traffic Control*

---

# Traffic Operations: Highway Capacity

*A peer-reviewed publication of the Transportation Research Board*

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1995

**Transportation Research Record 1484**
ISSN 0361-1981
ISBN 0-309-06123-7
Price: $26.00

Subscriber Category
IVA highway operations, capacity, and traffic control

Printed in the United States of America

**Sponsorship of Transportation Research Record 1484**

**GROUP 3—OPERATION, SAFETY, AND MAINTENANCE OF TRANSPORTATION FACILITIES**
*Chairman: Jerome W. Hall, University of New Mexico*

**Facilities and Operation Section**
*Chairman: Jack L. Kay, JHK & Associates*

Committee on Highway Capacity and Quality of Service
*Chairman: Adolf D. May, Jr., University of California at Berkeley*
*Secretary: Wayne K. Kittelson, Kittelson & Associates, Inc.*
*Rahmi Akcelik, James A. Bonneson, Werner Brilon, Kenneth G. Courage, Rafael E. DeArazoza, Richard Dowling, Daniel B. Fambro, Ronald K. Giguere, Fred L. Hall, Douglas W. Harwood, Michael Kyte, Joel P. Leisch, Douglas S. McLeod, John Morrall, Barbara Katherine Ostrom, Ronald C. Pfefer, James L. Powell, William R. Reilly, Roger P. Roess, Nagui M. Rouphail, Ronald C. Sonntag, Alex Sorton, Dennis W. Strong, Stan Teply, Pierre-Yves Texier, R. Troutbeck, Thomas Urbanik II, John D. Zegeer*

Committee on Methodology for Evaluating Highway Improvements
*Chairman: Ezra Hauer, University of Toronto*
*Secretary: Warren E. Hughes, Bellomo-McGee, Inc.*
*Karen K. Ajluni, Brian L. Bowman, Charles Philip Brinkman, Julie Anna Cirillo, Barbara Hilger DeLucia, William D. Glauz, Alfred-Shalom Hakkert, Jerome W. Hall, King K. Mak, Thomas L. Maleck, Ernst Meyer, Shaw-Pin Miaou, Olga Pendleton, Ramey O. Rogness, James M. Witkowski*

**Transportation Research Board Staff**
*Robert E. Spicher, Director, Technical Activities*
*Richard A. Cunard, Engineer of Traffic and Operations*
*Nancy A. Ackerman, Director, Reports and Editorial Services*

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1994.

# Transportation Research Record 1484

## Contents

# Foreword

The area of highway capacity is receiving considerable attention because of the 1994 update of the 1985 *Highway Capacity Manual* and the ongoing research effort within the Federal Highway Administration, Federal Transit Administration, and the TRB National Cooperative Highway Research Program, all leading toward the next edition of the *Highway Capacity Manual*, which will be published around 2000.

The highway capacity issues related to interrupted flow facilities are examined in papers on delays at oversaturated unsignalized intersections; capacity of yield-controlled intersections; capacity of two-way stop-controlled intersections; travel times on urban arterial streets; evaluating lane utilization for auxiliary through-lanes at signalized intersections; capacity for right turn on red; effects of trucks on left-turning queues at compressed diamond interchanges; and the development of a safety-based level-of-service criterion for signalized intersections.

Uninterrupted flow facilities are addressed in papers on speed-flow relationships on rural highways; providing acceptable levels of service at toll plazas; percentage time delay for rural highways; assessing the dynamics of freeway flows; and the nature of breakdowns at freeway merge junctions.

The papers presented in this volume were peer-reviewed and were sponsored by the Committee on Highway Capacity and Quality of Service and the Committee on Methodology for Evaluating Highway Improvements.

# Delays at Oversaturated Unsignalized Intersections Based on Reserve Capacities

WERNER BRILON

There is a practical need for estimating average delays at intersections, especially during peak periods, even if a temporary overload must be managed by the intersection. Suitable computation formulas must always be based on approximations. The formulas available are based on the degree of saturation $x$, where $x > 1$ describes the oversaturated situation. The application of $x$ has proven to be successful in the context of signalized intersections. For unsignalized intersections, the reserve capacity $R$ is a more elegant parameter. Here $R < 0$ describes the oversaturated situation. The coordinate transformation technique applied for $R$ to derive average delays during peak periods is explained. The complexity of many influencing parameters and the algebraic solutions make it impossible to solve the complete problem analytically. However, for two levels of approximation, a set of formulas is derived to estimate the average delay during a peak hour for the vehicles on the minor street of an unsignalized intersection. These formulas can be used for practical application.

A variety of formulas are available for estimating delays at intersections. The fact that so many formulas exist emphasizes that none describes the complete reality. Indeed, each of the formulas represents another approximation that focuses on different special capabilities. Table 1 characterizes the more well-known approaches.

## GENERAL QUEUEING MODEL

The basic sophistication of each delay formula is the understanding of traffic operations at an intersection as an analogy to a queueing system. As an example, Figure 1 illustrates this analogy for an unsignalized intersection with only two traffic streams ($q_n$ is the traffic flow on the minor street as the input to the queueing system). Here, the time, $s$, that a vehicle spends at the first position can be regarded as its service time. The time when a vehicle is waiting in higher positions of the queue than 1 can be regarded as the delay, $w$, in the sense of queueing theory. Therefore, the total time that a vehicle spends in the whole queueing system ($d = s + w$) is the delay, $d$, of the vehicle in the sense of traffic engineering.

The problem for traffic engineering is that the type of operation of the service counter cannot be described by one of the classical mathematical solutions from queueing theory, for neither signalized nor unsignalized intersections. Therefore, specific solutions for queueing problems in traffic engineering must be developed.

## ANALYTICAL SOLUTIONS

Analytical solutions have been sought for these service systems, which are established by intersections within traffic systems. A realistic chance for the development of such solutions, however, can be expected only for steady-state situations, in which $q_n$ and capacity $c$ are constant over time and $q_n < c$ for unsignalized intersections. For most analytical solutions, it must also be assumed that each of the traffic streams has Poisson properties (i.e., exponentially distributed headways).

For unsignalized intersections under steady-state conditions, several delay formulas have been proposed. The solution with the highest degree of sophistication appears to be the Kremser solution (*1,2*) in Brilon's formulation (*3*) (Equations 12 and 13), which is based on Yeo's formula (*4*):

$$D = \frac{E(W_1)}{v} + \frac{q_n}{2} \cdot \frac{y \cdot E(W_1^2) + z \cdot E(W_2^2)}{v \cdot y} \quad (1)$$

where

$$v = y + z$$
$$y = 1 - q_n \cdot E(W_2)$$
$$z = q_n \cdot E(W_1)$$

$$E(W_1) = \frac{1}{q_p}\left(e^{q_p t_c} - 1 - q_p t_c\right) + t_f$$

$$E(W_1^2) = \frac{2}{q_p}\left(e^{q_p t_c} - 1 - q_p t_c\right) \cdot \left(\frac{e^{q_p t_c}}{q_p} + t_f - t_c\right) + t_f^2 - t_c^2$$

$$E(W_2) = \frac{e^{q_p t_c}}{q_p}\left(1 - e^{-q_p t_f}\right)$$

$$E(W_2^2) = \frac{2 \cdot e^{q_p t_c}}{q_p^2}\left(e^{q_p t_c} - q_p t_c\right)\left(1 - e^{-q_p t_f}\right) - q_p t_f \cdot e^{-q_p t_f}$$

where $t_c$ equals critical gap, in seconds, and $t_f$ equals follow-up time, in seconds.

Here, however, the expressions for $E(W_2)$ and $E(W_2^2)$ are correct only if $t_c = t_f$ (*5*). For the more realistic situation of $t_c > t_f$, the improved expressions by Daganzo (*6*) or, even better, by Poeschl (*7*) could be used.

## TIME-DEPENDENT SOLUTIONS BASED ON DEGREE OF SATURATION

From the complexity of the equations mentioned before, it is evident that they are not useful for practical application. Moreover, steady-state situations are not realistic in road traffic operations. Instead, the input flows to street intersections fluctuate over the time of day. For instance, Figure 2 illustrates an example for a typical

Ruhr University, D-44780 Bochum, Germany.

**TABLE 1 Characterization of Delay Formulas for Unsignalized Intersections**

| | steady state | Time dependence with temporary oversaturation | |
|---|---|---|---|
| | | $q_0 = 0$ ; $q_1 = 0$ | $q_0 > 0$ ; $q_1 > 0$ |
| Combination of Kremser, Yeo, Brilon, Poeschl (eq. 1) | • | | |
| HCM 1994 *(13)* cf. (eq. 2) | | • | |
| Kimber, Hollis *(11)* | | | • |

workday. Here the peak hour is particularly important for the layout of the intersection and its control. Therefore, the average delay and its tolerable maximum during the peak hour determine the whole design procedure. Hence, it is most important to describe the quality of traffic operations at intersections for a peak period with sufficient precision. By the definition of a peak period, it is clear that lower input traffic flows exist before and after the peak.

As solutions for this problem at signalized intersections with fixed cycle times and greens, the well-known formula by Akçelik (*8,9*) or the formula by Wu (*10*) can be used. For unsignalized intersections the user can choose between the formula by Kimber and Hollis (*11*) and that by Akçelik and Troutbeck (*12*), which is also recommended by the new Chapter 10 of the *Highway Capacity Manual* (HCM) (*13*). The derivation of both formulas is based on $x$ = degree of saturation. For Akçelik and Troutbeck's solution (*12*, Equation 1.2), this is obvious.

The Kimber and Hollis formula (*11*) has proven to be especially useful to estimate delays, and it appears to have a reliable background, particularly for temporarily oversaturated conditions. The formula and its derivation are supposed to be based on the degree of saturation. Unfortunately, the derivation is not published anywhere, which is a remarkable drawback. It also has not been possible to redetermine the formula. From its description it can be obtained that the capacity $c_0$ and the input flow $q_0$, before and after the peak period, are taken into account.

The delay formula by Akçelik and Troutbeck (*12*) is given in modified version by the following equation:

$$d = \frac{1}{c} + \frac{T}{4}\left[(x-1) + \sqrt{(x-1)^2 + \frac{8 \cdot x}{c \cdot T}}\right] \quad (2)$$

where

$d$ = average delay (sec),
$T$ = duration of peak period (sec),
$x$ = degree of saturation = $q/c$,
$q = q_n$ = minor street traffic flow (veh/sec), and
$c$ = capacity (veh/sec).

The derivation of the formula can be found elsewhere (*14*).

The capacity, $c$, can be obtained from any useful capacity formula—for example, the formula according to Siegloch (*15*), which is also the basis of the HCM procedure (*13*):

$$\mu = \frac{3{,}600}{t_f} \cdot e^{-(t_c - 0.5 \cdot t_f) \cdot \frac{q_p}{3{,}600}} \quad (3)$$

where $\mu$ is given in vehicles per hour.

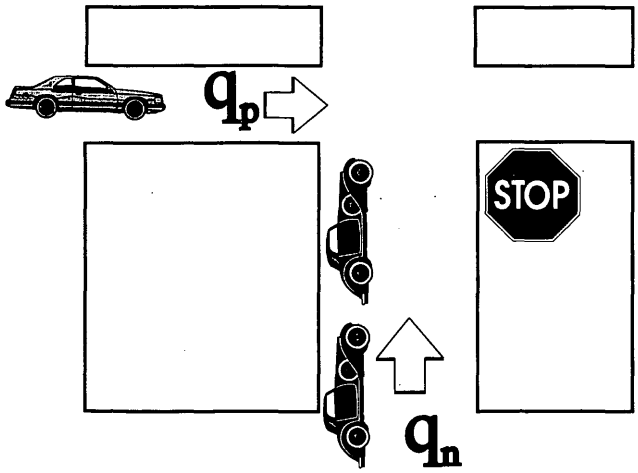



**FIGURE 1 General outline of queueing system that represents unsignalized intersection with one minor traffic stream (volume $q_n$) and one priority stream (volume $q_p$).**

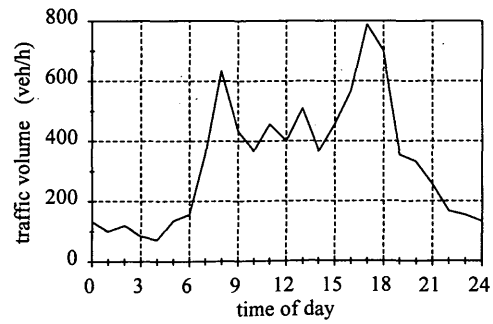



**FIGURE 2 Traffic flow pattern over time for typical workday on street with predominating afternoon peak hour.**

Equation 2 has one great drawback: it assumes that the traffic flow $q_n$ before and after the peak period of duration $T$ is 0, which is unrealistic. A set of formulas based on the same sophistication—including, however, positive traffic flows before and after the peak period—is described by Troutbeck (*14*).

The results from Equation 2 formulated in dependence on the degree of saturation $x$ can be obtained from Figure 3 as one example. This example has been prepared for $T$ = duration of the peak period = 3,600 sec. From the graph it can be seen that the curves are scattered over a very wide range if the capacity is altered.

As a counterpoint to this, Figure 4 should be compared. Here, the same relations are illustrated, but the reserve capacity $R = c - q_n$ is used as the independent variable. Again, the parameter of the curves is the capacity. All curves are nearly coinciding, and the parameter is not of high importance, especially in the range of delays that is useful in practice ($d < 60$ sec). Thus, the use of the reserve capacity, $R$, appears to facilitate some of the interrelations within the theory of unsignalized intersections. This fundamental idea originates from Harders' work (*16*). He was the first to find that the reserve capacity $R$ is a strong determinant for the traffic flow quality at unsignalized intersections. This was the reason that $R$ was used as the measure of effectiveness in Chapter 10 of the 1985 HCM (*17*), which was a realization of Harders' concept.

## CONCEPT OF RESERVE CAPACITY

The concept introduced in this paper tries to develop formulas for the average delay at an unsignalized intersection, only based on the reserve capacity $R$:

$$R = c - q \quad \text{(veh/sec)} \quad (4)$$

where $q$ is the traffic volume of the movement under observation in vehicles per second, and $R$ is given in vehicles per second. Here especially, the approximation of peak periods including temporary oversaturation (i.e., $R < 0$) should be solved.

### Simplification of Flow Pattern and D/D/1 System

Consider a queueing system with two traffic streams as in Figure 1. The minor street traffic flow, $q_n$, is the input flow to the system. For simplicity, it is now called $q$. Assume that the capacity of the system can be calculated from $q_p$ by any useful capacity equation; for example, the Siegloch formula (Equation 3).

Look at a traffic flow pattern over time as illustrated in Figure 5 (*top*). The variables being used are explained in Table 2, and $T$ equals the duration of peak period (e.g., $T$ = 3,600 sec). Of course, in reality, the headways of vehicles entering the system are distributed randomly. However, for further simplification in the case of oversaturation (i.e., $R < 0$), imagine—as a preliminary approach to the solution—that during the peak period of duration $T$, the queueing system is operating like a D/D/1 queueing system, in which arrival headways ($a$) of the minor street vehicles and their departure

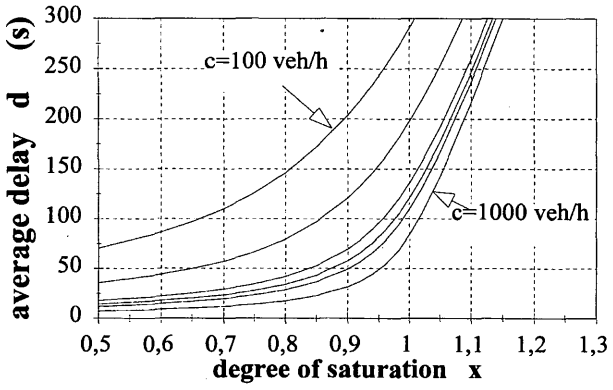



**FIGURE 3 Average delay of vehicles in front of stop line calculated from Equation 2 depending on $x$. Different lines apply to different capacities $c$ (100, 200, 300, 400, 500, 600, and 1,000 veh/hr); here $T$ is given as 1 hr.**

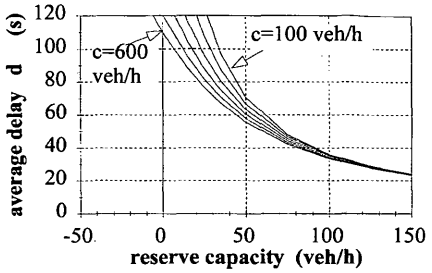



**FIGURE 4 Average delay of vehicles in front of stop line calculated from Equation 2 depending on $R$. Different lines apply to different capacities $c$ (100, 200, 300, 400, 500, and 600 veh/hr); here $T$ is given as 1 hr.**

Look at a traffic flow pattern over time as illustrated in Figure 5 (*top*). The variables being used are explained in Table 2, and $T$ equals the duration of peak period (e.g., $T$ = 3,600 sec). Of course, in reality, the headways of vehicles entering the system are distributed randomly. However, for further simplification in the case of oversaturation (i.e., $R < 0$), imagine—as a preliminary approach to the solution—that during the peak period of duration $T$, the queueing system is operating like a D/D/1 queueing system, in which arrival headways ($a$) of the minor street vehicles and their departure

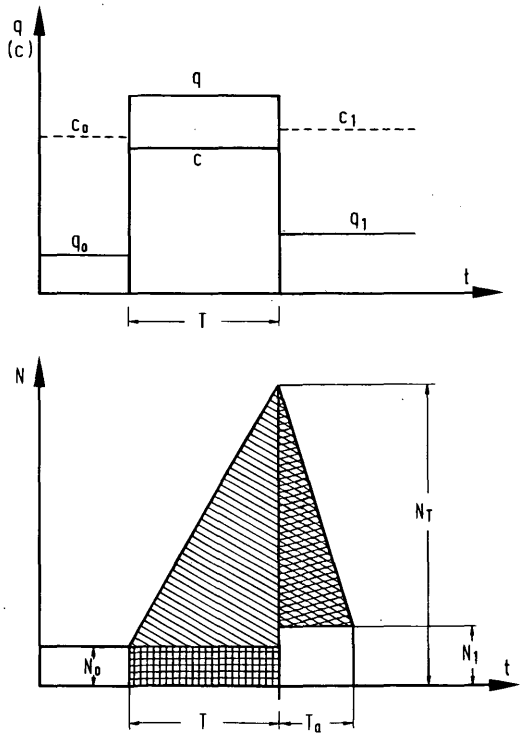



**FIGURE 5 Flow patterns over time including peak period: *top,* traffic flow $q$ and capacity $c$; *bottom,* queue length $N$.**

**TABLE 2 Variables for Traffic Flow Pattern**

| | before | during | after |
|---|---|---|---|
| | the peak period | | |
| traffic demand on the minor street | $q_0$ $q_0 < q$ | q | $q_1$ $q1 < q$ |
| capacity | $c_0$ | c | $c_1$ |
| reserve capacity | $R_0 = c_0 - q_0$ $R_0 > 0$ | $R = c - q$ $R > 0$ or $R \leq 0$ | $R_1 = c_1 - q_1$ $R_1 > 0$ |
| queue length | $N_0$ | N | $N_1$ |

All variables for q, c, and R are given in vehicles per second.

headways ($b$) from the stop line both are constant for all vehicles. For such a system, it can easily be imagined that for $R > 0$ the queue length and the delay both are 0.

However, for the time of temporary oversaturation with $R < 0$, queue length is constantly increasing (Figure 5, *bottom*). At the end of the oversaturated peak period, the queue length is $N_T$, with

$$N_T = (q - c) \cdot T + N_0$$
$$N_T = N_0 - R \cdot T \tag{5}$$

where $R = c - q$ has a negative value.

The time needed to clear the queue down to $N_1$ after the peak is

$$T_a = \frac{N_T - N_1}{c_1 - q_1}$$

$$T_a = \frac{N_T - N_1}{R_1} \tag{6}$$

where $N_1$ is the expectation of the queue length after the peak, on the assumption that no overload was observed within the peak period. Thus, $N_1$ is only a result of $q_1$ and $c_1$ (with $c_1 > q_1$) without regard to the results of the peak period.

In each queueing system, as a general property, the sum of all delays is the area under the function of the queue length. Before this basic idea can be applied, the type of average to be used must be defined. According to most of the authors mentioned earlier, the shaded area in Figure 5 (*bottom*) represents the sum of all delays induced in the system by the vehicles arriving during the peak period. Thus the sum $S$ of all delays is

$$S = N_0 \cdot T + \frac{(N_0 - N_1)^2}{2 \cdot R_1} - R \cdot \left( \frac{N_0 - N_1}{R_1} \cdot T + \frac{T^2}{2} \right)$$
$$+ \frac{T^2}{2 \cdot R_1} \cdot R^2 \tag{7}$$

The average delay, $d$, per vehicle caused by those vehicles arriving during the peak period, then, is

$$d = \frac{S}{q \cdot T} \tag{8}$$

This delay $d$ already includes the time spent in the service position (first position of the queue on the minor street) of the queue, because the vehicle in this position already has been included in the queue length according to Figure 5 (*bottom*). Therefore, $d$ is a representation of the delay in the sense of traffic engineering.

Equation 8 gives the delay, $d$, for the D/D/1 system. From alternating the parameters $N_0$, $N_1$, and $R_1$ by a series of sample calculations it is learned that the delay curve is a straight line for $R_1 = c$. For $R_1 < c$, the curve becomes concave (concave side above the curve). For $R_1 > c$, the curve would become convex. However, $R_1 > c$ is not a reasonable case, because it would be unusual that the reserve capacity, $R_1$, could become even greater after the peak than the peak-hour capacity, $c$. The function for $d$ is not very sensitive to $N_1$ as far as $N_1$ is varied over a reasonable range of values. Therefore, for practical cases, it could be sufficient to assume that $N_0 = N_1$ (see Case S1 later in the paper).

Equation 7, which is also a part of Equation 8, looks rather complicated. Hence, for better understanding, we also look at simplified special cases.

*Simplification Case S0*

The most simplified case is the one corresponding to the assumptions of the formula of Akçelik and Troutbeck (Equation 2):

$$c_0 = c = c_1$$
$$q_0 = 0 = q_1$$
$$N_0 = 0 = N_1$$
$$R_0 = c = R_1$$

For these conditions, Equation 7 can be written as

$$S = -R \cdot \frac{T^2}{2} \cdot \frac{q}{c} \tag{9}$$

Equation 8 then can be expressed as

$$d = -\frac{R \cdot T}{2 \cdot c} \tag{10}$$

Some results for $d$ as a function of $R$ are shown in Figure 6. Here it is clear that $d$ has a linear relationship to $R$, where the gradient depends on $T$ and $c$. The solution of Equation 10 toward $R$ is

$$R = -\frac{2 \cdot c \cdot d}{T} \tag{11}$$

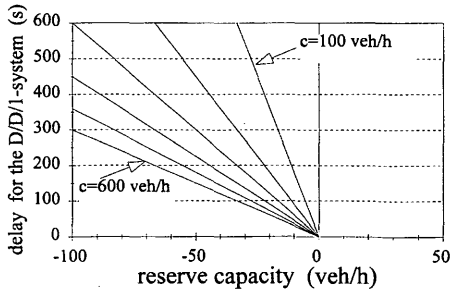



**FIGURE 6 Average delay for D/D/1 system as function of reserve capacity, $R$, during peak period for Case S0. $T$ has been fixed to 1 hr; $c$ is capacity of system.**

*Simplification Case S1*

A more general and realistic simplified case is to assume that the average queue lengths before and after the peak are of the same size:

$N_0 = N_1$
$c_0 = c = c_1$
$R_1 < c$

This case comes very close to reality, because evaluations of Equation 7 showed that $N_1$ has only a very minor influence on the result for the average delay (see previous discussion).

Under this assumption, Equation 7 becomes

$$S = N_0 \cdot T - R \cdot \frac{T^2}{2} + \frac{T^2}{2 \cdot R_1} \cdot R^2 \tag{12}$$

Equation 8 in this case can be written as

$$d = \frac{1}{c - R}\left[N_0 - \frac{R \cdot T}{2}\left(1 - \frac{R}{R_1}\right)\right] \tag{13}$$

Because the function is quadratic in $R$, it also has positive $d$ values for a large $R$. However, only the part for negative $R$ is of interest in this context. A solution of Equation 13 toward $R$ is possible:

$$R = \frac{R_1}{2 \cdot T} \cdot \left[A - \sqrt{A^2 + \frac{8 \cdot T}{R_1} \cdot (d \cdot c - N_0)}\right] \tag{14}$$

where $A = T - 2 \cdot d$.

But this solution cannot be used for further derivations because it leads to equations that cannot be solved (explained later). Therefore, at the moment, this solution is obsolete. Instead, it turns out that a simplified approximation of this equation is needed. To achieve this, approximate the curves for negative $R$ values by straight lines that have the same gradient $b$ for all $N_0$ values. This is possible with good approximation. More numerical evaluations of Equation 13 showed that this reduction to a uniform gradient of the curves does not cause much bias. This gradient, $b$, is given by the following equation, which is an application of Equation 13.

$$b = \left\{\frac{1}{c - R_f}\left[N_0 - \frac{R_f \cdot T}{2}\left(1 - \frac{R_f}{R_1}\right)\right] - \frac{N_0}{c}\right\} \cdot \frac{1}{|R_f|} \tag{15}$$

$R_f$ is an arbitrary point along the function of Equation 13 for $R << 0$, where the original function of Equation 13 should be met exactly by the linear approximation. Further derivations show that $R_f$ should be chosen in accordance with the other parameters, mainly the peak-hour duration $T$ and the reserve capacity $R_1$ after the peak. For application, the following is recommended:

$$R_f = \frac{100 \cdot 3{,}600}{T}$$

Then the approximation for Equation 13 is

$$d = \frac{N_0}{c} - b \cdot R \tag{16}$$

The solution toward $R$ is

$$R = \frac{1}{b} \cdot \left(\frac{N_0}{c} - d\right) \tag{17}$$

**Approximation for Steady-State Solution**

As pointed out earlier one could indicate an analytical solution for the delay in the steady-state queueing system that is established by an unsignalized intersection (Figure 1). This solution, however, is so complicated that it is not useful for further derivations. It has turned out in many investigations (*11,18*) that the M/M/1 queue is a very close approximation for an unsignalized intersection. In the M/M/1 queue, the total time that a customer spends in the queueing system is

$$d = \frac{1}{R} \tag{18}$$

This approach is used as an approximation of the delay at an unsignalized intersection in the steady state (i.e., $R >> 0$). The curve only can be used for $0 < R \leq c$. Equation 18 can be solved for $R$ as

$$R_s = \frac{1}{d} \tag{19}$$

where the index $s$ stands for steady state.

The M/M/1 queue also can be used, with rough approximation, to estimate the average queue length for steady-state conditions. The expectation for the number of vehicles in the system, then, is

$$N_0 = \frac{q_0}{R_0} = \frac{c_0 - R_0}{R_0}$$
$$N_1 = \frac{q_1}{R_1} = \frac{c_1 - R_1}{R_1} \tag{20}$$

This solution is assumed to apply for the periods before and after the peak when $R_0$ and $R_1$ are considerably larger than 0.

**Coordinate Transformation**

For longer peak periods or small $R$ values (e.g., $R << 0$), the delay in any type of queueing system tends toward the D/D/1 delay. Then the details of the arrival and departure process will be of less importance. The dominating property of the queueing system, then, is the tremendous increase of the queue during the oversaturation period. Therefore, the real delay must be found along a transition curve that connects the steady-state delay curve for $R >> 0$ with the D/D/1 delay for $R << 0$. This transition curve is illustrated in Figure 7. The equation for this curve cannot be derived analytically; again, only an approximation can be derived. A reasonable approach to the derivation of this approximation is to assume that $y$ is equal to $z$ (Figure 7). This is identical to

$$R_s = -R_D + R \tag{21}$$

where

$R_s$ = reserve capacity that causes average delay $d$ in steady-state system (Equation 19),

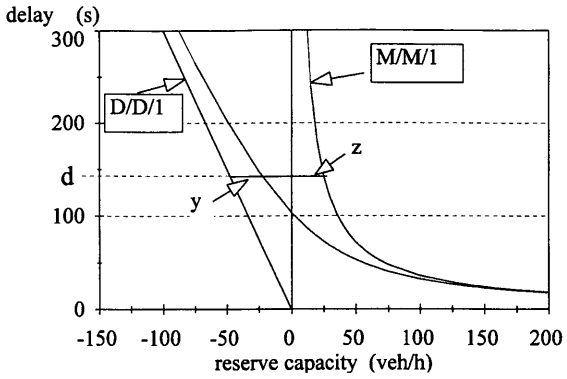



**FIGURE 7 Principle of coordinate transformation.**

$R_D$ = reserve capacity that causes average delay $d$ in D/D/1 system (solution of Equation 8 toward $R$ and simplified Equations 11 and 14 or 17, respectively), and
$R$ = actual reserve capacity during peak.

The expressions for both $R_s$ and $R_D$ include the delay, $d$. Therefore, Equation 21 establishes a relation between the delay, $d$, and the reserve capacity, $R$. The problem now is to solve Equation 21 toward the solution $d$ = function ($R$). It is not easy to solve this function by algebra for a more general case. Moreover, it can be shown that for the general case (Equation 7) this solution is not possible because it results in imaginary equations containing expressions for $\sqrt{-1}$. Therefore, one is forced to focus on the simplified cases S0 and S1.

### Case S0

Concentrate first on Case S0 to understand the method. Enter Equation 11 for $R_D$ and Equation 19 for $R_s$ into Equation 21, which results in

$$\frac{1}{d} = R + \frac{2 \cdot c \cdot d}{T} \tag{22}$$

The solution toward $d$ gives

$$d = -\frac{1}{4 \cdot c} \cdot \left[R \cdot T - \sqrt{(R \cdot T)^2 + 8 \cdot c \cdot T}\right] \tag{23}$$

An illustration of this formula is given in Figure 7 by the transition line for $T$ = 1 hr and $c$ = 600 veh/hr. To repeat: Equation 23 estimates the average delay at an unsignalized intersection with the same degree of approximation as Equation 2, which is based on a similar derivation using $x$ (degree of saturation) instead of the reserve capacity, $R$. The capacity, $c$, can be estimated by any useful type of formula [e.g., the Siegloch formula (Equation 3)]. Capacity formulas based on the empirical regression method (*14*) or any other method can be used. The new Equation 23 is a little bit shorter and thus easier for practical use than Equation 2.

### Case S1

To find a solution for Case S1, again the coordinate transformation technique must be used (Equation 21). For $R_s$, enter Equation 19, and for $R_D$, use Equation 17. Thus,

$$\frac{1}{d} = R - \frac{N_0}{b \cdot c} + \frac{d}{b} \tag{24}$$

with $b$ according to Equation 15.

The solution of this equation toward $d$ gives

$$d = -B + \sqrt{B^2 + b} \tag{25}$$

where $B = {}^1/_2 \cdot [b \cdot R - (N_0/c)]$. The result of this equation is illustrated in Figure 8. Equation 25 gives an estimation of the average delay during a peak period of duration $T$ when an initial queue of length $N_0$ exists at the beginning of the peak interval. Moreover, the reserve capacity $R_1$ after the peak is included via $b$ and Equation 15. The important independent parameter, however, is the reserve capacity, $R$, during the peak period.

One might argue that on the way to this result, many approximations were made. This, however, also is the case for each of the alternative approaches to solving the peak-hour delay problem. Thus, at the moment, Equation 25 is the most detailed formula for average delay at an unsignalized intersection for times of temporary oversaturation that can be recommended as a result of these derivations.

Of course, it would be desirable to use the more exact solution for $R_D$ (Equation 14) in the coordinate transformation (Equation 21). This, however, turns out to be impossible. The result would be a transcendental equation using imaginary numbers (containing $\sqrt{-1}$) as part of the result.

## DISCUSSION OF RESULTS

It would be even more desirable to enter the complete and general solution for the D/D/1 delay given by Equations 7 and 8 into the coordinate transformation technique. This attempt also results in transcendental equations. Thus, this most complete solution also is

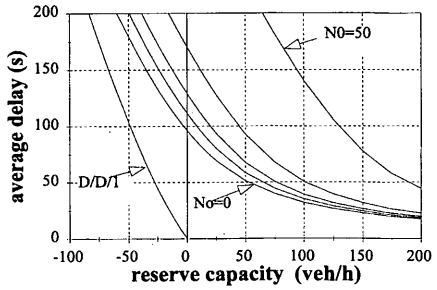



**FIGURE 8 Average delay as function of reserve capacity, $R$, during peak period of duration $T$ = 1 hr calculated from Equation 25 for Case S1 for $T$ = 1 hr, $R_1$ = 300 veh/hr, and $c$ = 600 veh/hr. Parameters of curves are $N_0$ = 0, 5, 10, 20, and 50 veh. Complete D/D/1 solution shown in figure is obtained from Equation 13.**

not possible now. To find a solution to that problem, a more general approximation to Equation 7 that does not have two roots for $R$ = function ($d$) (e.g., an exponential function) must be found. This would be useful only if all the parameters, $N_0$, $N_1$, $R_1$, $T$, and $R$, could be included in a realistic way.

Another possible improvement should be considered further: the attempt to allow a time dependency also within the peak period itself (e.g., a parabolic pattern could be used). For signalized intersections, this has been solved by Wu (*10*) on the basis of the standard philosophy of degree of saturation. For the unsignalized intersections, this task might have a chance to be solved together with the approximation mentioned before, as some preliminary numerical sample calculations showed.

Moreover, the same approach that has been developed here for unsignalized intersections could be used for signalized intersections, too. According to Kimber and Hollis (*11*) and subsequent publications by Kimber, one could try to model the signalized intersection delay by an M/D/1 queueing system. The time customers are in the system is

$$d = \frac{1}{2 \cdot c}\left(\frac{c}{R} + 1\right) \tag{26}$$

This equation also could be solved toward $R$. Nevertheless, compared with Equation 18, the same technique applied to signalized intersections promises to reveal more complications. Therefore, one could say that the degree of saturation, $x$, is a suitable parameter to describe signalized intersection performance. The reserve capacity, $R$, is a more suitable parameter for unsignalized intersections.

Finally, the number of possible approximative solutions for the peak-hour delay problem appears to be unlimited. Therefore, for the user it is of greatest importance to understand the sophistication of each of the provided solutions. The numerical results of these theoretically equivalent solutions could make a decisive difference, especially in situations with large overloads of the intersection. Thus, a confirmation of the validity of the solutions by either simulation studies or empirical evaluations would be desirable, a task that will initiated at the author's institute.

## CONCLUSIONS

The paper presents another approach for estimating average delays of minor street vehicles at an unsignalized intersection for oversaturated and nearly oversaturated peak periods. The derivations point out that the reserve capacity, $R$, is useful for application as the independent parameter to describe traffic performance. For the computation of average delays, Equation 23 can be used as a rough approach. It describes the delay problem with the same degree of sophistication as the delay equation of Chapter 10 in the latest edition of the HCM (*13*). However, it describes the average delay with considerable simplifications.

A more realistic solution is given by Equation 25 (including Equation 15 for $b$). This solution takes into account that before and after the peak period, only limited capacity reserves are available. Thus, the equation can be recommended for practical use. Of course, the whole set of equations is quite lengthy, as are all other solutions presented. However, in computer programs it is not a problem to apply this set of equations. The overall quality of the solution might be comparable to that of the solution of Kimber and Hollis (*11*). However, the derivation also is given here. To improve the applicability of the solution, the whole set of formulas is repeated step by step in the appendix.

# APPENDIX
## Steps for Applying Equation 25

It is assumed that all values for the variables mentioned in Table 2 are given, including the duration $T$ of the peak period. For practical cases, $T$ should be at least 15 min (900 sec). It is also assumed that the steady-state queue length before and after the peak period are nearly equal with sufficient approximation ($N_1 \approx N_0$). All variables in these equations should be used in units of seconds, number of vehicles (veh), and vehicles per second.

The average queue length during the time before the peak is

$$N_0 = \frac{q_0}{R_o} = \frac{c_0 - R_0}{R_0} \tag{20}$$

Then the sequence of the following equation must be applied:

$$R_f = -\frac{100 \cdot 3{,}600}{T}$$

$$b = \left\{\frac{1}{c - R_f}\left[N_0 - \frac{R_f \cdot T}{2}\left(1 - \frac{R_f}{R_1}\right)\right] - \frac{N_0}{c}\right\} \cdot \frac{1}{|R_f|} \tag{15}$$

$$B = \frac{1}{2} \cdot \left(b \cdot R - \frac{N_0}{c}\right)$$

$$d = -B + \sqrt{B^2 + b} \tag{25}$$

where $d$ is the average delay for vehicles arriving during the peak period.

The maximum of the average queue length must be expected at the end of the peak period (Figure 5, *bottom*). The expectation for this queue length at the end of an oversaturated peak period is

$$N_T = \max\begin{Bmatrix} N_0 - R \cdot T \\ 0 \end{Bmatrix} \tag{5}$$

On average, vehicles arriving at the end of the oversaturated peak period must face the longest delays. The expectation for their delay is

$$T_a = \frac{N_T - N_0}{R_1} \tag{6}$$

$N_T$ and $T_a$, in addition, are subject to random variation, which is not described in this paper.

## REFERENCES

1. Kremser, H. Ein einfaches Wartezeitproblem bei einem poissonschen Verkehrsfluss (A Simple Problem of Delay with a Poisson-Distributed Traffic Flow). *Oesterreichisches Ingenieur-Archiv*, No. 16, 1962.
2. Kremser, H. Ein zusammengesetztes Wartezeitproblem bei einem poissonschen Verkehrsfluss (A Composed Problem of Delay with a Poisson-Distributed Traffic Flow). *Oesterreichisches Ingenieur-Archiv*, No. 16, 1962.

3. Brilon, W. Recent Developments in Calculation Methods for Unsignalized Intersections in West Germany. In *Intersections Without traffic signals,* Springer-Verlag, Berlin, German, 1988.
4. Yeo, G. F. Single-Server Queues with Modified Service Mechanisms. *Journal of the Australian Mathematical Society,* Vol. 2, 1962, pp. 499–507.
5. Kremser, H. Wartezeiten und Warteschlangen bei Einfaedelung eines Poissonprozesses in einen anderen solchen Prozess (Delays and Queues with One Poisson Process Merging into Another One). *Oesterreichisches Ingenieur-Archiv,* No. 18, 1964.
6. Daganzo, C. F. Traffic Delay at Unsignalized Intersections: Clarification of Some Issues. *Transportation Science,* Vol. 11, 1977.
7. Poeschl, F. J. Die nicht signalgesteuerte Nebenstrassenzufahrt als verallgemeinertes M/G/1-Warteschlangensystem (The Unsignalized Minor Street Entry as a Generalized M/G/1 Queueing System). *Zeitschrift für Operations Research,* Vol. 27 B, 1983.
8. Akçelik, R. *Traffic Signals—Capacity and Timing Analysis.* Australian Road Research, Report 123. Australian Road Research Board. Victoria, 1981.
9. Akçelik, R. The Highway Capacity Manual Formula for Signalised Intersections. *ITE Journal,* March 1988.
10. Wu, N. *Wartezeiten und Leistungsfaehigkeiten von Lichtsignalanlagen unter Beruecksichtigung von Instationaritaet und Teilgebundenheit* (Delay and Capacity of Traffic Signals under Instationary and Partly Constraint Conditions). Dissertation. Institute for Traffic Engineering, Ruhr-University Bochum, Germany, 1990.
11. Kimber, R. M, and E. M. Hollis. *Traffic Queues and Delays at Road Junctions.* Report LR 909. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1979.
12. Akçelik, R., and R. Troutbeck. Implementation of the Australian Roundabaout Analysis Method in SIDRA. In *Highway Capacity and Level of Service.* (U. Brannolte, ed.), A. A. Balkema, Rotterdam, The Netherlands, 1991.
13. *Special Report 209: Highway Capacity Manual,* 3rd ed. TRB, National Research Council, Washington, D.C., 1994, Chapter 10.
14. Brilon, W., R. Troutbeck, and M. Tracz. Review of International Practices Used To Evaluate Unsignalized Intersections. *Transportation Research Circular.* TRB, National Research Council, Washington, D.C., in preparation.
15. Siegloch, W. Die Leistungsermittlung an Knotenpunkten ohne Lichtsignalsteuerung (Capacity Calculations for Unsignalized Intersections). *Strassenbau und Strassenverkehrstechnik,* No. 154, 1973.
16. Harders, J. Die Leistungsfaehigkeit nicht signalgeregelter staedtischer Verkehrsknoten (The Capacity of Unsignalized Urban Intersections). *Strassenbau und Strassenverkehrstechnik,* No. 76, 1976.
17. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985, Chapter 10.
18. Wu, N. An approximation for the Distribution of Queue Lengths at Unsignalized Intersections. *Proc., 2nd International Symposium on Highway Capacity,* Australian Road Research Board, Sydney, 1994.

*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Capacity of One-Way Yield-Controlled Intersections

HASHEM R. AL-MASAEID

An empirical model for estimating capacity of yield-controlled streams at a one-way minor street crossing a one-way major street was developed. For Jordan conditions, results of the empirical and gap acceptance models were compared. Data were collected from different cities in Jordan. The data consisted of 854 min of at-capacity operation and included both geometric and traffic characteristics. Also, for comparison purposes, data on critical gap and move-up time were collected. The results of analysis indicated that major traffic flow, visibility-to-speed ratio, and widths of the major and minor street had a significant effect on the capacity of each minor stream. For each minor stream, the results indicated that the traffic speed and the width of the major street significantly influenced the size of the critical gap. On the basis of field observations and results presented in this paper, gap acceptance models would significantly overestimate the capacity at low major traffic levels and underestimate at high levels. In addition, the results indicated that gap-acceptance models would provide unrealistic capacity values even if critical gaps are estimated for conditions in Jordan. Finally, a set of equations and figures were presented for practical applications.

One of the most important tasks of a traffic engineer is to estimate the capacity of unsignalized intersections. In Jordan traffic engineers have been using the 1985 *Highway Capacity Manual* (HCM) (*1*) in estimating the capacity of such intersections. Unfortunately, even if the critical gap is estimated for the local condition, the estimated reserve capacity is very high under low major traffic demand. On other hand, the use of the empirical models developed by Kimber and Coombe (*2*) is restricted to T-intersections. Furthermore, driver behavior and operating rules may affect the capacity estimation. For these reasons, an effort was made to estimate the capacity of one-way minor-major street intersections.

In this study an effort was made to estimate the capacity of yield-controlled streams at a one-way street crossing a one-way major street. This type of intersection is used widely to improve capacity and safety in urban areas. Stop-controlled intersections were excluded because of the wide variability in stop sign compliance among Jordanian drivers.

The empirical approach using multiple regression analysis was adopted to estimate capacity of minor street movements and investigate the effects of different geometric design variables on the estimated capacity. For each minor street movement, the critical gap was estimated and used to calculate capacity on the basis of the 1985 HCM procedure. The results of these approaches were compared and presented in this paper.

Civil Engineering Department, Jordan University of Science and Technology, P.O. Box 3030, Irbid-Jordan.

## BACKGROUND

Chapter 10 of the 1985 HCM contains a procedure for estimating capacity and level of service at unsignalized intersections. The procedure is based on a German guideline developed in 1972, which was based mainly on Harders' formula (*3*). Different problems and limitations with the 1972 guideline and Chapter 10 of the 1985 HCM are cited in the literature (*4,5*) such as difference to simple Poisson model, concept and size of the critical gap, impedance factors, and reserve capacity. Furthermore, the German guideline was developed only for a single lane for each movement. According to German practice, multilane approaches are always signalized for safety purposes. Although a series of research was started to develop a new German guideline for practical application, the simple Poisson model will still be the basis of the future German guideline (*4*). Instead of using Harders' formula to estimate capacity, the new guideline will use Siegloch's formula.

On the other hand, the Transport and Road Research Laboratory has made extensive studies to estimate capacity of nonpriority streams in the United Kingdom (*2*). The developed empirical models indicated that the capacities of the nonpriority streams at T-intersections depend linearly on the flow in the relevant priority streams. The developed relationships depend on the lane width available to the nonpriority stream, visibility to waiting drivers, and width of the major street. Kimber (*6*) compared the performance of the simple gap acceptance and the empirical models. He concluded that the simple gap acceptance models are poor predictors of the capacity of nonpriority traffic streams in the United Kingdom, for they seriously overpredict at low values of priority flow and underpredict at high values.

## METHODOLOGY

A number of one-way major-minor yield-controlled intersections were selected. The following criteria were adopted to determine the suitability of a given intersection for the purpose of this study:

1. The intersection operates at capacity during peak flow conditions. This criterion is achieved if a stable queue of vehicles is observed in the minor street. Furthermore, the intersection operates at capacity under low and heavy major traffic flow. This condition is necessary to develop a relationship that covers a wide range of major flow conditions.
2. The selected intersections should have different geometric design variables. This criterion was adopted to investigate the effects of the geometric design variables on the estimated capacity.
3. The selected intersections should be located in different cities to represent a wide range of driver populations and environmental conditions.

In this study, an empirical approach employing multiple regression analysis was used to estimate capacities of minor streams and identify variables that affect these capacities. The relationship between capacity of each minor stream and the major traffic flow was investigated from intersection-specific data. Once the basic form of this relationship was identified, the effects of geometric variables were included in the analysis.

For comparison, the same approach was used to develop critical gap and move-up time models for Jordanian drivers. The developed models were used to estimate the capacity of the minor stream using the gap acceptance approach (the 1985 HCM procedure).

## DATA COLLECTION

In this study, two independent sets of data were collected. The first set was collected to develop an empirical relationship for estimating capacities of minor streams. The second set was collected to estimate the critical gap and move-up time for different minor streams. The data were collected from cities in Jordan including Amman, Irbid, Zarqa, and Mafraq. The data were collected during the summers of 1992 and 1993. Three forms of one-way yield-controlled intersections were investigated; they are shown in Figure 1. The selected intersections are located in urban or suburban areas.

For the first set, data were collected using manual techniques. Similar to previous studies (*1*), capacity data for each minor stream were observed separately. The observations were taken with stable queueing in the minor stream. Capacity of the minor stream and flow in the major street (total approach volume) were observed at 1-min intervals. Field observation revealed that no separate lane was available for turning movements in the major street. In most cases, lane markings were not provided on major and minor street approaches.

To account for heavy vehicles, the capacity and major traffic flow were expressed in passenger car units (PCU). For conversion into PCU, single-unit trucks were rated as 1.5 PCU, other trucks and trailers as 2.0 PCU, and motorbikes as 0.5 PCU. These values are considered satisfactory in different studies (*7,8*). The traffic speed on the approach of the major street was measured over a trap and the average value was computed for each interval. Tables 1, 2, and 3 present the data characteristics for Intersection Forms 1, 2, and 3, respectively.

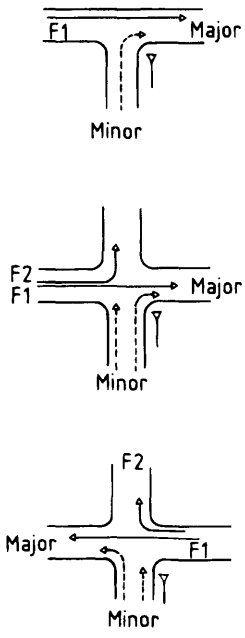



**FIGURE 1 Forms of one-way major-minor street intersections: *top,* Form 1; *middle,* Form 2; *bottom,* Form 3.**

For the second data set, a number of one-way, major-minor yield-controlled intersections were selected. A manual technique was used to collect gap and moving-up time data. Gaps, move-up time, and speed measurements were carried out by two observers equipped with stopwatches. In this study, no distinction was made between accepted gap and lag. A lag (first gap) is defined as the difference between the time at which a minor street vehicle arrives at the intersection and the time at which the next conflicting major street vehicle arrives. The move-up time represents the minimum headway between vehicles in a minor stream entering the same major stream gap. For each minor stream, the accepted gaps and move-up time were observed for 15-min intervals during peak and off-peak periods. Peak and off-peak periods were chosen to achieve variability in the major traffic flow. The collected data were used to determine the critical gap and the average value of the move-up time. The critical gap is defined as the median gap size accepted by minor street drivers. The measurements also included flow on the major traffic and the geometric variables of the major and minor streets.

## DEVELOPMENT OF CAPACITY MODELS

The main purpose of capacity modeling is to develop useful relationships between capacity of the minor stream and a set of major traffic and intersection geometric variables. The developed models should be easy for practical applications and sensitive to alternative policies and design. In the analysis, the relationship between minor stream capacity and major traffic flow was investigated from intersection-specific data. For all forms of the investigated intersections, the relationship cannot be considered linear. For illustration, Figure 2 presents the scatter plot of right-turn capacity and major traffic flow for a T-intersection (Form 1). In the next step of the analysis, traffic and geometric variables were included to establish the statistical correlation matrices among variables. This step enables the selection of traffic and geometric variables that are correlated strongly with the minor stream capacity.

For all forms of intersections and minor streams, the analysis showed that minor stream capacity had high correlation with major traffic flow, major traffic speed, visibility to waiting drivers, and major and minor street widths. The analysis also showed that the angle of intersection, radius of minor street vehicle path, and gradients did not have strong effects on the minor stream capacity. Kimber and Coombe (*2*) indicated that angle of intersection, turning radius, and gradients had no detectable effect on the capacity; however, in this study the ranges of these variables were very limited.

### Capacity of Right-Turn Stream

Multivariate regression analysis was conducted to determine the best form of the predictive equation for the right-turn capacity. A right-turn capacity model was developed for each intersection form.

**TABLE 1 Ranges of Variables for Intersection Form 1**

| Variable | Range |
|---|---|
| Number of Intersections | 32 |
| Number of Observations | 192 |
| Minor Street Width, m. | 3.0-7.8 |
| Major Street Width, m. | 6.0-9.0 |
| Approach Speed, km/hr. | 25-80 |
| Major Traffic Flow, PCU/hr. | 120-3000 |
| Capacity of Right-Turn, PCU/hr. | 60-960 |
| Visibility, m. | 20-150 |
| Angle of Intersection, Degree | 75-100 |

For Form 1, in Figure 1, the following regression equation was obtained:

$$C = 775\left(\frac{vis}{sp}\right)^{0.11}\left[1+\frac{(Wm-9)}{3.6}\right]^{0.94} \times\left[1+\frac{(W-3.6)}{3.6}\right]^{0.30}\left(\frac{1}{1+F_1^2}\right)^{0.82} \quad (1)$$

where

$C$ = capacity of right-turn stream (PCU/hr),
$vis$ = visibility to waiting drivers (m),
$sp$ = major traffic speed (km/hr),
$Wm$ = width of major street (m),
$W$ = width of minor street (m), and
$F_1$ = flow of through traffic in major street (1,000 PCU/hr).

Similarly, for Intersection Form 2, in Figure 1, the best predictive equation for the right turn was as follows:

$$C = 710\left(\frac{vis}{sp}\right)^{0.12}\left(1+\frac{W-9}{3.6}\right)^{0.97}\left(1+\frac{W-3.6}{3.6}\right)^{0.37} \times\left(\frac{1}{1+F_1^2}\right)^{0.80}\left(\frac{1}{1+0.4F_2^2}\right)^{0.78} \quad (2)$$

where $F_2$ is the flow of the left-turn traffic in the major street (in 1,000 PCU/hr). All parameters in these equations were significant at the 95 percent confidence level. Investigation of Equations 1 and 2 indicates that 775 and 710 can be interpreted as capacity under ideal conditions. The ideal conditions include visibility to speed ratio equal to 1, major street width of 9.0 m, minor street width of 3.6 m, and zero traffic flow in the major street. Although each minor stream was observed so that no minor street was lane sharing, the capacity of right-turn for Intersection Form 2 is significantly lower than that of Intersection Form 1.

In Equations 1 and 2, the value 3.6 m represents the lane width. Therefore, the effect of number of lanes on the estimated capacity can be evaluated easily. The value 9.0 m represents the maximum width of the major street.

**TABLE 2 Ranges of Variables for Intersection Form 2**

| Variable | Minor Street Stream | |
|---|---|---|
| | Right-Turn | Through |
| Number of Intersections | 39 | 39 |
| Number of Observations | 168 | 168 |
| Minor Street Width, m. | 3.0-7.8 | 3.0-7.8 |
| Major Street Width, m. | 5.6-9.6 | 5.6-9.6 |
| Approach Speed, km/hr. | 30-80 | 30-80 |
| Major Through Traffic Flow, PCU/hr. | 30-3280 | 30-3280 |
| Major Left-Turn Traffic Flow, PCU/hr. | 0-720 | 0-720 |
| Observed Capacity, PCU/hr. | 60-720 | 60-660 |
| Visibility, m. | 20-160 | 20-160 |
| Angle of Intersection, Degree | 80-100 | 80-100 |

**TABLE 3 Ranges of Variables for Intersection Form 3**

| Variable | Minor Street Stream | |
|---|---|---|
| | Left-Turn | Through |
| Number of Intersections | 19 | 19 |
| Number of Observations | 171 | 155 |
| Minor Street Width, m. | 5.0-7.0 | 5.0-7.0 |
| Major Street Width, m. | 6.6-9.0 | 6.6-9.0 |
| Approach Speed, km/hr. | 25-45 | 25-45 |
| Major Through Traffic Flow, PCU/hr. | 30-2220 | 30-2220 |
| Major Left-Turn Traffic Flow, PCU/hr. | 0-600 | 0-600 |
| Observed Capacity, PCU/hr. | 60-660 | 120-720 |
| Visibility, m. | 25-60 | 25-60 |
| Angle of Intersection, Degree | 80-100 | 80-100 |

### Capacity of Left-Turn Stream

A left-turn capacity predictive model was developed that is similar to the right-turn models. The best regression equation was as follows:

$$C = 675\left(\frac{vis}{sp}\right)^{0.11}\left(1+\frac{Wm-9}{3.6}\right)^{0.95}\left(1+\frac{W-3.6}{3.6}\right)^{0.30} \times \left(\frac{1}{1+F_1^2}\right)^{0.8}\left(\frac{1}{1+0.4F_2^2}\right)^{0.78} \quad (3)$$

The coefficient of determination ($R^2$) for Equation 3 was 0.92. All parameters were significant at a 95 percent confidence level. Compared with the capacity of the right-turn stream, the left turn has lower capacity, which is expected because a left-turn movement is much more complicated. However, traffic and geometric parameters are almost equal to the parameters of Equations 1 and 2.

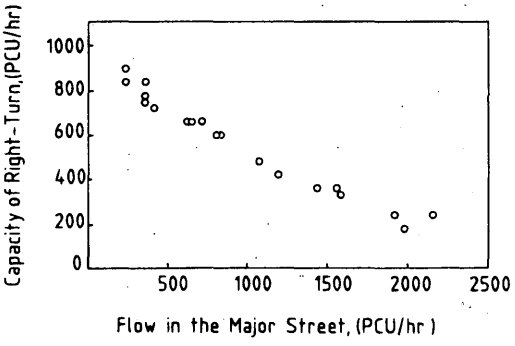



**FIGURE 2 Scatter plot of right-turn capacity for intersection Form 1 with major street width 8.6 m, minor street width 7.8 m, and approach speed 40 km/hr.**

### Capacity of Through Stream

Two predictive equations for estimating capacity of the through stream were developed. For Intersection Form 2, the best regression equation was as follows:

$$C = 580\left(\frac{vis}{sp}\right)^{0.07}\left(\frac{Wm}{3.6}\right)^{-0.25}\left(\frac{W}{3.6}\right)^{0.53}\left(\frac{1}{1+F_1^2}\right)^{0.93} \times \left(\frac{1}{1+0.8F_2^2}\right)^{1.19} \quad (4)$$

Similarly, for Intersection Form 3, the following equation was obtained:

$$C = 600\left(\frac{vis}{sp}\right)^{0.10}\left(\frac{Wm}{3.6}\right)^{-0.24}\left(\frac{W}{3.6}\right)^{0.57}\left(\frac{1}{1+F_1^2}\right)^{0.93} \left(\frac{1}{1+0.8F_2^2}\right)^{1.11} \quad (5)$$

The coefficient of determination values ($R^2$) were 0.94 and 0.91 for Equations 4 and 5, respectively. All parameters in Equations 4 and 5 were significant at a 95 percent confidence level. With one standard error of parameter estimates, the geometric and traffic parameters in both equations were almost equal.

### Modeling of Critical Gap and Move-Up Time

Regression analysis was used to identify the effect of major traffic and intersection geometric variables on the critical gap. The analysis indicated that the width of the major street and the speed of major traffic had a significant effect on the critical gap for right turn, left turn, and through stream. At the aggregate level, the analysis did not confirm the effect of major traffic flow on the estimated critical gap. But at the intersection level, the major traffic flow was correlated negatively with speed.

In this paper, the following regression equations for estimating the critical gap of right turn, left turn, and through stream were obtained:

$$\text{Ln}(CR) = 1.50 + 0.003\,(sp - 25) - 0.0423\,(Wm - 9) \quad (6)$$

$$\text{Ln}(CL) = 1.55 + 0.002\,(sp - 25) + 0.036\,Wm \quad (7)$$

$$\text{Ln}(CT) = 1.517 + 0.0017\,(sp - 25) + 0.035\,Wm \quad (8)$$

where

Ln = natural logarithm,
$CR$ = right-turn critical gap (sec),
$CL$ = left-turn critical gap (sec), and
$CT$ = through traffic critical gap (sec).

All parameters in Equations 6, 7, and 8 were found to be highly significant. The coefficient of determination values ($R^2$) were 0.72, 0.66, and 0.81 for Equations 6, 7, and 8, respectively. It is worth mentioning that the minimum average speed of major traffic was 25 km/hr.

The developed equations indicate that critical gaps increase substantially with increasing speed on the major street. Critical gaps for left-turn or through traffic increase with increasing width of the major street. Compared with the critical gap for minor through traffic, slightly longer gaps are needed for left-turn traffic. Field observations indicated that left-turning vehicles from the minor street merge with major traffic and tend to use the far right lane in the major street specifically under high major traffic speeds. The 1985 HCM estimated the critical gap on the basis of major traffic speed and street width as well as turning radius and gradient. Brilon (*9*) reported that Harders (*10*) found that critical gap was influenced by the speed of opposing vehicles at give-way intersections. Although recent studies (*5,11*) indicated that critical gap cannot be considered constant, other studies (*12*) have indicated that major traffic flow and queue length do not appear to have a significant impact on the length of critical gap. Therefore, the results of this study are compatible with most cited literature.

In the analysis of move-up time, it was found that it is correlated strongly with the relevant critical gap for each minor stream. For each stream, the move-up time was approximately 60 percent of the critical gap. This result is compatible with the Swedish Capacity Manual, which assumed that the move-up time makes up about 60 percent of the critical gap (*13*).

