TRANSPORTATION RESEARCH RECORD

538

Capacity and Quality of Service

7 reports prepared for the 54th Annual Meeting of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD

NATIONAL RESEARCH COUNCIL

Washington, D. C., 1975

Transportation Research Record 538
Price $3.40
Edited for TRB by Marianne Cox Wilburn

Subject areas
53 traffic control and operations
55 traffic measurements

Transportation Research Board publications are available by ordering directly from the Board. They are also obtainable on a regular basis through organizational or individual supporting membership in the Board; members or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Academy of Sciences, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

These papers report research work of the authors that was done at institutions named by the authors. The papers were offered to the Transportation Research Board of the National Research Council for publication and are published here in the interest of the dissemination of information from research, one of the major functions of the Transportation Research Board.

Before publication, each paper was reviewed by members of the TRB committee named as its sponsor and accepted as objective, useful, and suitable for publication by the National Research Council. The members of the review committee were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the subject concerned.

Responsibility for the publication of these reports rests with the sponsoring committee. However, the opinions and conclusions expressed in the reports are those of the individual authors and not necessarily those of the sponsoring committee, the Transportation Research Board, or the National Research Council.

Each report is reviewed and processed according to the procedures established and monitored by the Report Review Committee of the National Academy of Sciences. Distribution of the report is approved by the President of the Academy upon satisfactory completion of the review process.

LIBRARY OF CONGRESS CATALOGING IN PUBLICATION DATA

National Research Council. Transportation Research Board.
Capacity and quality of service.

(Transportation research record; 538)

1. Traffic engineering–Congresses. I. National Research Council. Transportation Research Board. II. Title. III. Series.

TE7.H5 no. 538 [HE332] 380.5'08s [388.3'14] 75-25527

ISBN 0-309-02389-0

CONTENTS

FOREWORD

The 7 papers in this RECORD reflect the broad interest of the sponsor, the Committee on Highway Capacity and Quality of Service, by covering capacity of pedestrian facilities and measurement of flow factors related to capacity and quality of service. Research designed to develop knowledge of weaving and signalized intersections also is reported and can be useful as the committee advances toward the updating of the Transportation Research Board's Highway Capacity Manual (HCM). Planners, designers, and operations and traffic engineers will find valuable information here that will be useful in addressing current operational problems.

Pushkarev and Zupan outline speed, flow, density, and levels of comfort for pedestrians on walkways, stairways, escalators, and moving walks and at street intersections. Walkway service levels in units of area per pedestrian are described in flow rates of pedestrians per minute per foot (meter) of walkway width. The application of these requirements to improvement of existing pedestrian facilities is discussed.

In a continuation of work in the National Cooperative Highway Research Program, Pignataro et al. developed a new procedure for the design and analysis of weaving sections. Current practices based on the HCM, as well as different ways of applying the new procedure, are discussed.

Messer and Berry studied the capacity of signalized diamond interchanges as affected by 4-phase signal timing plans and the spacing of ramp intersections. They describe a method for assessing interchange designs in terms of quality of service by using computer analysis, and they apply the method to some typical cases.

Reilly, Dommasch, and Jagganath tested a New Jersey method against the 1965 HCM method for estimating capacity of signalized intersection approaches. Their work was undertaken because of evidence that the HCM approach yielded erroneous results in many cases. They concluded that it is preferable to measure rather than estimate capacity for existing intersection approaches. They describe a method of determining capacity from field measurements.

Toronto researchers Rach et al. evaluated 4 off-line signal optimization techniques (SIGOP, TRANSYT, Combination Method, and a SIGRID-based, preferential street program) in both suburban area and central area settings. Network travel time, delay, and stop and volume data were collected and analyzed on both a network and link-by-link paired comparison basis. The results showed that the Combination Method provided slightly better on-street performance, but any of the 4 methods can provide reasonable signal network settings.

Gardner, Soliday, and Williamson report their development and test of a photooptic system to continuously measure and record lateral lane position of a test vehicle. The device, which is mounted on the vehicle, detects vehicle position with respect to the pavement shoulder line and is said to have been sensitive to differences in driving behavior when it was tested through shallow right and left curves.

In the final paper, Byrne and Roberts report a decline in mean vehicle speeds on an Interstate bridge from 61.0 mph (98 km/h) in 1973 to 54.5 mph (87 km/h) in 1974. This effect is attributed to the nationwide lowering of speed limits to 55 mph (89 km/h). Standard deviation of the distribution also was reduced from 9.3 mph (15 km/h) to 6.0 mph (10 km/h). The authors suggest that this effect was a factor contributing to the reduction in accidents after limits were reduced.

CAPACITY OF WALKWAYS

Boris Pushkarev and Jeffrey M. Zupan, Regional Plan Association

Flow and space standards for walking facilities and their application are focused on. Consistency is shown in a comparison of work done by various researchers on speed, flow, and density relationships. Levels of comfort at different fractions of maximum capacity are defined. The effect of short-term fluctuation of flow, known as platooning, is evaluated and related to average conditions. Levels of service for platooning are postulated based on available space per pedestrian. Key flow rates for defining walkway service levels are 2, 4, 6, and 10 pedestrians/min/ft (7, 13, 20, and 33 pedestrians/min/m) of walkway width corresponding to 130, 65, 40, and 24 ft^2/pedestrian (12.08, 6.04, 3.72, and 2.23 m^2/pedestrian) respectively.

•CAPACITY of pedestrian facilities, like capacity of vehicle facilities, usually means maximum ability of a facility to accommodate a flow, but, more often than not in vehicle traffic design, operation at maximum capacity is undesirable. So as not to establish imminent congestion as a design standard, researchers have defined levels of service that characterize the quality of traffic flow at various fractions of maximum capacity. Similarly, several pedestrian levels of service can be defined by indicating what kind of behavior is possible, or impossible, at various degrees of spaciousness or crowding. The selection of any particular level of service as a desirable design standard is, to a large extent, a matter of judgment and policy.

SPACE RELATED TO SPEED AND FLOW

Pedestrian travel requires enough room to allow for pacing and a buffer zone large enough to permit anticipating potential collisions and taking evasive action. For example, because of the angle the human eye encompasses, one has to be at least 7 ft (2.1 m) away from someone to be seen from head to toe and to have one's speed and direction of movement accurately judged. Pedestrians have been found to take evasive action anywhere from 2 to 17 ft (0.6 to 5.2 m) ahead of a stationary or moving obstacle. The longer the distance is, the less violent the evasion and the less likely a collision is. The spacing between pedestrians, like the spacing between vehicles, is related to the speed at which the objects are moving. More space is required for faster movements. The relationship of space requirements (density), speed of movement, and rates of flow in pedestrian streams has been studied by a number of investigators. Among the more recent ones are Fruin (1), Oeding (2), Older (3), and Navin and Wheeler (4). Their findings are generally consistent with those of several other researchers (5, 6, 7, 8).

The traditional equation describing traffic flow is (9)

$$\text{Flow} = \text{Speed} \times \text{Density} \tag{1}$$

where

flow = number of moving objects crossing a unit of channel width in a unit of time,
speed = number of units of distance the moving objects pass in a unit of time, and
density = number of moving objects per unit of channel area.

When the units by which channel area is measured are relatively small, such as square feet (m^2), density becomes an inconvenient concept that forces us to deal with fractions of pedestrians. Moreover, a density scale shrinks rapidly in the range in which we are most interested. The range is the one that has less than 0.1 pedestrians/square ft (1.0 pedestrians/m^2) and where varying degrees of comfort prevail. So the reciprocal of density, or available space per pedestrian, is a more useful unit for trying to arrive at comfort criteria. With that in mind, and by adding dimensions, we can rewrite equation 1 as follows:

$$\text{Space} = \frac{\text{Speed}}{\text{Flow}} \qquad (2)$$

The relationship between speed and flow can be approximated by a parabolic curve that is familiar from motor-vehicle flow analysis. Figure 1 shows a family of 5 speed-flow curves abstracted from measurements by the investigators cited previously and converted to common units.

The formula for the parabolas in Figure 1 is a quadratic equation:

$$\text{Speed} = \frac{A \pm A^2 - 4B\ \text{Flow}}{2} \qquad (3)$$

where A and B are constants. These constants can be calculated for any set of observations statistically by means of the least squares technique, or they can be estimated by inspection of a plot of speed versus density. Plotting density rather than space per pedestrian is useful in this case because the resulting relationship can be represented as a linear one. The straight-line form has been shown to represent a reasonable approximation of reality. It takes the form of the equation

$$\text{Speed} = A - B \times \text{Density} \qquad (4)$$

A represents the intercept on the y axis (speed in this case), and B represents the slope of a straight line or the rate at which speed declines with density as shown in Figure 2. The meaning of these 2 constants also can be interpreted as follows: A represents the theoretical speed attained by a traffic stream under conditions of completely free flow and an unlimited amount of space per pedestrian; B is a factor that, when divided by A, yields the theoretical minimum space allocation per pedestrian at a point where all movement in a traffic stream stops and speed is zero. The constants A and B for the curves in Figure 1 and 2 are given in Table 1.

To determine the maximum or capacity pedestrian flow and at what speed it occurs, all we have to do is find the maximums on the curves defined by equation 4. These calculated maximums are given in Table 2 along with extremes observed by the different investigators.

It is evident that the findings of Older (3), Oeding (2), and Fruin (1) on maximum pedestrian flow are in close agreement. In fact, Fruin's (1) calculated maximum of 24.7 pedestrians/min/ft (81 pedestrians/min/m) of walkway width at a speed of 134 ft (40.8 m)/min falls exactly halfway between the maximums derived by Older (3) and Oeding (2). The extremes observed by Older (3) in England and Oeding (2) in Germany are also in close agreement although speeds differ. These extreme flow rates are high and come close to those attainable in highly organized military formations as given in Table 3. The behavior of Navin and Wheeler's (4) student population is different; it has greater spacing between individuals and, accordingly, a lower flow at comparable speeds. One may speculate that the higher interpersonal distances adopted by the

Figure 1. Speed-flow relationships.

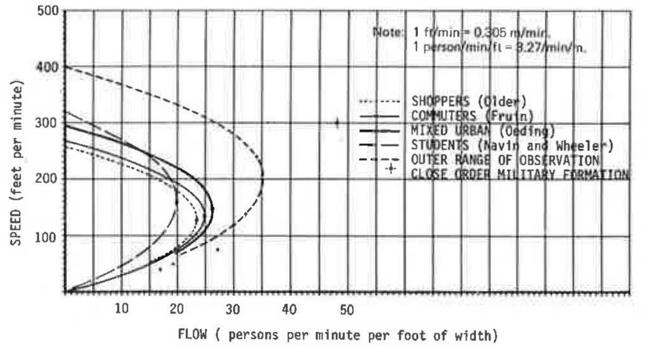


Figure 2. Speed-density relationships.

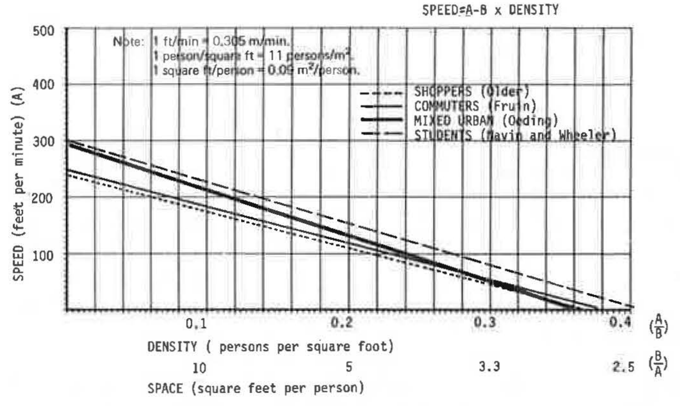


Table 1. Coefficients of pedestrian flow equations.

Type of Flow	Source	A (ft/min)	B	B/A (square ft)
Shoppers, average[a,b]	Older (3)	258	714	2.77
Commuters, average[a]	Fruin (1)	267	722	2.70
Mixed traffic, average[c]	Oeding (2)	295	835	2.83
Students, average[c]	Navin and Wheeler (4)	320	1,280	4.00
Mixed traffic, outer boundary[c]	Oeding (2)	400	1,132	2.83

Note: 1 ft/min = 0.305 m/min. 1 square ft = 0.09 m².

[a]Calculated.

[b]Extreme observations suggest a minimum space allocation of 2.1 square ft (0.2 m²)/pedestrian at zero speed.

[c]Estimated.

Table 2. Maximum pedestrian flow.

		Maximum Flow (pedestrians/ft)		Mean Speed at Maximum Flow (ft/min)	
Type of Flow	Source	Calculated Average	Observed Extreme	Calculated Average	Observed Extreme
Shoppers	Older (3)	23.3	33.0	129	170
Commuters	Fruin (1)	24.7	–	134	–
Mixed traffic	Oeding (2)	26.1	34.0	148	246
Students	Navin and Wheeler (4)	20.0	26.4	160	240
Close-order military drill		–	48.0	–	300

Note: 1 pedestrian/ft = 3.27 pedestrians/m. 1 ft/min = 0.305 m/min.

Table 3. Space per pedestrian at maximum flow.

Type of Flow	Source	Maximum Flow (pedestrians/ min/ft)	Space per Pedestrian at Maximum Flow (square ft)
Average			
Students	Navin and Wheeler (4)	20.0	8.0
Shoppers	Older (3)	23.3	5.5
Commuters	Fruin (1)	24.7	5.4
Mixed traffic	Oeding (2)	26.0	5.5
Extreme			
Students	Navin and Wheeler (4)	26.4	9.1
Shoppers	Older (3)	33.0	5.2
Mixed traffic	Oeding (2)	34.0	7.2
Close-order military drill		48.0	6.3

Note: 1 pedestrian/min/ft = 3.27 pedestrians/min/m. 1 square ft = 0.09 m².

students are more representative of a comfortable situation than are the close spacings found by Older (3), Oeding (2), and Fruin (1) in forced downtown flows.

To be able to make an evaluation for comfort, we must take a look at the relationship between flow and space per pedestrian. Following equation 3, if we take the speed at any point on the curves in Figure 1 and divide it by the flow at that point, we obtain the amount of space available per pedestrian at that point. For example, at a speed of 200 ft (61 m)/min and a flow rate of 20 pedestrians/min, the average space allocation is 10 square ft (0.93 m^2)/pedestrian. In this manner, the speed-flow diagrams shown in Figure 1 are converted into the flow-space diagrams shown in Figure 3. The formula for the flow-space curves is:

$$\text{Flow} = \frac{A \times \text{Space} - B}{\text{Space}^2} \qquad (5)$$

where A and B are the constants given in Table 1. The available space per pedestrian at maximum flow is given in Table 3.

It is apparent from Figure 3 and Table 3 that all the different observations of maximum flow previously listed fall in a very narrow range of density—that in which space allocation varies between 5.2 and 9.1 square ft (0.48 and 0.85 m^2)/pedestrian. As space is reduced to less than 5 square ft (0.46 m^2)/pedestrian, flow rate declines precipitously; all movement comes to a standstill at space allocations between 2 and 4 square ft (0.2 to 0.4 m^2) as the data given in Table 1 have shown.

Thus, if our objective is to maximize pedestrian flow, regardless of speed or comfort, the space allocation per pedestrian should be between 5.2 and 9.1 square ft (roughly 0.5 to 0.9 m^2). Letting space allocations drift below that level will lead to a crush; the crowd will grow in size as long as the number of incoming pedestrians is greater than what the bottleneck can release.

On the other hand, increasing space allocations above 10 square ft (0.9 m^2)/pedestrian will lead to declines in flow. It can be deduced from Figure 3 that at 40 square ft (3.7 m^2)/pedestrian, the flow rates are, depending on which curve one chooses, between 24 and 32 percent of maximum flow. At 100 square ft (9.3 m^2)/pedestrian, the flow rates are down to about 10 percent of maximum flow. Our concern is, of course, with quality of flow, not quantity. This leads us to look at average speed in relation to space per pedestrian.

Going back to equation 2, if we multiply the flow at any point of Figure 1 by the space per pedestrian at that point, we obtain the speed at which the flow is occurring. Thus, the flow-space diagram in Figure 3 can be transformed into the speed-space diagram shown in Figure 4. The equation of the speed-space curve is

$$\text{Speed} = A - \frac{B}{\text{Space}} \qquad (6)$$

A and B again are constants from Table 1. The form of equation 6 makes it clear why A equals B divided by the space allocation at zero speed. It also makes it clear that, as space per pedestrian increases toward infinity, speed increasingly approaches A, previously defined as the theoretical maximum speed at free flow for a given type of traffic stream. Thus, for example, at 100 square ft (9.3 m^2)/pedestrian, the average speed is between 96 and 97 percent of the theoretical speed at an infinite space allocation per pedestrian. At 40 square ft (3.7 m^2)/pedestrian, average speed drops to between 90 and 93 percent of this theoretical level. From then on, the reduction becomes sharper, and at 11 square ft (1 m^2)/pedestrian, average speed is down to between 64 and 75 percent of the theoretical maximum. In the range where flow is maximized to somewhere between 9 and 5 square ft (0.9 and 0.5 m^2) /pedestrian, speed drops drastically to between 27 and 50 percent of its theoretical level and then keeps declining to reach zero at space allocations between 2 and 4 square ft (0.18 and 0.36 m^2)/pedestrian.

Reductions in average speed come about as available space per pedestrian shrinks. Fewer people have the freedom to select their own rate of movement because of the interference from others in the traffic stream. The fastest walkers are slowed down first, but, eventually, even slow walkers are affected. Thus, the range of observed speeds shrinks as space per pedestrian is reduced. Some indication of this is given by the dotted lines shown in Figure 4 that portray the upper and lower limit of speeds observed by Oeding (2).

SERVICE LEVELS

Studies concerning the distribution of pedestrian speeds under conditions of free choice have been carried out by numerous observers, among them Fruin (1), MacDorman (10), Gehl (11), and Hoel (12). Biological limits govern both how fast and how slowly people can walk. The various investigators agree that virtually no one will voluntarily select speeds faster than 400 ft (122 m)/min, or slower than 145 ft (44 m)/min. Speeds below that range can be classified as shuffling. Oeding (2) points out that speeds in the shuffling range do not occur under unobstructed conditions because they require cramped movements, which are unnatural in terms of body balance.

On this basis we may note two things. First, average speeds on all the curves in Figure 4 are depressed into the unnatural shuffling range of less than 150 ft (46 m)/min at space allocations between 6 and 8 square ft (0.56 and 0.74 m^2)/pedestrian. Second, those who choose to walk at the minimum speed of about 150 ft (46 m)/min when space per walker is ample cannot maintain even that speed when space shrinks below 15 to 18 square ft (1.4 to 1.7 m^2). The fast walkers lose the ability to maintain their chosen speed as space drops below 30 to 40 square ft (2.8 to 3.7 m^2)/pedestrian.

There are other indicators of congestion, besides the inability to maintain a freely selected speed. An important one is the inability to choose one's path freely across the traffic stream. Fruin (1) studied pedestrian crossing conflicts in relation to available space per pedestrian. He defined conflicts as "any stopping or breaking of the normal walking pace due to a too close confrontation with another pedestrian" that requires adjustments in speed or direction to avoid collision. He found such situations inevitable when flow is dense—less than 15 square ft (1.4 m^2)/pedestrian. As gaps between pedestrians widen, crossing movements become easier and the probability of conflict drops to between 65 and 50 percent. However, the probability of conflict does not drop to zero until the space allocation reaches about 45 square ft (4.2 m^2)/pedestrian.

A related indicator is ability to pass slow-moving pedestrians, which Oeding (2) found to be relatively unrestricted at space allocations of more than 36 square ft (3.3 m^2)/pedestrian. He found the ability to pass to be considerably restricted in the range between 18 and 36 square ft (1.7 to 3.3 m^2)/pedestrian. At lower space allocations, he found passing to be possible only by physically pushing the slow-walking person aside.

Finally, an important consideration is the ability to maintain flow in the reverse direction. All of the data presented here, except some extreme observations by Oeding (2), refer to bidirectional flow. Bidirectional flow is not substantially different from 1-directional flow as long as the directional distribution is relatively balanced. Pedestrians spontaneously form directional streams that minimize conflict with the opposing flow. Each stream occupies a share of the walkway that is proportional to its share in the total flow, and reduction in speed or capacity is minimal—Fruin (1) found it to be less than 6 percent under maximum flow conditions. However, Navin and Wheeler (4) have shown that reduction in capacity increases as directional imbalance increases. Thus for directional distributions of 25 to 75 or better, reduction in capacity approaches 10 percent. For a 10 to 90 distribution, reduction in capacity rises to 14.5 percent, given a space allocation of 10 square ft (0.93 m^2)/pedestrian. As space allocations are reduced, maintaining a small flow in the opposite direction becomes more difficult (a problem acute on some rapid transit stairways) and effect on capacity becomes more pronounced. A summary of the different kinds of pedestrian behavior that are possible or impossible at different densities is given in Table 4 (1, 2). Fruin (1) brands space

Figure 3. Flow-space relationships.

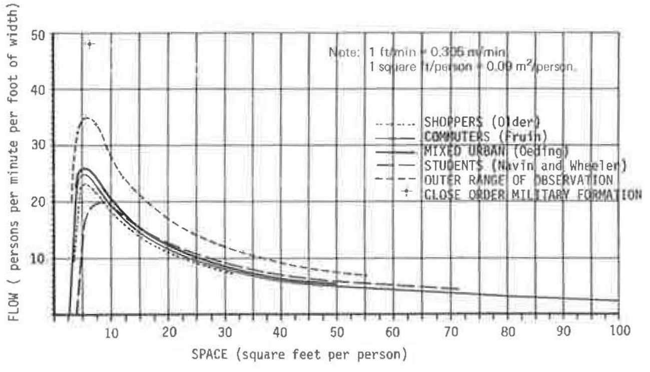


Figure 4. Speed-space relationships.

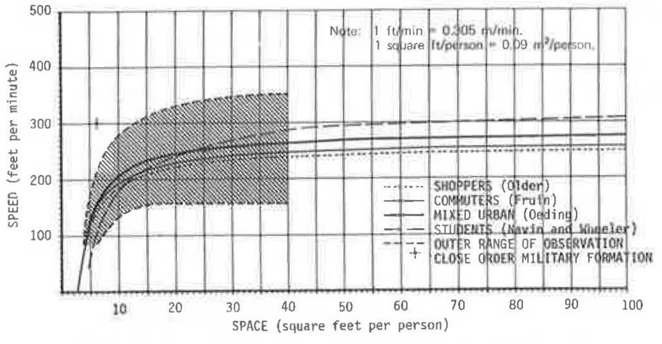


Table 4. Pedestrian behavior related to available space.

Average Area per Person (square ft)	Flow	Average Speed	Choice of Speed	Crossing or Reverse Movement	Conflicts	Passing
2 to 5	Erratic, on the verge of complete stoppage	Shuffling only	None	Impossible	Physical contact unavoidable	Impossible
5 to 7	Attains a maximum in traffic streams under pressure	Mostly shuffling	None, movement only with crowd	Most difficult	Physical contact probable, conflicts unavoidable	Impossible
7 to 11	Attains a maximum in more relaxed traffic streams	About 67 percent of free flow	Practically none	Severely restricted with collisions	Physical contact probable, conflicts unavoidable	Impossible
11 to 15	65 to 80 percent of maximum capacity	About 75 percent of free flow	Restricted, constant adjustments to gait necessary	Severely restricted with conflicts	Unavoidable	Rarely possible without touching
15 to 18	56 to 70 percent of maximum capacity	About 80 percent of free flow	Restricted except for slow walkers	Restricted with conflicts	Highly probable	Rarely possible without touching
18 to 25	Roughly 50 percent of maximum capacity	More than 80 percent of free flow	Partially restricted	Possible with conflicts	Highly probable	Difficult without abrupt maneuvers
25 to 40	Roughly 33 percent of maximum capacity	Approaching free flow	Occasionally restricted	Possible with occasional conflicts	Probably 50 percent of the time	Possible with interference
More than 40	20 percent of maximum capacity or less	Virtually as chosen	Virtually unrestricted	Free	Maneuvering needed to avoid conflicts	Free with some maneuvering

Note: 1 square ft = 0.09 m².

allocations of less than 5 square ft (0.5 m^2) completely unacceptable; Oeding (2) brands those of less than 7 square ft (0.66 m^2) completely unacceptable. Both (1, 2) agree that space allocations below 10 or 11 square ft (1 m^2) can easily lead to flow stoppages and a buildup of crowds; pedestrians should not be required to endure this degree of congestion. Yet it is at this level that the maximum flow of 25 to 28 pedestrians/min/ft (82 to 93 pedestrians/min/m), which is frequently accepted as design capacity, occurs. Oeding (2) and Fruin (1) point out, however, that such crush loads can, on occasion, be difficult to avoid in short-term bulk situations, such as when a crowd leaves a sports stadium.

Oeding (2) calls space allocations between 18 and 36 square ft (1.7 and 3.3 m^2)/pedestrian tolerable. Fruin (1) subdivides the range between 15 and 35 square ft (1.4 and 3.3 m^2) into 2 service levels, B and C, which he recommends conditionally for transportation terminals and similar heavily used facilities. Fruin's (1) B and C levels represent respective flow volumes of up to 10 and 15 pedestrians/min/ft (33 and 49 pedestrians/min/m) of walkway. Oeding (2) cites similar flow volumes—14 pedestrians/min/ft (45 pedestrians/min/m) for shoppers and 18 pedestrians/min/ft (60 pedestrians/min/m) for commuters—as the upper limit of his tolerable range.

Although they are tolerable occasionally in tight circulation areas, space allocations of 15 to 35 square ft (1.4 to 3.3 m^2)/pedestrian still impose serious restrictions on pedestrian flow, as evident from the data given in Table 4.

Both Oeding (2) and Fruin (1) characterize only space allocations greater than 35 or 36 square ft (3.3 m^2)/pedestrian as permitting free flow. Stramentov, based on observations in Moscow, suggests in effect a similar range between 25 and 55 square ft (2.3 and 5.1 m^2)/pedestrian. Thus 40 square ft (3.7 m^2)/pedestrian, corresponding to a flow rate of 6 pedestrians/min/ft (20 pedestrians/min/m) of walkway width, can be accepted as a reasonable threshold, beyond which pedestrian behavior no longer is constrained physically by the traffic stream.

One should note, however, that a space allocation of 40 square ft (3.7 m^2)/pedestrian, although it allows a relatively free choice of speed and direction of movement, does not really represent an uncrowded situation. The lateral spacing adopted by people under conditions approaching free flow was found by Fruin (1) to be roughly 3.5 ft (1 m); if one assumes this, the longitudinal spacing on the threshold of Oeding's (2) and Fruin's (1) comfortable density is a little over 11 ft (3.5 m). At such close spacing, people, although they are able to avoid physical collisions or restrictions in speed, are acutely aware of others in the traffic stream and must continuously interact with them.

For example, Wolff (13) points out that at distances of less than 15 ft [which represents a space of at least 60 square ft (5.6 m^2)] people normally do not walk behind each other but rather walk in a checkerboard pattern, looking over the shoulder of the person in front. Thus, if any person in a group of walkers changes his or her lateral position, others are forced to accommodate to maintain the checkerboard spacing. A similar phenomenon also can be observed in the lateral direction. People prefer not to walk side-by-side with a stranger for any length of time, and they either accelerate or slow down if someone else is walking next to them. Navigating in the fluid, dense pedestrian stream thus requires constant attention and interaction with others. Psychologists suggest that it is this kind of effort that makes walking in crowded places tiresome, especially if other walkers are uncooperative, as shoppers with bags tend to be.

Exactly at what point flow on a walkway becomes sufficiently sparse to induce no stress is a good subject for further study. Only fragmentary pieces of evidence are available. Wolff (13) shows that the distance at which evasive action is taken in the face of an imminent collision increases from about 2 ft (0.6 m) at a space allocation of 40 square ft (3.7 m^2)/pedestrian to an average of about 7 ft (2.1 m) at 100 square ft (30 m^2)/pedestrian and then stays constant, suggesting that evasion at that distance may be sufficiently smooth. However, Wolff (13) cautions that the latter distance may have been foreshortened by the conditions of the experiment; he found 16.5 (5 m) to be the distance at which evasive maneuvers from fixed objects began.

