TRANSPORTATION RESEARCH RECORD 644

Highway Capacity, Traffic Flow, and Traffic Control Devices

TRB
TRANSPORTATION RESEARCH BOARD
NATIONAL ACADEMY OF SCIENCES

Contents

TECHNIQUE FOR MEASURING DELAY AT INTERSECTIONS
William R. Reilly and Craig C. Gardner 1

CONSISTENCY OF MAXIMUM FLOW CHARACTERISTICS AND CONGESTION PATTERNS ON AN URBAN FREEWAY DURING MORNING PEAK PERIODS
Avishai Ceder and Adolf D. May 8

PROCEDURE FOR ESTIMATING DEMAND FOR REGIONAL FRINGE PARKING FACILITIES
R. K. Mufti, L. S. Golfin, and C. D. Dougherty 15

DESIGNING A PARKING MANAGEMENT PROGRAM
Ann B. Rappaport 19

CRITICAL LANE ANALYSIS FOR INTERSECTION DESIGN
Carroll J. Messer and Daniel B. Fambro 26
Discussions
Stephen G. Petersen 33
Frederick A. Wagner 34
Authors' Closure 35

REVIEW OF ROAD TRAFFIC NETWORK SIMULATION MODELS
Paul Ross and David Gibson 36

SIMULATION OF BUS LANE OPERATIONS IN DOWNTOWN AREAS
R. W. Bowes and J. van der Mark 41

MICROSCOPIC TRAFFIC SIMULATION PACKAGE FOR ISOLATED INTERSECTIONS
Thomas W. Rioux and Clyde E. Lee 45

POSTOPTIMALITY ANALYSIS METHODOLOGY FOR FREEWAY ON-RAMP CONTROL
(Abridgment)
Menahem Eldor and Ilan Adler 51

AREAWIDE IMPACT OF TRAFFIC CONTROL DEVICES
(Abridgment)
Gerard M. Ebbecke and James J. Schuster 54

APPROACH TO REAL-TIME DIVERSION OF FREEWAY TRAFFIC FOR SPECIAL EVENTS
Graeme D. Weaver, Conrad L. Dudek, Donald R. Hatcher, and William R. Stockton 57

VEHICLE PLATOON PARAMETERS: METHODOLOGY FOR TRAFFIC CONTROL
(Abridgment)
Jose L. Rodriguez and Roy C. Loutzenheiser 62

COMPARISON OF TWO TYPES OF LEFT-TURN CHANNELIZATION
(Abridgment)
Judy C. Chang, Judson S. Matthias, and Mary R. Anderson 64

ACCIDENT EXPERIENCE WITH RIGHT TURN ON RED
Hugh W. McGee 66

DRIVER BEHAVIOR DURING THE YELLOW INTERVAL
(Abridgment)
William L. Williams 75

OPTIMIZATION OF PRETIMED SIGNALIZED DIAMOND INTERCHANGES
Carroll J. Messer, Daniel B. Fambro, and Stephen H. Richards 78

VIRGINIA'S CRASH PROGRAM TO REDUCE WRONG-WAY DRIVING
N. K. Vaswani 84

EFFECT OF FLASHING BEACONS ON INTERSECTION PERFORMANCE
R. B. Goldblatt 91

EFFECTS OF SIGNAL PHASING AND LENGTH OF LEFT-TURN BAY ON CAPACITY
Carroll J. Messer and Daniel B. Fambro 95

OPTIMIZING SETTINGS FOR PEDESTRIAN-ACTUATED SIGNAL CONTROL SYSTEMS
Feng-Bor Lin 102

LEVEL OF SERVICE AT SIGNALIZED INTERSECTIONS
T. C. Sutaria and J. J. Haynes 107

ESTIMATION OF UNPROTECTED LEFT-TURN CAPACITY AT SIGNALIZED INTERSECTIONS
Daniel B. Fambro, Carroll J. Messer, and Donald A. Andersen 113

EFFECT OF DOTTED EXTENDED LANE LINES ON SINGLE DECELERATION LANES
(Abridgment)
W. D. Keck and A. W. Roberts 120

EVALUATING LOCATION EFFECTIVENESS OF FREEWAY DIRECTIONAL AND DIVERSION SIGNS
(Abridgment)
Moshe Levin 122

INFLUENCE OF CONTROL MEASURES ON TRAFFIC EQUILIBRIUM IN AN URBAN HIGHWAY NETWORK
Nathan H. Gartner 125

OPTIMUM CONTROL OF TRAFFIC SIGNALS AT CONGESTED INTERSECTIONS
(Abridgment)
Panos G. Michalopoulos 129