## DISCUSSION OF RESULTS

In this study an empirical approach using regression techniques was used to develop capacity models for yield-controlled streams at a one-way minor street crossing a one-way major street. For all minor streams, the results indicated that the relationship between capacity and both major traffic and intersection variables had a multiplicative form. Unlike the results of the study by Kimber and Coombe (*2*), here the relationship between minor stream capacity and major traffic flow would not be considered linear. If linearity exists, the capacity of minor stream will be zero at a high major traffic level. But field observations revealed that the actual capacity of the minor stream is at least 60 PCU/hr. Taking the sample size into consideration, Figure 2 may not confirm the linearity. However, further studies are recommended to highlight this issue.

Results of this study indicated that minor stream capacity increases with an increase in the visibility-to-speed ratio. For the same visibility level, doubling the major traffic speed would reduce the capacity by about 8 percent. This value is very small compared with values reported in the new German guideline (*4*). For right- or left-turning vehicles, increasing the width of the major street would increase capacity substantially. Increases in the width of a major street reduce the interactions between major and minor traffic and increase the turning capacity. For the same major traffic level, increasing the number of lanes from two to three in a major street would raise the right-turn capacity by about 45 percent (Figure 3). This gain may be explained by the fact that major street drivers tend to follow the far lane in the major street to avoid possible conflict with minor street vehicles. This behavior might provide better opportunities for minor vehicles to merge into the near-side lane of the major street.

The results also indicated that right- and left-turn capacity can be improved by increasing the number of minor street lanes. Unlike other highway facilities, increasing the number of lanes does not necessarily result in a corresponding increase in the capacity of the turning stream. For example, doubling the number of minor street lanes (from one to two lanes) would increase the right-turn capacity by 25 percent. This is expected, because right-turning vehicles rarely turn simultaneously from both lanes, as observed in the field. Despite the difference in traffic operation, the effects of doubling the number of minor street lanes on the right-turn and circular entry capacities can be compared. The circle can be considered as a series of T-shaped entries into a one-way cir-cular street. Brilon and Stuwe (*14*) found that the entry capacity increased by 30 to 40 percent when the number of entry lanes was doubled.

Furthermore, the results of the empirical and gap acceptance models were compared. Figure 4 shows the relationship between right-turn capacity and major traffic flow for an intersection of typical geometry. The intersection had a one-lane minor street (3.6 m), a three-lane major street (9.0 m), and a visibility-to-speed ratio of 2.0. Two curves are shown in Figure 4. The first curve (Curve A) was obtained using Equation 1 of the empirical approach, and the second curve (Curve B) was obtained using the gap acceptance model. For Jordanian drivers, the critical gap was estimated using Equation 6 with major traffic speed of 50 km/hr. Equation 6 indicates that the critical gap is 4.83 sec. Accordingly, the second curve was obtained using 4.83 and 2.9 sec for the critical gap and move-up (60 percent of the critical gap) parameters of Siegloch's formula. Siegloch's formula is as follows (*4*):

$$C = (3{,}600/t_f) * e^{-p[tg - (tf/2)]} \quad (9)$$

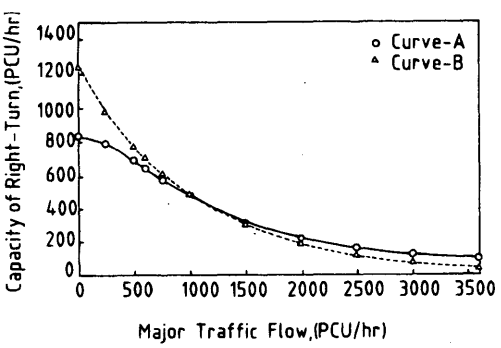



**FIGURE 3 Estimation of right-turn capacity according to empirical and gap acceptance models.**

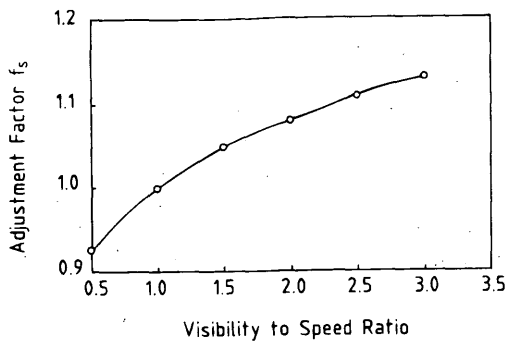



**FIGURE 4 Effect of visibility-to-speed ratio on right- or left-turn capacity.**

where

$C$ = capacity of minor stream (PCU/hr),
$p$ = $q$/3,600,
$q$ = volume of major stream,
$t_g$ = critical gap (sec), and
$t_f$ = move-up time (sec).

Compared with empirical results, Figure 4 indicates that the gap acceptance model significantly overestimates the predicted capacity at low major traffic levels and underestimates it at high levels. As shown in Figure 4, the empirical approach indicates that the estimated capacity is 835 PCU/hr at low major traffic levels. In contrast, the gap acceptance model indicates that the estimated capacity is 1,240 PCU/hr. The difference, 405 PCU/hr, in the estimated capacity values is very large. This large difference would result in a poor estimate of the reserve capacity and level of service. For Jordanian conditions, agreement between the results of the empirical and gap acceptance approaches can be achieved if the move-up time is set at 4.31 sec. This value is very large compared with field observations and cited literature (*15*). Accordingly, gap acceptance models would provide unrealistic capacity values at low major traffic levels.

At high major traffic levels, the gap acceptance model underestimates the predicted capacity, suggesting that longer delay is associated with higher traffic flow and that right-turning drivers are willing to accept shorter gaps to enter the major street. In addition, field observations indicated that at high traffic flows, some major street drivers yield the right of way to the entering drivers, specifically under low speed levels. Therefore, to achieve correspondence between the results of the empirical and gap acceptance approaches, it is necessary to adjust gap acceptance parameters to reflect the effect of traffic flow level, which could be done by adjusting the speed in the critical gap models according to the major traffic flow level.

Also, Figure 4 shows that the minimum right-turn capacity is about 100 PCU/hr. This value is compatible with the suggested limit of practical capacity for minor traffic streams at unsignalized intersections in the future German guideline (*4*).

## PRACTICAL APPLICATIONS

The results of the statistical analysis described in this paper provide a strong foundation for a method to estimate the capacity of one-way major-minor yield-controlled streams. The method is based on the models developed for each minor stream. It takes into account the effects of different geometric and traffic variables. Similar to other transportation facilities, the capacity of right- or left-turn stream in PCU/hr can be expressed as follows:

$$C = C_o * f_s * f_{Wm} * f_w * f_{F1} * f_{F2} \quad (10)$$

where $C_o$ represents the capacity of the turning stream under ideal conditions [(vis/sp) = 1.0, Wm = 9.0, $W$ = 3.6, $f_1$ = $\phi$, and $f_2$ = $\phi$). The value of $C_o$ is 775 for an exclusive right-turn minor street. For a nonexclusive right-turn minor street, $C_o$ values are 710 and 675 for right and left turns, respectively. The adjustment factor $f_s$ represents the effect of the visibility-to-speed ratio on the estimated capacity. Values of $f_s$ for different speeds and visibility-to-speed ratio are shown in Figure 5. The adjustment factors $f_{Wm}$ and $f_w$ represent the effects of major and minor street widths on the estimated capacity. Figure 3 provides values of $f_{Wm}$ and $f_w$ for different widths. Finally, $f_{F1}$ and $f_{F2}$ represent the effects of through and turning major traffic flows on the right-or left-turn capacity. Values of these adjustment factors are shown in Figure 6. For practical application, similar expression could be written to estimate the capacity of through stream from minor street.

In this study, capacity was estimated for each minor stream individually. However, if a lane (or lanes) is shared by more than one minor stream, the shared capacity is computed from the individual capacities according to the individual streams using Equation 10-1 in the HCM (*1*).

## CONCLUSIONS

An empirical approach using regression techniques was used to develop capacity models for yield-controlled streams at a one-way minor street crossing a one-way major street. The results of this study led to the following conclusions:

1. Capacity models for yield-controlled streams were developed. In addition to the widths of the minor and major streets, the visibility-to-speed ratio and the major traffic flow level had a significant influence on the minor stream capacity. The relationship between the estimated capacity and both major traffic and intersection geometric variables had a multiplicative form.

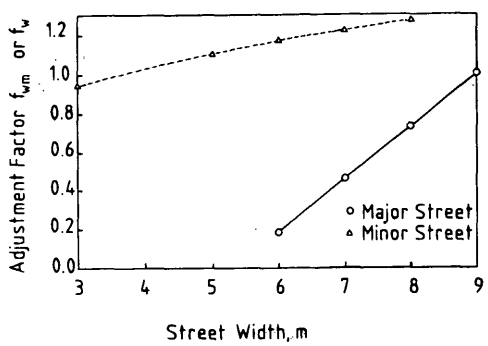



**FIGURE 5 Effect of major and minor street widths on right- or left-turn capacity.**

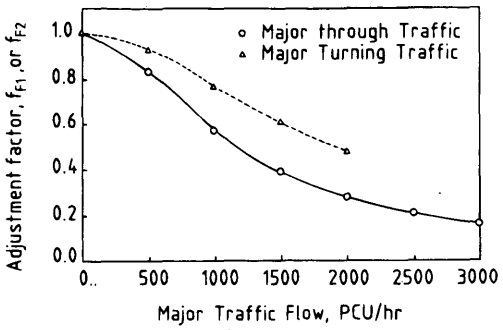



**FIGURE 6 Effect of major traffic flows on right- or left-turn capacity.**

2. For each minor stream, critical gap model was developed. Both the traffic speed and width of the major street were found to be significant for estimating the critical gap for each stream. The results indicated that the increase in the traffic speed tends to increase the critical gap for all minor streams. In contrast to the critical gap for right-turn stream, the critical gaps for the left-turn and through streams increase with the increase in the major street width.

3. Compared with results of the empirical approach, the gap acceptance models significantly overestimate the predicted capacity at low major traffic levels and underestimate them at high levels. The results indicated that even if critical gap is estimated for local conditions, the difference in estimating capacity is considerably large.

4. The 1985 HCM and the new Germany Guideline, which are based on gap acceptance models, would provide unrealistic minor stream capacity even if critical gaps were estimated for local conditions.

## REFERENCES

1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
2. Kimber, R., and R. Coombe. *The Traffic Capacity of Major/Minor Priority Junctions.* TRRL Report SR582. Department of the Environment, Department of Transport, Crowthorne, Berkshire, England, 1980.
3. Harders, J. Die Leistungsfahigkeit nicht signalgeregelter Stadtischer Verkehrsknoten. *Schriftenreihe Strassenbau und Strassenverkehrstechink,* Heft 76, Bonn, Germany, 1968.
4. Brilon, W., M. Grossmann, and B. Stuwe. Toward a New German Guideline for Capacity of Unsignalized Intersections. In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 168–174.
5. Kyte, M., C. Clemow, N. Mahfood, B. K. Lall, and C. J. Khisty. Capacity and Delay Characteristics of Two-Way Stop-Controlled Intersections. In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991, pp. 160–167.
6. Kimber, R. Gap-Acceptance and Empiricism in Capacity Prediction. *Transportation Science,* Vol. 23, No. 2, May 1989, pp. 100–111.
7. Troutbeck, R. J. Effect of Heavy Vehicles at Australian Traffic Circles and Unsignalized Intersections. In *Transportation Research Record 1398,* TRB, National Research Council, Washington, D.C., 1993, pp. 54–60.
8. Brilon, W., and M. Grossmann. The New German Guideline for Capacity of Unsignalized Intersections. In *Intersection Without Traffic Signals II,* Springer-Verlag, Germany, 1991.
9. Brilon, W. Recent Development in Calculation Methods for Unsignalized Intersections in West Germany. In *Intersections Without Traffic Signals* (W. Brilon, ed.), Springer-Verlag, New York, 1988, pp. 111–153.
10. Harders, J. Grenz-und Folgezeitlucken als Grundlage für die leistungsfahigkeit von landstrassen. *Schriftenreihe Strassenbau und Strassenverkehrstechnik,* Heft 216, Bonn, Germany, 1976.
11. Madanat, S., M. Cassidy, and Mu-Han Wang. Probabilistic Delay Model at Stop-Controlled Intersection. *Journal of Transportation,* ASCE, Vol. 120, No. 1, 1994, pp. 21–36.
12. Thomas, H. A Probabilistic Model of Gap Acceptance Behavior. In *Transportation Research Record 795,* TRB, National Research Council, Washington, D.C., 1981, pp. 8–13.
13. Hansson, A. Swedish Capacity Manual, Part 2. Capacity of Unsignalized Intersections. In *Transportation Research Record 667,* TRB, National Research Council, Washington, D.C., 1973, pp. 4–11.
14. Brilon, W., and B. Stuwe. Capacity and Design of Traffic Circles in Germany. In *Transportation Research Record 1398,* TRB, National Research Council, Washington, D.C., 1993, pp. 61–67.
15. Owens, D. *Flow Measurements at a Number of Uncontrolled T-Junctions.* RRL Report LR171. Ministry of Transport, England, 1968.

*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Unsignalized Intersection Capacity and Level of Service: Revisiting Critical Gap

MICHAEL J. CASSIDY, SAMER M. MADANAT, MU-HAN WANG, AND FAN YANG

Operational performance at a minor street stop-controlled intersection is a function of motorist gap acceptance behavior. Issues in modeling gap acceptance are reexplored using discrete choice methods. Logit models of varying levels of sophistication are used in simulation to generate average delays at the intersection stop bar. Comparison of simulated and empirical delays suggests that deterministic methods for modeling gap acceptance may represent a reasonable trade-off between accuracy and ease of application, but two potential concerns are at issue—namely, delay estimates are very sensitive to the value used for mean critical gap, and the use of a single-valued critical gap necessitates the exclusion of disaggregate factors influencing the gap acceptance decision. Logit models estimated for intersection traffic movements have identified a number of such influential factors. Further research to explore fundamental issues of gap acceptance should be undertaken before adopting a capacity and level-of-service methodology for minor street stop-controlled intersections.

Operating conditions at a two-way stop-controlled intersection are a function of driver choice characteristics. The propensity of motorists traveling on the minor street to use available gaps in the major street traffic streams will dictate operational performance. Efforts to model the gap acceptance behavior of motorists have been the focus of considerable research. There may be value in reexamining these previous efforts in light of newly revised capacity and level-of-service procedures for stop-controlled intersections (*1*). This paper reexplores key issues in modeling gap acceptance.

## BACKGROUND

The term "critical gap" is defined as the minimum time gap (exhibited by major street vehicles) allowing one vehicle to enter the intersection from a minor street. The gap acceptance process is probabilistic in nature. Each driver has his or her own perception of a critical gap, and the value of this "minimum acceptable" gap may change with changing conditions at the intersection. Functions have been developed from suggested distributions of critical gap to relate the probability of gap acceptance to the gap length (*2–4*). Miller (*5*) assumed that critical gaps conform to a normal distribution and used probit modeling techniques to estimate the probability of accepting a given gap on the basis of its length. These works addressed the variation in gap acceptance tendencies from one driver to the next.

Daganzo (*6*) extended Miller's work by accounting for variation within drivers as well as across drivers. Daganzo used multinomial probit to estimate the parameters of the distribution of critical gaps. Mahmassani and Sheffi (*7*) modeled the gap acceptance process as a series of independent, sequential choices to either accept or reject each gap in a conflicting traffic stream. By modeling the probability of gap acceptance using a probit function, Mahmassani and Sheffi demonstrated that an *individual* motorist's propensity to accept gaps varies as a function of the time spent waiting at the stop bar (*or* the number of gaps previously rejected). Recent empirical studies by Kittelson and Vandehey (*8*) support this finding. Most recently, Madanat et al. (*9*) used logit modeling to demonstrate that the delay spent in queue (before arriving at the stop bar) also influences drivers' gap acceptance behavior.

These previous findings underscore two important features:

1. Drivers are not homogeneous. Different drivers display different gap acceptance tendencies.
2. Drivers are not consistent. Drivers display time-dependent gap acceptance tendencies (e.g., a driver may ultimately accept a gap smaller than gaps that were previously rejected).

There appears to be little dispute concerning the probabilistic nature of gap acceptance. The literature does, however, document efforts to model gap acceptance decisions using deterministic methods (*1,10*). Single-valued, mean *critical gaps* are estimated from a distribution of gaps. The motorist is assumed to reject all prevailing gaps smaller than the critical gap, and all gaps larger than the critical gap are presumed to be accepted. These deterministic methods for capturing gap acceptance behavior are often assumed to possess adequate predictive strength or the benefits of exploiting deterministic models (which are easy to apply) are often considered to outweigh potential inaccuracies.

The need may exist to examine more carefully the trade-offs between the simplicity of a deterministic methodology and the robustness provided by a probabilistic, properly specified gap acceptance function. And if through careful examination the traffic engineering community eventually elects to adopt a deterministic model, the methodology used for estimating values of critical gap should be based on behaviorally defensible theories. This paper presents evidence concerning variation in gap acceptance behavior from one driver to the next, as well as time-dependent factors influencing the gap acceptance decision. The potential significance of this variability across and within drivers is presented both statistically and through simple example.

The work exploits a very limited empirical data base for model estimation. As such, gap acceptance functions presented herein are not definitive. The purpose of this paper, however, is not to propose the adoption of any particular model but to explore the relative merits of deterministic and probabilistic gap acceptance functions.

## EMPIRICAL DATA

Empirical data used to estimate gap acceptance functions were collected from two neighboring T-intersections in Indiana. The geo-

M. J. Cassidy, University of California, 109 McLaughlin Hall, Berkeley, Calif. 94720. S.M. Madanat, M.-H. Wang, and F. Yang, Purdue University, Civil Engineering Building, West Lafayette, Ind. 47907.

metric configurations of these suburban intersections are illustrated in Figure 1. The stop-controlled approach at Intersection 1 consists of separate lanes for left- and right-turning traffic, and the stop-controlled approach at Intersection 2 consists of a single shared-turn lane. The major (uncontrolled) street has one lane in both directions. No traffic control devices, other than the minor street stop signs, influence operation at the intersections.

Fifteen minutes of operation were recorded at each intersection using video. Data manually extracted from videotape included

- The lengths of all gaps observed in the major street traffic streams and whether each gap was accepted or rejected by motorists on the minor street,
- The time that minor street motorists waited in queue (before arriving at the stop bar),
- Vehicle move-up times to the stop bar, and
- The amount of time that individual motorists waited at the stop bar before accepting a gap.

## HIERARCHY OF GAP ACCEPTANCE FUNCTIONS

The authors first estimated a series of gap acceptance functions ranging from the very simple to the more sophisticated. Each function was estimated through discrete choice techniques (*11*). The application of discrete choice methods produced models estimated from disaggregate observations of individual behavior. Thus, the logit models estimated herein reflect the probabilistic nature of the gap acceptance process (i.e., the variability across and within drivers).

### Single-Valued Critical Gap Function

The simplest gap acceptance model recognizes variations in critical gap values across drivers; each driver is assumed to have his or her own critical gap. All gaps confronting a motorist that are smaller than his or her specified critical gap are invariably rejected. Conversely, motorists accept all gaps greater than or equal to their critical gaps. It is further assumed that these critical gap values follow some probability distribution across the population of motorists. The mean and variance of this distribution are the relevant parameters in this model.

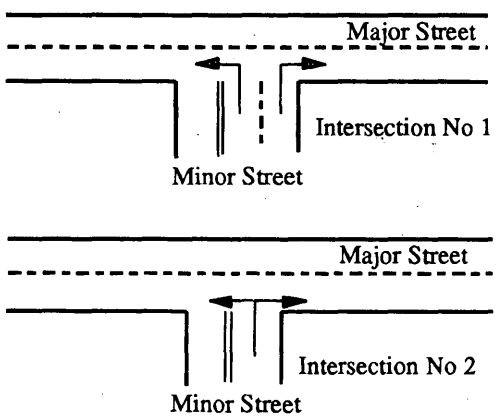



**FIGURE 1 Data collection sites: *top*, Intersection 1; *bottom*, Intersection 2.**

The mean critical gap cannot be observed directly. To estimate this value, the authors express the probability that a motorist accepts a given gap, Pr($a$), as the probability that a gap of length $t$ is greater than or equal to the motorist's critical gap, $T_{cr}$.

$$\Pr(a) = \Pr(t \geq T_{cr})$$

Denoting by $\bar{T}_{cr}$ the mean critical gap, and by $\epsilon$ the random deviation of each motorist's critical gap from the mean, this probability can be rewritten as

$$\Pr(a) = \Pr(t \geq \bar{T}_{cr} + \epsilon) = \Pr(\epsilon \leq t - \bar{T}_{cr})$$

Depending on the distribution of $\epsilon$, this probability function can take different forms; if $\epsilon$ is assumed to be distributed logistically with scale parameter $\mu$ (which implies that the distribution of critical gaps across drivers is logistic with mean $\bar{T}_{cr}$ and variance $\pi^2/3\mu^2$), then the probability takes the logit form

$$\Pr(a) = \frac{1}{1 + e^{-\mu(t-\bar{T}_{cr})}} = \frac{1}{1 + e^{-\mu t + \mu \bar{T}_{cr}}}$$

setting $\mu\bar{T}_{cr} = \alpha$

$$\Pr(a) = \frac{1}{1 + e^{\alpha - \mu t}}$$

Values of $\hat{\alpha}$ and $\hat{\mu}$ (the estimates of $\alpha$ and $\mu$) are obtained through maximum likelihood estimation (*12*). From the relationship $\hat{\bar{T}}_{cr} = \hat{\alpha}/\hat{\mu}$, the mean value of critical gap is estimated. The resulting model can be used to predict motorist gap acceptance behavior as a homogeneous and consistent process: all drivers can be assumed to accept gaps larger than $\hat{\bar{T}}_{cr}$ and to reject other gaps.

### Probabilistic Gap Acceptance Function

The deterministic model just presented fails to exploit the full capabilities of logit models. The logit model used for estimating mean critical gap also provides a *distribution* of critical gaps. As shown, these critical gaps are distributed logistically with a mean of $\bar{T}_{cr}$ and a variance of $\pi^2/3\mu^2$. Because the model identifies only the distribution of $T_{cr}$ and not the actual critical gap for each driver, the model can only generate probabilistic statements. The logit model yields the probability of accepting a specific gap as a function of its length, $t$. Thus, the function recognizes that drivers are not homogeneous, although drivers are still assumed to behave in a consistent manner.

### Probabilistic Function with Disaggregate Factors

Discrete choice methods (e.g., logit models) facilitate the identification of factors influencing gap acceptance as well as the inclusion of these factors in the resulting gap acceptance function. Thus, explanatory variables in addition to gap length can be incorporated into a logit model to further enhance estimation capabilities. The output of this more sophisticated model is the motorist's probability of gap acceptance as a function of relevant prevailing conditions. Through the inclusion of time-dependent factors that further explain gap acceptance decisions, the resulting function recognizes that a motorist's critical gap may change with changing conditions at the

intersection. Effectively, the function still assumes that motorists respond to critical gaps, but the value of critical gap varies with each motorist and with each new situation. As such, the logit function captures nonhomogeneous and inconsistent gap acceptance behavior among motorists.

### Estimating Gap Acceptance Functions

To demonstrate further the characteristics of each aforementioned gap acceptance function, the authors estimated specific models using empirical observations of right-turning vehicles at Intersection 1. To estimate the mean critical gap for a single-valued function, the logit model was derived incorporating only gap length $t$ as an explanatory variable:

$$\Pr(a) = \frac{1}{1 + e^{5.212 - 0.89934t}}$$

| *Independent Variable* | *Estimated Coefficient* | t-*Statistic* |
|---|---|---|
| 1 (constant) | −5.21200 | −7.65608 |
| $t$ | 0.89934 | 7.01240 |
| *Auxiliary Statistics* | *At Convergence* | *Initial* |
| Log likelihood | −67.129 | −150.41 |
| Number of observations | 217 | — |
| Adjusted rho-square | 0.547 | — |

The factor $t$ is highly significant in explaining gap acceptance propensity at the 95 percent level (i.e., the $t$-statistic is well above 2). Moreover, the model's overall fit is very satisfactory as evidenced by the adjusted rho-square value of 0.547.

From the relationship $\hat{T}_{cr} = \hat{\alpha}/\hat{\mu}$, the estimated mean critical gap is 5.212/0.899, or 5.8 sec. This single-valued function is illustrated at the top of Figure 2. The characteristics of the logit model are illustrated in the middle of Figure 2.

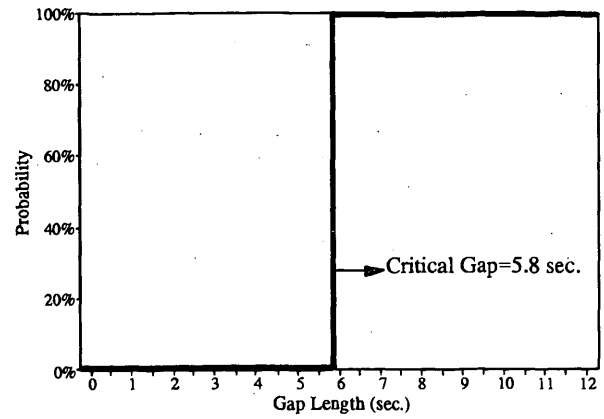
The logit model incorporating all explanatory variables found to be significant is as follows:

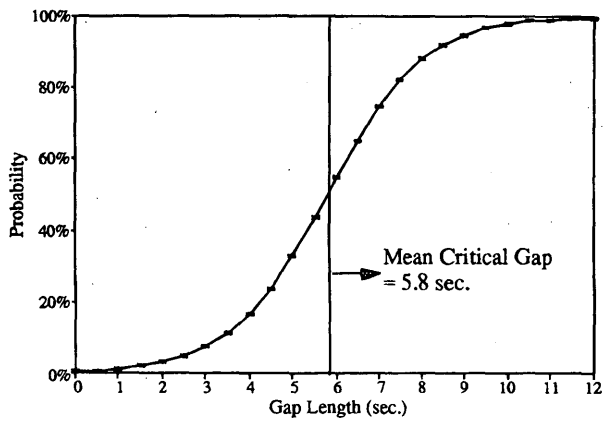
$$\Pr(a) = \frac{1}{1 + e^{8.109 - 1.373\,t - 0.042\,td + 1.720dv}}$$

where

$t$ = gap length confronting the right-turning motorist,
$td$ = total individual delay imparted to motorist up to occurrence of subject gap (where $td$ is sum of queueing and stop bar delays), and
$dv$ = dummy variable distinguishing gaps and lags;
$dv$ = 1 where motorist is confronted with a lag,
$dv$ = 0 otherwise (lag is defined as the elapsed time between the arrival of the minor street vehicle to the stop bar and the arrival of the first conflicting vehicle to the intersection).

| *Independent Variable* | *Estimated Coefficient* | t-*Statistic* |
|---|---|---|
| 1 (constant) | −8.10861 | −6.45629 |
| $t$ | 1.37346 | 6.56972 |
| $td$ (total delay) | 4.19540$e$−002 | 3.02295 |
| $dv$ | −1.71950 | −2.86030 |
| *Auxiliary Statistics* | *At Convergence* | *Initial* |
| Log likelihood | −53.828 | −150.41 |
| Number of observations | 217 | — |
| Adjusted rho-square | 0.620 | — |



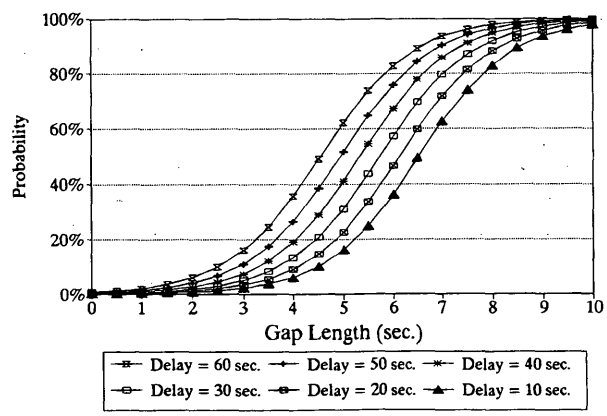




**FIGURE 2 Three hierarchal models of gap acceptance.**

The right-turn model indicates that a motorist's propensity to accept a gap increases with increasing gap length, consistent with virtually all previous studies of gap acceptance. Likewise, the model indicates that gap acceptance propensity increases with delay incurred on the intersection approach, a finding consistent with previous research evidence (*7–9*). Finally, the model indicates that, all else

**TABLE 1 Mean Stop Bar Delays**

| Measure | Empirical | Disaggregate Logit Model | Simple Logit Model | 5.8 sec. Critical Gap |
|---|---|---|---|---|
| Mean Stop Bar Delay (secs) | 7.5 | 7.0 | 5.6 | 6.0 |
| Percent Error | | 6.7 | 25.3 | 20.0 |

being equal, motorists have a greater tendency to accept gaps than to accept lags, consistent with the findings of Daganzo (*6*).

All independent variables in the model are statistically significant at the 95 percent level. The model's overall fit is very satisfactory, as indicated by the adjusted rho-square of 0.620, a higher value than that of the simpler logit model. The bottom of Figure 2 illustrates the gap acceptance probabilities estimated by the disaggregate logit model for a range of total individual delays. Lag acceptance probabilities are not displayed in this figure.

## DELAY ESTIMATION

Using the functions just described in conjunction with simulation, the potential impacts of gap acceptance functions on delay prediction will be explored. Moreover, the sensitivity of predicted delay to single-valued gap acceptance functions is demonstrated. On the basis of this sensitivity, the authors argue the importance of estimating a mean critical gap through behaviorally defensible techniques.

### Estimation Method and Simulation Model

The gap acceptance functions were incorporated into a microscopic, stochastic simulation model. Delay estimates generated from each function were evaluated.

Simulated vehicle arrivals on all approaches conformed to a Poisson distribution and were based on the observed mean arrival rates. Vehicle move-up times on the stop-controlled approach conformed to empirically identified distributions.

The initial simulation experiments separately used each of the three gap acceptance functions presented earlier. When a logit function was used in the simulation model, the gap acceptance probability of each right-turning vehicle at the stop bar was computed at the onset of each gap or lag. If the gap acceptance probability exceeded a randomly generated number from the [0,1] uniform distribution, the gap was accepted. When a single-valued function was used in the simulation, the process was purely deterministic (i.e., all gaps less than the specified critical gap were rejected, all gaps greater than or equal to the critical gap were accepted).

### Simulation Findings

Each gap acceptance function was evaluated by comparing empirical and simulated stop bar delays (i.e., the delays incurred by motorists waiting at the stop bar for a suitable gap). Table 1 presents average stop bar delays as obtained (*a*) empirically, (*b*) through simulation with the logit function incorporating disaggregate explanatory variables, (*c*) through simulation with the simple logit function accounting only for the influence of gap length, and (*d*) through simulation with a single-valued critical gap of 5.8 sec (as estimated previously).

The logit model incorporating disaggregate factors generated an average delay prediction that most closely matches the empirical value. Statistical tests indicated that discrepancies between simulated and empirical means and variances were not significant at the 95 percent level.

The discrepancies between stop bar delay mean and variance as generated with the single-valued critical gap of 5.8 sec likewise were not statistically significant from the empirical values. Thus, from the example scenario evaluated in this paper, there is no evidence that a single-valued gap acceptance function cannot be used to model driver behavior reliably at a stop sign. As is explained in the following section, however, a deterministic approach to gap acceptance may be reliable only if the specified value of critical gap is an appropriate estimate.

Finally, significant differences did not exist between delay values generated from the simple logit model accounting only for gap length and from the single-valued critical gap function. This was to be expected as the long-run estimates generated from an average value of critical gap will be equivalent to the outcomes generated from a distribution of critical gaps.

### Delay Sensitivity to Critical Gap

Where the gap acceptance model is a single-valued function, simulation experiments suggest that predicted delay is very sensitive to the specified value of critical gap. The following table presents the simulated estimates of average stop bar delay for various single-valued critical gaps:

| *Critical Gap (sec)* | *Average Stop Bar Delay (sec)* |
|---|---|
| 5.5 | 5.1 |
| 5.8 | 6.0 |
| 6.0 | 6.6 |
| 6.5 | 8.8 |
| 7.0 | 13.0 |

Marginal changes in the specified critical gap value produce relatively large differences in estimates of average stop bar delay, consistent with the tendencies of analytical queueing models.

Further simulation experiments revealed that the specified value of critical gap substantially alters estimates of average approach delay, a common measure of effectiveness. Table 2 presents simulated steady-state values of average approach delay as a function of critical gap. Critical gap values that vary slightly from 5.8 sec yield sizable differences in estimated approach delay.

**TABLE 2 Simulated Average Approach Delays Using Critical Gap**

| | Probabilistic | Deterministic | | | | |
|---|---|---|---|---|---|---|
| Gap Acceptance Function | Disaggregate Logit Function | 5.5 sec | 5.8 sec | 6.0 sec | 6.5 sec | 7.0 sec |
| Average Approach Delay (secs) | 11.7 | 8.6 | 10.9 | 12.1 | 17.5 | 28.2 |

### Estimating Mean Critical Gap

If a single-valued gap acceptance function is to be used for analysis, the apparent sensitivity of predicted delay dictates that critical gap values reflect proper estimates. Thus, any published values of critical gap should be estimated by means consistent with motorist behavior.

The revised *Highway Capacity Manual* procedures (*1*), for example, adopt a method for estimating critical gap previously described by Miller (*5*) and Troutbeck (*13*). With this method, a mean critical gap (for a particular maneuver) is inferred statistically from sequences of observed gaps at a stop-controlled approach using the assumption that the observed largest gap rejected by a motorist is smaller than the driver's critical gap, which, in turn, is smaller than the gap actually accepted by the motorist (i.e., drivers are assumed to be consistent). Estimating mean critical gap using this assumption has several shortcomings:

- All gaps rejected by the motorist except for the largest rejected gap are not included in the estimation of critical gap. This results in a loss of important information.
- Important information is also lost if the observed motorist accepts a *lag*. As no gaps are rejected, data specific to the driver are discarded. The loss of such information may cause bias in the estimated critical gap (i.e., sample selectivity bias) given that previous research (*6*) and findings reported in this paper indicate that motorists respond differently to gaps than to lags.
- A problem occurs whenever drivers reject gaps larger than the one that they eventually accept, a frequent occurrence (*6–9*, also the disaggregate logit model). Data specific to these drivers are either discarded or "*modified*" to be consistent with the assumption of motorist homogeneity and consistency. Discarding or changing observations to match postulates is a concern.

In contrast to the method just described, the application of discrete choice techniques to estimate critical gap is consistent with observable phenomena. By exploiting all observations, the resulting estimates of mean critical gap capture the variability across and within motorists. Given the apparent sensitivity of delay, discrete choice methods should be used for estimating mean critical gap. Such estimates can be derived easily with standard software packages, as demonstrated earlier.

A logit function estimated with sample data in which the fraction of rejected gaps differs significantly from that of the population will be biased in the estimated constant term. If the population's fraction of rejected gaps is known, a correction can be applied (*11*). Bias becomes an issue when estimating functions that are to be generalized. The concern can be avoided by developing gap acceptance models for intersections operating under specified sets of conditions.

## FURTHER EVALUATION OF GAP ACCEPTANCE FACTORS

The example scenario does not suggest that exploiting a single-valued gap acceptance function is inappropriate for intersection analysis. Nonetheless, it will be demonstrated that the application of a mean critical gap leads to a potential dilemma: excluding disaggregate factors that influence gap acceptance erodes estimation power. A likelihood ratio test indicated that the predictive strength of the disaggregate logit model is significantly greater at the 95 percent level than that of the simpler logit model. (This finding was inevitable given that all coefficients in the disaggregate function are statistically significant.)

For further exploring the significance of influential factors, gap acceptance functions estimated for the remaining minor street movements at Intersections 1 and 2 are presented.

### Intersection 1

The gap acceptance function estimated for left-turn minor street vehicles at Intersection 1 is as follows:

$$\Pr(a) = \frac{1}{1 + e^{7.909 - 1.382\,ming - 0.013\,td + 1.192 dvnf}}$$

| *Independent Variable* | *Estimated Coefficient* | t-*Statistic* |
|---|---|---|
| 1 | −7.90869 | −6.81553 |
| *ming* | 1.38196 | 6.79695 |
| *td* | 1.26032e-002 | 2.13013 |
| *dvnf* | −1.19245 | −2.10543 |
| *Auxiliary Statistics* | *At Convergence* | *Initial* |
| Log likelihood | −53.204 | −218.34 |
| Number of observations | 315 | |
| Adjusted rho-square | 0.740 | |

Disaggregate total delay, *td*, increases driver propensity to accept smaller gaps. The influence of gap length is complicated in that left-turn maneuvers are executed through two conflicting traffic streams. A specification search indicated that a more powerful model results from the inclusion of a single coefficient, *ming*, representing the smaller of the two gaps in both traffic streams. This suggests that left-turning motorists evaluate opposing gap lengths collectively and react to the smaller of the two gaps.

Finally, the estimated coefficient *dvnf*, a dummy variable, is equal to 1 if the smaller prevailing gap is in the near-side lane (*dvnf* = 0 otherwise). The sign of this coefficient indicates that drivers have a reduced propensity to accept a smaller gap occurring in the near-side lane, contrary to an earlier empirical finding (*8*).

### Intersection 2

The estimated gap acceptance functions for right- and left-turn movements at Intersection 2, the T-intersection with a shared turn lane, are as follows:

$$\Pr(a) = \frac{1}{1 + e^{6.126 - 1.11\,t - 1.258\,dv1}}$$

| *Independent Variable* | *Estimated Coefficient* | t-*Statistic* |
|---|---|---|
| 1 | −6.12569 | −6.75654 |
| *t* | 1.10980 | 6.23020 |
| *dv*1 | 1.25750 | 2.17838 |
| *Auxiliary Statistics* | *At Convergence* | *Initial* |
| Log likelihood | −46.553 | −160.12 |
| Number of observations | 231 | — |
| Adjusted rho-square | 0.69 | — |

$$\Pr(a) = \frac{1}{1 + e^{7.531 - 0.916\,ming - 0.011\,td - 1.466\,dvnf - 1.209\,dv2}}$$

| *Independent Variable* | *Estimated Coefficient* | t-*Statistic* |
|---|---|---|
| 1 | −7.53077 | −7.88504 |
| *ming* (min. gap) | 0.91590 | 8.62597 |
| *td* | 1.05003e-002 | 2.18858 |
| *dvnf* | 1.46597 | 2.82062 |
| *dv*2 | 1.20854 | 1.94173 |
| *Auxiliary Statistics* | *At Convergence* | *Initial* |
| Log likelihood | −79.989 | −302.91 |
| Number of observations | 437 | — |
| Adjusted rho-square | 0.720 | — |

For right-turn movements, driver propensity to accept a gap increases with gap length. The dummy variable *dv*1 reflects driver propensity to follow behind a "leading" right-turn vehicle (an influence unique to shared turn lanes)—that is, *dv*1 = 1 where the preceding vehicle executed a right turn. The sign of this coefficient indicates that right-turning drivers are motivated to accept lags remaining from a previous right turn.

The right-turn function in this model does not have a coefficient reflecting the effect of individual delay at the stop-controlled approach. The apparent exclusion of this influence is most likely attributable to a lack of variability in total delay observed in the data set. The actual influence of delay on gap acceptance may not be insignificant.

In the gap acceptance function for left-turn movements at Intersection 2, the influence of individual delay, *td*, and minimum gap, *ming*, are interpreted as in the logit model for left-turn vehicles at Intersection 1. At Intersection 2, however, the sign of the dummy variable *dvnf* indicates that a motorist has a lower propensity to perform left-turn maneuvers where the smaller of the two opposing gaps occurs in the far-side lane, a finding consistent with an earlier study (*8*). Finally, the dummy variable *dv*2 characterizes a left-turn driver's propensity to accept the lag "left behind" by a preceding left-turn vehicle. The sign of this coefficient is compatible with the factor identified for right-turn movements at the intersection.

### Pooling Models Across Intersections

Tests of taste variation were conducted to assess gap acceptance behavior across intersections. The assessments indicated that gap acceptance functions for either turning movement should not be combined across intersections, implying that differences in geometrics may create differences in gap acceptance behavior.

### Summarizing Model Estimation

The coefficients found to affect gap acceptance decisions include disaggregate measures such as individual approach delay, influences of near- and far-side gaps in the conflicting traffic streams, motorist propensity to follow closely behind leading motorists executing the same maneuver, and a general preference for gaps over lags. Because these factors were all significant at the 95 percent level, the disaggregate gap acceptance functions have much greater predictive power than models including only gap length. Thus, for the scenarios evaluated in this paper, the estimation capabilities of mean critical gaps are inferior to those provided by disaggregate models.

## CONCLUSIONS

This paper has highlighted issues relevant to modeling gap acceptance behavior at stop-controlled intersections. Findings from this study do not suggest that deterministic methods for modeling gap acceptance are unacceptable. Because using deterministic functions leads to analysis techniques that are easy to apply, modeling gap acceptance using a single-valued critical gap may be justified.

However, this paper illustrates two concerns. First, if the traffic engineering community adopts a deterministic gap acceptance methodology, values of mean critical gap should be estimated using techniques consistent with motorist behavior. Critical gap values that differ only marginally from proper estimates produce dramatic delay prediction errors.

Second, the use of single-valued functions necessitates the exclusion of disaggregate factors influencing the gap acceptance decision. The limited data exploited in this paper provide some insight into how the exclusion of these factors harms estimation. More conclusive assessments (using larger empirical data bases) are required. Expanded empirical evaluations would probably identify additional factors that affect gap acceptance. Such discrete influences might include socioeconomic driver characteristics, conflicting vehicle speeds and flows, and intersection geometrics.

Before the trade-offs between deterministic and probabilisitic gap acceptance functions can be identified, the estimation capabilities of both function types should be evaluated carefully. The application of discrete choice methods may represent the appropriate means for satisfying research needs in gap acceptance modeling. The relative strengths of deterministic and probabilistic gap acceptance functions may be evaluated through discrete choice. No matter which function type is ultimately adopted, the gap acceptance model can be estimated by logit or probit.

Should probabilistic functions be warranted, incorporating the gap acceptance model into an intersection assessment procedure becomes a consideration. Perhaps the only practical means of applying probabilistic models is through computer simulation. If manual evaluation techniques are desired, nomographs or some

other graphical-based method can be constructed from simulation experiments.

## ACKNOWLEDGMENT

The authors thank Jon Fricker, Purdue University, for providing the empirical data used in this work.

## REFERENCES

1. *Special Report 209: Highway Capacity Manual,* 3rd ed. TRB, National Research Council, Washington, D.C., 1994, Chapter 10.
2. Cohen, E., J. Dearnaley, and C., Hansel. The Risk Taken in Crossing a Road. *Operations Research Quarterly,* Vol. 6, 1955, pp. 120–128.
3. Drew, D., L. LaMotte, J. Buhr, and J. Wattleworth, *Gap Acceptance in the Freeway Merging Process.* Report 430-2. Texas Transportation Institute, Texas A&M University, College Station, 1967.
4. Solberg, P., and J. Oppenlander. Lag and Gap Acceptance at Stop-Controlled Intersections. In *Highway Research Record 118,* HRB, National Research Council, 1966, pp. 48–67.
5. Miller, A. Nine Estimators of Gap Acceptance Parameters. *Proc., 5th International Symposium on Theory of Traffic Flow and Transportation* (G.F. Newell, ed.), 1972, pp. 215–235.
6. Daganzo, C. Estimation of Gap Acceptance Parameters Within and Across the Population from Direct Roadside Observation. In *Transportation Research Part B,* Vol. 15B, 1981, pp. 1–15.
7. Mahmassani, H., and Y. Sheffi. Using Gap Sequences to Estimate Gap Acceptance Functions. In *Transportation Research Part B,* Vol. 15B, 1981, pp. 243–248.
8. Kittelson, W., and M. Vandehey. Delay Effects on Driver Gap Acceptance Characteristics at Two-Way Stop-Controlled Intersections. In *Transportation Research Record 1320,* TRB, National Research Council, Washington, D.C., 1991.
9. Madanat, S. M., M. J. Cassidy, and M. H. Wang. A Probabilistic Model of Queueing Delay at Stop-Controlled Intersection Approaches. *Journal of Transportation Engineering,* ASCE, Vol. 120, No. 1, Jan.–Feb. 1994, pp. 21–36.
10. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
11. Ben-Akiva, M., and S. Lerman. *Discrete Choice Analysis.* MIT Press, Cambridge, Mass., 1985.
12. Benjamin, J. R., and C. A. Cornell. *Probability, Statistics and Decision.* McGraw-Hill Publishing, New York, 1970.
13. Troutbeck, R. J. Estimating the Critical Acceptance Gap from Traffic Movements. Report 92/3. Physical Infrastructure Center, Queensland University of Technology, Australia, 1992.

# DISCUSSION

ROD TROUTBECK
*Queensland University of Technology*, P.O. Box 2434, Brisbane, Australia.

The paper by Cassidy et al. is an interesting one in that it discusses a problem that has relevance today as researchers look at the performance of unsignalized intersections and of permitted right-turn movements at signalized intersections. The authors presented a well-documented case, but omitted an important point: the proportion of gaps accepted is influenced by driver characteristics and by flows. The reason is, drivers whose perceptions of a critical gap is longer than other drivers' will reject a number of gaps. Similarly, as the major stream flow increases, the number of shorter gaps will increase, and the drivers who need longer gaps will reject even more gaps. Hence, as the priority stream flow increases, the gap size with a 50 percent probability of being accepted increases.

An approximate relationship exists between the gap size with a 50 percent probability of being accepted and the priority stream flow, as a function of the mean and variance of the critical gap distribution (*1*). Ashworth's correction was based on the critical gap distribution for drivers whose critical gaps follow a normal distribution and on a probit function for the probability that a gap will be accepted. If the critical gap distribution has a log-normal distribution and the gaps in the priority stream are exponentially distributed, the proportion of gaps of a particular size that are accepted, Pr($a$), does not have a log-normal distribution (*2*).

However, the equation

$$\mu_p \approx \mu_f + q\sigma_f^2 \tag{1}$$

can be applied for most critical gap distributions, including normal, log-normal, and gamma distributions. Here, $\mu_p$ is the mean used to quantify the *P(a)* distribution. If the critical gaps are normally distributed, the gap size with a 50 percent chance of being accepted is $\mu_p$. If the critical gap distribution has a log-normal distribution, the gap size with a 50 percent chance of being accepted, $t_{50}$, is given by

$$t_{50} = \frac{\mu_p}{\sqrt{1 + \sigma_p^2/\mu_p^2}}$$

where

$s_p^2$ = variance used to quantify Pr($a$) curve,
$\mu_f$ and $\sigma_f^2$ = mean and variance of the critical gap distribution, and
$q$ = flow (vehicles/sec).

Miller also indicated that the coefficient of variation of the Pr($a$) distribution is approximately the same as the critical gap distribution. Hence

$$\sigma_p/\mu_p \approx \sigma_f/\mu_f \tag{2}$$

For example, if the driver's critical gap has a log-normal distribution with a mean of 6$s$ and a standard deviation of 2$s$, Ashworth's technique indicates that $t_{50}$ would be given by

$$t_{50} = \mu_p/\sqrt{1.1111} = 0.949\ \mu_p$$

In addition, by using Equation 1

$$t_{50} \approx 0.95\ (\mu_f + q\ \sigma_f^2)$$

or

$$t_{50} \approx 5.69 + 3.79q \tag{3}$$

To demonstrate that this equation is reasonable, I took a sample of 500 drivers whose critical gaps followed a log-normal distribution and presented them with simulated gaps with a Cowan M3 distribution of 100 times. For each of these times, a logit analysis was applied to the accepted and rejected gaps, and a $t_{50}$ value was estimated.

There was a small difference between the logit function I used and the one used by the authors in that the logit function was assumed to be a function of logarithm of gap size. That is,

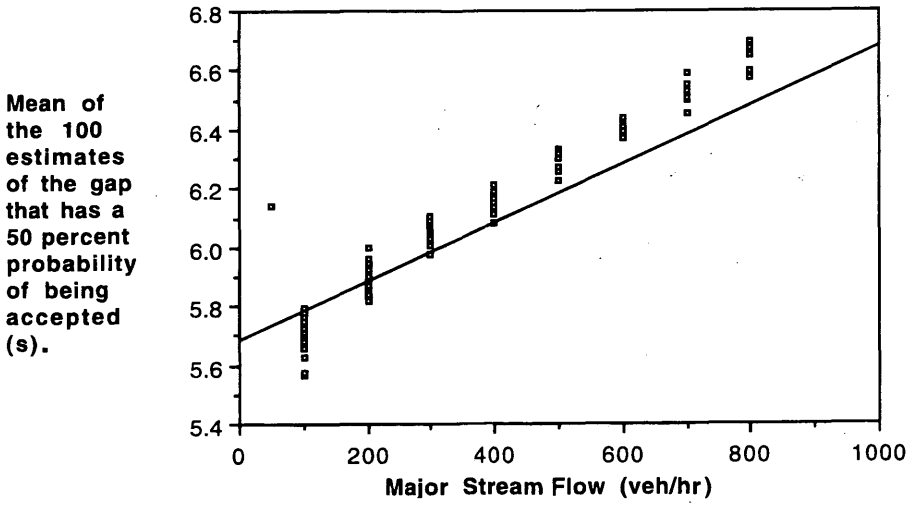



**FIGURE 3 Ashworth's line determined by Equation 3.**

$$\ln\left\{\frac{\Pr(t)}{1-\Pr(t)}\right\} = a\ln(t) + b \tag{4}$$

hence

$$\Pr(t) = \frac{1}{1 + e^{-a\ln(t)-b}} \tag{5}$$

This ensures that as $t$ approaches zero $\Pr(t)$ approaches zero. The functions in the Cassidy et al. paper indicate that there is a probability that a gap of zero will be accepted.

By using Equation 3 or 4,

$$t_{50} = e^{-b/a} \tag{6}$$

Thus, for each set of minor stream and major stream arrival flows, I obtained 100 estimates of $t_{50}$. The mean of these estimates, determined by using Equation 3, appear in Figure 1. The drivers were assumed to be consistent with a mean critical gap of 6*s* and a standard deviation of 2*s*. If this condition is relaxed so that drivers have a degree of inconsistency, the results show a similar trend.

Figure 3 indicates that Ashworth's equation provides a reasonable fit, and explains that there is a relationship between the proportion of gaps with a 50 percent chance of being accepted and the major stream flow.

The conclusions I reach are, first, that there is a monotonic, increasing relationship between the proportion of accepted gaps and the major stream flow if driver behavior remains constant and flows change. This can be explained using Ashworth's method. Second, logit or probit analyses should account for this trend, but more important, researchers should expect terms such as total delay to be statistically significant in the logit or probit analyses. However, this delay term also is a proxy for the flow term. I caution others (using these simplified logit or probit analyses) not to assume that a delay term has a substantial affect on the critical gap function.

## REFERENCES

1. Ashworth, R. A Note on the Selection of Gap Acceptance Criteria for Traffic Simulation Studies. In *Transportation Research Record 2,* TRB, National Research Council, Washington, D.C., 1968, pp. 171–175.
2. Miller, A.J. Nine Estimators of Gap Acceptance Parameters. *Proc., 5th International Symposium on the Theory of Traffic Flow and Transportation,* (G.F. Newell, ed.) 1972, pp. 215–232.

# AUTHORS' CLOSURE

We thank Troutbeck for his discussion concerning the relationship between flow on major streets and gap acceptance. We do not disagree with his assertion that "as the priority stream flow increases, the gap size with a 50-percent probability of being accepted increases." We do not have data substantiating this claim because our study relied solely on a small data set; that is 15 minutes of observations from each of two intersections. This small data set did not provide a wide range of flows.

We suspect that if we had estimated gap acceptance models by using a larger data base with a range of major street flows, we would have found this factor to be significant. We note the likelihood of this in the conclusions of the manuscript.

We emphasize that the delay term in our models is not a proxy for the flow term, and our finding that delay is a significant predictor of gap acceptance is not attributed to a change in major street flows. Our models' delay term is driver specific. It is the disaggregate delay imparted to a motorist who, while waiting on a minor street, is confronted with a nearly fixed flow on a major street. We found that drivers who experienced longer delays had a propensity to accept shorter gaps and found this difference in gap acceptance behavior to be statistically significant. This observed effect was independent of the major street traffic flow.

It is worth reiterating that two factors—minor street delay and major street flow—influence gap acceptance in opposite ways. The discussant's formulas and accompanying figure point out that the gap size with a 50-percent chance of being accepted increases with major street flow. Conversely, we found that added delay leads to a decrease in the estimated value of critical gap.

*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Development of Speed-Flow Relationships for Indonesian Rural Roads Using Empirical Data and Simulation

KARL L. BANG, ARNE CARLSSON, AND PALGUNADI

Highway capacity manuals from developed countries cannot be applied successfully in Indonesia because of large differences in driver behavior, traffic composition, and level of roadside activities. The Indonesian Highway Capacity Manual project was started in 1990 and has resulted in interim manuals for urban traffic facilities, interurban roads, and superhighways. Speed-flow relationships for interurban rural roads have been obtained from a combination of direct speed-flow measurements and simulation, using the VTI microscopic simulation model for two-lane, two-way undivided roads. The empirical data were used mainly for calibrating and validating the simulation model and for analyzing the effect on speed and capacity of cross-section and environmental conditions. The simulation model was used for determining speed-based light-vehicle units (used instead of passenger-car units and speed-flow relationships for flat, rolling, and hilly terrain. The results can be summarized as follows: (*a*) the light-vehicle free-flow speed for a flat two-lane, two-way road at ideal conditions is considerably lower in Indonesia than in developed countries; (*b*) free-flow speed is reduced by road width and side friction such as public transit stops, pedestrians, nonmotorized vehicles, and entries and exits from roadside properties and minor roads; (*c*) Indonesian drivers tend to overtake at short sight distances, which reduces the slope of the speed-flow curve; and (*d*) the capacity for two-lane, two-way roads is slightly higher in Indonesia than in developed countries.

Capacity values and speed-flow relationships used for planning, design, and operation of highways in Indonesia mainly have been based on manuals from Europe and the United States. Studies at the Institute of Technology in Bandung in the 1980s, however, showed that these sources might produce misleading results, possibly because of the high content of small utility vehicles and motorcycles on Indonesian roads and the side friction caused by roadside activities. The Indonesian Highway Capacity Manual (IHCM) project started in 1990 and so far has resulted in interim manuals for urban traffic facilities (*1*) and interurban roads and superhighways (*2*). The last phase of the IHCM project, including development of traffic engineering guidelines and computer software and implementation in national road management systems, will be completed in 1996.

Although the term "rural roads" is used in the title of the paper, it must be noted that the data base for the analysis also includes roads with considerable roadside development, residential and commercial. This development can be nearly continuous even far away from major urban areas because of the very high population density and difficult topography in many parts of Indonesia. Java, for instance, has a population of 115 million in an area smaller than Florida.

K. L. Bang, SWEROAD, P. O. Box 27, Ujung Berung 40600, Indonesia. A. Carlsson, Swedish Road and Traffic Research Institute, Linkoping 58101, Sweden. Palgunadi, Directorate General of Highways, Jl. Pattikura 20, Jakarta Selatar, Indonesia.

## DATA COLLECTION

### Analysis Strategy

In determining speed-flow relationships from empirical speed-flow data for even a limited number of standard road classes, traffic, and environmental (roadside) conditions, it is necessary to conduct data collection at a very large number of sites. Another way is to use a simulation model calibrated for actual driver behavior and vehicle characteristics. A combination of these two methods was used in the IHCM project. The empirical data were used mainly for calibrating and validating the simulation model and for analyzing the effect on speed and capacity of cross-section and environmental conditions. The simulation model was used primarily for determining light-vehicle units (LVUs, instead of passenger-car units, as explained later) and speed-flow relationships at different horizontal and vertical alignments.

### Vehicle Classes

The Indonesian Highway Administration distinguishes between 13 classes of vehicle for its routine classified counts. In the data collection for the IHCM project, the following seven vehicle classes and their average traffic compositions were distinguished:

- Light vehicles (LVs): passenger cars, jeeps, minibuses, pick-ups, microtrucks (58 percent).
- Medium heavy vehicles (MHVs): two-axle trucks with double wheels on the rear axle, buses shorter than 8 m (22 percent).
- Large trucks (LTs): three-axle trucks (2 percent).
- Truck combinations (TCs): truck plus full trailer, articulated vehicle (2 percent).
- Large buses (LBs): buses longer than 8 m (5 percent).
- Motorcycles (MCs): (11 percent).
- Unmotorized vehicles (UMs): mainly tricycles and bicycles ($<$1 percent).

The number of vehicle classes considered in IHCM was later reduced by including the truck combinations in the large-truck category and by considering unmotorized vehicles as elements of side friction rather than traffic flow.

### Data Collection Methodology

Data were collected in the field during 1991–1993 at 150 sites, including 35 interurban road sections with continuous residential or

commercial roadside development that were surveyed during the urban phase of the project. The basic survey equipment at each site included one or more short-base measurement stations equipped with pairs of pneumatic tubes (spacing of 3 m) connected to data loggers for recording vehicle axle passage times. By means of specialized software traffic flow and composition, space mean speed and headways were obtained automatically and cross-checked with the backup video recordings.

All side friction events were recorded manually, and all vehicles passing the short base were continuously video recorded. Through video data reduction (visual matching) of vehicles passing successive short-base stations, travel time and frequency of overtaking vehicles were also obtained for longer road sections in different terrain types.

Nine overtaking surveys using continuous, stationary video recording of about 1-km segments of roads in combination with overtaking observations from a moving observer vehicle equipped with cameras pointing forward and backward were also made. Sight distances were measured with the help of roadside markers or existing electrical poles that were visible in the video recordings (Figure 1).

## SIMULATION MODEL

### Description of Model

The VTI Road Traffic Simulation Model System (*3*), developed by the Swedish Road and Traffic Research Institute, was used in the IHCM project. The VTI model is "event-controlled" and programmed in SIMULA. It describes traffic operation on a two-way, single-carriageway road at a microscopic level, in which the movements of individual vehicles are modeled as they "progress" along a defined road segment in the computer (Figure 2).

Each movement is determined from a set of stochastic attributes defining the driver behavior and the vehicle characteristics. The position and current speed of each vehicle is calculated on the basis of driver decisions, which in turn are related to external factors such as road alignment and interference with other vehicles in their own or the opposing stream.