Another method for analyzing the quality of flow in the lower density range, into which neither Fruin (1) nor Oeding (2) ventured, is the maximum pedestrian technique. The maximum pedestrian sets out to walk as fast as he or she can. Both the speed and the number of conflicts (sharp evasive maneuvers or near collisions) that he or she encounters at different flow rates are observed. In one experiment on Fulton Street in Brooklyn (14), the maximum pedestrian generally was unable to walk faster than 300 ft (91 m)/min, which is at the threshold of Oeding's (2) and Fruin's (1) comfortable density [about 5 pedestrians/min/ft (16 pedestrians/min/m)], and encountered an average of 12 conflicts/250 ft (76 m) of walking distance. As average hourly flow declined to less than 3 pedestrians/min/ft (10 pedestrians/min/m), the number of conflicts declined linearly to about 4, and the maximum possible speed increased to 380 ft (116 m)/min, at an average space allocation on the order of 90 square ft (8.4 m^2)/pedestrian in the traffic stream.

Qualitative observations as a part of this study, both in transit corridors and on outdoor walkways, suggest that a space allocation on the order of 130 square ft (12 m^2)/pedestrian may be a reasonable minimum limit for truly unimpeded walking; only negligible influence will come from the traffic stream. That represents a flow rate of 2 pedestrians/min/ft (6.5 pedestrians/min/m) of walkway, which feels comfortable yet retains a busy appearance. However, involuntary bunching or platooning still occurs at this flow rate and does not disappear until flow falls below 0.5 pedestrians/min/ft (1.6 pedestrians/min/m) and space allocation increases to roughly 500 or 600 square ft (50 m^2)/pedestrian. When space allocations are beyond this range, one can no longer talk about pedestrian flow, but only about isolated pedestrians.

Let us now define walkway width. Some people in the past have described a pedestrian "lane" as a strip as narrow as 22 in. (56 cm) (15). However, the lane is irrelevant to capacity calculations. The lane can only be meaningful if one wishes to calculate how many people can walk abreast or pass each other simultaneously along a walkway of a given width. The lateral spacing to avoid interference with a passing pedestrian, according to Oeding (2) and the observations of this study, is at least 30 in. (75 cm). Pedestrians who know each other and are walking together will walk as close as 26 in. (65 cm) center-to-center; at this distance there is considerable likelihood of touching. Lateral spacing of less than 24 in. (60 cm) between strangers occurs, as Fruin (1) has shown, under jammed conditions, when there is less than about 5 square ft (0.5 m^2)/pedestrian. [In contorted evasive maneuvers on narrow stairs, people, if necessary, can squeeze by in about 20 in. (50 cm) of space.] Under normal conditions, even the 2.5-ft (0.75-m) lateral spacing is tolerated only momentarily to pass a person or to walk alongside a person through a stairway. Otherwise, a spacing of 3 to 4 ft (0.9 to 1.2 m) or more is adopted by walking in a checkerboard pattern.

Multiples of about 2.5 ft (0.75 m) can be used to calculate clear walkway width for a given number of people to walk abreast in a voluntary group and to be able to pass a group, but clear walkway width deserves more emphasis. People shy away from walking along the very edge of a curb or against building walls. Therefore, dead space along the edges of a walkway must be excluded from effective width when one calculates design flow. Also excluded must be a strip preempted by physical obstructions, such as light poles, mail boxes, and parking meters, although their exact effect on pedestrian flow has not been sufficiently investigated. The area preempted by standing pedestrians also is not available for walking.

In a study of shopping walkways in Leeds, O'Flaherty and Parkinson (8) found that a speed-density relationship calculated on the basis of curb-to-wall sidewalk width could not be meaningfully converted into a flow-space relationship because of a large number of standing pedestrians who occupied space but did not contribute to flow. Only by subtracting the space occupied by those standing from total sidewalk space could a useful relationship be obtained. The width preempted by window shoppers was between 1.6 and 2.5 ft (0.5 and 0.75 m), and that by standees at a bus stop, about 3.6 ft (1.1 m). The implicit space allocations per window shopper were roughly between 5 and 7 square ft (0.5 and 0.7 m^2). These findings are in agreement with the lateral clearances from building walls suggested by Oeding (2). The clearance from the curb suggested by him is 1 to 1.5 ft (0.3 to 0.5 m).

On the basis of these observations we can now proceed to summarize the characteristics of pedestrian flow at different levels of spaciousness. This is given in Table 5, which goes beyond the range investigated by Oeding (2) and Fruin (1); the boundaries of the various conditions are slightly adjusted for arithmetical convenience. These data in Table 5 assume that the pedestrian flow is even, or homogeneous, in time. The flow rate is expressed in terms of 1 min and should not be extrapolated to longer periods of time until the considerations presented in the next section are taken fully into account.

Essentially no interaction among pedestrians occurs at the open flow level. At the unimpeded level some bunching begins to occur, but an individual is generally not influenced by others in the traffic stream, and walking is carefree. At the impeded level progress is possible only by constant interaction with the movement of others. At the constrained level interaction turns into physical restrictions on freedom of movement, speed is limited, and conflicts occur. The crowded level is rarely reached except for short periods of time on urban sidewalks and is more typical of heavily used transportation terminals where movement may still be fluid but has a lot of friction and depressed speed.

SPACE FOR PLATOONS

To have defined possible flow rates at different levels of pedestrian comfort will do us little good unless we know what time spans these rates should be applied to. Flow is uneven so a flow rate of 10 pedestrians/min does not necessarily equal 600 pedestrians/hour.

Platoon Effect

A good picture of minute-by-minute variation can be obtained from data collected by Okamoto and Beck (16) in their time-lapse photography studies of 2 walkways in Lower Manhattan. These data, shown in Figure 5, cover the morning rush hour on Nassau Street and the morning rush hour and lunch hour at the entrance to the Chase Manhattan Plaza. The maximum 15-min flow rate at the Nassau Street location averaged 10 pedestrians/min/ft (32.8 pedestrians/min/m). The maximum 15-min flow rate at the Chase Plaza entrance during the morning rush hour averaged 1.4 pedestrians/min/ft (4.6 pedestrians/min/m); during the lunch hour it averaged 1.9 pedestrians/min/ft (6.2 pedestrians/min/m).

The diagrams indicate that flow during 1 minute can, on occasion, be more than twice as high as flow during the next minute, particularly when overall volume is low. Even during the peak 15-min periods, 1 minute can be 1½ times different from another minute even during what would appear to be, on the average, an unimpeded flow to the plaza and a constrained flow on Nassau Street. Relating the scatter in the diagrams to the 15-min average, we find that the highest minute within a 15-min period exceeds the average by at least 20 and up to 75 percent. The 3rd highest minute exceeds the average by at least 10 and up to 30 percent. Even the 7th highest minute can be up to 20 percent higher than average. In general, at least 6 and up to 9 min of every 15-min period experience an above-average rate of flow. As a result, more than 50 percent (up to 73 percent) of the people walk during minutes when flow exceeds the 15-min average. For them, the flow on Nassau Street is no longer constrained, but rather is crowded, and the lunch-hour flow in the Chase Manhattan Plaza entrance is not unimpeded but impeded.

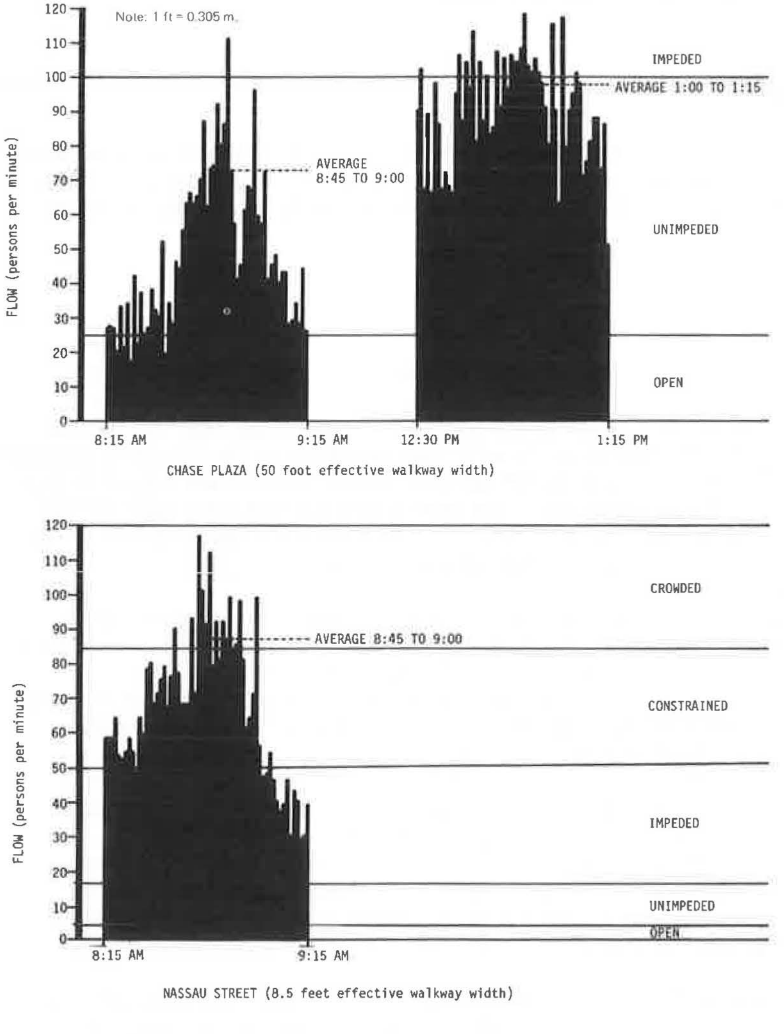
These findings are supported by manual minute-to-minute counts in Midtown Manhattan that fall in the same range. It is clear that any facility designed for the average flow in a 15-min period will be underdesigned for a sizable portion of the pedestrians using it. At the same time, it would be extravagant to design a facility for 1 peak minute that may be 150 percent of the average but that may occur with only a 1 or 2 percent probability. To resolve that dilemma and to find a relevant time period, we must take a closer look at short-term fluctuation.

Table 5. Characteristics of pedestrian flow in a homogeneous stream.

Quality of Flow	Space per Pedestrian (square ft)	Flow Rate (pedestrians/ min/ft)
Open	More than 530	Less than 0.5
Unimpeded	530 to 130	0.5 to 2
Impeded	130 to 40	2 to 6
Constrained	40 to 24	6 to 10
Crowded	24 to 16	10 to 14
Congested	16 to 11	14 to 18
Jammed	2 to 11	0 to 25

Note: 1 square ft = 0.09 m^2.
1 pedestrian/min/ft = 3.27 pedestrians/min/ft.

Figure 5. Minute-by-minute variation in pedestrian flow.



Short-term fluctuation is generally present in any traffic flow that is not regulated effectively by a schedule, and its underlying cause is that participants in a traffic stream arrive at a given spot at random. Thus, purely by chance, one minute a section of sidewalk may receive many pedestrians, and the next minute it may receive few. In an urban situation, this random unevenness is exaggerated by 3 additional factors. First, if passing is impeded because of insufficient space, faster pedestrians will slow down behind slow-walking ones, and a random bunch of pedestrians will snowball into a platoon. Second, subway trains and, to a lesser extent, elevators and buses release groups of people in very short intervals of time with pauses during which no flow may occur. Until they have a chance to dissipate, these groups proceed together more or less as a platoon. Finally, and most importantly, traffic signals release pedestrians in groups that tend to proceed as groups along a sidewalk.

Platoons represent involuntary groupings of pedestrians, and, as such, should be distinguished from groups who walk together by choice. Of course, a voluntary group of people strolling leisurely together and chatting can cause others to form a platoon when opportunities for passing are limited.

One of the reasons why platoons have been neglected by previous researchers may be that they are hard to define. In this exploration, we have tried both a positive and a negative definition. In the positive definition, platoons were timed and counted when it appeared to the observer that a wave of above-average density was swelling up in the traffic stream. In the negative definition, gaps in flow were timed and the stragglers walking during these lulls were counted; then the nonplatoon time and flow were subtracted from total time and flow to determine performance in platoons. The total time of an observation was generally 5 to 6 min except at subway exits, where hourly counts were taken. The platoons were timed in seconds to avoid the arbitrary mixing of periods of flow with periods of no flow that results from choosing longer units of time. Some 58 observations are summarized in the scatter diagram in Figure 6. The symbols in Figure 6 distinguish between observations that defined platoons positively and those that defined them negatively. The duration of platoons defined either way generally ranged from 5 to 50 s, but the average time in platoons was shorter for the positive definition. According to the positive definition, 53 percent of the flow occurred in platoons roughly 20 percent of the time. By the negative definition, 84 percent of the flow occurred in platoons 63 percent of the time. The flow rate in platoons was about 2.5 times greater than the average flow rate for the positive definition and about 1.3 times greater than that for the negative definition. Platooning tends to be more pronounced during the morning and evening rush hours than during midday.

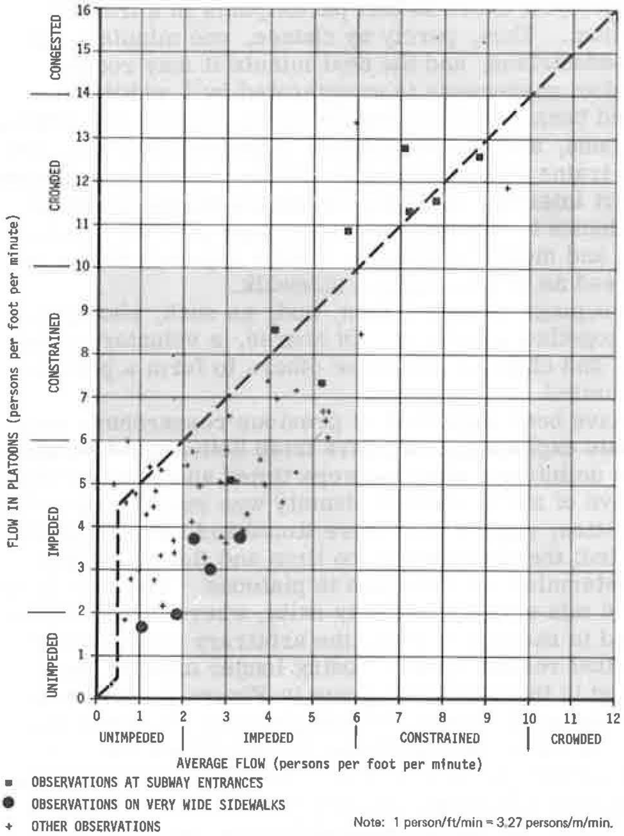
The most important influence on platoons at the street surface is traffic signals. Platoons generally follow signal cycles. To explore a different situation, counts also were taken during the morning arrival period at light-flow subway station exits. When platoons were strictly defined, 75 percent of the flow occurred in platoons 47 percent of the time, which is about 1.6 times the average flow rate. When platoons were more loosely defined, 95 percent of the flow occurred in platoons 60 percent of the time, which also is about 1.6 times the average flow rate.

It is clear that an average flow rate, even if it refers to a period as short as 1 min, is of little relevance to defining the condition of most of the pedestrians in a traffic stream. The time period truly relevant for design does not appear to be 15 min, 1 min, or any other arbitrary time span, but rather it appears to be that period during which flow in platoons occurs. Because time in platoons is composed of short spans of variable length, the most convenient way to deal with it is to take a time interval that is appropriate from the viewpoint of cyclical variation, say 15 to 30 min, and then design not for the average, but for the platoon flow rate during that period.

Revised Service Levels

Our task thus becomes one of showing those flow rates in platoons that occur at certain average flow rates so that the characteristics given in Table 6 can be applied to platoons. A comprehensive way of going about this would be to plot distributions for a range of

Figure 6. Flow in platoons related to average flow.



Note: 1 person/ft/min = 3.27 persons/m/min.

Table 6. Characteristics of average flow and flow in platoons.

	Average Flow		Possible Flow in Platoons	
Quality of Flow	Space per Pedestrian (square ft)	Flow Rate (pedestrians/min/ft)	Space per Pedestrian (square ft)	Flow Rate (pedestrians/min/ft)
Open	More than 530	Less than 0.5	More than 530	Less than 0.5
Unimpeded	530 to 130	0.5 to 2		
Impeded	130 to 40	2 to 6	60 to 40	4.5 to 6
Constrained	40 to 24	6 to 10	40 to 24	6 to 10
Crowded	24 to 16	10 to 14	24 to 16	10 to 14
Congested	16 to 11	14 to 18	16 to 11	14 to 18
Jammed	2 to 11	0 to 25	Less than 11	More than 18

Note: 1 square ft = 0.09 m^2. 1 pedestrian/min/ft = 3.27 pedestrians/min/m.

pedestrian densities by type of facility and time of day showing the percentage of people that have to walk at densities exceeding the average and by what amount the average is exceeded. Then a cutoff level can be chosen to serve a specified percentage of the walkers at a specified level of service. This detail, however, could not be attained, so a shortcut method was used.

In Figure 6 a time was drawn approximating the upper limit of all the platoon observations for 51 out of the 58 cases. Above it are 3 observations typifying small platoons during periods of light flow. One observation shows extreme conditions on an approach to the Port Authority bus terminal shortly after 5:00 p.m.; 3 of the 8 observations were at subway exits. The equation of this line relating maximum platoon flow to average flow is:

$$\text{Platoon Flow} = 4 + \text{Average Flow} \quad (7)$$

Application of this equation shows that an average flow rate of 6 to 10 pedestrians/min/ft (20 to 33 pedestrians/min/m) of walkway width, which we have previously described as constrained, will result in a crowded flow of 10 to 14 pedestrians/min/ft (33 to 46 pedestrians/min/m) in platoons. A flow between 2 and 6 pedestrians/min/ft (6.5 and 20 pedestrians/min/m), which all preceding authors have unanimously categorized as free flow and which we have called impeded flow, can result in platoon flow of between 6 and 10 pedestrians/min/ft (20 and 33 pedestrians/min/m), which by common consensus, is constrained.

To ensure a platoon flow rate of less than 6 pedestrians/ft (20 pedestrians/m) or a space allocation of more than 40 square ft (3.7 m^2)/pedestrian in platoons, the average flow rate must drop below 6.5 pedestrians/ft (21.2 pedestrians/m) and the average space allocation must rise above 130 square ft (12 m^2)/pedestrian, especially when sidewalks are narrow. An average of more than 500 square ft (46 m^2)/pedestrian would prevent formation of platoons on sidewalks 12 to 15 ft (3.7 to 4.6 m) wide, but this criterion would be impossible to meet in downtown areas. However, average space allocations between 80 and 200 square ft (7.4 and 18.5 m^2)/pedestrian on sidewalks wider than 30 ft (9 m) cause platoons to be substantially attenuated. Four such observations are shown separately in Figure 6.

The form that equation 7 takes (constant added to average flow) indicates that platooning has a much greater impact on light flow volumes than heavy flow volumes. Thus for an average flow rate of 2 pedestrians/min/ft (6.5 pedestrians/min/m), the additional margin necessary to accommodate platoons is 200 percent; at a flow rate of 10 pedestrians/min/ft (33 pedestrians/min/m) it is 40 percent. This pattern is not illogical because gaps between platoons tend to fill up as flow increases. It does, however, point to the following design conclusion: Minimum walkway standards that can be applied regardless of actual flow volume are necessary when flows are small because large platoons could arise suddenly. For example, an entrance to an apartment house may experience zero flow for many minutes until an elevator arrives with a platoon. As average flow increases space requirements do not grow proportionally but rather at a retarded rate, which is fortunate for the design of such high-intensity pedestrian facilities as shopping malls or transportation terminals. There are clear economies of scale in providing walkway space. Table 6 gives a comparison of average flow and platoon flow.

If the designer wants to attain for platoons what Oeding calls free flow and what Fruin calls service level A then 130 square ft (12 m^2)/pedestrian is the minimum average space allocation and 2 pedestrians/min/ft (6.5 pedestrians/min/m) of walkway width is the maximum flow except for wide walkways or where the absence of platooning can be demonstrated.

If unusual cost limitations are present, such as for underground passageways, or if a degree of crowding is desirable, such as in intensive shopping areas, then the average space allocation can be lowered and the flow rate can be raised accordingly. But, if overcrowding and congestion in platoons are to be avoided, space allocation should never fall below 40 square ft (3.7 m^2)/pedestrian on the average [below 24 square ft/ pedestrian (2.2 m^2/pedestrian) for platoons], and flow rate should never rise above 6 pedestrians/min/ft (20 pedestrians/min/m) on the average [above 10 pedestrians/ min/ft (33 pedestrians/min/m) for platoons].

REFERENCES

1. J. J. Fruin. Designing for Pedestrians; a Level of Service Concept. Polytechnic Institute of Brooklyn, PhD dissertation, Jan. 1970.
2. D. Oeding. Verkehrsbelastung and Dimensionierung von Gehwegen und anderen Anlagen des Fussgaengerverkehrs. Strassenbau und Strassenverkehrstechnik, Bonn, West Germany, No. 22, 1963, 62 pp.
3. S. J. Older. Movement of Pedestrians on Footways in Shopping Streets. Traffic Engineering Control, Aug. 1968, pp. 160-163.
4. F. P. D. Navin and R. J. Wheeler. Pedestrian Flow Characteristics. Traffic Engineering, June 1969, pp. 30-36.
5. B. D. Hankin and R. A. Wright. Passenger Flow in Subways. Operational Research, London, Vol. 81, No. 2, June 1958.
6. H. Kirsch. Leistungsfaehigkeit und Dimensionierung von Gehwegen. Strassenverkehr and Strassenverkehrstechnik, No. 33, 1964.
7. M. P. Ness, J. F. Morall, and B. G. Hutchinson. An Analysis of Central Business District Circulation Patterns. Highway Research Record 283, 1969, pp. 11-18.
8. C. A. O'Flaherty and M. H. Parkinson. Movement on a City Centre Footway. Traffic Engineering and Control, Feb. 1972, pp. 434-438.
9. Highway Capacity Manual. HRB Special Rept. 87, 1965.
10. L. C. MacDorman. An Investigation of Pedestrian Travel Speeds in the Business District of Washington, D.C. Catholic University of America, PhD dissertation, May 1967, 58 pp.
11. J. Gehl. Mennesker til fods. Arkitekten, Copenhagen, No. 20, 1968.
12. L. A. Hoel. Pedestrian Travel Rates in Central Business Districts. Traffic Engineering, Jan. 1968, pp. 10-13.
13. M. Wolff. Notes on the Behavior of Pedestrians. In People in Places: The Sociology of the Familiar (Birenbaum, Arnold, Sagarin, and Edward, eds.), Praeger Publishers, New York, 1973.
14. J. Rock, L. Greenberg, P. Hill, and J. Meyers. Aspects of Pedestrian Walkways. Graduate School of Public Administration, New York University, 1971.
15. J. Baerwald, ed. Traffic Engineering Handbook, 3rd Edition. Institute of Traffic Engineers, Washington, D.C., 1965, pp. 113-120.
16. R. Y. Okamoto and R. J. Beck. Preliminary Report on the Urban Density Study. Regional Plan Association, New York, May 1970.
17. P. Fausch. Simulation Tools for Designing Pedestrian Movement Systems in Urban Transportation Facilities. Pedestrian-Bicycle Planning and Design Seminar, San Francisco, Dec. 15, 1972.
18. Pedestrian Impact Study for the Proposed 175 Park Avenue Building, New York. Wilbur Smith and Associates, May 1968.
19. G. H. Winkel and G. D. Hayward. Some Major Causes of Congestion in Subway Stations. Environmental Psychology Program, City University of New York, May 1971.
20. G. R. Strakosch. Vertical Transportation. John Wiley and Sons, Inc., New York, 1967.
21. R. S. O'Neil. Escalators in Rapid Transit Stations. Transportation Engineering Journal, American Society of Civil Engineers, Vol. 100, No. TE1, Proc. Paper 10333, Feb. 1974, pp. 1-12.
22. J. M. Tough and C. A. O'Flaherty. Passenger Conveyors. Ian Allan, London, 1971.

23. Moving Way Transit. In Lea Transit Compendium, Huntsville, Alabama, Vol. 1, No. 2, 1974.
24. R. Kuner. The Boston Moving Walk Study. Barton-Aschman Associates, Chicago, 1972.
25. High Speed, Continuous Flow Person Conveyors. Tri-State Regional Planning Commission, New York, Interim Technical Rept. 4089-0600, 1968.
26. Peoplemover Systems for Mid and Lower Manhattan. Kaiser Engineers, New York, Jan. 1973.
27. Preliminary Application—Demonstration Project—High Speed Moving Sidewalks. Tri-State Regional Planning Commission, New York, 1974.
28. J. R. Allison et al. A Method of Analysis of the Pedestrian System of a Town Centre (Nottingham City). Journal of the Town Planning Institute, London, Vol. 56, No. 8, Sept.-Oct. 1970.
29. B. Pushkarev and J. M. Zupan. Pedestrian Travel Demand. Highway Research Record 355, 1971, pp. 37-53.

RECOMMENDED PROCEDURE FOR WEAVE-AREA OPERATIONS AND DESIGN

Louis J. Pignataro, William R. McShane, Roger P. Roess, Bumjung Lee, and Kenneth W. Crowley, Department of Transportation Planning and Engineering, Polytechnic Institute of New York

ABRIDGMENT

•TWENTY-FIVE years have passed since the original Highway Capacity Manual (1) first appeared in print. Since then, the procedures developed in the first manual as well as the modifications, extensions, and new methodologies presented in the 1965 edition of the Highway Capacity Manual (HCM) (2) have become national guides in the design and analysis of highway sections. As such, they have been exposed, through constant application, to detailed scrutiny by traffic planning, design, and operation specialists. Such exhaustive on-the-job evaluation has exposed problem areas, instructions that may be subject to misinterpretation, procedures that are complex and difficult to apply, and results that sometimes appear unreasonable.

In 1969, the National Cooperative Highway Research Program (NCHRP) authorized Project 3-15. The project statement stated: "Design criteria for weaving sections on multilane controlled-access highways require revision and updating, taking into account such variables as roadway geometrics, composition of traffic, volumes of mainline vehicles, and volumes of weaving vehicles."

As a result of an extensive evaluation of the accuracy and consistency of existing weave-area design and analysis procedures, it was recommended that a new procedure be developed. This paper reports on that procedure.

The end result of research under NCHRP auspices should be of direct use to the practicing engineer. The final recommended procedure is written as a self-contained document within the final report (3) in which a computer program implementing the procedure is also described.

The program handles both design and analysis problems, ramp weave, and major weave. It includes a feature by which consecutive analysis problems may be done without intermediate headings, so that comparison is simplified. Another feature allows one to step through a range of weave volumes and design an appropriate length for each. In this way, one may plot required length as a function of weave volume when all other parameters are fixed.

As part of the research, 1 multiple weave site was filmed. On the basis of this and other data, guidelines for application of the recommended procedure to multiple weaves were generated.

DEVELOPMENT OF WEAVE PROCEDURE

The following are some of the general concepts or ideas integral to the weave procedure:

1. Mean space speeds rather than operating speeds are used to define levels of operation.
2. Service volume concepts of the HCM are adapted and used for nonweaving traffic.
3. Volumes are considered in passenger car equivalents in units of passenger cars per hour. Adjustments of vehicles per hour to passenger cars per hour are made according to the HCM.
4. Levels of service are defined separately for weaving and nonweaving flows.

5. Although balanced design (comparable levels of service) is sought, configuration may prevent it from being realized.
6. As far as basic relations are concerned, there are 2 sets of equations: 1 for major-weave sections and 1 for ramp-weave sections.

Consideration and awareness of configuration (section lane arrangement, including number of lanes on each leg) are important and essential elements of the recommended weaving procedure and should be kept in mind while all research is done.