SIMULATION AND CONTROL OF TRAFFIC ON A DIAMOND INTERCHANGE
C. H. Knapp, D. Ghosh, and R. Jain 132

Technique for Measuring Delay at Intersections

William R. Reilly, JHK and Associates, Tucson, Arizona
Craig C. Gardner, Tucson Department of Traffic Engineering, Arizona

This report presents our findings on a design for a simple, accurate technique for measuring vehicular delay on the approach to a signalized intersection. Precise definitions were established for four measures of performance: stopped delay, time-in-queue delay, approach delay, and percentage of vehicles stopping. Approach delay was selected as being most representative of intersection efficiency. Four manual methods using film taken at 10 intersections were tested in the laboratory. The values thus obtained were statistically compared with true values from time-lapse photography. The point sample, stopped delay procedure and the percentage of vehicles stopping method were selected as the most promising methods for practical use and were performed in the field at three sites. A user's manual for application of these two methods was produced but is not included in this report. Correction factors were developed to allow the field results to more accurately estimate the true values of stopped delay and percentage of vehicles stopping. Interrelationships among the four measures of performance were established so that approach delay could be estimated from a value for stopped time.

It has always been difficult to quantitatively define operational efficiency for approaches to signalized intersections by using traffic engineering techniques. In 1965, the Highway Capacity Manual (1) introduced the concepts of load factor and level of service in an attempt to relate intersection efficiency to some factor obtainable in the field. However, load factor has not received widespread acceptance, and several researchers have postulated that this measure may not be a reliable indicator of intersection performance. It has been generally agreed that some measure of vehicle delay should provide a practical and meaningful measure of performance.

There are many reasons for considering delay (and perhaps stops) a better measure than load factor. First, load factor is by definition a measure applied to each individual approach. To date, no method has been devised in which load factor can be used to provide a single measure of overall intersection operation. Second, load factor is not a good measure for locations with traffic-actuated signals. For example, a phase is considered "loaded" if traffic continually enters the intersection on green, but at the first major gap in traffic the green is terminated. This can be considered an efficient traffic-actuated controller and does not imply congested or near capacity conditions. Third, small changes in volume appear to cause large changes in the value for load factor; for example, relatively small increases in volume can bring load factor from 0.0 or 0.1 to 0.7 or higher in a short period of time. Also, conditions near capacity flow are not well defined, and conditions of over capacity flow are not described by load factor.

In view of these deficiencies, the need for a different measure of operating effectiveness of signalized intersections is apparent. This need has prompted several researchers to explore the use of delay to measure performance.

The scope and objectives of the present research were limited to the definition of several delay types and to the formulation of a practical and accurate method that can be used in the field to measure delay. A complete description of the research, including a user's manual, is given in a three-volume report of the Federal Highway Administration (2).

STUDY PROCEDURE

Most previous work related to delay measurement did not clearly describe either the phenomenon to be measured or the details of the measurement technique. The Federal Highway Administration, recognizing this deficiency, set out objectives for a research project that would provide a precisely defined technique for measuring vehicle delay at signalized intersections. The objectives of the study were

1. To identify and define various measures of vehicle delay on approaches to signalized intersections,
2. To select the delay measure most appropriate for use by practicing traffic engineers, and
3. To develop a field method for collecting data that would lead to the most appropriate measure.

Synopsis of Related Work

One early effort to quantify vehicle flow characteristics on approaches to intersections was reported by Greenshields (3). The use of a 16-mm camera to capture traffic flow for subsequent analysis in the laboratory made this early work particularly noteworthy. In 1940, Rivett (4) presented a report on the use of mechanical aids (a desk calculator) in collecting data on vehicle delay at intersections.

Certainly the most complete and probably most important work related to field measurement, in contrast to theoretical modeling, was conducted by Berry and others (5). This work led to the establishment of several techniques for measuring intersection delay. Also, Berry's work included the first major effort to define different types of delay and to estimate the interrelationships among delay types.

In a 1957 paper, Solomon (6) described a measurement technique that related to Berry's procedure but was applied to a different type of delay. In the late 1960s, May and Pratt (7) published an article that can be considered as the beginning of the search for a measure of intersection performance more easily applicable and more meaningful than load factor. May and Pratt suggested that performance would be better described by a measure of delay than by load factor. Then Sagi and Campbell (8) described a new technique that could be relatively easily applied and would give a value for vehicle delay on the approaches to traffic signals. Buehler, Hicks, and Berry (9) used a questionnaire to survey existing practices for delay studies, but the results did not point conclusively to one method or one delay type as being most widely used and accepted in the United States.