**FIGURE 1 Short-base measurement station along road section equipped with roadside markers for overtaking survey.**

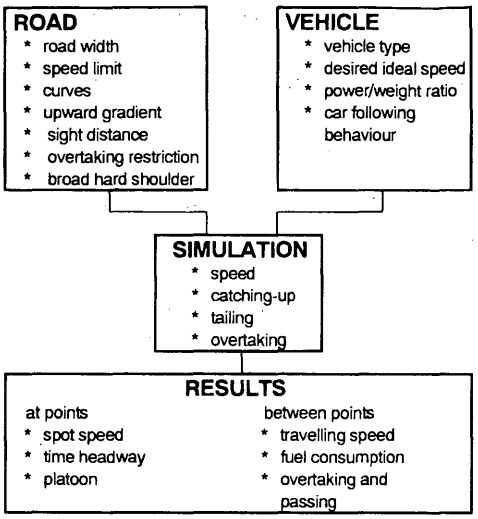



**FIGURE 2 Overview of VTI simulation model for two-way roads.**

### Calibration of Model

The following calibrations of the VTI model for Indonesian conditions were performed on the basis of the field studies:

- Base free-flow speed for ideal road and environmental conditions;
- Free flow speed at different road cross sections, roadside land use, and side friction;
- Speed in horizontal curves;
- Variation of speed around the average value (median speed);
- Distribution of used driving power for each vehicle class (determines the ability to retain the speed of a vehicle in an upgrade);
- Mean time headway value between vehicles in a platoon (mean time headway for constrained vehicles); and
- Overtaking functions, on the basis of a function named the Gompertz model. The expression for this function is as follows:

$$\omega = \exp[-A \exp(-ks)] \quad (1)$$

where

$\omega$ = overtaking probability (proportion of drivers accepting overtaking opportunity), $A$,
$k$ = calibration constants, and
$s$ = free sight distance (m).

The median value $s_{50}$, the sight distance that 50 percent of the drivers accept and the other 50 percent reject, is called the critical sight distance. It is described by the following equation:

$$S_{50} = \frac{\ln(A/\ln 2)}{k} = \frac{\ln A}{k} + \frac{\ln 1.443}{k} \quad (2)$$

Table 1 presents the calibration constants and the calculated overtaking probabilities for accelerating or flying overtaking with a visible oncoming vehicle. Drivers performing flying overtakings accepted very short sight distances, and there was only a weak influence of the speed of the overtaken vehicle.

**TABLE 1 Functions for Overtaking Behavior with Visible Oncoming Vehicles**

| Overtaken vehicle | Speed at overtaking | Type of overtaking | Calibr const Equation (1) | | Critical sight distance (m) |
|---|---|---|---|---|---|
| | | | A | $k*10^3$ | $s_{50}$ |
| LV, LB | ≥60 | acc. | 4.71 | 3.40 | 564 |
| LV, LB | <60 | acc. | 6.20 | 7.25 | 302 |
| MHV,LT | ≥45 | acc. | 6.71 | 4.00 | 568 |
| MHV,LT | <45 | acc. | 2.61 | 3.45 | 384 |
| LV | >0 | flying | 3.38 | 21.0 | 75 |
| LB | >0 | flying | 4.65 | 10.0 | 190 |
| MHV | >0 | flying | 3.50 | 18.0 | 90 |
| LT | >0 | flying | 3.55 | 11.9 | 137 |

### Validation of Model

The calibrated model was validated using observed journey speed, overtaking ratio, and degree of bunching data from three specially designated long-base sites ranging from 3 to 7 km. The field data were compared with corresponding data from simulation runs with the same road and traffic characteristics. Few differences above 3 km/hr between observed and simulated speed occurred; the average difference was about 1 km/hr. Similarly good correspondence between observed and simulated data was obtained regarding overtaking ratios and degree of bunching (leading headway <5 sec). It was therefore concluded that the model could be used to simulate the traffic process on Indonesian two-lane, two-way undivided (2/2 UD) roads.

## FREE-FLOW SPEED ANALYSIS

### Base Free-Flow Speed

Free-flow speed was determined for unobstructed vehicles defined as vehicles with a headway to the nearest vehicle in front of more than 8 sec and no recent or immediate meeting with a vehicle in the opposing direction (±5 sec). To evaluate the effect on free-flow speed of different site conditions, regression analysis was performed with travel time (TT) as dependent variable with the following regression equation:

$$TT = 1/V_{LV} = \text{constant} + B * X + C * Y + D * Z \ldots \quad (3)$$

where

$V_{LV}$ = speed of light vehicles (km/hr)
$X,Y,Z, \ldots$ = selected independent variables, and
$B,C,D, \ldots$ = regression coefficients.

When significant independent variables had been identified with stepwise multiple regression, the data base for 2/2 UD roads was normalized for a set of ideal conditions as follows:

- Carriageway 7 m wide,
- Shoulders 1.5 m wide and usable for parking but not driving,
- Undeveloped roadside land use,
- No side friction,
- No minor road access,
- Arterial road function, and
- Sight distance of 75 percent of the road section more than 300 m.

The free-flow speeds calculated for these conditions, called base free-flow speeds, were reviewed with simulation results and checked for consistency. The resulting base free-flow speeds used in IHCM are given in Table 2.

Terrain type describes the hilliness of the area through which a road passes and is defined by the total rise plus fall (in meters per kilometer) and the total horizontal curvature (in radians per kilometer) over the road segment (Table 3). (Values in parentheses were used to develop the graphs for standard terrain types in IHCM).

### Free-Flow Speed for Actual Conditions

The regression analysis described in the previous section showed that free-flow speed was affected primarily by roadway width, side friction, and road functional class (arterial, collector, or local). Figure 3 illustrates empirical results of the influence of carriageway width.

In Indonesia often a great deal of activity occurs at the edge of the road, both on the roadway and on shoulders and sidewalks, which interacts with the flow of traffic, causing it to be more turbulent and hurting capacity and performance. The following side friction events were recorded manually in the IHCM field surveys:

- PED: number of pedestrians, whether walking or crossing,
- PSV: number of stops by small public transport vehicles (motorized as well as nonmotorized) plus the number of parking maneuvers,
- EEV: number of motor vehicle entries and exits into and out of roadside properties and side roads, and
- SMV: slow-moving vehicles (bicycles, trishaws, horsecarts, oxcarts, etc).

**TABLE 2 Base Free-Flow Speed $FV_0$, 2/2 UD Road**

| | Base Free-Flow Speed FV0 (km/hr) | | | | |
|---|---|---|---|---|---|
| Terrain Type | LV | LB | MHV | LT | MC |
| Flat | 68 | 73 | 60 | 58 | 55 |
| Rolling | 61 | 62 | 52 | 49 | 53 |
| Hilly | 55 | 50 | 42 | 38 | 51 |

**TABLE 3 Definition of General Terrain Types**

| Terrain Type | Vertical Curvature: Rise + Fall (m/km) | Horizontal Curvature (rad/km) |
|---|---|---|
| Flat | <10 (5) | <1.0 (0.25) |
| Rolling | 10–30 (25) | 1.0–2.5 (2.00) |
| Hilly | >30 (45) | >2.5 (3.50) |

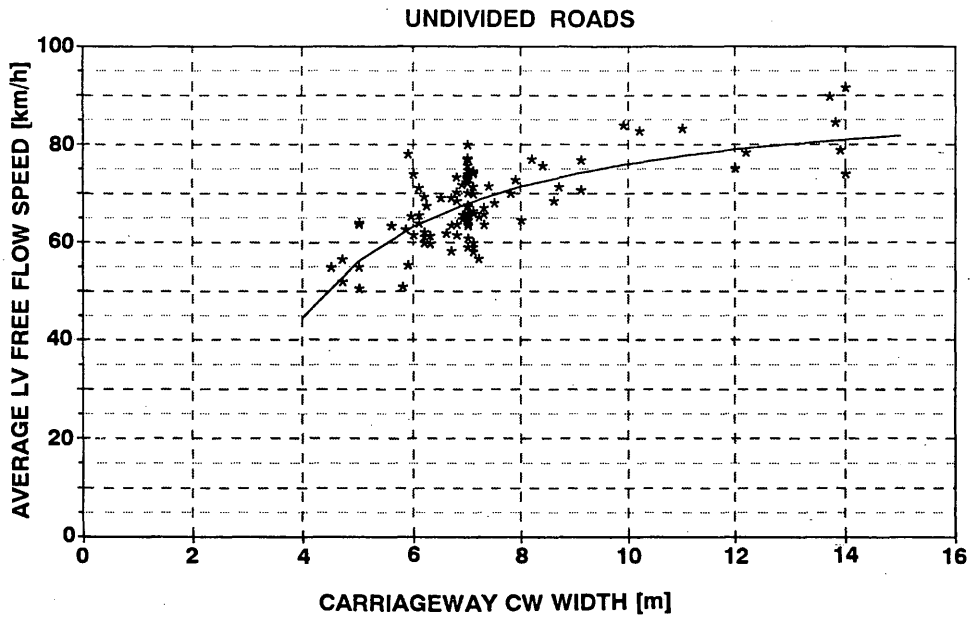



**FIGURE 3 Relationship between carriageway width and free-flow speed for light vehicles.**

To reduce the number of variables in the speed-flow analysis, a single measure of side friction (FRIC) was determined empirically equal to the sum of the weighted impacts of each of the four frictional items just described.

$$FRIC = 0.6 \times PED + 0.8 \times PSV + 1.0 \times EEV + 0.4 \times SMV$$

Five side friction classes relating to the value of FRIC were also predetermined:

| *Class* | *Value of FRIC* |
|---|---|
| Very low | <50 |
| Low | 50–149 |
| Medium | 150–249 |
| High | 250–350 |
| Very High | >350 |

The impact of side friction was shown to be related to shoulder width, with a 16-km/hr speed reduction for very high side friction at shoulder widths below 1.0 m, a 12-km/hr speed reduction at 1.5-m shoulder widths, and a 5-km/hr speed reduction at shoulders wider than 2 m. If there was no side friction, shoulder widths (range 0.5–3.0 m) had no significant impact on free-flow speed.

Road function class and land use had a speed reduction range from 0 km/hr for arterial and 0 percent land use to 11 km/hr for local and 100 percent land use (roadside development).

The actual free-flow speed for each vehicle type can be calculated in IHCM as follows:

$$FV = (FV_0 + FFV_W) \times FFV_{SF} \times FFV_{RC} \quad (4)$$

where

FV = free-flow speed for actual conditions (km/hr),
$FV_0$ = base free-flow speed for predetermined standard (ideal) conditions,
$FFV_W$ = adjustment for effective carriageway width,
$FFV_{SF}$ = adjustment for side friction, and
$FFV_{RC}$ = adjustment for road functional class and land use.

## DETERMINATION OF LIGHT-VEHICLE UNITS

In determining general speed-flow relationships for mixed traffic flows, it is necessary to convert the different vehicle types into a uniform unit. This is very important in a developing country because of the large variations in traffic compositions. Light-vehicle units, or LVU, were determined for each category instead of passenger-car units (PCU) because of the low frequency of passenger cars outside major cities in Indonesia. Free-flow speed for a passenger car is typically 5 to 10 km/hr higher than for an average light vehicle.

The primary methodology used for determining LVU was by means of simulation using the Swedish VTI model. Additional analysis using 5-min speed-flow data from selected sites in flat terrain was performed. The criterion of equivalency for speed-based LVU was the effect of different vehicle types on the speed of light vehicles, with LVU for light vehicles = 1.0.

Analysis was also performed using a capacity equivalency criterion for determining capacity-based LVU as described in the following.

### Speed-Based LVUs

*Determination of LVU by Simulation*

The simulation model was used to determine LVU for different general terrain types for roads with 7.0-m roadway width and medium side friction. The journey time for light vehicles ($TT_{LV}$) was observed as the dependent variable and calculated for each subsection of the roads and for their total length. The LVU value for each vehicle type was calculated according to following formula:

1. Define the journey-time-flow constant for LV at 100 percent LV in the traffic flow as

$$\Delta T_{LV} = \left(\frac{1}{V_{1,200}} - \frac{1}{V_{600}}\right) \cdot 3{,}600/600 \quad (5)$$

where $V_{600}$ and $V_{1,200}$ are the space mean speeds (km/hr) for LV at traffic flow 600 and 1,200 veh/hr with 100 percent LV. The flow levels represent a normal traffic flow range for interurban roads. The calculation is done in journey time since the speed-based LVU is determined regarding the effect of different traffic compositions on space mean speed.

2. Define in the same way the journey-time-flow constant for LV at the proportion $p$ of LV and the proportion $1 - p$ of Type $X$ as

$$\Delta T_{p,X} = \left(\frac{1}{V_{1,200}} - \frac{1}{V_{600}}\right) \cdot 3{,}600/600 \quad (6)$$

where $V_{600}$ and $V_{1,200}$ now are the space mean speeds (km/hr) for LV at traffic flows with the proportion $p$ of LV.

3. The LVU for vehicle type X can now be calculated as

$$p\Delta T_{LV} + (1 - p)\,\alpha\Delta T_{LV} = \Delta T_{p,X} \quad (7)$$

where $\alpha$ is the LVU value.

In this way LVU values for three terrain types were obtained, as shown in Table 4. Since the simulation model did not include motorcycle traffic, the LVU for MC could not be determined using this method.

*Determination of LVU from Empirical Speed-Flow Data*

The assumptions underlying this regression analysis were that the speed-flow relationship is linear and that LVU therefore could be determined from least-square fits of speed-flow samples with different traffic composition:

$$V_{LV} = A - K_{LV} \cdot Q_{LV} - K_{MHV} \cdot Q_{MHV} - \ldots - K_{MC} \cdot Q_{MC}$$

where

$V_{LV}$ = speed (km/hr),
$A$ = constant representing free-flow speed,
$Q$ = traffic flow for each vehicle type (veh/5 min), and
$K$ = speed reduction effect caused by specific vehicle type.

The LVU were obtained as the ratio between the $K$-coefficient for a specific vehicle type and for light vehicles: for example,

$$LVU_{MHV} = K_{MHV}/K_{LV}$$

For flat roads, MHV = 1.5, LB = 1.2, LT = 2.7, and MC = 0.8. The LVUs for MHV, LB and LT were thus very similar with the results obtained with simulation.

### Capacity-Based LVUs

At levels of high traffic flow, most vehicles are traveling in platoons. Speed-based LVUs do not represent the relative impact of different vehicle types at such conditions, leading to a need for a second set of LVUs based on a capacity equivalency criterion.

**TABLE 4 Speed-Based LVU Determined by Means of Simulation**

| Terrain Type | Vehicle Class | | |
|---|---|---|---|
| | MHV | LB | LT |
| Flat | 1.5 | 1.4 | 2.7 |
| Rolling | 2.0 | 1.6 | 4.5 |
| Hilly | 4.5 | 1.8 | 12.5 |

Capacity-based LVU can be determined by analyzing time headways of a single stream of vehicles during congested conditions (*4*). According to this method, the headway for a particular vehicle type (e.g., MHV) can be determined as follows:

$$LVU_{MHV} = H_{MHV}/H_{LV}$$

where $H_{MHV}$ is the mean headway between an MHV following an MHV during the passage of the stop line or conflict point, and $H_{LV}$ is the mean headway between an LV following an LV during the passage of the stop line or conflict point.

Since the Indonesian surveys covered only a few cases of single-stream flow at capacity level on rural roads, headway data for bunching conditions (headway < 5 sec) on such roads were analyzed. Significant LVU results were obtained only for MHV in flat terrain (LVU = 1.2). The other values were obtained in combination with engineering judgment (Table 5). Transition from speed-based to capacity-based LVU in IHCM is based on the actual traffic flow (in vehicles per hour) in a number of predefined flow classes for different road types. Capacity-based LVU are used for determining the degree of bunching.

### Synthesis of LVU Results

The simulation results were compared with LVU results obtained from regression analysis of the speed-flow data. The resulting recommended speed-based LVU are presented in Table 5. The table also gives the capacity-based LVU obtained as described earlier.

## SPEED-FLOW RELATIONSHIPS

### Speed-Flow Modeling Using Simulation

The VTI simulation model was used to produce speed-flow relationships for two-lane, two-way roads in different terrain types. For

**TABLE 5 Synthesis of LVU Results for 2/2 UD Roads**

| Terrain Type | Criterion for Equivalency | LVU | | | | |
|---|---|---|---|---|---|---|
| | | LV | MHV | LB | LT | MC |
| Flat | Speed | 1.0 | 1.5 | 1.2 | 2.7 | 0.8 |
| | Capacity | 1.0 | 1.2 | 1.5 | 2.0 | 0.25 |
| Rolling | Speed | 1.0 | 2.0 | 1.3 | 4.0 | 0.6 |
| | Capacity | 1.0 | 1.3 | 1.7 | 2.5 | 0.25 |
| Hilly | Speed | 1.0 | 3.5 | 1.5 | 5.5 | 0.4 |
| | Capacity | 1.0 | 1.5 | 2.0 | 3.0 | 0.25 |

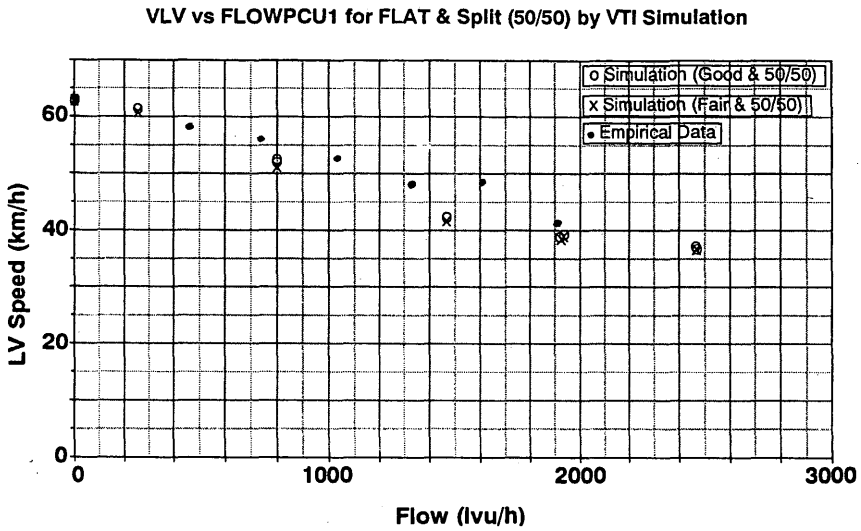



**FIGURE 4 LV speed (km/hr) as function of flow (lvu/hr) for flat roads.**

flat roads the simulations were performed with two sight conditions, one with 75 percent sight distance above 300 m ("good"), and one with 40 percent sight distance above 300 m ("fair").

Simulations were performed for five traffic flows: 250, 600, 1,200, 1,800, and 2,200 veh/hr. At every flow level a traffic composition of 63 percent LV, 25 percent MHV, 8 percent LB, and 4 percent LT was used, which represented average conditions for the studied roads. Motorcycle traffic was not included in the simulation. Simulations were also made with different directional splits in the traffic flow: 50/50, 60/40, and 70/30. For each case of flow and directional split, two simulations were made to decrease the random effects in the simulation. The results were plotted as speed-flow diagrams, with the flow presented in speed-based LVUs per hour.

Figure 4 presents the speed for light vehicles as a function of the flow for the flat road with 50/50 directional split. The speed-flow curve is almost linear up to 1,700 LVU/hr, at which level there appears to be a knee in the curve. The free-flow speed is 63 km/hr, and the speed at 2,800 LVU/hr is 37 km/hr. The difference between good and fair sight distance conditions is noticeable only in the middle flow range, which may be due to the tendency of Indonesians to overtake with very small accepted sight distances. Empirical speed-flow observations for other roads with the same general characteristics are also shown in the figure; they appear to indicate a more linear relationship.

Simulations of the flat road with different directional splits showed that this variable had very little impact on average speed, maybe because that speed decrease in the direction with more traffic is compensated by a speed increase in the other direction.

Figure 5 presents the speed-flow curve for a road in rolling terrain at different directional splits. The free-flow speed is 54 km/hr. The

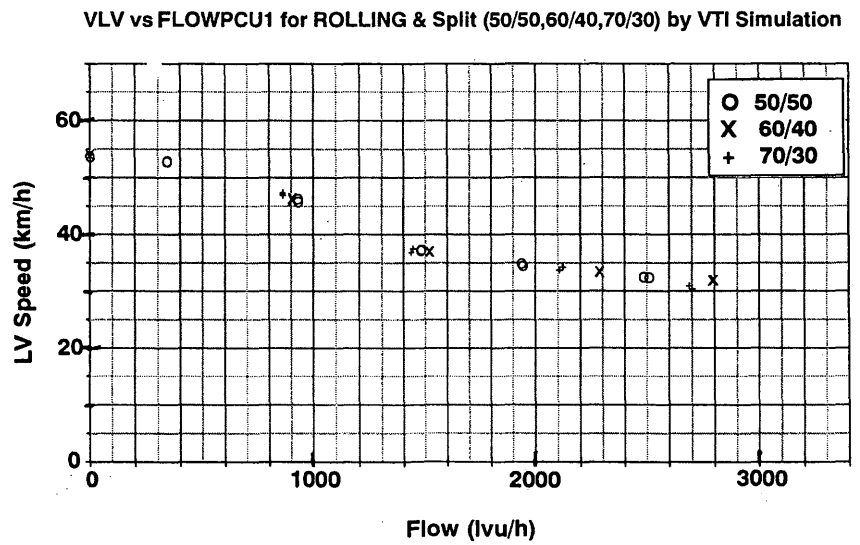



**FIGURE 5 LV speed for road in rolling terrain with different directional splits.**

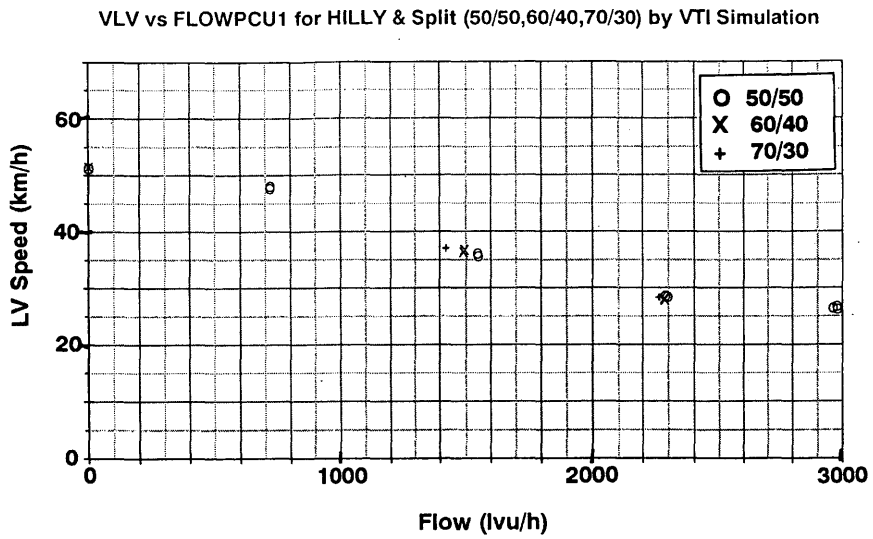



**FIGURE 6 LV speed for road in hilly terrain as function of directional split.**

shape of the curve is essentially the same as for the flat road, but there is a more clear S-shape, with the steepest drop in speed between 800 and 1,600 LVU/hr. At 2,800 LVU/hr the speed is about 31 km/hr.

The result for the hilly road is presented in Figure 6. The free-flow speed is 51 km/hr. The curve looks very much the same for the rolling road. The steepest drop in speed is between 800 and 1,800 LVU/hr. At a flow of 3,000 LVU/hr, the speed is 27 km/hr.

### Speed-Flow Modeling Using Empirical Data

Aggregated short-base 5-min speed-flow data were used to test different speed-flow and speed-density models for flat conditions. Each sample in this data base represents the average speed and flow value for all observed 5-min periods falling in predetermined flow classes for each site. The data base covered 123 sites with a total of 546 sets of flow class average observations.

The impact of site conditions (road width, side friction, land use, road function class, sight distance class) were analyzed with multiple regression. The samples in the data base were then normalized for site conditions deviating from the predetermined ideal conditions for each road class as described earlier.

Speed-density and speed-flow regressions were then made for each road class (e.g., 2/2 UD road 6.5 to 7.5 m wide) with the following models: linear speed-flow model, single-regime, Underwood, May. The linear speed-flow model ($R^2 > 0.6$) was selected for 2/2 UD roads (Figure 7). There was no apparent knee in the relationship as obtained with the simulation. Similar linear

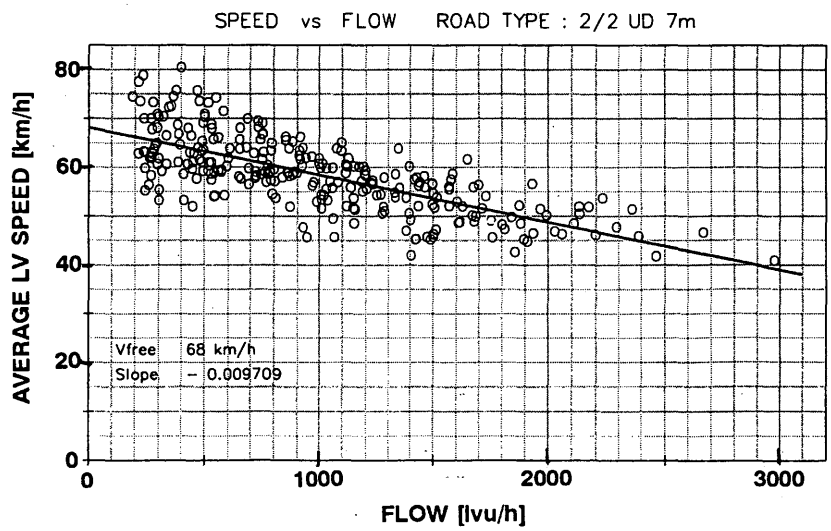



**FIGURE 7 Speed-flow relationship for LV, undivided 7-m-wide flat terrain.**

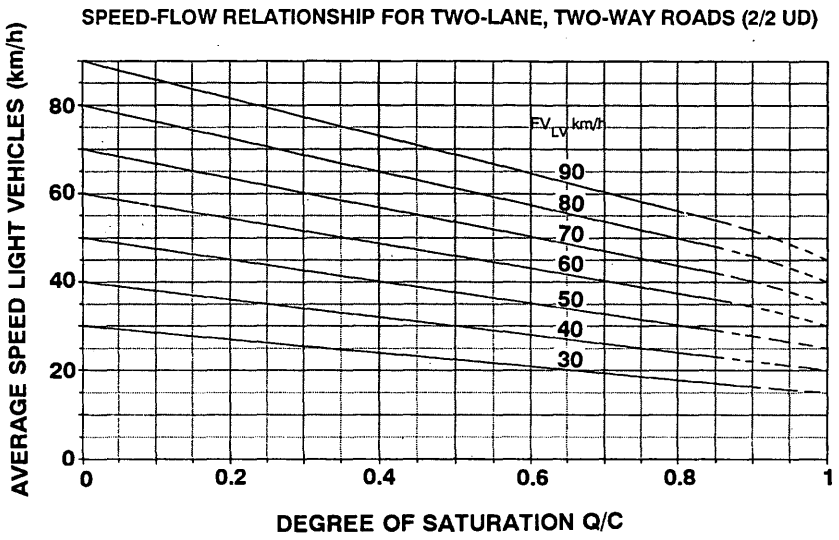



**FIGURE 8 Diagram for determination of speed as function of free-flow speed and degree of saturation ($Q/C$).**

relationships were obtained for each vehicle type, with the lines converging at a speed of 35 to 40 km/hr at a flow level of 2,900.

## Estimation of Capacity

The capacity of 2/2 UD roads was estimated in the following ways:

1. Direct observation of speed and flow rate averages per 5 min. Because of the lack of road sections where the observed maximum flow could be clearly identified as representing the capacity of the road section itself and not of an adjacent bottleneck, only a few observations had been made, of which the highest ranged from 2,800 to 3,000 LVU/hr (Figure 7).
2. Observation of flow rates during short periods of simultaneous bunching conditions in both directions (headways $< 5$ sec). These observations indicated a capacity ranging from 2,800 to 3,100 LVU/hr.
3. Theoretical estimation from speed-flow-density modeling, showing capacity of about 3,000 LVU/hr occurring at a density of 81 LVU/km.

The conclusion at this stage of the IHCM project was that the base capacity of a 2/2 UD, straight, 7-m-wide road with no side fric-

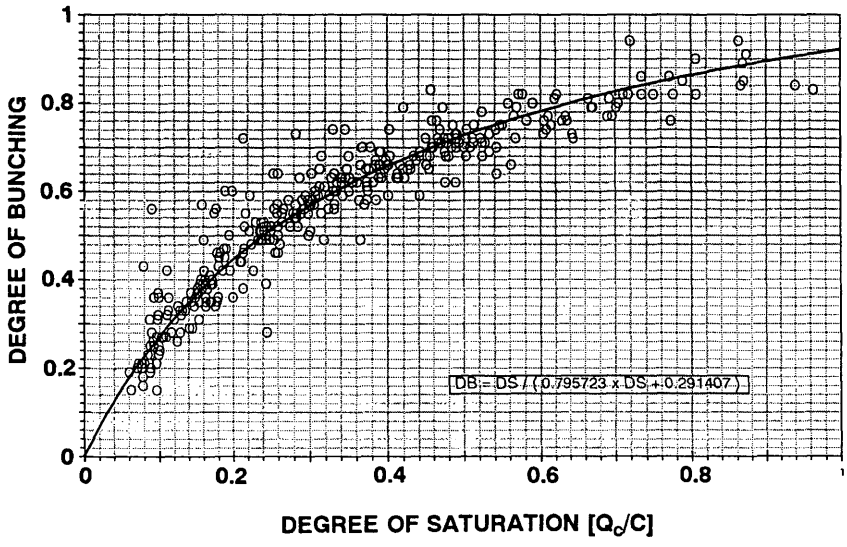



**FIGURE 9 Diagram for determination of degree of bunching for 2/2 UD roads.**

tion and shoulders of more than 1 m is 2,900 LVU/hr. Capacity is calculated in IHCM as a function of road width (5 to 10 m: adjustment factor 0.81–1.21), side friction/shoulder width [very high (< 0.5 m) to very low (>2 m)]: adjustment factor 0.74–1.03, and directional split (50/50 to 70/30: 1.0–0.88).

## SYNTHESIS

The results from the simulation and empirical analysis of speed-flow data taken with the capacity estimations showed that the actual average speed for a particular vehicle type in a mixed flow could be predicted from information of free-flow speed and flow/capacity equals degree of saturation. A diagram for this purpose, which is used in the interim IHCM (*2*), is shown in Figure 8, using a linear speed-flow curve. The method has some similarity to the approach used in the 1992 revision of the HCM (*5*) and in the HDM-Q model (*6*). Similar diagrams based on single-regime speed-flow curves (*7*) have been proposed for multilane roads.

The degree of bunching can also be predicted in IHCM from information on *Q/C* as shown in Figure 9.

The speed-flow relationships and parameter values reported in this paper are the results from the interim IHCM for interurban roads. They are being reviewed and may be revised in the final version, to be published in 1996. The speed-flow relationship may therefore be revised in the shape of a two-part linear model with a breakpoint at a degree of saturation of 0.85. The proposal to use both speed- and capacity-based LVU is also subject to further study.

## REFERENCES

1. Bang, K.-L., T. Bergh, and N. W. Marler. *Highway Capacity Manual: Part I, Urban Roads.* 09/T/BNKT/1993. Directorate General of Highways, Indonesia, Jan. 1993.
2. Bang, K.-L., and A. Carlsson. *Interim Manual for Interurban Roads and Motorways.* Indonesian Highway Capacity Manual Project, Directorate General of Highways, Bandung, Indonesia, Aug. 1994.
3. Brodin, A., and A. Carlsson. *The VTI Traffic Simulation Model: A Description of the Model and the Program System.* Report 321 A. Swedish Road and Traffic Research Institute, Linköping, 1986.
4. Salter, R. J. *Highway Traffic Analysis and Design,* rev. ed. MacMillan Press Ltd., London and Basinstoke, England, 1976.
5. *Special Report 209s: Supplement to Highway Capacity Manual,* rev. Chapter 7, 1st ed. TRB, National Research Council, Washington, D.C., 1992.
6. Hoban, C. J., and Archondo-Callao. *Highway Design and Maintenance Model HDM-III with Congestion Analysis Capabilities.* Infrastructure and Urban Development Department, World Bank, Washington, D.C., 1994.
7. Easa, S. M., and A. D. May. Generalized Procedure for Estimating Single- and Two-Regime Traffic-Flow Models. In *Transportation Research Record 772,* TRB, National Research Council, Washington, D.C., 1980.

*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Simulation-Based Approach to Evaluating Optimal Lane Staffing Requirements for Toll Plazas

VICTOR GULEWICZ AND JOHN DANKO

A comprehensive approach is presented, using a General Purpose Simulation System World simulation model, to evaluate the optimal lane staffing requirements necessary to satisfy, at an acceptable level of service, off-peak demand at a toll plaza. The approach provides an effective tool to analyze requirements, at a quantitative level, for varying levels of traffic volume and mix, thereby providing an acceptable level of service to patrons while maximizing the efficiency of the toll plaza operation. In addition, the evaluation resulted in the identification and application of a level-of-service criterion for determining toll plaza lane staffing requirements that would result in an acceptable level of performance for Port Authority toll plazas. Last, the approach allowed for modeling the toll plazas to an extremely detailed level of accuracy, which resulted in a high degree of confidence in the model's capabilities to predict plaza operations.

The Port Authority of New York and New Jersey is a public agency responsible for promoting and facilitating trade, commerce, and transportation in the New York–New Jersey region. It is a self-supporting bistate agency that provides, operates, and maintains many transportation facilities. The current demands for these facilities are especially severe on the six tunnel and bridge crossings that link New York and New Jersey at key locations within the metropolitan region. These crossings are the George Washington Bridge, which connects northern New Jersey to Manhattan; the Lincoln and Holland tunnels, which link New Jersey and the Manhattan central business district; and the Outerbridge, Goethals, and Bayonne bridges, which connect Staten Island with New Jersey. Patrons traveling toward New York through each of these crossings must stop and pay a toll.

By its nature, a toll plaza can become a major congestion bottleneck that can severely impede vehicular movements in a metropolitan area. Long queues are particularly evident during the morning and evening peak travel periods, when the number of vehicles arriving sometimes greatly exceeds the tolls processing capacity at a given crossing. Conversely, during the off-peak periods, opportunities exist to maximize the efficiency of the plaza operation by determining the minimum lane staffing requirements needed to meet an acceptable service standard.

Recently, the Interstate Transportation Department (ITD) of the Port Authority requested that the corporate industrial engineering unit, Management Engineering and Analysis (ME&A), assist in evaluating and updating toll lane and staffing requirements at its tunnel and bridge crossings. Historically, these efforts have been accomplished with mathematically based models using manual or computer-assisted computations. Although they provided generally acceptable results, the models often failed to predict toll plaza operations accurately.

Recognizing that traffic volumes at crossings vary over time, ITD staff were interested in developing an analytical approach that could be used to assess alternative staffing scenarios and to select the most cost-effective strategy for each facility. Specifically, the effectiveness included vehicle queueing levels, rate of vehicle throughput, and wait time in queue. On the basis of research and past experience, ITD and ME&A staff concluded that a lane-by-lane assessment was required and could be accomplished using computer-based simulation modeling techniques.

## STUDY OBJECTIVE AND SCOPE

The objective was to develop an approach for determining the toll plaza lane and staffing requirements needed to provide an acceptable service level during off-peak periods of varying levels of traffic volume and mix. The focus on the off peak was purposeful since peak-period volumes require staffing of all available toll lanes, eliminating the possibility of evaluating alternative lane staffing schedules.

Given the scope and complexity of the study, a phased project plan was agreed upon, beginning with a data collection program and the development and validation of the simulation model for the eight-lane toll plaza at the Outerbridge Crossing (OBX). This paper will focus on the results of the Phase 1 effort for the OBX.

## SERVICE LEVEL STANDARDS

As stated previously, the operational impacts of alternative scenarios for lane and toll collector staffing must be predicted with a reasonable degree of accuracy. Furthermore, these impacts should define the performance of the toll plaza in terms of the level of service, or LOS (normally expressed as an index of discomfort), that the user experiences. As an initial step, the study team conducted research on the use of service level standards for toll plazas.

A literature search identified a number of documents for review, including the *Highway Capacity Manual* (HCM) (*1*). In addition, the team contacted four research organizations, 16 transportation properties, and six consulting firms with transportation and traffic engineering experience in an attempt to identify any applicable standards being used in the design, evaluation, and management of toll plazas.

Port Authority of New York and New Jersey, Management Engineering and Analysis Group, One World Trade Center, Floor 63 South, New York, N.Y. 10048.

**TABLE 1 LOS Criteria for OBX**

| LOS | Average Waiting Time (sec) | Average Queue (no. of veh) |
|---|---|---|
| A | ≤ 5 | ≤ 1 |
| B | ≤ 15 | ≤ 3 |
| C | ≤ 25 | ≤ 5 |
| D | ≤ 40 | ≤ 8 |
| E | ≤ 60 | ≤ 11 |
| F | > 60 | > 11 |

As a result of the search, the study team found that generally recognized or accepted service level standards do not exist for evaluating toll plaza performance. Furthermore, for the organizations contacted, most indicated that they depend on the experience and judgment of management and operating personnel instead of formal standards for determining toll lane and staffing requirements that would result in an acceptable level of performance.

A noted exception to this approach was identified for the New Jersey Highway Authority (NJHA), which operates the Garden State Parkway. In the late 1980s NJHA retained the services of Vollmer Associates to develop an approach for determining the number of toll booths required at each of its plazas to achieve an acceptable level of performance (*2*). Vollmer recommended the application of LOS criteria for a signalized intersection, finding very similar processing characteristics to that of a toll lane. For a signalized intersection the measure of user discomfort, and therefore LOS, is the amount of time stopped at the signal. Similarly, for a toll lane, user discomfort can be measured as the time stopped in a queue waiting to be processed. This stopped time is equal to the total of the transaction time(s) for each vehicle in the queue ahead of that user.

On the basis of the results of the search, the team recommended and gained ITD's concurrence to apply the approach identified by Vollmer for evaluating toll plaza performance, using the LOS values in the HCM of average stopped delay per vehicle for signalized intersections (Table 1) (*1*). The LOS values for stopped delay were translated by the study team into the number of vehicles queued by considering the average transaction time per vehicle. The parameter of "vehicles queued" relates well to the physical characteristics of a toll plaza, thereby providing a less abstract characterization of plaza performance.

After considering this research, ITD management decided that the range of service level C to D would be its operating goal.

## DATA COLLECTION APPROACH

To identify applicable time frames for toll lane and staffing analyses and obtain the required data for model development, the study team developed and conducted an aggressive data collection program. An initial step in this evaluation was obtaining a definition of seasonality of traffic demand for the facility. It was determined that October through March and April through September were most representative of the "off-season" and "season" months, respectively. The study team then obtained daily traffic volumes for the facility from July 1991 to June 1992. Distributions of the daily vehicle volumes were calculated, and the 85th-percentile demand days given here (which have been the standard used by the transportation industry to design and evaluate facilities) were selected for each:

| | *Season* | *Off-Season* | *Variance (%)* |
|---|---|---|---|
| Weekday | 39,190 | 35,773 | −9.5 |
| Weekend day | 42,522 | 38,653 | −10.0 |

The percentage variance between the season and off-season and for each day type was then calculated; if it exceeded 5 percent, the team and facility staff concurred that this difference was attributable to seasonal aspects of traffic volume. On the basis of the percentile demand analysis, four data collection days were identified for the OBX for the weekday and weekend day scenarios during the season and off-season time frames. For each scenario, a data scanning was performed, which involved documenting at 5-min intervals the lane status (open or closed) and the vehicle queue. The vehicle queueing data for each of the eight lanes were then averaged and plotted against the number of lanes opened. An example of a scanning plot can be found in Figure 1.

Those time frames exhibiting low queue conditions and a high number of lanes open were identified as candidates for evaluating alternative lane and staffing requirements. As such, these time frames were identified for data collection:

| | *Season* | *Off-Season* |
|---|---|---|
| Weekday | None | 4:00 p.m.–12:00 a.m. |
| Weekend day | 6:00–11:00 a.m. | 11:00 p.m.–11:00 a.m. |

## MODEL STRUCTURE

The toll staffing simulation model was developed to provide the ability to analyze lane-by-lane queueing as a result of modifications to lane staffing. To achieve this level of detail, the model was devel-

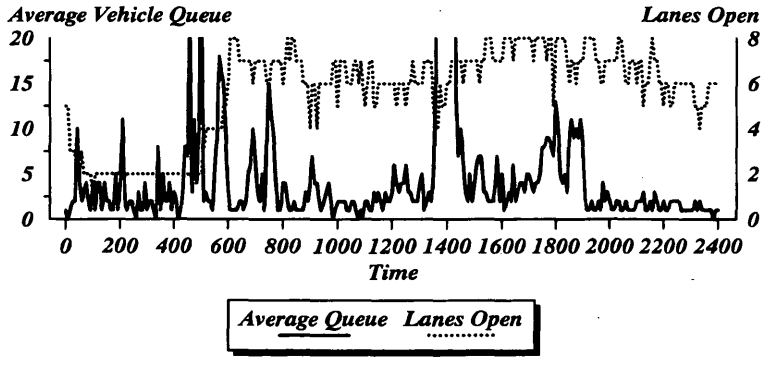



**FIGURE 1 Example of OBX scanning graph.**

oped using GPSS World, which is a general application simulation language run on a high-speed personal computer.

In the model, each vehicle is generated and processed individually, which allows for a greater level of accuracy. However, this level of detail resulted in the model structure containing approximately 15,000 lines of programming code, requiring that the team investigate ways to compact the model structure to keep it manageable. Through the use of several high-level GPSS World coding techniques, the structure (number and position of lanes) of the toll plaza being simulated in each model is created in memory as the model runs, rather than physically defining the plaza structure using standard coding methods. Doing this saves a substantial amount of programming lines and allows the simulation of a typical 8-hr period to run in approximately 1 min.

For model input, vehicle interarrival data were collected independently for each arrival lane and by vehicle type (car, bus, light truck, and heavy truck). The data were then fed into a statistical software package to determine if the arrivals "fit" a standard theoretical distribution (pattern) of arrivals. The package, CurveFit, was used to determine which standard distribution fits the closest to the field data. The software selects the distribution on the basis of the "shape" of the arrival pattern and whether it passed both the Kolmogorov-Smirnov and Anderson-Darling tests. By passing these tests, the data are confirmed to be a good fit. Use of the standard theoretical distributions in the model was preferred because it accounts for all possibilities of randomness in that pattern and eliminates the statistical anomalies that occur in raw data, which is the case for when the actual empirical distributions are used. Different arrival distributions were created (fit) for each 1/2-hr period to allow for changes in the volume of vehicles that arrive at different times during the day.

The lane selection data, which describe how vehicles choose which toll lane to enter, were also collected for each scenario at the OBX. To collect these data, a random sample of vehicles entering the toll plaza was observed. For each observed vehicle, the bridge arrival lane (left or right), queue in each of the eight toll lanes (even Numbers 2 through 16), and the destination toll lane were recorded. The OBX toll plaza layout can be found in Figure 2.

From the recorded data, four general driving habits were observed:

1. Most drivers enter a toll lane on the same side of the toll plaza from which they exited the span.
2. Most drivers, once they have selected which half of the plaza to enter, select the lane with the shortest queue on that side.
3. Some drivers were observed entering the lane with the shortest queue, even though an empty lane was available. It was believed that these patrons focused on only the other vehicles in the plaza and not on the signals above the toll lanes. This caused them to believe that the empty lanes were closed.
4. A small percentage of drivers appeared to choose a toll lane randomly.

These driving habits were replicated in the simulation model by using percentages to create a lane selection algorithm, developed by the study team, that determines the probability of any vehicle's entering any lane in the toll plaza based on the level of queueing. An example of the lane selection algorithm used in the model is shown in Figure 3. This algorithm for the OBX season weekend day revealed that drivers tend to stay on the same side of the toll plaza from which they arrived, with 84.2 percent of the drivers arriving from the left lane choosing Lanes 2 through 8 (left side) and 66.3 percent of the drivers arriving from the right lane choosing lanes 10 through 16 (right side). The 33.7 percent that arrived from the right lane and moved to the left side was higher than the 15.8 percent moving from the left to the right side. The study team observed that this is caused by trucks that dominate the right arrival lane. Cars arriving in the right lane tend to move more to the left to avoid the truck queueing on the right side of the toll plaza.

The processing rate data are composed of both the move time (the time required for the next vehicle to move into and out of the toll booth) and the transaction time (the time required for the patron to pay the toll). Empirical distributions were developed for the move time for each of the four vehicle classes and transaction times for three possible payment types (using cash with change returned, exact change, or a pass/ticket).

The OBX toll relief and meal record, which is the staffing plan (schedule) for the facility, was used to simulate minute by minute when each toll lane was scheduled to be open, accounting for personal breaks, meal breaks, and lane closures.

## MODEL VALIDATION

A simulation model is beneficial only if its results are known to be accurate. Because of the significant impacts of the results of the toll

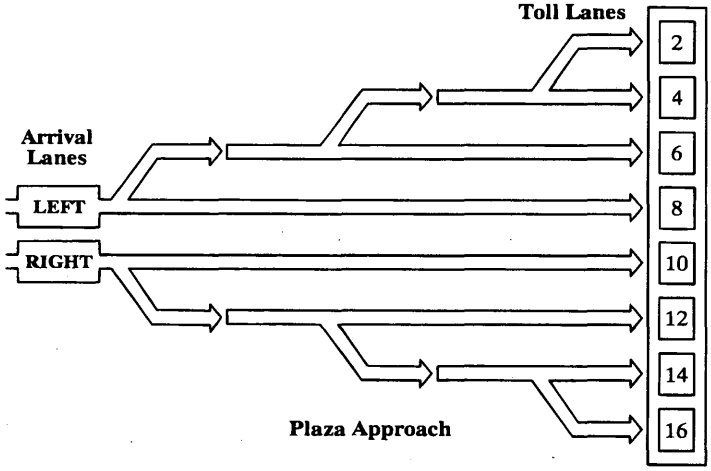



**FIGURE 2 OBX toll plaza layout (not to scale).**

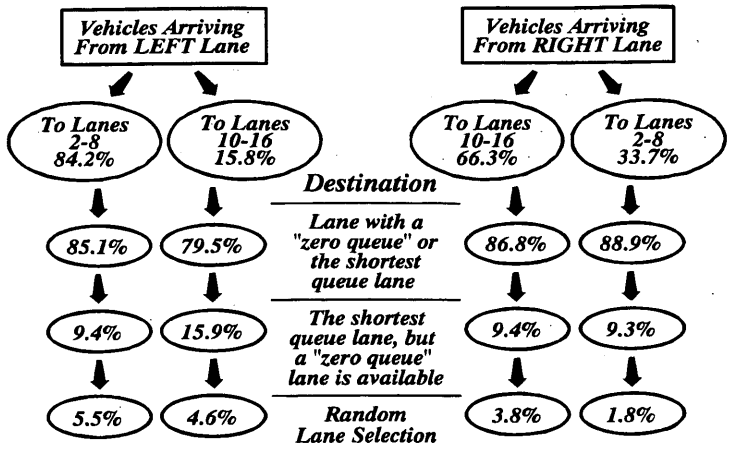



**FIGURE 3 Lane selection algorithm.**

staffing simulation model, the study team decided to validate the model using two variables: vehicle queue and throughput.

For the OBX, the season weekend scenario was selected for validation. During the data collection on Sunday, August 30, 1992, the vehicle queue in each lane and the total vehicle throughput were recorded at 1-min intervals from 6:00 to 11:00 a.m. Additionally, all lanes were recorded as being open (green light) or closed (red light) at 1-min intervals.

The OBX model was configured on the basis of data recorded on August 30, 1992, and was then programmed to tabulate the individual lane queue and total toll plaza throughput at 1-min intervals to match the field data collection method. The results of the queue validation can be seen in Figure 4. The study team thought that it was necessary to create one value that would indicate the level of the model's accuracy. So in addition to the visual comparison, the absolute differentials in average vehicle queue between the field-measured data and the model's calculated value were tabulated at 1-min intervals. The average for the scenario time frame (6:00–11:00) was then calculated for this variance, which the study team defined as the average deviation (average error) value for the model. For the OBX validation, the average deviation in queueing was 1.2 percent. This can be translated as the model having an average of 1.2 percent error in estimating vehicle queue or, conversely, the model being 98.8 percent accurate in predicting vehicle queue, which is evidence that the model capabilities are confirmed. In addition, the throughput graph (Figure 5) reveals similar results, with an average deviation in vehicle throughput of 0.52 percent.

From the results of the validation, the study team was confident that the model can accurately predict vehicle processing at the OBX.

## MODEL SCENARIOS

Using the four OBX scanning scenarios mentioned earlier, the study team determined that three scannings, off-season weekday, season weekend day, and off-season weekend day, offered the potential for modifying lane staffing requirements.

For each scenario, the OBX model structure was modified to simulate the vehicle arrival, lane selection, and processing rate patterns based on the data collected for each. Once the structures were developed, a batch of 10 model runs using varying random number streams was conducted for each to create the "typical" demand during that time frame at the facility. The results of the 10 runs were then averaged, creating the typical day results.

## RESULTS

For each of the scenarios, the base condition model was run first to determine the baseline results. The base condition models simulate

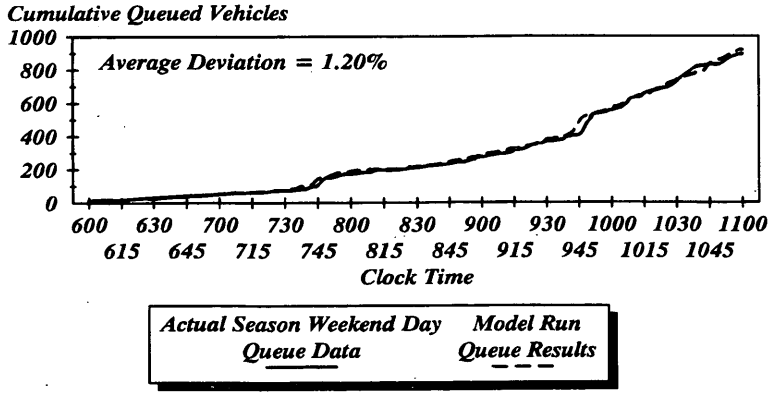



**FIGURE 4 OBX validation queue graph.**

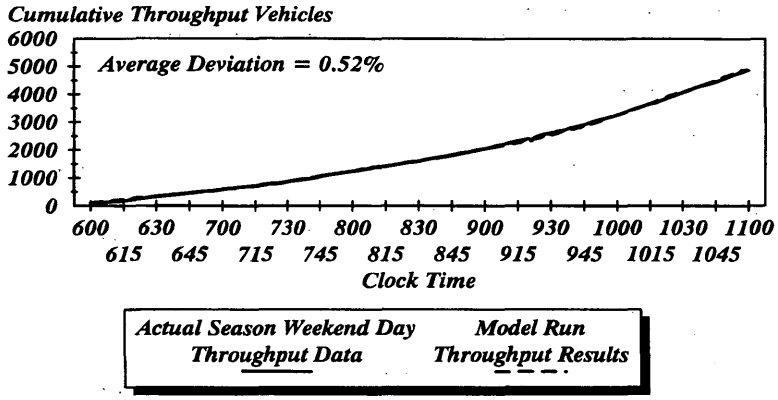



**FIGURE 5 OBX validation throughput graph.**

toll plaza operations using the existing OBX toll collector staffing plan. The average vehicle queue per lane was tracked for each run throughout the scenario time frame. This variable was chosen by the team because it is a good indicator of the overall operating condition of the toll plaza.

### OBX Off-Season Weekday Results

For the OBX off-season weekday scenario, the analysis time frame based on the results of the scanning was identified as 4:00 p.m. through 12:00 a.m. The results for all vehicles in the base condition (no staffing changes) can be found in Figure 6.

For the base condition, over the time frame being analyzed, the plaza was operating within LOS A. The lowest level resulting from the existing staffing plan was LOS C, albeit for a very short duration.

For all of the scenarios, the study team decided that the lane that was open for the longest time during the analysis time frame would always be the first to be closed to determine the queueing impacts. This was done to represent the worst-case possibility, since it would create the largest change in lane availability. For the OBX off-season weekday, the first lane to be closed was Lane 14. With Lane 14 closed from 4:00 p.m. to midnight, the simulated results indicated minimal change in queueing from the base condition.

The next step was to determine the impact of closing an additional lane. Lane 8 was open the next longest in the time frame, so it was the next to be closed from 4:00 p.m. to midnight. Preliminary runs with Lane 8 closed revealed that severe queueing occurred between 4:00 and 7:00 p.m., resulting in LOS F, but for the rest of the time frame, only two small spikes into LOS D were the worst queueing obtained. From these results, the team decided that Lane 8 could be closed only from 7:00 p.m. to midnight. With Lanes 14 and 8 closed, LOS A or B was maintained throughout most of the time frame (Figure 7). Since the two small spikes created by closing Lane 8 moved beyond the maximum LOS C into LOS D for only a total of about 5 min, and no other significant change was found, the team concluded that closing Lane 8 would be acceptable. Further lane closures were not possible, as additional runs revealed that severe queueing would result.

### OBX Season Weekend Results

For the OBX season weekend scenario, the analysis time frame based on the results of the scanning was identified as 6:00 to 11:00 a.m. For the base condition, operating conditions during most of the time frame were within LOS A. The results for all vehicles in the base condition (no staffing changes) can be found in Figure 8.

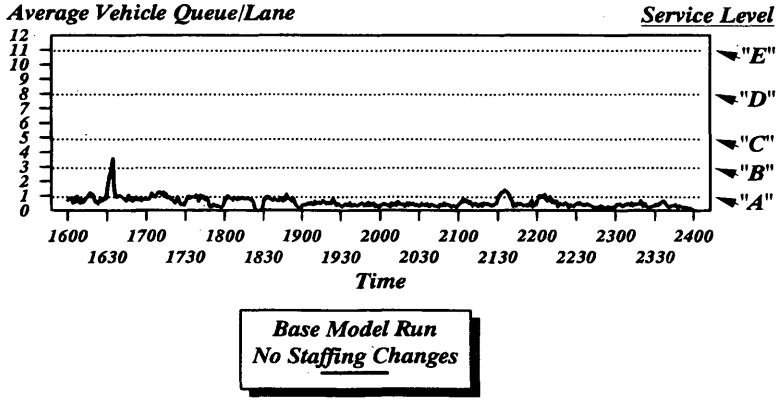



**FIGURE 6 OBX base condition queue graph, off-season weekday.**

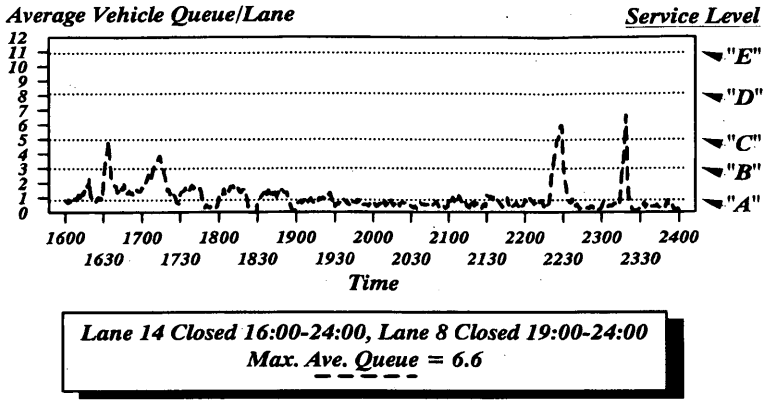



**FIGURE 7 OBX modified lane staffing queue graph, off-season weekday.**

For this scenario, the first lane to be closed was Lane 12. With Lane 12 closed, the results indicated minimal change in queueing from the base condition.

Since Lane 8 was opened the second longest amount of time, it was the next to be closed. Preliminary runs with Lane 8 closed revealed that the queue spiked into LOS F from 7:45 to 8:30 a.m. with a maximum average queue of 29.3 vehicles (Figure 9). Otherwise, LOS C or better was maintained. Since Lanes 12 and 8 will be closed in the modified staffing plan, the existing relief toll collectors will have fewer lanes to cover. If the relief schedule is modified so that relief toll collectors open Lane 8 between 7:45 and 8:30 p.m., the queueing pattern will revert to LOS B during these 45 min, which was observed during the previous run. Thus, the team determined that Lane 8 could be closed (with some minor shifting of breaks) from 6:00 to 11:00 a.m. with no detrimental customer impacts. Further lane closures were not possible, as additional runs revealed excessive queueing to LOS F.

### OBX Off-Season Weekend Results

For the OBX off-season weekend scenario, the analysis time frame was identified as 11:00 p.m. to 11:00 a.m. For the base condition, the average queue was within LOS A or B for almost the entire time. The results for all vehicles in the base condition (no staffing changes) can be found in Figure 10.

For this scenario, the first lane to be closed was Lane 12. With Lane 12 closed, the results indicated no change in queueing from the base condition.

Lane 8, being open the second longest, was then closed. Preliminary runs with Lane 8 closed revealed that the queue spiked into LOS F from 11:00 p.m. to midnight, 8:00 to 8:30 a.m., and 10:00 to 11:00 a.m. Otherwise, the queue levels maintained LOS C or better. If Lane 8 were not closed until midnight and adjustments could be made for relief toll collectors to open Lane 8 from 8:00 to 8:30 a.m. and 10:00 to 11:00 a.m., the queueing would, at its worst point, revert to LOS C. The team determined that Lane 8 could be closed from midnight to 11:00 a.m. (Figure 11). Further lane closures were not possible, as additional runs revealed that substantial queueing would result in LOS F.

## CONCLUSIONS

On the basis of the results of the modified lane staffing simulation analyses, the team recommended that the current lane staffing plan for the OBX be modified as described in this paper.