It is of prime importance in design that the configuration be such that

1. The computed weaving width can be delivered,
2. Lanes required for each outer flow (nonweaving flow) can be delivered, and
3. Lanes on each input-output leg can handle volumes at the level of service desired.

One of the prime results of the research leading to the recommended procedure was the determination of the maximum width that can be used by weaving traffic. It was found that this depended on configuration.

DEVELOPMENT OF BASIC RELATIONS

Extensive analysis of both the macroscopic data (6-min or greater flows and speeds) and the microscopic data and models developed within this research project led to development of the regression-based relations that form the core of the recommended procedure.

Some of the characteristics of the calibration, beyond those already noted, are as follows:

1. During calibration, one should distinguish between ramp weaves and major weaves.
2. The proper range of the calibration was found to be 30 mph (48 km/h) or greater for nonweave speeds (S_{nw}s). This limit, the common limit for level of service, was found as a result of investigation; it was not an a priori assumption.
3. For major weaves, weave speed (S_w) can go as low as 20 mph (32 km/h) for S_{nw}s $\geq$ 30 mph (48 km/h). This can be, and is, used to define a lower limit for weave level of service.
4. The resulting relations include S_{nw}s and S_ws (sometimes by $\Delta S = S_{nw} - S_w$) such that a continuum results rather than subcases for each of a set of levels of service. As a result, levels of service can be, and are, specified exogenously. Definitions that considered existing uses were selected.
5. Data aggregated in 18-min periods yielded better regularity than did 6- or 12-min periods. Longer periods did not improve regularity but did reduce the number of data points available. Calibration was based on 18-min time periods.

The best relationships describing weaving traffic were developed from the assumption that the ratio of weaving to total lanes is proportional (functionally related) to the ratio of weaving to total volume. That is to say, the percentage of width required by weaving vehicles is directly related to the percentage of total traffic that the vehicles constitute. Note that this relation involves both weaving and nonweaving types of flow in the determination of weaving. This is reasonable because, although flows are significantly segregated as vehicles enter the section, a physical overlap and interaction exist in the space the vehicles occupy.

MECHANISMS OF WEAVING: RESULTS

The project data base was used for a wide range of microscopic studies, and a number of microscopic models for various purposes were formulated. These investigations served 2 purposes: (a) they were a guide and a control in the macroscopic investigations, and (b) they provided a better understanding of the basic mechanisms of weave section operation.

These studies affirm that

1. There is a substantial presegration of weaving and nonweaving traffic as it enters the weave section. The degree of presegregation lessens as section length increases, but the sensitivity is significant [under 2,000 ft (610 m)] for ramp weaves.
2. Configuration is important.
3. The benefit of increasing length dissipates rapidly.
4. Weave sections often are controlled by specific concentrations of vehicles or "hot spots" within the weave section. Conversely, some areas within the weave section are underused.
5. As far as can be discerned, lane-change probabilities are not dependent on volume, longitudinal position within the weave section, or section length. They do vary according to essential or nonessential lane changing and, for nonessential changes, according to direction of movement.
6. A weave section may be, and frequently is, subjected to a wide range of conditions regarding flow levels. This range can cover a typical day, a few hours, or seasons.
7. In addition to substantial presegregation, the multiple weave site in the project data base also gave evidence that the allocation of weaving according to subsection lengths recommended in the HCM does not hold.
8. The difference in speed between the 2 weaving movements is such that heavier volume is almost always faster. This pattern is more pronounced for ramp weaves than for major weaves.
9. Although the accident rate is greater in weave sections than on open freeway sections, attributing this rate specifically to length, weave volume, or any other factor is not possible according to available data. In addition to the limited quantity of data, other factors such as signing may be predominant, and an investigation should take all factors into account.

SUMMARY

A new procedure for weave section design and analysis has been developed, and is recommended for use. A complete methodology and sample problems, which are given in the project report (3), explicitly recognizes the importance of configuration because configuration controls the maximum weaving width that can be delivered.

Although the formulation allows for an analytic solution, the nomographic approach or the computer program should be used. For analysis problems, nomographic approval requires an iterative solution with which a user can rapidly become facile. In design, such iterations are not required. The NCHRP report (3) contains results of the evaluation of previously existing procedures that are both accurate and consistent. It also presents a validation of the recommended procedure. Guidelines are presented for multiple weaves, but they are based on limited data. Although the data are limited, the practicing engineer must cope with the design and analysis of multiple weave sections. Therefore, the best possible guidelines should be developed from existing knowledge, and the engineer should be advised to use them with caution.

ACKNOWLEDGMENTS

This study was conducted under National Cooperative Highway Research Program Project 3-15. The opinions and findings expressed or implied in this paper are those of the authors. They are not necessarily those of the Transportation Research Board, the National Academy of Sciences, the Federal Highway Administration, the American Association of State Highway and Transportation Officials, nor of the individual states

participating in the National Cooperative Highway Research Program.

REFERENCES

1. Highway Capacity Manual. Bureau of Public Roads, U.S. Department of Commerce, 1950.
2. Highway Capacity Manual. HRB Special Rept. 87, 1965.
3. Weaving Areas—Design and Analysis. NCHRP Rept. 159, 1975.

EFFECTS OF DESIGN ALTERNATIVES ON QUALITY OF SERVICE AT SIGNALIZED DIAMOND INTERCHANGES

Carroll J. Messer, Texas Transportation Institute, Texas A&M University; and David J. Berry,* Country Roads Board, Victoria, Australia

This paper examines the effects of minimum phase length and variations in the spatial arrangement of ramp intersections on the capacity of diamond interchanges operated with 4-phase-overlap signalization. The introduction of practical minimum constraints on phase lengths can significantly alter the performance characteristics of this type of interchange. This paper also illustrates a method by which interchange design can be assessed in terms of quality-of-service measures. A computer program, written in FORTRAN IV, was developed to speed up the analysis. Examination of some typical cases illustrates that an efficient design can be selected if one uses the program for a given set of minimum green-phase lengths and traffic demands.

•THE DIAMOND interchange is a widely used form of freeway-to-highway interchange. It may be designed to operate in either a rural or urban environment. Traffic flow in rural diamond interchanges normally is controlled by signing whereas, in urban areas, signalization generally is used. The urban diamond interchange typically has high traffic volumes and is a critical traffic facility in the urban freeway corridor. Efficient movement of traffic through the interchange is desired. The quality of service provided motorists by the signalized diamond interchange depends, to a large extent, on the geometric features of the interchange and on signalization strategy.

This paper examines some of the more important relationships between traffic performance measures and geometric design alternatives of diamond interchanges with 4-phase-overlap signalization. An example of a diamond interchange and its associated traffic movements are shown in Figure 1. The design includes 1-way frontage roads and U-turn bays. As a consequence, the frontage-road U-turn traffic has no effect on signal operation. The design also may have free right-turn lanes, advanced left-turn bays, and acceleration and deceleration lanes.

The 4-phase-overlap signalization strategy used in this paper is shown in Figure 2 for the controlled traffic movements within the diamond interchange shown in Figure 1. This signalization strategy has been widely implemented, but recent literature illustrates that questions remain concerning the operational characteristics and the resulting quality of service afforded motorists (1, 2). Interchange signalization results in a nearly perfect progression between the 2 closely spaced intersections that make up the interchange (2).

To evaluate traffic operations at the interchange, we first calculated signal timings for a given set of geometric and traffic inputs. Durations of green plus yellow phases are calculated by using equations derived from Webster's method (3). Interchange traffic performance is described by using Webster's expression for delay (3) and Miller's expression for load factor (4). Interchange performance is examined first for the case in which there are no minimum constraints on phase lengths. Then the effects of introducing minimum phase lengths and varying some aspects of the physical

*Mr. Berry was with the Texas Transportation Institute when this research was performed.

arrangement of the intersections are examined. The notations used in the formulations in this paper are as follows:

- i = an approach traffic movement shown in Figure 1 (external approaches are $i = 1, 2, 3, 4$);
- ϕ_i = phase overlap on movements 7 and 8 as related to intersection spacing shown in Figure 3 in seconds;
- ϕ = total overlap = $\phi_7 + \phi_8$ in seconds;
- C = cycle length in seconds;
- G_i = sum of green plus amber time on approach i in seconds;
- ℓ_i = lost time on external approach i (assumed to be 4.0 s);
- L = lost time over the 4 external approaches = $\Sigma\ell_i$ (16 s);
- g_i = effective green time on external approach i, $G_i - \ell_i$ in seconds;
- $\lambda_i = \frac{g_i}{C}$ = proportion of cycle that is effectively green on approach i in seconds;
- M_i = minimum green plus amber time on approach i in seconds;
- s_i = saturation flow on external approach i in vehicles/second of green;
- q_i = average arrival rate on approach i in vehicles/second;
- $y_i = \frac{q_i}{s_i}$ = ratio of average arrivals to saturation flow on external approach i (y ratio);
- Y = sum of y_i's taken over the 4 external approaches (Y value);
- $x_i = \frac{q_i C}{s_i g_i}$ = saturation ratio (average number of arrivals to maximum number of departures per cycle on external approach i);
- d_i = average delay to vehicles on external approach i in seconds/vehicle;
- $\bar{d}$ = average delay for all vehicles on external approaches in seconds/vehicle; and
- LF_i = load factor on external approach i.

MODEL DEVELOPMENT

Pinnell and Capelle (1) established the basic relationships and developed a design method to analyze diamond interchanges by using critical lane analysis. Messer, Whitson, and Carvell (5) further modeled the operations and limitations of 4-phase-overlap operation. Munjal (2) illustrated that 4-phase strategy results in a nearly perfect progression between 2 closely spaced signalized intersections within the interchange. The signal overlap times needed to provide smooth progression are related directly to the travel times between the intersections. In this paper, the 2 phase overlaps were assumed to be equal ($\phi_7 = \phi_8 = \phi/2$) and to increase with increasing spacing between intersections within the interchange, as shown in Figure 3 (6, 7, 8).

Pinnell and Capelle (1) and Messer, Whitson, and Carvell (5) have shown that certain relationships describing 4-phase-overlap signalization exist among C, ϕ, and G_i for the numbered movements identified in Figure 1. These relationships may be expressed as

$$G_1 + G_2 + G_5 = C \tag{1}$$

$$G_3 + G_4 + G_6 = C \tag{2}$$

The sum of the external phases is equal to

$$G_1 + G_2 + G_3 + G_4 = C + \phi \tag{3}$$

Figure 1. Diamond interchange with traffic movements.

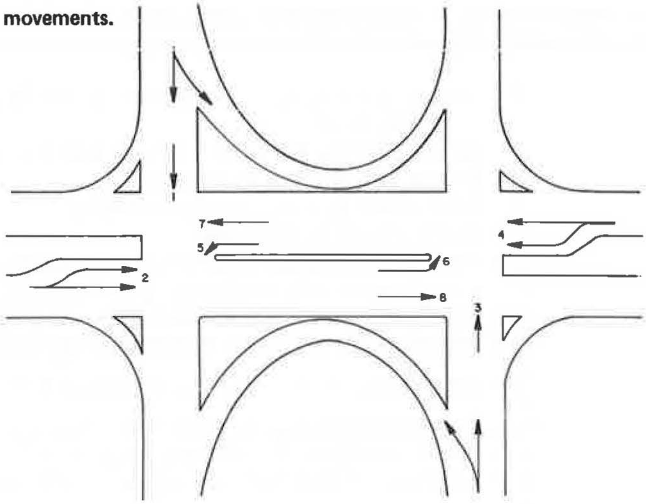


Figure 2. Four-phase-overlap signalization.

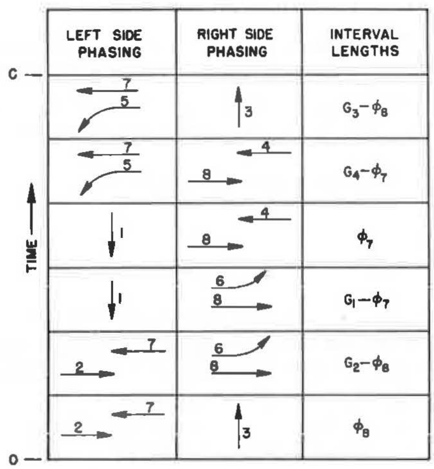
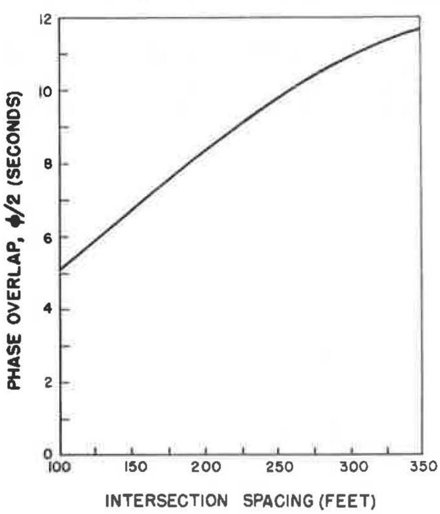


Figure 3. Relationship between phase overlap and center-to-center intersection spacing.



which results in the sum of the internal left-turn phases being

$$G_5 + G_6 = C - \phi \tag{4}$$

Equal saturation ratios and minimum delay conditions for a given cycle length are obtained over all external movements if their effective green times are calculated according to the following relationship:

$$g_1 = \frac{y_1}{Y}(C + \phi - L) \tag{5}$$

Sometimes it may be more desirable to calculate first the internal left turns from the following equations (5):

$$G_5 = \frac{(C - L/2)(y_3 + y_4) - (y_1 + y_2)(\phi - L/2)}{Y} \tag{6}$$

$$G_6 = \frac{(C - L/2)(y_1 + y_2) - (y_3 + y_4)(\phi - L/2)}{Y} \tag{7}$$

and then the external movement phase lengths by using relationships of the form:

$$g_1 = \frac{y_1}{y_1 + y_2}(C - G_5 - L/2) \tag{8}$$

Equations 6, 7, and 8 may be used either for general cases or for particular cases in which phase lengths calculated for internal left turns do not satisfy predetermined minimums. In actual field applications, internal left-turn times calculated by using Eqs. 6 and 7 must be restrained to conform to predetermined minimum phase lengths such that total green plus amber time equals or exceeds M_1, which is determined from considerations of minimum pedestrian requirements, or minimum practical vehicular movement times (5). These practical restraints are shown in Figure 4.

A computer program that uses FORTRAN IV was written to solve for the desired movement green times for a given set of demand volumes, saturation flow rates, overlaps, minimum green constraints, and cycle lengths. After results were obtained, a detailed mathematical analysis of the traffic performance of the interchange was possible, thus minimizing the need for simulation or extensive field studies (2, 6, 7, 9).

Two measures of the quality of traffic service afforded motorists—delay and load factor—were calculated. Traffic flow toward the interchange was assumed to be of a Poisson type with regular, uniform flow departures during effective green time. Delays on the 4 external approaches were estimated by using Webster's well-known expression (3):

$$d = \frac{C(1 - \lambda)^2}{2(1 - \lambda x)} + \frac{x^2}{2q(1 - x)} - 0.65\left(\frac{C}{q^2}\right)^{1/3} x^{(2 + 5\lambda)} \tag{9}$$

Again, because of the near perfect progression that results for traffic moving within the interchange (2) and because of the assumed design that provides U-turn bays for frontage-road traffic, negligible delay within the interchange was assumed to occur. Thus average interchange delay may be calculated from the 4 external movements by

$$\bar{d} = \frac{\sum_{i=1}^{4} q_i d_i}{\sum_{i=1}^{4} q_i} \tag{10}$$

For each external movement, load factors were estimated by using Miller's expression (4):

$$LF = \exp(-1.3\phi) \tag{11}$$

where $\phi = \frac{1 - x}{x} \sqrt{sg}$.

RELATIONSHIPS BETWEEN DESIGN ALTERNATIVES AND QUALITY-OF-SERVICE MEASURES

The majority of the discussion in the literature on 4-phase-overlap signalization deals with the analysis of capacity and the quality of traffic service afforded motorists under unconstrained conditions. In practice, however, interchanges are rarely unconstrained. Physical, land use, and economic factors frequently limit the range of ramp spacings and cross-street configurations that can be considered in design. Operational constraints such as the requirement to provide minimum green times necessary for pedestrian crossings frequently result in green times longer than those theoretically needed to minimize delay. Nevertheless, it is useful to examine the unconstrained interchange to establish some optimum performance values.

Unconstrained Interchange

The first case to be analyzed is that of no minimum phase-length constraints. In this case, the G_i in Eq. 3 can be as small or large as desired, but it must be nonnegative. Webster (3) has shown that minimum delay occurs when all saturation (demand-to-signal-capacity) ratios at a signalized intersection are the same. Because the saturation ratio is

$$x_i = \frac{q_i C}{s_i g_i} = \frac{y_i C}{g_i}$$

or

$$g_i = \frac{y_i C}{x_i} = G_i - \ell_i$$

then

$$\frac{y_1 C}{x_1} + \frac{y_2 C}{x_2} + \frac{y_3 C}{x_3} + \frac{y_4 C}{x_4} = C + \phi - \Sigma \ell_i$$

If we assume that all x_is are equal to X, then

$$\frac{C}{X} \Sigma y_i = C + \phi - \Sigma \ell_i$$

or

$$Y = \frac{X(C + \phi - L)}{C} \tag{12}$$

which expresses the allowable Y value over all external approaches that can exist and still provide a particular saturation ratio, X, for a given set of C, ϕ within the interchange, and total L. Capacity conditions may be described by setting X in Eq. 12 equal to 1. From expressions similar to Eq. 12, it has been deduced that the capacity of the 4 external approaches of a 4-phase-overlap signalization strategy increases as cycle length decreases if total overlap is greater than total lost time (1, 7).

Figure 5 shows the relationships between the allowable total saturation flow ratio, Y, and the related variables of C and X. Two sets of curves are presented. The upper curves are for a total overlap of 20 s and a total lost time of 16 s. The lower curves have a total overlap time of 12 s and a total lost time of 16 s. Because the total Y value is directly related to demand volume for a fixed capacity, Figure 5 shows that the allowable demand volume on the interchange decreases as cycle length increases for a total overlap that is greater than lost time; it increases as cycle length increases for a total overlap that is less than lost time.

The 2 unconstrained interchanges were further analyzed with respect to delay, again assuming uniform y_is, equal flows over all external approaches, and a total lost time of 16 s. Figure 6 shows the variation in delay when cycle length and total Y value for the case of ϕ equal 12 s. Figure 7 shows the variation in delay when cycle length and total Y value for the case of ϕ equal 20 s. For the case of ϕ assumed equal to lost time (16 s), relationships similar to those shown in Figure 7 are obtained.

When ϕ is less than lost time (total overlap 12 s), the delay curves in Figure 6 take on the characteristic shape expected at conventional signalized intersections. In contrast, interchanges that have their intersections spaced further apart such that the total overlap equals or exceeds lost time (Figure 7) exhibit approximately linear relationships between delay and cycle length. For a given total Y value, delay in the unconstrained interchange is reduced by increasing overlap or ramp-intersection spacing. It is suggested that the linear reduction in delay with cycle length for ϕs equal to or exceeding the lost time will be limited by the validity of the constant lost time and uniform departure model used in the delay expression and might be expected to break down for cycle lengths that are less than 45 s.

Constrained Interchange

In practice, there are necessary constraints that limit the range of phase lengths. In general, minimum phase lengths will be determined either by considerations of pedestrian movement intervals or by consideration of minimum vehicle-phase intervals. To

Figure 4. Solution area for phase (movement) 6 as limited by overlap and minimum greens.

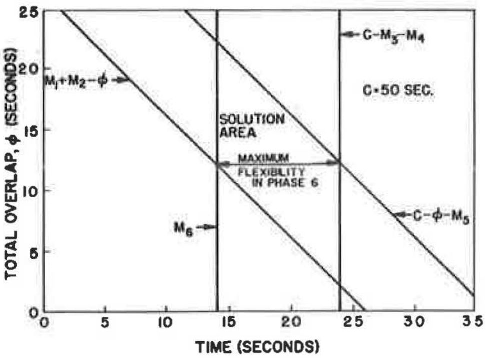


Figure 5. Variation in Y value with cycle length and saturation ratio, X, for $\phi > L$ and $\phi < L$.

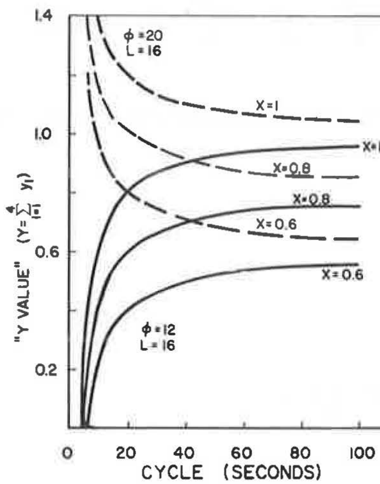


Figure 6. Average unconstrained interchange delay when $\phi < L$.

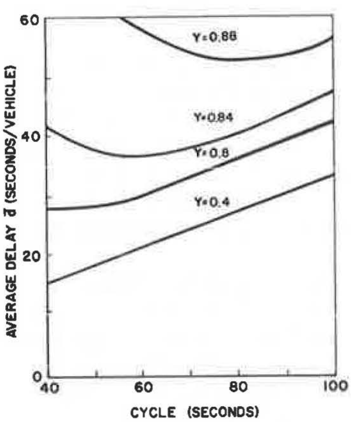


Figure 7. Average unconstrained interchange delay when $\phi > L$.

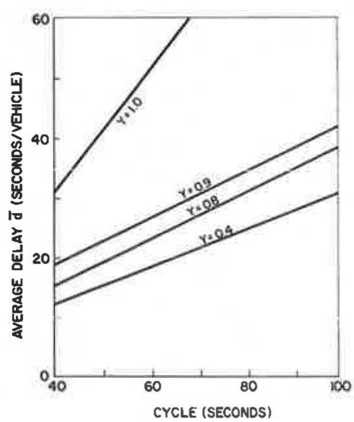
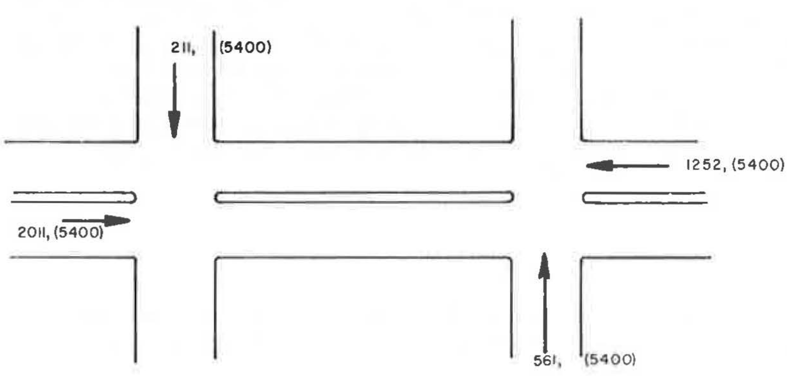


Figure 8. Example interchange design problem given volumes and saturation flows.



illustrate the effect of the minimum green constraint conditions, consider an interchange with the demand and saturation flows shown in Figure 8. The demand volumes have been taken from a previous study of an interchange at I-610 and Richmond Road in Houston, Texas (1). For simplicity, saturation flows have been assumed to be 1,800 vehicles/h of green/lane. The cross street is a 6-lane divided facility and the freeway includes 3-lane frontage roads. Minimum phase times were assumed to be 16 s. Intersection spacing was assumed to be 260 ft (79 m) with a phase overlap of 10 s and a total overlap of 20 s.

The variation of average delay to cycle length at the interchange is shown in Figure 9. When there is no minimum green requirement, the resultant delay curve has the same form as do those for unconstrained interchanges because the available effective green time may be apportioned over the approaches to obtain uniform saturation ratios. The introduction of minimum green restraints significantly alters the performance of the interchange. The performance curves for delay take on the form expected at conventional intersections exhibiting a well-defined optimum cycle. Delay escalates rapidly as cycle length is reduced below the optimum value.

The effect of introducing minimum green requirements is also shown in Figure 10 in a before and after comparison of load factors for the most severely affected approach (in this case, movement 2 on the cross street), when the load factor exceeds about 0.4 at a cycle length of 70 s.

Perhaps the simplest explanation of the change in performance between the constrained and the unconstrained interchanges can be obtained by drawing an analogy to the effect of lost time on performance. When total lost time exceeds total overlap, interchange performance takes on some of the characteristic features of normal signalized intersections. In the case of the constrained interchange operating under nonuniform y_i ratios, the apportionment of effective green time is constrained to satisfy certain minimum values. As a result, more time is apportioned to the low demand movements than is necessary to carry traffic. The resultant "lost" time does not move traffic efficiently and is similar in effect to actual lost time. If this time is increased such that the total effective lost time exceeds the total overlap, delay will escalate rapidly as cycle length is reduced below optimum (at moderate to high Y values). When y_i ratios are approximately uniform, total effective green time is apportioned approximately uniformly to all movements, and minimum phase lengths are more easily satisfied. For these cases, interchange performance curves take on characteristics similar to those described for unconstrained interchanges.

Effect of Varying Intersection Spacing

The effect of varying the spacing between the intersections within the interchange can be illustrated by using the design volumes given in Figure 8 and by assuming constant minimum phase lengths of 16 s.

Figure 11 shows the effect on average delay at the interchange that is due to variations in cycle length and total overlap (equivalent intersection spacing given by Figure 3). In general, both overlap and cycle length significantly affected the resulting delay. The effects are complex and interrelated. In the cycle length range of 70 to 100 s, significant reductions in delay occur when total overlap is increased from 12 to 24 s. This reduction is due to increasing the proportion of effective green time per cycle on the external movements. That is, the external approach capacity of the interchange increases as predicted by Pinnell and Capelle (1) and Woods (7). These increases are contrary to what Munjal (2) said in his general statement on capacity of diamond interchanges. However, for total overlaps that are less than lost time (ϕ = 12 s), external capacity and phase flexibility do increase with cycle length, thus permitting a reduction in delay.

For cycle lengths less than 70 s, increasing the total overlap does not necessarily reduce the delay. In fact, at a cycle length of nearly 55 s, there will be no satisfactory overlap for these minimum greens. These results are due to a lack of phase-length flexibility that arises from the minimum green requirements in Eqs. 1, 2, 3, and 4.

Figure 9. Average interchange delay with and without minimum green constraints.

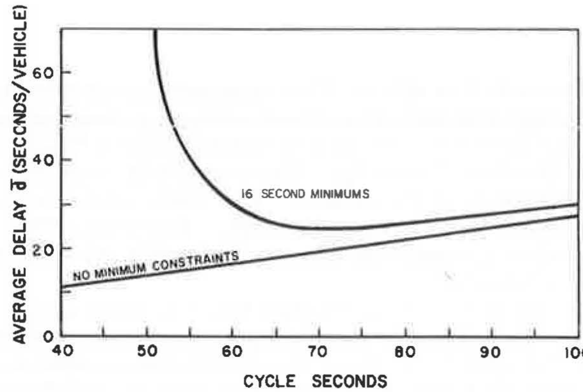


Figure 10. Resulting load factor on movement 2.

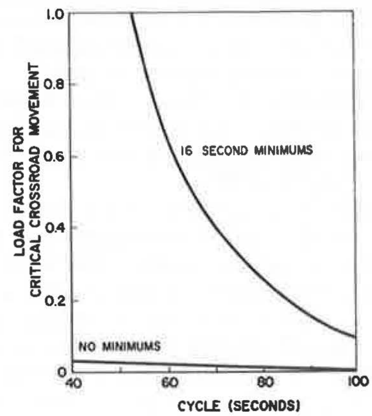


Figure 11. Average interchange delay for different overlaps or intersection spacings.