Much research has been directed toward theoretical models for estimating stops and delay. These models have been based on assumptions related to patterns of vehicle arrival and departure on an intersection approach. All the modeling work has suffered in one important aspect: No matter how simple the model, the basic assumptions have not been generally applicable to a wide variety of intersections. For example, one common assumption has been that vehicles arrive randomly. However, in an interconnected signal system this assumption does not normally hold true. Most of the modeling developed has contributed to a better understanding of traffic flow theory but has not been of practical use. The work by Sagi and Campbell was perhaps the best effort toward combining a certain amount of traffic flow theory with a practical field procedure.

In summary, most previous studies related to delay measurement techniques have had one or more of these deficiencies. (a) No clear definition of delay measures was given. (b) Much attention was paid to mechanical or electronic devices that merely serve as aids, while relatively little attention was paid to the validity of the procedure itself in providing good estimates of the delay measure. (c) Several new techniques reported in the literature have simply been old techniques applied in a slightly different manner. (d) Many studies relied on techniques that could be applied easily only under certain traffic or geometric conditions (for example, at intersections with a pretimed traffic signal control). These techniques are not well suited to general application. (e) Most modeling techniques used in the past are somewhat cumbersome for the practicing traffic engineer and are not particularly suited for application to a wide range of intersection types.

Study Execution

The approach used to achieve our research objectives was composed of the following steps.

1. A review of pertinent literature and previous research was made.
2. Several types of vehicle delay of the type found on approaches to signalized intersections were identified, and a precise definition was developed for each type.
3. The delay type that appeared to best portray the efficiency of intersection operation was selected. The term "approach delay" was used to identify this best measure.
4. Methods for field data collection were identified and defined. These methods were designed to collect and reduce traffic data so that the selected measure of approach delay could subsequently be estimated.
5. Using combinations of delay types and field data collection methods, a selection was made of the most promising methods to be tested during the research.
6. Ten intersection approaches from four urban areas and with differing physical and operational characteristics were selected for study.
7. For each study approach, two periods were filmed: 50 min during off-peak conditions and 50 min during peak conditions. Two cameras, one for time-lapse photography at 1 frame/s and one for real-time photography at 16 frames/s were run simultaneously during each study period.
8. The real-time film was viewed in the laboratory for simulation of manual field studies. The selected delay types and manual methods were used to collect data from the film.
9. The time-lapse film was used to obtain precise data on each vehicle observed on the study approach. For each study period a precise (also referred to as "true" in this report) value for each measure of delay and stops was obtained.
10. The true values were compared with the values obtained from the manual methods for each delay type and for each study technique. Statistical tests were used for the analyses.
11. The manual methods that appeared to best meet the objectives of the research were selected for field validation.
12. Three new study approaches were used to train a field crew in the selected methods, and two 1-h studies (covering peak and off-peak) were conducted at each approach. Time-lapse film was taken at the same time the field crew was performing each study.
13. In the laboratory the time-lapse film was analyzed and the measures of stops and delay were obtained. Statistical analyses were used to compare these values with those obtained by the field crew.
14. A final report and a user's manual were prepared.

VEHICLE DELAY

Characteristics and Definitions of Delay

Figure 1 describes the movement of vehicles along the approach to a signalized intersection. Three types of typical vehicle movements are shown. Also shown is the time-space relationship for an unimpeded vehicle with no stops or delays on the approach.

In the analysis of delay, at least two points along the approach to a signalized intersection must be found. First, the point at which a moving vehicle is considered to be leaving the approach should be fixed. Noting that one objective of this research was to develop an efficient and relatively simple method of manual data collection, the STOP line or the first crosswalk line traversed by the approaching vehicle was considered the most obvious point to use for definition of a leaving vehicle. If neither of these lines exists, some type of mark to denote the location of a STOP line could be made for the purposes of the field study. The second point is located upstream from the intersection under study. This point would be situated so as to include all delay (including deceleration-acceleration cycles) created by the signalized intersection under normal peak flow conditions. The location of this point would vary from intersection to intersection but would be based on the same criterion, that is, located far enough upstream to include all delays caused by the traffic signal but not so far upstream to include delays caused by other traffic signals or major cross-street flows.