The lane modifications can be phased into the current staffing plan by not back-filling the lanes indicated to be closed in the event of an unexpected schedule vacancy (e.g., when a toll collector calls in sick or requests a day off). In this way, monitoring of the queue-

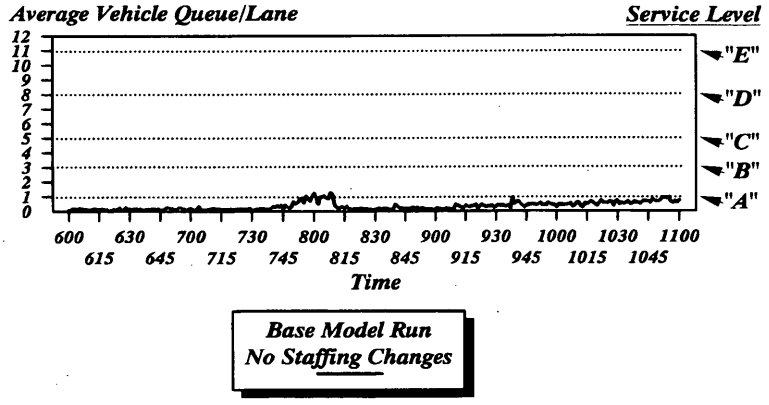



**FIGURE 8 OBX base condition queue graph, season weekend day.**

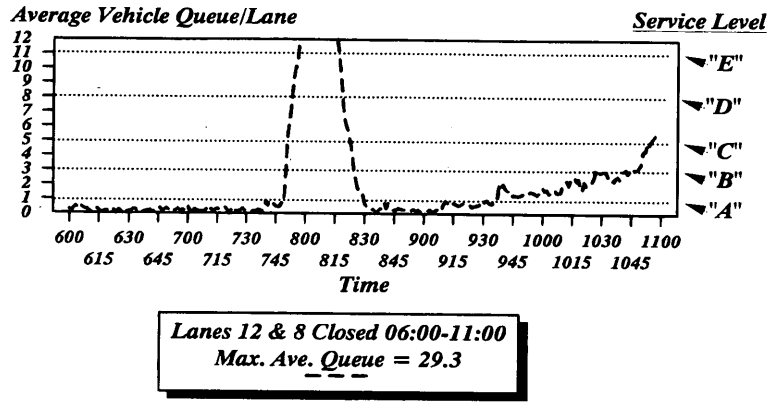



**FIGURE 9 OBX modified lane staffing queue graph, season weekend day.**

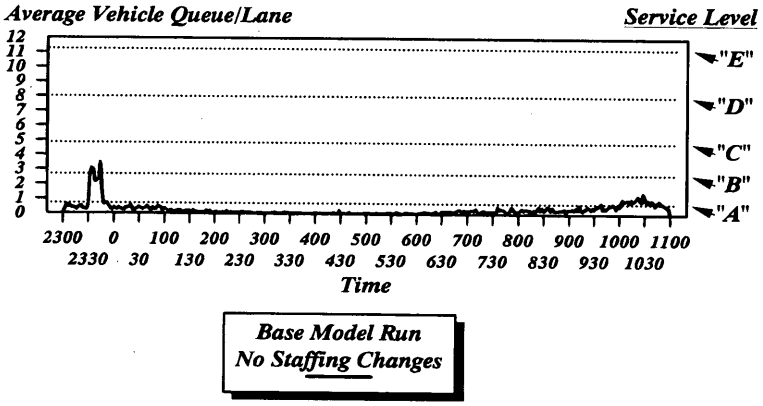



**FIGURE 10 OBX base condition queue graph, off-season weekend day.**

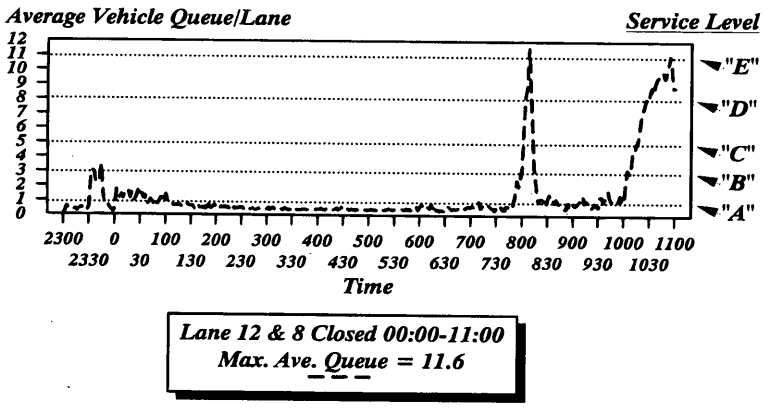



**FIGURE 11 OBX modified lane staffing queue graph, off-season weekend day.**

ing impacts of lane closures as vacancies occur, before full implementation, can further field-validate the schedule changes at the facility.

The recommended plan will reduce toll lane staffing hours by 5.8 percent a year, resulting in an annual cost savings of $114,000. The next step for the study team will be to continue the analysis at the remaining Port Authority tunnel and bridge facilities so that a systemwide service improvement plan can be developed and implemented.

## REFERENCES

1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
2. Nielsten, G. *Evaluation of the Traffic, Level of Service and Needs at the Hillsdale, Bergen, Essex, Union and Raritan Toll Plazas.* Technical Memorandum. Vollmer Associates, New York, 1988.

*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Reconciling Estimated and Measured Travel Times on Urban Arterial Streets

KENNETH G. COURAGE, RANDALL H. SHOWERS, AND DOUGLAS S. MCLEOD

Although the *Highway Capacity Manual* (HCM) modeling process is widely accepted, there is evidence of significant disparity between estimated arterial travel speeds and speeds measured in the field. The HCM suggests that speed and travel time values measured in the field are preferable to the computed estimates. Often the validity of the model estimates may be challenged by competing interests. The primary objective was to reconcile the differences between estimated and measured travel times on arterial streets. The principal product is a set of recommendations for modifying the estimation and measurement procedures to reduce the disparity between them. Four tasks were involved: (*a*) examine the arterial speed computational methodology to identify sources of disagreement with field measures performed with moving vehicles, (*b*) compare a large sample of measured travel speeds with travel speed estimates carried out using the HCM methodology, (*c*) develop candidate adjustment factors that can be applied in practice to improve the agreement between estimated and measured speeds, and (*d*) test the candidate adjustment factors against the field data and recommend specific modifications to the travel time procedures and the HCM model. Although the HCM models were intended for analyses on the planning and operational level, the focus here is on planning applications. The main difference between planning and operational analyses is the levels of detail of the input data and in the required level of accuracy of the results. It is important in either case that the travel time estimates be unbiased, that is, the procedures should not consistently underestimate or overestimate the travel times. The results offer a reasonable explanation for the apparent discrepancies between estimated and measured travel times and delays on arterial streets. They also provide a practical means of adjusting the estimated delay values to produce a very close agreement with the corresponding measurements.

Florida's efforts at growth management have gained national attention and respect for their policy content as well as their technical methodology. One of the main features of the system is a mandate for periodic assessment of the level of service (LOS) for public facilities—more specifically, roads. For arterial streets the current LOS evaluation criterion is average travel speed. The *Highway Capacity Manual* (HCM) (*1*) provides a technique for estimating the average traffic speed based on known values of traffic volumes and signalized intersection capacities. The Florida methodology relies heavily on the HCM technique.

Florida has its own LOS manual (*2*) that assists local agencies in applying the HCM model at a planning level. The Florida LOS manual includes software for performing the computations, tables for deriving approximate estimates, guidelines for preparing input data, and limitations on the acceptable values for assumed parameters.

K. G. Courage and R. H. Showers, University of Florida, Transportation Research Center, 512 Weil Hall, Gainesville, Fla. 32611. D. S. McLeod, Florida Department of Transportation, 605 Suwannee Street, Tallahassee, Fla. 32399.

## STATEMENT OF PROBLEM

The HCM modeling process is widely accepted, but there is evidence of significant disparity between estimated arterial travel speeds and speeds that are measured in the field (*3*). The HCM suggests in Chapter 11 that speed and travel time values measured in the field are preferable to the computed estimates. Often the validity of the model estimates may be challenged by competing interests, especially in growth management applications.

In a properly timed arterial traffic control system, it is usually possible to travel progressively through several signals without stopping. Field studies often show no delay at all for specific travel time runs. The HCM model always predicts some delay at each intersection. It must be understood that the HCM model is a deterministic approximation of a stochastic process. It is not expected to produce an accurate prediction of the travel time for each run; however, it should be able to produce an unbiased estimate of the average travel time over several runs. Recent evidence indicates that in some cases the HCM method tends to overestimate average travel times to a degree that cannot be overlooked (*4*).

A more reliable method is required for estimating vehicular delay and travel times on arterial streets without the need for moving-vehicle studies. The accuracy of such a method must be adequate at least for planning purposes. Moving-vehicle studies are very labor-intensive and cannot be applied to the hypothetical situations that generally are involved in planning applications.

## STUDY OBJECTIVES AND TASKS

The primary objective of the study was to reconcile the differences between estimated and measured travel times on arterial streets. The principal product of the study is a set of recommendations for modifying both the estimation and measurement procedures to reduce the disparity between them. Four tasks were involved:

1. Examine the arterial speed computational methodology to identify sources of disagreement with field measures performed with moving vehicles.
2. Compare a large sample of measured travel speeds with travel speed estimates carried out using the HCM methodology.
3. Develop candidate adjustment factors that could be applied in practice to improve the agreement between estimated and measured speeds.

4. Test the candidate adjustment factors against the field data and recommend specific modifications to the travel time procedures and the HCM model for use in the Florida's LOS manual.

The HCM models were intended for analyses at the planning and operational levels, but this study focused on planning applications. The main difference between planning and operational analyses is the levels of detail of the input data and in the required level of accuracy of the results. It is important in either case that the travel time estimates be unbiased: the procedures should not consistently underestimate or overestimate the travel times. This is because they are treated as deterministic models for decision-making purposes.

## BACKGROUND DISCUSSION

### HCM Chapter 11 Model

The structure of the HCM Chapter 11 model is illustrated in Figure 1. Note that the intersection delay, as determined in HCM Chapter 9, is an important element of this model. The average arterial speed is determined by dividing the distance between intersections (Points 1 and 2) by the total travel time between the points. The total travel time is determined as the sum of the running time and the total intersection delay. The running time is obtained from HCM Table 11-4 as a function of arterial classification, signal density, and free speed. The intersection delay is obtained from the HCM Chapter 9 analysis.

Figure 1 also shows a typical time-space trajectory for a moving vehicle between Points 1 and 2. Each vehicle traveling on this segment of roadway will have a different time-space trajectory. The essence of the Chapter 11 model is a representation of the two travel time elements shown in Figure 1 as a deterministic approximation of the "average" vehicle's trajectory.

### Moving-Vehicle Studies

Moving-vehicle studies may be carried out in several ways with different levels of instrumentation. Since the LOS basic measure of effectiveness as specified by the HCM is average travel speed, it is theoretically necessary to measure only the total time required to travel a given roadway segment or section for comparison purposes. This can be done easily with nothing more than a stopwatch. However, since the HCM model computes the total travel time as the sum of two components—running time and stopped delay time—the candidate adjustment factors needed to reconcile the two techniques must be applied separately to each travel time element. This requires that both travel time elements be available from both techniques. For this reason, detailed time-space trajectories were obtained for each moving-vehicle run.

### Sources of Bias

There are three general reasons for discrepancy between estimated and measured travel times on an arterial roadway: errors in the data, deficiencies in computational procedures, and conceptual differences between procedures.

#### *Errors in Data Used by Computational Procedures*

Data errors can be caused by field errors; however, it is more likely that they will result from the use of assumed values for data items that are very difficult or costly to measure accurately in the field—examples are saturation flow rates, turning movement volumes, and traffic signal timing. The study procedures for each of these data items are simple and straightforward, but it is very difficult to cover an entire roadway section simultaneously with moving-vehicle studies. Furthermore, it is not possible to guarantee that each data item is applied at the exact moment of passage of the moving vehicle through the system. Thus, particularly for planning studies, it is necessary to rely on assumptions and approximations in developing the input data for travel time estimates.

It could be argued that errors in the input data could affect the results either way (i.e., underestimate or overestimate), but most of the data items have a direct effect on the volume/capacity (v/c) ratio. The v/c ratio, in turn, has a nonlinear influence on delay. An overestimation in the v/c ratio will produce a much larger error in the delay estimate than a corresponding degree of underestimation. This introduces a bias toward overestimation of travel times, an

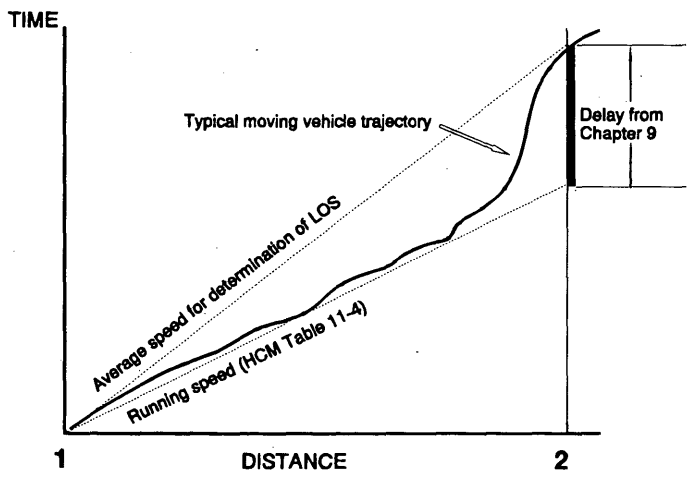



**FIGURE 1 Existing Chapter 11 model structure.**

effect that can be worse during the peak periods when the operation is fully saturated (i.e., v/c approaching 1.0).

### *Deficiencies in Computational Procedures*

Some apparent deficiencies in the HCM methodology could produce large discrepancies between estimated and measured travel times. For example, the running time estimates obtained from HCM Table 11-4 are based only on free speed and signal density for a given arterial classification. They are assumed to be independent of traffic volume, number of lanes, and other parameters that could be expected to influence running speeds. The running speed estimates are particularly vulnerable to situations in which speeding occurs because most agencies are reluctant to recognize an operating speed that exceeds the speed limit.

The direct application of the Chapter 9 delay model to estimate the stopped delay at signalized intersections is also questionable. The main problem is that the delay is computed in Chapter 9 as the sum of two components. The first component estimates the delay that would occur if all vehicles arrived at the intersection with uniform spacing. The second component is a correction factor that accounts for randomness in the arrival pattern, including the "cycle failures" that result when the arrivals exceed the approach capacity for one or more cycles. This component is called the incremental delay term.

This model is entirely appropriate when applied at isolated signals; however, the direct extension to arterial routes with coordinated signals is somewhat difficult to rationalize. In coordinated arterial systems the "metering" effect of the upstream signal can be expected to reduce the randomness of the arrivals at the downstream signal. In particular, it is reasonable to anticipate a much lower occurrence of cycle failures at the downstream signal because temporary overcapacity situations are cushioned by the upstream signal. Therefore, the application of the incremental delay term equally at all intersections can be expected to overestimate the total delay, and therefore the average travel time.

Another important factor is the effect of progression quality on delay. This is incorporated in the HCM model as the progression factor (PF). It is common practice to assume Arrival Type 4 on coordinated arterial streets. This implies a progression correction factor ranging from 0.7 to 0.9. The limits and default values for the computation of the progression factor are explained in HCM Table 9-13, uniform delay ($d_1$) adjustment factor, DF.

Figure 2 shows the distribution of progression factors obtained by running TRANSYT-7F (*5*) on several available data sets to generate a sample of approximately 100 links. The progression factor was established by running TRANSYT-7F twice for each data set: once with coordination and once without. Note that a large proportion of the observations fell outside of what is generally accepted as the range for Arrival Type 4.

### *Conceptual Differences Between Estimation and Measurement Procedures*

In the discussion of the first two sources of bias, discrepancies between the estimated and measured values generally would be resolved in favor of the measured values. In other words, both procedures are addressing the same phenomena and the differences would be attributed to shortcomings in the input data or the estimation procedure. In this case, the two procedures are addressing different phenomena by the same name.

The average travel time computed by the HCM model applies to all vehicles on the approach, regardless of arrival time. On the other hand, the measured travel time applies primarily to the subset of vehicles within the progression band. Clearly, the vehicles receiving the benefits of progression may be expected to have higher overall speeds than the rest of the vehicles. Therefore, it should not be surprising that the results of moving-vehicle studies are more optimistic than the HCM estimates.

This raises an interesting philosophical question: which of the two speeds is more appropriate as an LOS criterion? Since the HCM defines LOS, it could be argued that the estimated values are the only ones compatible with the HCM. On the other hand, the LOS criterion is intended to be based on motorist perception of disutility, and it is reasonable to propose that it is best applied to the coordinated arterial traffic flow. Theoretically, the two definitions will converge as traffic volumes approach their capacity.

## STUDY DESCRIPTION

The main objective of this paper is to identify sources of bias and recommend adjustments that will eliminate the bias between the field data and estimation models. To make comparisons, moving-vehicle data and arterial data are needed. The overall study procedure is illustrated in Figure 3. Field data on travel time trajectories and arterial characteristics were furnished by consultants under contract to the Florida Department of Transportation (FDOT). Five counties in the southeast part of Florida containing both urban and rural roadways were represented; Miami and Fort Lauderdale urban areas were predominant. The moving-vehicle study locations and general characteristics of the sample are described in general in Table 1 and in more detail in Table 2. A total of 656 runs were included, covering 161 km (100 mi) of arterial routes containing 316 signalized intersections. A sample summary of the data and graphics for each route is presented in Figure 4. The moving-vehicle equipment and study methodology are described elsewhere (*6*).

The level of detail in the arterial characteristics varied from location to location. In general, the normal planning level requirements for field data were greatly exceeded, but default values were used where necessary to perform the estimates of travel time and delay. In some cases $g/C$ ratios were observed in the field; in other cases they were determined from signal timing records. Some observations of progression quality were made by observing the proportion of vehicles arriving on the green phase. Where field observations were not carried out, a default arrival type was assigned on the basis of peak direction. Concurrent mid-block traffic counts were carried out to obtain representative 15-min volumes on all routes. Recent peak-period turning movement counts were used where available to estimate the proportion of turns leaving the arterial approaches at each intersection. Where no turning movement counts were available, a default value (usually 12 percent) was applied. Saturation flow rates were observed in some cases, and representative default values were applied in others.

The HCM estimates of travel time and delay were performed using the ART-PLAN spreadsheet program (*7*). ART-PLAN performs a straightforward implementation of the HCM Chapter 11 methodology. A sample ART-PLAN analysis summary is presented in Figure 5. Versions of ART-PLAN have been developed for both the 1985 and 1994 methods. The modified estimates of travel time

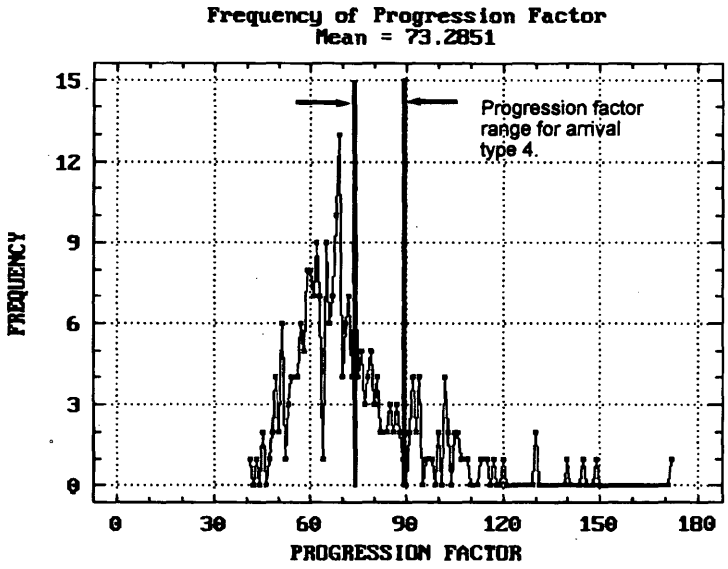



FIGURE 2 Progression adjustment factor variation.

and delay resulting from adjustment factors were performed using a combination of standard programming methods for data analysis.

The reduced data for all of the results for each run (i.e., measured travel time estimated by various methods) were combined into a unified data base. The analysis of the reduced data was performed by a combination of dBase and SAS procedures.

## INITIAL COMPARISON OF ESTIMATED AND MEASURED TRAVEL TIMES

The comparison of unadjusted travel time estimates with the field data produced very discouraging results. Considering the complete sample, the estimated travel time estimates averaged 3,000 percent higher than the corresponding measured values. Substituting the 1994 HCM Chapter 11 method, the degree of overestimation was reduced to 974 percent. Clearly, the unadjusted data require further attention.

Inspection of the data indicates that the discrepancy is concentrated in a relatively small proportion of the cases, each of which has an unreasonably high v/c ratio. The v/c ratio is very sensitive to the values used for traffic volume, saturation flow rate, and *g/C* ratio. Accurate modeling requires very precise data for all of these items. The level of required accuracy generally exceeds the accuracy normally associated with planning level data, which rely heavily on assumptions and approximations. Planning estimates of these

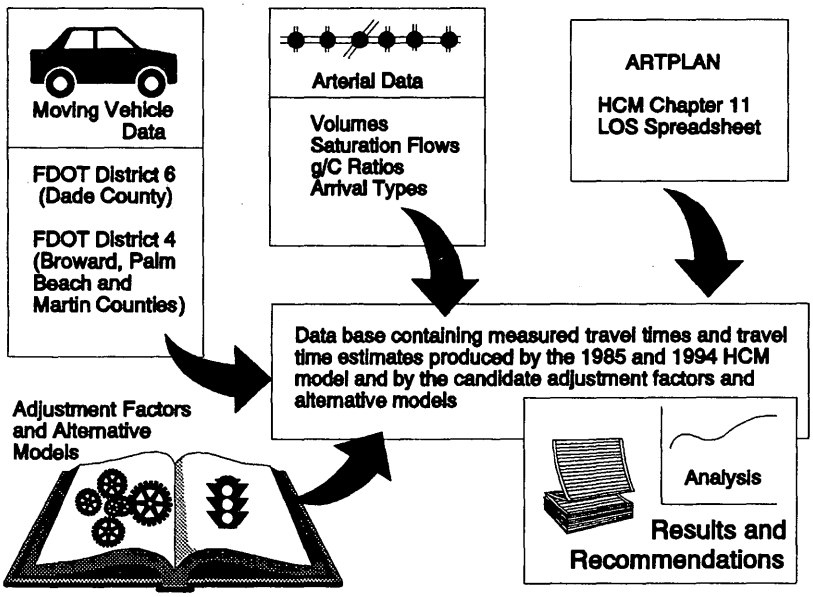



FIGURE 3 Study procedure and data flow.

**TABLE 1 Summary of Travel Time and Delay Study Characteristics**

| | Total | Breakdown by FDOT District | | Breakdown by HCM Arterial Class | |
|---|---|---|---|---|---|
| | | 4 | 6 | I | II |
| Route kilometers[1] | 161 | 95 | 66 | 102 | 60 |
| Total runs | 656 | 512 | 144 | 405 | 251 |
| Run kilometers[1] | 2223 | 1589 | 685 | 1542 | 732 |
| Number of signals | 316 | 182 | 134 | 178 | 138 |
| Signals per route | 6.9 | 6.2 | 9.4 | 7.0 | 6.7 |
| Signals per km[1] | 2.0 | 2.0 | 2.0 | 1.9 | 2.3 |

[1] 1 km = 0.6 mi.

data items generally will not support accurate modeling. This is very difficult to address in an adjustment factor.

The HCM model is defined to be valid for v/c ratios of less than 1.2. When v/c ratios exceed this threshold in the field, the result is extensive and prolonged congestion. Several of the routes represented in the field data had computed v/c ratios much greater than 1.2. A logical alternative would be to discard all cases in this category; however, the proposed model must be able to deal with such cases. A more practical candidate adjustment would involve placing an upper limit of 1.2 on the computed v/c ratio.

In a report describing the travel time data collection and capacity analysis performed in District 4 (*8*), the consultant indicated that the calculated speeds were reasonably close to the measured speeds, except when the *g*/*C* ratio was less than 0.4. The consultant recommended that future projects of this nature devote more effort to collecting turning movement counts to reduce the dependence

**TABLE 2 Summary of Routes for Travel Time and Delay Studies**

| Area/Route Name | Number of Signals | Signal Density sig/km[1] | Route Length km[1] | Class & Speed kph[1] | Number of Studies |
|---|---|---|---|---|---|
| DISTRICT 6 (Dade County, Florida) | | | | | |
| NW 42 St/LeJune Rd A | 4 | 2.6 | 1.5 | II/65 | 12 |
| NW 42 St/LeJune Rd B | 4 | 2.4 | 3.4 | II/65 | 12 |
| SW 152 St | 5 | 1.3 | 4.0 | I/73 | 12 |
| NW 125 St (1) | 6 | 2.4 | 2.4 | II/56 | 16 |
| NW 125 St (2) | 11 | 2.6 | 4.2 | I/56 | 16 |
| NW 125 St (3) | 9 | 3.9 | 2.3 | II/56 | 16 |
| NW 79 St East | 15 | 2.1 | 7.0 | I/65 | 16 |
| SR 860 MGD | 16 | 2.1 | 7.7 | II/65 | 12 |
| Bird Rd (1) | 5 | 2.7 | 1.9 | II/65 | 15 |
| Bird Rd (2) | 11 | 2.2 | 5.0 | II/65 | 12 |
| SW 87 Ave (1) | 6 | 1.3 | 4.7 | I/65 | 16 |
| SW 87 Ave (2) | 14 | 1.6 | 9.0 | I/65 | 12 |
| Red Road/SW 57 Ave | 7 | 1.1 | 6.2 | I/65 | 16 |
| NW 72 Ave | 2 | 0.6 | 3.4 | I/56 | 14 |
| NW 107 Ave South | 5 | 2.1 | 2.3 | I/65 | 13 |
| DISTRICT 4 (Broward, Palm Beach and Martin Counties, Florida) | | | | | |
| SR 5/Federal Hwy | 4 | 0.9 | 4.5 | I/73 | 42 |
| SR 5/US 1 | 6 | 1.0 | 6.3 | I/65 | 18 |
| SR 870/Commercial Blvd | 4 | 1.2 | 3.2 | I/73 | 51 |
| SR 7/US 441 | 6 | 1.5 | 3.2 | I/73 | 62 |
| SR 845/Powerline Rd | 8 | 2.0 | 4.0 | II/73 | 39 |
| SR 858/Hallandale | 4 | 1.7 | 2.4 | II65 | 46 |
| SR 808/Glades Blvd | 4 | 1.9 | 2.1 | I/73 | 48 |
| SR 870/Commercial Blvd | 6 | 1.8 | 3.2 | I/56 | 41 |
| ST 805/S Dixie Hwy | 3 | 1.8 | 1.6 | II/56 | 40 |
| SR 824/Pembroke Rd | 4 | 1.9 | 2.1 | II/56 | 49 |
| SR 816/Oakland Pk Blvd | 12 | 3.0 | 4.0 | I/65 | 50 |
| SR 7/US 441 | 9 | 2.8 | 3.2 | I/65 | 52 |
| SR 814/Atlantic Blvd | 8 | 3.2 | 2.4 | II/56 | 60 |
| SR 824/Broward Blvd | 7 | 3.6 | 1.9 | II/56 | 47 |
| SR 858/Hallandale | 12 | 3.6 | 3.2 | I/65 | 66 |

[1] 1 km = 0.6 mi.

| END POINT | LINK LENGTH (ft) | TRAVEL TIME (sec) | DELAY (sec) | STOPS | ACCELE-RATION NOISE | SPEED (MPH) AVG. | RUN. |
|---|---|---|---|---|---|---|---|
| NW 6TH AVE | 1438 | 34.4 | 1.7 | 0.2 | 1.99 | 28.5 | 30.0 |
| NW 4TH AVE | 656 | 18.0 | 2.5 | 0.2 | 1.86 | 24.7 | 28.7 |
| DIXIE HWY | 1245 | 122.7 | 85.1 | 1.0 | 1.98 | 6.9 | 22.6 |
| NE 1ST AVE | 194 | 7.3 | 0.0 | 0.0 | 2.40 | 18.1 | 18.2 |
| US 1 SR 5 | 1752 | 73.6 | 31.4 | 0.6 | 2.16 | 16.2 | 28.3 |
| NE 8TH AVE | 715 | 22.6 | 2.4 | 0.2 | 2.39 | 21.5 | 24.1 |
| NE 10TH AVE | 593 | 24.0 | 7.6 | 0.1 | 1.92 | 16.8 | 24.6 |
| NE 14TH AVE | 1364 | 50.6 | 17.6 | 0.6 | 1.98 | 18.3 | 28.2 |
| LAYNE BLVD | 1366 | 40.6 | 9.0 | 0.4 | 2.40 | 22.9 | 29.5 |
| GOLDEN ISLES | 305 | 21.5 | 11.3 | 0.4 | 2.49 | 9.6 | 20.3 |
| DIPLOMAT PKWY | 490 | 15.7 | 2.2 | 0.0 | 1.89 | 21.2 | 24.6 |
| THREE ISLES BLV | 472 | 23.4 | 9.3 | 0.3 | 2.43 | 13.7 | 22.8 |
| TOTAL | 10597 | 455.0 | 180.7 | 4.4 | 2.26 | 15.8 | 26.3 |

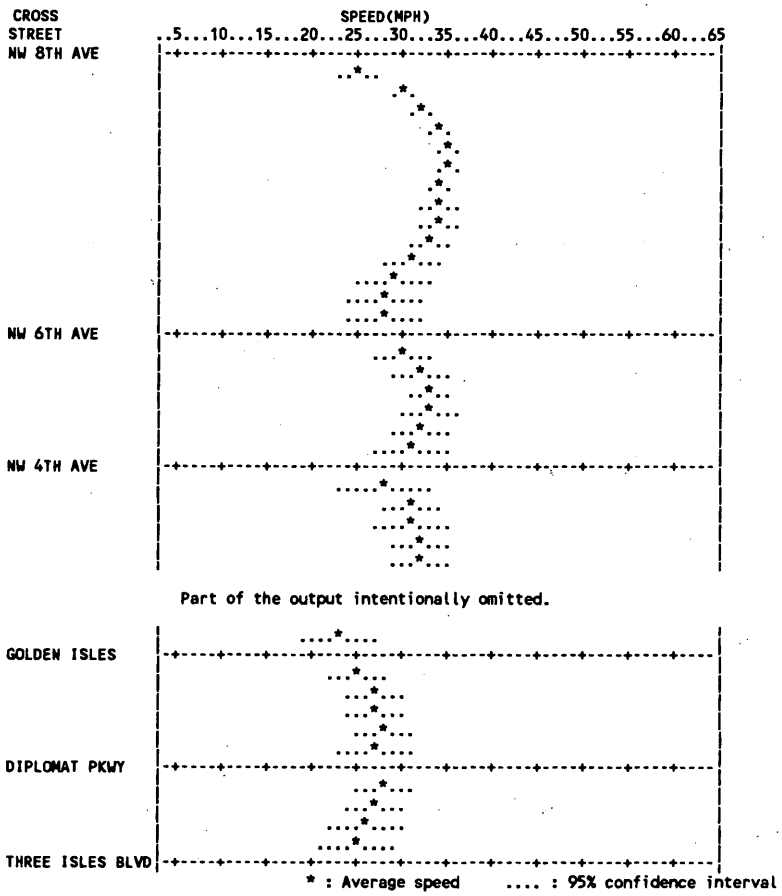



**FIGURE 4 Sample MVRAP output summary of travel time study sample.**

on assumed values for turning movements. It was also suggested that deficiencies in the computational methodology could be at fault.

As a first-level screening technique, the proposed v/c limit of 1.2 was applied to the data. This reduced the travel time discrepancy to 43 percent for the 1985 HCM model and 37 percent for the 1994 model. Although neither of these results could be considered satisfactory, they do set the stage for the development of adjustment factors that could reconcile the discrepancy. The difference between the 1985 and 1994 computational methods is relatively small. Since the 1994 method has been approved for use in the HCM, it is a logical choice over the now obsolete 1985 method.

In the rest of the analyses described in this paper, the 1994 HCM Chapter 11 model will be used and an upper limit of 1.2 will be imposed on all v/c ratios. The results associated with these conditions will be referred to as the base values.

## DEVELOPMENT OF CANDIDATE ADJUSTMENT FACTORS

Each of the sources of bias described previously must be addressed independently in the development of adjustments. The goal of this exercise is to improve the travel time estimation procedures within the existing structure of the HCM model. Departures will be proposed only when they can be shown to produce worthwhile improvements in accuracy and when they can be reconciled with the existing model.

EB PEAK DIRECTION RESULTS

| LINK | THRU FLOW RATE | V/C RATIO | THRU DELAY | THRU APPROACH LOS | SPEED (MPH) | ARTERIAL LINK LOS | |
|---|---|---|---|---|---|---|---|
| 1-2 | 1749 | 0.39 | 2.7 | A | 30.7 | B | |
| 2-3 | 1749 | 0.34 | 0.7 | A | 29.4 | B | |
| 3-4 | 1749 | 1.07 | 102.9 | F | 5.9 | F | *Warning |
| 4-5 | 1749 | 1.04 | 56.1 | E | 11.4 | F | *Warning |
| 5-6 | 1749 | 0.42 | 3.5 | A | 23.9 | C | |
| 6-7 | 1749 | 0.44 | 4.6 | A | 22.0 | D | |
| 7-8 | 1749 | 0.44 | 6.5 | B | 25.1 | C | |
| 8-9 | 1749 | 0.39 | 3.4 | A | 28.5 | B | |
| 9-10 | 1749 | 0.46 | 8.5 | B | 12.0 | F | |
| 10-11 | 1749 | 0.41 | 5.3 | B | 19.4 | D | |
| 11-12 | 1749 | 0.39 | 1.8 | A | 25.7 | C | |

EB Arterial Speed = *** mph
LOS = F
NOTE: Intersection Capacity Exceeded

WB OFF-PEAK DIRECTION RESULTS

| LINK | THRU FLOW RATE | V/C RATIO | THRU DELAY | THRU APPROACH LOS | SPEED (MPH) | ARTERIAL LINK LOS |
|---|---|---|---|---|---|---|
| 12-11 | 1,614 | 0.38 | 7.0 | B | 16.9 | E |
| 11-10 | 1,614 | 0.43 | 11.0 | B | 13.9 | E |
| 10-9 | 1,614 | 0.36 | 4.5 | A | 17.0 | E |
| 9-8 | 1,614 | 0.40 | 8.5 | B | 23.7 | C |
| 8-7 | 1,614 | 0.40 | 8.4 | B | 23.5 | C |
| 7-6 | 1,614 | 0.39 | 6.4 | B | 19.8 | D |
| 6-5 | 1,614 | 0.96 | 52.8 | E | 5.6 | F |
| 5-4 | 1,614 | 0.99 | 90.5 | F | 8.0 | F |
| 4-3 | 1,614 | 0.32 | 1.0 | A | 32.1 | B |
| 3-2 | 1,614 | 0.36 | 3.5 | A | 24.1 | C |
| 2-1 | 1,614 | 0.60 | 28.4 | D | 15.3 | E |

WB Arterial Speed = 14.4 mph
LOS = E

**FIGURE 5 Sample ART-PLAN output summary.**

### Incremental Delay Adjustment

It has already been pointed out that when one intersection effectively controls the arrival of vehicles at the next intersection downstream, it is not appropriate to apply the full incremental delay term to the second intersection. Consider, for example, the hypothetical case in which the second intersection is located only a few meters downstream of the first. In this case, each vehicle leaving the first intersection would arrive more or less immediately at the second, and there would be no random component. Oversaturation of the second intersection would also be impossible, because all the excess demand would be absorbed at the first intersection. In this extreme case, no incremental term should be applied.

Now, as the distance between the intersections is increased, the random element in the arrival pattern will reappear. At some point the influence of the first intersection will be eliminated and the full incremental delay will apply. For a given classification of arterial the proportion of the incremental delay that should be applied is clearly dependent on the signal spacing.

Lacking any theoretical basis to describe this effect, a very simplistic model was proposed and tested as a candidate adjustment factor. The proportion of the incremental delay term to be applied at an intersection on a coordinated arterial route was assumed to be directly proportional to the distance between signals. The full value of the incremental term was applied when the distance reached a specified threshold. The threshold values, based somewhat on judgment, were established at 0.81 km (0.5 mi) for Classification I routes and 0.4 km (0.25 mi) for Classification II routes. So the incremental term was multiplied by a factor of

$$\text{Min}\left(\frac{\text{segment length}}{\text{reference length}}, 1.0\right) \quad (1)$$

This reduced the estimated delay at signals with short spacing. The overall effect on the data collected for this study was a reduction of the overestimate of travel time from 37 to 27 percent.

### Floating Car Advantage

Another suggested source of bias was the advantage given to the travel time study vehicle as compared with the "average" vehicle because of its position in the progression band. In a properly timed arterial system, the test vehicle, or floating car, usually arrives on the green signal. This does not mean that there will be no delay to the test vehicle, because often it will be impeded by a queue of vehicles that are still waiting to be serviced. However, the ratio of uniform delay to vehicles arriving on the green as compared with all vehicles arriving during the cycle should be a good indicator of the

extent of the floating car advantage. This ratio is therefore proposed as a candidate adjustment factor.

Consider an approach to a signalized intersection. Let

$g$ = length of green phase (sec)
$C$ = cycle length (sec)
$L$ = green ratio ($g/C$)
$q$ = average arrival rate over whole cycle (veh/sec)
$s$ = average steady-state departure rate on green phase (veh/sec)
$q_g$ = average arrival rate on green phase (veh/sec)
$q_r$ = average arrival rate on red phase (veh/sec)
$P$ = proportion of arrivals on green phase
$R_p$ = platoon ratio = $P/L$
$A_g$ = total arrivals on green (veh/cycle)
$A_r$ = total arrivals on red (veh/cycle)
$X$ = degree of saturation = $q/(Ls)$
$d1$ = average uniform delay to all vehicles on approach = $0.38C(1 - L)^2/(1 - LX)$ by HCM delay equation
$d1_g$ = average uniform delay to all vehicles arriving on green phase
$F_{FC}$ = floating car advantage factor = $d1_g/d1$

Then,

$$A_g = qCP = qCLR_p \tag{2}$$

$$q_g = A_g/LC = qR_p \tag{3}$$

$$A_r = qC(1 - P) = qC(1 - LR_p) \tag{4}$$

$$q_r = qC(1 - LR_p)/C(1 - L)$$

$$= q(1 - LR_p)/(1 - L) \tag{5}$$

Referring to Figure 6, the area of the triangle ABC represents the total delay to all vehicles arriving over the entire cycle. The smaller area, A′B′C represents the total delay to vehicles arriving on the green phase only, as a subset of the total delay.

To obtain average delay values, the total delay values given by the areas of the respective triangles must be divided by the number of vehicles represented per cycle, that is,

$$d1 = \text{Area ABC}/qC \tag{6}$$

and

$$d1_g = \text{Area A′B′C}/qCLR_p \tag{7}$$

Since the concern is the ratio of $d1{:}d1_g$, the factor of 0.38 will be dropped from the HCM equation. This factor is used to convert total delay to stopped delay and will be the same for both $d1$ and $d1_g$.

Referring again to Figure 6, the queue at the end of the red phase (QEOR) may be computed as

$$\text{QEOR} = qR_pC(1 - L) \tag{8}$$

$$\text{QEOR} = qrC(1 - L) = \frac{qC(1 - LR_p)(1 - L)}{(1 - L)}$$

$$= qC(1 - LR_p) \text{ veh} \tag{9}$$

The time required to clear the arrivals on red (GQR) will be

$$\text{GQR} = \frac{\text{QEOR}}{S} = \frac{qC(1 - LR_p)}{\frac{q}{LX}} = CLX(1 - LR_p) \text{ sec} \tag{10}$$

Now, during the period GQR, the arrivals on green will accumulate at a rate of $q_g$ veh/sec. The time GQA required to clear all of the vehicles, including those that arrived on the red and the queued portion of the green, will be

$$\text{GQA} = \frac{\text{QEOR}}{s - q_g} = \frac{qC(1 - LR_p)}{\frac{q}{LX - qR_p}} = \frac{qC(1 - LR_p)}{\frac{q(1 - LR_pX)}{LX}}$$

$$= \frac{CLX(1 - LR_p)}{1 - LR_pX} \text{ sec} \tag{11}$$

Now the area of the triangle A′B′C may be computed as

$$0.5\, q_g\, \text{GQR GQA} = \frac{0.5\, q_g\, CLX(1 - LR_p) \cdot CLX(1 - LR_p)}{1 - LR_pX}$$

$$\text{veh-sec/cycle} \tag{12}$$

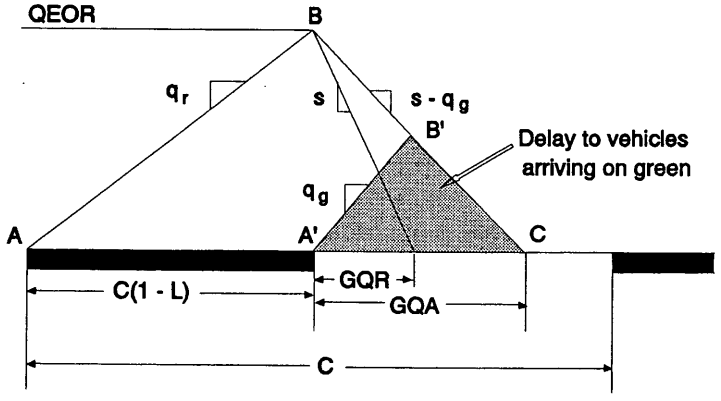



**FIGURE 6 Queue accumulation polygon for floating car adjustment factor.**

To determine a unit delay (sec/veh) the number of vehicles per cycle arriving on the green must be computed by $CLq_g$, and the uniform delay per vehicle arriving on the green becomes

$$d1_g = \frac{0.5q_g\cancel{CL}X(1 - LR_p)\text{CLX}(1 - LR_p)}{\cancel{CL}q_g(1 - LR_pX)}$$

$$d1_g = \frac{0.5CLX^2(1 - LR_p)^2}{1 - LR_pX} \quad (13)$$

The floating car advantage factor, $F_{FC}$, may be computed as

$$\frac{d1_g}{d1} = \frac{\dfrac{0.5CLX^2(1 - LR_p)^2}{1 - LR_pX}}{\dfrac{0.5C(1 - L)^2}{1 - LX} \cdot \dfrac{(1 - LR_p)}{(1 - L)}}$$

$$= \frac{LX^2(1 - LR_p)(1 - LX)}{(1 - LR_pX)(1 - L)} \quad (14)$$

for the special case of random arrivals (i.e., $R_p = 1$), this equation simplifies to

$$\frac{d1_g}{d1} = \frac{LX^2(1 - L)(1 - LX)}{(1 - LX)(1 - L)} = LX^2 \quad (15)$$

It should be appropriate to apply this factor to the uniform delay as long as there is some discernable progression. By definition, there is no discernable progression with Arrival Type 1. So, as a matter of judgment, no floating car adjustment was applied to those cases in the study for which Type 1 was indicated. The adjustment was applied for all other arrival types, which reduced the discrepancy between the estimated and measured travel times to a negligible level.

## STUDY RESULTS

To summarize the preceding discussion,

1. The measured and estimated (HCM Table 11-4) *running times* agreed surprisingly well without further adjustment. Virtually all of the discrepancy in *travel times* was in the intersection delay values.
2. The imposition of an upper limit of 1.2 on the v/c ratio was necessary as a first-level filter to eliminate gross discrepancies between the measured and estimated delays. This created a base condition with a 37 percent overestimation of travel time.
3. The application of the incremental delay reduction factor for closely spaced intersections reduced the travel time overestimate to 27 percent.
4. The additional application of the floating car adjustment factor to the uniform delay term effectively eliminated the discrepancy between measured and estimated travel times.

These results are presented graphically in Figure 7 and Table 3. Figure 7 also shows the degree of overestimation of the intersection delay in addition to the travel times. A breakdown of the estimation error for the fully adjusted results is also provided by FDOT district (Districts 4 and 6) and by HCM arterial classification (Classes I and II). Both of these breakdowns indicate minimal errors for any of the categories. This lends additional credibility to the validity of the results.

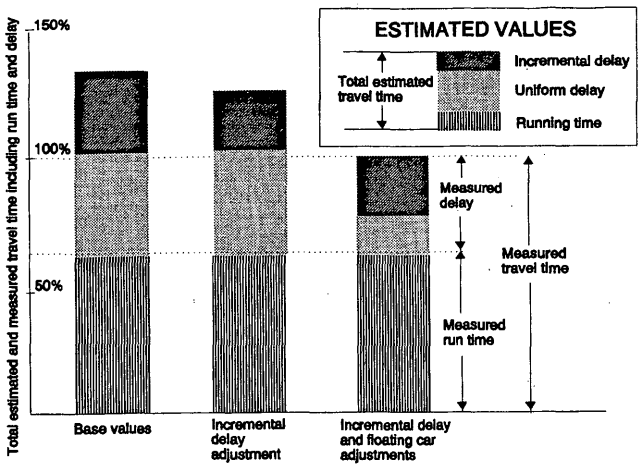



**FIGURE 7 Comparison of measured and estimated delays.**

## CONCLUSIONS AND RECOMMENDATIONS

The results offer a reasonable explanation of the apparent discrepancies between estimated and measured travel times and delays on arterial streets. They also provide a practical means of adjusting the estimated delay values to produce a very close agreement with the corresponding measurements.

These results will be most useful for planning level analyses. The adjustment factors appear to eliminate the bias from the travel time and delay estimation models, but there is a substantial error and variability in the comparison of several of the individual runs. More accurate values would be required for the field data to be used in the computational models if the results were intended for operational analysis purposes. The findings of this study should be implemented as follows:

1. The travel time data collection program should be modified to compute the running speed in a manner compatible with the HCM. The stopped delay should be multiplied by the HCM factor of 1.3 before being subtracted from the total travel time.
2. An incremental delay reduction factor should be considered for the HCM Chapter 9 and 11 methodology. This modification should not, however, be used to justify the operation of an arterial route beyond its capacity.

**TABLE 3 Summary of Results**

| | Error (%) | |
|---|---|---|
| | Total Travel Time | Intersection Delay |
| Base data | 37/O | 101/O |
| Incremental delay adjustment | 28/O | 77/O |
| Uniform and incremental delay adjustments | 0 | 1/O |
| District 4 | 4.1/O | — |
| District 6 | 1.4/U | — |
| Classification I | 2.2/O | — |
| Classification II | 3.2/U | — |

NOTE: O = overestimate, U = underestimate.

3. The uniform delay reduction factor should be incorporated as a supplementary ART_PLAN output, labeled specifically as the estimated speed or travel time for a floating car study.

The data should be analyzed further to explore alternative models that could reduce the variability of the estimates and produce better agreement between estimated and measured travel time and delay for individual runs.

## REFERENCES


1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
2. *Florida Highway System Plan: Level of Service Manual.* Topic:525-000-005-c. Florida Department of Transportation, Tallahassee, April 1992.
3. Ewing, R., and A. Khoury. Predicted vs. Actual Levels of Service: Evaluation of Three Arterial Analysis Procedures. In *Compendium of Technical Papers*, Institute of Transportation Engineers, Washington, D.C., 1991, pp. 229–223.
4. Ewing, R. A Fresh Look at Roadway Level-of-Service Issues. In *Transportation Research, Record 1364*. TRB, National Research Council, Washington, D.C., 1992.
5. Wallace, C. E., K. G. Courage, and E. C. P. Chang. Methodology for Optimizing Signal Timing—M|O|S|T Reference Manual. FHWA, U.S. Department of Transportation, Dec. 1991.
6. University of Florida Transportation Research Center. A System for Evaluating Urban Congestion with Moving Vehicle Studies. Florida Department of Transportation, Governor's Energy Office, Tallahassee, Feb. 1990.
7. Prassas, E. V., W. R. McShane, and D. S. McLeod. *ART_PLAN Program Documentation.* Polytechnic University, Brooklyn, N.Y., 1993.
8. Barton-Aschman Associates, Inc. *District-Wide Capacity Analysis Study.* Florida Department of Transportation, Tallahassee, Aug. 1993.


*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Utilization of Auxiliary Through Lanes at Signalized Intersections

JAMIE W. HURLEY

The capacity of signalized intersections is sometimes increased by adding an auxiliary lane for use by through traffic. The effectiveness of an auxiliary lane depends on the amount of traffic using it. Equal distribution of traffic between a continuous and an auxiliary through lane would result in the greatest total capacity of this lane pair, but traffic, land use, and geometric factors are usually such that this does not occur. The 1985 *Highway Capacity Manual* does not address this situation. A concept of captive and choice lane users was used in modeling auxiliary lane use for intersection configurations with a single continuous through lane and an auxiliary lane beginning upstream of the intersection and extending downstream of it. Stepwise multiple regression was performed on data collected at sites in Tennessee to determine, from a candidate list of factors, those that significantly affect choice use of the auxiliary lane. These factors were found to be (*a*) through flow rate, (*b*) right turns off of the facility in the last 500 ft of the auxiliary lane, (*c*) downstream auxiliary lane length, and (*d*) urban area size. For the sites studied, it was found that traffic distribution between lanes for intersection configurations with a single continuous through and an auxiliary lane is much different from the value given in the *Highway Capacity Manual* for two continuous through lanes.

A common practice for increasing the capacity of signalized intersections is to use exclusive turning lanes. Although less common, signalized intersection capacity may also be increased by adding an auxiliary lane for use by through traffic. A lane configuration typically used in urban areas is illustrated in Figure 1. The effectiveness of the continuous and auxiliary through lanes depends on the amount of traffic using the auxiliary lane. Equal lane distribution between these lanes, if it could be achieved, would result in the greatest total capacity of this lane pair.

Intersection configurations such as that shown in Figure 1 are not addressed in the 1985 *Highway Capacity Manual* (HCM) (*1*). The HCM procedures do, however, treat configurations with two continuous through lanes, which could be thought of as the Figure 1 configuration with infinite upstream and downstream auxiliary lane lengths.

This paper defines, on the basis of sample site data, the factors that significantly affect lane distribution for the continuous and auxiliary through lanes as shown in the Figure 1 intersection configuration. It also presents, using these same data, a means of estimating the corresponding volume by lane.

## BACKGROUND

The 1985 HCM uses lane utilization factors to account for the distribution of traffic across multiple lanes continuing through an intersection. These factors are based on the assumption that the most heavily traveled lane in a group of two serves 52.5 percent of the total flow. Although the basis for these factors is not mentioned in the 1985 HCM, the factors themselves are the same as those in *Transportation Research Circular 212* (*2*). It is stated (*2*) that the 52.5 percent value is a compromise between the 55/45 percent volume split assumed by Messer and Fambro (*3*) and an assumption of equal (50/50 percent) lane distribution during peak conditions.

The first known capacity-related effort for the study intersection configuration is that of Leisch (*4*). In this work, Leisch developed a nomographic procedure for solving intersection capacity problems that was based on the 1965 HCM, similar to that which he developed and based on the 1950 HCM. In addition to these basic intersection capacity nomographs, Leisch included procedures for certain "special conditions" not covered in the 1965 HCM, one of which is the Figure 1 configuration. Unfortunately, Leisch did not discuss the methods used to develop these procedures.

McCoy and Tobin (*5*) observed the use of auxiliary through lanes, evaluated the effect of the length of these lanes on their use by through vehicles, and incorporated their findings into the critical movement analysis technique of *Transportation Research Circular 212* (*2*). As part of their work, however, McCoy and Tobin also studied the work of Leisch (*4*). Using stepwise multiple regression analysis of the data, McCoy and Tobin developed a linear model for estimating the mean number of through vehicles discharging from the additional through lane as a function of green time for the through and right movements and the total length of the lane addition (upstream plus downstream). They noted that the number of right-turning vehicles did not significantly affect the usage of the auxiliary through lane by through vehicles. Finally, McCoy and

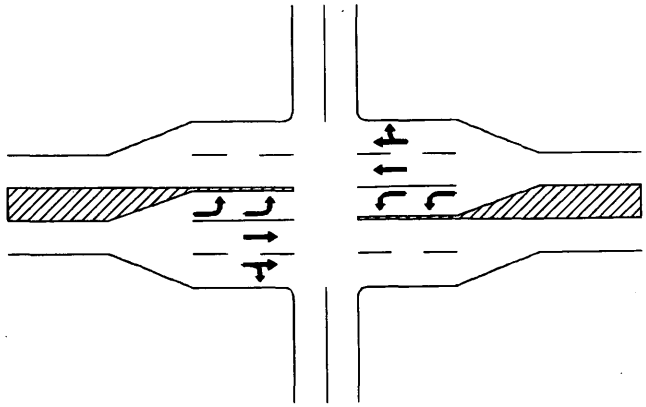


**FIGURE 1 Adding auxiliary lane to increase capacity.**

Department of Civil Engineering, University of Memphis, Memphis, Tenn. 38152.

Tobin also concluded that length requirements for the additional lane as determined using the Leisch guidelines are too short to achieve an average use by through vehicles of more than 1.5 passenger cars per cycle.

## STUDY APPROACH

As stated previously, the effectiveness of a through auxiliary lane depends on the number of drivers using it. Lieberman (*6*), in developing a lateral deployment model for the TRAFLO macroscopic traffic simulation program, employed a variation of Wardrop's first principle: that every motorist will select a lane on an approach consistent with his or her intended turn maneuver and with any specified lane channelization so as to minimize his or her perceived travel time. In reviewing research efforts related to the factors that most influence intersection capacity, Stokes (*7*) concluded that although there is general agreement on the effects of certain physical factors on capacity, a class of factors characterized as "driver behavior" does not lend itself to quantification in any systematic fashion. Stokes believes that the effects of these factors are site-specific.

It is suggested here that the concepts of Lieberman and Stokes are substantially correct with regard to the study intersection configuration, and probably with regard to any other intersection situation involving lane choice. This is best explained in terms of "captive" and "choice" users of the auxiliary through lane. Captive users are those through movement drivers who *must* use the auxiliary lane because of their need to turn right downstream of the intersection (into driveways). Some auxiliary lane users are captive because they turn into the auxiliary lane from the right side of the roadway immediately upstream of the intersection. There are captive users of the continuous through lane as well. These are those turning left into driveways downstream of the intersection and those turning left onto the facility from driveways immediately upstream of the intersection. The primary characteristic of captive users of either the continuous or the auxiliary lane is that they use those lanes because of their association with adjacent land use. Traffic demand, signal timing, and (in a sense) auxiliary lane length, for example, have nothing to do with their use of continuous lanes. There is obviously some zone of influence upstream and downstream of the intersection beyond which drivers entering and exiting the facility, should the auxiliary lane be long enough, have enough space to change lanes. It is assumed here that captive drivers are those who, within the length of the auxiliary lane, exit driveway downstream of the intersection.

Choice users of an auxiliary lane are drivers who travel continuously through the intersection and who have decided to use the auxiliary lane. This choice may or may not be based on perceived travel time alone. Turning movements onto and off of the left side of the roadway will cause interference with traffic in the leftmost continuous lane, thereby making the auxiliary lane a more attractive choice. Conversely, turning movements onto and off of the right side of the roadway tend to inhibit use of the auxiliary lane by drivers having a choice of lanes. Proper analysis of these inhibiting effects required that a determination be made as to *which* turning movements most affect lane choice: those near the intersection, those near the downstream end of the auxiliary lane, or the total over the entire downstream length.

It was considered that other factors in addition to turning movements could affect auxiliary lane use by choice drivers. These include geometrics, urban area size, demand magnitude, and the number of heavy vehicles in the traffic stream. The total portion of through traffic in the auxiliary lane may be expressed as

$$P_{total} = \frac{P(Q - RTOFF - LTOFF) + RTOFF}{Q}$$

where

$P_{total}$ = total portion of through traffic in auxiliary lane,
$P$ = portion of choice users in auxiliary lane,
$Q$ = total through flow rate (vph),
RTOFF = right turns off of facility downstream of intersection (per hour), and
LTOFF = left turns off of facility downstream of intersection (per hour).

For the study intersection configuration, the lane utilization factor, $U$, used in the HCM will be the larger value of $2P_{total}$ or $2(1 - P_{total})$.

## MODEL DEVELOPMENT

Since no theory existed by which lane distribution could be predicted, a model was developed using stepwise multiple regression analysis. The basic model form for regression analysis is of the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_i + \ldots + \beta_k X_k$$

where

$Y$ = value of dependent variable (in this case, some measure of auxiliary lane use),
$X_1, \ldots, X_k$ = values of independent variables (i.e., for factors that affect lane distribution), and
$\beta_0, \ldots, \beta_k$ = numerical coefficients (determined in regression process).

A problem with this model structure is that it is linear—that is, if one of the $X$'s is flow rate, the portion in the auxiliary lane can (theoretically, at least) increase to infinity. This, of course, is not realistic. It would be more realistic to expect voluntary lane use to increase as demand increases, but up to a maximum limit. For example, under very low demand, one would expect the portion of through traffic in the auxiliary lane to be very small, since nothing could be gained by using it. With increasing demand, however, one would expect the portion of choice users in the auxiliary lane to increase.