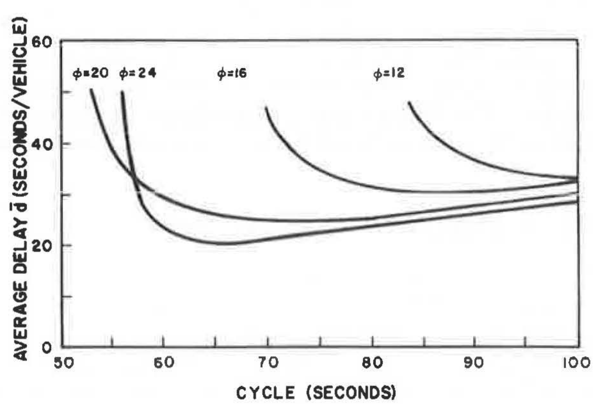
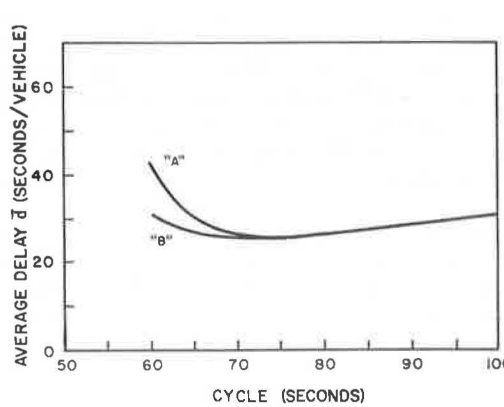


Figure 12. Average interchange delay for 2 different design configurations.



Total overlaps exceeding 24 s are not shown in Figure 10 for 2 reasons. First, larger overlaps did not significantly reduce the external delay below those of 24-s overlaps for cycle lengths exceeding 60 s. Second, total overlaps exceeding approximately $\frac{C}{3.5}$ begin to develop internal delays on movements 5 and 6 (Figure 1), and the delays become larger as overlaps increase. This critical total overlap is about 20 s for a rush-hour cycle length of 70 s. Thus a phase overlap of 10 s again appears to be about the optimal value because good capacity and phase flexibility are provided together with nearly minimum external delay and practically no internal delay (7).

Effect of Varying Number of Lanes

Occasionally design engineers find that they must sacrifice ramp spacing and phase overlap because of social and economic constraints. If the operations staff insist on attaining a total overlap time of 20 s there may, in fact, be little to trade. This type of situation might occur when wide intersection spacings would intrude undesirably into school grounds or other areas of local community significance. Then, it would be desirable to establish whether concessions in intersection spacing could be made against the number of lanes provided on the cross street.

To illustrate a hypothetical situation, consider the design volumes used previously. Compare the performance of the original configuration that consisted of a 6-lane cross street, 3-lane frontage roads, and an intersection spacing corresponding to a total overlap of 20 s (case A in Figure 12). This configuration had the effect of reducing the overlap to 12 s and widening the cross street to 8 lanes (case B in Figure 12). The resultant performance curves also are shown in Figure 12. That the performances under the design volumes are remarkably similar can be seen. The additional approach capacity provided by the wider crossroad, shorter overlap design, and more nearly uniform y_i ratios appears to give a small theoretical advantage over the original design alternative (case A) that had more overlap time available. This advantage may not be fully realized in practice because the flared lane may not be fully used.

Bay Designs With No Provision for U-Turns

This analysis has assumed that the effects of freeway-ramp (frontage-road) traffic making U-turns through the interchange could be overlooked because U-turn bays had been provided. Essentially, any interchange having low U-turn volumes could be analyzed as done in this paper and practical accuracy could be maintained. Diamond interchanges without frontage roads would tend to have low U-turn volumes because access to adjacent traffic generators across the freeway normally would be provided by the existing street system, resulting in left turns rather than U-turns from the ramps. For practical purposes, the analyses presented hold for U-turn volumes, q_u, in vehicles/hour, that are at least as large as

$$q_u = 270\left(1.00 - \frac{13.3 + 2.67\phi}{C}\right) \tag{13}$$

or 100 vehicles/h for an 80-s cycle and a 14-s total overlap. Equation 13 was developed from Eqs. 1, 2, 3, and 4 by assuming that most of the traffic making U-turns would have stopped initially at the external ramp signal and then progressed through the U-turn movement while experiencing a total lost time of 4 s. Maximum U-turn flow rates were assumed to be 900 vehicles/h of green. Equal green times between the ramp and adjacent cross-street approaches together with a 60:40 split between the 2 internal left-turn phases were assumed.

SUMMARY

This study has confirmed that the operational performance of 4-phase-overlap signalized diamond interchanges is complex and that many interrelated factors affect the performance of the signal system. As a result, predicting the quality of service afforded motorists by making a simple examination of a given situation is difficult. Any diamond interchange proposed to improve operations should be analyzed by using methods that will accurately describe the effects of different geometric alternatives and signalization strategies. If 4-phase-overlap signalization is being considered, a comprehensive analysis can be rapidly made by using computer programs similar to the one developed in this research.

The consequences of increasing intersection spacing, adding additional approach lanes, reducing pedestrian crossing times, and varying the cycle length of the system can be efficiently determined in terms of approach delay, load factor, and total interchange delay.

Specific observations and findings may be drawn from this study, and we hope that they will be of assistance to design and operations engineers working with 4-phase-overlap signalized diamond interchanges. The experiences of the authors in freeway design and computer traffic control of diamond interchanges support the following study results (5, 10):

1. Signal performance depends on both the proportion of effective green time per cycle and the phase flexibility available to allocate the total time in proportion to existing movement-demand-to-capacity ratios.
2. Increasing the overlap for a fixed cycle increases the proportion of effective green time per cycle while tending to reduce phase flexibility between intersections. These results become more pronounced at shorter cycles.
3. Longer minimum green times reduce phase flexibility as do shorter cycles.
4. Unbalanced demand volumes usually call for considerable phase flexibility, but their effects can be reduced by judicious selection of the number of approach lanes.
5. The proportion of effective green time per cycle will be greater than 1.0 if total overlap is greater than lost time. When this exists, the proportion of effective green time per cycle increases as the cycle length is reduced.
6. Signal performance seldom will be improved by decreasing cycle length below 70 s during rush-hour conditions.
7. Experience has indicated that cycle length selection usually depends more on phase flexibility than on the total external approach capacity of the interchange.
8. For most practical cases, there will be a minimum delay cycle length.
9. Phase overlaps of 10 s appear to be near the optimal phase overlap for a wide range of conditions. However, diamond interchanges having phase overlaps of 7 s will operate effectively at rush-hour cycle lengths.
10. In selecting a design alternative, sums of Y values should not exceed 0.75, 0.80, and 0.85 for phase overlaps of 12, 16, and 20 s respectively.
11. Alternative interchange designs should be analyzed for a range of design and traffic operation variables in terms of both system and driver quality-of-service measures.

ACKNOWLEDGMENTS

We wish to express our appreciation to D. L. Woods and other members of the design and operations staff of the Texas Transportation Institute for their assistance and review of this research. Mr. Berry wishes to express his gratitude to the Country Roads Board of Victoria, Australia, and the International Road Federation whose support was instrumental in enabling him to study at Texas A&M University.

This paper discusses 1 phase of the research project, "Effects of Design on the Operational Performance of Signal Systems," conducted by the Texas Transportation Institute and the Texas Highway Department in cooperation with the Federal Highway

Administration. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented. The contents do not reflect the official views or policies of the Federal Highway Administration.

REFERENCES

1. C. Pinnell and D. C. Capelle. Operational Study of Signalized Diamond Interchanges. HRB Bulletin 324, 1962, pp. 38-72.
2. P. K. Munjal. An Analysis of Diamond Interchange Signalization. Highway Research Record 349, 1971, pp. 47-64.
3. F. V. Webster. Traffic Signal Settings. Her Majesty's Stationery Office, London, Road Research Technical Paper 39, 1958.
4. A. J. Miller. On the Australian Road Capacity Guide. Highway Research Record 289, 1969, pp. 1-13.
5. C. J. Messer, R. H. Whitson, and J. D. Carvell. A Real-Time Frontage Road Progression Analysis and Control Strategy. Transportation Research Record 503, 1974, pp. 1-12.
6. D. L. Woods. Digital Simulation of a Diamond Interchange. Texas A&M Univ., PhD dissertation, Aug. 1967.
7. D. L. Woods. Limitations of Phase Overlap Signalization for Two-Level Diamond Interchanges. Traffic Engineering, Sept. 1969.
8. J. D. Barnett and C. Pinnell. At-Grade Intersection Spacing on Conventional Two-Level Diamond Interchanges. Texas Transportation Institute, Texas A&M Univ., Research Rept., Project 2-8-58-6, 1963.
9. J. F. Torres, J. A. Nemeczky, P. K. Munjal, and B. D. Widdice. Before and After Simulation and Field Studies of Diamond Interchange Operations. System Development Corporation, May 1972.
10. C. J. Messer and J. L. Gibbs. Computer Control of the Wayside-Telephone Arterial Street Network. Texas Transportation Institute, Texas A&M Univ., Research Rept. 165-3, April 1973.

CAPACITY OF SIGNALIZED INTERSECTIONS

Eugene Reilly, Ivan Dommasch, and Munireddy Jagannath,
Division of Research and Development, New Jersey Department of Transportation

Standardized methods for estimating capacity at signalized intersections have been sought for at least 25 years. The publication by the Bureau of Public Roads of the 1950 Highway Capacity Manual was the first extensive effort made toward this objective. In 1960, W. R. Bellis, of the New Jersey Department of Transportation, developed a more simplified technique. The results of the research discussed in this paper indicate that the techniques in the Highway Research Board's 1965 Highway Capacity Manual yield greater than 20 percent error in estimating the through capacity of the approach to a signalized intersection for at least half the locations studied. When the Bellis technique was used, less than 20 percent of the locations showed this error. Revisions were made to techniques in the 1965 Highway Capacity Manual, and 20 percent error then was found in only 25 percent of the locations studied. It was concluded that measuring the capacity of an approach in the field is preferable to estimating capacity by using the 1965 Highway Capacity Manual. A method of determining capacity from field measurements is described briefly in this paper.

•STANDARDIZED methods for estimating capacity at signalized intersections have been sought for at least 25 years. The 1950 Highway Capacity Manual (1) was the first extensive effort made toward this objective. The 1965 Highway Capacity Manual (HCM) (2) added parameters not covered in the 1950 publication. Meanwhile, many engineers were continuing to use their own techniques for capacity estimation primarily because of the HCM's lack of accuracy.

The HCM has been the subject of several studies (3, 6, 7, 8) that have investigated the applicability of the parameters, accuracy of results, and modification of the approach used to determine capacity.

A study by Reilly and Seifert (3) was conducted to determine the accuracy of the HCM's capacity estimation. The results of this study indicated that the HCM capacity estimates were inaccurate by at least 20 percent for half the approaches studied. The results indicated an inaccurate quantification of some of the parameters in the HCM for a limited study sample. As a result of Reilly and Seifert's study (3), an expanded work plan was developed to study and revise the parameters of the HCM to improve its reliability for estimating capacity. More intensive analysis was performed by the Bellis method (4) on capacity estimation.

STUDY METHODOLOGY

This paper compares the HCM and Bellis techniques to a field measurement of capacity [average loaded-cycle expanded (ALE) measurement]. ALE is considered to be the most accurate method of determining the capacity of a signalized approach, and it represents an expansion of the average number of vehicles/loaded cycle. For example, if there were an average of 20 vehicles/loaded cycle and 60 cycles/hour, ALE would be 20 × 60 or 1,200 vehicles/hour.

The HCM method combines factors for several environmental and traffic conditions and then applies these factors to a basic volume. The basic volume is determined by the approach width and parking conditions at the site.

The Bellis method is simpler. It has been modified in this study to include a factor for right and left turns. Basically, a lane capacity volume is given for each of the 4 types of approach for a specific green period. This volume is adjusted by the number of lanes, number of cycles, and percentage of trucks and turns.

Data Collection

The primary object of this study is the improvement of the reliability of the HCM parameters. It is evident that the interaction of many variables would require an extremely large sample size. The estimate of approach volume is dependent on 10 separate variables, many of which have a wide range of application. We decided to try to incorporate our data collection with another program because it would be impossible to amass the information with our own forces.

The data collected in 21 samples by 1 consultant's field crews were found to have a variety of inaccuracies which resulted in our using only 38 percent of the data collected. A second consultant volunteered input from 24 samples. Saturation flow information and HCM parameters were collected. However, only 50 percent of these data were useful in the analysis. New Jersey Department of Transportation field crews tabulated data on 106 sample sites. Some of these samples were reruns because inaccuracies were found in the original data. Sixty percent of these data were used in the final analysis.

Tables 1 and 2 give the characteristics of 122 sites. The following abbreviations are used in these tables:

Abbreviation	Definition
LF	Load factor
PHF	Peak-hour factor
CBD	Central business district
OBD	Outlying business district

Eighty-five of the sites were used in the analysis. Although 151 sites were studied originally, some were not tabulated in Tables 1 and 2 usually because LF was zero or the data were collected inaccurately.

Because of the variety and enormousness of unusable data, the original plan for the study had to be modified to accommodate the limited input that was available for the modification of so many parameters in the HCM.

Because of the lack of available 1-way streets, only approaches on 2-way streets were used. The method of collection was similar to that used in the Reilly and Seifert study (3).

Data Analysis—Highway Capacity Manual Method

The major steps in the analysis are as follows:

1. Capacity is estimated by using the adjustment factors in the HCM. This includes use of the factors for peak hour and metropolitan area population. The population used for this factor is the population of the municipality in which the intersection is located plus the population of adjacent municipalities.
2. Estimate of capacity is again computed, but no adjustment is made for PHF. The reason for not making an adjustment is that the PHF correction accounts for delays due to peaking traffic and not for reductions in capacity. Because ALE does not make subjective judgments concerning delays, the HCM estimate for capacity is computed without a PHF adjustment, thereby putting the HCM estimate on the same basis as ALE.
3. HCM parameters are studied for their rationale and accurateness (as determined

Table 1. Physical, environmental, and traffic characteristics of study locations for 2-way streets with no parking.

Site	Approach Width (ft)	Lanes	Location	Population (thousands)	PHF	LF	Percent Turning Left: Expanded Loaded Cycle	Percent Turning Left: Peak Hour	Percent Turning Right: Expanded Loaded Cycle	Percent Turning Right: Peak Hour	Truck Percent: Expanded Loaded Cycle	Truck Percent: Peak Hour	Bus Percent: Expanded Loaded Cycle	Bus Percent: Peak Hour	Cycle (s)	Green Time (s)	Service Volume (vehicles/ h green)	ALE Capacity[a] (vehicles/ h green)
RS-16	9	1	Residential	250	0.89	0.35	0		7		0		–	–	90	63	1,291	1,486
RS-22	9.5	1	Residential	250	0.84	0.90	31		3		2		–	–	70	26	1,402	1,455
MEP-1W	10	1	Residential	75	0.89	0.63	0	0	19	23	1	2	–	–	70	27	1,129	1,173
SB-2S[b]	14	1	Residential	75	0.78	0.51	25	27	12	10	2	1	–	–	70	26	1,035	1,246
EN-1S	14	1	Residential	250	0.79	0.27	8	9	21	23	2	3	–	–	60	26	954	1,385
MET-1E	15	1	Residential	250	0.57	0.20	5	9	17	18	2	1	1[c]	1[c]	70	28	877	1,683
MET-1N[d]	15	1	Residential	250	0.93	0.67	2	3	67	65	0	0	–	–	70	33	1,359	1,454
RS-17	18	1	Residential	250	0.79	0.46	10		11		6		–	–	70	31	1,022	1,156
SB-1E[b]	18	2	Residential	75	0.85	0.31	27	27	8	8	1	2	–	–	70	36	1,370	1,605
EM-1E	18	1 and 2[e]	Residential	250	0.85	0.78	6	5	22	23	0	0	–	–	60	24	1,852	2,014
SB-1S[b]	20	2	Residential	75	0.81	0.59	11	11	25	27	1	1	–	–	70	25	2,308	2,570
RS-14(A)	20	2	Residential	250	0.80	0.12	0		0		3		5[c]		90	31	2,520	3,314
RS-14(B)	20	2	Residential	250	0.91	0.40	0		8		4		0[c]		90	31	2,616	3,227
RS-20	20	2[f]	Residential	250	0.89	0.78	4		7		2		0[c]		90	55	1,751	1,833
RS-12	10	1	CBD	250	0.87	0.21	0		2		3		15[c]		70	39	839	1,107
RS-23[g]	10	1	CBD	250	0.81	0.27	16		0		3		11[h]		70	32	783	978
MEP-2E[d]	11	1	CBD	75	0.87	0.41	62	59	15	17	6	6	–	–	70	29	1,077	1,260
MEP-2EII[d]	11	1	CBD	75	0.84	0.18	70	60	15	18	14	7	–	–	70	29	911	1,204
EB-1S	30	3	CBD	250	0.81	0.07[i]	41	39	3	3	8	5	3[c]	12[c]	80	48	1,792	2,200
EO-4E[j]	14	1	OBD	250	0.90	0.23	23	18	18	20	8	6	–	–	60	21	577	808
EO-4S	16	1	OBD[k]	250	0.76	0.27	3	2	5	8	11	8	–	–	60	31	862	1,191
MNB-1N	18	2	OBD	100	0.90	0.12	28	34	2	3	0	1	–	–	70	31	1,330	1,708
RS-13	21	2	OBD	250	0.75	0.02[j]	0		11		0		0[c]		70	42	1,150	2,317
MET-2W[d]	22	2	OBD	250	0.87	0.39	17	15	53	52	6	6	–	–	70	21	1,420	1,827
MET-2N	26	2	OBD[k]	250	0.83	0.31	20	18	1	2	6	6	–	–	70	39	1,977	2,294

Note: 1 ft = 0.3048 m.

[a] Average loaded cycle volume expanded to full volume.
[b] Data were not used because of large variation in green phase.
[c] Near-side stop.
[d] Data were not used because few samples had more than 60 percent turns.
[e] Data were not used because this operation combined 1- and 2-lane streets.
[f] Data were not used because 2-lane approach narrows to 1 lane downstream.
[g] Data were not used because left-turn lane is also used for through movement.
[h] Far-side stop.
[i] Data were not used because load factor was too low to allow accurate expansion of loaded cycle volume.
[j] Data were not used because of excessive delays during loaded cycles.
[k] Reclassified as CBD.

Table 2. Physical, environmental, and traffic characteristics of study locations for 2-way streets with parking.

Site	Approach Width (ft)	Lanes	Location	Population (thousands)	PHF	LF	Percent Turning Left: Expanded Loaded Cycle	Percent Turning Left: Peak Hour	Percent Turning Right: Expanded Loaded Cycle	Percent Turning Right: Peak Hour	Truck Percent: Expanded Loaded Cycle	Truck Percent: Peak Hour	Bus Percent: Expanded Loaded Cycle	Bus Percent: Peak Hour	Cycle (s)	Green Time (s)	Service Volume (vehicles/h green)	ALE Capacity[a] (vehicles/h green)
PC-3E[b]	16	1	Fringe	250	0.85	0.13	23	20	36	33	5	5	—	—	90	33	669	851
PC-3ERE	16	1	Fringe	250	0.83	0.13	12	15	8	20	8	5	—	—	90	33	885	1,287
PC-3WRE	16	1	Fringe	250	0.84	0.35	28	28	18	18	5	5	1[c]	2[c]	90	33	1,148	1,379
RS-27	17	1	Fringe	250	0.85	0.52	0		37		6		—	—	70	28	1,000	1,184
RS-29	17	1	Fringe	250	0.83	0.41	4		12		3		—	—	70	28	934	1,184
MNB-2W	17	1	Fringe[d]	100	0.71	0.12	13	12	10	9	3	3	—	—	70	30	697	1,249
RS-27II	17	1	Fringe	250	0.84	0.24	7	4	24	26	9	8	0[e]	2[e]	70	25	675	904
RS-29II	17	1	Fringe	250	0.78	0.16	25	25	9	13	8	6	0[e]	0[e]	70	25	888	1,142
MNB-4WII	18	1	Fringe	100	0.84	0.59	5	5	8	8	1	1	—	—	70	28	1,185	1,313
MNB-4WIII	18	1	Fringe	100	0.82	0.45	2	2	12	12	0	1	—	—	70	28	1,117	1,392
MNB-4W	18	1	Fringe	100	0.88	0.33	2	3	9	10	4	3	—	—	70	35	998	1,176
SS-1N	18	1	Fringe	75	0.90	0.18	7	11	16	13	4	4	—	—	80	36	813	1,175
PC-2E	18	1	Fringe	250	0.85	0.38	8	8	14	12	5	5	—	—	90	43	1,146	1,194
PH-1N	18	1	Fringe	100	0.96	0.40	2	3	2	4	5	5	—	—	80	40	1,328	1,386
PH-1NRE	18	1	Fringe	100	0.94	0.69	2	2	2	2	2	3	—	—	80	40	1,428	1,553
PC-2ERE	18	1	Fringe	250	0.75	0.52	10	10	9	9	2	2	2[c]	2[c]	90	43	1,218	1,451
MNB-4SII	19	1	Fringe	100	0.86	0.43	31	30	11	11	1	1	—	—	70	31	850	1,094
MNB-4SIII	19	1	Fringe	100	0.80	0.41	32	33	13	11	1	1	—	—	70	31	866	1,143
MNB-4S	19	1	Fringe	100	0.83	0.45	17	16	26	21	0	0	—	—	70	25	1,157	1,372
EI-1SRE[f]	19	1	Fringe	250	0.87	0.25	27	17	26	37	47	22	—	—	90	33	760	851
EI-1S	19	1	Fringe	250	0.84	0.45	22	25	18	15	1	2	—	—	90	33	1,145	1,394
UP-1SII	20	1	Fringe	75	0.91	0.13	3	3	12	10	5	4	—	—	80	36	933	1,250
UP-1WII	20	1	Fringe	75	0.87	0.07	4	10	6	8	2	4	—	—	80	36	1,147	1,633
EI-4W	20	1	Fringe[d]	250	0.90	0.55	10	9	7	8	4	3	—	—	70	28	1,348	1,557
CC-1W	20	1	Fringe	100	0.88	0.24	7	5	14	13	3	5	—	—	70	31	962	1,324
SP-1E	20	1	Fringe	75	0.82	0.22	12	11	15	13	5	3	—	—	60	30	852	1,182
SP-1S	20	1	Fringe	75	0.89	0.28	17	18	16	20	2	1	—	—	60	23	929	1,160
RS-30-a	20	1	Fringe	250	0.81	0.41	13		17		4		—	—	70	32	1,022	1,267
RS-30-b	20	1	Fringe	250	0.83	0.58	16		14		3		—	—	70	32	1,044	1,244
RS-31	20	1	Fringe	250	0.82	0.41	4		10		4		0[c]		70	36	1,118	1,372
RS-36	20	1	Fringe	250	0.81	0.18	7		7		1		6[c]		70	36	1,104	1,333
RS-30II	20	1	Fringe	250	0.87	0.39	18	17	19	18	6	5	—	—	70	31	1,007	1,244
PC-3SRE	20	1	Fringe	250	0.92	0.15	19	15	10	7	9	11	0[e]	0[e]	90	51	954	1,259
RS-34	21	1	Fringe	500	0.86	0.75	8		25		1		0[c]		90	29	1,247	1,327
PP-1WRE	22	1	Fringe	250	0.88	0.68	11	12	6	6	1	1	—	—	60	24	1,162	1,324
PP-1W	22	1	Fringe	250	0.86	0.70	16	16	9	9	5	5	—	—	60	24	1,105	1,246
RS-38	22	1	Fringe	250	0.85	0.88	3		9		5		6[c]		70	28	1,350	1,375
RS-37	22	1	Fringe	250	0.94	0.35	7		2		1		3[c]		70	42	1,467	1,583
RS-37II	22	1	Fringe	250	0.83	0.06[g]	12	8	10	6	6	6	—	—	70	42	840	1,388
RS-38II	22	1	Fringe	250	0.94	0.63	14	12	12	13	7	7	—	—	70	28	1,140	1,235
PP-1S	23	1	Fringe	250	0.85	0.10	6	10	6	8	5	5	—	—	60	30	900	1,260
PP-1N	23	1	Fringe	250	0.83	0.65	7	8	23	22	5	5	—	—	60	30	1,058	1,185
PP-1NRE	23	1	Fringe	250	0.88	0.83	8	8	18	18	2	3	—	—	60	30	1,158	1,217
PP-1E	23	1	Fringe	250	0.87	0.58	8	7	4	4	5	5	—	—	60	24	1,172	1,371
CC-1N	24	1	Fringe	100	0.87	0.33	2	3	9	6	3	3	1[c]	5[c]	70	31	1,406	1,666
MNB-5E	24	1 and 2[h]	Fringe	100	0.84	0.47	9	8	11	12	1	2	—	—	70	25	1,381	1,618
RS-32	25	1 and 2[h]	Fringe	250	0.87	0.18	26		15		1		—	—	70	31	1,512	2,032

Table 2. Continued.

Site	Approach Width (ft)	Lanes	Location	Population (thousands)	PHF	LF	Percent Turning Left: Expanded Loaded Cycle	Percent Turning Left: Peak Hour	Percent Turning Right: Expanded Loaded Cycle	Percent Turning Right: Peak Hour	Truck Percent: Expanded Loaded Cycle	Truck Percent: Peak Hour	Bus Percent: Expanded Loaded Cycle	Bus Percent: Peak Hour	Cycle (s)	Green Time (s)	Service Volume (vehicles/ h green)	ALE Capacity[a] (vehicles/ h green)
EN-1W	18	1	Residential	250	0.88	0.15	17	14	19	16	3	3	–	–	60	28	1,004	1,400
MNB-3W	18	1	Residential	100	0.85	0.61	35	32	6	8	2	1	–	–	70	38	1,050	1,127
SB-2E[i]	18	1	Residential	100	0.85	0.24	10	8	8	8	3	2	–	–	70	37	1,159	1,496
MNB-3N	20	1	Residential	100	0.86	0.55	14	13	3	2	1	1	–	–	70	25	1,179	1,382
PP-6NRE	20	1	Residential	250	0.83	0.05[g]	12	12	0	2	0	1	–	–	60	33	727	1,491
PP-5E	20	1	Residential	250	0.82	0.15	24	30	26	25	5	5	–	–	60	19	611	1,137
PP-4W[b]	20	1	Residential	250	0.81	0.20	12	13	17	19	5	5	–	–	60	20	697	1,035
PP-5W	20	1	Residential	250	0.82	0.22	22	17	10	14	5	5	–	–	60	19	778	1,253
PP-4WRE	20	1	Residential	250	0.84	0.18	14	15	19	20	5	2	–	–	60	20	946	1,358
PP-5WRE	20	1	Residential	250	0.84	0.27	8	8	14	12	4	4	–	–	60	19	1,063	1,397
PP-4E	20	1	Residential	250	0.79	0.17	38	39	13	15	5	5	–	–	60	20	586	1,152
PP-6N[b]	20	1	Residential	250	0.72	0.17	9	6	13	11	5	5	–	–	60	33	624	993
MNB-2S	22	1	Residential	100	0.84	0.22	10	7	21	24	5	3	–	–	70	30	1,296	1,634
PP-6W	22	1	Residential	250	0.87	0.27	27	20	14	13	5	5	–	–	60	21	1,131	1,402
PP-6WRE	22	1	Residential	250	0.92	0.30	19	19	12	14	3	3	–	–	60	21	1,214	1,467
PP-2E[j]	17	1	CBD	250	0.85	0.52	0	0	10	9	5	5	–	–	90	33	740	873
MNB-1W	19.5	1	CBD	100	0.87	0.20	12	11	16	19	6	2	–	–	70	29	978	1,145
UP-1S	20	1	CBD[k]	75	0.83	0.18	5	12	11	13	4	4	–	–	80	36	1,067	1,400
UP-1W	20	1	CBD[k]	75	0.83	0.18	11	9	9	10	2	5	0[c]	3[c]	80	36	1,147	1,762
RS-33	20	1	CBD	250	0.86	0.58	0		18		5		18[c]		70	32	1,152	1,239
MEP-4S	21	1	CBD	75	0.90	0.29	0	0	0	0	6	8	5[c]	19[c]	70	42	927	1,139
MEP-4SII	21	1	CBD	75	0.88	0.10	0	0	0	0	7	5	0[c]	0[c]	70	42	783	1,190
PP-3N[j]	22	1	CBD	250	0.90	0.20	0	0	0	0	5	5	0[c]	0[c]	90	53	1,292	1,622
RS-26	25	1	CBD	250	0.91	0.10	0		12		5		4[c]		120	47	1,200	1,436
MEP-4N	26	1 and 2[h]	CBD	75	0.92	0.37	0	0	28	25	2	2	–	–	70	42	1,275	1,445
MEP-4NII	26	1 and 2[h]	CBD	75	0.88	0.12	0	0	27	28	8	6	–	–	70	42	1,102	1,346
MEP-2S	28	1 and 2[h]	CBD	75	0.86	0.24	3	4	42	38	8	6	–	–	70	36	1,307	1,702

Note: 1 ft = 0.3048 m.