A curve shape that better represents lane choice behavior is the hyperbolic tangent function given here and illustrated in Figure 2.

$$\tanh X = \frac{e^X - e^{-X}}{e^X + e^{-X}}$$

The basic hyperbolic tangent function has a range from $-1$ to $+1$ and passes through the origin. However, the function can be shifted both horizontally and vertically by replacing $X$ with a mathematical function. It was desired to model $P$, the choice portion of drivers using the auxiliary lane such that $P$ ranged from 0 (at the origin) up to the "ideal" or maximum value of choice auxiliary lane use. (The

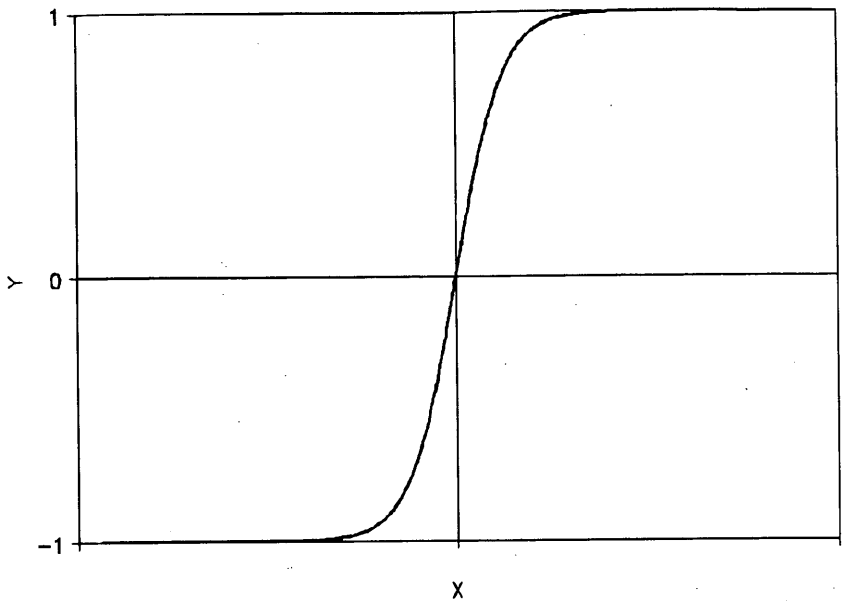



**FIGURE 2 Hyperbolic tangent function.**

ideal value is based on field data collected as part of the study.) This variation is illustrated in Figure 3 for an ideal choice auxiliary lane use of 50 percent. To force the curve to behave in this manner, the $X$ in the hyperbolic tangent function is replaced by

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

Mathematically then, the model for the choice users of the auxiliary lane is

$$P = P_{\text{ideal}} \tanh(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)$$

Another positive quality of the hyperbolic tangent function is that it is inverted easily. In general terms, the inverse is written in the following form:

$$\tanh^{-1} X = \frac{1}{2} \ln\left(\frac{1 + X}{1 - X}\right)$$

It is convenient here to define a dependent variable, $Y$, as

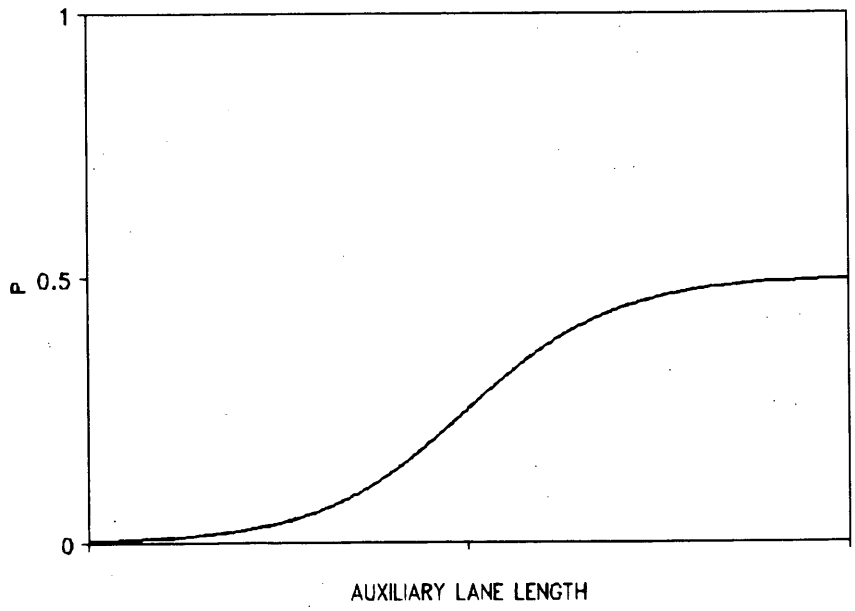



**FIGURE 3 Expected auxiliary lane usage.**

$$Y = \frac{1}{2} \ln\left(\frac{P_{ideal} + P}{P_{ideal} - P}\right)$$

Describing the dependent variable, $Y$, in this manner makes it possible to write the following:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_i + \ldots + \beta_k X_k$$

This model form made it possible to perform stepwise multiple regression on data collected in the field beginning with a number of "candidate" independent variables that may or may not influence auxiliary lane choice and ending with a model containing only those variables that do (statistically) influence it. The candidate independent variables considered in the analysis are:

- Peak 15-min flow rate,
- Area size (0 = large cities, 1 = small towns),
- Product of peak 15-min flow rate and area size,
- Percentage trucks,
- Upstream auxiliary lane length,
- Downstream auxiliary lane length,
- Total auxiliary lane length,
- Right turns on at intersection,
- Right turns off at intersection,
- Left turns off at intersection,
- Total right turns on downstream of intersection,
- Total right turns off downstream of intersection,
- Total left turns on downstream of intersection,
- Total left turns off downstream of intersection,
- Right turns on in first 107 m (350 ft) downstream of intersection,
- Right turns off in first 107 m (350 ft) downstream of intersection,
- Left turns on in first 107 m (350 ft) downstream of intersection,
- Left turns off in first 107 m (350 ft) downstream of intersection,
- Right turns on in first 152 m (500 ft) downstream of intersection,
- Right turns off in first 152 m (500 ft) downstream of intersection,
- Left turns on in first 152 m (500 ft) downstream of intersection,
- Left turns off in first 152 m (500 ft) downstream of intersection,
- Right turns on in last 107 m (350 ft) downstream of intersection,
- Right turns off in last 107 m (350 ft) downstream of intersection,
- Left turns on in first 107 m (350 ft) downstream of intersection,
- Left turns off in last 107 m (350 ft) downstream of intersection,
- Right turns on in last 152 m (500 ft) downstream of intersection,
- Right turns off in first 152 m (500 ft) downstream of intersection,
- Left turns on in last 152 m (500 ft) downstream of intersection,
- Left turns off in last 152 m (500 ft) downstream of intersection,
- Right turns on in first 122 m (400 ft) upstream of intersection,
- Right turns off in first 122 m (400 ft) upstream of intersection,
- Left turns on in first 122 m (400 ft) upstream of intersection,
- Left turns off in first 122 m (400 ft) upstream of intersection,
- Right turns on in first 91 m (300 ft) upstream of intersection,
- Right turns off in first 91 m (300 ft) upstream of intersection,
- Left turns on in first 91 m (300 ft) upstream of intersection,
- Left turns off in first 91 m (300 ft) upstream of intersection,
- Total right turns on upstream of intersection,
- Total right turns off upstream of intersection,
- Total left turns on upstream of intersection, and
- Total left turns off upstream of intersection.

## DATA COLLECTION

To determine and evaluate the model coefficients, lane use and upstream and downstream turning movement studies were conducted at a number of sites.

### Study Sites

It was desired that study data be obtained from five sites. To assess the effect, if any, of urban area size, two sites were to be in small towns. Five sites in western and middle Tennessee were selected for analysis of the study intersection configuration:

1. Southbound US-45 at US-64 (Selmer),
2. Southbound US-43 at SR-50 (Columbia),
3. Westbound Quince at Kirby (Memphis),
4. Eastbound Quince at Kirby (Memphis), and
5. Northbound Kirby at Quince (Memphis).

The data collected at Site 5 were not used in the analysis, primarily because several equipment failures were encountered while attempting to collect data at that site. During this period, mud tracked by vehicles from a nearby upstream construction site obscured the upstream pavement markings to the extent that the beginning of the upstream auxiliary lane could not be seen. It was thought that, since drivers could not ascertain the beginning of the auxiliary lane, measurements of upstream section length (taken before markings were obscured) were meaningless.

### Data Collection Techniques

The data required for intersection analysis are of two types: (*a*) an inventory of site geometric and land use data and (*b*) traffic data collected in the field. The inventory data, obtained from drawings or physical measurements (or both) at the sites included

- Length of auxiliary lane on approach to intersection (neglecting taper),
- Length of auxiliary lane downstream of intersection (measured from stop bar to beginning of taper),
- Existence of parking in vicinity of intersection,
- Location of driveways and identification of land use adjacent to the facility, and
- Presence of local bus stops in the intersection area.

Traffic data collection involved volumes, turning movements both at and in the vicinity of the intersection, vehicle mix, and, of course, lane distribution. Data were collected for at least 2 hr at each intersection. At locations with short peak periods, it was necessary to revisit the site to obtain more data under high-volume conditions.

The primary tool for data collection was the video camera with a character generator, an option that displays lapsed time to 1/10 sec. The videotape provides a permanent record of the basic data and contains heavy-vehicle data in addition to basic lane distribution data. With proper camera positioning, intersection turning movements may be recorded as well.

Where applicable, data collection personnel were stationed upstream and downstream of the intersection to record turning movements from and into the traffic stream. Where possible, these data were recorded by hand on data forms. Where there were a relatively large number of driveways or a high rate of driveway activity, the data were recorded verbally onto microcassette tapes. These tapes were also used to record any unusual activity such as cycle failures. Using stopwatches synchronized with the video camera character generator, the end of each 15-min period of the study was announced so that all data collected would be consistent by time.

## DATA ANALYSIS

To be consistent with the 1985 HCM procedures, the data collected were analyzed in 15-min segments so as to represent peak 15-min flow rates. For a given site, then, each 2-hr period represented eight data points.

A preliminary screening analysis to assess the candidate independent variables was undertaken before the stepwise regression analyses were performed. This not only sheds light as to which variables might not be of statistical importance, but provides information as to candidate variable interaction. In Figure 4, for example, the variation in choice use of the auxiliary lane is plotted with flow rate. Examination of the data might lead to the conclusion that the data do not appear to follow any particular form. However, when the data are taken in conjunction with their origin, two patterns emerge. One is data on the left side of the figure that come from smaller towns and the other is the data on the right that come from large towns. There are actually two patterns, then, on the same graph, which leads to the suspicion that the product of urban area size (a value of 0 for large cities and 1 for small ones) and flow rate might, when treated as a single variable, significantly affect choice of the auxiliary lane.

### Prediction Model

After preliminary screening and the stepwise regression process, the following model was obtained:

$$P = 0.22 + 0.22 \tanh (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$$

where $\beta_0 = -2.36191$, $\beta_1 = 1.391615$, $X_1$ = (peak 15-min flow rate)/1,000, $\beta_2 = 3.587243$, $X_2$ = (peak 15-min flow rate × area size)/1,000, $\beta_3 = -0.94191$, $X_3$ = [right turns off in last 152 m (500 ft) downstream]/100, $\beta_4 = 2.975039$, and $X_4$ = (downstream length in meters)/1,000.

The 0.22 terms appearing in the model are based on a value of $P_{ideal} = 0.44$. This value for $P_{ideal}$ was developed from the model for a value of zero right turns off in the last 500 ft downstream, a small town, and the maximum 15-min flow rate measured at any of the intersections studied. It is emphasized that $P_{ideal}$ applies only to *choice* users of the auxiliary lane.

It is desirable that the *t*-values used in the stepwise regression process be at least 2.0 in magnitude. The interpretation of this is that one would be at least 95 percent confident that that particular coefficient ($\beta$) is not 0. The *t*-value for $\beta_4$ was only 1.62. However, since that was the only variable representing either upstream or downstream section length, and since one would still be almost 90 percent confident that $\beta_4$ is not 0, it was included in the final model. $R^2$, the coefficient of multiple determination for this model is 0.80.

Because the model has four independent variables, the degree to which the model follows the data is difficult to understand when plotted in two dimensions. However, by examining Figure 5, the manner in which the model tries to duplicate each data point may be observed to some extent. The means absolute error for choice users of the auxiliary lane was found to be approximately 2.8 percent. It is believed that the model performs reasonably well.

It should be pointed out that the upstream auxiliary lane length did not appear in the equation for choice use of the auxiliary lane.

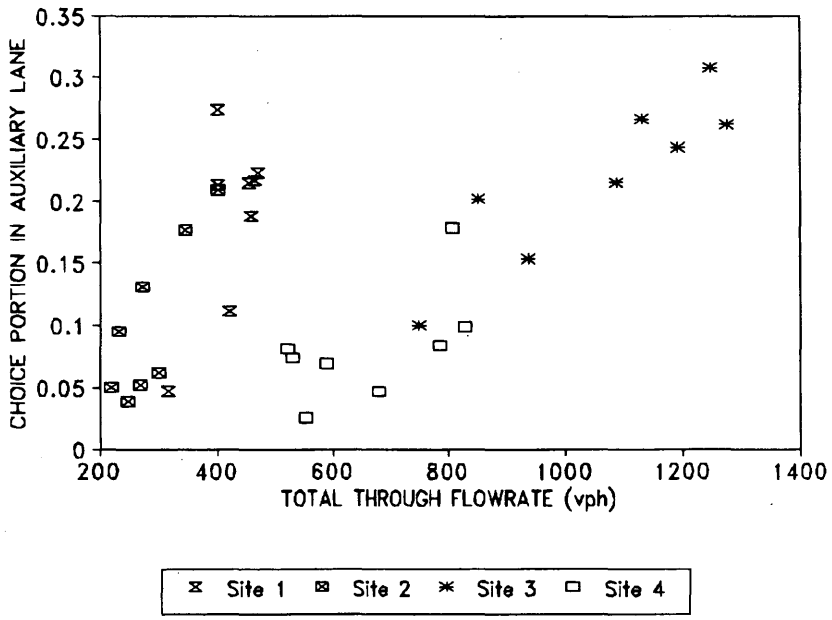



**FIGURE 4 Variation of site data with flow rate.**

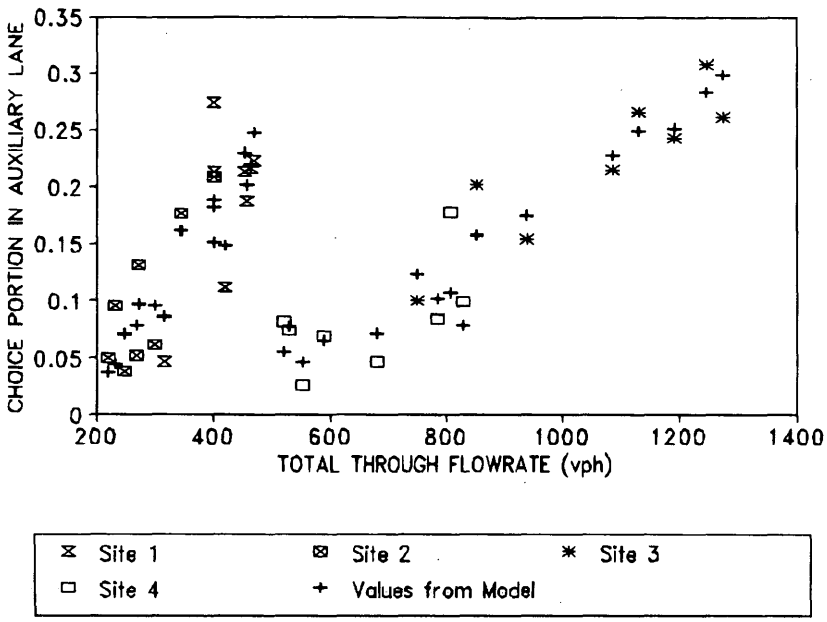



**FIGURE 5 Model and site data.**

This does not mean that upstream lane length plays no role in lane choice. It is probable that the upstream section lengths at the study sites were long enough that the effects of upstream length could not be ascertained.

### Observation

During the data collection portion of this study, the total observed auxiliary lane use varied from 14 to 32 percent of the through traffic movement, which means that 68 to 86 percent of the through traffic (for the data collected) was in the more heavily traveled lane. This value differs markedly from the 52.5 percent used in the HCM for two continuous through lanes. It should not be assumed that the maximum auxiliary lane usage that can be obtained in practice is 32 percent, for it is not. There are a variety of reasons for the values measured, including auxiliary lane lengths too short to achieve greater use.

## CONCLUSIONS

A new approach has been used for determining lane utilization—that of choice and captive auxiliary lane users. This concept could be used for intersections without auxiliary lanes as well. The potential increase in accuracy of intersection capacity analyses for these cases is not yet known, although this paper has shown that field measurements involving auxiliary lanes differ greatly from the HCM value for two continuous lanes. Although the lane distribution model developed is based on the behavior of Tennessee drivers only, the approach could be used for developing a similar model on a larger geographical scale—perhaps nationally. The model could also be used in the design process of these intersection configurations by including the effect of downstream auxiliary lane length on intersection operation.

The greatest difficulty in using the model presented is the estimation of driveway turning movements downstream of the intersection. Actual driveway count data at the site being evaluated are obviously superior to anything else, but it is possible that other data sources requiring less effort could be adequate. The ITE *Trip Generation* manual (*8*) is a possible source of data, although no attempt has been made to assess the adequacy of its data for this purpose. Another approach would be to make a reasonable assumption of driveway movements and then assess the sensitivity of the operation on the basis of these estimates with plus and minus deviations from these estimates. To provide some insight as to the magnitude of these turning movements, the turning movement data measured during this study are presented in Table 1. Should an attempt be made to use these data in some manner, it is particularly important to consider the time of day during which the data were collected and the location. For example, trips to and from shopping centers are much greater in the p.m. peak period than in the a.m. peak.

The need for some additional research has already been indicated: a model based on "national" driver behavior, and an assessment of the captive/choice approach for intersections without auxiliary lanes. A lane utilization model is also needed for intersection configurations involving two continuous through lanes and a through auxiliary lane. Finally, if the approach taken herein is of value, some attempt should be made to develop default driveway turning movement data so that excessive effort is not required to use this type of model.

## ACKNOWLEDGMENTS

The research on which this paper is based was sponsored by the Tennessee Department of Transportation and FHWA. The author would like to express his appreciation to Don Dahlinger of the Ten-

TABLE 1 Measured Downstream Turns

| TOTAL THROUGH FLOWRATE (vph) | RIGHT TURNS OFF (vph) | | LEFT TURNS OFF (vph) | |
|---|---|---|---|---|
| E.B. Quince at Kirby (Memphis) - **P.M. Peak** | | | | |
| | Service Station/Convenience Store | Small Shopping Center | Small Office Center (Side Entrance) | Townhouse Driveway |
| 528 | 8 | 28 | 0 | 4 |
| 520 | 24 | 48 | 4 | 0 |
| 552 | 16 | 64 | 0 | 4 |
| 588 | 16 | 48 | 0 | 4 |
| 680 | 20 | 56 | 0 | 0 |
| 784 | 16 | 48 | 0 | 4 |
| 808 | 16 | 48 | 0 | 4 |
| 828 | 12 | 72 | 0 | 16 |
| W.B. Quince at Kirby (Memphis) - **A.M. Peak** | | | | |
| | Service Station/Convenience Store | Small Shopping Center | Multi-Story Office Building | |
| 1088 | 56 | 12 | 36 | |
| 1132 | 56 | 16 | 40 | |
| 1192 | 32 | 24 | 100 | |
| 748 | 56 | 44 | 88 | |
| 1764 | 40 | 20 | 56 | |
| 1700 | 28 | 28 | 128 | |
| 1652 | 56 | 40 | 88 | |
| 1360 | 48 | 24 | 48 | |
| S.B. U.S. 43 at SR 50 (Columbia) - **Mid Day** | | | | |
| | Fast Food Restaurant/ Small Shopping Center Driveway | | Vision Center | Bank |
| 476 | 32 | | 0 | 4 |
| 436 | 48 | | 8 | 8 |
| 356 | 28 | | 12 | 0 |
| 512 | 52 | | 8 | 0 |
| 372 | 36 | | 4 | 0 |
| 516 | 20 | | 8 | 0 |
| 492 | 20 | | 4 | 4 |
| 524 | 36 | | 0 | 0 |
| S.B. U.S. 45 at U.S. 64 (Selmer) - **P.M. Peak** | | | | |
| | Service Station/Convenience Store | Fast Food Restaurant | Service Station/Convenience Store | Hardware Store |
| 464 | 36 | 16 | 12 | 12 |
| 400 | 36 | 16 | 32 | 16 |
| 452 | 24 | 16 | 12 | 8 |
| 420 | 48 | 16 | 20 | 12 |
| 468 | 32 | 8 | 36 | 16 |
| 400 | 28 | 8 | 20 | 8 |
| 456 | 40 | 16 | 12 | 4 |
| 316 | 28 | 20 | 4 | 8 |

nessee Department of Transportation for his support throughout the conduct of the study.

## REFERENCES

1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
2. *Transportation Research Circular 212: Interim Materials on Highway Capacity.* TRB, National Research Council, Washington, D.C., Jan. 1980.
3. Messer, C. J., and D. B. Fambro. A New Critical Lane Analysis for Intersection Design. In *Transportation Research Record 644*, TRB, National Research Council, Washington, D.C., 1977, pp. 26–33.
4. Leisch, J. E. Capacity Analysis Techniques for Design of Signalized Intersections, Installment 1. *Public Roads*, Vol. 34, No. 9, Aug. 1967, pp. 171–209.
5. McCoy, P. T., and J. R. Tobin. Use of Additional Through Lanes at Signalized Intersections. In *Transportation Research Record 869*, TRB, National Research Council, Washington, D.C., 1982, pp. 1–5.
6. Lieberman, E. B. Determining the Lateral Deployment of Traffic on an Approach to an Intersection. In *Transportation Research Record 772*, TRB, National Research Council, Washington, D.C., 1980, pp. 1–5.
7. Stokes, R. W. Some Factors Affecting Signalized Intersection Capacity. *ITE Journal*, Jan. 1989, pp. 35–40.
8. *Trip Generation*, 5th ed. Institute of Transportation Engineers, Washington, D.C., 1991.

---

*The opinions presented herein are those of the author and are not necessarily those of either of the sponsoring agencies.*

*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Approximation of Percentage Time Delay with Local Measurements

MATTI PURSULA

In the 1985 *Highway Capacity Manual* (HCM), percentage time delay (PTD) is used as the main indicator of the level of service for two-lane, two-way highways. HCM defines PTD as the average percentage of time that all vehicles are delayed while traveling in platoons due to the inability to pass. For field measurement purposes, the HCM states that PTD is approximately the same as the percentage of vehicles traveling in platoons at headways less than 5 sec. The relationship between local platoon percentage, measured with a fixed 5-sec headway as the platooning criterion, and the previous estimate of PTD is analyzed. Theoretical considerations indicate that local platoon percentage is a biased estimate of PTD and must be corrected with a factor calculated as the ratio of space mean speed of all traffic and that of the platooned vehicles. If several measurements are made along a road, PTD can be estimated as a weighted average of local values using mean travel times and traffic flows as weights. Analysis of simulation and field data indicate that the local platoon percentage is usually about 3 to 5 percent lower than the corresponding PTD value. To make local measurements correspond more precisely to the basic definition of PTD, one could exclude from measurements situations in which a slower vehicle is behind a faster one inside the 5-sec headway. The effect of this exclusion, however, is not tested.

In the 1985 *Highway Capacity Manual* (HCM) (*1*), percentage time delay (PTD) is used as the main indicator of the level of service for two-lane, two-way highways. According to the HCM, PTD is defined as the average percentage of time that all vehicles are delayed while traveling in platoons because of the inability to pass.

In previous works, the author has used the percentage of vehicles driven with a headway of fewer than 5 sec, when measured at a fixed point, as the approximation of PTD (*2–4*). The aim of this paper is to analyze the relationships between PTD and this locally measured platoon percentage. The analysis is based on theoretical considerations, simulations, and real measurements.

## THEORETICAL RELATIONSHIPS

### Basic Considerations

The 1985 HCM (*1*) and the background report (*5*) state that in estimating PTD, motorists are defined to be delayed when traveling behind a platoon leader at speeds less than their desired speeds and at headways of less than 5 sec. The reports also state that for field measurement purposes, PTD on a road section is approximately the same as the percentage of all vehicles traveling in platoons at headways of less than 5 sec.

Helsinki University of Technology, Transportation Engineering, Rakentajanaukio 4A, FIN-02150 Espoo, Finland.

The 5-sec criterion is commonly used as the limit value for time headway between free and platooned vehicles. As can be seen in Figure 1, field measurements indicate that when a faster vehicle is reaching a slower one, the adaptation of speed—and thus the hindrance—begins at headways between 5 to 10 sec (*3*). The same result is given by several researchers (*6,7*). The adaptation is, of course, dependent on the speed difference of the vehicles and the driver in question. So, a wide range of platooning criteria can be justified, among them the 5-sec criterion used in the HCM.

PTD is defined as the average share of time spent in platoons. From a road section, the number of vehicles in platoons is a dynamic variable that changes in value all the time. Thus, PTD can be calculated by averaging instantaneous road section values over time.

From the point of view of an individual driver, PTD is a variable related to a certain section of road. The percentage of time driven in platoons of the total travel time of the road section is the platooning experienced by the driver. Calculating PTD from individual total travel time values means averaging over drivers.

In the following, the PTD of the total traffic flow is expressed as the sum of the time that drivers are in platoons divided by the sum of the total travel time of all drivers on the same road section during the period in question.

The aim of the paper is to analyze the relationship of this PTD to the locally measured platoon percentage based on the 5-sec headway criterion. For simplicity, even though the word "percentage" is used in the text, PTD and platoon percentage in all equations are expressed as ratios, without multiplication to percentage values.

### Simple Analysis with Shock Wave theory

First analyze a stable traffic flow given in Figure 2. The flow is divided into two densities, that is, to platoons (density $k_1$) and to free vehicles (density $k_2$). The speeds ($u_1$ and $u_2$) and the lengths ($l_1$ and $l_2$) of the platoons are given. In a stable situation, as many vehicles per unit time move from the faster traffic flow to the slower (and more dense) one as move from the slower traffic to the faster. This way, the lengths of the platoons do not change when traffic propagates along the road. The points of density change travel along the road with constant and equal speeds. This can be seen very easily from the equation of the shock wave speed ($c$ in Figure 2), which gives the same value for both edges of density change.

Thus, using the symbols in Figure 2, one can estimate $r$, the PTD on the road section with one cycle of high- and low-density traffic. The estimation (Equation 1) is based on the fact that at any moment the number of vehicles on the road section, as well as the number of vehicles in platoons and outside platoons, is constant.

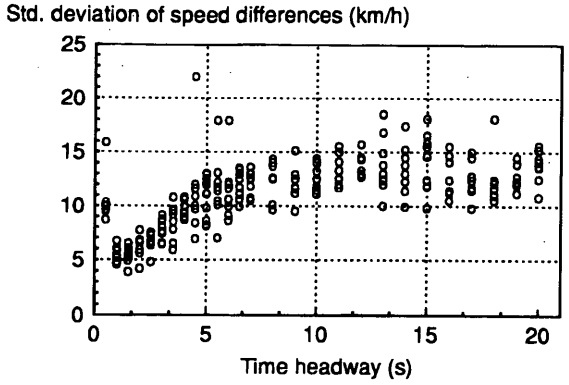



**FIGURE 1 Standard deviation of speed differences of successive vehicles as function of time headway (*3*).**

$$r = \frac{\text{veh-hr in platoon in section}}{\text{veh-hr in section}} = \frac{k_1 l_1 (t_1 + t_2)}{k_1 l_1 (t_1 + t_2) + k_2 1_2 (t_1 + t_2)}$$

$$= \frac{k_1 l_1}{k_1 l_1 + k_2 l_2} \tag{1}$$

where veh-hr equals vehicle hours.

If a local measurement is made at the end of the road section, the resulting local platoon percentage, $p$, can be calculated. For simplicity, the measurement is made over a period during which one pulse of high- and low-density traffic passes the measuring point.

In the local measurement if can be seen that the number of vehicles passing the observation point in platoon ($N_1$) and the number of vehicles passing outside the platoon ($N_2$) are

$$N_1 = q_1 t_1 \tag{2}$$

$$N = q_2 t_2 \tag{3}$$

These numbers are not the same as the number of vehicles in platoon ($k_1 l_1$) and outside the platoon ($k_2 l_2$) along the road. This is because during the time that the platoon passes the measuring point, new vehicles enter the platoon. The same applies for nonplatooned traffic.

Using Equations 2 and 3, the share of vehicles in platoons in the local measurement can be calculated:

$$p = \frac{N_1}{N_1 + N_2} \tag{4}$$

The times $t_1$ and $t_2$ can be calculated by the help of the speed of the density change, $c$, that is,

$$t_1 = \frac{l_1}{c} \tag{5}$$

$$t_2 = \frac{l_2}{c} \tag{6}$$

The fundamental flow relationship of traffic gives

$$q_1 = u_1 k_1 \tag{7}$$

$$q_2 = u_2 k_2 \tag{8}$$

Combining all these equations gives

$$p = \frac{N_1}{N_1 + N_2} = \frac{q_1 t_1}{(q_1 t_1 + q_2 t_2)} = \frac{u_1 k_1 \frac{l_1}{c}}{u_1 k_1 \frac{l_1}{c} + u_2 k_2 \frac{l_2}{c}} = \frac{u_1 k_1 l_1}{u_1 k_1 l_1 + u_2 k_2 l_2} \tag{9}$$

For the duration of the measurement,

$$t = t_1 + t_2 = \frac{l_1 + l_2}{c} = \frac{l}{c} \tag{10}$$

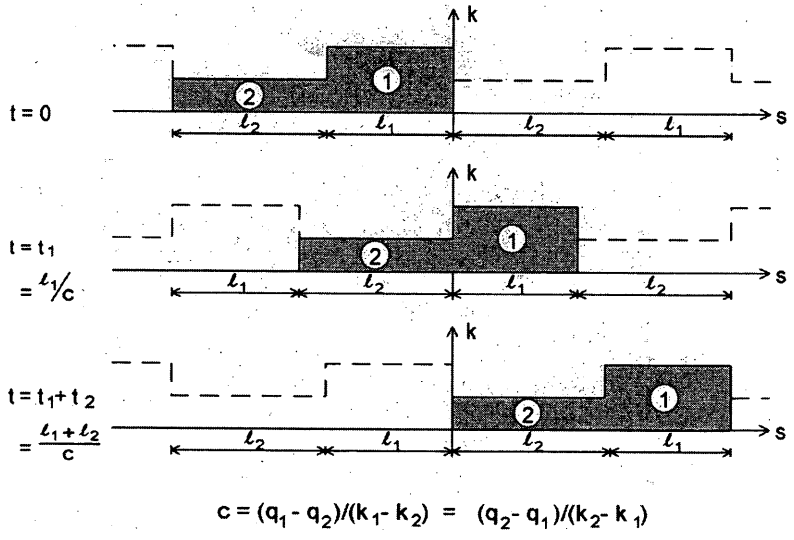



**FIGURE 2 Simplified description of platooned and free traffic flow along a road.**

and for the total flow rate,

$$q = \frac{N_1 + N_2}{t} = \frac{q_1 t_1 + q_2 t_2}{t} = \frac{u_1 k_1 \frac{l_1}{c} + u_2 k_2 \frac{l_2}{c}}{\frac{l}{c}}$$

$$= \frac{u_1 k_1 l_1 + u_2 k_2 l_2}{l} \tag{11}$$

This gives the local platoon percentage

$$p = \frac{q_1 l_1}{ql} = \frac{u_1 k_1 l_1}{ukl} \tag{12}$$

where $u$ is the space mean speed of the total traffic flow and can, according to the definition, be calculated as

$$u = \frac{u_1 k_1 l_1 + u_2 k_2 l_2}{k_1 l_1 + k_2 l_2} \tag{13}$$

The space mean speed $u$ in Equation 13 can be calculated from the locally measured speeds as the harmonic mean, that is,

$$u = \frac{N_1 + N_2}{\frac{N_1}{u_1} + \frac{N_2}{u_2}} = \frac{q_1 t_1 + q_2 t_2}{\frac{q_1 t_1}{u_1} + \frac{q_2 t_2}{u_2}} = \frac{u_1 k_1 \frac{l_1}{c} + u_2 k_2 \frac{l_2}{c}}{k_1 \frac{l_1}{c} + k_2 \frac{l_2}{c}}$$

$$= \frac{u_1 k_1 l_1 + u_2 k_2 l_2}{k_1 l_1 + k_2 l_2} \tag{14}$$

which is the same result as given in Equation 13.

Consequently, $k$ in Equation 12 is the mean density of the total traffic flow and can be calculated as follows:

$$k = \frac{q}{u} = \frac{k_1 l_1 + k_2 l_2}{l} \tag{15}$$

Equation 1 for the PTD can now be rewritten as

$$r = \frac{k_1 l_1}{k_1 l_1 + k_2 l_2} = \frac{k_1 l_1}{kl} = \frac{k^p}{k} \tag{16}$$

where $k^p = k_1 l_1 / l$ is the mean density of platooned vehicles over the whole road section $l$.

If the equation of the local platoon percentage (Equation 12) is compared with that of the PTD (Equation 16),

$$p = \frac{u_1}{u} \frac{k_1 l_1}{kl} = \frac{u_1}{u} r = \frac{u^p}{u} r \tag{17}$$

where $u^p = u_1$ is the space mean speed of platooned vehicles.

In this way the relationship between the PTD, $r$, and the local platoon percentage, $p$, has been derived:

$$r = \frac{u}{u^p} p \tag{18}$$

This equation gives the result that PTD is calculated from the local platoon percentage by multiplying the local percentage by the ratio of the space mean speed of the whole traffic and the space mean speed of the platooned traffic.

## Generalization of Relationship

In the preceding analysis, the traffic flow was simplified to two densities traveling at constant speeds typical of the density in question. In this kind of traffic flow, no overtakings are made. In real traffic, all sorts of situations exist and they can be described with vehicle trajectories in time-space domain (Figure 3).

The fundamental flow relationship of traffic can be generalized for situations given in Figure 3 [for example, see the work by Leutzbach (*8*)]. The three variables—flow, density, and mean speed—are then defined in the following way:

- Traffic flow $q$ = total vehicle kilometers of travel ($S$) in the domain divided by the area of the domain, that is,

$$q = \frac{S}{LT} \tag{19}$$

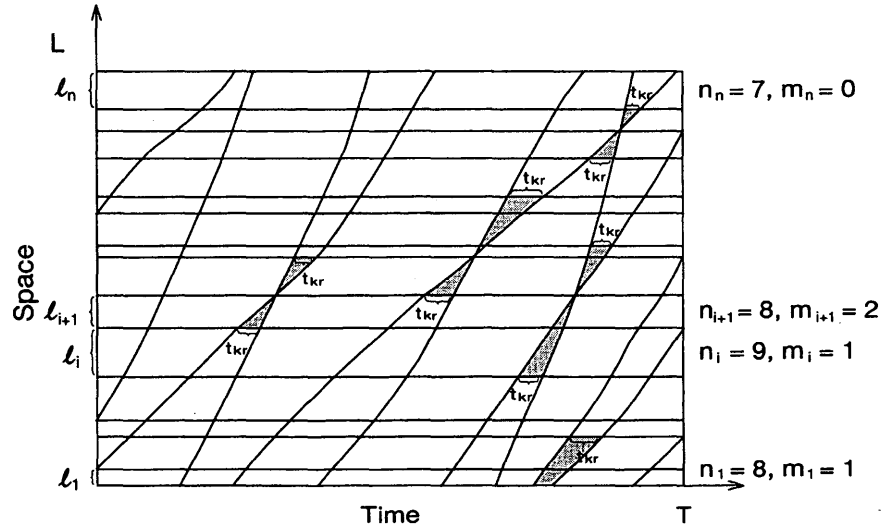



**FIGURE 3 Time-space trajectories of vehicles in generalized description of traffic flow.**

• Traffic density $k$ = total vehicle hours of travel ($T_{tot}$) in the domain divided by the area of the domain, that is,

$$k = \frac{T_{tot}}{LT} \tag{20}$$

• Space mean speed of traffic $u$ = total vehicle kilometers of travel in the domain divided by the total vehicle hours of travel in the domain, that is,

$$u = \frac{S}{T_{tot}} \tag{21}$$

The aim now is to generalize the relationship of local platoon percentage, $p$, and PTD, $r$. To do that, first define the criterion of platooning as a constant headway of $t_{kr}$. After that, divide the time-space domain (Figure 3) with horizontal lines into sections of road (length $l_i$) with a constant number of vehicles ($n_i$) and a constant number of vehicles in platoons ($m_i$) inside each section during the whole period of time $T$. Because of different speeds of vehicles, the vertical (space) distances between platooned vehicles at the borderlines of the road sections do not have a constant value.

For each of the sections $i$,

$$p_i = \frac{m_i}{n_i}$$

or (22)

$$m_i = p_i n_i$$

where $p_i$ is the local platoon percentage at any cross section inside the road section $i$.

The PTD $r_i$ is calculated as the sum of travel times in platoons divided by the total travel time in section $i$ during the time $T$.

The sum of the travel times of individual vehicles in section $i$ is $T_i$:

$$T_i = \sum_{j=1}^{n_i} t_{ij} = n_i t_i = \frac{n_i l_i}{u_i} \tag{23}$$

where

$i = 1 \ldots, n$ = section index,
$j = 1 \ldots, n_i$ = vehicle index inside section $i$,
$u_i$ = mean travel speed of vehicles in section $i$, and
$t_i$ = mean travel time of vehicles in section $i$.

For the travel time in platoons in section $i$, these corresponding relationships exist:

$$T_i^p = \sum_{h=1}^{m_i} t_{ih}^p = m_i t_i^p = p_i n_i t_i^p = \frac{p_i n_i l_i}{u_i^p} \tag{24}$$

where $h = 1 \ldots, m_i$ is the platoon vehicle index in section $i$, and superscript $p$ indicates platooned vehicles only.

For the PTD,

$$r_i = \frac{T_i^p}{T_i} = \frac{\frac{p_i n_i l_i}{u_i^p}}{\frac{n_i l_i}{u_i}} = p_i \frac{u_i}{u_i^p} \tag{25}$$

Furthermore,

$$r_i = p_i \frac{u_i}{u_i^p} = \frac{p_i \frac{q_i}{k_i}}{\frac{q_i^p}{k_i^p}} = \frac{k_i^p}{k_i} \tag{26}$$

So, for a single road section $i$, the same relationship is valid that was found before for the simplified case using the shock wave theory.

For the whole road section $L$, the following relationships are valid:

$$S = \sum_{i=1}^{n} n_i l_i \tag{27}$$

where $S$ gives the total amount of vehicle kilometers of travel inside the domain,

$$S^p = \sum_{n=1}^{n} m_i l_i = \sum_{n=1}^{n} p_i n_i l_i \tag{28}$$

where $S^p$ is the total amount of vehicle kilometers in platoons inside the domain,

$$T_{tot} = \sum_{i=1}^{n} n_i t_i \tag{29}$$

where $T_{tot}$ gives the total number of vehicle hours inside the domain, and

$$T_{tot}^p = \sum_{i=1}^{n} m_i t_i^p = \sum_{i=1}^{n} p_i n_i t_i^p \tag{30}$$

where $T_{tot}^p$ gives the total number of vehicle hours in platoons inside the domain.

Before going further, it is worth noting that the general definitions of traffic flow variables given in Equations 19 through 21 are valid for the platooned part of the traffic in the same way as for the traffic flow as a whole.

For the mean platoon percentage of the whole domain, $p$,

$$p = \frac{\sum_{i=1}^{n} m_i l_i}{\sum_{i=1}^{n} n_i l_i} = \frac{S^p}{S} = \frac{q^p LT}{qLT} = \frac{q^p}{q} \tag{31}$$

Equation 31 has used the general definitions of traffic flow (Equation 19) for platooned traffic.

For the PTD of the whole domain, $r$,

$$r = \frac{\sum_{i=1}^{n} T_i^p}{\sum_{i=1}^{n} T_i} = \frac{T_{tot}^p}{T_{tot}} = \frac{k^p LT}{kLT} = \frac{k^p}{k} \tag{32}$$

Here, again, the general definition of traffic density (Equation 20) for platooned traffic has been used.

Now, by using the fundamental flow relationship of traffic and Equations 31 and 32,

$$p = \frac{u^p k^p}{uk} = r \frac{u^p}{u} \tag{33}$$

or

$$r = p\frac{u}{u^p} = p\frac{\frac{l}{u^p}}{\frac{l}{u}} = \frac{pt_{av}^p}{t_{av}} \tag{34}$$

where $t_{av}$ is the average travel time of vehicles in the whole time-space domain, and $t_{av}^p$ is the average travel time of platooned vehicles in the whole time-space domain.

In this way the relationship between the local platoon percentage, $p$, and the PTD, $r$, has been generalized to any kind of traffic flow. For the relationship to be valid, one must bear in mind the way in which the variables in the relationship are defined. The considerations have been made for one direction of traffic only, but they easily generalize for two-way flow also.

In general, the mean platoon percentage of a time-space domain is calculated by dividing the vehicle kilometers driven in platoons by the total vehicle kilometers driven in the domain. Consequently, the PTD is calculated by dividing the vehicle hours driven in platoons by the total vehicle hours driven in the domain. So platoon percentage is proportional to vehicle kilometers (or traffic flow rates), and PTD is proportional to vehicle hours (or traffic densities).

Furthermore, the mean speed of all vehicles in the domain is calculated by dividing the total vehicle kilometers by the total vehicle hours in the domain. Consequently, the mean speed of platooned vehicles is calculated by dividing the vehicle kilometers driven in platoons by the vehicle hours driven in platoons inside the domain.

Usually it can be expected that the space mean speed of platoons is lower than the space mean speed of the whole traffic. In principle, this gives the relationships between PTD and platoon percentage given in Figure 4. The difference between the curves can be quite small if the speed variation of traffic flow is small. On the other hand, if there are great differences in speeds of different vehicle categories, the difference between platoon percentage and PTD can be quite remarkable. For example, this type of situation can exist on roads where the percentage of heavy vehicles is high and the terrain is hilly.

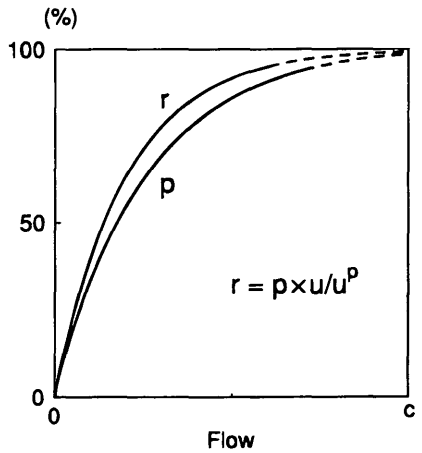



**FIGURE 4 Hypothetical relationship between PTD ($r$) and local platoon percentage ($p$).**

### Calculation of PTD from Several Local Measurements

If several local measurements are made along the road, then the mean platoon percentage, $p$, can be calculated using Equation 35 (see Equation 31).

$$p = \frac{\sum_{i=1}^{n} m_i l_i}{\sum_{i=1}^{n} n_i l_i} = \frac{\sum_{i=1}^{n} p_i n_i l_i}{\sum_{i=1}^{n} n_i l_i} = \frac{\sum_{i=1}^{n} p_i q_i l_i}{\sum_{i=1}^{n} q_i l_i} \tag{35}$$

(Here, and in the following, the same notation is used for local measurements as in the former analysis for road sections.)

If the traffic flow rate along the road section is constant, Equation 35 simplifies to Equation 36:

$$p = \frac{\sum_{i=1}^{n} p_i l_i}{L} \tag{36}$$

In this case the mean platoon percentage can be calculated as the weighted average of the platoon percentages of consecutive road sections by using the section lengths as weights.

Consequently, the mean PTD can be calculated using Equation 37:

$$r = \frac{\sum_{i=1}^{n} T_i^p}{\sum_{i=1}^{n} T_i} = \frac{\sum_{i=1}^{n} r_i T_i}{\sum_{i=1}^{n} T_i} = \frac{\sum_{i=1}^{n} r_i n_i t_i}{\sum_{i=1}^{n} n_i t_i} = \frac{\sum_{i=1}^{n} r_i q_i t_i}{\sum_{i=1}^{n} q_i t_i}$$

$$= \frac{\sum_{i=1}^{n} r_i q_i \frac{l_i}{u_i}}{\sum_{i=1}^{n} q_i \frac{l_i}{u_i}} = \frac{\sum_{i=1}^{n} r_i k_i l_i}{\sum_{i=1}^{n} k_i l_i} \tag{37}$$

If the basic relationship between $p_i$ and $r_i$ is used (Equation 25), then on the basis of Equation 37,

$$r = \frac{\sum_{i=1}^{n} r_i k_i l_i}{\sum_{i=1}^{n} k_i l_i} = \frac{\sum_{i=1}^{n} p_i \frac{u_i}{u_i^p} k_i l_i}{\sum_{i=1}^{n} k_i l_i} = \frac{\sum_{i=1}^{n} p_i q_i \frac{l_i}{u_i^p}}{\sum_{i=1}^{n} q_i \frac{l_i}{u_i}} = \frac{\sum_{i=1}^{n} p_i q_i t_i^p}{\sum_{i=1}^{n} q_i t_i} \tag{38}$$

If $q_i$ along the road is constant, then Equations 37 and 38 can be simplified to Equation 39:

$$r = \frac{\sum_{i=1}^{n} r_i t_i}{\sum_{i=1}^{n} t_i} = \frac{\sum_{i=1}^{n} p_i t_i^p}{\sum_{i=1}^{n} t_i} \tag{39}$$

So, in the case of constant flow, the mean PTD can be calculated as the weighted average of the PTD values of consecutive road sections by using the section travel times as weights. The mean PTD can in this case also be calculated on the basis of local platoon percentages, mean travel times in platoons, and mean travel times of the whole traffic.

By comparing Equations 37 and 38,

$$r = \frac{\sum_{i=1}^{n} r_i q_i t_i}{\sum_{i=1}^{n} q_i t_i} = \frac{\sum_{i=1}^{n} p_i q_i t_i^p}{\sum_{i=1}^{n} q_i t_i} \quad (40)$$

Equations 35 through 40 describe the estimation for the mean platoon percentage and mean PTD on the basis of local measurements. The speed values $u_i^p$ and $u_i$ must be calculated as the harmonic means of the spot speeds of the individual vehicles at each measuring point to represent the corresponding travel speeds.

Equation 40 describes the relationship of PTD to mean travel times of platooned and all vehicles through the individual road sections. The travel times, $t_i$ and $t_i^p$, can be calculated by using the section lengths and harmonic mean speeds of platooned vehicles and all vehicles at each measuring point.

## SIMULATION RESULTS AND MEASUREMENTS

### Simulation Results with TRARR

Simulation is a suitable tool for estimating the relationship between PTD and platoon percentage. In simulation one can register exactly when vehicles are hindered by others and how long they do travel in platoons, thus resulting in a PTD estimate over the whole simulated road section. For platoon percentage, local registration of platooned vehicles can be made at several cross sections. As far as we know, no program uses the previous derived definition based on vehicle kilometers to create an average platoon percentage for the whole simulated road section.

Hoban has described the difference of PTD and percentage following observed in simulations (*9*). He uses a 4-sec headway as the platoon criterion. Whether the platoon leaders are included in the platoon percentage is not clear because the report (*9*) refers only briefly to the research in question. Hoban also states that the average of several consecutive cross sections is used, but he gives no indication of any kind of weighing in the calculation of the average platoon percentage.

The reported platoon percentages in Hoban's work are, with one minor exception, lower than the percentage following values that basically are PTD values (*9*). This is in accordance with the preceding theory if it is assumed that the mean speed in platoons usually is lower than the mean speed of the whole traffic. The report does not give any results about the relationship between the average speed in platoons and traffic as a whole.

From the data of some Finnish simulation studies on a highclass rural road (*10*), simulated local platoon percentages and simulated PTD values can also be compared. The simulation program used was the Australian TRARR. A short English description of the calibration and use of the program, including some results, is given elsewhere (*11*). The platoon percentages given here were calculated as the weighted average of local percentage values given by the simulation program. A 5-sec time headway was used as the platoon criterion, and local platoon percentages were printed at about 1-km intervals. The program calculated the actual PTD values over the whole simulated road section on the basis of travel times in platoons and outside platoons, thus enabling the comparison with the local platoon percentages. The headway distributions for the incoming flows were given as measured in the field.

In Figure 5, the Finnish and Australian results are compared. The Finnish values are for one direction of traffic only. The simulated local platoon percentages in the figure are usually 3 to 5 percent lower than the simulated PTD values (maximum difference in the Australian data is 7.0 and in the Finnish data, 7.8 percent), but in some cases the PTD value is up to 2 percent lower than the corresponding platoon percentage. The results appear to correspond with each other quite well, despite the differences in platoon criteria and the flow in question.

In the Finnish simulation data, the mean platoon percentage was calculated as the weighted average of 14 to 16 local measurements over 16 to 22 km of road, and PTD values were taken directly from the simulation output. The variation of the individual local platoon percentages ranged from 3 to 19 percent; usually it was about 10 percent. So, quite a lot of variation in the local platoon percentages can exist, even on a high-class road with a low percentage of heavy vehicles.

The TRARR outputs of the Finnish simulations do not give any information about the speeds of platooned vehicles. In the following, Finnish field measurement data are used to analyze more closely the relationship between the speeds of platooned vehicles and all vehicles.

### Results of Field Measurements

The field measurements used here were done on ordinary two-lane rural roads in Finland in 1984 (*4*). The data were gathered with several individual local measurements on different kinds of roads by using double induction loops. The space mean speeds were calculated as the harmonic means of the measured individual speeds.

Figure 6 presents the ratio of the speeds ($u/u^p$) as a function of flow in both directions of travel together. It can be seen that the ratio usually varies between 1.00 and 1.05. The ratio is decreasing when the flow is increasing. This is natural because in higher flow rates, more vehicles drive in platoons.

The analysis of one-way traffic shows very similar values for the $u/u^p$ ratio for the main direction of flow. In the direction of the

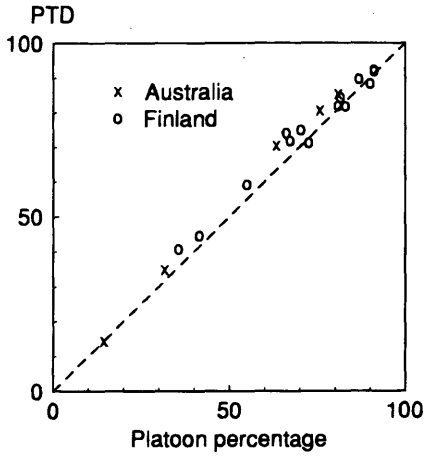



**FIGURE 5 Comparison of simulated mean platoon percentage and simulated PTD on Australian (*9*) and on Finnish two-lane roads.**

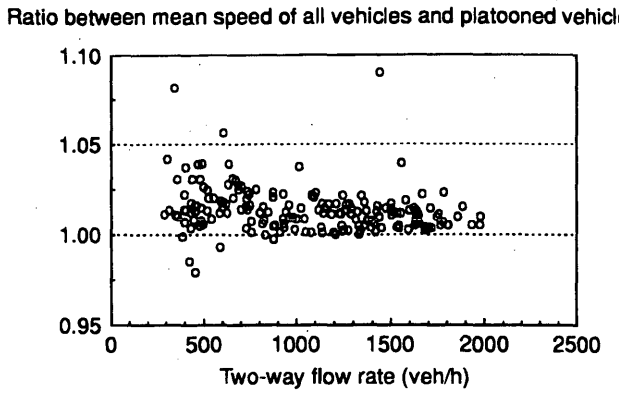



**FIGURE 6 Ratio between space mean speed of all vehicles and space mean speed of vehicles in platoons as function of flow (15-min intervals).**

minor flow, the ratio varies somewhat more—the usual range being from 1.00 to 1.10. (Figures of these analyses are not given here.)

As Figure 6 indicates, the $u/u^p$ ratio can also have values less than 1.00. Thus in some 15-min intervals the platoons travel faster than free vehicles. These values are, logically, located in the area of low volumes; at high volumes platoons appear to be slower than free vehicles.

A simple linear regression analysis for the ratio $u/u^p$ was made, resulting in Equations 41 and 42. For both directions together,

$$\frac{u}{u^p} = 1.0204 - 0.00000656q \quad (41)$$

For the main direction of travel,

$$\frac{u}{u^p} = 1.0171 - 0.00000984q \quad (42)$$

Because of the very low slope of the regression lines, the coefficients of determination ($r^2$) are very low, between 0.05 and 0.09. The equations are given here as rough approximations of the interdependence between the $u/u_p$ values and flow. Equation 41 gives a $u/u^p$ value of 1.00 at a flow rate of 3100 veh/hr; the corresponding flow value for the main direction, from Equation 42, is 1740 veh/hr. These values can be seen as approximate capacity values for the whole set of measuring locations.

In Figure 7 the platoon percentage observations and the corresponding PTD values calculated with Equation 34 are given for one location. The $r$- and $p$-points do not differ very much from each other, as could be expected from the low values of the $u/u^p$ ratio in the data in Figure 6.

## DISCUSSION OF RESULTS

Locally measured platooning is a biased estimate of platooning along the road. This bias is of the same origin as the difference in all local and space variables of traffic flow, like the difference of time mean speed and space mean speed. The paper shows how this bias can be corrected. The correction is important if local platooning data are used to approximate the PTD and level of service on two-lane, two-way roads.

The general relationship derived between the local platoon percentage and the approximate PTD is quite simple. In a local measurement with induction loops, all the variables needed for the unbiased estimation of the approximate PTD value can easily be achieved. The harmonic means of speeds can be used to estimate the space mean speeds, and platoon percentage is directly calculated as the share of vehicles driving with headways of fewer than 5 sec.

For a long road section, the estimate of the PTD can be calculated as the weighted average of the results of different local measurements inside the section. The measuring sites must then be chosen with care to represent different road and traffic characteristics. The length of road that each measuring site is representing must also be known.

Simulation study results are well in accordance with the theoretical analysis of the relationship of PTD and local platoon percentage. The differences between corrected and uncorrected local platoon percentages, as estimated from field measurements, appear

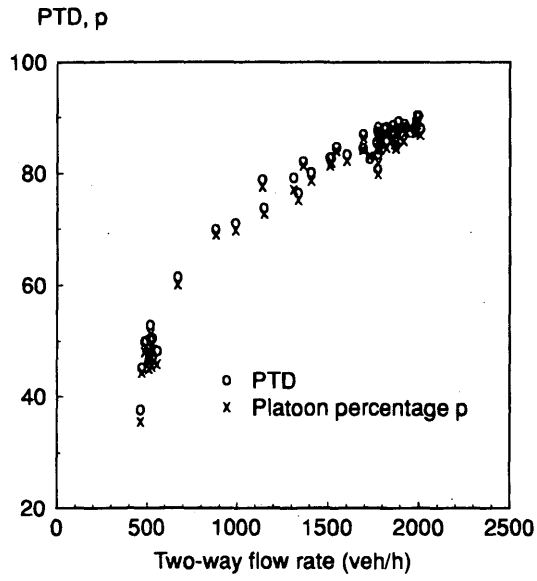



**FIGURE 7 Example of observed local platoon percentages and corresponding PTD estimates as function of flow rate on two-lane road.**

to be similar to the differences between PTD and local platooning observed in simulation. On roads with poor alignment and high percentages of heavy vehicles, the differences can be clearly higher than those observed in the data of this work. Even small (2 to 3 percent) corrections to the PTD values can greatly affect the level-of-service estimation because of the low slope of the PTD-flow curve in the critical level-of-service areas (LOS D and E).

According to the assumptions in the estimation, with a fixed-time headway as the platooning criterion, vehicles with a faster vehicle close in front of them are classified as platooned (or delayed). In principle, this can lead to an overestimation of the PTD. The general equation for the relationship of local platooning and the PTD estimate derived earlier is valid also if these situations are excluded from the measurement data with a speed difference criterion. The effects of such a change in the measurements, as well as a suitable criterion for speed difference, were not analyzed in this paper.

There are many problems in the precise definition and measurement of platooning and hindrance experienced by drivers. So a simple and locally measurable estimate of PTD, like the corrected local platoon percentage described here, is needed for practical purposes, at least until more thorough field data of real driver experience are available.

## REFERENCES


1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
2. Enberg, Å., and M. Pursula. Traffic Flow and Level of Service on High-Class Two-Lane Rural Roads in Finland. In *Highway Capacity and Level of Service*, (U. Brannholte, ed.) A. A. Balkema, Rotterdam, the Netherlands, 1991, pp. 119–126.
3. Enberg, Å., and M. Pursula. *Basic Characteristics of Traffic Flow on High-Class Two-Lane Rural Roads in Finland* (in Finnish). Research Report 12/1992. The Finnish National Road Administration, Helsinki, 1992.
4. Pursula, M., and Å. Enberg. Characteristics and Level of Service Estimation of Traffic Flow on Two-Lane Rural Roads in Finland. In *Transportation Research Record 1320*, TRB, National Research Council, Washington, D.C., 1991, pp. 135–143.
5. Messer, C. J. *Two-Lane, Two-Way Rural Highway Capacity.* Final Report, prepared for NCHRP. TRB, National Research Council, Washington, D.C., 1983.
6. Luttinen, R. T. Statistical Properties of Vehicle Time Headways. In *Transportation Research Record 1365*, TRB, National Research Council, Washington, D.C., 1992, pp. 92–98.
7. McLean, J. R. *Two-Lane Highway Traffic Operations. Theory and Practice.* Gordon and Breach Science Publishers, Melbourne, Australia, 1989.
8. Leutzbach, W. *Introduction to the Theory of Traffic Flow.* Springer-Verlag, Berlin, Germany, 1988.
9. Hoban, C. J. *Low-Cost Methods for Improving Traffic Operations and Safety on Rural Roads.* Draft Report DN 1512. Australian Road Research Board, Nunawading, 1987.
10. Siimes, H., and M. Pursula. *Simulation of Two- and Three-Lane Road Traffic with TRARR* (in Finnish). Publication 80. Department of Transportation Engineering, Helsinki University of Technology, Otaniemi, 1994.
11. Pursula, M., and H. Siimes. A Simulation Study of a High-Class Three-Lane Rural Road. In *ITE 1993 Compendium of Technical Papers* (Publication PP-037), Institute of Transportation Engineers, Washington, D.C., 1993, pp. 16–20.


*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Capacity for Right Turn on Red

MARK R. VIRKLER AND RAMANA RAO MADDELA

Right turn on red (RTOR) can have a significant effect on intersection operation, but RTOR volume data are seldom collected and are not available for solving intersection design problems. Two techniques have been suggested for analyzing RTOR in the absence of field data. The first is to assume that during a protected left-turn phase, the RTOR movement that is "shadowed" by the protected left turn can have a volume equal to the per-lane volume of the shadowing left turn. The second technique suggests that the movement of an RTOR vehicle is analogous to the movement of a right-turning vehicle at a stop sign–controlled, unsignalized intersection. Extra capacity is present for an RTOR vehicle to move through the unsaturated green portions of movements that currently have a green indication. These two approaches are examined with data from 40 intersections to determine ways to provide a more realistic estimate of intersection operations when significant RTOR volumes may occur. Both approaches yield significant changes in reported intersection operation. For instance, shadowing improved the reported level of service for almost a third of the exclusive right-turn lanes. The stop sign analogy drastically reduced the number of right-turn lanes reported as over capacity. Neither approach is modeled correctly by the *Highway Capacity Manual* delay equation used for estimating level of service, but modeling of RTOR with the stop sign analogy could lead to a more realistic description of intersection performance and a more efficient use of green time as well as eliminate the construction of unneeded lanes.

The 1985 *Highway Capacity Manual* (HCM) states that when right turn on red (RTOR) is allowed at a signalized intersection, the analyst may reduce the right-turn volume by the RTOR volume (*1*). To implement this concept in the HCM operational analysis, one needs an estimate of the RTOR volume. However, such data are seldom collected during intersection traffic counts and would not be available for the design or retiming of an intersection.

While analyzing existing traffic counts with the HCM operational procedure, Virkler and Chen (*2*) found that RTOR may have a significant influence on the resulting flow-to-capacity ratios (v/c) and level of service (LOS). In several cases, right-turn volumes were much greater than the estimated capacity on green. Since actual volume cannot exceed capacity, the most likely explanation was that a significant RTOR volume was present. In these cases a good estimate of RTOR flow could dramatically change the estimated v/c and LOS.

Two techniques have been suggested for analyzing RTOR in the absence of field data. The first is to assume that during a protected left-turn phase, a parallel RTOR movement can take place because there is no conflicting traffic (e.g., during a protected left-turn phase for traffic approaching from the south and turning to the west, RTOR traffic approaching from the west and turning to the south will have no conflicting traffic). This approach is included in the updated version of the HCM intersection operational procedure (*3*). The second technique, proposed by Luh and Lu, suggests that the movement of an RTOR vehicle is analogous to the movement of a right-turning vehicle at stop sign–controlled, unsignalized intersection (*4*). The HCM's procedure for a right-turn at an unsignalized intersection can therefore be modified to estimate the RTOR capacity.

The objective of this research was to examine these procedures to determine how to provide a more realistic estimate of intersection operations when significant RTOR volumes may occur. The procedures were applied to the data on 40 intersections used by Virkler and Chen. The procedures, data, analysis, results, and conclusions are described.