[a]Expanded to full volume.
[b]Data were not used because loaded cycle data were inaccurately tabulated.
[c]Near-side stop.
[d]Reclassified as residential.
[e]Far-side stop.
[f]Data were not used because of the high percentage of trucks.
[g]Data were not used because the load factor was too low to allow accurate expansion of loaded cycle volume.
[h]Data were not used because this operation combined 1- and 2-lane streets.
[i]Data were not used because of the large variation in green phase.
[j]Data were not used because of the unusual layout.
[k]Reclassified as fringe.

from prior studies), and changes are made accordingly. Estimates of capacity again are made by using the revised factors; again, this is done without a correction factor being applied for PHF.

4. The estimates of HCM capacity derived from these 3 steps are compared to the ALE values for statistical differences.

5. ALE capacities as determined from the field conditions by those basic conditions defined in the HCM charts (2, Figures 6.8 and 6.9) as revised in the third step are adjusted. A new curve should be computed for the 1.0 LF condition with the adjusted ALE values.

The resulting set of curves and adjustment factors can be checked accurately only with an entirely new set of field data. However, the utility of the HCM method of capacity estimation is currently under question and soon may be completely revised.

Although LF data were included in the data collected, no examination of the HCM estimate at various levels of service was attempted for 2 reasons. First, it was not possible to define adequately a loaded cycle such that 2 individuals would each have equal assurance of its loaded condition. The resulting situation would reflect a wide variation in the load factor for equal service volumes. Although the level-of-service variation would be large, the average number of vehicles/loaded cycle would not be large, which would thus maintain reliable capacity estimates.

The second reason is the variability encountered in unloaded cycles. If one assumes that the loaded cycles were accurately defined, then the unloaded cycles could contain any number of vehicles up to the number needed to load a cycle.

The variability results from an inability to measure the volume-to-capacity ratio of individual unloaded cycles. For a given period of time at a given intersection, this ratio will not vary significantly. However, when many intersections studied at different periods are considered, this ratio can range from near 0 to 1. Because of this, it would be possible for 2 intersections with similar physical and environmental characteristics to have great differences in volumes with equal load factors.

Data Analysis—Bellis Method

The revisions made to HCM parameters also were made to those parameters that apply to the Bellis method. The most prominent correction was for turns. Otherwise the procedure for estimating capacity by this method is similar to that outlined by Reilly and Seifert (3).

Revisions to HCM Parameters

Turning Movements

The quantified adjustments for turns (2, Tables 6.4 and 6.5) reflect basic approach volume decreases for increases in the width of approach for 0 percent turns at 2 points of each curve. Figures 1 and 2 show the actual differences in service volume at the smaller widths. To overcome this apparent disparity for less than 10 percent turns, the turn factors of the HCM will be discarded for lower approach widths up to 10 percent turns. For example, for a 14-ft (4.3-m) approach with no parking, the factors for 16 to 24 ft (4.9 to 7.3 m) will be used for 0 to 10 percent turns; the factors for 15 ft (4.6 m) will be used for more than 10 percent turns.

Metropolitan Area Size

Three metropolitan area size choices exist for the engineer:

Figure 1. HCM turn factors for 2-way streets with no parking.

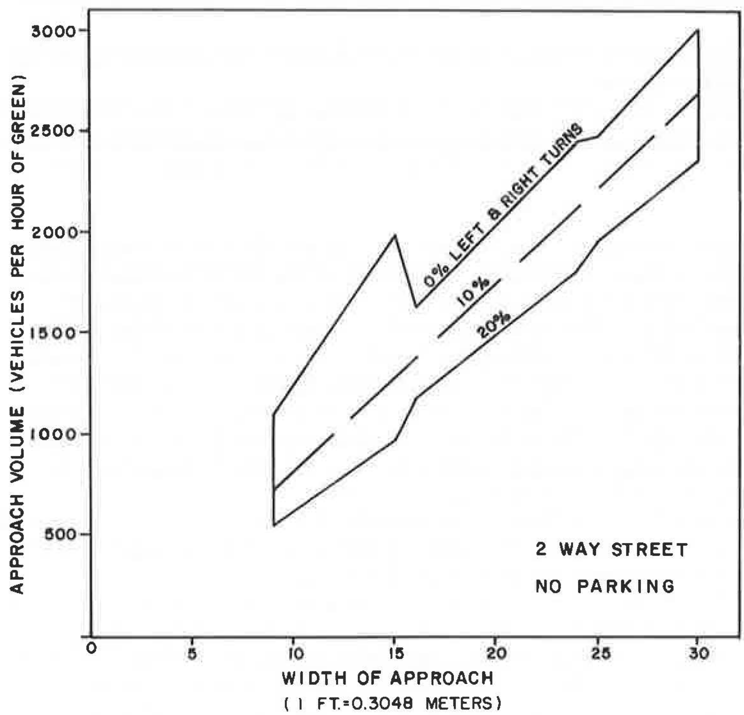
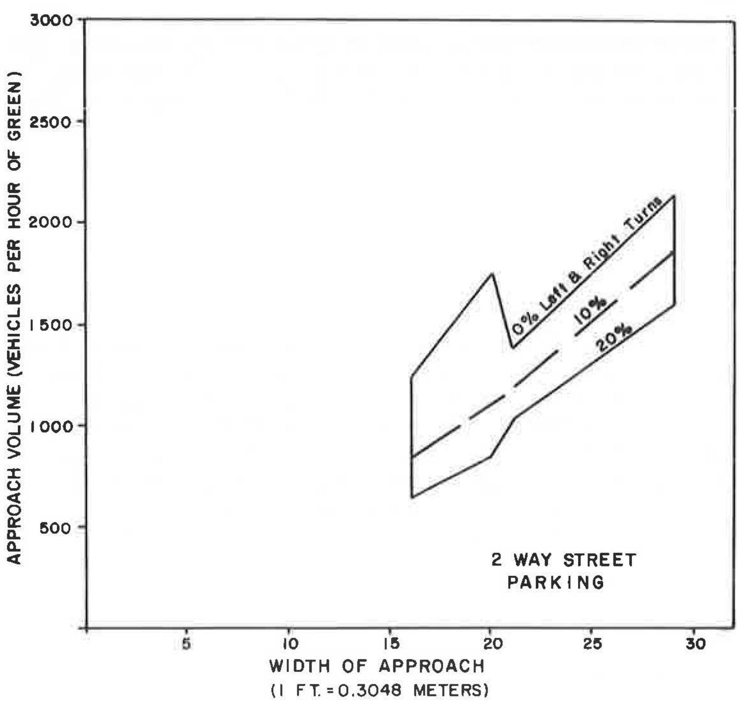


Figure 2. HCM turn factors for 2-way streets with parking.



1. Population of the municipality itself,
2. Population of the municipality and the surrounding municipalities, or
3. Population of the region in which the municipality exists.

The first 2 choices are straightforward, but judgment is called for in making the third choice. For the current study, all 3 choices were tried and the regional approach made use of 3 main areas: New York City (more than 1 million), Trenton (1 million), and Philadelphia (more than 1 million). The HCM estimate of capacity for all samples was tabulated by using each of the 3 choices but without adjusting for PHF. The mean and standard deviation then were compared for each choice.

The smallest factor used in this study was 75,000 because it is the lower limit to population found in the HCM charts.

One last difference for the peak-hour and metropolitan population factors is the adjustments for 2-way streets with no parking. These adjustments are 3 to 4 percent lower than those on all the other charts. No tests were made on the data to check this difference because it is too small in relation to the magnitude of adjustments for all other parameters.

Location Within Metropolitan Area

Two primary reasons exist for adjusting the HCM factors for metropolitan location. As indicated by Chang and Berry (7), the disparity in the basic approach volume curves between 1-way and 2-way streets is made even greater when fringe-area adjustments are applied (Figure 3). In addition, the HCM description of the various areas defines levels of pedestrian and commercial vehicle activity in the different areas. So, for this study, the following factors are used with the basic HCM curves for 2-way streets:

Location	Factor
CBD	1.00
Fringe	1.10
OBD	1.15
Residential	1.25

These factors do not consider the pressure of traffic in the busy areas, but they attempt to equalize the disparity between approaches, 1-way streets, and 2-way streets according to the HCM description of these areas.

Local Transit Buses

The Reilly and Seifert study (3) indicated that large HCM capacity estimate errors resulted when bus adjustment factors were used for near-side bus stops on streets with parking. For the study discussed in this paper, adjustment factors for near-side bus stops with parking were cut off at 1.0.

Parking Conditions

An attempt was made to overcome some of the judgment problems that exist in determining whether an approach has parking and to what extent parking affects capacity. The sites with parking are plotted by width against ALE capacity in Figure 4. A visual inspection of the distance of parking from the stop line would be incomplete without including the associated percentages of left and right turns and the indication of a bus stop.

Figure 3. HCM factor for metropolitan location.

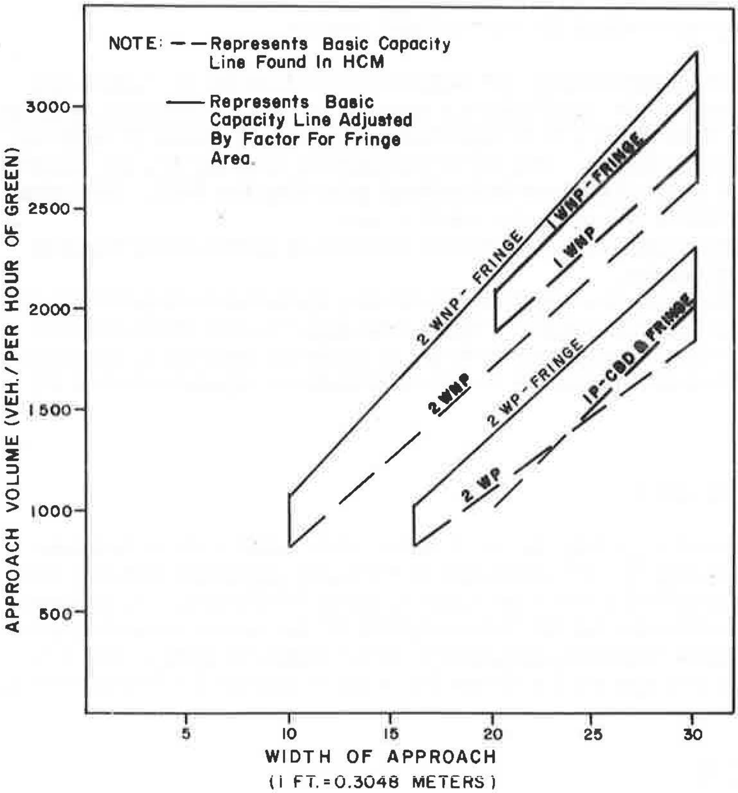
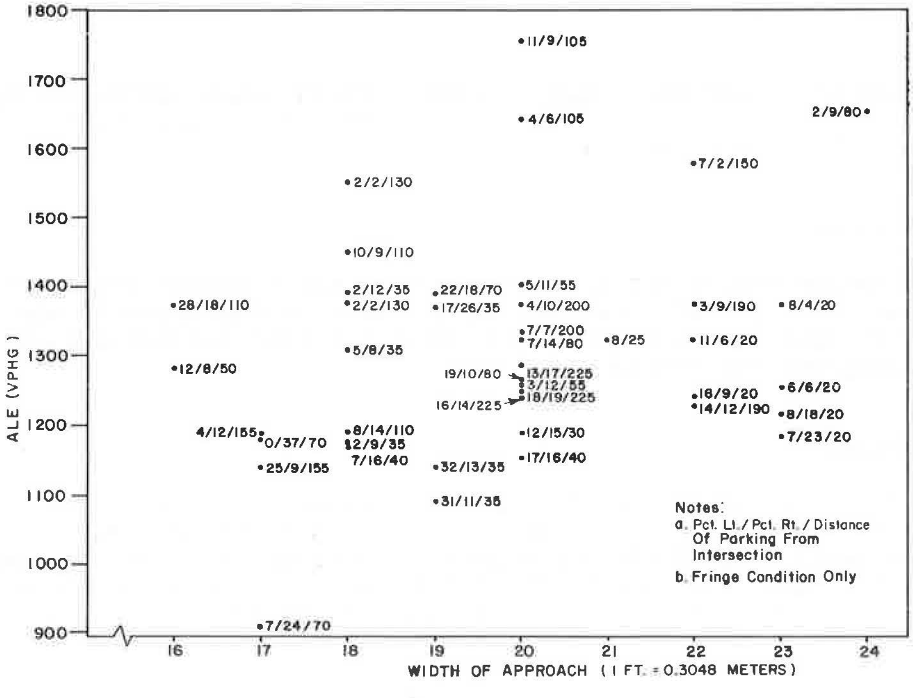


Figure 4. Parking conditions.



The impedance to traffic movement that is offered by parked vehicles is similar to that of a narrow street. But parked vehicles appear to have less effect at 200 ft (61 m) than at 100 ft (30.5 m).

A multiple linear regression analysis was made of the aforementioned variables to determine the effect of the distance of parking from the intersection.

RESULTS

Highway Capacity Manual

Error of Estimate

Mean and standard deviation percentage differences of the HCM estimates and ALE are as given in Table 3. The error of the unrevised HCM was expected to be negative, assuming that the correction factors in the HCM are accurate. The reason to expect a negative error is that the HCM uses an adjustment for PHF that is similar to the PHF itself; hence an adjustment that is less than 1.0 is made to the approach volume. The comparable volume for ALE has no adjustment for the peaking effect of traffic at the intersection approach.

The second set of errors listed in Table 3 is for the HCM estimate without an adjustment for PHF. Hence it should have a zero error with ALE, assuming that the correction factors in the HCM are accurate. Because the data used to develop the unrevised HCM estimate had a PHF average of approximately 0.85, the error for the HCM without PHF adjustment can be expected to be approximately 15 percent higher than the unrevised HCM method. However, the error percentage shown indicates that the capacity estimate for the HCM without PHF adjustment is from 20 to 25 percent higher than actual field conditions indicated. Moreover, the error as shown gives no indication of which parameters and to what extent their adjustments are inaccurate.

The third set of errors are those for revised HCM estimates with no adjustment for PHF. Adjustments can be made for the PHF at the discretion of the user with the understanding that they represent a subjective reduction in the intersection's ability to handle traffic on an hourly basis. In effect, we made no peak-hour adjustment if we were willing to let drivers wait on the approach. The effectiveness of the adjustments to the HCM factors is evident when the error estimates of revised HCM and HCM without PHF adjustments are compared. Further comparison can be made in Figure 5.

There is no conclusive evidence that adjustments we chose were the best ones to make. To more accurately test the effect and subsequent adjustments for some of the factors in the HCM, one would have to control the variation of a parameter at a single intersection approach. The ability to exert this control may never be within a researcher's power because 10 distinct parameters are used to vary the estimate of capacity in the HCM. Only 2 parameters could be tested in this study: parking and metropolitan area size.

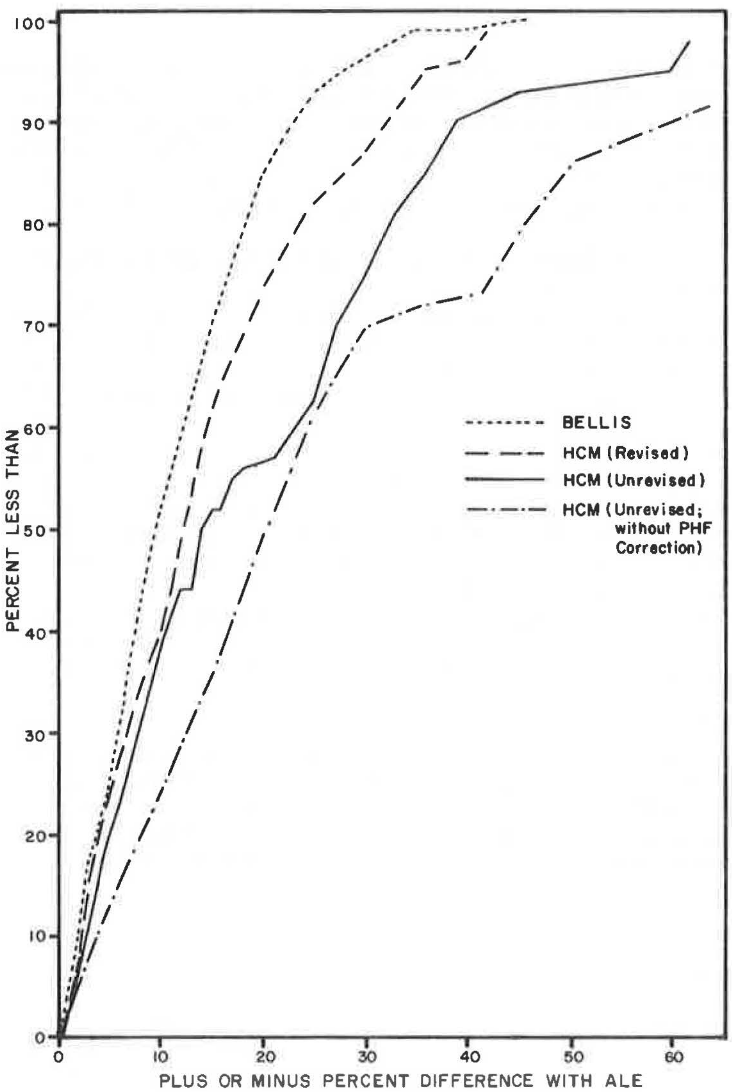
Distance of Parking to Intersection

We have defined as closely as possible those features common to an intersection in an attempt to determine the effect of parking on an intersection's approach. A multiple linear regression analysis of the data shown in Figure 4 showed that the use of the distance to parking from the stop line had no significant effect in a regression equation defining ALE. However, the previous reasoning that implied an improved ability of an approach to handle demand when parking was removed to 200 ft (61 m) [as opposed to 20 ft (6.1 m)] is more logical than the statistical results indicate. The results of the regression analysis highlight the need for a more controlled testing of this parameter.

Table 3. Average error of 1965 Highway Capacity Manual capacity estimates.

	2-Way, No Parking		2-Way, Parking	
Estimate Method	Error Percent	Standard Deviation	Error Percent	Standard Deviation
Unrevised HCM	+10	27	+7	24
HCM without PHF adjustment	+23	30	+22	25
Revised HCM	+4	24	+5	17

Figure 5. Cumulative frequency curves of capacity estimation errors.



Metropolitan Area Size Factor

Another comparison was made on the choice of any of 3 areas for the metropolitan area size. The error percentage of HCM versus ALE when these values are used is as follows:

Area	2-Way, No Parking (percent)	2-Way, Parking (percent)
Municipality	17	17
Municipality and surrounding area	23	22
Region	40	40

As might be expected, the difference in error percentage for the first 2 areas is approximately equal to the difference in factors. However, the error for the third area far exceeds the proportional difference of the factors. Essentially, the logic for using a metropolitan area size factor equivalent to population of the municipality and the surrounding area appears to be reason enough to overlook the small difference in error percentage.

Adjustment to Approach Volume Curves

The final step in the HCM analysis was the attempted derivation of a revised approach volume curve by using the values for ALE modified by the adjusted HCM parameters, which would thus allow a direct comparison to the HCM curves. A review of data available elsewhere (9, Tables 4 and 5) indicates the need for repositioning the basic curves relative to approach width. Table 4 in this report (9) has 1- and 2-lane data combined. For example, the first 7 fringe sites and the 9th fringe site were used as 1-lane approaches, and the 8th and 10th through 13th sites were used as 2-lane approaches. Within each of these groupings, the approach width increased from 14 ft (4.3 m) to 19 ft (5.8 m) for the 1-lane approaches and from 19 ft (5.8 m) to 26 ft (7.9 m) for the 2-lane approaches. The regression analyses grouped the data accordingly and were performed for each of the following groupings:

1. Two-way streets with parking for fringe areas only (40 samples);
2. Two-way streets with parking for all sites combined (56 samples);
3. Two-way streets with no parking for 1-lane approaches (12 samples);
4. Two-way streets with no parking for 2-lane approaches (8 samples); and
5. Two-way streets with no parking for all sites combined (20 samples).

The error percentage of the revised HCM method varied from a negative error for the streets with low widths to a positive error for streets with high widths. In effect, this would indicate a flattening of the basic approach volume curve. Figure 6 shows the results of this analysis against the background of the curves currently shown in the HCM.

The adjusted ALE curves for 1- and 2-lane 2-way streets with no parking yield reasonably high correlation coefficients. However, the curve for 1-lane streets shows a decrease in capacity for an increase in width. The curve for 2-lane streets also is questionable.

Extremes in the values for ALE for the higher and lower width streets account for the shape of the computed curves.

Inspection of ALE capacities as collected under field conditions (Tables 1 and 2) shows why the adjusted curves are questionable. Under constant physical and

Figure 6. Capacity curves.

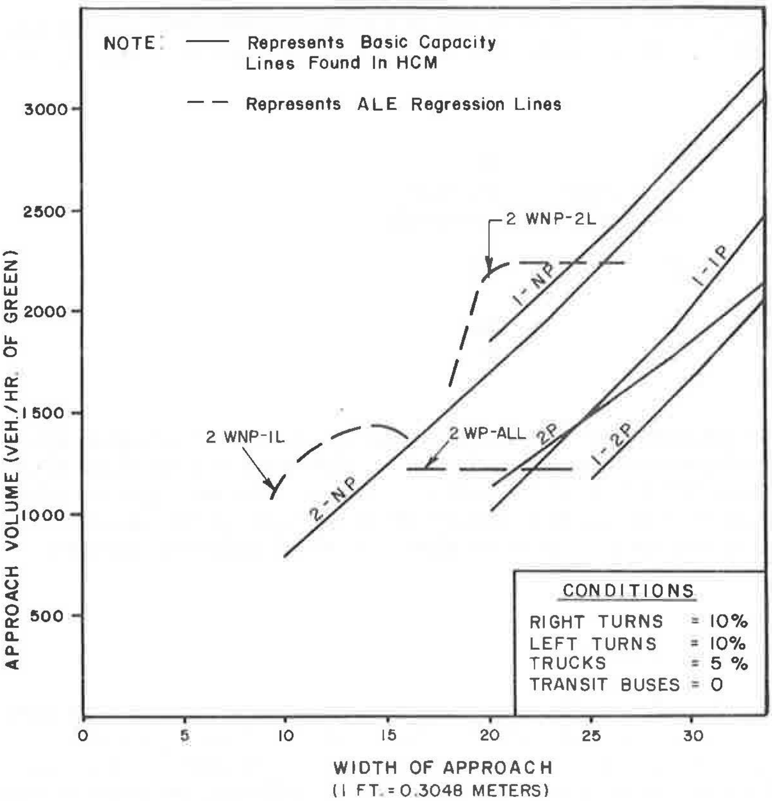
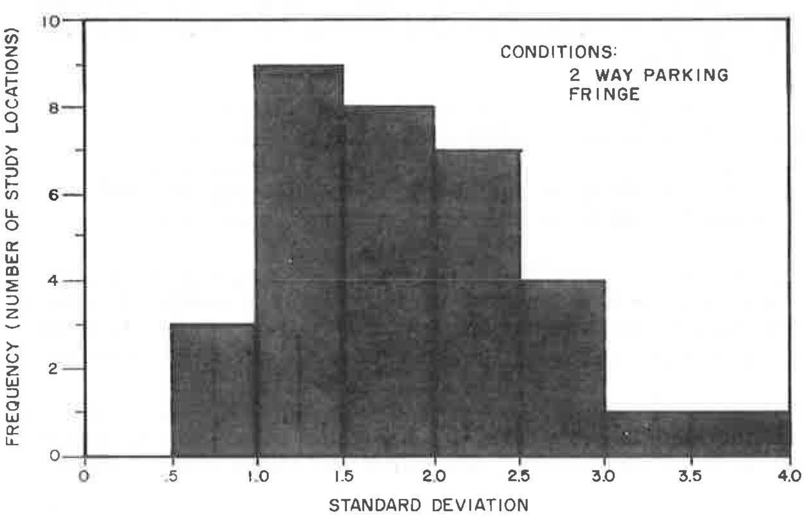


Figure 7. Frequency distribution of standard deviation of number of vehicles/loaded cycle.



environmental conditions, there is great variability in the actual capacity (as measured by ALE) for similar types of intersection approaches. There are several cases in which increased percentages of turns yield increased capacities. This appears to be contrary to the rationale behind the HCM turn correction factors, which holds that turns have a decreasing effect on capacity.

The variability found for the actual capacity of similar intersection approaches leads to the conclusion that capacity estimation, in its present state, can be subject to large inaccuracies. It is the opinion of the authors that these inaccuracies render the HCM useful for design estimates only. The capacity of specific locations should be determined by using intersection data and not the generalized information in the HCM.

Bellis Method

As defined by Reilly and Seifert (3) and Bellis (4), the Bellis method for estimating capacity does not consider street width and parking. It considers only the type of street and number of lanes. Although there are 4 types of approaches for Bellis curves, only the CBD (type 1) and those streets outside the CBD that allow both right and left turns (type 2) had sample sizes sufficient for consideration in this study. The error percentages of capacity estimate are as follows:

Type	Number of Samples	Error Percentage	Standard Deviation
1	10	-19	8
2	72	2	13

These errors are almost identical to those found by Reilly and Seifert (3). The resulting upward adjustment to the type 1 capacity curve would be sufficient to satisfy the discrepancy. Although the standard deviation of the error percentage for type 2 approaches is high (indicating the need for further refinement to the technique), the standard deviation is far less than that experienced with the HCM capacity estimates.

Data Collection

A very low number of intersections were used compared to the number that were studied. Hence an extremely costly and time-consuming effort would be required to extend the data collection effort in this study. Only 56 percent of all sites samples were used, and there would have to be substantial reasons to continue this study under the initial data collection format. A more productive approach may be found by varying parameters at individual sites.

Some of the reasons that data could not be used are as follows:

Reason	Percent
Zero load factor	13
Inaccurate	8
Lane use variation	5
High percentage of turns or trucks	5
Large green-phase variation	4
Miscellaneous	9

The variation of the number of vehicles/loaded cycle for any 1 intersection approach can be defined by the standard deviation of the number of vehicles/loaded cycle for that approach. Figure 7 shows the frequency distribution of the standard deviations for the intersections in the fringe-area 2-way street with parking category. As expected, a skewed distribution to the right results. There was absolutely no correlation of the standard deviation with the average number of vehicles/loaded cycle. Hence the distribution of the standard deviation, as shown in Figure 7, could result for most of the range of average loaded-cycle data (6.5 to 17.5 vehicles/loaded cycle). Inspection of the distribution only substantiates the variability of cycle-to-cycle activity that could result at an intersection.