M. R. Virkler, Department of Civil Engineering, University of Missouri–Columbia, Columbia, Mo. 65211. R. R. Maddela, Andhra Pradesh State Engineering Services, Vijayawada, India 520010.

## ALTERNATIVE PROCEDURES

### New HCM Signalized Intersection Operational Analysis (Shadowing)

An analyst can estimate an expected RTOR volume with the revised signalized intersection operational analysis. This expected volume is recommended for use if the field-counted RTOR volume is not available. The protected left-turn volume (on a per-lane basis) is deducted from the "shadowed" RTOR volume, if an exclusive right-turn lane is available. For example, if dual left-turn lanes carry 300 left-turning vehicles from the northbound approach (150 left turns per lane) during a protected left-turn phase, then 150 RTOR vehicles can be subtracted from the eastbound approach right-turn lane volume. If the RTOR approach has a shared right/through lane, then this number is reduced according to the likelihood that the RTOR will be blocked by a through vehicle (*3*).

### HCM Stop Sign Analogy

Luh and Lu (*4*) demonstrated that the RTOR movement is similar to a right turn made at a stop sign–controlled approach. The HCM procedure for an unsignalized intersection can therefore be used to estimate the capacity for RTOR. An abbreviated version of the steps are shown in the list. Since a complete description of the steps would be lengthy, the reader is referred to specific tables and figures of the HCM (as cited) for more detailed discussion of the concepts.

1. *Identify conflicting traffic.* During the red phase, the right-turn vehicle can make an RTOR maneuver after stopping (if it is controlled by a red ball) or it can turn right by yielding to other movements that have the right of way (if it is controlled by a yield sign). The traffic to which the RTOR vehicle yields is called conflicting traffic (HCM, Figure 10-2). This conflicting traffic could be through traffic from the left side approach, protected left turns from the opposite approach, or no traffic (e.g., during a shadowed phase). Depending on the signal phasing, RTOR might be made during one

or two or all of these conflicting flows during a cycle. Each phase would be analyzed separately.

2. *Compute unsaturated red time.* If the conflicting movements have unsaturated green time during their phases (i.e., they are not operating at capacity), then the RTOR becomes possible. The amount of the time that could be used for RTOR movements is the unsaturated green time of the conflicting traffic (referred to as the "unsaturated red time" for RTOR). The calculation for unsaturated green time is demonstrated in Figure 9-9 of the HCM.

3. *Find critical gap.* Critical gap is the 50th-percentile gap used for the right turn at a stop or yield sign, as provided by HCM Table 10-2.

4. *Compute conflicting flow rate.* The conflicting traffic rate of flow during each phase's unsaturated red time is determined. The RTOR maneuver will be similar to a right turn from a stop or yield sign onto a street having this rate of flow. The conflicting flow will equal the arrival flow rate of the subject movement, since these conflicting vehicles will not have been part of an approach queue at the intersection (i.e., these conflicting vehicles arrived at the intersection during the green for their phase when no queue was present).

5. *Find potential capacity during each unsaturated red time.* The potential capacity is the capacity under ideal conditions (HCM, Figure 10-3). Each part of the red time (i.e., through traffic from the left side, left turns from the opposite approach, and no conflicting traffic during a shadowed phase) can have a potential capacity.

6. *Find adjustment factor for pedestrians.* The HCM adjustment for pedestrians blocking the right turn (HCM, Table 9-11) is applied to the potential capacity. Use the pedestrian volume that would interfere with the RTOR vehicle because of the signal indication

7. *Compute actual capacity.* The potential capacity for RTOR during each unsaturated red time is summed. This is the RTOR capacity of an exclusive right-turn lane.

For a shared right/through lane this number is reduced according to the likelihood that the RTOR will be blocked by a through vehicle.

### Updated HCM Stop Sign Analogy

The HCM procedure for unsignalized intersections, like that for signalized intersections, is being revised. The newer unsignalized approach (*5*) is similar to the earlier version, but the resulting capacity numbers are different. Therefore the stop sign analogy was also applied using the new the HCM unsignalized intersection analysis procedure.

## DESCRIPTION OF DATA AND DATA ANALYSIS

Virkler and Chen (*2*) examined pretimed and actuated signals for 40 typical intersections of the Missouri state highway system. Half of the data were from a large city (St. Louis), and the other half were from three smaller cities (Columbia, Jefferson City, and Sedalia). The data included 15-min turning movement counts, phase plans, and intersection condition diagrams. No RTOR volumes were available. The data contained both a.m. and p.m. peak-hour traffic counts, so 80 peak-period data sets were available.

### Application of Shadowing Procedure

The 1994 HCM RTOR treatment can be applied to the right-turn lane groups that are shadowed by protected left-turn phases. The data set included 45 intersections with 112 approaches having shadowing phases. Of the 112 approaches, 71 approaches had exclusive right-turn lanes and 41 had shared right/through lanes.

The expected RTOR volumes for the lane groups were calculated by the shadowing procedure. In some cases the RTOR volume exceeded the field-counted right-turn volume. The *Highway Capacity Software* (*6*) would not allow an RTOR volume greater than the right-turn volume. In such cases the expected RTORs were set equal to the right-turn volume. Occasionally with a right-turn volume equal to 0, the software gave inconsistent values for delay (i.e., different answers were provided by subsequent runs of the software, apparently due to a memory problem caused by a volume equal to 0). To gain consistent output, a minimum value of 1 was assigned to the right-turn volume on green.

### Application of HCM Stop Sign Analogy

Although the shadowing procedure applies only to shadowed RTOR, the stop sign analogy can be applied to right turns with or without a shadowing phase. The stop sign analogy was applied to all the exclusive right-turn lanes in the data. Because of the large time requirement for data analysis, shared right-turn lanes were omitted from this application. There were 99 exclusive right-turn lanes in the data. In cases in which the conflicting flow was very low–potential capacity values beyond 1,000 passenger cars per hour (pcph)—the HCM nomographs did not show potential capacity values. In these cases the curves were extrapolated.

### Application of Updated HCM Stop Sign Analogy

The updated HCM stop sign analogy was applied to the 99 exclusive right-turn lanes. The updated unsignalized intersection parameters gave higher potential capacities in most of the cases.

## RESULTS

The different effects of the procedures complicates the comparison of results. The shadowing method reduces the right-turn volume; the HCM stop sign analogies increase the capacity of the right-turn lane group. These effects are described separately.

### Right-Turn Volume Reduction from Shadowing

The right-turn volume reduction from the shadowing procedure ranged from 0 to 100 percent. In exclusive right-turn lanes the mean reduction was 67 percent, and in shared through/right-turn lanes the mean reduction was 36 percent. Figures 1 and 2 are comparisons of lane group v/c ratios with and without the right-turn volume reduction. Figure 1 deals with exclusive right-turn lanes. Many exclusive right-turn lane groups showed large v/c reductions. Figure 2 shows shared right-turn lane groups. The v/c reduction was dramatic for only a few shared lanes. Since the original lane group volumes (without shadowing) were actual flows, no v/c ratio should have exceeded unity if the no-RTOR assumption was correct. The shadowing procedure appears to make the large v/c ratios more reasonable. On the other hand, not all right turns will occur on red. With the shadowing procedure, the right-turn volume reductions can equal 100 percent of the original right-turn volume.

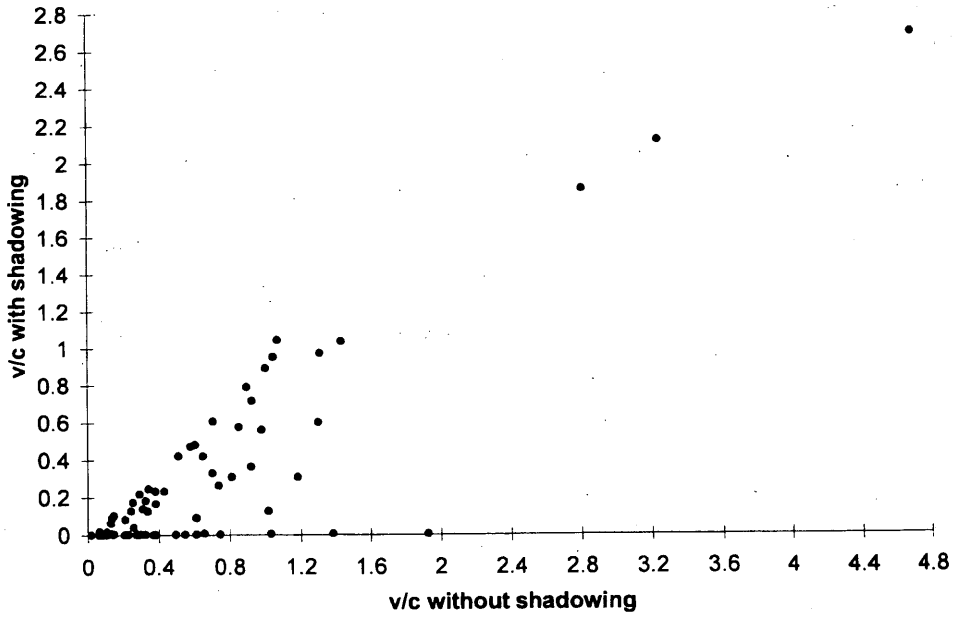



**FIGURE 1 Effects of shadowing on exclusive right-turn lane groups.**

The HCM average stopped delay equation is applicable for a 15-min period when v/c is less than 1 and can be applied to somewhat higher v/c ratios if the resulting queue would clear during the next 15-min period. If the v/c is too high, the delay and LOS (which is based on average stopped delay) are both reported as an asterisk (*) rather than in seconds and in an LOS category. A description of delay reductions would be incomplete because of the large number of situations in which the delay equation was not calculated because the v/c ratios exceeded the allowable maximum. However, the impact of shadowing on reported LOS is clearly visible. After the right-turn volume reductions from shadowing, many LOS results changed and many asterisks were replaced by a calculated LOS.

The delay equation used to determine LOS was not developed to consider RTOR. The shadowing procedure simply eliminates RTOR vehicles from the analysis. The following results should be read as a description of what the procedure will calculate, rather than as an accurate picture of the true LOS situation.

Figure 3 shows the changes in LOS for the 45 intersections (based on the average delay of all vehicles said to use the intersection and, therefore, not including subtracted RTOR vehicles). Four of the intersections, which originally included right-turn lane groups having v/c ratios too high for use of the delay model (in the before condition), were changed to LOS B, C, D, and E by shadowing (the after condition). Two intersections improved from LOS C to B, and one inter-

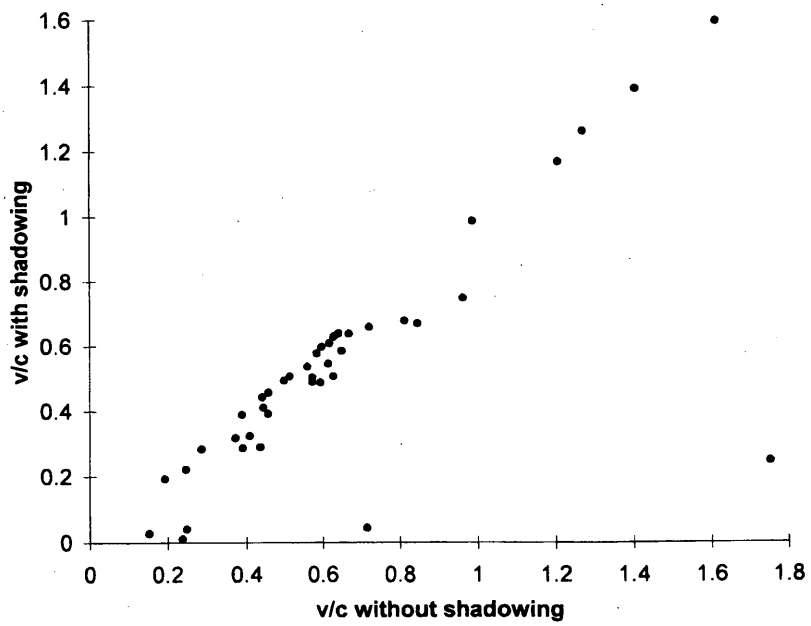



**FIGURE 2 Effects of shadowing on shared right-turn lane groups.**

| | AFTER | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| BEFORE | A | B | C | D | E | F | * | TOTAL |
| A | 0 | | | | | | | 0 |
| B | | 6 | 1 | | | | | 7 |
| C | | 2 | 10 | | | | | 12 |
| D | | | | 2 | | | | 2 |
| E | | | | | 1 | | | 1 |
| F | | | | | | 0 | | 0 |
| * | | 1 | 1 | 1 | 1 | | 19 | 23 |
| TOTAL | 0 | 8 | 12 | 3 | 2 | 0 | 19 | 45 |

**FIGURE 3 Changes in intersection LOS due to shadowing.**

section regressed from B to C. This negative impact on LOS was due to the reduction in the right-turn volume in a low-delay right-turn lane. The average delay of all vehicles included in this intersection's analysis increased from 14.7 to 15.1 sec (LOS C begins at 15.0 sec). Intersection delay sometimes increased because of the removal of low-delay right-turning vehicles, but the increases were all small.

The LOS impact of shadowing on the 112 intersection approaches (based on the average delay of all left through and right-turning vehicles said to use the approach and, therefore, not including subtracted RTOR vehicles) is shown in Figure 4. The top portion of the figure shows that for the 71 approaches with exclusive right-turn lanes, LOS improved in 12 cases (including 3 cases in which LOS could now be calculated because the v/c ratio had been reduced to within the range of the delay model). Two approaches jumped from LOS F to D, and one approach improved from E to C. The 41 approaches with shared right-turn lanes are described in the bottom half of Figure 4. LOS improved in seven cases (including three cases in which LOS could now be calculated because of the right-turn volume reduction). One approach leaped from LOS F to LOS D.

The impact on LOS was most dramatic within exclusive right-turn lane groups, as shown at the top of Figure 5. Among the 71 right-turn lanes, 22 (or 31 percent had improved LOS (including 5 that were now within the range of the delay model). Five lanes improved by two levels and two lanes improved from LOS F to LOS C.

The 41 shared right-turn lane groups are described at the bottom of Figure 5. Two of the five that originally were not within the range of the delay model could now be categorized. Three lane groups improved by one LOS.

| | AFTER | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| BEFORE | A | B | C | D | E | F | * | TOTAL |
| A | 0 | | | | | | | 0 |
| B | | 8 | 1 | | | | | 9 |
| C | | 1 | 16 | | | | | 17 |
| D | | | 3 | 10 | 1 | | | 14 |
| E | | | 1 | 2 | 0 | | | 3 |
| F | | | | 2 | | 6 | | 8 |
| * | | | | 1 | 1 | 1 | 17 | 20 |
| TOTAL | 0 | 9 | 21 | 15 | 2 | 7 | 17 | 71 |

| | AFTER | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| BEFORE | A | B | C | D | E | F | * | TOTAL |
| A | 0 | | | | | | | 0 |
| B | | 8 | | | | | | 8 |
| C | | | 12 | | | | | 12 |
| D | | | | 5 | | | | 5 |
| E | | | | 3 | 4 | | | 7 |
| F | | | | 1 | | 0 | | 1 |
| * | | 1 | 1 | | | 1 | 5 | 8 |
| TOTAL | 0 | 9 | 13 | 9 | 4 | 1 | 5 | 41 |

**FIGURE 4 Changes in approach LOS due to shadowing: *top*, exclusive right-turn lanes; *bottom*, shared right-turn lane groups.**

| | AFTER | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| BEFORE | A | B | C | D | E | F | * | TOTAL |
| A | 0 | | | | | | | 0 |
| B | | 14 | | | | | | 14 |
| C | | 2 | 23 | | | | | 25 |
| D | | 1 | 4 | 9 | | | | 14 |
| E | | | 2 | 3 | 0 | | | 5 |
| F | | | 2 | 2 | 1 | 0 | | 5 |
| * | | | | 3 | 1 | 1 | 3 | 8 |
| TOTAL | 0 | 17 | 31 | 17 | 2 | 1 | 3 | 71 |

| | AFTER | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| BEFORE | A | B | C | D | E | F | * | TOTAL |
| A | 0 | | | | | | | 0 |
| B | | 11 | | | | | | 11 |
| C | | | 15 | | | | | 15 |
| D | | | | 4 | | | | 4 |
| E | | | | 2 | 3 | | | 5 |
| F | | | | | 1 | 0 | | 1 |
| * | | | 1 | | | 1 | 3 | 5 |
| TOTAL | 0 | 11 | 16 | 6 | 4 | 1 | 3 | 41 |

**FIGURE 5 Changes in lane group LOS due to shadowing: *top*, exclusive right-turn lanes; *bottom*, shared right-turn lane groups.**

### Capacity Increase from Stop Sign Analogies

The capacity increase from the HCM stop sign analogy ranged from 3 to 483 percent (with a mean of 113 percent) and from 22 to 875 vehicles. The capacity increase from the updated HCM stop sign analogy ranged from 4 to 561 percent (with a mean of 130 percent) and from 23 to 1,328 vehicles. No direct means to estimate the change in delay (and therefore LOS) caused by the capacity increase was apparent. The HCM delay equation is based on the assumption that vehicles depart from the intersection during their green phase. The stop sign analogies add capacity during the red phase. Whereas delay will be reduced by RTOR, the amount of the reduction cannot be modeled correctly by the HCM delay equation. Therefore, the discussion of results focuses on the change in v/c ratios.

Figure 6 shows the before and after v/c ratios for lane groups that originally had v/c ratios between 0 and 1.0. Data points are shown for the HCM stop sign analogy, the updated HCM stop sign analogy, and the shadowing procedure. However, the shadowing procedure results have been changed from volume reductions to capacity increases. Consider a right-turn lane group with a volume of 200 right turns, a capacity of 400 right turns on green, and a volume reduction from shadowing of 100 right turns during red. The v/c without shadowing would be 200/400, or 0.5. The v/c with the volume reduction from shadowing would be (200 − 100)/400, or 0.25. If the shadowing were interpreted as a capacity increase rather than a volume decrease, then the v/c would be 200/(400 + 100), or 0.40. The latter interpretation is used in Figure 6. Many of the changes in v/c from shadowing appear small, but almost all of the stop sign analogy changes appear fairly large.

The results are most dramatic for the 12 lane groups that originally had v/c ratios greater than unity (Figure 7). Three lane groups that originally had incredible v/c ratios of 2.8, 3.2, and 4.7 were reduced to ratios below 1.7 by all three applications. The shadowing procedure left six lane groups significantly above unity, while the stop sign analogies each left three significantly above unity.

### Critical v/c Ratios for Intersections

Five of the 80 original intersection analyses indicated that the critical v/c for the intersection was greater than unity. In such cases the present intersection and timing arrangement would be judged to be incapable of serving the demand. The analysis of RTOR, however, indicated that some of these intersection v/c values were too high. Table 1 presents the intersection v/c values before and after consideration of RTOR. In Case 1 the v/c was reduced moderately. In Cases 2 and 3 all three methods gave a result below or nearly below capacity. In Case 4 no change occurred because the right-turn lane group was not a critical movement. In Case 5 no shadowing protected left turn was present, so only the stop sign analogies led to indications of below-capacity operation.

## CONCLUSIONS

The results indicate that many right-turn lane groups and intersections will be deemed to be over capacity unless RTOR is considered explicitly. It is likely that in some of these cases, analysts would consider adding unneeded lanes, providing unnecessary green time, prohibiting left turns, or implementing other measures when demand could be handled without these actions.

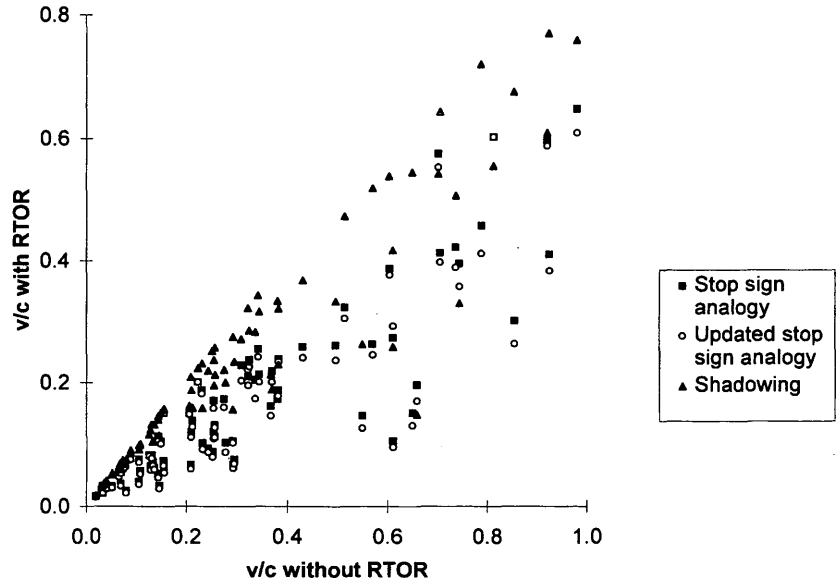



**FIGURE 6 RTOR effect on lane group v/c (v/c without RTOR < 1).**

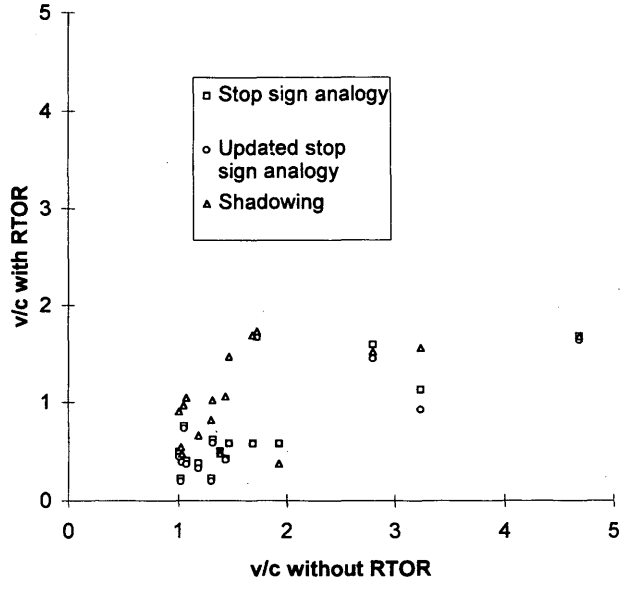



**FIGURE 7 RTOR effect on lane group v/c (v/c without RTOR > 1).**

Two rational approaches are available to consider RTOR. The shadowing procedure is conservative in its logic, simple to implement, and available with the new HCM signalized intersection procedure. The stop sign analogy is consistent with the HCM analysis of unsignalized intersections and yields higher RTOR capacity, but it would require a significant effort to be placed within a computer application of the HCM. Either approach would probably be better than assuming zero RTOR volumes. However, both methods need refinements for estimating delay and LOS.

The shadowing procedure eliminates 0 to 100 percent of right-turning vehicles from demand. The procedure can underestimate the RTOR volume by ignoring RTOR through a conflicting stream having significant unsaturated green time. On the other hand, the shadowing procedure can indicate that all right turns occur on red, which is unrealistic. In reality, the right turn on green and RTOR vehicles will experience some stopped delay, but less than that which would be determined by assuming no RTOR.

The stop sign analogy increases the right turn capacity. Since this extra capacity would be available during the red phase, the HCM delay equation is not directly applicable. The stop sign analogy also does not indicate how much of the demand will use RTOR. Since the unsaturated red time (unsaturated green of the conflicting flow) is not uniform throughout the red phase, estimating RTOR volume

**TABLE 1 Changes in Intersection Critical v/c Ratio**

| | Intersection Critical v/c Ratio | | | |
|---|---|---|---|---|
| CASE | Without RTOR | With Shadowing | With Stop Sign Analogy | With Updated Stop Sign Analogy |
| 1 | 1.68 | 1.55 | 1.55 | 1.55 |
| 2 | 1.31 | 1.00 | 0.97 | 0.97 |
| 3 | 1.31 | 0.99 | 0.96 | 0.96 |
| 4 | 1.18 | 1.18 | 1.18 | 1.18 |
| 5 | 1.01 | 1.01 | 0.86 | 0.86 |

and delay becomes a complicated task. However, it appears likely that a procedure could be created to estimate delay with RTOR.

A computer version of the updated stop sign analogy should be developed. Until this is available, an analyst should apply the shadowing procedure when a right-turn lane group, without RTOR, is found to be over capacity. If the shadowing procedure leaves a right-turn lane group significantly over capacity, then the analyst should manually apply the updated stop sign analogy (at least to the nonshadowed phases) to estimate the capacity situation, if there is unsaturated green time within the conflicting flows.

A procedure for estimating delay under the stop sign analogy should be developed and tested. HCM users familiar with the treatment of protected plus permitted left turns can appreciate the complexity of a procedure to estimate RTOR flows. However, this effort could lead to a more realistic description of intersection performance. In many cases it will yield more efficient use of green time and avoid the construction of unneeded lanes.

## REFERENCES

1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985, Chapters 9 and 10.
2. Virkler, M. R., and C. C. Chen. Potential Accuracy of a Planning Application for the HCM Signalized Intersection Operational Procedure. In *Transportation Research Record 1365*, TRB National Research Council, Washington, D.C., 1992, pp. 40–53.
3. *Special Report 209: Highway Capacity Manual*, 3rd ed. TRB, National Research Council, Washington, D.C., 1994, Chapter 9.
4. Luh, J. Z., and Y. J. Lu. Capacity Computations of Right-Turn-on-Red Using HCM. *ITE Journal,* April 1990, pp. 34–39.
5. *Special Report 209: Highway Capacity Manual*, 3rd ed. TRB, National Research Council, Washington, D.C., 1994, Chapter 10.
6. *Highway Capacity Software*, Version 2.3. McTrans Center, University of Florida, Gainesville, 1990.

*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Methodology for Assessing Dynamics of Freeway Traffic Flow

MICHAEL J. CASSIDY AND JOHN R. WINDOVER

A methodology for the detailed evaluation of measured freeway traffic stream features is presented. The method compares cumulative vehicle arrival curves at multiple locations, and empirical data illustrate traffic flow dynamics. However, conclusions with regard to traffic flow features will not be drawn until ongoing research is completed.

The paper presents a methodology for performing a detailed assessment of features of measured freeway traffic stream. Application of the proposed methodology identifies how disturbances propagate in time and space. Empirical data are used to present examples of traffic flow features revealed by the proposed method and to illustrate the methodology's advantages over conventional techniques for evaluating freeway data. The paper is methodological in nature; the authors, therefore, defer drawing conclusions on traffic flow dynamics until the ongoing research is completed.

## BACKGROUND

Traffic flow on any freeway system cannot exceed the capacity of its most severe restriction (i.e., bottleneck). Thus, bottlenecks often characterize freeway operating conditions (*1*, p.288). The measurement and assessment of bottleneck flow has been the subject of much research.

Past studies of bottleneck operation often have relied on observations measured at a single location along the freeway. Such observations might have included measured values of flow, $q$; speed, $v$; and density, $k$ (or occupancy) from which $q$-$k$ or $q$-$v$ scatterplots were constructed (*2–7*). More recent work has sought to assess capacity flow by measuring vehicle arrival rates at locations presumed to be downstream of restrictions and comparing these rates before and after the observed onset of queueing (*8–11*).

The evaluation of operating states measured at a single location is, for lack of a better term, myopic. Restricting assessments of traffic stream behavior to a single location obscures flow dynamics occurring over space and time. A number of studies have constructed $q$-$k$ or $q$-$v$ scatterplots for multiple locations along a freeway and compared the relative features of these fundamental relations in an effort to assess the influence of physical location (*12–15*). Likewise, past work has examined changing traffic patterns in response to time-variant conditions by constructing plots of $q(t)$, $k(t)$ versus time $t$ (*12,15–17*). Yet these techniques do not identify explicitly the propagation of changing flow states in the traffic stream. As such, bottleneck flow dynamics may have yet to be identified in a definitive manner.

Institute of Transportation Studies, University of California, 109 McLaughlin Hall, Berkeley, Calif. 94720.

## PROPOSED METHOD

The method described herein is based on the work by Newell (*18*), who used assumptions about wave motion to predict the features of cumulative vehicle arrival curves. Analogously, the authors use the observed features of cumulative arrival curves to identify the motion of changing traffic states.

The cumulative vehicle arrival curve plots cumulative arrival number to time $t$ (*19–22*). In Figure 1, the value $j$ on the vertical axis is the cumulative number of vehicle arrivals to the given location by time $t_j$. Analogously, $t_j$ is the time that the $j$th vehicle arrives at the location. In constructing cumulative curves, the authors plot smooth, differentiable interpolations through the stepwise function illustrated in Figure 1. The derivative (i.e., slope) of this interpolation is flow.

The cumulative arrival curve is a visual representation of observations collected directly from the highway. The measure *flow*, on the other hand, requires specification of a time interval, and the interval selected can influence the magnitude of flow. Moreover, cumulative curves do not model relationships, as is often the intent of $q$-$k$ and $q$-$v$ scatterplots.

The methodology herein uses cumulative curves constructed in series. The input-output diagram in Figure 2 shows cumulative curves measured at two locations along the highway. Curve $A(x_o, t)$, the cumulative vehicle arrivals past upstream location $x_o$ to time $t$, is constructed from the same collection of vehicles used for $A(x, t)$, the cumulative curve at downstream location $x$. That is, an upstream observer records (and cumulatively graphs) the arrival times of vehicles as they pass $x_o$. The times at which these same vehicles pass $x$ are also recorded (and plotted). The vertical distance between curves at some time, say $t_1$ for example, is the number of vehicles in section $x$-$x_o$ at $t_1$. In the absence of vehicle overtaking maneuvers, the horizontal distance between curves at height $j$, for example, is $j$'s trip time, $tt_j$, from $x_o$ to $x$.

The input-output diagram is an effective tool for tracing the motion of disturbances in time and space. As an example, the "thick" portions of the arrival curves in Figure 3 depict a short-term fluctuation in arrival rate. The fluctuation on arrival curve $A(x_o, t)$ is passed horizontally to downstream curve $A(x, t)$, indicating that this fluctuation propagates forward among the same collection of vehicles. The phenomenon of changing flow states moving forward with vehicles has been observed consistently in this study.

The input-output diagram in Figure 3 can be transformed into a queueing diagram by translating upstream curve $A(x_o, t)$ horizontally to the right by a distance equal to the average free-flow trip time from $x_o$ to $x$. Translated curve $A(x_o, t)$ is a "desired" arrival curve mapping what would be vehicle arrival times to downstream location $x$ in the absence of delay. Where a desired arrival curve is superimposed on the downstream arrival curve, traffic is flowing

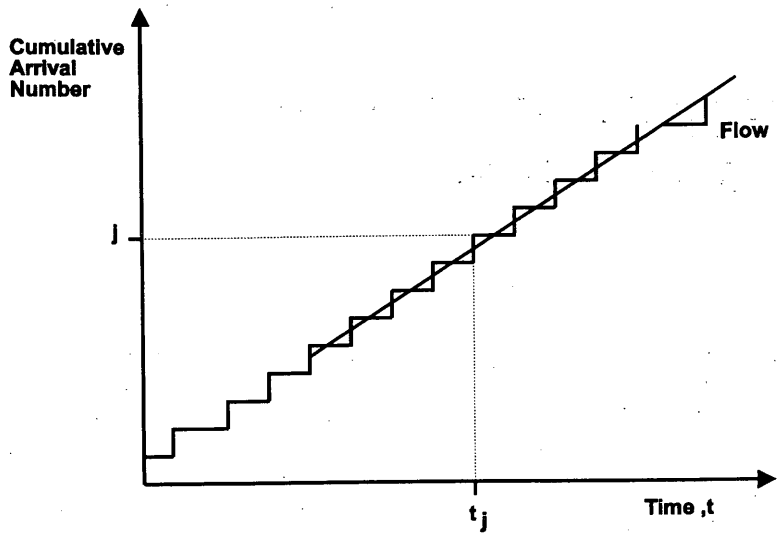



**FIGURE 1 Cumulative vehicle arrival curve.**

without delay (i.e., desired and actual arrival times to the downstream location are equal). In the presence of delay, displacement will exist between the desired and the downstream arrival curves. The horizontal distances between these displaced curves at unit heights define vehicle delays.

A system of moving time coordinates (*18*) is used to describe conveniently the process of translating upstream curve $A(x_o, t)$ by the appropriate free-flow trip time. In the moving time coordinate system, time advances forward over space at a pace equal to average free-flow trip time. "Moving" time, $t'$, at any downstream location, $x$, lags behind "actual" time, $t$, by the free-flow trip time from an upstream reference point, $x_o$. That is,

$$t' = t - u_f(x - x_o)$$

where $u_f$ is the average free-flow trip time per unit distance.

The use of moving time facilitates the presentation of desired and downstream arrival curves with a single time axis, as in a queueing diagram. Free-flow vehicles exhibit zero trip time to downstream locations (i.e., curves are superimposed), whereas displacements between curves reveal added trip times (i.e., delays).

Figure 4 presents the motion of a forward-moving wave propagating at a rate slower than prevailing vehicle speed, as described by *Kinematic Wave Theory* (*23*) for moderately heavy, uncongested flow conditions. As vehicles advance downstream faster than the wave, the disturbance past location $x_o$ manifests itself at downstream location $x$ among a collection of vehicles of higher arrival number. A horizontal translation of upstream curve $A(x_o, t)$ will not result in the superposition of both curves in Figure 4.

Figure 5 illustrates the motion of backward-moving waves. At time $t_1$, Curve $A(x, t)$ exhibits a dramatic discontinuity in flow states created by a sudden flow reduction past point $x$. This flow reduction

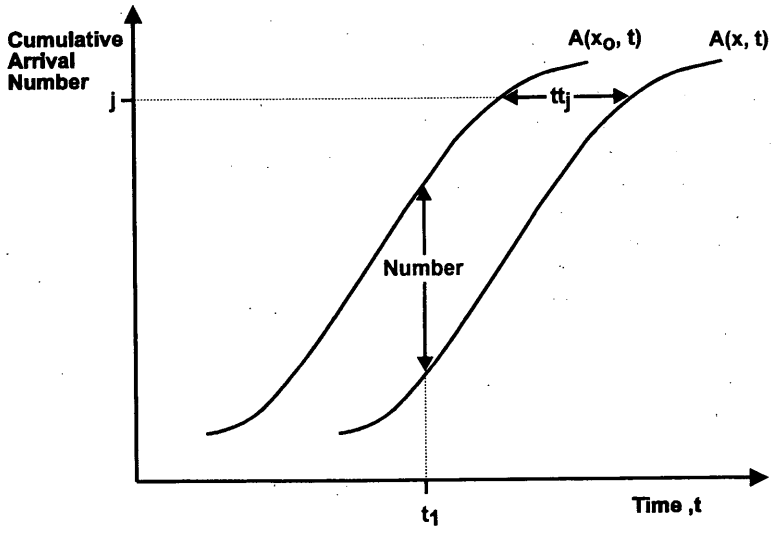



**FIGURE 2 Input-output diagram.**

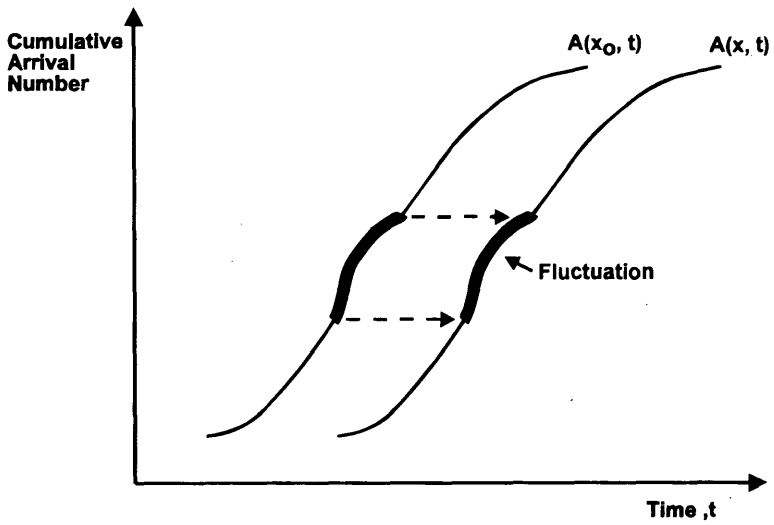



**FIGURE 3 Disturbance traveling with vehicles.**

might be caused by a downstream incident or a surge in on-ramp flow downstream. The resulting discontinuity in flow, called a shock wave, propagates in the upstream direction. Upon the shock wave's arrival at $x_o$, Curve $A(x_o, t)$ presumably conforms to the shape of $A(x, t)$. Figure 5 indicates that the shock's trip time from $x$ to $x_o$ is given by $w$ and that the number of vehicles traveling through the shock during this trip is $\Delta A$.

Later, at time $t_2$, a disturbance is created by a rise in flow through the downstream bottleneck (e.g., the incident is partially cleared or downstream on-ramp flow slightly diminishes). The backward motion of the resulting wave in Figure 5 describes, according to *Kinematic Wave Theory*, how disturbances propagate in congestion.

This methodology is used to study the evolution of traffic flow. Upstream curves are translated horizontally, as in a queueing diagram, thereby employing a system of moving time coordinates. Subsequently, the features of traffic disturbances are evaluated by comparing the attributes of the cumulative curves in series.

The arrival curves presented in Figures 3 through 5 idealized as empirical count data seldom reveal changing flow states in a pronounced or obvious manner. Thus, the methodology incorporates a simple but important graphical "trick": cumulative counts used for arrival curves are reduced uniformly by a "background" flow. A fixed number is cumulatively subtracted from the vehicle counts in each count interval. The reduction is applied sequentially to counts at each observation location starting with intervals that correspond to the same moving time at all locations.

The process used for background flow reduction can be visualized using Figure 6, which already displays a horizontally translated $A(x_o, t)$. Once in moving time, a fixed reduction is cumulatively applied to both curves simultaneously (e.g., starting from $t' = 0$).

Using this technique, the vertical distances between consecutive curves are preserved following the background flow reduction. Similarly, a background flow reduction will not alter the occurrence times of flow changes on the cumulative curve, which is the feature

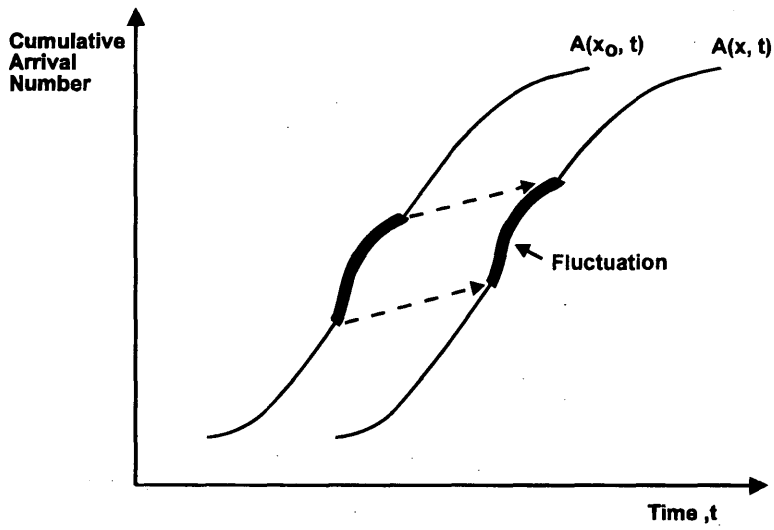



**FIGURE 4 Wave propagating forward slower than vehicle speed.**

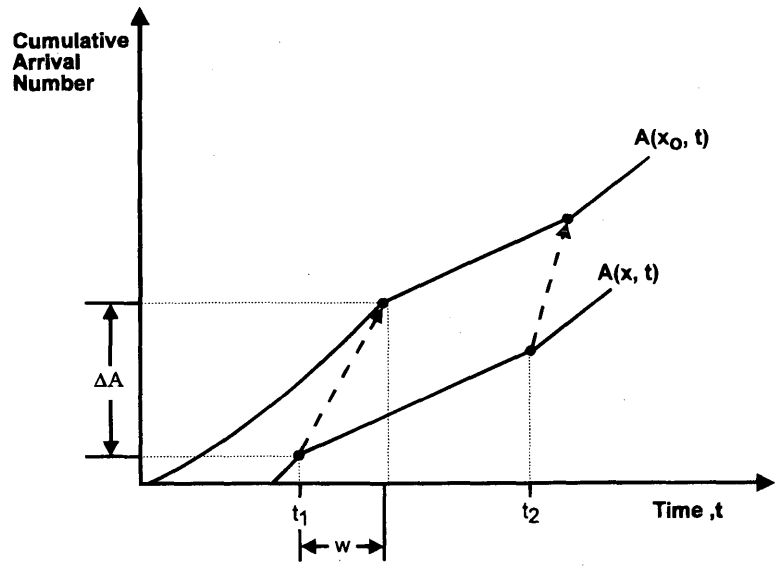



**FIGURE 5 Backward-moving waves.**

of interest in this analysis. Note, however, that where displacements between curves prevail because of delay (as in Figure 6), a background flow reduction enlarges the horizontal distances between consecutive curves. The number "reassigned" to the *j*th vehicle on upstream curve $A(x_o, t)$ is different from the number reassigned to vehicle *j* on downstream curve $A(x, t)$, and horizontal distances no longer equal delays.

A background flow reduction amplifies flow changes on the cumulative curve. Following a background flow reduction of sufficient magnitude, changing flow states can be identified visually on the cumulative curve. The resulting curve may exhibit negative slopes denoting prevailing flows less than the specified background flow.

## EMPIRICAL DATA

The methodology is applied using data measured on a section of the Queen Elizabeth Way near Toronto, Canada. Data collected during multiple weekdays in March 1994 were generously provided by personnel at the Ontario Ministry of Transportation.

The study site is illustrated in Figure 7. Mainline and on-ramp demands at the Cawthra Road junction create recurring congestion during the morning commute. Detector stations for measuring traffic stream data are located throughout the system and have been labeled in Figure 7 according to the numbering strategy adopted by the Ontario Ministry of Transportation.

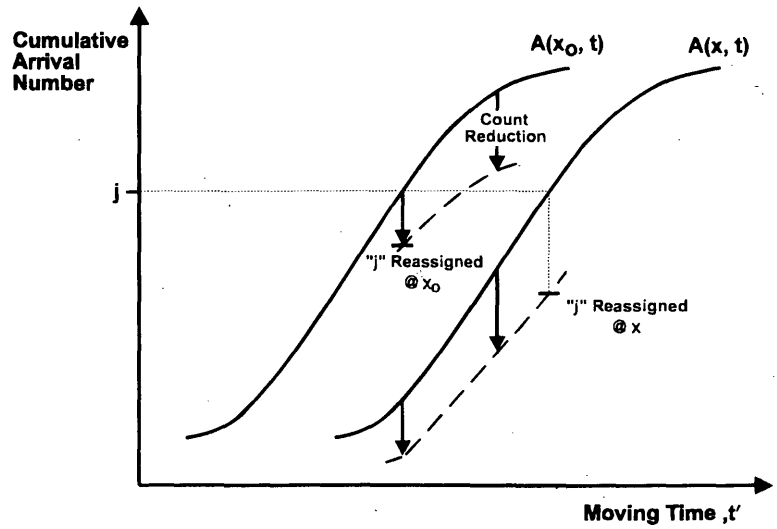



**FIGURE 6 Background flow reduction.**

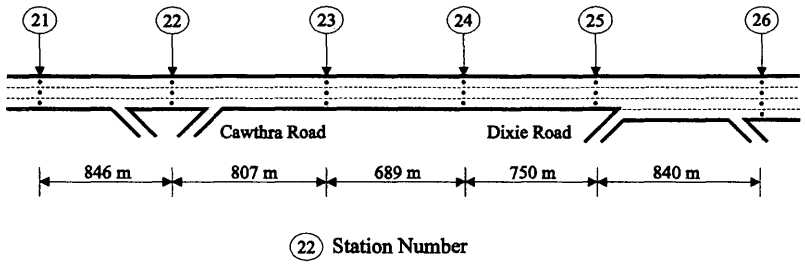



**FIGURE 7 Queen Elizabeth Way, Ontario, Canada.**

Data measured at Stations 23, 24, and 25 are presented Detectors at these locations record counts, average speeds, and occupancies over 30-sec intervals. Average free-flow trip time between these stations is, likewise, 30 sec. Data presented herein are aggregated across all travel lanes.

Observations are presented from only Stations 23 through 25 purely in the interest of brevity. The methodology can be applied even in the absence of conservation. Consider, for example detector Stations 25 and 26. These consecutive stations exist upstream and downstream of ramp junctions. Where ramp counts are not available, cumulative curves at Stations 25 and 26 would not be superimposed. One could, however, readily evaluate the motion of disturbances by comparing the relative changes in slope of these consecutive curves.

## FORWARD-MOVING FLUCTUATIONS IN ARRIVAL RATE

Our initial example presents the motion of forward-moving disturbances as revealed by the proposed method. In this example, the method is applied to a 25-min period on a single observation day (labeled "Day 1" in Figures 8 through 10). Figure 8 illustrates cumulative arrival curves constructed from traffic counts at Stations 23, 24 and 25. These arrival curves illustrate a key issue: the "eye" does not readily identify subtle changes in a function's slope. Examining how disturbances propagate is almost impossible because changing flow states are not apparent from the curves in Figure 8.

Figure 9, on the other hand, presents the cumulative curves previously shown in Figure 8 following (*a*) the use of moving time by translating curves at Stations 23 and 24 by the respective free-flow trip times to Station 25, and (*b*) a background flow reduction applied to all three curves. Cumulative curves at Stations 24 and 25 were reduced uniformly by a rate of 4,300 vehicles per hour (vph). A slightly higher background flow reduction of 4,436 vph was applied to the curve at Station 23 as the detectors at this station were found, on this day, to be overcounting vehicles at a rate of 136 vph. Having applied background flow reductions, changing vehicle arrival rates are now displayed prominently as "wiggles" on the curves in Figure 9.

In Figure 9 a sudden flow reduction (manifest as a near-zero average slope) occurs at approximately $t' = 6{:}43$ a.m. and prevails for approximately 10 min. The general superimposition of curves denotes an absence of delay between Stations 23 and 25. Thus, the observed flow reduction initially occurs upstream and the resulting disturbance propagates forward past the observation locations. If the 10-min flow reduction is the consequence of an upstream incident (a plausible explanation), then including data collected during this 10-min interval could corrupt certain experiments, such as maximum flow measurements to estimate capacity.

Conventional wisdom suggests that the 10-min flow reduction in Figure 9 can be identified by constructing time-series plots of flow at each detector station. There are potential shortcomings with this approach, however. Vehicle count is a random variable exhibiting a variance-to-mean ratio comparable with 1. Figure 10 shows the time-series plots of flow computed from 30-sec vehicle counts at

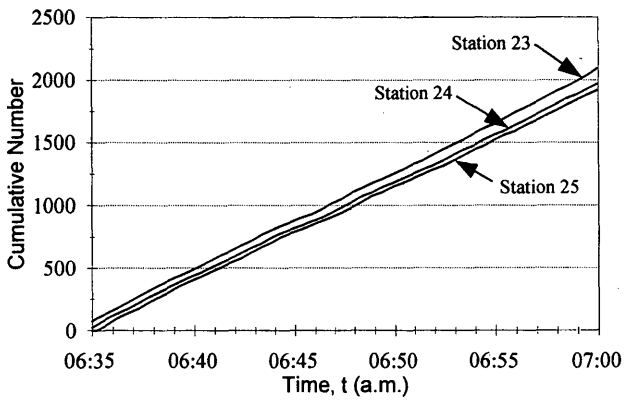



**FIGURE 8 Arrival curves constructed from "raw" data, Day 1.**

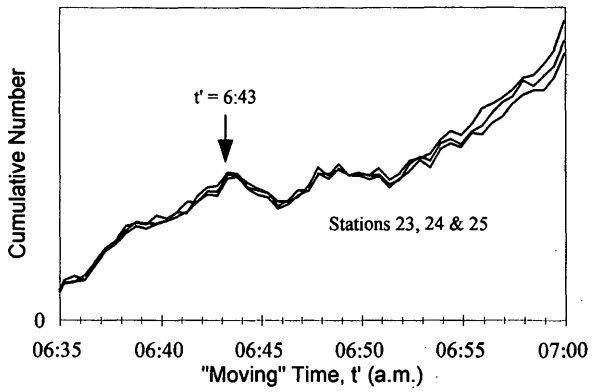



**FIGURE 9 Arrival curves in moving time with background flow removed, Day 1.**

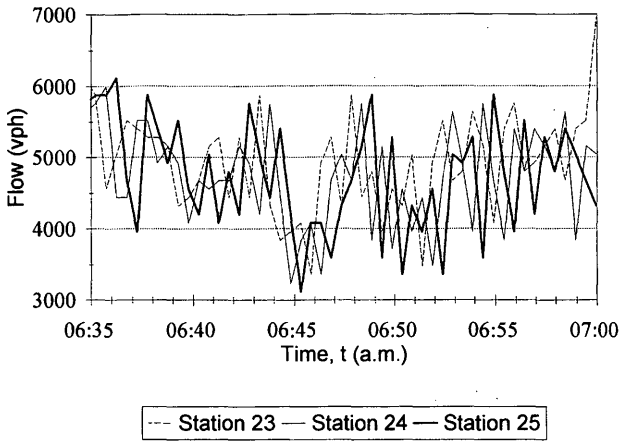



**FIGURE 10 Time-series plot of 30-sec flows, Day 1.**

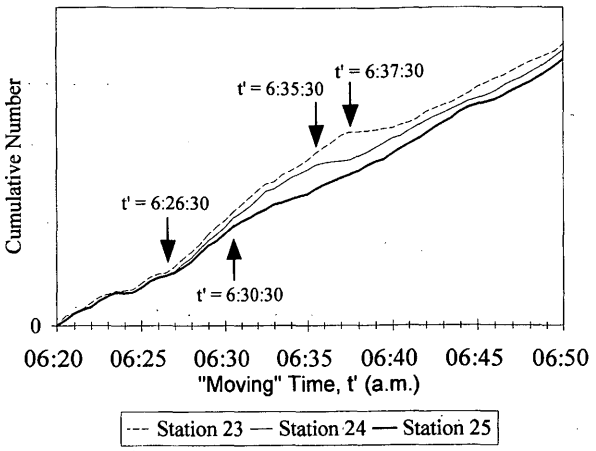



**FIGURE 11 Backward-moving shock wave, Day 2.**

each station during the 25-min period of interest. Given the observed variability, the 10-min drop in flow is not readily apparent in Figure 10.

One can reduce the dispersion in Figure 10 by computing flow on the basis of vehicle counts taken over longer time intervals. The problem here, however, is that resulting flow values are average magnitudes occurring during each specified interval, not the actual rates prevailing at any given time. This approach to "smoothing" flow obscures the details of traffic dynamics.

The advantage of the cumulative curve is its representation of detail. Referring to Figure 9, short-term changes in vehicle arrival rate (i.e., wiggles) replicate across cumulative curves. The ability to superimpose wiggles using a horizontal translation denotes that changing arrival rates are propagating among the same collection of vehicles. That is, disturbances travel forward with vehicles. These disturbances are not waves that propagate across vehicles creating velocity changes in the traffic stream. Instead, the wiggles in Figure 9 are arrival rate fluctuations created by the varying headways chosen by different motorists. The replication of wiggles across cumulative curves reveals that motorists "remember" and maintain their respective headways while traversing the freeway segment. This flow feature is not predicted by any conventional continuum model of freeway flow (*23–25*).

## BACKWARD-MOVING WAVES

The forward-moving fluctuations in vehicle arrival rate described previously are not the only type of disturbance that can occur in the traffic stream. Other disturbances, such as shock waves, will create velocity changes by propagating across vehicles. Using data from a different observation day (Day 2 in Figures 11 and 12), detailed features of shock wave propagation can be demonstrated.

Figure 11 presents cumulative arrival curves in moving time with background flow reductions of 4,300 vph for Stations 24 and 25 and a slightly larger background flow reduction of 4,450 vph for Station 23, as detectors were again overcounting vehicles at this upstream station. Starting at $t' =$ 6:26:30 a.m., cumulative curves in Figure 11 exhibit their maximum flows (i.e., slopes). At approximately $t' =$ 6:30:30 a.m., the curve at Station 25 begins to diverge in a pronounced manner from the others, depicting added vehicle delay between Stations 24 and 25. Substantial flow reductions occur at upstream Stations 24 and 23 at $t' =$ 6:35:30 a.m. and at $t' =$ 6:37:30 a.m., respectively. These sudden flow reductions, which occur sequentially in time and space, depict the motion of a backward-moving shock wave. These dramatic, short-term "collapses" in flow (manifest as near-zero slopes) appear to reflect initial motorist tendency to overreact to the shock's arrival.

From *Kinematic Wave Theory*, one would expect that as the shock arrives at each station, the respective cumulative curve would take on the slope of its downstream counterpart. After the shock arrived at Stations 24 and 23, the curves would exhibit fixed displacements denoting delay and the presence of additional vehicles between detector stations.

The occurrence of the flow collapse, however, creates conditions that are different than expected. Namely, a flow collapse "starves" the downstream freeway section as seen in the bulges displayed by the cumulative curves in Figure 11. Note that the additional vehicles accumulated during the shock's propagation are depleted and the cumulative curves tend to reconverge. This unstable flow behavior might obscure the bottleneck's location. Figure 12 presents measured speed profiles at Stations 23, 24, and 25 during an extended

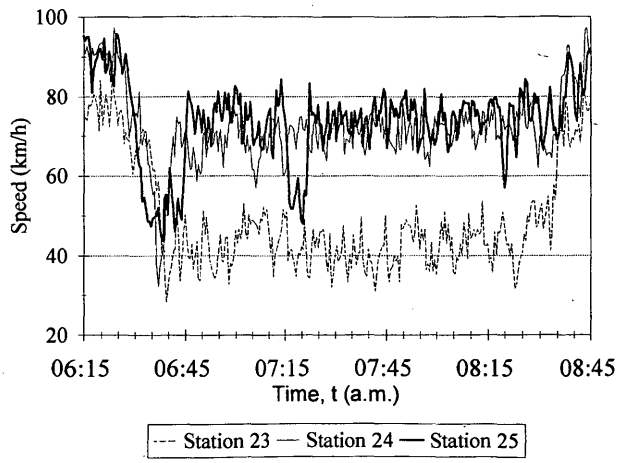



**FIGURE 12 Measured speed profiles, Day 2.**

period of Day 2. During the shock's propagation just after 6:30 a.m., all three detector stations exhibit speed reductions associated with congestion. Eventually, however, speeds at downstream Stations 24 and 25 recover to higher levels while speeds at Station 23 remain substantially lower. By observing this prolonged difference between upstream and downstream speeds, one might identify a bottleneck in close proximity to the Cawthra Road on-ramp junction while overlooking the initial bottleneck well downstream, as shown in Figure 11.

## CONCLUSIONS

The paper has presented a methodology for the dynamic assessment of freeway traffic flow. The method facilitates identification of the details of flow features. The objective of this paper has not been to draw conclusions or conjectures with regard to freeway traffic stream dynamics. Instead, the authors have described an assessment methodology. For demonstration, the methodology has been applied to assess freeway operation on two days. Some of the flow features identified by the method and presented herein are not yet completely understood. Further research to investigate these dynamics is ongoing. In the future, the authors' intent is to demonstrate that the proposed method is a valuable tool for assessing bottleneck capacity, speed-flow-density relationships, and highway traffic flow dynamics in general.

## ACKNOWLEDGMENTS

The authors are indebted to many people. They thank Gordon Newell and Carlos Daganzo, University of California, for their advice throughout the study. The authors also thank David Tsui and Mark Fox, Ontario Ministry of Transportation, for supplying the data used herein, and they likewise thank James Banks, California State University; Kirk Barnes and Tom Urbanik, Texas Transportation Institute; and Fred Hall, McMaster University, for their assistance in acquiring data.

## REFERENCES

1. May, A. D. *Traffic Flow Fundamentals.* Prentice Hall, Englewood Cliffs, N.J., 1990.
2. Greenshields, B. D. A Study of Traffic Capacity. *HRB Proc.*, Vol. 14, 1935, pp. 448–477.
3. Drake, J. S., J. L. Schofer, and A. D. May. A Statistical Analysis of Speed Density Hypothesis. In *Highway Research Record 154*, HRB, National Research Council, Washington, D.C., 1967, pp. 53–87.
4. Ceder, A., and A. D. May. Further Evaluation of Single- and Two-Regime Traffic Flow Models. In *Transportation Research Record 567*, TRB, National Research Council, Washington, D.C., 1976, pp. 1–15.
5. McDermott, J. M. Freeway Surveillance and Control in Chicago Area. *Journal of Transportation Engineering*, ASCE, No. 106. 1980, pp. 333–348.
6. Koshi, M., M. Iwasaki, and I. Ohkura. Some Findings and an Overview of Vehicular Flow Characteristics. *Proc., 8th International Symposium on Transportation and Traffic Theory*, Elsevier, Amsterdam, the Netherlands, 1983, pp. 403–426.
7. Payne, H. Discontinuity in Equilibrium Freeway Traffic Flow. In *Transportation Research Record 971*, TRB, National Research Council, Washington, D.C., 1984, pp. 140–146.
8. Hurdle, V. F., and P. K. Datta. Speeds and Flows on an Urban Freeway: Some Measurements and a Hypothesis. In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp. 127–137.
9. Banks, J. H. Flow Processes at a Freeway Bottleneck. In *Transportation Research Record 1287*, TRB, National Research Council, Washington, D.C., 1990, pp. 20–28.
10. Agyemang-Duah, K., and F. L. Hall. Some Issues Regarding the Numerical Value of Capacity. *Proc., International Symposium of Highway Capacity*, Karlsruhe, Germany, 1991, pp. 1–15.
11. Ringert, J., and T. Urbanik. *An Evaluation of Freeway Capacity in Texas.* Report FHWA/TX-92/1196-2F. Texas Transportation Institute, College Station, 1992.
12. Ceder, A. *Investigation of Two-Regime Traffic Flow Models at the Micro- and Macroscopic Levels.* Ph.D. thesis. University of Calif., Berkeley, 1975.
13. Persaud, B. N., and V. Hurdle. Some New Data That Challenge Some Old Ideas About Speed-Flow Relationships. *Transportation Research Record 1194.* pp 191–198.
14. Banks, J. H. Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations. In *Transportation Research Record 1225*, TRB, National Research Council, Washington, D.C., 1989, pp. 53–60.
15. Hall, F. L. and M. A. Gunter. Further Analysis of the Flow-Concentration Relationship. In *Transportation Research Record 1091*, TRB, National Research Council, Washington, D.C., 1986, pp. 1–9.
16. Hall, F. L., B. L. Allen, and M. A. Gunter. Empirical Analysis of Freeway Flow-Density Relationships. *Transportation Research*, Vol. 20A, 1986, pp. 197–210.
17. Hall, F. L., and L. M. Hall. Capacity and Speed Flow Analysis of the QEW in Ontario. In *Transportation Research Record 1287*, TRB, National Research Council, Washington, D.C., 1990, pp. 108–118.
18. Newell, G. F. A Simplified Theory of Kinematic Waves in Highway Traffic, Part I: General Theory. *Transportation Research*, Vol. 27B, 1993, pp. 281–287.
19. Edie, L. C., and R. S. Foote. Effect of Shock Waves on Tunnel Traffic Flow. *HRB Proc.*, Vol. 39, 1960.
20. Makagami, Y., G. F. Newell, and R. Rothery. Three-Dimensional Representation of Traffic Flow. *Transportation Science*, Vol. 5, No. 3, 1971, pp. 303–313.
21. Newell, G. F. *Applications of Queueing Theory*, 2nd ed. Chapman & Hall, London, 1982.
22. Persaud, B. N. Study of a Freeway Bottleneck to Explore Some Unresolved Traffic Flow Issues. Ph.D. dissertation. Department of Civil Engineering, University of Toronto, Canada, 1986.
23. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads. *Proc., Royal Society*, Vol. A229, 1955, pp. 317–345.
24. Payne, H. FREFLO: A Macroscopic Simulation Model of Freeway Traffic. In *Transportation Research Record 722*, TRB, National Research Council, Washington, D.C., pp, 68–77.
25. Michalopoulos, P. G., Y. Ping, and A. Lyrintzis. Continuum Modeling of Traffic Dynamics for Congested Freeways. *Transportation Research B*, Vol. 27B, No. 4, 1993, pp. 289–303.