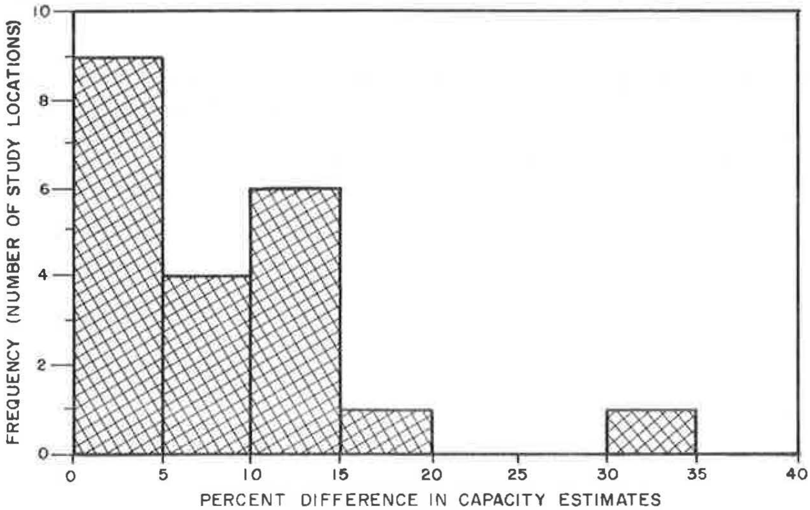
An indication of the repetitiveness of the field measurement of the actual capacity of an intersection approach, which is expressed as ALE, is shown in Figure 8. Eighteen intersections were sampled a second and third time. The percentage difference between repeat samples, as indicated by the frequency distribution of Figure 8, averaged 10 percent. Expressed in other terms, 90 percent of the repeated estimates of capacity (made from field measurements) were within 10 percent of the original estimate. The variation of field measurement of capacity for any 1 site can only be hypothesized, but the reasons may vary from differences in drivers through environmental and traffic changes.

SUMMARY AND CONCLUSIONS

The accuracy of the HCM in estimating capacity of approaches at 2-way signalized intersections was tested. In addition, revisions were made to 4 of the factors used in the HCM, and the accuracy of the HCM method was tested by using the revised factors. As a result, certain revisions to the HCM factors are suggested.

1. Reduce turn corrections for narrow approaches [less than 15 ft (4.6 m) for no parking areas and less than 20 ft (6.1 m) for parking areas] for up to 10 percent turns.
2. For estimating the population, use the population of the municipality plus the population of the surrounding municipalities in dense suburban areas.
3. Use factors for metropolitan area 2-way street locations that are consistent with the factors for 1-way streets.
4. Use a maximum correction factor of 1.0 for near-side bus stops.

Figure 8. Frequency distribution of percentage of difference in capacity estimates.



Twenty percent is the accepted error range. Results indicate that the unrevised HCM estimating method has errors in excess of approximately 20 percent for half of the study samples. The revised HCM method has errors in excess of approximately 20 percent for a quarter of the samples.

The large variation that was found to exist in the capacity of signalized intersections could not be satisfactorily explained even though certain parameter adjustments to the HCM resulted in dramatic reductions for error percentage estimates. We can only conclude that it is far more preferable to measure capacity in the field than it is to estimate from the HCM.

The sample data also were used to test the accuracy of the modified Bellis method of data collection, and it was found that an error in excess of approximately 20 percent existed for 15 percent of the sites.

Data at 150 sites were collected, but less than 60 percent of the sites had data that were considered useful. It is evident that continued efforts to extend this program by using the current methods of data collection would be extremely time-consuming.

ACKNOWLEDGMENT

This paper was prepared in cooperation with the Federal Highway Administration.

REFERENCES

1. Highway Capacity Manual. Bureau of Public Roads, U.S. Department of Commerce, 1950.
2. Highway Capacity Manual. HRB Special Rept. 87, 1965.
3. E. F. Reilly and J. Seifert. Capacity of Signalized Intersections. Highway Research Record 321, 1970, pp. 1-15.
4. W. R. Bellis. Capacity of Traffic Signals and Traffic Signal Timing. HRB Bulletin 271, 1960, pp. 45-67.
5. O. K. Norman. Variations in Flow at Intersections Related to Size of City, Type of Facility and Capacity Utilization. HRB Bulletin 352, 1962, pp. 55-99.
6. An Evaluation of the 1965 HCM Method of Computing Intersection Capacities. Missouri State Highway Commission, Study 68-2, Aug. 1969.
7. Y. B. Chang and D. S. Berry. Examination of Consistency in Signalized Intersection Capacity Charts of the Highway Capacity Manual. Highway Research Record 289, 1969, pp. 14-24.
8. A. D. May and D. Pratt. A Simulation Study of Load Factor at Signalized Intersections. Traffic Engineering, Feb. 1968.
9. E. Reilly, I. Dommasch, and M. Jagannath. Capacity of Signalized Intersections. Division of Research and Development, New Jersey Department of Transportation, 1974.

EVALUATION OF OFF-LINE TRAFFIC-SIGNAL OPTIMIZATION TECHNIQUES

Leonard Rach, Joseph K. Lam, David C. Kaufman, and David B. Richardson, Metropolitan Toronto Roads and Traffic Department

Four off-line traffic-signal optimization techniques (SIGOP, TRANSYT, Combination Method, and a preferential-street program that is SIGRID-based) were evaluated in both a suburban area and central area network environment within metropolitan Toronto. For each network, fixed-time and time-of-day signal timing patterns were developed for the 7:00 a.m. to 9:00 a.m., 10:00 a.m. to 12:00 noon, 1:00 p.m. to 3:00 p.m., and 4:00 p.m. to 6:00 p.m. time periods. To evaluate the various timing patterns, network travel time, and delay, researchers collected stop and volume data over a 12-week period in the fall of 1973. These data served as the base for a series of comprehensive statistical analyses oriented primarily toward a network evaluation of travel time and service rate. The data later were evaluated by a link-by-link paired comparison analysis. The network analysis of travel time and service rate did not provide conclusive results because of the nature of the study data. On the other hand, the link-by-link paired comparison analyses were more conclusive, relatively simple to use, and easy to interpret. Although the Combination Method settings provided Toronto motorists with a slightly better on-street performance level, any 1 of the 4 methods can provide reasonable signal-network settings.

•OVER the past several years there has been significant progress achieved in the development of off-line optimization programs for pretimed signal networks. The following 3 computer programs are widely regarded as the best approaches available (1):

1. Traffic signal optimization program (SIGOP) (2, 3, 4), which was developed under the sponsorship of the Federal Highway Administration;
2. Traffic network study tool (TRANSYT) (5, 6, 7, 8), which was developed cooperatively by Plessey Automation and the Road Research Laboratory of Great Britain; and
3. Combination Method (9, 10, 11, 12), which was developed by the Road Research Laboratory of Great Britain and later modified by the Greater London Council.

SIGOP contains a split calculation routine in which green times at each signal are computed in proportion to their respective critical-lane flows, total approach flows, or combinations of both. It also contains an offset optimization algorithm that minimizes the discrepancy between the actual signal offsets and a set of given or calculated ideal offsets. The resultant optimized settings are evaluated in terms of delay, stops, and cost values.

TRANSYT consists of a traffic-flow model that computes network flow patterns and associated delay and stop values and allows for platoon dispersion. It also includes a hill-climbing optimization procedure that optimizes splits and offsets alternately by minimizing a network performance index expressed as an aggregate function of link delays and stops. Computing an initial set of splits, which is optional, is based on the Webster method.

The Combination Method, in the form used in this study, does not contain a split computation routine and applies only to condensable networks. Given a set of traffic-signal splits, the program calculates a relationship of delay to difference of offset for each link, and, by combining all of these link functions to form an overall system function of delay to difference of offset, it selects an array of offsets that gives minimum system delay. A combination of delay and stops also may be used as the optimization criterion.

During the fall of 1973, the Metropolitan Toronto Roads and Traffic Department in cooperation with the Ontario Ministry of Transportation and Communications and the Federal Transportation Development Agency compared on-street performance of optimized settings produced by SIGOP, TRANSYT, and the Combination Method with the performance of Toronto's SIGRID-based, fixed-time, time-of-day settings (13). To properly evaluate the various optimization techniques over a range of varying conditions, they selected 4 time periods to cover both peak-hour periods (7:00 a.m. to 9:00 a.m. and 4:00 p.m. to 6:00 p.m.) and representative off-peak periods (10:00 a.m. to 12:00 noon and 1:00 p.m. to 3:00 p.m.). In addition, 2 subnetworks of signals were selected within metropolitan Toronto's grid network of 240 square miles (624 km^2) and 1,100 traffic signals to field test the various signal settings under actual operating conditions. These 2 areas contained a wide range of activities and representative traffic control situations found in most urban areas. The central area subnetwork, as shown in Figure 1, is a grid approximately 2.5 by 1.5 miles (4.0 by 2.4 km) with 68 traffic signals, 174 links, and 245 loop detectors. Within this area, land use varies from high-density commercial activities along Bloor Street to outlying business and medium- to high-density residential neighborhoods between St. Clair and Eglinton Avenues. Traffic signal spacing varies from 460 ft (140 m) to 3,973 ft (1211 m); average spacing is 1,657 ft (505 m). High pedestrian activity and stable volume flows are other typical characteristics of this area.

The second subnetwork is located in a suburban area and contains light industrial-commercial development and relatively low-density residential neighborhoods. Figure 2 shows the 3.5 by 2.5-mile (5.6 by 4.0-km) suburban grid of 51 traffic signals, 114 links, and 257 loop detectors. There are 2 freeways that bisect this subnetwork causing arterial volumes to have a tendency to fluctuate from day to day depending on freeway level of service. Average signal spacing in this area is 2,742 ft (836 m) ranging from 541 ft (165 m) to 6,875 ft (2,096 m).

STUDY METHODOLOGY

For each subnetwork, 3 signal timing plans, encompassing the morning peak period, the evening peak period, and the midday off-peak period, were generated by SIGOP, TRANSYT, and the Combination Method. Rather than predetermine cycle length, we tested a range of cycle lengths for each optimization time period (OTP). However, the magnitudes of the trial cycle lengths were chosen with reference to the existing system and practical constraints such as pedestrian walk requirements and storage problems. The evaluation block within each program was employed to evaluate the potential performance of candidate cycle lengths and associated signal settings. Vehicular delays and stops were used as evaluation criteria, and the optimum solution was the set of cycle lengths, splits, and offsets that provided the lowest system performance index given by total system delay plus 4 times the total number of stops. Table 1 gives a summary of the cycle lengths that were evaluated for each OTP together with the optimum cycle length selected for field evaluation in the central and suburban areas. For each plan, a small number of intersections had to be isolated from the system and be given a higher cycle length to satisfy pedestrian and special phasing requirements. In the case of the Combination Method, a few links were deleted manually from the test subnetworks so that the network could be condensed.

Although the input requirements for the various off-line optimization programs were generally similar in nature from program to program, it is interesting to compare the manpower requirements and computer processing time on Toronto's Univac 1107 system. Knowing these requirements is necessary to prepare typical network traffic-signal plans.

Table 2 gives a summary of average computer time required for processing 1 pro-

Figure 1. Central area subnetwork.

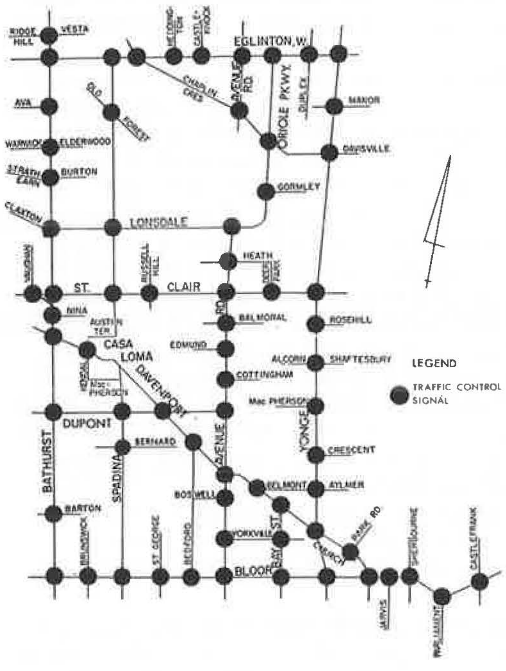


Figure 2. Suburban area subnetwork.

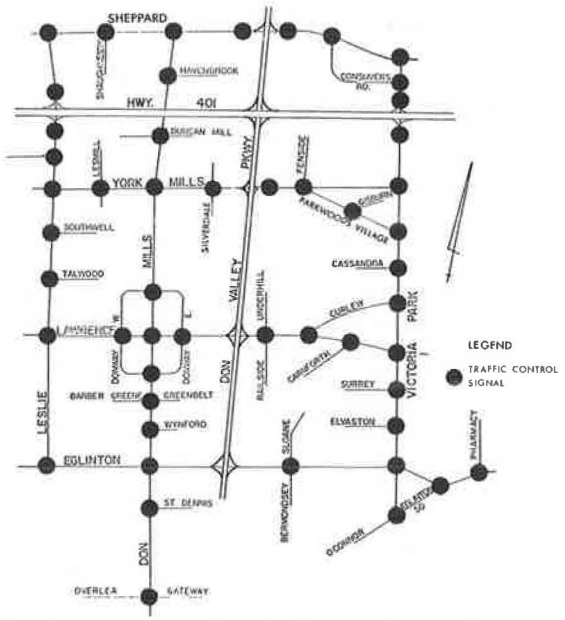


Table 1. Cycle length selection.

		Central Area			Suburban Area		
		Cycle Length (s)			Cycle Length (s)		
Program	Time Period	Tried	Selected	Exceptions	Tried	Selected	Exceptions
Existing plan	Morning peak	–	80	1	–	80 to 110	–
	Off peak	–	60 to 70	0	–	70 to 90	–
	Evening peak	–	80	3	–	70 to 120	–
SIGOP	Morning peak	60, 65, 70, 75, 80, 85, 90	80	1	65, 70, 75, 80, 85, 90, 95, 100	95	2
	Off peak	55, 60, 65, 70, 75, 80	65	0	65, 70, 75, 80, 85, 90	80	4
	Evening peak	70, 75, 80, 85 90, 95	80	3	70, 75, 80, 85, 90, 95, 100, 110, 120	90	7
TRANSYT	Morning peak	80, 90, 100, 120	90	1	95, 100, 110	100	2
	Off peak	60, 65, 70, 80	70	1	70, 80, 90, 100	80	2
	Evening peak	70, 80, 90, 100	90	1	95, 100, 110	100	5
Combination Method	Morning peak	80, 90, 100	80	2	95, 100, 110	100	2
	Off peak	65, 70, 80	70	2	80, 90, 100	80	2
	Evening peak	80, 90, 100	80	2	95, 100, 110	100	3

Table 2. Time requirements for preparing and testing off-line optimization programs.

		Subnetwork Signal Setting Requirements (person hours)			
Program	Average Computer Time per Run (min)	Senior Engineer	Engineer	Technician	Total
SIGOP	12	70	280	360	710
TRANSYT	165	42	250	634	926
Combination Method	6	30	142	715	887
SIGRID	4	10	93	360	463

gram run and the total number of person hours required to prepare both subnetwork signal settings. It should be noted that for TRANSYT and the Combination Method 1 program run was required for each trial cycle length; for SIGOP, however, a range of cycle lengths could be tested within a single program run. Although they are estimated, these figures serve as a useful indicator of the relative cost and effort required in using the computer programs. TRANSYT appears to be the most time-consuming, at least for the networks under study. However, the computer time required for processing TRANSYT is extremely sensitive to the size of the networks. For a small network, it compares favorably with SIGOP and the Combination Method. Although SIGOP seems to require the least coding effort, more effort by professional personnel is necessary because most of the SIGOP input is based on arbitrary decisions and engineering interpretations. On the other hand, the coding of TRANSYT and the Combination Method involves more precise measurements and is relatively routine when the procedure is set up and understood. It also should be noted that the coding effort for SIGOP was lower than normal because the staff involved was familiar with SIGRID, an older version of SIGOP. The Combination Method seems to be the easiest to use, but it does not calculate signal splits. Therefore, additional effort would be needed to prepare the split data. In the Toronto study, the splits as calculated by TRANSYT were used as input to the Combination Method, which reduced overall manpower requirements.

The network settings produced from each of the off-line programs were field tested over a 12-week period in accordance with the schedule given in Table 3. Over this period, data from a comprehensive speed and delay survey were assembled to derive estimates of network performance. Ten field crews were assigned to specific routes during each OTP in both the central and suburban areas. These routes were developed by using a number of general guidelines.

1. Routes should be designed to represent the main flow patterns throughout the test network, including turning movements.
2. High-volume links and links with heavy turning movements should be sampled at least once per OTP.
3. Link samples should be distributed uniformly throughout each OTP.
4. Travel time of all subroutes should not be longer than 20 min in the central area and 30 min in the suburban area. Therefore, 6 sample time periods (STPs) are in each OTP in the central area and 4 STPs are in each OTP in the suburban area.
5. Amount of travel outside the test network should be minimized, but this should not restrict the routes from representing the direction of the major flow.

A typical route layout for one of the 5 speed and delay crews assigned to the suburban area during a certain OTP is shown in Figure 3. Each of the 10 crews consisted of a driver and an observer equipped with 2 stop watches (1 for route time and 1 for stop time), a route map, and a field data form (Fig. 4).

The resultant speed and delay information was edited and merged with real-time traffic volume data from 2 sources (inductive loop detectors and temporary pneumatic counters), and with a complete historical volume file for each link within the 2 subnetworks. The links were classified into 2 types or strata. All links with real-time traffic volume data were categorized under stratum 1. The remaining links were placed in a stratum 2 file because they provided less reliable data. The following gives a summary of the number of links in each strata within the central and suburban subnetworks:

Subnetwork	Real-Time Volume Data	Historical or Extrapolated Volume Data
Central	73	101
Suburban	59	55

Table 3. Schedule of field tests.

Program	Test Period	Reruns
Existing plan	September 24 to 28 October 1 to 5	October 9 and 10
SIGOP	October 15 to 19 October 22 to 26	October 29 and 30
TRANSYT[a]	November 5 to 9 November 13 to 15 November 19	November 20 to 22
Combination Method	November 26 to 30 December 3 to 7	December 10

[a]On November 7, data tape was scratched by malfunctioning tape drive. On November 15, the computer was down.

Figure 3. Suburban area speed and delay route map.

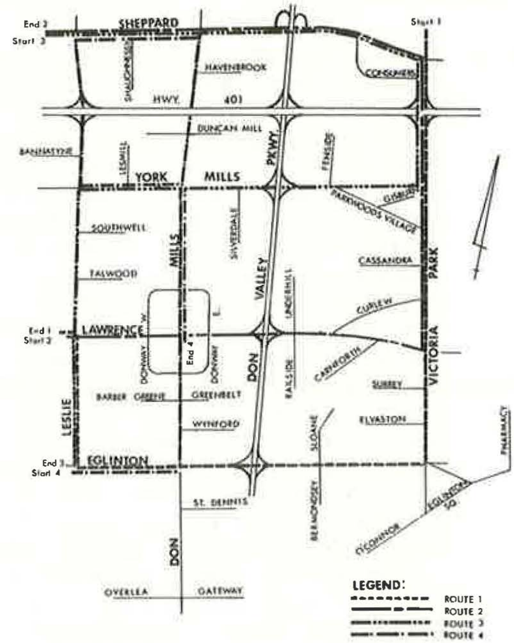
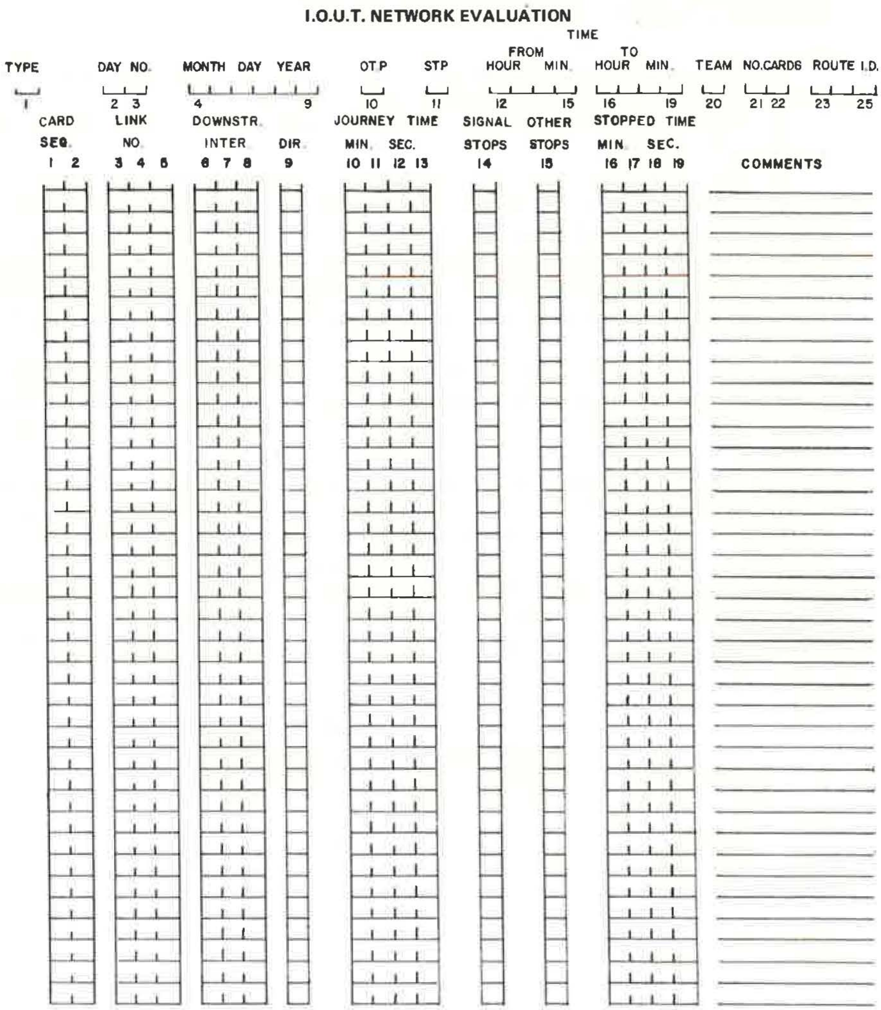


Figure 4. Field data form.



The real-time traffic volume data then were subjected to a high-low quality-control-limit test. If the real-time count fell below the low limit or above the high limit, either an extrapolated volume or a historical count was substituted. A range of ±4 standard deviations was considered to be a sufficiently wide acceptance limit to permit natural fluctuations of real-time counts without rejection.

ANALYSIS OF DATA

The overall evaluation phase of the study was divided into 2 major components. Initially, a service-rate analysis of network travel time was undertaken (14, 15). However, because of the generally inconclusive nature of the results from this analysis, a link-by-link paired comparison test was carried out afterward.

Service-Rate Analysis

Linear regression analyses were performed on the data, which were stratified by subnetwork; system travel time, in vehicle hours/hour, was the dependent variable, and system service rate in vehicle miles (kilometers)/hour was the independent variable. The data were organized into the following categories:

1. All teams for all OTPs,
2. All teams by OTP,
3. All teams by STP, and
4. Individual OTP and individual plan by team.

The analyses were conducted on network estimates of travel time and service rate based on sample averages per mile (kilometer) and per link by using data from stratum 1 only and from both strata combined. Although the stratum 1 data were considered to be more reliable, analysis results indicated that there was no apparent difference between the 2 sets of data.

Figures 5 and 6 are simplified scatter diagrams of system travel time versus system service rate and derived regression lines for the central and suburban subnetworks. In almost all cases, the regression lines do not exhibit any significant relationships. About 80 percent of the correlation ratios are lower than 0.50, and the regression slopes also are insignificant as shown in Figure 5. The system service rate computed on an OTP basis varies by only a few percent, and this is overwhelmed by the relatively large fluctuations in the system's travel-time values. As a result, data points tend to cluster around the mean. For system service rate computed on an STP basis, a greater range is observed, but the fluctuations in the system's travel-time values also are increased because of greater errors in the network travel-time estimates from a smaller number of link samples. The poor regression results indicate that no valid comparisons based on the regression equations can be made of the different signal timing plans. To complete this phase of the analysis, regression lines also are derived from aggregated plots of data for OTPs 1 to 4 (Fig. 6). Although these regression equations are significant, it is doubtful that they are valid in practice because different OTPs have different traffic behavior, operational characteristics, and signal timing designs.

Although the clustering effects of service-rate data are a weakness for regression analysis, the data lend themselves to a straight comparison of the network travel-time values used in the previously mentioned service-rate analysis. In other words, if the network traffic demand expressed in vehicle miles (kilometers)/hour during a particular OTP on a certain day is not significantly different from that on another day, then the respective network travel times expressed in vehicle hours/hour can be compared directly and tested for significant differences. To ensure that network demand is relatively constant, the service rates for different plans were paired and tested for significant differences by Student's t-test for difference of means. As indicated in the data given in Table 4, almost no significant differences in the network vehicle miles (kilo-

Figure 5. Regression analysis per mile for OTP 1, all teams and STPs: (a) central area and (b) suburban area.

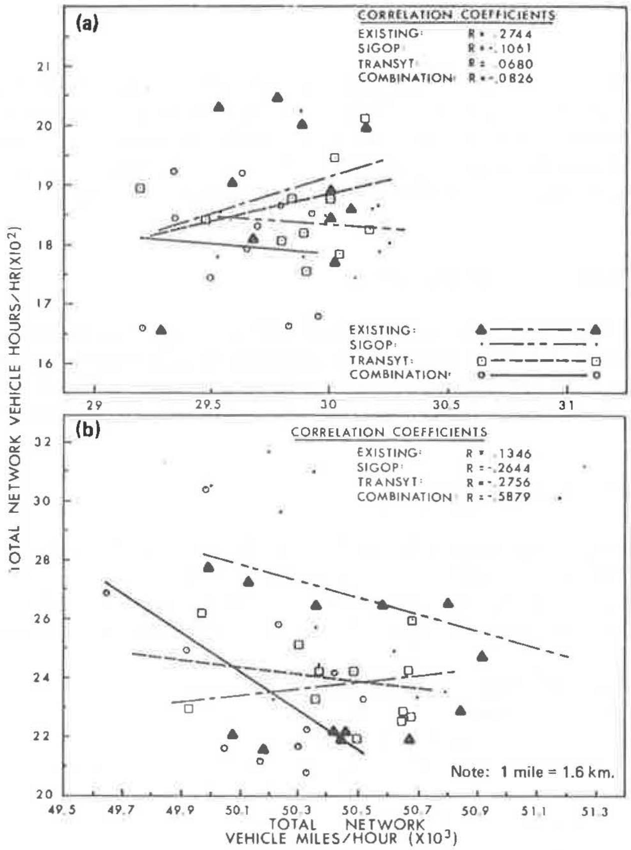
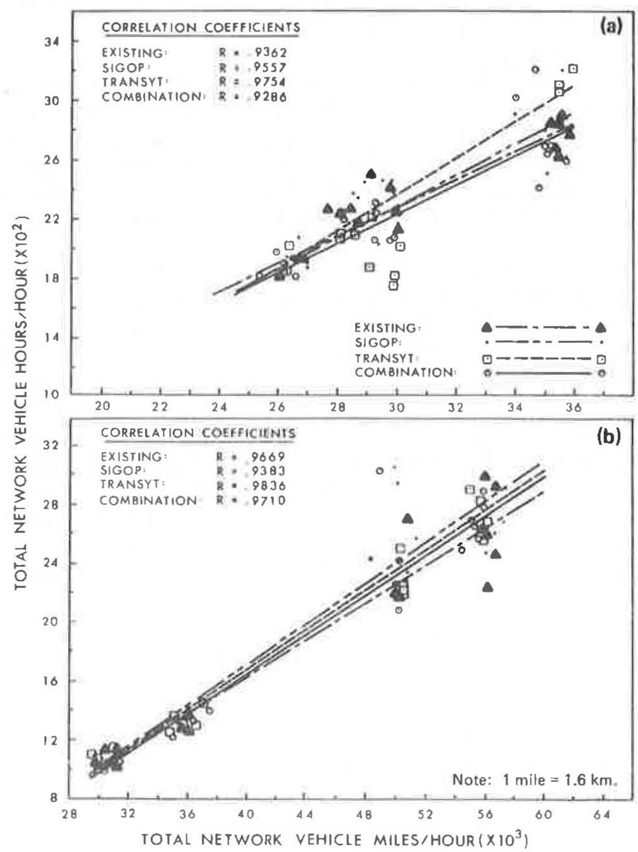


Figure 6. Regression analysis per mile for all OTPs, teams, and days: (a) central area and (b) suburban area.



meters)/hour were detected among all the signal plans for all test periods. It therefore was felt that a straight comparison of network vehicle hours/hour was justified. The results of the straight comparison test, as given in Table 5, indicate that there was no significant difference among the signal plans in terms of their effectiveness. There are however 2 exceptions—between TRANSYT and the existing plan and between the Combination Method and the existing plan during the evening rush period. During this period, TRANSYT seemed to be significantly less effective than the existing plan and the Combination Method.