*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Probabilistic Nature of Breakdown at Freeway Merge Junctions

LILY ELEFTERIADOU, ROGER P. ROESS, AND WILLIAM R. MCSHANE

Observation of field data collected as part of NCHRP Project 3-37 showed that at ramp merge junctions, breakdown may occur at flows lower than the maximum observed, or capacity, flows. Furthermore, it was observed that at the same site and for the same ramp and freeway flows, breakdown may or may not occur. After visual examination of traffic operations at sites where breakdown occurred, it was observed that immediately before breakdown, large ramp-vehicle clusters entered the freeway stream and disrupted traffic operations. It was concluded that breakdown is a probabilistic rather than deterministic event and is a function of ramp-vehicle cluster occurrence. Subsequently, a probabilistic model for describing the process of breakdown at ramp-freeway junctions was examined. The model gives the probability that breakdown will occur at given ramp and freeway flows and is based on ramp-vehicle cluster occurrence. Simulation of a data collection effort was conducted to establish the data requirements for model validation. It was concluded that the amount of data available was not adequate for precise validation of the probabilistic model.

Various mathematical models have been used to describe the relationships among flow, speed, and density on freeways for any given instance. Such models provide the basis for selecting measures of effectiveness and defining level-of-service ranges (*1*). They are also used for estimating capacity and the operating conditions under which capacity is reached. Doing so requires the identification of the maximum volume point on a speed-flow or flow-density curve, a process often questionable because of the vague range of data generally observed before breakdown on most facilities. Because of these inconsistencies in the data, the process of flow breakdown in merge areas has not been documented adequately. The mechanism through which the operation switches from stable to unstable flow has not been modeled effectively.

Existing models predict breakdown as a deterministic function of a given flow rate, or speed, or density. These models generally assume that breakdown will occur at capacity flows only, and therefore capacity can be measured in the field immediately before breakdown. However, it is clear neither what triggers the breakdown, nor when and how the facility will eventually break down. The development of a quantitative model describing the process of flow breakdown around entrance ramps will be very useful for the operational analysis of ramp-freeway junctions. It can also help in establishing freeway control strategies to maximize flow and optimize operations on the freeway.

The probabilistic aspect of ramp merge breakdown was developed through examination and analysis of traffic data at ramp merge junctions. It was observed that breakdown was not the direct result of peak volumes. Data from NCHRP-Project 3-37, Capacity of Ramp-Freeway Junctions (*2*), showed that breakdown often occurs at flows lower than observed "capacity." One of the most interesting observations was that at a site where data were collected during two peak morning periods, breakdown occurred only once while volumes on the freeway and the ramp remained at the same levels. Closer examination of videotapes revealed that when a large cluster of vehicles entered the freeway from the ramp, queues were created on the ramp or on the freeway (or on both). Furthermore, the higher the number of ramp vehicles entering the freeway in platoons, the bigger the impact on freeway operations. On some occasions this series of vehicles caused a shift of freeway traffic to the left lanes, as they tried to avoid the turbulence and conflicts in the merge area.

The unpredictability of breakdown during data collection for NCHRP Project 3-37 led to this attempt to describe breakdown as a probabilistic function. The model that has been developed is based on the occurrence and size of on-ramp vehicle clusters instead of on ramp volume, as is done in common practice. The probability of breakdown is estimated as a function of the clusters on the ramp, which are, however, directly related to the ramp volume. The freeway flow and the respective gaps available on each freeway lane, as well as drivers' actions as they approach the merge area, were considered in developing the model. Some implications in data collection of the existence of a probabilistic model are subsequently examined, and the data requirements for validating the model are calculated.

## OBSERVATIONS ON CAPACITY, BREAKDOWN, AND SPEED-FLOW RELATIONSHIPS

Capacity is defined in the 1985 *Highway Capacity Manual* (HCM) (*1*) as the maximum flow rate that can reasonably be expected to pass through a section of a roadway under prevailing roadway and traffic conditions. According to the HCM,

> at capacity there are no usable gaps in the traffic stream, and any perturbation from vehicles entering or leaving the facility, or from internal lane changing maneuvers, creates a disturbance that cannot be effectively damped or dissipated. Thus operation at, or near capacity is difficult to maintain for long periods of time, and the switch from stable to unstable flow occurs rapidly.

This definition of capacity implies that breakdown occurs immediately after capacity has been reached and is a direct consequence of high traffic volumes. The field data for this study, however, show that capacity and breakdown are not necessarily interconnected. Clearly, the way in which the stable and unstable flow branches are joined, and the operational nature of transitions between the

L. Elefteriadou, Civil and Environmental Engineering Department, Pennsylvania State University, 212 Sackett Building, University Park, Pa. 16802. R. P. Roess, Polytechnic University, Six Metrotech Center, Brooklyn, N.Y. 11201. W. R. McShane, Polytechnic University, 901 Route 110, Farmingdale, N.Y. 11735.

branches, must be investigated further. If capacity does not always occur immediately before breakdown, the shape of the speed-flow curve will appear to be discontinuous around capacity. The data presented here demonstrate that breakdown may occur at flows lower than capacity.

## Field Observation of Breakdown Conditions

When studying capacity and breakdown issues at ramp-freeway merge junctions, the site selection is critical. The section downstream of the merge should be free of constraints. A downstream bottleneck would cause spillback of queues into the merge section and create the impression that the speed-flow and flow-density relationships are discontinuous near capacity. The merge area must regularly experience breakdown conditions as a direct result of the ramp volume, not because of geometric design deficiencies. Traffic operations are dependent on the specifics of the location, so that the sites selected must be as similar as possible in terms of geometrics and general environment. Taking into account geometric factors would unnecessarily complicate the study of breakdown. The field data were collected using video cameras at various locations along the freeway in the vicinity of the ramp. It was possible to observe directly the number of ramp vehicles approaching in clusters, as well as traffic operations in general. A detailed description of the data collection effort for this study can be found elsewhere (*3*).

Two sites were selected from the NCHRP Project 3-37 data base. At the first one, Site 28, data were collected during one afternoon peak period; at the second, Site 21, data were collected during two morning peak periods (Sites 21 and 59, respectively). Both sites are on six-lane freeways and involve a single-lane on-ramp. The acceleration lane at both junctions is of the parallel type. Both sites are also the middle ramps in an off-on-off sequence of ramps along a freeway.

## Data Analysis

The shape of the curves around capacity depends heavily on the time intervals over which the traffic variables are averaged. If the transition period is smaller than the time intervals used, the process of averaging will create some false data points, especially if data from many days are used (*4*). Persaud observed that there is no gradual drop in speeds and flows. He showed that speed-flow observations during breakdown indicate a false pattern because they are averaged over conditions with and without a queue present, which results in a gradual decrease of speeds and flows. A similar study at a metered location undertaken by Allen et al. (*5*), who investigated transitions in speed-flow relationships, resulted in graphs implying a continuous relationship, but the data used came from detectors, so there is no information about the queue. The sudden drop in speeds that Persaud advocates is also in accordance with other researchers' observations (*6,7*) that breakdown is initiated by a slower-moving vehicle. It is also in accordance with the probabilistic model of breakdown, since a ramp vehicle cluster forces lower speeds on the freeway as soon as it occurs. Since one slow-moving vehicle makes the following vehicles reduce their speeds, the result is an overall sudden speed drop. Therefore, for this study, field data were analyzed in 1-min intervals.

Observation of videotapes on a second-by-second basis helped identify transitions and their possible causes, as well as traffic conditions in general. Transitions can best be observed with plots of traffic variables for each period of observation. For these plots the variables of interest are plotted versus the time period to examine how they change during the period of data collection. There appear to be clear advantages to looking at daily traces of traffic behavior rather than relying on scatter diagrams of many days of accumulated data (*5*). The first advantage is that from the daily plots one can obtain some idea of actual behavior of the variables, as well as their relationship in time, which the scatter diagram cannot provide. The second advantage is that inspection of the daily plots along with the videotapes permits one to identify points that represent transition between congested and uncongested flow. Visual examination also helps identify points of localized congestion on the freeway or on the ramp, not easily distinguished otherwise.

### *Site 28*

Site 28 is located on I-290 southbound, at the intersection with Biesterfield Road, in Chicago, Illinois. Data were collected for 100 min at five locations along the freeway. Starting with the 79th min of data collection, there is a temporary disruption of traffic for 5 to 6 min, during which the flow is unstable, and then the facility returns to stable operation. During this disruption, queues form occasionally in Lanes 1 (shoulder lane) and 3 (median lane) along the merge section. The fact that no queues appeared at the last downstream camera confirmed that breakdown was not caused by a downstream bottleneck. Queues form first along the acceleration lane area, and then the disruption spreads upstream. Visual examination showed that during the 79th to 81st min, clusters of vehicles entering from the ramp disrupted traffic at this section and caused temporary breakdown.

Speed-flow time-connected diagrams for Site 28 were constructed for each camera location and lane. Figure 1 shows the speed-flow graph for a location close to the end of the acceleration lane. It can be observed that a relatively flat section representing stable conditions and speed fluctuating between 80 and 97 km/hr and a section where flow decreases somewhat with decreasing speed. After the end of the acceleration lane, on the other hand, speed only drops to 68 km/hr, with a small effect on flow (Figure 2). The numbers on the data points represent the period of observation.

Contrary to common belief, the speed drop shown in Figure 1 does not correspond to the highest observed flows. At this site the breakdown does not occur at capacity flows, when 1-min analysis intervals are used. As shown in Figure 2, flow at this location is stable at all times. There is, however, a slight speed drop (68 km/h) starting with Period 79, approximately 20 sec before the breakdown is observed at the upstream location. No queues are observed at this location, and there is no other noticeable change in operations other than the slight speed drop. This subject is investigated in greater detail in the following section, in conjunction with the observation of clusters at this site.

A general examination of Figures 1 and 2 shows that flows exceed conventional capacity estimates by far, which may be expected since these numbers represent 1-min flows. Freeway flow after the end of the acceleration lane reaches 8,500 passenger cars per hour (pcph). Most of the higher flows occur before the breakdown, but there are several intervals of high flow after the breakdown. The single-lane flows are very high, approaching 3,000 pcph per lane (pcphpl) in some cases. Finally, Lane 3 flow is consistently the highest among the three lanes.

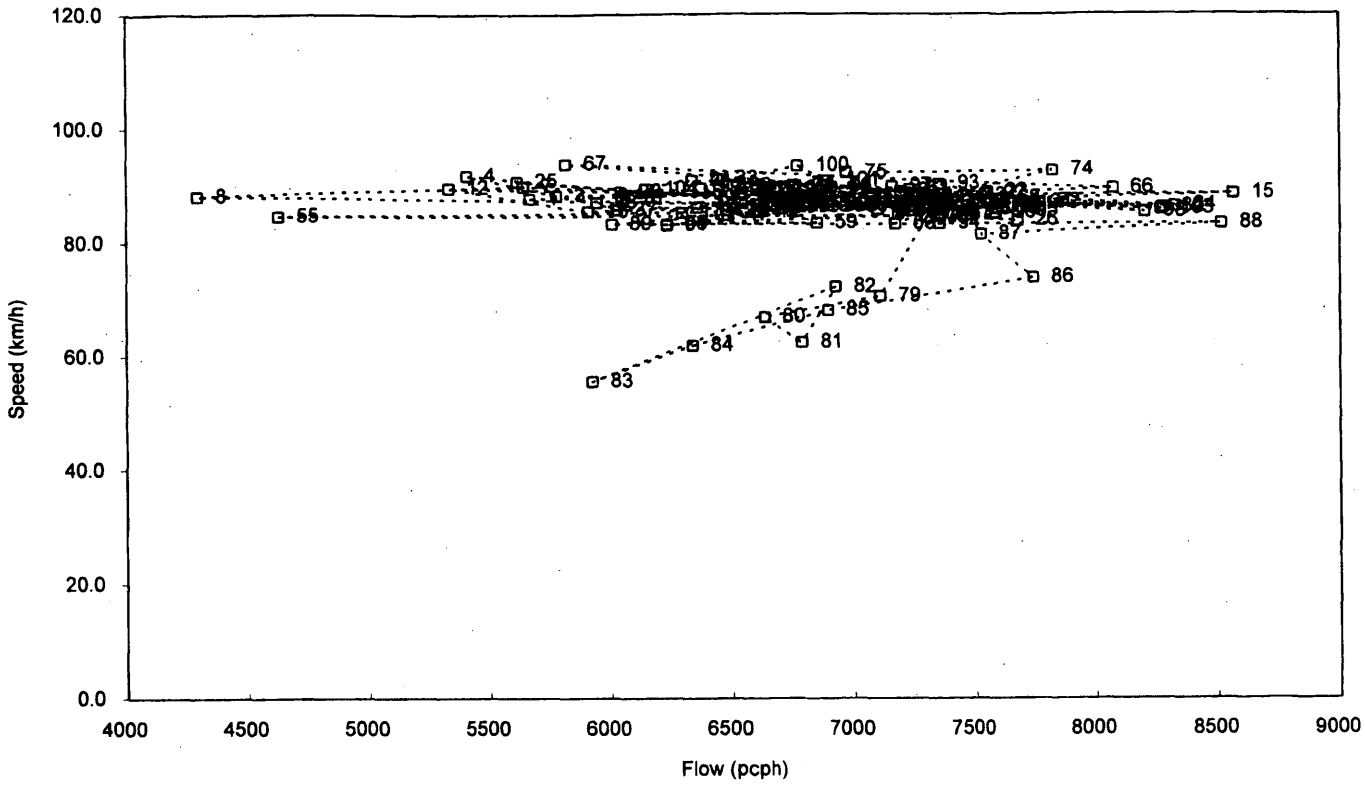



**FIGURE 1 Flow versus speed at acceleration lane section, Site 28.**

Speed-flow time-connected graphs with 5-min analysis intervals, especially the ones at the downstream end of the acceleration lane area, resemble much closer the conventional curves. However, two intervals preceding the breakdown have higher flows than the breakdown interval. It is also interesting that the flows downstream of the merge exceed the conventional 6,000-pcph capacity throughout the data collection period, including both stable and unstable operations. At the time of breakdown the flow at the last downstream camera is near 7,400 pcph.

*Site 21*

Site 21 is in Orlando, Florida, at the junction of I-4 eastbound and Princeton Street. Data were collected for approximately 1 1/2 hr at three locations along the freeway. At the beginning of data collection, the flow at the facility is stable; it switches to unstable after 1 hr and remains so until the end of data collection. Observation of the videotapes and the speed data revealed that the first location where speed drops at this site is Lane 3, the leftmost lane, downstream of the end of the acceleration lane. Subsequently, the other lanes are affected, and the breakdown spreads to the sections upstream. At this site queues were also observed downstream of the merge section, but after the formation of queues upstream. However, the last downstream camera is only 61 m from the end of the acceleration lane, whereas Site 28 is 153 m downstream. Therefore, the last downstream camera for Site 21 is closer to the merge, and the operations at this section are much more affected by the turbulence of the merge. It is speculated that at the section farther downstream from the merge area, the speed dropped only slightly, as at Site 28. It is possible that congestion and unstable flow spread downstream, with speeds increasing farther away from the merge.

As at Site 28, the transition from stable to unstable flow does not occur during the interval with the highest flow. The drop in speeds starts during Period 57, during the same interval that a cluster of 12 vehicles enters the freeway. The speed starts dropping when the freeway flow is approximately 7,500 pcph, while flows had reached 8,500 pcph before breakdown.

As expected, curves for 5-min intervals are much smoother than the ones for smaller analysis intervals. Again, the speed drop caused by the breakdown does not correspond to the peak volume.

*Site 59*

At Site 59 there is no breakdown during the data collection period, even though this is a site that breaks down regularly. This is the same ramp merge location as Site 21 at which breakdown was observed on a similar weekday morning peak period. Several data variables for this site were compared with the previous two sites to identify similarities and differences. The comparisons between Sites 21 and 59 were for the same real-time periods. It was concluded that the flows at Site 59 are as high as, or even higher than, they are at Site 21. The fact that Site 21 breaks down even with lower flows supports the hypothesis that the breakdown is not a deterministic function of freeway flow. The 1-min flow at the last downstream camera at Site 59 reached 9,500 vehicles per hour (vph), whereas the respective number for Site 21 is 8,800 vph. This clearly demonstrates that breakdown does not necessarily occur at

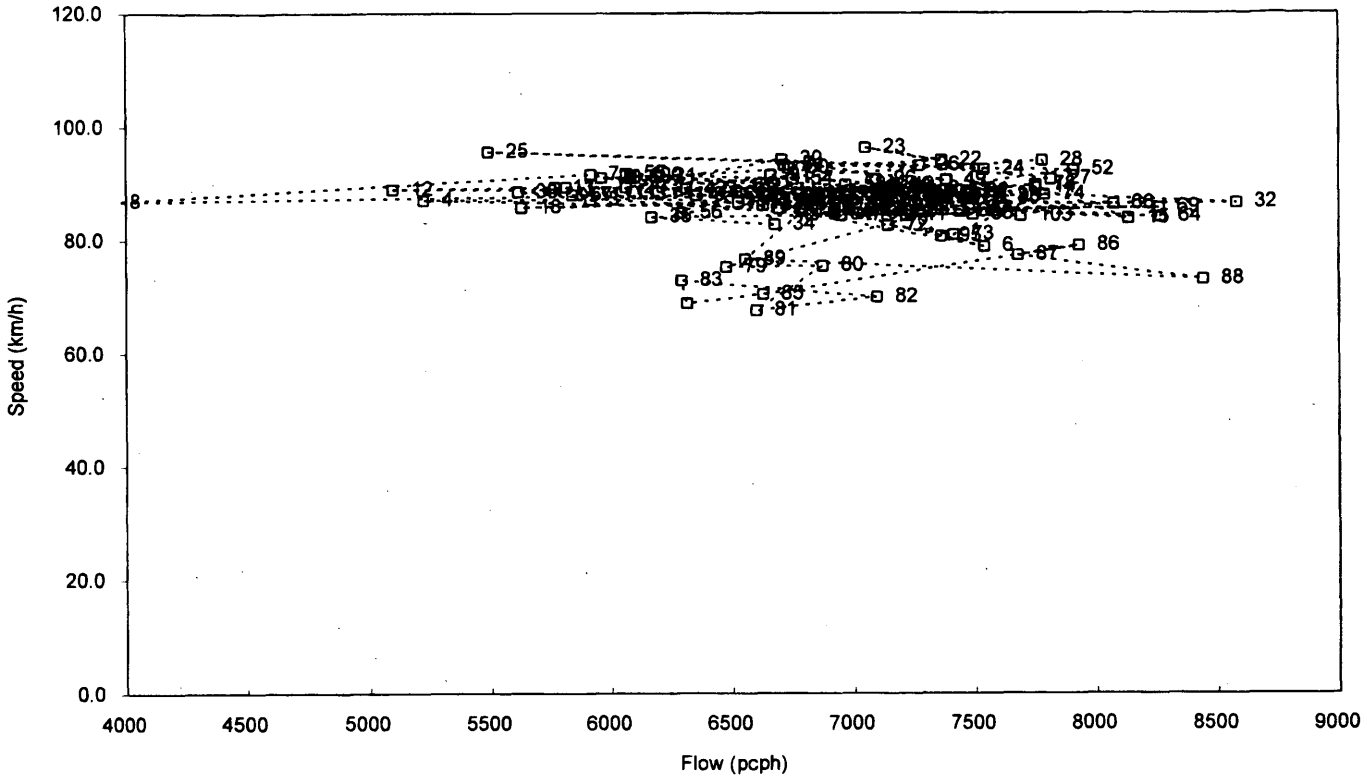



**FIGURE 2 Flow versus speed after end of acceleration lane, Site 28.**

capacity, therefore, previous assumptions that capacity can be measured immediately before breakdown are not true.

## Field Observation of On-Ramp Vehicle Clusters

Observation of traffic operations at the gore/merge area before the breakdown led to the conclusion that when a large number of vehicles in clusters enters the freeway stream, it disrupts traffic and may result in breakdown. Therefore, the clusters were observed at the three sites to document their sizes and establish their relationship to breakdown.

A cluster is defined here as three or more vehicles traveling together so that their headway does not exceed 3 sec, or a spacing of 54 m at a speed of approximately 64 km/h. The vehicles in a cluster entering the freeway from the ramp were counted at each site. The ramp vehicles were observed from Camera 2 as they were approaching the gore area from the ramp. The time that the first vehicle in a cluster crossed the gore was recorded, along with the number of vehicles in the cluster.

### *Site 28*

Clusters at Site 28 were observed starting at real time 3:13 p.m. The size of clusters ranges between 3 and 11 vehicles, up until the 78th min (period) of data collection. During the 78th period the cluster entering the freeway includes 15 vehicles, the largest cluster so far, and breakdown occurs during the 79th period. During the 78th interval the ramp flow is 1,320 pcph and the freeway flow is 6,449 pcph. Note that earlier, during Period 68, the freeway flow was higher (7,080 pcph) and the ramp flow was the same as for Period 78, but there was no breakdown. The only difference in operations was the cluster size, which during the 68th period never exceeded the six vehicles. There appears to be a 1-min interval during which the high-flow, high-concentration cluster must travel downstream to the beginning of the bottleneck, at which point the speed drops. Then the low-speed wave travels upstream, resulting in the temporary disruption of traffic.

As noted, the speed drops somewhat at the section downstream from the end of the acceleration lane, starting approximately 20 sec before the breakdown. Observation of videotapes verified that at the camera located after the end of the acceleration lane, 20 to 30 sec into the 79th interval there is a transition from high-speed low flow to lower-speed higher flow and density. Therefore, at this site, the sequence of events before breakdown is (*a*) entrance of a large cluster to the freeway stream, (*b*) subsequent small speed drop at the beginning of the bottleneck, which (*c*) spreads upstream and creates the breakdown.

At the freeway section along the acceleration lane, the speed starts dropping 40 sec into the interval, with the most notable drop in the shoulder lane during the last 15 sec of the interval. The speed drop at this interval is not justified by the corresponding freeway flows, since there were intervals preceding the breakdown during which flows were higher. Again, the data reinforce the notion of breakdown as a probabilistic variable. After the breakdown, during the 81st period, there is a new maximum of 16 vehicles in a ramp cluster. During this time the ramp flow is 1,020 pcph and the freeway flow is only 5,521 pcph. Operations recover after the 85th period, with the ramp flow dropping dramatically (180 pcph) dur-

ing several intervals. The size of the ramp clusters does not exceed the 11 vehicles thereafter.

*Site 21*

Clusters were observed at Site 21, but only until breakdown occurred. Afterward, the speed drops dramatically and a queue is created on the ramp, precluding observation of cluster sizes. In general, at this site, the distances between cars were longer. It was found that the largest cluster was 12 vehicles, and it entered the freeway during the same interval that speed started dropping at the section after the end of the acceleration lane. As at Site 28, as soon as a large cluster of ramp vehicles enters the freeway, the speed starts dropping at the beginning of the bottleneck. After that, it is a matter of time for the wave to reach the merge area and for breakdown to occur. Just as at Site 28, the ramp and freeway flows were not the highest during the time of breakdown. At Site 28, though, operations became stable after 5 or 6 min, because of the low ramp volume at the time immediately following the breakdown. At Site 21, after the breakdown, there is a queue created on the ramp that does not dissipate after the first conflict. This demand on the ramp is not reflected in the ramp flows, which represent discharge flows. The location of the cameras, unfortunately, did not allow for measuring queue length on the ramp.

*Site 59*

Clusters were observed at Site 59 to compare operations with Sites 28 and 21 and to determine whether there is a different cluster pattern that may be crucial to the breakdown. It was determined that where the largest cluster at Site 21 was 12 vehicles, the largest cluster at Site 59 was 10 vehicles. The difference is small, and it illustrates the point that breakdown is not deterministic—that is, large clusters have a high probability to cause breakdown, but again it is not imperative that they do. Cluster size is a very important factor in causing breakdown, but breakdown is not a deterministic event.

### Breakdown As a Probabilistic Event

The data presented here clearly demonstrate that breakdown does not necessarily occur immediately after capacity is reached, and high flows do not necessarily result in breakdown. During the data collection for NCHRP Project 3-37, Capacity of Ramp Freeway Junctions (*2*), several sites recommended by various agencies as frequently experiencing breakdown operated under free-flow conditions during the videotaping. This phenomenon supports the hypothesis that breakdown is a probabilistic event. Taking into consideration the probabilistic nature of merge operations, it can be explained why capacity has not been identified and why breakdown, as a direct result of the merge, is so difficult to observe.

The cluster size of on-ramp vehicles plays a very important role in breakdown. At two of the three sites, large clusters were entering the freeway stream immediately before breakdown occurrence. At the third site breakdown did not occur even though the volumes on both the freeway and the ramp were as high. It was concluded that even a large cluster does not guarantee breakdown; instead, the probability of breakdown increases with increasing cluster sizes. In the following section, a probabilistic model of breakdown (*3*) is examined, and the implications of the existence of a probabilistic model are investigated.

## PROBABILISTIC MODEL OF BREAKDOWN

For the purposes of this paper, breakdown is defined as a significant reduction in speed for vehicles traveling on the mainline (i.e., 16 km/h). If vehicles on the freeway must reduce their speeds because of ramp vehicles, a traffic wave is created that propagates upstream and triggers breakdown. A traffic wave can be described as a cluster of vehicles traveling at similar speeds along a highway. In general, traffic flow is not homogeneous, and traffic waves can influence operations at any point or section of a highway.

The development of the model was based on the observation that a large cluster of vehicles entering the freeway from the ramp triggered breakdown, while at the same site, comparable freeway and ramp flows on a similar weekday peak hour had no effect on operations. Even though ramp clusters certainly affect operations, breakdown is not a deterministic function of cluster occurrence. Therefore, the probability of breakdown was computed as a function of the cluster size, which is dependent on the ramp flow rate, and the freeway flow. The objective was to develop a model that gives the probability that breakdown will occur, based on the cluster sizes, and as a function of given ramp and freeway flow rates. A brief description of the model is given in the following section; a detailed analysis can be found elsewhere (*3*).

## MODEL DEVELOPMENT

The objective in developing this model was to calculate the probability of a disruption created from the ramp clusters at the ramp merge area of the freeway. The model development process is shown in Figure 3. First, the cluster effect is taken into consideration by calculating the probability of occurrence for all possible cluster sizes; clusters of 3 to 15 vehicles were considered. Then the presence of vehicles in the most critical lane of the freeway, the shoulder lane, is entered into the model with the calculation of the probability that at least one vehicle is present at the critical area of the freeway. The effect of drivers' possible actions is estimated by assigning probabilities to different actions that a driver in the critical area of the freeway may take as the cluster of vehicles approaches the freeway from the ramp. These three factors are taken into account in calculating the probability of breakdown, given that a cluster of a specific size has occurred. It is assumed that breakdown occurs if at least one vehicle on the freeway is forced to reduce its speed by 16 km/h or more. Finally, the probability of breakdown in a 15-min period is calculated, considering the expected number of clusters in the specified interval, as a function of the ramp and freeway flow.

Figure 4 shows the probability of breakdown in 15 min versus ramp flow, for freeway flows ranging from 1,400 to 2,200 vphpl over three freeway lanes. Only clusters of 3 to 15 vehicles were considered in calculating the probability of breakdown. As expected, Figure 4 shows that the probability of breakdown increases with increasing freeway flow, and effect that becomes more pronounced for high ramp flows. For ramp flows of 0 to 600

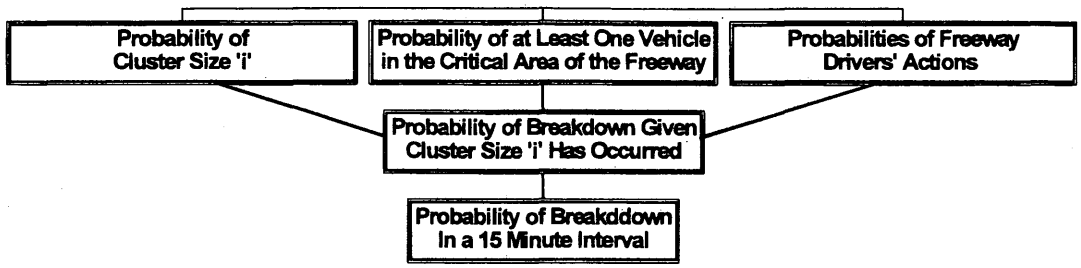



**FIGURE 3 Model development flow chart.**

vph, the probability of breakdown increases slowly and the effect of freeway flow is minimal. For ramp flows of 600 to 1,200 vph, the increase is sharper, especially for the higher freeway flows. The figure shows what is expected: the probability of breakdown increases with increasing ramp and freeway flows. In light of this, it is understandable why it was thought that high flows are the single cause of breakdown. Even though high ramp and freeway flows increase the probability of breakdown, they are not the direct cause of it.

In Figure 4 the probability of breakdown does not increase much beyond 0.70 for ramp flows approaching 1,500 vph, even for near-capacity freeway flows. This explains why there are few instances in which one can observe breakdown that is a direct result of the merging maneuvers. In most cases, there is a downstream constriction that creates congestion and breakdown, which eventually spreads upstream into the merge area.

## Data Implications of Probabilistic Model

The existence of a probabilistic model at ramp-freeway merge junctions implies new considerations for the data collection process. The data requirements for validating the probabilistic model must be examined. In light of the probabilistic nature of breakdown, for future studies examining breakdown, the data collection requirements will change as a function of the expected probability that breakdown will occur. Calculating the sample size required for observing any aspect of breakdown operations must take into account the probabilistic model. These aspects of data collection at ramp-freeway merge junctions are examined here.

The required sample size for estimating the probability of breakdown at a single point, with given ramp and freeway flows, is determined first. The sample size $N$ needed to estimate the true percentage within $h$ is given by the equation

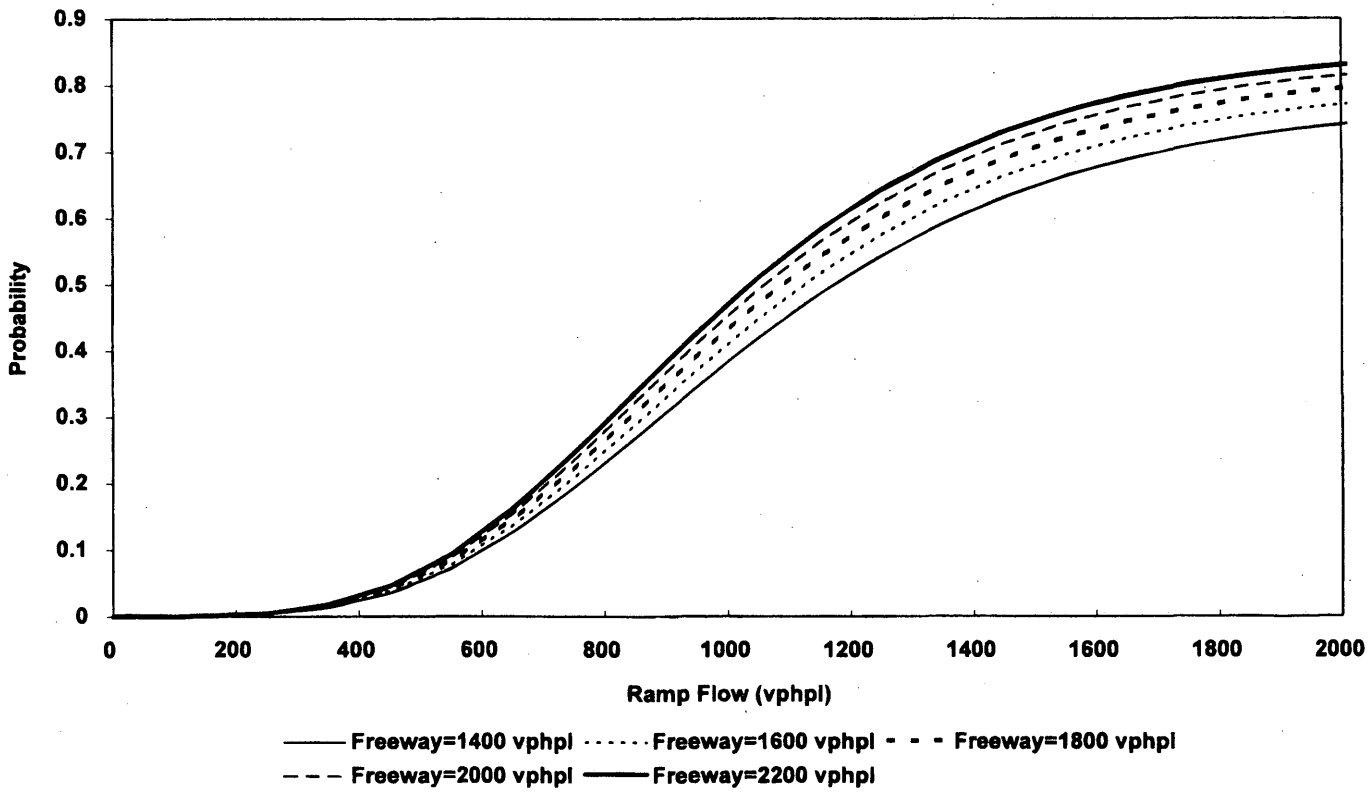



**FIGURE 4 Probability of breakdown in 15 min.**

$$N = \frac{z^2 p(1-p)}{h^2}$$

where

$p$ = true percentage,
$z$ = standard normal distribution parameter, and
$h$ = precision requirement, or amount of deviation from true value allowed.

From Figure 4, for a ramp volume of 1,100 vph and a freeway volume of 1,800 vphpl, the probability of breakdown is 50 percent. The normal approximation to the binomial distribution can be used here, since the percentage $p$ is not close to $p$ or 1. For a confidence level of 95 percent and a precision requirement of 5 percent, the required sample size is

$$N = \frac{1.96^2\, 0.5(1-0.5)}{0.05^2} \approx 384$$

It is noted that these samples are required to validate one point of the probabilistic curve, corresponding to a particular value of ramp flow and a particular value of freeway flow. Each of these data points represents a 15-min interval. From these calculations, it is obvious that the data requirements for validating the probabilistic model are enormous.

In the preceding example, the confidence intervals for a sample size of 10 are

$$h = z\sqrt{\frac{p(1-p)}{N}} = 0.310 = 31.0 \text{ percent}$$

They do not provide the desired precision, given that they cover more than half of the range of probabilities.

### Demonstration of Data Requirements Using Simulation

After the data requirements for validating each point of the probabilistic model are examined, the corresponding requirements for a linear regression equation fitted to the model are studied. The Far fewer data points are needed for calibrating a linear relationship than are needed to calibrate each point of the $x$-axis. Assuming that the model and its equation are true, simulated field data are generated and compared with the probabilistic model.

First, regression was performed to derive the equation of the line describing the probabilistic model of Figure 4 for a freeway flow of 1,800 vph per lane (vphpl). This equation is almost a straight line between the ranges of 300 and 1,500 vph. For this range, and for freeway flow 1,800 vphpl, the probability of breakdown is approximated by the linear equation

$$y = -0.29723 + 0.000653\, x$$

Then, using Monte Carlo simulation, it can be shown what the typical data scatter will be in estimating the probability of breakdown from field data. It is assumed that a typical data base will contain 50 data points, each of them representing whether breakdown occurred at a particular ramp flow (yes or no). It is also assumed that these points are divided equally among five ramp flow rates; 300, 600, 900, 1,200 and 1,500 vph.

For the simulation, a random number generator is used for predicting whether breakdown occurs or not. The boundaries established in the probabilistic model are used to distinguish between breakdown and free-flow conditions for each volume. It is assumed that the freeway volume is constant, at 1,800 vphpl. Twenty such experiments were conducted to illustrate the possible outcomes of such an experiment; the resulting probability curves are shown in Figure 5. As shown, a sample of 50 data points can produce drastically different results. Each of the 20 experiments, if seen alone, can give a totally different picture than the others: some of them result in lower probabilities of breakdown for higher flows, others show exactly the opposite. It is clear from the simulation experiment that limited amounts of data will give misleading results.

Regression was used to derive the linear equations that describe each of these experiments, so that they can be compared with the original probabilistic model. For each experiment, the two linear equation coefficients $b_0$ and $b_1$ are calculated as follows:

$$b_1 = \frac{\sum x_i y_i - nXY}{\sum x_i^2 - nX^2}$$

$$b_0 = Y - b_1 X$$

where

$x_i$ = ramp flow,
$y_i$ = probability of breakdown, and
$X, Y$ = respective averages.

Subsequently, the regression confidence bounds are calculated by calculating the confidence bounds of the parameters $b_0$ and $b_1$. The confidence interval for $b_1$ is

$$\beta_1 = b_1 \pm z s_{b_1}$$

where $z$ is the normal distribution parameter and the standard deviation for $b_1$ is calculated as

$$s_{b_1}^2 = \frac{s^2}{N s_x^2}$$

The confidence interval for $b_0$ is

$$\beta_0 = b_0 \pm z\, s_{b_0}$$

The standard deviation of $b_0$ is estimated as

$$s_{b_0}^2 = \frac{s^2}{n}\left(1 + \frac{x^2}{s_x^2}\right)$$

The confidence intervals for these parameters are too wide for the accuracy needed in the model. The constant $b_0$ of the probabilistic model is $-0.29723$, whereas the ranges calculated vary from $-0.30$ to 0.10. The $b_1$ coefficient of the probabilistic model is 0.000653, with ranges between $-0.00027$ and 0.00080. Figure 6 shows the frequency distribution for $b_0$ and Figure 7 shows the frequency distribution for $b_1$. These figures show the variability of the coefficients and demonstrate again the inadequacy of a limited data base to validate the probabilistic model.

As noted, establishing the true probability of breakdown at ramp-freeway merge junctions is required for estimating the sample size needed for observing breakdown. The probability of breakdown must be known before the data requirements can be estimated. For example, if the probability of breakdown is 50 percent, then a sam-

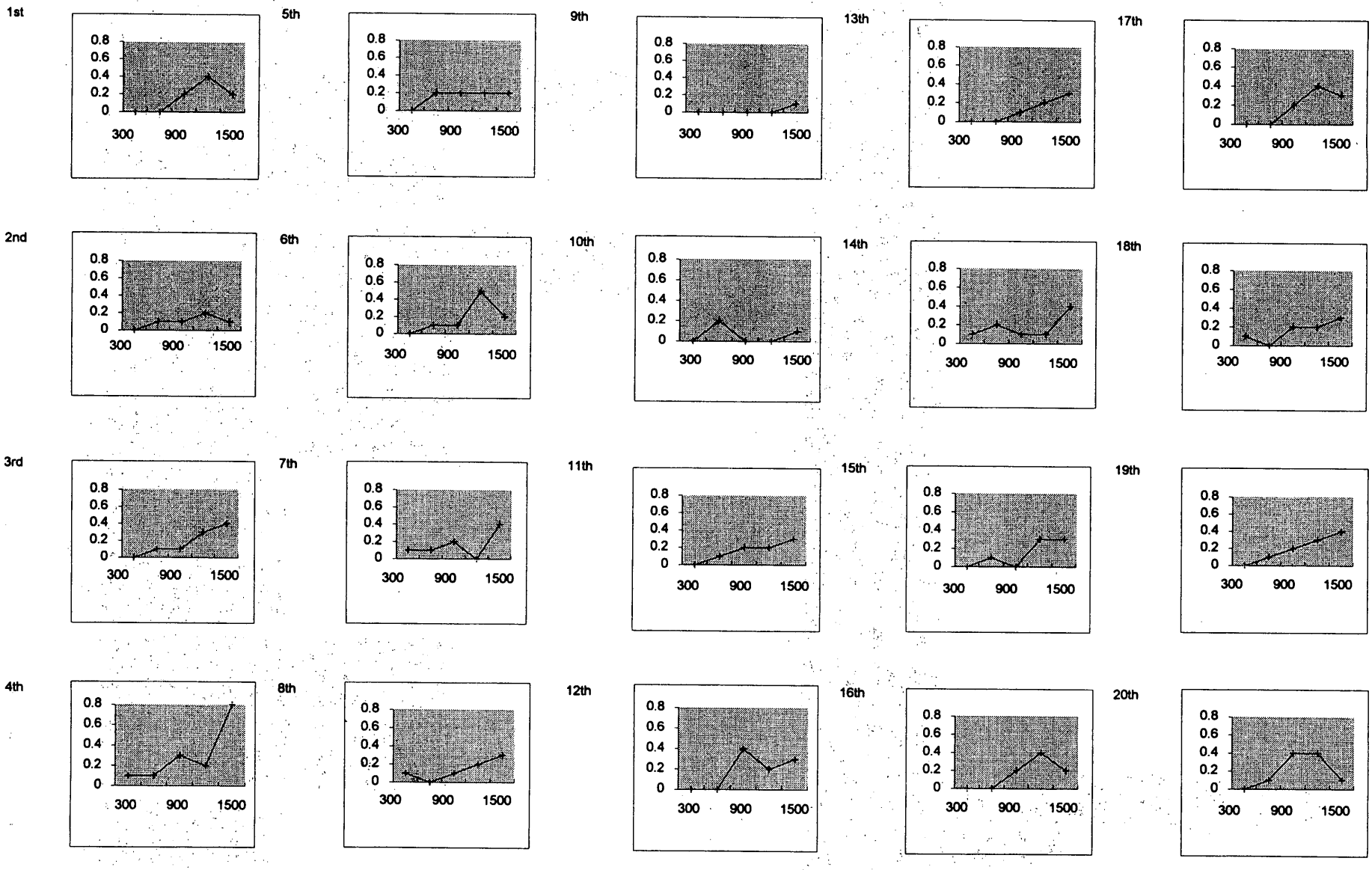



**FIGURE 5** Simulation results (probability is shown on *y*-axis; flow rate in vph on *x*-axis).

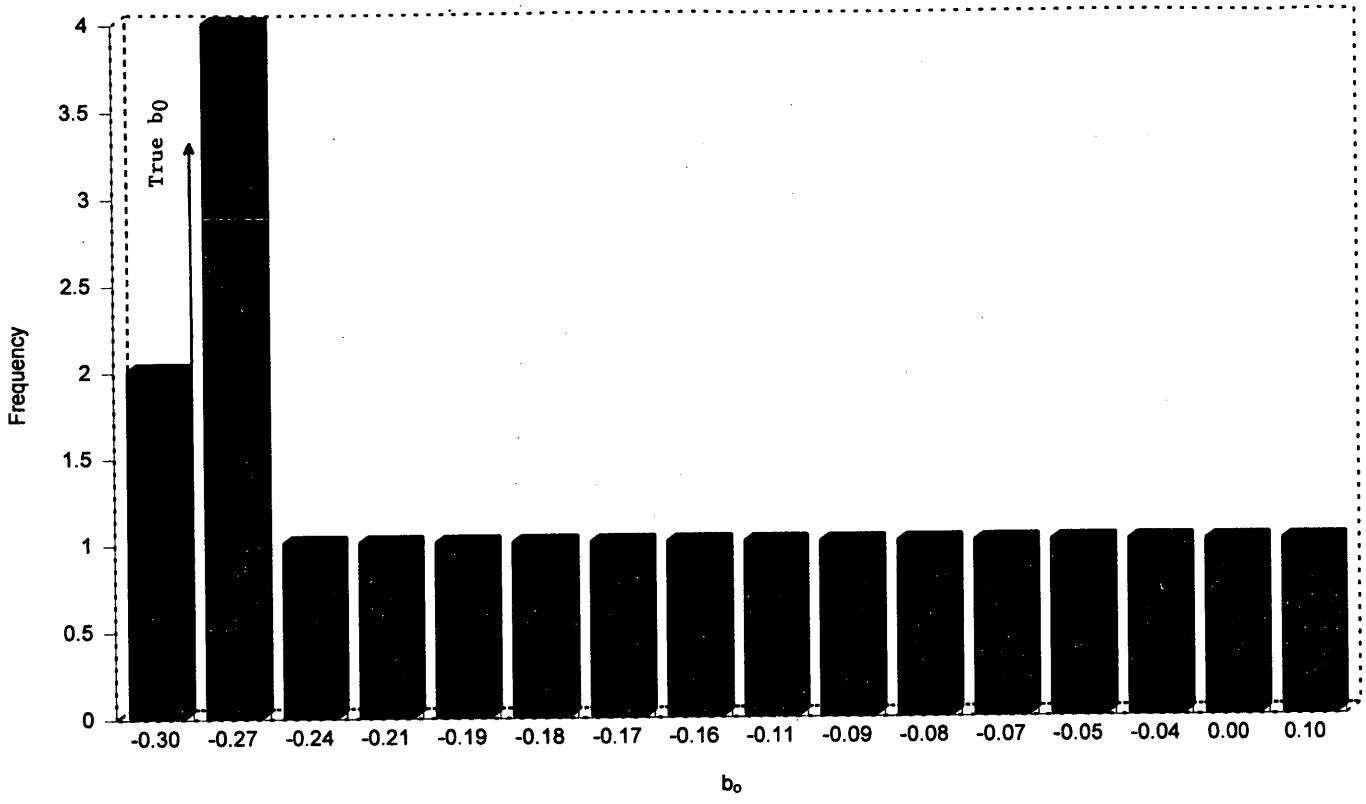



**FIGURE 6** Frequency distribution of $b_0$.

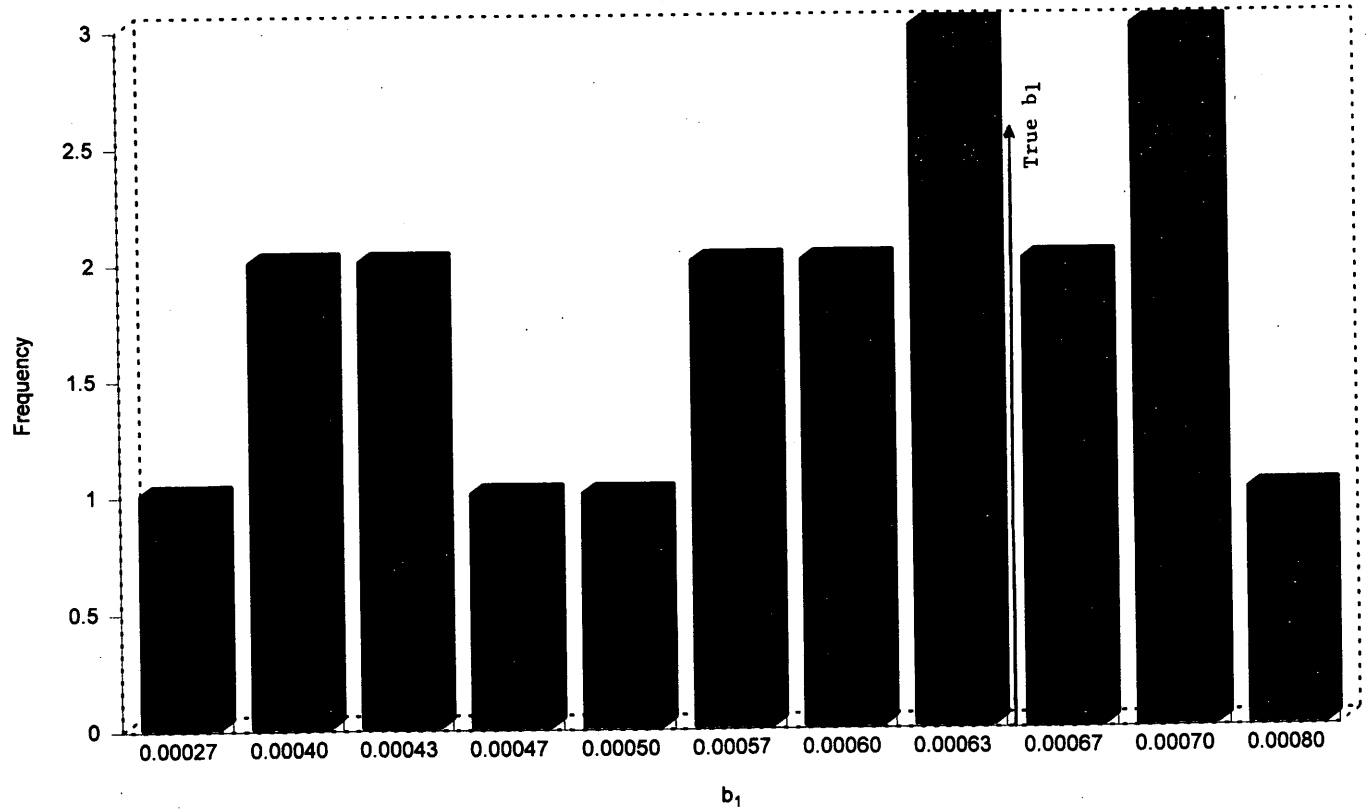



**FIGURE 7** Frequency distribution of $b_1$.

ple of size 10 will on the average result in observing only five breakdown occurrences.

## CONCLUSIONS AND FUTURE RESEARCH

Field data, taken during three peak periods at two merge junctions, were examined to study capacity, breakdown, and speed-flow issues. The first one, Site 28, experienced a disturbance of traffic operations for only 10 min, but it gave valuable information about breakdown conditions. The fact that the speed never drops below 68 km/h downstream of the merge area precludes the occurrence of a downstream bottleneck affecting the section. The flows at the time of breakdown were not the highest observed at this site. Observation of videotapes showed that a ramp cluster entered the freeway stream immediately before breakdown and caused the speed drop at the merge.

At the second site, Site 21, the last camera was not far enough downstream for free-flow conditions to be observed after the bottleneck section. Nevertheless, Site 21 data have many similarities to Site 28 data. The flows were not the maximum observed at this site when breakdown occurred. Again, a large cluster was observed entering the freeway immediately before breakdown. During the second data collection period at this site (Site 59), breakdown did not occur, even though the ramp and freeway flows were at comparable levels. Observation of field conditions verified that there were no large ramp clusters occurring throughout the data collection at this site.

Following these observations, a probabilistic model that was developed to predict breakdown at ramp-freeway junctions as a function of ramp clusters and freeway flow was examined. In general, the following conclusions can be drawn:

- The breakdown at ramp-freeway junctions is a probabilistic variable and not deterministic, as it has been thought. The data showed that breakdown does not always occur at given volumes, even at the same site.
- Capacity does not necessarily occur immediately before breakdown. Data showed that capacity flows are not a prerequisite for breakdown and are not the only factor in breakdown occurrence.
- Clusters of vehicles from the ramp, rather than ramp flow, may cause breakdown. Even though clusters are a function of ramp flow, it is the clusters that affect operations at ramp-freeway junctions.
- Ramp metering, which helps break up the ramp vehicle clusters, improves traffic operations at merge junctions. Even though it is assumed that ramp metering is beneficial because it decreases the ramp volume, in reality its greatest benefit is that it alleviates large clusters. The same effect may be achieved by using small cycle lengths at the signals upstream from the ramp.
- Since the probability of breakdown does not increase much beyond 70 percent for ramp flows of 1,500 vph, it can be explained why there are few instances in which one can observe breakdown that is a direct result of the merging maneuvers. In most cases, a downstream constriction creates congestion and breakdown, which eventually spreads upstream into the merge area.
- Speed remains almost constant at both sites and all camera locations until the breakdown. At the last cameras downstream, there is a relatively gradual speed drop before the breakdown, but it does not occur until operations are about to become unstable.
- More data are needed to validate the probabilistic model developed here with more precision, or to establish other factors that may influence the model. Establishing the true probability of breakdown at merge junctions is required for estimating the sample size needed for observing breakdown.

## ACKNOWLEDGMENT

The data used for this paper were obtained from the NCHRP Project 3-37 data base.

## REFERENCES

1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
2. Roess, R. P., and J. M. Ulerio. *Capacity of Ramp-Freeway Junctions.* NCHRP Project 3-37, Final Report. Polytechnic University, Brooklyn, 1994.
3. Elefteriadou, A. *A Probabilistic Model of Breakdown at Freeway-Merge Junctions.* Dissertation. Department of Civil and Environmental Engineering, Polytechnic University, Brooklyn, N.Y., June 1994.
4. Persaud, B. N. *Study of a Freeway Bottleneck To Explore Some Unresolved Traffic Flow Issues.* Dissertation. Department of Civil Engineering, University of Toronto, Canada, 1986.
5. Allen, B. L., F. L. Hall, and M. A. Gunter. Another Look at Identifying Speed-Flow Relationships on Freeways. In *Transportation Research Record 1005,* TRB, National Research Council, Washington, D.C., 1985, pp. 54–64.
6. Banks, J. H. Flow Processes at a Freeway Bottleneck. In *Transportation Research Record 1287,* TRB, National Research Council, Washington, D.C., 1990, pp. 20–28.
7. Kühne, R. D. Freeway Control Using a Dynamic Traffic Flow Model and Vehicle Re-Identification Techniques. In *Transportation Research Record 1320,* TRB, National Research Council, Washington D.C., 1991.



*Publication of this paper sponsored by committee on Highway Capacity and Quality of Service.*

# Developing Passenger-Car Equivalents for Left-Turning Trucks at Compressed Diamond Interchanges

JAMES E. WEST AND GLEN S. THURGOOD

A study on the effects of light and heavy trucks on left-turning queues at compressed diamond interchanges is described. The headway method was used for determining passenger-car equivalents (PCEs) of various left-turning light and heavy truck classes, including some specialty vehicles. Inside turn radii were 15 to 18 m (50 to 60 ft), which is typical of many left turns for compressed diamond interchanges, where freeway on/off ramps are brought in as close to the freeway as possible. Composite PCEs for standard light and heavy trucks are reported as 1.7 and 4.4, respectively. PCEs for specialty vehicles are given as well.

Of all the elements that make up a freeway, the interchange is commonly one of the greatest capacity-limiting factors in the system. To minimize the bottleneck problem, several types of grade-separated interchanges have been used with varying degrees of success, the most common being variations of the diamond interchange. Within the last 15 years or so, a modified version of the diamond interchange has gained popularity. It is commonly referred to as the single-point urban interchange (SPUI) and has been touted by some as the universal answer to the problem of efficiently moving vehicles on and off the freeway. Others are cautious and are involved in research to determine exactly how efficient and safe the SPUI really is (*1–3*). As a result, there is considerable interest in comparing the operating characteristics of the SPUI with those of the diamond interchange.

Observing heavy trucks turning left at SPUI interchanges leaves one with the impression that the SPUI geometry handles left-turn maneuvers much more efficiently than do the short-radius turns permitted by the compressed diamond configuration. On the other hand, the lower clearance interval between phase transitions at compressed diamond interchanges appears to favor vehicles making through movements. If this assumption is correct, one interchange design may have a capacity advantage over the other, depending on the size of the interchange and the number of trucks in the traffic stream and their predominant movement. Recent research into SPUI operation supports this hypothesis by showing that SPUIs are not always more efficient at moving traffic than standard diamond interchanges (*2*). (NCHRP Project 3-47, which is ongoing, would consider the diamond interchange in this research to be a tight urban diamond rather than a compressed diamond.)

The research in this paper describes the method used in calculating left-turning passenger-car equivalents (PCEs) for trucks at diamond interchanges, with the hope that the procedure will also be used at SPUIs, whereby left-turning PCEs may be compared between the two interchanges to determine which design has the least negative impact on truck operation. The results of the research may be particularly important at locations where truck operations are high, such as interchanges near industrial and warehouse areas, since an improper evaluation of the effect of heavy vehicles on interchange operation could substantially affect the accuracy of any operational analyses.

J. E. West, Kimley-Horn and Associates, Inc., 9320 S.W. Barbur Boulevard, Suite 130, Portland, Oreg. 97219. G. S. Thurgood, Department of Civil Engineering, Brigham Young University, 368-J Clyde Building, Provo, Utah 84602.

## FACTORS INFLUENCING PCEs

In understanding the capacity of signalized intersections, it is assumed that the ideal traffic stream consists of only passenger cars and such other factors as 0 percent grade, 3.7-m (12-ft) lanes, no parking, and dry pavement conditions. In most instances, the traffic stream is less than ideal and contains a mixture of cars and trucks. As trucks are introduced into the traffic flow, the ability of the roadway to carry vehicles is reduced because of the increased size of the trucks and their lower performance characteristics.

In defining the effect of trucks on traffic flow, the term "PCE" was introduced in the 1965 *Highway Capacity Manual*. PCE referred to "the number of passenger cars displaced in the traffic flow by a truck or bus, under the prevailing roadway and traffic conditions" (*4*). A review of current literature indicates that three general factors affect the size of the PCE: truck length, truck turning and acceleration characteristics, and behavior of following drivers (*5*). PCEs are also influenced by traffic conditions, weather, and other environmental factors; however, these factors were not considered specifically in this analysis.

The first factor (truck length) is relatively easy to understand. Passenger cars typically are shorter than 5.5 m (18 ft), whereas it is common for large five-axle trucks to be longer than 15 m (50 ft) (*6*). Physically, trucks take up more space than passenger cars; therefore, as the length of the truck increases, that portion of the PCE increases also.

Truck turning and acceleration characteristics are based on the performance capabilities of the truck, such as minimum turning radii and weight-to-horsepower ratio. Vehicle performance curves show that passenger cars can accelerate more than twice as fast as heavy trucks on a 0 percent grade (*6,7*). Differences in truck acceleration characteristics and their inability to negotiate small-radius turns cause them to have longer travel times through an intersection, thus increasing the PCE. If it were possible to improve truck performance so that the weight-to-horsepower ratio was essentially that of a car, this factor could be reduced; however, even with the latest engine power improvements, the difference is still significant because of turning limitations.

The third factor—the behavior of a following driver—is the most difficult element to quantify. Large trucks often are difficult to see around and create an uncomfortable feeling for a driver following behind closely. As a result, it is common for drivers to increase the separation between the two vehicles until a comfortable gap is achieved. Increasing the distance between vehicles further increases the PCE (*5*). The ability of following cars to accelerate is also limited by the acceleration and speed of the leading truck.

## CURRENT VALUES FOR PCEs

Since the introduction of the term PCE almost 30 years ago, much research has been performed to quantify acceptable values; however, almost all of the studies have examined trucks moving in a straight path.

The basis for much of the research was pioneered in studies by Greenshields, et al. that dealt with saturation flow theory and vehicle start-up time (*8*). Their results, although slightly refined since their introduction, have proved to be remarkably accurate and consistent with later studies (*7*). Their findings and the later research by others give a clear understanding of the dynamics of starting a standing queue of vehicles.

When a stopped queue of vehicles prepares to move, the driver of the first vehicle must see the green signal indication and react to the change by removing his or her foot from the brake and then accelerating across the stop line. The process requires a relatively long period for the first vehicle. The second vehicle makes the same perception and reaction response but is able to initiate it at almost the identical time as the first vehicle. This allows the second vehicle to reduce its headway (the time to cross the stop line after the first car crosses). A similar procedure occurs for the remaining vehicles in the queue, with succeeding vehicles further reducing their headways. At about the sixth vehicle, the effects of starting up the queue are dissipated and the remaining vehicles travel at a constant headway and speed (*7*). As trucks are introduced into the traffic stream, they increase the time that it takes for the queue of stopped vehicles to achieve saturation flow.

Since the introduction of PCEs, researchers have developed their own suggested values, each taking into account some of the factors that affect PCEs. Miller found that compared with a car, it took an additional 1.79 sec for a commercial truck to cross the stop line. He divided the average headway of the truck by the average headway of a passenger car and obtained a PCE value of 1.85. Miller's work was one of the first to define quantitatively a PCE value (*9*).

Carstens essentially repeated Miller's procedure by measuring headways, where measurements were made as the front bumpers passed a stop line (known as leading headway). His research resulted in a PCE of 1.63, a value that supports Miller's findings (*10*).

Later, Branston and van Zuylen measured headways as the rear bumpers crossed the stop line (lagging headway). They attempted to separate trucks into two classes. Their results showed PCEs for medium and heavy trucks to be 1.35 and 1.68, respectively. It should be noted that their classification of heavy trucks did not include any trucks with five or more axles (*11*).

The 1985 *Highway Capacity Manual* (HCM) uses a PCE of 1.5 to 1.6 for signalized intersections to determine its heavy vehicle factor; this method is used by many engineers. The HCM factor makes no attempt to separate heavy vehicles in any way, but instead groups trucks, buses, and recreational vehicles into one category. Therefore, the PCE is an average value for all the types of vehicles (*7*).

Not all PCE research has focused on trucks traveling in a straight path. Although it is unclear what parameters were involved in obtaining the results, an Australian PCE methodology recommends using a combined value of 2.5 for a nonconflict turning truck and 3.9 for a conflict turning truck (*12*).

Although each researcher attempted to define a PCE value, it was not until later that a methodology was developed that included essentially all the factors that make up a PCE.

## STUDY METHODOLOGY

### General

In May 1987 a report by Molina et al. suggested that PCEs for light and heavy trucks traveling along a straight path should be 1.7 and 3.7, respectively (*5*). Their research gives strong reason to believe that current PCEs are low, but unfortunately their work did not evaluate the effect of trucks turning left at an intersection. Their methodology was the model for conducting this research into PCEs for left-turning trucks.

### Study Model

The headway method is the most common method used for calculating PCEs. Equation 1 describes the difference between truck and passenger-car headways:

$$\text{PCE} = h_t/h_c \tag{1}$$

where $h_t$ is the headway of the truck, and $h_c$ is the saturation flow headway of a passenger car.

Equation 1 describes the effect of the increased size of the truck and its lower acceleration characteristics. Although Equation 1 is the most common method used to calculate PCEs, it does not consider the generally slower discharge rate of trucks, which affects start-up lost time and saturation flow headways.

To account for the effect of the slower truck acceleration that propagates down the queue of vehicles, causing increased delay, a factor must be added to Equation 1. This additional factor, which includes start-up lost time, is perhaps one of the major differences from early PCE studies. The inclusion in the model of start-up lost time is intentional and was expected to yield higher PCEs than those found by other researchers.

The additional factor is the incremental increase in the headways caused by the truck. Not all of the incremental increase is caused by slower truck acceleration: some of the increase is a result of passenger cars shying away from the large truck rather than the physical impedance. The increase is measured to the point at which the vehicles following the truck are no longer impeded and are able to travel at a speed as if the truck were not present in the traffic stream.

Unfortunately, there is no simple method for directly measuring that incremental increase. However, as a substitute, the problem may be solved by measuring the discharge time of a queue of passenger cars with a truck in the queue and then comparing it with the discharge time of a queue of the same size, which consists only of passenger cars.

The differences between the light and dark bars in Figure 1 represent the incremental increased headway accruing to each vehicle

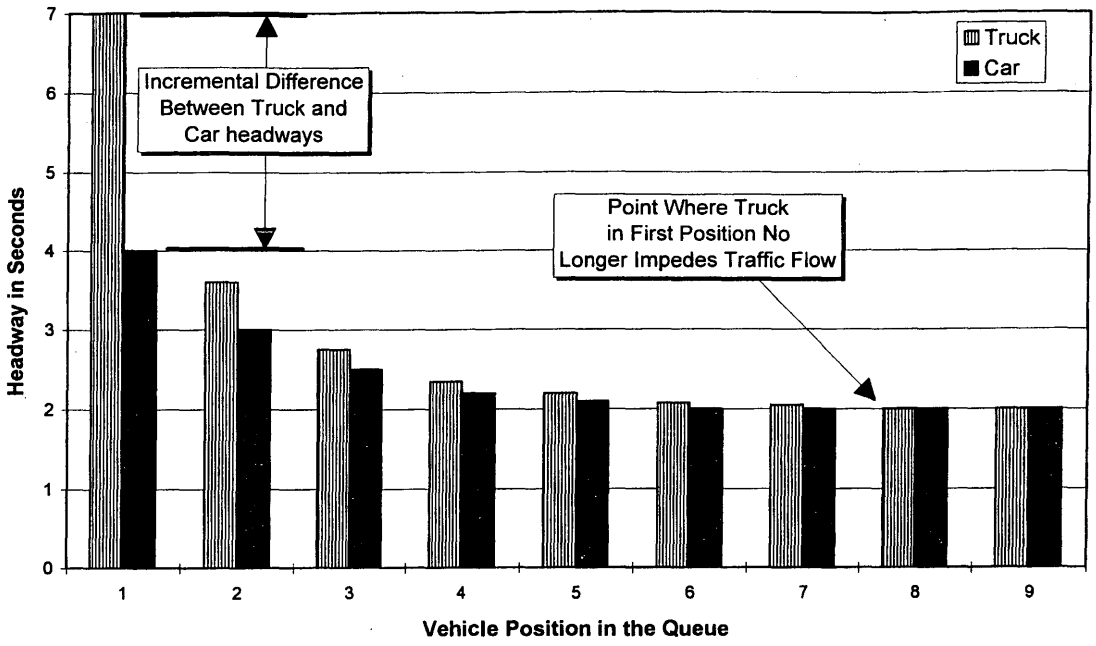



**FIGURE 1 Theoretical difference between truck and passenger-car headways.**

in the queue as a result of a truck being in the first position. In this example, the truck no longer impedes following vehicles at the eighth position. At that point, the headways between the two queues are essentially the same as if no truck were present. The sum of the incremental differences in discharge times between the two queues quantifies the total impact of the truck on the other vehicles in the queue (i.e., a passenger car queue with a truck in the first position versus a passenger car queue only).

By considering the incremental increase in headways, one can account for all the factors that affect the size of the PCE (truck size, acceleration, and driver behavior). In other words, the method accounts for the fact that the truck in the queue occupies more space than a passenger car and that the effects of the lower performance characteristics, coupled with driver behavior, propagate down the queue and cause a number of following vehicles to be delayed. Eventually the truck reaches the normal travel speed and the delay effects caused by the truck dissipate (*5*). At this point, the headways in the queue are essentially the same as the headways of a queue containing all passenger cars, except for the longer headway of the truck due to its longer length.

The entire theory is based in the assumption that only one truck is present in the queue at a time. Often this is not the case, but again, a direct method to account for additional trucks is not available. However, it is possible to quantify the impact of additional trucks by applying the same methodology. Even though the effect of additional trucks cannot be measured directly, it is unlikely that a second truck would double the PCE value. Judgment suggests that the truck that has the lower performance capabilities will control the PCE value. The other trucks following the control vehicle will at least be able to approach the speed of saturation headway as they cross the stop line. However, the effect of an additional truck on the moving queue due to its increased size and following driver behavior still remains.

Since it is not a simple matter to add the incremental increase in delay to Equation 1, a new equation was written that describes the desired effect. It is assumed that the PCE for a truck in a queue is a number greater than 1. Thus, the final equation that describes the effect of a truck in the traffic stream is shown in a simplified form (*5*):

$$PCE = [(TT_t - TT_c)/h_c] + 1 \quad (2)$$

where

$TT_t$ = total discharge time of truck queue;
$TT_c$ = total discharge time of passenger-car queue,
$h_c$ = saturation flow headway of all-passenger-car queue;

In addition,

- PCEs are calculated with the location of the truck ranging from Positions 1 to 10,
- Headways are measured at the point where the rear wheels of the vehicle cross the stop bar,
- Saturation flow headway ($h_c$) is based on stable moving queues of passenger cars at this study location ($h_c$ was found to be 2.0), and
- Vehicle queues contain an equal number of vehicles.

During their research, Molina et al. were able to obtain hundreds of truck samples from three intersections (*5*). The left-turn research for this study was conducted on a limited budget and under a time constraint that did not allow for a large study sample. Fortunately, Molina et al. did offer guidelines on the minimum size of the sample (i.e., number of queues) to ensure that the results would be statistically correct. For the truck data to be statistically valid, it was determined that each truck position being examined must have at least five observations (*5*). Therefore, if the examination was to include 10 positions and four truck types, the minimum number of observations would be $5 \times 4 \times 10 = 200$, plus all the passenger-car data that could be obtained. All passenger-car queues needed at

least seven cars to be usable (*5*). In addition, traffic conditions needed to be at or near capacity (saturated) for the left-turn movement. Sporadic arrivals during the green indication would not give accurate results.

## STUDY PROCEDURE

### General

The study procedure essentially replicates the method used by Molina et al. during their analysis. The site used in the research was a diamond interchange located at 4500 South and Interstate 15 in Salt Lake City, Utah. The interchange was in a location where heavy trucks often used the facility. One nearby traffic generator was a concrete ready-mix plant. As a result, a large portion of the truck traffic consisted of multitrailer dump trucks, combination hopper trucks, and concrete mix trucks. The abundance of these "special vehicles" allowed an opportunity to calculate which standard truck class they most closely resembled in terms of PCEs. Figure 2 shows the special truck classes observed during the data collection period. Data were collected during good weather and daylight conditions on each side of the diamond interchange during the morning and evening peak periods, as well as during midday, when sufficient traffic volumes permitted. Grade at the location is level.

### Equipment

Two methods were used to collect the raw data: on-site observation and videotape. The on-site observations were recorded using a personal computer (PC) and adapting the Traffic Data Input Program (TDIP) software (*13*). The software package was intended to be used in analyzing delays at stop-controlled intersections, but it was useful in recording the discharge times (headways) between successive vehicles. The software allowed the observer to start the count at the beginning of the left-turn green indication and record when the rear wheels of each vehicle in the queue crossed the stop bar.

At the start of the green indication, a specific key was depressed on the keyboard. As the rear tire of each vehicle crossed the stop line, a second key was pressed. Pressing the keys caused the times to be recorded in an ASCII file. After the complete observation of a queue, it was noted on a field sheet whether any trucks were present and, if so, the type, number of axles, queue position, and a brief description. Keeping accurate field records was essential since they would enable the identification of the data set that matched each queue.

With the video camera, the information was recorded on tape in the field and later extracted using the PC and the TDIP program. The camera was positioned so that the stop line and as much of the standing queue as possible could be seen clearly. Since the camera position did not allow the green signal indication to appear in the viewfinder at the same time as the stop line, a simple light tree was hard-wired into the signal controller and placed in the camera's field of vision to facilitate the precise determination of the beginning time of each left-turn movement.

### Data Analysis

When all available headway information was recorded in the data files, each file was loaded into a spreadsheet program for analysis.

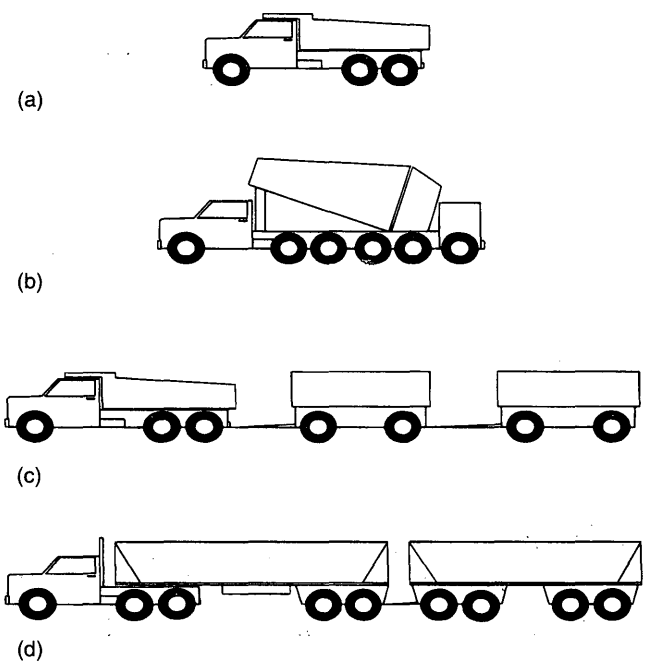



**FIGURE 2 Special vehicle classes: *a*, three-axle dump truck; *b*, six-axle concrete mix truck; *c*, seven-axle dump truck with two trailers; *d*, nine-axle hopper truck with one trailer.**

One column of the spreadsheet contained the times at the beginning of the green indications. A second column contained the times at which each vehicle in the queue discharged across the stop line.

The analysis procedure required that each start time be paired with the appropriate vehicle crossing times and that they be identified as to the type of truck involved, if applicable.

Once the time differences between vehicles were calculated, corresponding data sets were identified. In other words, there was one grouping for the all passenger-car queues, one for two-axle trucks in the first position, one for two-axle trucks in the second position, and so on for all trucks and their specific positions.

The next step was to compute average headways for each vehicle position in each grouping. For example, in the all passenger-car queues, the average headway was calculated for the first cars, the second, the third, and so forth. This same procedure was also done for every truck position in every different truck grouping. Unfortunately, headways for some of the truck types and some of the truck positions were not obtained because of insufficient data.

Referring to Equation 2, $TT_t$ is the total discharge time for the truck queue being considered. It is measured as the sum of the headways from the start of the green indication to the point at which the first passenger car behind the truck reaches saturation flow headway. Figure 3 graphically illustrates the differences between the headways for some of the truck types in the first position. For the five-axle truck in the first position, saturation flow headway occurred at the seventh vehicle with the sum of the headways being 23.4 sec. The average headways were then applied to Equation 2 to determine a PCE for each vehicle class and position in the queue.

$TT_c$ is the total discharge time for a passenger-car queue. It is measured as the sum of the headways from the start of the green indication to the same point as in the queue with the truck. In this example it occurred at the seventh vehicle, the sum being 16.0 sec.

The saturation flow headway ($h_c$) was already identified as occurring at the seventh vehicle; it is 2.0 sec.

Applying the equation for a five-axle truck in the first position yields

$$PCE = (23.4 - 16.0)/2.0 + 1 = 4.7$$

## STUDY RESULTS

### General

It was anticipated that the PCEs determined for left-turning trucks would be slightly higher than the PCEs determined by Molina et al. who examined trucks traveling in a straight path. The results from the data showed that PCEs were approximately as expected. Figure 3 shows that three-axle trucks compare closely to passenger cars. On the other hand, the heavy trucks take considerably longer to get moving.

### PCEs versus Position in Queue

Table 1 gives the PCE for each truck class and position in the queue, as well as the number of truck queues in each data set. The results show that the highest PCEs occur with the truck in first position in the queue, and as the platoon of vehicles approaches saturation flow, the PCE reaches its minimum value.

It can be seen that the light trucks (two-axle truck, three-axle truck, three-axle dump truck, and six-axle concrete mix truck) have comparable PCEs. It is also evident that the heavy trucks (five-axle combination truck, seven-axle dump truck with trailers, and nine-axle combination hopper truck with one trailer) have about the same PCEs.

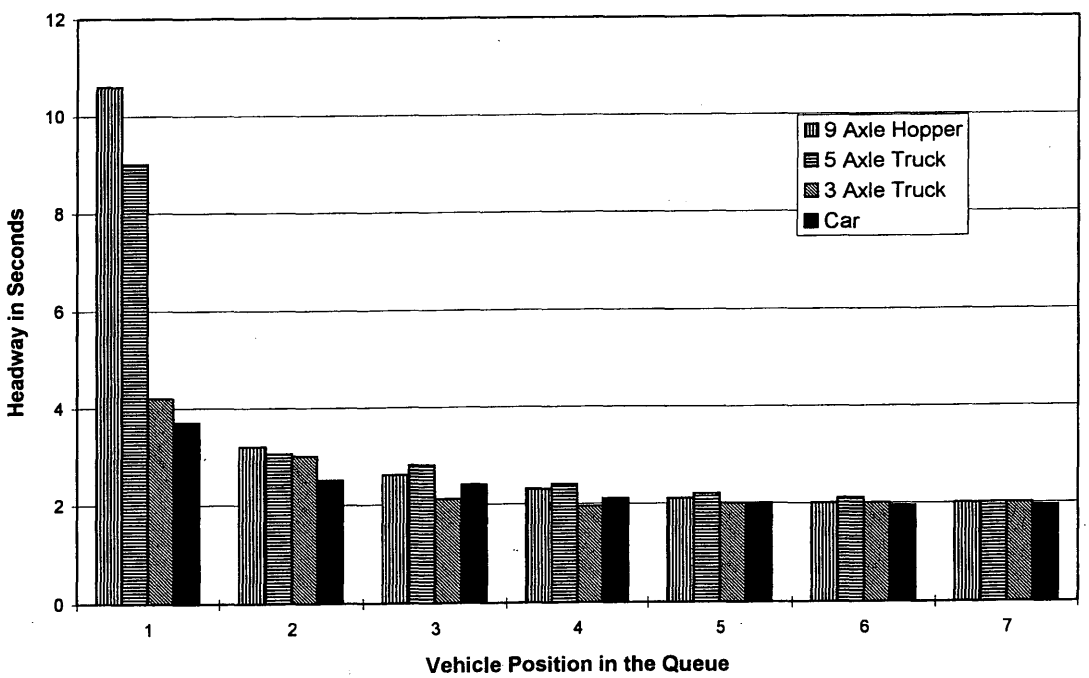



**FIGURE 3 Headways for various trucks in first position versus passenger-car headways.**

TABLE 1 PCEs versus Truck Position in Queue

| Truck Class | Queue Position | Number of Observations | PCE |
|---|---|---|---|
| 2-Axle Single | 1 | 13 | 2 |
| | 2 | 12 | 1.8 |
| | 3 | 11 | 1.7 |
| | 4 | 14 | 1.7 |
| | 5 | 14 | 1.6 |
| | 6 | 11 | 1.6 |
| | 7-10 | * | ** |
| 3-Axle Single | 1 | 13 | 2.1 |
| | 2 | 11 | 2 |
| | 3 | 12 | 1.8 |
| | 4 | 11 | 1.7 |
| | 5 | 12 | 1.6 |
| | 6 | 11 | 1.6 |
| | 7-10 | * | ** |
| 5-Axle Combination | 1 | 17 | 4.7 |
| | 2 | 12 | 4.7 |
| | 3 | 13 | 4.5 |
| | 4 | 17 | 4.3 |
| | 5 | 12 | 4 |
| | 6-10 | * | ** |
| 3-Axle Dump Truck | 1 | 13 | 2.2 |
| | 2 | * | * |
| | 3 | * | * |
| | 4 | 12 | 1.7 |
| | 5 | 11 | 1.6 |
| | 6-10 | 11 | 1.6 |
| 7-Axle Dump Truck w/ 2 Trailers | 1 | 14 | 5.7 |
| | 2 | 12 | 5 |
| | 3 | 11 | 4.1 |
| | 4 | * | ** |
| | 5 | 11 | 3.5 |
| | 6-10 | * | ** |
| 6-Axle Concrete Mix Truck | 1 | 12 | 2.8 |
| | 2 | 13 | 2.5 |
| | 3 | * | ** |
| | 4 | * | ** |
| | 5 | * | ** |
| | 6 | 11 | 1.7 |
| | 7-10 | * | ** |
| 9-Axle Combination Hopper Truck w/ 2 Trailers | 1 | 15 | 5 |
| | 2 | 12 | 4.9 |
| | 3 | 13 | 4.4 |
| | 4 | * | ** |
| | 5 | 11 | 4.1 |
| | 6 | 11 | 4 |
| | 7-10 | * | ** |

* Insufficient number of observations to calculate PCE.

** No PCE calculated due to insufficient number of observations.

Because of similarities in PCEs, the truck types were grouped into two categories. Grouping the PCEs is logical since it is doubtful that anyone performing a traffic count would have the resources to apply a separate PCE for each truck type and for each position in the queue.

For the location used in this research, it was appropriate to calculate a weighted average for both the light and heavy truck groups and include the special vehicles unique to the study site (Table 2). Doing so produced a left-turn PCE of 1.9 for the light trucks and a PCE of 4.5 for the heavy trucks. These values are averages for all queue positions. If this information is used elsewhere, the weighted average PCE should be based only on those types of trucks that use the facility and whether the intersection geometrics are similar to the study location. If the special trucks are excluded from the data set (i.e., not present), the weighted average left-turn PCE is 1.7 for light trucks and 4.4 for heavy trucks.

If the results are to be used in traffic applications other than at this location, note that certain geometric factors unique to this interchange are reflected in the results. The approximate inside radii for the left-turn movements were 15 to 18 m (50 to 60 ft). Interchanges with three lanes in each direction along the arterial road would probably have an even larger turning radius. The larger-radius turns should allow trucks to negotiate the turns at higher speeds, which would lower the PCE.

The PCEs reflect a mix of loaded and unloaded trucks. No rigid effort was made to note which trucks were carrying a load, but general observation of the traffic mix conditions indicates that about half the trucks were loaded.

A comparison of these results with those of Molina et al., who recommended using PCEs of 1.7 for light trucks and 3.7 for heavy trucks for a through movement on level grade, suggests that light trucks are relatively unaffected by the left-turn maneuver whereas heavy trucks are slowed by it. The difference may at first seem small, but if many trucks are in the traffic stream, using an inappropriate PCE can make a significant difference in capacity calculations.

## APPLICATION OF RESULTS

### General

The PCE results are intended to be used in computer modeling and analysis techniques that evaluate intersection capacity and level of service. The information is also expected to be helpful to traffic engineers who design, maintain, and operate signalized intersections. PCEs for left-turn movements at modern SPUIs appear to approximate those for through movements at intersections because of the much larger turning radii provided. Analysis of left turns at compressed diamond interchanges should reflect the higher PCEs resulting from the shorter-radius left turns. The left-turn PCEs determined in this paper are just that—PCEs for left-turning trucks with an inside radius of turn of 15 to 18 m (50 to 60 ft).

## CONCLUSIONS AND RECOMMENDATIONS

### General

The conclusions and recommendations in this paper are intended to serve as a model for further research. They are to encourage a procedure for obtaining more accurate truck influence information using relatively simple procedures and equipment.

### Conclusions

- As initially expected, PCEs for trucks turning left are higher than those for trucks traveling in a straight path.
- Light trucks have only a small effect on the traffic stream because of their good performance characteristics. A low weight-to-horsepower ratio allows for the trucks to accelerate quickly up to saturation flow speed.
- Heavy trucks have a large influence on the traffic stream, especially when turning left, because of (*a*) their high weight-to-horsepower ratio, which causes them to take longer to accelerate, and (*b*) their limited ability to negotiate sharp turns. It was noted that after the heavy trucks did get moving, the PCE decreased noticeably.
- Concrete mix trucks and single-unit dump trucks closely resemble the PCE for lightweight trucks.
- Multitrailer dump trucks and hopper trucks with trailers have PCEs approximately the same as heavy trucks.
- Left-turning PCEs for standard light and heavy trucks should be 1.7 and 4.4, respectively.
- If many special trucks are present in the traffic stream, the above PCEs should be modified as described in this paper.

### Recommendations

- More data collection and study are recommended to calculate PCEs for other specialty vehicles.
- Similar studies are recommended at SPUIs to allow further comparisons between interchange types.
- More studies are recommended to determine the effect on PCE values of two or more trucks in a queue.

**TABLE 2 Weighted Average PCEs for Each Truck Class**

| Truck Class | Weighted Average PCE |
|---|---|
| 2-Axle Single | 1.7 |
| 3-Axle Single | 1.8 |
| 3-Axle Dump Truck | 1.8 |
| 6-Axle Concrete Mix Truck | 2.4 |
| 5-Axle Combination | 4.4 |
| 9-Axle Combination Hopper Truck w/ 2 Trailers | 4.5 |
| 7-Axle Dump Truck w/ 2 Trailers | 4.7 |

## ACKNOWLEDGMENT

Funding support for this research was provided by the Utah Department of Transportation.

## REFERENCES

1. Leisch, J. Innovations of Diamond Interchanges in the Urban Development. *ITE Journal*, Nov. 1985, pp. 24–26.
2. Messer, C. J., J. A. Bonneson, S. D. Anderson, and W. F. McFarland. *NCHRP Report 345: Single Point Urban Interchange Design and Operations Analysis.* TRB, National Research Council, Washington, D.C., Dec. 1991.
3. Leisch, J., T. Urbanik II, and J. P. Oxley. A Comparison of Two Diamond Interchange Forms in Urban Areas. *ITE Journal*, May 1989.
4. *Special Report 87: Highway Capacity Manual.* HRB, National Research Council, Washington, D.C., 1965.
5. Molina, C. J., C. J. Messer, and D. B. Fambro. *Passenger Car Equivalencies for Large Trucks at Signalized Intersections.* Report FHWA/TX-87/397-2. May 1987.
6. *A Policy on Geometric Design of Highways and Streets.* AASHTO, Washington, D.C., 1984.
7. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
8. Greenshields, B. D., D. Shapiro, and E. L. Erickson. *Traffic Performance at Urban Intersections. Technical Report 1.* Bureau of Highway Traffic, 1947.
9. Miller, A. J. *The Capacity of Signalised Intersections in Australia. Report 3.* Australian Road Research Board, March 1968.
10. Carstens, R. L. Some Traffic Parameters at Signalized Intersections. *Traffic Engineering*, Vol. 41, No. 11, Aug. 1971, pp. 33–36.
11. Branston, D., and H. van Zuylen. The Estimation of Saturation Flow, Effective Green Time and Passenger Car Equivalents at Traffic Signals by Multiple Linear Regression. *Transportation Research*, Vol. 12, No. 1, Feb. 1978, pp. 47–53.
12. *Transportation and Traffic Engineering Handbook,* 2nd ed. ITE, 1982, pp. 511–512.
13. Kyte, M. *Traffic Data Input Program (TDIP) Documentation and User's Guide.* University of Idaho, May 1988.

---



*Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.*

# Development of Safety-Based Level-of-Service Criteria for Isolated Signalized Intersections

TAE-JUN HA AND W. D. BERG

The *Highway Capacity Manual* specifies procedures for evaluating intersection performance in terms of delay per vehicle. What is lacking in the current methodology is a comparable quantitative procedure for assessing the safety-based level of service (LOS) provided to motorists. The objective of the research was to develop a computational procedure for evaluating the safety-based LOS of signalized intersections based on the relative hazard of alternative intersection designs and signal timing plans. Conflict opportunity models were derived for those crossing, diverging, and stopping maneuvers associated with left-turn and rear-end accidents. Safety-based LOS criteria were then defined on the basis of distribution of conflict opportunities computed from the developed models. A case study evaluation of the LOS analysis methodology revealed that the developed safety-based criteria were not as sensitive to changes in prevailing traffic, roadway, and signal timing conditions as the traditional delay-based measure. However, the methodology did permit a quantitative assessment of the trade-off between delay reduction and safety improvement. The results of the research are considered to be of an exploratory nature.

The *Highway Capacity Manual* (HCM) specifies procedures for evaluating intersection performance in terms of a wide variety of prevailing conditions such as traffic composition, intersection geometry, traffic volumes, and signal timing (*1*). Performance, however, is only measured in terms of delay per vehicle. It is a parameter that is widely accepted as a meaningful and useful indicator of the efficiency with which an intersection is serving traffic needs.

What is lacking in the current methodology is a comparable quantitative procedure for assessing the safety-based level of service (LOS) provided to motorists. For example, it is well-known that the change from permissive to protected left-turn phasing can reduce left-turn accident frequency. However, the HCM permits a quantitative assessment of the impact of this alternative phasing arrangement only on vehicle delay. It is left to the engineer or planner to judge subjectively the level of safety benefits and to evaluate the trade-off between the efficiency and safety consequences of the other phasing plans. Many examples of other geometric design and signal timing improvements could also be given.

At present, the principal methods available to the practitioner for evaluating the relative safety at signalized intersections are (*a*) the application of engineering judgment (*b*) accident analyses, and (*c*) traffic conflicts analysis. Reliance on engineering judgment has obvious limitations, especially when placed in the context of the elaborate HCM procedures for calculating delay. Accident analyses generally require some type of before-after comparison, either for the case study intersection or for a large set of similar intersections. In either situation, there are problems associated with compensating for regression-to-the-mean phenomena (*2*) as well as with obtaining an adequate sample size. Research has also pointed to potential bias caused by the way in which exposure to accidents is measured (*3,4*). Because of such problems with traditional accident analyses, some have promoted the use of the traffic conflicts technique (*5*). However, this procedure also has shortcomings in that it requires extensive field data collection and trained observers to identify the different types of conflicts occurring in the field.

The objective of the research described herein was to develop a computational procedure for evaluating the safety-based LOS of signalized intersections that would be compatible and consistent with that presently found in the 1985 HCM for evaluating efficiency-based LOS as measured by delay per vehicle (*6*). The intent was not to develop a new set of accident prediction models but to design a methodology to quantitatively predict the relative hazard of alternative intersection designs and signal timing plans.

## RESEARCH APPROACH

It was assumed that by adapting and enhancing the accident exposure models developed for FHWA by Council et al. (*3*), a practical safety-based LOS indicator could be designed. The work of Council et al. was founded on the premise that a quantitative estimate of the number of conflict opportunities for a given accident type is a preferable measure of exposure to accidents than that of simply summing the number of vehicles entering an intersection. Although that work was focused on developing more sensitive and less biased accident rate expressions, the resulting conflict exposure equations offered an excellent starting point for the development of a safety-based LOS indicator that might be incorporated in current capacity analysis procedures.

The models formulated by Council et al. estimate the number of conflict opportunities for each of the following conflict types: single vehicle, rear-end, head-on, angle, and sideswipe. It was assumed that for an accident to occur, first the opportunity for it to occur must be present. The opportunity for an accident consists of the presence of certain prerequisite conditions related to vehicle speeds and relative positions. Without these conditions, the opportunity and therefore the likelihood of a given type of accident do not exist. For example, there is a greater likelihood that angle collisions will occur if left turns are allowed against through traffic. However, if this maneuver is prohibited, the opportunity for such an accident would no longer exist, nor would its likelihood. The prerequisite condition that makes up the opportunity in this case is the simultaneous pres-

Department of Civil and Environmental Engineering, University of Wisconsin–Madison, 2208 Engineering Hall, 1415 Johnson Drive, Madison, Wis. 53706.

ence of both a through and left-turning vehicle within the physical limits of the intersection.

The opportunity models specified by Council et al. did not account for the full range of geometric, traffic flow, and signal timing variables that are input parameters to the HCM procedures. Therefore, a major task of this research was the modification and enhancement of these models. A second major task was to analyze the magnitude and distribution of the resulting estimated conflict opportunities to permit the specification of threshold values that would reflect the relative safety LOS being provided.

## DEVELOPMENT OF CONFLICT OPPORTUNITY MODELS

The 24 possible conflict points at a typical four-leg signalized intersection are illustrated in Figure 1. They include crossing, diverging, merging, and stopping maneuvers. Depending on the signal phasing, several of these conflict points can effectively be eliminated. For example, protected left-turn phasing would eliminate the crossing conflict points. Prohibiting right turns on red would eliminate most merging conflicts associated with right-turn maneuvers.

On the basis of a literature review of accident frequency data by type of maneuver as well as considering of those maneuvers most likely to be influenced by intersection geometrics and signal timing, it was decided to concentrate the modeling of conflict opportunities on those crossing, diverging, and stopping maneuvers associated with left-turn and rear-end accidents. This focus resulted in 16 possible conflict points for a four-leg intersection. Mathematical models were then developed to estimate the frequency of these left-turn and rear-end conflict opportunities.

### Left-Turn Conflict Opportunity Model

Left-turn conflict opportunities involve target vehicles turning left within the intersection proper. They are exposed to traffic flows from the opposing approach entering the intersection proper while the turn is being made. There are two possible scenarios for left-turning vehicles arriving at an intersection. The first is where the left-turning drivers find an acceptable gap when they arrive at the intersection. In this case, they will be able to clear the intersection without a complete stop. The second scenario is where the left-turning drivers are not able to find a suitable gap and have to slow down and eventually come to a stop at the intersection. Two conditions must be present for the opportunity for the latter to occur. The first is that left-turning vehicles are present in the intersection proper and, second, the left-turning vehicles will not be able to find an acceptable gap in the opposing traffic lanes immediately.

Gap is one of the most important factors in determining left-turn opportunities. Very small gap sizes leave little probability for any left-turn conflict opportunity to occur since there would not be enough time for a vehicle to complete a turn. There is also little probability for any left-turn conflict opportunity to occur when the gap sizes are very large, since there would be ample time for a vehicle to make a turn and clear the intersection. The problem, however, lies in identifying the range of the gap sizes that would produce a significant conflict opportunity.

Research on gap acceptance for left-turning vehicles (*7,8*) indicates that a typical accepted gap has a mean of 4 to 5 sec and a variance of approximately 2 sec. Therefore, it was assumed that the range of gaps in opposing traffic that would create a conflict opportunity would be represented by the intersection clearance time ±2.0 sec to reflect the variance of the acceptable gap. The duration

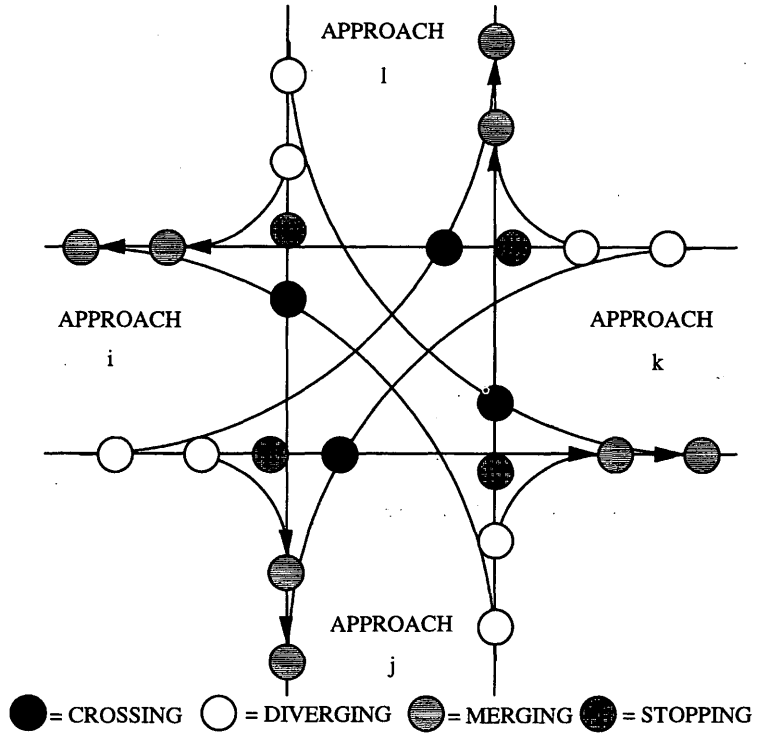



**FIGURE 1 Signalized intersection conflict points.**

of the intersection clearance time varies depending on the width of the opposing lanes, the acceleration rate of the left-turning vehicles, and the length of vehicles. If the headway distribution of the opposing traffic stream on an intersection approach is known, it is then possible to calculate the probability of a left-turn conflict opportunity. However, a few important parameters should be defined and estimated before the necessary equations for a left-turn conflict opportunity measure can be developed.

The first parameter is the estimated turning time of left-turning vehicles at an intersection. Figure 2 shows the assumed typical path of a left-turning vehicle as well as several geometric characteristics of the intersection. The clearance time for an average vehicle 6.7 m (22 ft) long can be calculated as

$$t_i = \sqrt{2(d_i + 22)/a} \tag{1}$$

and

$$d_i = \frac{\pi}{2}\left(W_K + WM_i + \frac{W_i}{2N_i}\right) \tag{2}$$

where

$W_K$ = entire width of approach $K$ (m),
$WM_i$ = width of median on approach $i$ (m),
$W_i$ = entire width of approach $i$ (m),
$N_i$ = number of lanes at approach $i$, and
$a$ = acceleration rate (m/sec$^2$).

Depending on the situation or time at which a vehicle intending to turn left arrives at an intersection, it may make the turn from a stationary or nonstationary position. However, for modeling purposes, all vehicles were assumed to turn from a stationary position, as this would require the longest oncoming gap. It was also assumed that the average acceleration rate of these left-turning vehicles is 1.3 m/sec$^2$ (4.4 ft/sec$^2$), consistent with values used in calculating intersection sight distance requirements.

If a left-turning vehicle takes $t_i$ sec to clear the intersection from approach $i$, the total maneuver time will be $t_i + 2$ sec, assuming a 2-sec driver perception-reaction time. Thus, any through vehicles on the opposing approach that would arrive at the intersection within a $(t_i + 2) - 2$ and $(t_i + 2) + 2$ sec maneuver interval were counted as left-turn conflict opportunities. However, opposing vehicles arriving at headways greater than $t_i + 4$ sec or less than $t_i$ sec were not considered in the calculation of left-turn conflict opportunities.

The negative exponential distribution was used to estimate the probability of a headway between the lower ($t_{li}$) and upper ($t_{ui}$) bound of the intersection clearance time:

$$\begin{matrix} P(h \geq t_{li}) \\ P(h \leq t_{ui}) \end{matrix} = \{e^{-N_k\lambda(t_{li})} - e^{-N_k\lambda(t_{ui})}\} \tag{3}$$

and

$$\lambda = v_k/(3{,}600N_k) \tag{4}$$

where

$t_{li}$ = lower bound of intersection clearance time on approach $i$ (sec),
$t_{ui}$ = upper bound of intersection clearance time on approach $i$ (sec),
$t_i$ = time required for left-turning vehicle from approach $i$ to clear intersection (sec),
$v_k$ = total hourly flow rate on opposing approach $k$ [vehicles per hour (vph)], and
$N_k$ = number of through lanes on opposing approach $k$.

The number of left-turn conflict opportunities on approach $i$ was expressed as

$$C_{LT_i} = \frac{E_{LT_i}[P(h \geq t_{li})]}{E_{LT_i}[P(h \leq t_{hi})]} \tag{5}$$

where $E_{LT_i}$ is the number of left-turning vehicles on approach $i$ that are exposed to opposing through traffic.

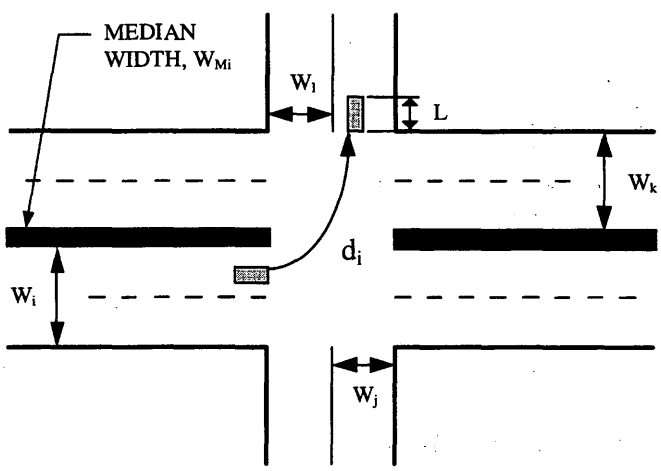



**FIGURE 2 Typical path of left-turning vehicle.**

## Rear-End Conflict Opportunity Model

The continuum model was chosen as the basis for describing the behavior of stopping traffic at a signalized intersection. As illustrated in Figure 3, traffic is assumed to arrive at a uniform rate $v_i$ on approach $i$, stop during an effective red period $r_i$, and discharge at a saturation rate $s_i$ during the effective green period $g_i$ until the accumulated queue disappears. During the red period, $r_i$, all vehicles arriving on approach $i$ will be forced to stop. Each of these vehicles, while decelerating and coming to a stop, has the possibility of colliding with the vehicle ahead of it except for the first vehicle. As the green interval begins, it will take $g_{qi}$ time for the queue of stopped vehicles to clear the intersection. The new vehicles arriving at the intersection during this portion of the green will also be forced to decelerate because of the presence of the queue at the approach and, thus, will have the potential to collide with the vehicle waiting at the end of queue. Finally, the vehicles arriving during the remaining green period, $g_{ui}$, were considered to have the potential to collide with another vehicle that is slowing to turn left or right.

The number of rear-end conflict opportunities was then calculated in three steps corresponding to the flow conditions shown in Figure 3.

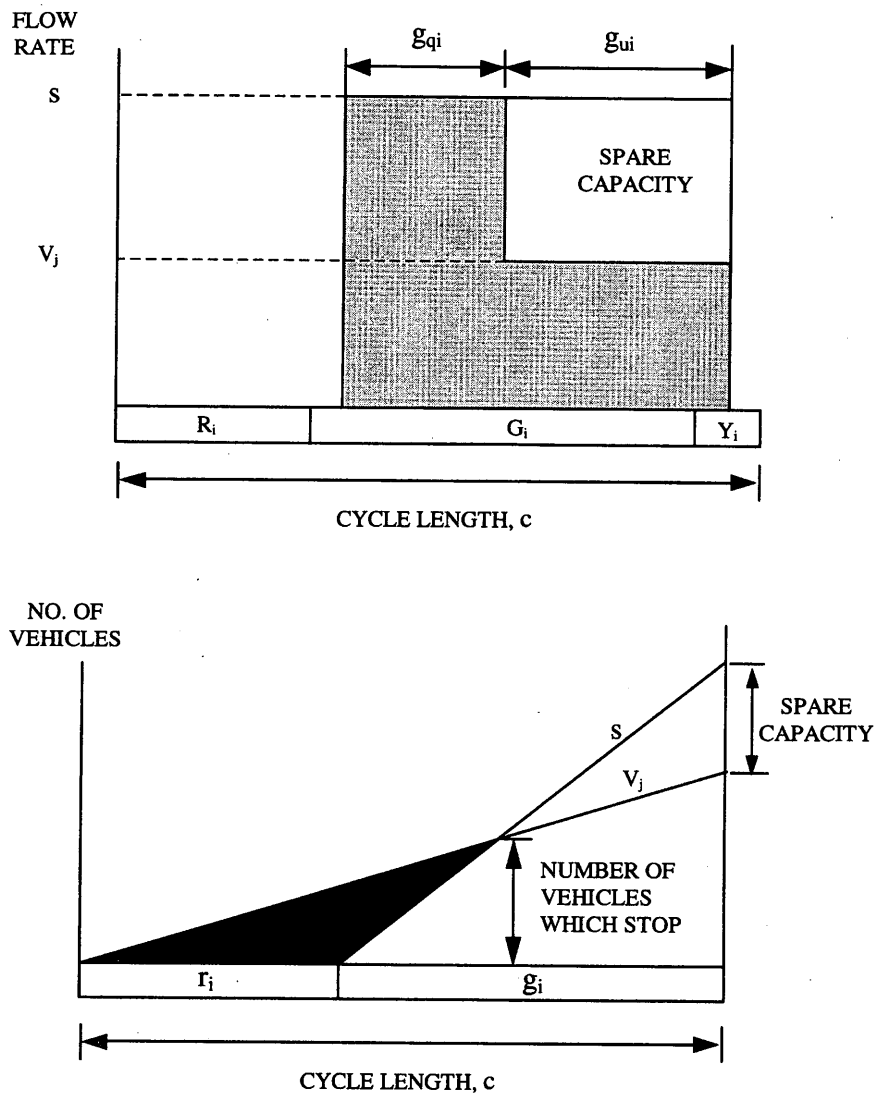



**FIGURE 3 Continuum model.**

1. Red period ($r_i$) stopping maneuver: Vehicles arriving during the red period will be forced to come to a stop and will have the opportunity to collide with the vehicle ahead of them, except for the first vehicle. The number of rear-end conflict opportunities per hour during the red period on approach $i$ can be expressed as

$$C_{RE,r} = \{(v_i r_i) - 3{,}600\}/c \quad (6)$$

where

$v_i$ = flow rate at approach $i$ (vph),
$r_i$ = red period at approach $i$ (sec), and
$c$ = cycle length (sec).

2. Green period ($g_{qi}$) stopping maneuver: As the queue begins to discharge at the saturation rate, $s_i$, the new vehicles arriving at the intersection will also be forced to decelerate until the queue has dissipated. These vehicles will join the rear of the existing queue. Each of these vehicles will thus have the potential to collide with the vehicle waiting at the end of queue. The number of rear-end conflict opportunities per hour during the green period, $g_q$, on approach $i$ was expressed as

$$C_{RE,gq} = (s_i g_{qi} - v_i r_i)/c \quad (7)$$

where $s_i$ is the saturation flow rate on approach $i$ (vph), and $g_{qi}$ is the time to clear queue on approach $i$ (sec).

3. Green period ($g_{ui}$) diverging maneuver: Vehicles moving during this portion of green period were considered to have the potential to collide with vehicles preparing to turn left or right. It was assumed that the number of rear-end conflict opportunities can be estimated as the product of the number of vehicles arriving during the remaining green period, $g_{ui}$, and the percentage of right- and left-turning vehicles on the approach.

$$C_{RE,gu} = [v_i g_{ui}](P_{LT} + P_{RT}) \quad (8)$$

where

$g_{ui}$ = portion of effective green after queue has dissipated (sec),
$P_{LT}$ = percentage of left turns (decimal fraction), and
$P_{RT}$ = percentage of right turns (decimal fraction).

## SENSITIVITY ANALYSIS OF MODELS

A sensitivity analysis of the conflict opportunity models with respect to major input variables was undertaken as a means of evaluating the general reasonability of the models. Conflict opportunities per hour were calculated for several combinations of intersection geometrics and left-turn phasing for single approach $i$ (Table 1). The following input data were used:

$v_i$ = 500 vph,
$P_{LT}$ = 10 percent,
$v_o$ = 250 vph,
$C$ = 100 sec, and
$g$ = 50 sec.

The data in Table 1 indicate that the type of signal phasing is a very important factor affecting the number of left-turn conflict opportunities. For example, for protected left-turn phasing, there will be no left-turn conflict opportunities because no vehicle is exposed to the opposing through traffic. However, for permissive left-turn phasing, left-turn conflict opportunities will arise because left-turning vehicles will be exposed to opposing traffic when they attempt to cross the intersection. For protected/permissive phasing, left-turn conflict opportunities will occur because left-turning vehicles will be exposed to opposing traffic during the permissive phase when they attempt to cross the intersection. The number of left-turn conflict opportunity counts is at its peak when all left-turns are made during a permissive green interval.

Protected left-turn phasing has the advantage of reducing left-turn conflict opportunities. Its main disadvantage, however, is that it increases rear-end conflict opportunities. Therefore, there is a trade-off between left-turn and rear-end conflict opportunities when choosing left-turn phasing. Protected phasing remains the best option for reducing left-turn conflict opportunities, whereas permissive phasing is best for reducing the rear-end conflict opportunities. The addition of exclusive turn lanes will also reduce rear-end conflict opportunities regardless of the type of signal phasing.

## DEVELOPMENT OF LOS CRITERIA

Safety-based LOS criteria for isolated signalized intersections were based on the distribution of conflict opportunities computed from the developed models. In general, the total hazard (or safety) at an intersection can be expressed as the number of accidents per time period multiplied by the average cost per accident. Because accident frequency and severity were not being modeled directly, it was assumed that the number of accidents could be approximated as the number of conflict opportunities multiplied by the average number of accidents per conflict opportunity, and that cost per accident could be accounted for by using the kinetic energy associated with the conflict opportunity as a surrogate measure. These assumptions can be expressed as follows:

Number of accidents

$$= \text{number of conflict opportunities} * \left( \frac{\text{number of accidents}}{\text{conflict opportunity}} \right)$$

and

$$\text{Cost/accident} = f\left( \frac{\text{kinetic energy of}}{\text{conflict opportunity}} \right)$$

However, left-turn and rear-end conflict opportunities are not the same in terms of expected accident occurrence. For example, the number of accidents occurring per conflict opportunity may be greater for left turns, or vice versa. As a consequence, conflict opportunities were compared with number of accidents for different types of accidents using data from the city of Madison, Wisconsin, for 15 intersections. Conflict opportunities were calculated for a typical hour during the a.m., p.m., and off-peak periods. Five years of accident data for the same periods were also collected. An analysis of these data did not yield any models with even a modest level of variance explanation. As a consequence, the relative frequency of accident occurrence per conflict opportunity was defined in terms of the ratio of the mean values for accidents per year and

**TABLE 1 Comparison of Conflict Opportunities**

| Intersection Geometrics | Signal Phasing | Number of Conflict Opportunities/Hours | |
|---|---|---|---|
| | | Left-Turn | Rear-End |
| Single Lane Approach | Permitted Left-Turn | 8.1 | 321.1 |
| 2 Lanes/Shared Left-Turn Lane | Permitted Left-Turn | 7.5 | 256.1 |
| 2 Lanes Plus Exclusive Left-Turn Lane | Permitted Left-Turn | 7.5 | 185.1 |
| 2 Lanes Plus Exclusive Left-Turn Lane | Protected Left-Turn | 0.0 | 247.2 |
| Lanes Plus Exclusive Left-Turn Lane | Protected/ Permitted Left-Turn | 3.0 | 208.6 |

conflict opportunities per hour. The resulting ratios were 0.054 and 0.00049 accidents per year per conflict opportunity for left-turn and rear-end collisions, respectively.

The level of accident severity would be expected to differ when comparing left-turn and rear-end accidents. In the absence of actual accident severity data, the kinetic energy associated with the conflicting vehicles was used as a surrogate measure of the relative severity of the collision. The total initial kinetic energy that might be dissipated in a collision was expressed as

$$E = {}^1/_2\, m_1 v_1^2 + {}^1/_2\, m_2 v_2^2 \quad (9)$$

where

$E$ = kinetic energy (kg-m$^2$/sec$^2$),
$m_1$ = mass of Vehicle 1 (kg),
$m_2$ = mass of Vehicle 2 (kg),
$v_1$ = relative speed of Vehicle 1 (m/sec), and
$v_2$ = relative speed of Vehicle 2 (m/sec).

The severity of a left-turn accident depends on the speed of opposing traffic. To account for possible collision avoidance braking, the speed of opposing traffic was assumed to be 67 percent of the typical approach speed. It was also assumed that the weight of a typical passenger car is 1362 kg (3,000 lb) and that of a typical truck is 13 620 kg (30,000 lb). The potential severity of a left-turn collision can then be calculated as

$$E = {}^1/_2\, (1{,}362P_p + 13{,}620P_t/100)v_o^2 \quad (10)$$

where

$P_p$ = passenger cars (%),
$P_t$ = trucks (%), and
$v_o$ = 67 percent of speed of opposing traffic (m/sec).

The severity of rear-end accidents also depends on the speed of the colliding vehicles. It was assumed that the speed of the lead vehicle is zero and that of the following vehicles at the time of collision is 33 percent of the approach speed. The potential severity of a rear-end collision can then be calculated:

$$E = {}^1/_2[(1{,}362P_p + 13{,}620P_t)/100]v_f^2 \quad (11)$$

where $v_t$ is 33 percent of the prevailing approach speed in meters per second. For example, if there were one left-turn conflict opportunity and one rear-end conflict opportunity with 100 percent passenger cars in the traffic stream and 64-km/hr (40-mph) approach speeds, the ratio between the left-turn and rear-end severity measures is about 4 to 1, meaning that the potential severity of a left-turn conflict is about four times greater than that of a rear-end conflict.

Finally, the measure for total hazard at an isolated signalized intersection was calculated as follows:

$$\text{Total hazard} = (\text{number of conflict opportunities} * \frac{\text{number of accidents}}{\text{conflict opportunity}} * \text{kinetic energy of conflict opportunity})_{\text{rear-end}} + (\text{number of conflict opportunities} * \frac{\text{number of accidents}}{\text{conflict opportunity}} * \text{kinetic energy of conflict opportunity})_{\text{left-turn}}$$

The 15 case study intersections were then evaluated using this expression for total hazard for 1-hr a.m., p.m., and off-peak periods. Because the resulting numbers were very large, each value was divided by 211 times the total number of entering vehicles. These values were then referred to as the total hazard rate. The range in these values served as the basis for subjectively defining six safety-based LOS (Table 2). The six levels were intended to be conceptually similar to those currently found in the HCM.

A worksheet was developed to assist in performing the necessary calculations to evaluate the safety-based LOS at an isolated signalized intersection. The format is similar to that found in the HCM and permits the safety LOS to be evaluated and compared by both lane group and intersection approach for a selected 1-hr control period.

## COMPARISON OF LOS CRITERIA

A highway capacity analysis case study presented in the traffic engineering textbook by McShane and Roess (*9*) was used to analyze the trade-off of delay versus safety LOS for a set of given conditions. Using a hypothetical four-leg intersection having two approach lanes per direction and two-phase signal control, the case study evaluates the impacts on delay per vehicle and LOS associated with the following changes in conditions:

1. Increase flow rate on one approach,
2. Add a leading protected left-turn phase for onapproach,
3. Add left-turn lanes on one of the arterials, and
4. Add leading protected left-turn phasing in conjunction with the added left-turn lanes.

For the delay-versus-safety comparison, it was also assumed that approach speeds on each arterial were 48 km/hr (30 mph).

The results of the application of the conflict opportunity models and safety-based LOS criteria to these alternatives clearly demonstrated the trade-off between achieving reduced delay and increased safety. The delay-based measures ranged from LOS B to E (13.8 to 40.2 sec/veh), whereas the safety-based measures ranged from LOS C to C (0.38 to 0.49), based on a scale from 0 to 1. The fact that the safety-based LOS measure was not as sensitive as the delayed-based measure (meaning that the safety-based LOS did not change dramatically when the input data such as geometrics, signal timing, and phasings were changed) was somewhat disappointing. However, because the two methods of intersection analysis do not use the same units to determine LOS, a judgment must be made concerning how the A-through-F LOS rating based on delay should be weighted with that of the safety-based analysis.

Two approaches might be taken with respect to how these two performance measures should be interpreted. The first approach,

**TABLE 2 Safety-Based LOS Criteria for Isolated Signalized Intersections**

| LOS | Total Hazard Rate |
|---|---|
| A | <0.10 |
| B | 0.11–0.30 |
| C | 0.31–0.50 |
| D | 0.51–0.70 |
| E | 0.71–0.90 |
| F | >0.91 |

which was not addressed within the scope of the research, would categorize intersections by total intersection volume and recognize that the safety resulting at an intersection will be strongly tied to the number of users of the intersection. Therefore, a different range of total hazard rate values and LOS criteria might be appropriate for different levels of total intersection volumes. For the case study intersection this might simply mean that a range in total hazard rate of 0.38 to 0.49 would reflect a range in LOS of, say, B to D.

A second approach to interpreting the delay-versus-safety trade-off would be to accept the values as computed. For the case study intersection, this would imply that a large change in the delay-based LOS does not produce a comparable change in the magnitude of the safety-based LOS. If this result were to hold for a wide range of intersections, it would suggest that large changes in delay do not necessarily produce dramatic changes in safety.

## CONCLUSIONS AND RECOMMENDATIONS

Because of a lack of resources and the fact that the research was exploratory, no additional work was undertaken. Nevertheless, it was concluded that the HCM delay-based LOS criteria are probably not a good surrogate for the level of safety offered at a signalized intersection. The methodology developed for evaluating the safety-based LOS at isolated signalized intersections is preliminary and requires further testing and development. However, it is believed that the results offer a useful starting point for further research that hopefully would produce an implementable tool for practicing engineers.

## REFERENCES

1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
2. Hauer, E., and J. Lovell. New Directions for Learning About the Safety Effect of Measures. In *Transportation Research Record 1068,* TRB, National Research Council, Washington, D.C., 1986.
3. Council, F. M., J. R. Steward, D. W. Reinfurt, and W. W. Hunter. *Exposure Measures for Evaluating Highway Safety Issues,* Vol. 1. Final Report. FHWA/RD-83/088. Highway Safety Research Center, University of North Carolina, Chapel Hill, 1988.
4. Plass, M., and W. D. Berg. Evaluation of Opportunity-Based Accident Rate Expressions. In *Transportation Research Record 1111,* TRB, National Research Council, Washington, D.C., 1987.
5. Glauz, W. D., and D. J. Migletz. *NCHRP Report 219: Application of Traffic-Conflict Analysis at Intersections.* TRB, National Research Council, Washington, D.C., 1980.
6. Ha, T.-J. Development of Safety-Based Level-of-Service Criteria for Isolated Signalized Intersections. Ph.D. thesis. University of Wisconsin, Madison, 1994.
7. Blunden, W. R., C. M. Clissold, and R. B. Fisher. Distribution of Acceptance Gaps for Crossing and Turning Maneuvers. *Proc., Australian Road Research Board,* Vol. 1, 1962, pp. 188–205.
8. McNeil, D. R., and J. H. T. Morgan. Estimating Minimum Gap Acceptance for Merging Motorists. *Transportation Science,* Vol. 2, No. 3, 1968, pp. 265–277.
9. McShane, W. R., and R. P. Roess. *Traffic Engineering,* Prentice-Hall Polytechnic Series in Transportation, Englewood Cliffs, N.J., 1990.

*Publication of this paper sponsored by Committee on Methodology for Evaluating Highway Improvement.*