Based on these findings, the service-rate analysis technique does not appear to be sufficiently sensitive to actual differences (if any) among the alternative signal plans. The effect of vehicular volume and the aggregation of field data may have had the most significant effect on the data results. Because vehicular volume in vehicles/hour is part of the dependent as well as the independent variable, the independent variable in vehicle miles (kilometers)/hour is not truly independent. In addition, errors in link travel-time measurements and network estimates may greatly influence the data such that travel-time differences between the various signal plans are rendered insignificant. Thus we felt that a more sensitive statistical routine for the comparison of signal plans must be used for this study.

Paired Comparison Analysis

For the reasons outlined above, a link-by-link paired comparison analysis based on Student's t-test for mean of differences was carried out on both link volume and link travel time. SIGOP, TRANSYT, and Combination Method data were compared against Toronto's existing fixed-time, time-of-day signal plans, thus minimizing the number of tests to be carried out. The links were compared on a 1-to-1 basis over a similar time period. To determine which links had no significant change in volume, a t-test at a 1 percent significance level was employed. The nature of this test required that a population (in this case an STP or an OTP) be totally accepted or rejected. Although the test was a link-by-link comparison, a complete OTP or STP had to be tested as a unit for volume comparison. When the time periods with significant volume differences had been eliminated, a t-test on link travel time at a 5 percent significance level was carried out on the remaining data.

Table 6 gives a summary of the results of the paired comparison tests on link volume. At a 1 percent significance level, the rejection rate was generally higher on an OTP basis than an STP basis because an STP with significantly different volumes could cause the rejection of an entire OTP data set. In all cases, the accepted data were more than adequate to continue a paired comparison analysis for link travel time.

In terms of the travel time, the Combination Method provided a slightly better level of service than the other 3 methods did (Table 7). For example, on an OTP basis in the central area, the Combination Method produced a 4.5 percent improvement in system travel time over the existing plan; TRANSYT and SIGOP increased travel time over existing times by 0.64 and 0.75 percent respectively. In the suburban area, the Combination Method on an OTP basis produced a 2.8 percent improvement over the existing plan, which in turn effected travel time reductions of 3.0 percent over TRANSYT and 3.2 percent over SIGOP.

SUMMARY AND CONCLUSIONS

In SIGOP, signal splits are calculated from critical lane flows or total approach flows without allowing for the capacity or saturation flow characteristics of the lanes or approaches. The program requires the use of many arbitrary factors, and the instructions for choosing their appropriate values are, for the most part, not presented clearly in the program manual. Most of these factors do not have any apparent theoretical basis; the success of their use depends on the interpretation and judgment of the user. The program also has a number of oversimplifying assumptions that tend to reduce its

Table 4. Comparison of system service rate for no significant difference.

		Significant Difference											
		SIGOP				TRANSYT				Combination Method			
Subnetwork	Program	OTP1	OTP2	OTP3	OTP4	OTP1	OTP2	OTP3	OTP4	OTP1	OTP2	OTP3	OTP4
Central	Existing plan	No	No	No	No	No	No	No	No	No	No	No	No
	SIGOP	—	—	—	—	No	No	No	No	Yes[a]	No	No	No
	TRANSYT	—	—	—	—	—	—	—	—	No	No	No	No
Suburban	Existing plan	No	No	No	No	No	No	No	No	Yes[a]	No	No	Yes[a]
	SIGOP	—	—	—	—	No	No	No	No	No	No	No	Yes[a]
	TRANSYT	—	—	—	—	—	—	—	—	No	No	No	No

[a]Significant at the 1 percent level.

Table 5. Comparison of system travel time for a 5 percent significant difference.

		Significant Difference											
		SIGOP				TRANSYT				Combination Method			
Subnetwork	Program	OTP1	OTP2	OTP3	OTP4	OTP1	OTP2	OTP3	OTP4	OTP1	OTP2	OTP3	OTP4
Central	Existing plan	No	No	No	No	No	No	No	Yes[a]	No	No	No	No
	SIGOP	—	—	—	—	No	No	No	No	NT[b]	No	No	No
	TRANSYT	—	—	—	—	—	—	—	—	No	No	No	Yes[a]
Suburban	Existing plan	No	No	No	No	No	No	No	No	NT[b]	No	No	NT[b]
	SIGOP	—	—	—	—	No	No	No	No	No	No	No	NT[b]
	TRANSYT	—	—	—	—	—	—	—	—	No	No	No	No

[a]Significant at the 5 percent level. [b]No test undertaken because of significant difference in system service rate.

Table 6. Summary of paired comparison results for link volume.

			Samples With Significant Difference (percent)		
Subnetwork	Time Period	Sample Size	Existing Plan Versus SIGOP	Existing Plan Versus TRANSYT	Existing Plan Versus Combination Method
Central area	OTP	40	25.0	27.5	42.5
	STP	240	11.7	19.6	20.0
Suburban area	OTP	40	12.5	27.5	35.0
	STP	160	15.0	26.2	22.5

Table 7. Central and suburban area paired comparison results for link travel time.

			Samples Tested		Significant Difference (percent)		Travel Time Difference			Time Improvement	
Time Period	Comparison	Total Samples	Number	Percent	Yes	No	Hour	Minute	Second	Percent	Plan
Central Area											
OTP	Existing plan versus SIGOP	40	29	72.0	17.2	82.8	0	23	4	0.75	Existing
	Existing plan versus TRANSYT	40	28	70.0	14.3	85.7	0	19	7	0.64	Existing
	Existing plan versus Combination Method	40	23	57.5	34.8	65.2	1	49	53	4.5	Combination
STP	Existing plan versus SIGOP	240	205	85.4	8.8	91.2	0	38	50	0.47	Existing
	Existing plan versus TRANSYT	240	192	80.0	10.4	89.6	0	56	1	0.71	Existing
	Existing plan versus Combination Method	240	191	79.6	13.6	86.4	3	50	33	2.9	Combination
Suburban Area											
OTP	Existing plan versus SIGOP	40	34	85.0	29.4	70.6	1	48	15	3.2	Existing
	Existing plan versus TRANSYT	40	28	70.0	28.6	71.4	1	25	41	3.0	Existing
	Existing plan versus Combination Method	40	24	60.0	29.2	70.8	1	9	42	2.8	Combination
STP	Existing plan versus SIGOP	160	135	84.4	20.7	79.3	4	48	42	3.4	Existing
	Existing plan versus TRANSYT	160	119	74.4	26.1	73.9	3	1	42	2.5	Existing
	Existing plan versus Combination Method	160	122	76.3	18.0	82.0	2	29	4	2.0	Combination

effectiveness as a signal optimization model. For example, vehicular arrivals at downstream intersections are assumed to follow a square wave pattern; platoon dispersion effects are not considered.

Despite these weaknesses, SIGOP performed surprisingly well according to the Toronto results mainly because the study personnel were experienced in using the test signal system. They also were familiar with SIGRID, a predecessor of SIGOP. So it was relatively easy for them to interpret the program manual and choose appropriate values for the various arbitrary factors required as input to the program. Also any unreasonable output from the program was detected easily and corrected by making the necessary adjustment in the program input. Because of the arbitrary nature of SIGOP, familiarity with both the program and the system is such an important factor that it could well explain why SIGOP has been used with varying degrees of success.

TRANSYT has been regarded as a logical and theoretically sound program. Its success has been demonstrated in a number of research studies (1). The strength of this program lies in its traffic-flow model that accurately traces flow patterns from signal to signal and allows for the effects of platoon dispersion by means of a platoon prediction model.

However, the superiority of TRANSYT was not evident from the results of the Toronto study. This perhaps occurred because the program was used without prior calibration of some of the program parameters for local conditions (such as the smoothing factor used in the platoon dispersion model). This would have reduced the effectiveness of the program. For example, in a subsequent and separate study in Toronto (16), Robertson's platoon dispersion model was found to be satisfactory, but the parameters had to be calibrated to suit local conditions to obtain the best fit between actual and predicted platoon structures. Because TRANSYT was not found to be decisively inferior to the other programs, one expects that it would perform much better if the program were calibrated.

The Combination Method contains a flow model similar to that existing in TRANSYT, but its optimization process is radically different. It is assumed that no flow continuity exists between links; each link is treated as a separate entity with its own distinct relationship of delay to difference of offset. This influences the program's effectiveness in dealing with undersaturated signals where vehicular queues seldom exist. The Combination Method is not a comprehensive signal optimization package because it does not calculate signal splits. However, this permits the user to intervene freely and introduce his or her own judgment in screening the split data input prior to program execution. This may be the primary reason for the slightly better performance of the Combination Method in some cases.

The Toronto study results also indicated that the existing signal settings compared favorably with those obtained from the sophisticated computer program packages. This was expected because the existing timings are the results of years of experience with the signal system and continuous engineering efforts.

All of the signal timing plans provided similar levels of service in the test subnetworks based on the analysis of travel time. It should be noted, however, that the various signal optimization programs are based on the criteria of delay and stops, which may constitute only a minor proportion of total system travel time, particularly for a large network with relatively long links and smooth flow characteristics.

Whichever optimization program is chosen to design urban network signal settings, the user must have a thorough understanding of the selected program and a comprehensive knowledge of the signal system. In addition, a commitment must be made to carefully review the program output to ascertain its validity. Although these off-line signal optimization programs can be used as engineering aids in network signal-setting design, they should not be used as replacements for engineering judgment and expertise.

ACKNOWLEDGMENTS

We would like to acknowledge the contribution of M. M. Dillon, Limited, Consulting Engineers, who supervised the collection of speed and delay data, and F. Wagner of Alan M. Voorhees and Associates for his assistance in the planning of the speed and

delay survey and for supplying a computer listing of the service-rate-analysis program package.

REFERENCES

1. L. Rach et al. Traffic Signal Control Strategies—A State-of-the-Art. In Improved Operation of Urban Transportation Systems, Metropolitan Toronto Roads and Traffic Department, Vol. 1, March 1974.
2. Traffic Research Corporation. SIGOP: Traffic Signal Optimization Program. Bureau of Public Roads, U.S. Department of Commerce, PB 173 738, 1966.
3. Peat, Marwick, Livingston and Co. SIGOP: Traffic Signal Optimization Program, User's Manual. Bureau of Public Roads, U.S. Department of Commerce, PB 182 835, 1968.
4. Peat, Marwick, Livingston and Co. SIGOP: Traffic Signal Optimization Program, Field Tests and Sensitivity Studies. Bureau of Public Roads, U.S. Department of Commerce, PB 182 836, 1968.
5. D. I. Robertson. TRANSYT Method for Area Traffic Control. Traffic Engineering and Control, Vol. 11, No. 6, Oct. 1969.
6. D. I. Robertson. TRANSYT: A Traffic Network Study Tool. Road Research Laboratory, Crowthorne, Berkshire, England, RRL Rept. 253, 1969.
7. D. I. Robertson. TRANSYT: Traffic Network Study Tool. Fourth International Symposium on the Theory of Traffic Flow, Karlsruhe, West Germany, 1968.
8. D. I. Robertson. User Guide to TRANSYT Version 5. Road Research Laboratory, Crowthorne, Berkshire, England, TN 813, March 1973.
9. J. A. Hillier. Appendix to Glasgow's Experiment in Area Traffic Control. Traffic Engineering and Control, Jan. 1966.
10. Hillier and Lott. A Method of Linking Traffic Signals to Minimize Delays. International Study Week in Traffic Engineering, Barcelona, Spain, 1966.
11. Huddart and Turner. Traffic Signal Progressions—G.L.C. Combination Method. Traffic Engineering and Control, Vol. 11, No. 7, Nov. 1969.
12. E. D. Turner. A Simple Guide to the G.L.C. Computer Program for Linking Traffic Signals. Department of Planning and Transportation, Greater London Council, Research Memo 66, April 1968.
13. Traffic Research Corporation. SIGRID Program: Notes and User's Manual. Toronto Traffic Control Centre, 1965-1973.
14. TRW, Inc. System Analysis Methodology in Urban Traffic Control Systems. Bureau of Public Roads, U.S. Department of Commerce, PB 185 422, June 1969.
15. F. A. Wagner. SIGOP/TRANSYT Evaluation: San Jose, California. Federal Highway Administration, Rept. FH-11-7822, July 1972.
16. J. K. Lam. Calibration of Robertson's Platoon Dispersion Model. Metropolitan Toronto Roads and Traffic Department, Sept. 1974.

DESIGN AND IMPLEMENTATION OF A SYSTEM TO RECORD DRIVER LATERAL POSITIONING

James A. Gardner,* Honeywell, Inc.;
Stanley M. Soliday,* Midwest Research Institute; and
Glen A. Williamson, North Carolina State University

An important indicator of driving performance is steering or tracking accuracy. Previous attempts to measure this variable have suffered from a variety of limitations. Previous attempts are reviewed, and numerous criteria are specified for an optimal tracking-system design. These criteria are used to develop a novel photooptic system that mounts on a test vehicle and continuously records lateral lane positions. The device uses a photodiode array to detect the position of the vehicle relative to the pavement shoulder line. System design, data reduction procedures, and practical implementation considerations are discussed. System performance is demonstrated by recording driver tracking performance over shallow right and left curves. The system is sensitive to the differences in driving behavior on the 2 curves. The device is presented as a flexible, reliable, and accurate instrument for recording vehicular track.

•ONE of the most important criteria of driving performance is steering or tracking accuracy. A driver must steer his or her vehicle along a traffic lane and keep deviations to a minimum regardless of vehicle speed, traffic flow, or roadway conditions. Although the importance of this performance has been recognized for decades, practical methods for its measurement have eluded researchers. It is therefore the purpose of this study to examine alternative solutions and develop a practical, low-cost measurement system.

The purposes were threefold: (a) to identify design criteria for a system to measure vehicle lateral positioning based on a review of previous attempts; (b) to develop an electronic system to continuously record a driver's tracking behavior by using these criteria as a set of design requirements; and (c) to evaluate the resulting system in the field to determine whether it performs according to design specifications.

Early research in lateral displacement primarily attempted to determine the minimum road width necessary for safe driving. Steinbaugh (13) subjectively estimated the lateral position of vehicles on a straight, nearly level section of highway and found that the average distance from the right front tire to the edge of the pavement was 4.8 ft (1.46 m) at speeds less than 40 miles/h (64.37 km/h). When speed increased beyond 60 miles/h (96.56 km/h), the average lateral distance from the roadside was 8.6 ft (2.62 m). On the basis of these findings, he recommended a minimum roadway width of 23.5 ft (7.16 m), which is only 6 in. (15.24 cm) less than present specifications. The present roadway design width of 24 ft (7.32 m) evolved from the results of numerous similar studies.

Researchers soon recognized the subjectivity of early observational methods and turned to photographic techniques to obtain more objective data on driver performance. Thompson and Hebden (14) used a movie camera to study passing behavior. Probably the most ambitious and perhaps dangerous attempt was described by Forbes (3), who

*Messrs. Gardner and Soliday were with North Carolina State University when this research was performed.

took movies of drivers while he flew directly above them in a blimp. Although procedures such as these were an improvement over simple estimation, they did not permit determination of measurement accuracy.

Subsequently, Green (5) developed a photographic technique to measure the lateral position of a moving vehicle by taking several photographs in rapid succession. A gun-sight camera was modified to mount on a bridge truss and was operated by a remote push-button switch. Target boards containing sawtooth calibration scales were photographed at each of the observation points on the highway. These boards were then removed and traffic was photographed at each of the same points. By using 2 projectors, one could superimpose the calibration scale on the traffic image, thus allowing accurate measurement of lateral position. Measurement accuracies ranged from 6 in. (15.24 cm) at the farthest location [180 ft (54.86 m) from the camera] to 3 in. (7.62 cm) at the nearest point [40 ft (12.19 m) from the camera].

In later research, Case, Hulbert, Mount, and Brenner (2) were able to increase the accuracy of Green's technique (5) to within 3 in. (7.62 cm) at distances of up to 400 ft (121.92 m) from the camera by using a more sophisticated calibration technique. Here the calibration scales were permanently inscribed on a lucite screen coated on 1 side with magnesium oxide. Edge lighting of the screen was used to permit brightness adjustment of the calibration scale for maximum contrast with the highway vehicle.

Another technique that attempted to improve on subjective observation was reported by Holmes and Reymer (7). They modified a positive-contact traffic detector to permit recording of the vehicle path at specified points on the roadway. The original device consisted of two 20-ft (6.10-m) metal strips encased in rubber so that one was superimposed over the other. The strips made electrical contact when the top strip was depressed by the passage of a vehicle. Lateral position was measured by dividing 1 of the metal strips into individual segments and running a separate wire to each. By these means, vehicle position was indicated by the particular segments making electrical contact. A major problem with this approach, however, was that the system was accurate to within only 1 ft (0.30 m). Furthermore, 2 or more adjacent sections could be activated by vehicles with wide or dual tires.

A popular technique still in use involves lining the measurement course with traffic cones and recording the number of cones touched by the vehicle (6, 12). This procedure obviously has little relevance to actual dynamic traffic situations.

Although these techniques could indicate the driver's lateral position at specific points along the road, none were capable of describing steering or tracking behavior. Considerably more measurements taken over a length of roadway were needed to achieve valid measures.

One approach to obtaining such measures would be to take actual measurements of the driving track. This technique would yield greater experimental flexibility by permitting evaluation of the effects of numerous roadway characteristics on driving behavior. For example, the effects of curves and hills on lateral positioning must be assessed by recording many measurements over a relatively long stretch of roadway.

Perhaps the earliest technique for recording the actual track was described by Brown (1). Water from a tank dripped on the rear wheels, and a movie camera photographed the resulting trial. Brown (1) reported 2 primary difficulties with the technique that led to its rejection. First, the test drivers thought that the gas tank was leaking, and second, the resulting tracking scores could not be readily quantified.

In an attempt to increase flexibility and accuracy, Michaels and Cozan (9) developed an electronic system to study the perceptual and field factors that cause lateral displacements. A row of independent photodetector units was mounted facing downward on the rear bumper of an automobile. Each detector contained a photoresistor, the output of which varied with the level of light reflected from the roadway. When a properly biased transistor was connected to each photoresistor, no output was given unless the amount of reflected light reached a specified minimum level. Thus when a single detector passed over a highly reflective line painted in the center of the driving line, that detector produced a signal. With this system, lateral displacement could be estimated to the nearest inch (2.54 cm) over a range of 6 ft (1.83 m). The Michaels and Cozan device provided several advantages over previous systems. It directly and continuously measured the vehicle's path as it traveled down the test track, and it gave a very accurate description of the vehicle's track and required no subsequent manual calculations or time for film development. A major drawback, however, was that the system required a center-lane reflective paint strip that was clearly visible to the driver. No way was found to camouflage this line and still maintain sufficient contrast ratios to ensure reliable data from the placement detectors. Therefore, the driver's perceptual task was not the same as that for driving on normal highways. In addition, this tracking system could be used only on specially painted road surfaces and thus could not be used to make measurements on normal highways.

A more versatile tracking system was described by Matanzo and Rockwell (8). This system measured vehicle position relative to the white shoulder line. A single photoresistor, mounted in an arm positioned on the right side of the instrumented vehicle, scanned back and forth beside the vehicle and produced a large resistance change whenever the photodiode sensed the highly reflective shoulder line. Lateral position could be assessed by determining that point in the sweep where this change occurred. Because this system used the standard white shoulder line to determine vehicle position, it did not require a specially painted road surface. It also was capable of recording the vehicular track regardless of traffic conditions or speed. Measurement accuracy was relatively high, and no manual calculations were required because the position was determined electronically.

A serious problem existed for the Matanzo and Rockwell system because it could not compensate for varying illumination levels. It therefore functioned best at night when a pair of floodlights was used to illuminate the scanned region. Because most driving and research occur during daylight, this is a major drawback. A mechanically driven scan arm also is quite sensitive to buffeting at high speeds, thus reducing measurement accuracy. System adjustments (such as increased scan rate) require additional mechanics. Recording more position measurements in one experimental condition than in another might be desirable.

A similar tracking system described by Michon and Koutstaal (10) also employed a single detector at the side of the car to measure the distance to the shoulder line. Scanning was achieved by a rotating mirror that provided 28 data points/s. This high sampling rate minimized the effects of occasional detection errors caused, for example, by wet spots. The system worked both day and night when it was used with an auxiliary light source.

This approach presented several advances over previous measurement techniques.

1. Relatively high resolution was possible.
2. It could be used over a wide range of roadway configurations.
3. It performed well over a range of environmental conditions.

However, this device also employed mechanical scanning methods that continued to limit flexibility and increased mechanical design complexity.

A major weakness of most tracking system designs has been a lack of reliability. Rockwell has stated (11, p. 159): "No method for measuring lateral placement along the highways has been reliably demonstrated." A number of alternative techniques have been attempted, but no simple, reliable, and accurate system has resulted. Nevertheless, the designs of these previous systems have suggested a number of areas in which design improvements can be made in an attempt to attain a truly flexible system. This type of system would (a) work both day and night over a range of illumination levels, (b) contain no moving parts, and (c) permit simple, inexpensive data recording.

As noted earlier, variable scan rate is a highly desirable design element that minimizes the effects of occasional spurious data. The experimenter should have the option of making this measurement or sampling rate either independent of or dependent on vehicle speed. If it were independent, lateral position data would be recorded at a constant rate regardless of vehicle speed. If it were dependent, rate of data collection would depend on vehicle speed. This latter approach would permit, for example, the same number of position measurements to be taken over the same segment of highway at both 30 mph (48.28 km/h) and 60 mph (96.56 km/h). Hence driving behavior at the 2 speeds could be compared point by point. Adjustable data sampling rates also would be highly desirable if measurement or sampling rate were a dependent variable. An improved tracking system also should permit rapid transfer from 1 test vehicle to another. Because driving behavior is thought to be strongly biased by the particular vehicle model used, this feature would allow the experimenter to assess the extent of such influences.

Common sense indicates certain other desirable design criteria. The tracking system data collection technique should permit automatic recording and analysis of various external events, treatments, and experimenter observations. The system should be relatively inexpensive. The design should be as simple as possible and should require few expensive components. The system should

1. Operate over existing roads,
2. Not affect driving behavior,
3. Be accurate,
4. Operate over a range of illumination levels,
5. Have a high and variable measurement rate,
6. Have no moving parts,
7. Provide for simple data recording and reduction,
8. Be inexpensive,
9. Provide for simultaneous event recording, and
10. Be transferrable among vehicles.

TRACKING SYSTEM DESIGN

The heart of this system is a photodiode array mounted in a box on the side of the car that views the white shoulder line. A lens mounted in this box receives an image from the roadway and projects the scene on the array. Each photodiode produces an output voltage proportionate to the amount of light reflected from the area of road or shoulder line it views. The higher reflectance of the shoulder line distinguishes it from the darker roadway.

Circuit Description

The central component of the system is an integrated circuit containing a linear array of 50 independent silicon planar photodiodes and an associated 51-bit shift register. Also contained are presensors and postsensors located before the first and after the last diodes.

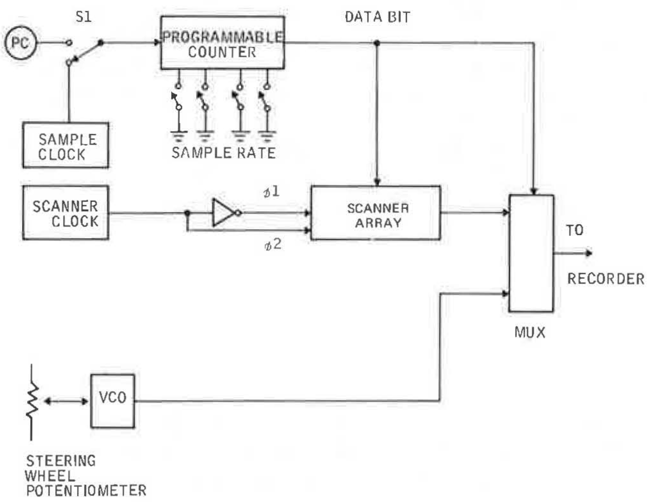
The array views the road and sequentially scans all 50 photodiodes. The sequence of output pulses varies in voltage with the amount of light that is reflected from the highway and falls on the diode being scanned. Because it is lighter than the surrounding pavement, the white line distinguishes itself by a higher reflectance and resultant higher pulse voltage.

The shift register permits a sequential output from the 50 photodiodes. As shown in Figure 1, an external clock controls the multiplexing rate.

The 50-element photodiode array also requires a data pulse to begin the sampling sequence. The frequency of this pulse determines the number of vehicle position measurements taken per second because all 50 photodiodes must be sampled to determine the position. This data-pulse frequency may be determined by either an adjustable oscillator or a wheel rotation detector on the rear tire. Thus measurement rate can be either dependent on or independent of vehicle speed.

The wheel-rotation-detector approach consists of a single photodiode mounted on the vehicle's rear fender. This detector senses a white line or piece of tape placed on the tire sidewall. As the wheel rotates, the photodiode increases its output voltage each time the white strip passes. The frequency of this pulse therefore is controlled

Figure 1. Tracking system.



directly by vehicle speed. The switch (S1 in Figure 1) selects either the oscillator or the rotation detector. When normalized, wheel-pulse frequency very accurately indicates vehicle speed.

To increase system versatility, a variable modulo counter permits division of the data-pulse frequency by any number from 1 through 16. Hence the frequency of lateral position measurement may be changed easily. In this case, it is done by switches in the car.

Because clock frequency exceeds data-pulse frequency by 50, a dead period occurs between each data series transmittal. This interim provides a convenient time to transmit information about the steering wheel position. Measurements of steering wheel position are made by a potentiometer connected to the steering column. Adjustments in resistance that are due to steering motions may then be transformed into frequency changes by a voltage-controlled oscillator. As the resistance of the oscillator input changes, the frequency of its output changes. It is this information that is introduced during the dead period of the data series.

The proposed tracking system can record the vehicle's lateral position, speed, and driver steering motions on a single-channel recorder. No moving parts are employed, and the experimenter may change the rate of sampling with the flip of a switch. All components, except those in the array, are relatively inexpensive and readily available. The system output is compatible with any high-quality AM recorder or wide-bandwidth FM recorder.

Optical System

An optical system that uses a relatively inexpensive 16-mm C-mount movie camera lens projects an image of the road surface on the linear diode array. A wide-angle (10-mm focal length) lens yields a measurement range of 41.88 in. (1.06 m) in the current configuration.

The photodiode array, housed in a dust-proof, electronically shielded container called the optical assembly box, mounts on the right side of the test vehicle just above the passenger door. A metal plate covers the side facing downward toward the shoulder white line. A C-mount lens thread centered in this plate just over the photodiode array inside the box permits rapid lens changes if necessary.

The sensitivity of the photodiode array is so great that a Wratten II neutral density filter is needed to prevent diode saturation at high light levels. With the filter removed, sensitivity is sufficient to allow normal operation, even at twilight.

Mechanical Design

A luggage rack anchors the photodiode array and optical system to the vehicle. This rack consists of 2 horizontal steel bars that connect to the automobile roof gutters. A 2 by 4-in. (5.1 by 10.2-cm) wooden beam mounts atop each of the 2 racks, and these structures are then joined by 2 wooden braces to form a rigid rectangular unit. The front wooden beam extends beyond the sides of the steel rack assembly to allow a mounting area for the photodiode array circuit box. The luggage rack size is adjustable from beneath, and thus, the tracker mounting system may be adjusted for most vehicle models. The system mounted on the test vehicle is shown in Figures 2 and 3.

Electronic connections with the photodiode array are made by multiple-pin connectors mounted on the rear of the optical assembly box. Interconnecting cables enter the car through the top of the right front passenger window and travel to the rear seat where all of the peripheral circuitry is located.

IMPLEMENTATION

Calibration

Each of the 50 photodiodes is independently calibrated by drawing a line between the right sidewalls of the front and rear right tires and by measuring the perpendicular distance from this line to the pinpoint of light activating each photodiode.

Recording Techniques

The output signal from the tracking system may be recorded on either an AM or FM tape recorder. In this study, an AM system was chosen because it is much less expensive than an FM system. The composite output signal is of a relatively high frequency and requires a bandwidth capability of at least 10 000 kHz. Only the most expensive FM recorders can approach this bandwidth.

Traditional AM taping techniques are not usable because the 25-Hz measurement frequency is seen as a 25-Hz pulse, which is slightly below the low-frequency capability of the recorder. Thus the recorder introduces a certain degree of signal distortion that affects both the output signal fidelity and the accuracy of the recording level indicators. The data analysis techniques, which will be described later, are capable of coping with distortions in signal fidelity, but they cannot compensate for incorrect recording levels. This problem was resolved by presetting signal levels with an oscilloscope.

DATA REDUCTION

A computer converts each data tape from analogue to digital form for further processing. For each of the 25 position measurements taken per second, the analysis program determines which of the 50 photodiode output voltages represents the inside edge of the white line. This signal is then translated into the distance in inches (centimeters) from the outside edge of the right front wheel to the inside edge of the shoulder line according to the 50 previously described calibration distances.

Because the tracking system is adjusted to make 25 position recordings each second [representing about 1 data point every 3.81 ft (1.16 m)] at 65 miles/h (104.6 km/h), excessive amounts of data are amassed at this measurement rate. To overcome this problem, medians are calculated for each successive group of 8 measurements. A cathode-ray-tube display plots sequential groups of 40 median data scores, and suspicious data are deleted by typing their assigned number on the computer keyboard. All subsequent statistical analyses are then performed on these edited median scores.

SYSTEM PERFORMANCE

The system is relatively easy to operate. After it is turned on and calibrated, it requires no further attention until data collection is completed. Measurement accuracy

Figure 2. Front view of tracking system mounted on test vehicle.

Figure 3. Side view of tracking system mounted on test vehicle.

is greater than 0.5 in. (1.27 cm), and all data reduction is done by using appropriate computer techniques. The system is quite sensitive and requires only small differences in contrast between the white line and the surrounding pavement. In fact, reliable lateral position measurements are obtainable from the shoulder line on concrete roads. The equipment is highly reliable; it has performed flawlessly through many hours of data collection.

The test results of 10 subjects will illustrate system performance and versatility. Each subject drove a Ford Custom 500 Sedan equipped with power steering over a 0.28-mile (0.45-km) curve on a rural 24-ft-wide (7.32-m-wide) 2-lane asphalt road. The curve was a level 3-min transitional spiral preceded and followed by straight pavement. Subjects drove both directions around the curve at a speed of 55 mph (88.51 km/h) but did not know that their track was being recorded.

Of the 1,112 median lateral position measurements recorded for both directions across all subjects, 984 yielded useful data for subsequent analyses. Eleven and a half percent of the data was lost because of tree shadows and a badly worn white shoulder line.

A number of statistics were calculated on these basic data. Mean and median measures of central tendency were analyzed. The mean lateral position for subjects driving the right curve was 31.66 in. (80.42 cm) measured from the edge of the right tire sidewall to the inside edge of the roadway shoulder line. For subjects driving in the opposite direction, the mean position was 35.31 in. (89.69 cm). Thus, the tracking system detected a difference of 3.65 in. (9.27 cm) in mean lane position on left and right curves. The median position was 30.99 in. (78.71 cm) for the right curve and was 35.16 in. (89.31 cm) for the left curve. The 2 different measures of central tendency showed the same results.

The standard deviation was calculated to describe oscillations about the mean lane position. The standard deviation averaged across all 10 subjects was 5.28 in. (13.41 cm) for the right curve and 7.05 in. (17.91 cm) for the left curve. Drivers showed greater variability in tracking behavior on the left curve.

Skew, calculated across subjects on both curves to describe the shapes of the distributions, showed little difference: 0.41 for the right curve and 0.53 for the left curve. Deviations were positively biased on both left and right curves.

Maximum and minimum lateral deviations were calculated for these 2 curves. Averaged across subjects, the maximum position was 41.66 in. (105.8 cm) from the shoulder line on the right curve and 47.68 in. (121.1 cm) for the left curve. These scores indicate that the subjects deviated 6.02 in. (15.29 cm) farther to the left on left curves than they did on right curves. Minimum lane position was 23.19 in. (58.90 cm) for the right curve and 23.10 in. (58.67 cm) for the left curve. The measurement system did not detect any meaningful differences in any subject's minimum lane position on the 2 different curves.

The value of the first position measurement recorded when the curve was entered analyzed to determine differences in initial position between right and left curves. The mean initial position was 35.63 in. (90.50 cm) for right curves and 26.07 in. (66.22 cm) for left curves. Thus drivers tracked 9.56 in. (24.28 cm) farther to the left when they first entered a right curve than they did when they first entered a left curve. The direction of this difference is opposite to that found when the average track through the complete curve was recorded. An instrument capable of recording only the driver's position on entering a curve would lead to entirely different conclusions about driving behavior than would have been made from an analysis of the complete track. The present system configuration permits measurements to be taken both continuously and at separate points along the roadway. This flexibility is necessary to properly interpret driver tracking behavior.

From this experiment, 4 basic statistics emerge as being sensitive to differences in tracking behavior. Either the mean or median may be used to describe the driver's central tendency within a section of roadway. In addition, the standard deviation seems to be sensitive to differences in lateral variability within right and left curves. Maximum lateral position is affected by the direction of curvature. And a measure of the initial position of the vehicle on entering a curve seems to be quite sensitive to the direction of curvature. Although skew and minimum position measures were not sensitive to differences in road curvature, they could become important descriptors of behavior if other roadway characteristics are measured. For example, minimum lane position could differ for various lane widths or types of vehicles.

CONCLUSIONS

This experiment demonstrates the feasibility of using a fixed photodiode array to scan the roadway for white line detection. This highly accurate instrument performs well over existing roads and does not appear to influence driving behavior. The system is very flexible and fulfills all of the original design criteria.

The findings of this study illustrate the sensitivity of the system to basic differences in driver tracking behavior on shallow right and left curves. These differences were found by using a small number of subjects driving a full-sized, slow-response automobile with limited proprioceptive feedback. Even though the study took place on existing roads under various environmental conditions, only a small amount of data were lost. A more complete study (4) that used this measurement system showed similar driver behavioral discrimination in evaluating driver tracking performance over a variety of Interstate and secondary roadway configurations. This system, then, has been shown to be a sensitive new tool for the study of driving performance over a range of variables.

This instrument could be employed in a number of alternate ways to study driving behavior. For example, actual vehicle track data could be compared with steering wheel positioning data to validate and extend earlier research based on steering wheel movements. A frequency analysis of variously sized lateral deviations could be used to compare male and female drivers. Also, a study of tracking performance could provide useful information about the effects of age, alcohol, or drugs on driving behavior. A careful evaluation of tracking skills over various roadway configurations could be conducted to determine the potential for screening novice and accident-prone drivers. Such a study of driver skill elements could prove invaluable to driver training and licensing programs.

A new versatility in the study of driving behavior is therefore possible by recording vehicular track. The present instrument provides a flexible, reliable, and accurate means of recording this track.

ACKNOWLEDGMENTS

Support for this project came from the Department of Industrial Engineering and a grant from the School of Education at North Carolina State University. Appreciation is expressed to Raymond Stroh for his assistance in reducing the data.

REFERENCES

1. I. D. Brown. Studies of Component Movements, Consistency and Spare Capacity of Car Drivers. Annals of Occupational Hygiene, Vol. 5, 1962, pp. 131-143.
2. H. W. Case, S. F. Hulbert, G. E. Mount, and R. Brenner. Effect of Roadside Structure on Lateral Placement of Motor Vehicles. HRB Proc., Vol. 32, 1953, pp. 364-370.
3. T. W. Forbes. Methods of Measuring Judgment and Perception Time in Passing on the Highway. HRB Proc., Vol. 19, 1939, pp. 218-231.
4. J. A. Gardner. Design, Implementation, and Use of a System to Record Driver Lateral Positioning on Interstate and Secondary Highways. North Carolina State University, PhD dissertation, 1974.
5. F. H. Green. Method for Recording Lateral Position of Vehicles. HRB Proc., Vol. 26, 1946, pp. 397-404.

6. E. R. Hoffman and P. N. Joubert. The Effect of Changes in Some Vehicle Handling Variables on Driver Steering Performance. Human Factors, Vol. 8, 1966, pp. 245-263.
7. E. H. Holmes and S. E. Reymer. New Techniques in Traffic Behavior Studies. Public Roads, Vol. 21, 1940, pp. 29-45.
8. F. Matanzo and T. H. Rockwell. Driving Performance Under Nighttime Conditions of Visual Degradation. Human Factors, Vol. 9, 1967, pp. 427-432.
9. R. M. Michaels and L. W. Cozan. Perceptual and Field Factors Causing Lateral Displacement. Public Roads, Vol. 32, 1963, pp. 233-240.
10. J. A. Michon and G. A. Koutstaal. An Instrumented Car for the Study of Driver Behavior. American Psychologist, Vol. 24, 1969, pp. 297-300.
11. T. H. Rockwell. Skills, Judgment and Information Acquisition in Driving. In Human Factors in Highway Traffic Safety Research (T. W. Forbes, ed.), John Wiley and Sons, Inc., New York, 1972.
12. K. U. Smith, R. Kaplan, and H. Kao. Experimental Systems Analysis of Simulated Vehicle Steering and Safety Training. Journal of Applied Psychology, Vol. 54, 1970, pp. 364-376.
13. V. B. Steinbaugh. Separating Traffic on Multiple-Lane Highways. In Highway Planning, 24th Annual Michigan Highway Conference, Michigan Department of State Highways, 1928, pp. 53-63.
14. J. T. Thompson and N. Hebden. A Study of the Passing of Vehicles on Highways. Public Roads, Vol. 18, 1937, pp. 121-137.

EFFECT OF 55-MPH SPEED LIMIT ON AVERAGE SPEEDS OF FREE-FLOWING AUTOMOBILES ON AN INTERSTATE BRIDGE IN WEST VIRGINIA

Bernard F. Byrne, Department of Civil Engineering, West Virginia University; and Robert R. Roberts, Traffic and Transportation Center, University of South Carolina

Data on average speeds of free-flowing automobiles were collected from a study of vehicle speed and placement on an Interstate bridge in West Virginia. Because the study included data collected in the summers of 1973 and 1974, it was possible to compare mean speeds for each summer to determine whether the nationwide speed limit of 55 mph (89 km/h) had any effect. It was found that mean speed declined from 61.0 mph (98.2 km/h) in 1973 to 54.5 mph (87.7 km/h) in 1974. The standard deviation of the distribution also declined from 9.2 mph (14.8 km/h) to 6.0 mph (10.0 km/h), thus providing another possible explanation for the reduction in automobile accidents.

•IN RESPONSE to the shortage of gasoline, the U.S. Congress enacted legislation requiring all states to impose a general speed limit of 55 mph (89 km/h) on Interstate highways as a fuel conservation measure. How effective this speed limit has been in reducing average traffic speed has been questioned. We will attempt to answer that question as it pertains to a local situation. Hopefully, a number of local answers may help to formulate a general answer.

The situation examined in this paper concerns an Interstate bridge on I-79 east of Fairmont, West Virginia. The West Virginia Department of Highways is sponsoring a research project to examine the effect on speed and lateral placement of vehicles of narrowing a bridge by means of various combinations of curbing and guardrails. This research has been in progress for a number of years; specific studies on speed and lateral placement were performed in the summers of 1973 and 1974. However disastrous the imposition of the 55-mph (89-km/h) speed limit may have been to the comparability of the data between 1973 and 1974, an opportunity is presented to study average speeds before and after imposition of the speed limits.

DATA COLLECTION

The effect of various combinations of curbing and guardrail on vehicle speed and placement was examined by investigating the study site, data collection methods, data analysis, and consequent data limitations. Data collection procedures require sight distances of 1,500 ft (457.2 m) upstream and 750 ft (228.6 m) downstream of the bridge and tangent roadway. The roadway throughout the test section is a 4-lane highway divided by a grass median. Each lane is 12 ft (3.7 m) wide. During the summer of 1972, the West Virginia Department of Highways repaired this section with an overlay of bituminous concrete. The study site is located in the southbound lane and has a 5 percent downgrade. Its average daily traffic was 6,400 in 1972. The speed limit for this section in 1973 was 70 mph (112.7 km/h).

To meet the objectives of the study, we made measurements of vehicle speed and placement by using a tape-switch system 1,000 ft (304.8 m) and 500 ft (152.4 m) upstream of the bridge, at the upstream end, middle, and downstream end of the bridge, and 500 ft (152.4 m) downstream of the bridge. The tape switches, shown in Figure 1, were numbered from 1 to 6 beginning at the upstream end of the study section.

The recording instruments were set up in the median at the downstream end of the bridge so that traffic could be observed throughout the test section and the operator could be hidden. When a free-flowing vehicle was observed approaching the first set of tape switches, the operator activated the recording system. A free-flowing vehicle was defined as a vehicle whose performance was not affected by any nearby vehicle. The maximum distance between any 2 traps was 500 ft (152.4 m), so vehicles selected for the sample were always 500 ft (152.4 m) behind any preceding vehicle because the preceding vehicle would activate the trap. Vehicles traveling behind the study vehicle could not be recorded until the system was reset. A vehicle ahead of the study vehicle could come into the system only if the study vehicle were to approach the lead vehicle within 500 ft (152.4 m). The data thus collected excluded trucks, motorcycles, and automobiles less than 500 ft (152.4 m) apart.

A series of tape-switch vehicle detectors together with the appropriate electronics and recording equipment can be used to determine the time-position signature of a vehicle over some distance, from which data speed and lateral placement values then can be computed. The tape switch is a vinyl-clad electrical strip switch that is flat, flexible, and pressure-sensitive. It is $^3/_4$ in. (19.1 mm) wide and $^3/_8$ in. (9.5 mm) thick and can be cut to any length and fitted with a connecting electrical lead. The system in this study consisted of 4 major parts

1. Tape-switch detectors and their connectors and cables,
2. Electronics package built by FHWA in their laboratory,
3. A printing Hewlett-Packard 5050B digital recorder, and
4. Power supply (12-Vdc battery and 12-Vdc to 110-Vac power inverter).

The tape-switch detectors were connected to the electronics package by special, shielded cables. The electronics package consisted of gating circuits, a 100-μs clock, and storage registers. The gating circuits accepted inputs from the tape-switch detectors in sequential order to prevent signal overlap. Activation of the second tape-switch detector produced no response in the system unless the first tape-switch detector had been activated; the same principle applied to the third tape-switch detector. Activation of a tape-switch detector caused the input channel of that particular tape-switch detector to close. This process continued until the last detector had been activated or until the system was reset. In this manner, the tape-switch system tracked only 1 vehicle at a time through the system.

In its initial state, the clock was set at zero. When the first tape-switch detector was activated the clock was started. Each time a detector was activated current clock time was stored in a buffer register, from which place it was transmitted to the printer. The printer printed the tape-switch sequence number and the current clock time to the nearest 100 μs each time a switch was activated. The system had controls for resetting and starting the tape-switch system and for indicating good and bad data. It also had a series of switches for turning on or off each individual input channel.

Each trap consisted of 3 tape-switch detectors: 2 were installed perpendicular to the direction of travel 18 ft (5.49 m) apart, and 1 was installed in the middle of the road at a 45-deg angle to the direction of travel. The detectors extended 6 ft (1.83 m) into the roadway so that traffic in the right lane only was recorded.

Figure 2 shows the layout of the tape-switch trap. The formula for speed is

$$V = L/t_L$$

where

V = speed,
L = distance from tape switch 1 to tape switch 3,
$t_L = t_3 - t_1$,

t_1 = time vehicle passed over tape switch 1, and
t_3 = time vehicle passed over tape switch 3.

The formula for lateral placement is

$$P = (Lt_D/t_L - R) \tan \theta$$

where

P = lateral placement,
$t_D = t_2 - t_1$, and
t_2 = time vehicle passed over tape switch 2.

An error analysis of the tape-switch system was undertaken, and it was found to be accurate to within ±0.1 mph (0.16 km/h).

RESULTS

Before recording speeds and placements for the curb and guardrail conditions on the bridge, we established the base condition to provide a basis for comparison. Because curb condition tests were run in 1973 and guardrail condition tests were run in 1974, running a base condition test in 1974 to determine whether the 55-mph (88.6-km/h) speed limit had any effect on speed was thought necessary. The base condition data for 1973 were collected on June 8, 11, 12, and 14. The base condition data for 1974 were obtained on June 13. Because these 2 periods were at the same time of year, speed variation due to time of year was eliminated. A related study (1) showed that days of the week could be combined for statistical analysis. In each case data were taken on days with good weather (no rain). For data collection with curb and guardrail in place, a warning sign stating TEST BRIDGE AHEAD was placed 1,500 ft (457.2 m) upstream of the site. For data collection during the base conditions discussed in this paper, no such sign was on the site.

The speed profiles for the 2 different base conditions for the first trap were calculated and are shown in Figure 3. Trap 1 was used in all comparisons because it was least affected by the grade through the site and proximity to the bridge.

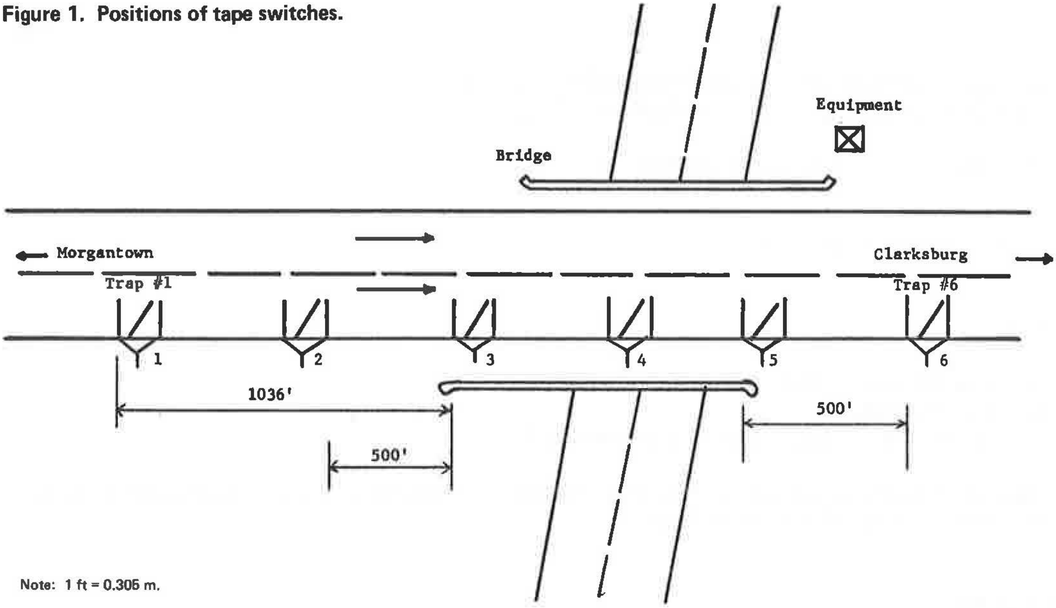
In 1973, the sample size was 851; in 1974, it was 245. For 1974 the mean speed was 54.5 mph (87.7 km/h), and for 1973 it was 61.0 mph (98.2 km/h). Two statistical tests were run: one to determine the equality of variances, the other to determine the equality of means. In both cases a 5 percent level of confidence was used. To determine the equality of variances an F-test was used. The F-ratio was formed. In this case $F = 2.22$. If this is less than the critical F-value, then the hypothesis of equal variances is accepted. In this instance, for a 5 percent level of confidence and $n - 1$ degrees of freedom (df), the critical F-value is 1.23. Therefore, the hypothesis is rejected and the variances are not equal.

To test the hypothesis of equal means when the variances may not be equal, the Smith-Satterthwaite t' statistic is used (2). This approximates Student's t-distribution with n df. The values of t' and n are

$$t' = (\bar{x}_1 - \bar{x}_2)/(S_1^2/n_1 + S_2^2/n_2)^{1/2}$$

$$n = (S_1^2/n_1 + S_2^2/n_2)^2/[(S_1^2/n_1)^2/(n_1 - 1) + (S_2^2/n_2)^2/(n_2 - 1)]$$

Figure 1. Positions of tape switches.



Note: 1 ft = 0.305 m.

Figure 2. Typical tape-switch trap.

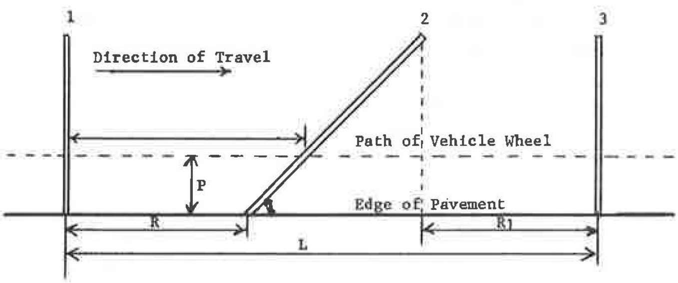
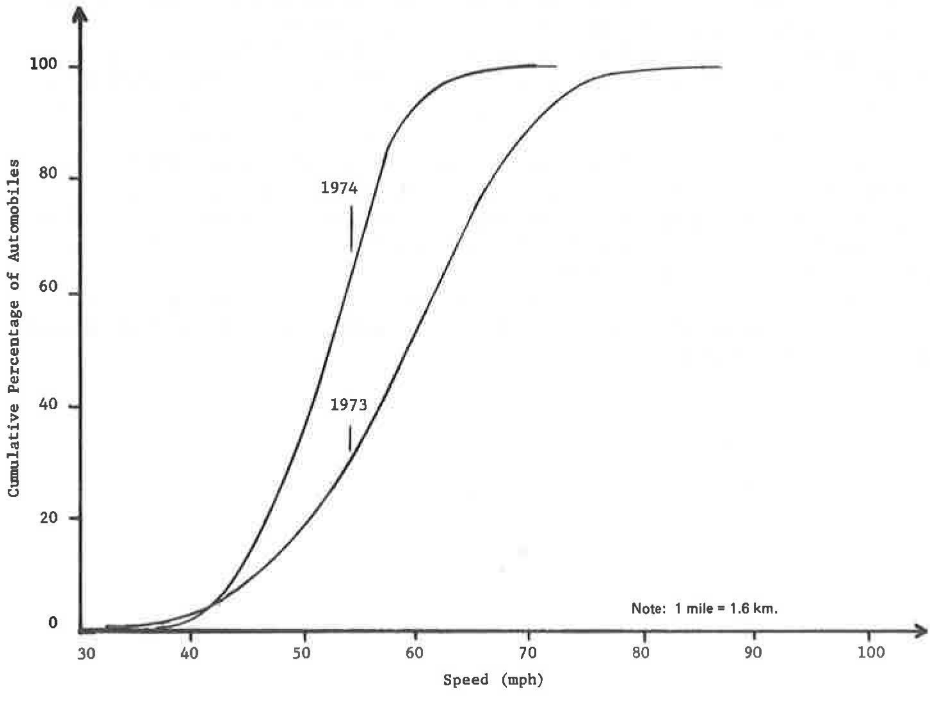


Figure 3. Speed profile for trap 1, 1973 and 1974.



Note: 1 mile = 1.6 km.

where

X_1 = sample mean,
S_1^2 = sample variance, and
n_1 = sample size.

In this case t' = 12.89 with 587 df. Using a 2-sided t-test at a 5 percent level of confidence, we can reject the hypothesis of equal means if $t' < t_{0.025}$. The critical value of t is 1.90, so the hypothesis of equal means is rejected.

The standard deviation was 6.2 mph (10 km/h) for 1974 and 9.2 mph (14.8 km/h) for 1973. One may establish ranges for the 2 years by using the usual figures for a normal distribution in which 95 percent of the sample lies within 2 standard deviations of the mean and 99 percent of the sample lies within 3 standard deviations of the mean. These and other distributions are given below (where 1 mph = 1.6 km/h).

Item	1973 Speed (mph)	1974 Speed (mph)
95 percent of sample	42.6 to 79.4	42.1 to 66.9
99 percent of sample	33.4 to 88.6	35.9 to 73.1
15th percentile	47	49
85th percentile	68	58

With regard to the established speed limits 50 percent of the drivers exceeded 55 mph (88.6 km/h) and 8 percent exceeded 60 mph (96.6 km/h) in 1974 and 10 percent exceeded 70 mph (112.7 km/h) and 3 percent exceeded 75 mph (120.8 km/h) in 1973.

CONCLUSIONS

From the examination of the results, one may conclude that not only has the mean speed been reduced but also the distribution of speeds has been compressed so that differences in speeds of automobiles on the highway have been reduced. An additional explanation for the reduction in accidents and fatalities may be that not only have speeds been reduced, which lowers the severity of accidents, but also speed differences have been reduced, which provides fewer conflicts.

Although the reduction in the variance of speeds also may affect level of service, no effect on highway capacity is envisioned because capacity flow occurs at speeds that are somewhat less than those discussed here. However, significant effects on service volumes for various levels of service may be anticipated. A lower variance in the speed distribution would imply that fewer passing maneuvers need be performed. Because the levels of service are defined qualitatively (ease of passing is one of the measures), the need for fewer passing measures would imply that service volumes could be raised within each level of service without adversely affecting the quality of flow, at least for those levels of service that are less than capacity.

The limitations of these data also should be kept in mind when one draws conclusions. These data were taken at 1 place on an Interstate highway in West Virginia and apply only to free-flowing automobiles. Despite these limitations we feel that the results herein are significant and offer a possible explanation for the recent reduction in accidents.

ACKNOWLEDGMENTS

This paper is part of a report prepared for the West Virginia Department of Highways in cooperation with the Federal Highway Administration. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the state or the Federal Highway Administration.

We wish to acknowledge the help of Fern Wood, Karen Ruckle, Majid Kabariti, J. S. Duggal, Mike Romansky, Kulanand Jha, and Charles Kona.

REFERENCES

1. R. R. Roberts. The Effect of Bridge Shoulder Width on Traffic Operational Characteristics. West Virginia University, PhD dissertation, 1974.
2. I. Miller and J. E. Freund. Probability and Statistics for Engineers. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1960.

SPONSORSHIP OF THIS RECORD

GROUP 3—OPERATION AND MAINTENANCE OF TRANSPORTATION FACILITIES
Lloyd G. Byrd, Byrd, Tallamy, MacDonald, and Lewis, chairman

Committee on Highway Capacity and Quality of Service
Robert C. Blumenthal, Alan M. Voorhees and Associates, Inc., chairman
Arthur A. Carter, Jr., Federal Highway Administration, secretary
Donald S. Berry, H. A. Mike Flanakin, B. D. Greenshields, Joseph W. Hess, Jack A. Hutter, Thomas D. Jordan, James H. Kell, Jerry Kraft, Jack E. Leisch, Adolf D. May, Jr., Karl Moskowitz, Louis J. Pignataro, Carlton C. Robinson, John L. Schlaefli, Gerald W. Skiles, T. Darcy Sullivan, Joseph A. Wattleworth

K. B. Johns, Transportation Research Board staff

The organizational units and the chairmen and members are as of December 31, 1974.