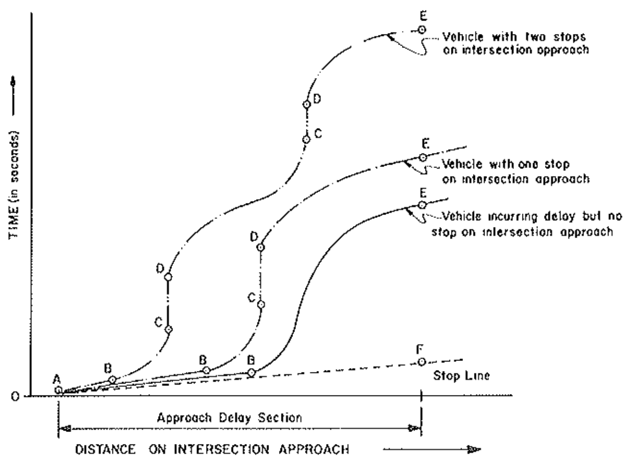
Various definitions and terms were utilized during the research and are recommended for future work in the analysis of intersection delay. The most important terms are defined in the following table.

Term	Description
Approach delay section	Section where most or all approach delay is incurred
Approach free flow time	Time used by unimpeded vehicle to traverse approach delay section
Approach time	Time used by any vehicle to traverse approach delay section
Approach delay	Approach time minus approach free flow time
Stopped delay	Time vehicle is stopped, with locked wheels, equal to stopped time
Time-in-queue delay	Time from first stop to vehicle's exit across STOP line, equal to time in queue
Percentage of vehicles stopping	Number of vehicles incurring stopped delay divided by number of vehicles crossing STOP line

One important distinction should be noted here: Vehicles moving along an approach experience a series of "times," the sum of which will be equal to the approach time. The term "delay" is not always synonymous with time. For example, for a given vehicle approach, delay is the difference between two measured times, while stopped delay is in fact equal to stopped time.

The above definitions can be better understood by referring to Figure 1, in which the approach delay section lies between point A and the STOP line. The approach

Figure 1. Time-space relationships.



time of a vehicle incurring delay is the time in seconds from zero to point E. The approach delay is the difference between points E and F. Stopped delay is the difference between points C and D, and time-in-queue delay is the difference between points C and E.

Descriptive Analysis of Delay Types

Of the many types of delay, past research and present practice suggest that three are of practical use. In addition, the number or percentage of vehicles forced to stop is another important flow characteristic that should be considered. Thus, the four measures of intersection performance selected for study were (a) approach delay, (b) stopped delay, (c) time-in-queue delay, and (d) percentage of vehicles stopping.

Most researchers agree that, although the best indicator of intersection performance is approach delay, this is difficult to obtain in the field. Therefore it was abandoned in favor of more easily measured factors.

Approach delay, unlike other measures, relates to the total time during which drivers and passengers are delayed and thus can easily be used in analyses of road-user time costs. Stopped delay is an obvious measure to the driver but can overstate efficiency of operation under conditions where the length of stop is short and is followed by slow movement in a long sluggish queue. Time-in-queue delay will often lie between the values for stopped delay and approach delay. However, for any given vehicle, time-in-queue delay can be greater than approach delay. Under certain traffic conditions, this measure can overstate the amount of delay being incurred by motorists.

During the study, approach delay was considered the best measure to describe operation effectiveness on the approach to a signalized intersection. Approach delay, although technically more difficult to collect in a direct manner, appeared to be better than either stopped delay or time-in-queue delay for describing and comparing intersection operation.

Percentage of vehicles stopping was not directly comparable to the delay measures. In the research, this measure of intersection performance was considered useful and was selected for testing along with three manual methods related to delay.

INTERSECTION DELAY MEASUREMENT PROCEDURES

Definition and Analysis of Basic Procedures for Estimating Delay

Review of the literature and of possible techniques not previously described led to the establishment of four basic procedures that could be used in estimating delay. All past and present efforts were categorized as one of the four. In several cases, a method contained elements from two or more of the basic procedures. The four procedures are (a) point sample, (b) input-output, (c) path trace, and (d) modeling.

The point sample method is based on a periodic sampling of some factor (such as number of stopped vehicles) on the intersection approach. In essence, it is a series of instantaneous samples having an interval of time between each sample. An example of this technique is the method commonly known as the Berry-VanTil procedure.

The term "interval sample" might also be used to describe the input-output method. It is similar to the point sample method but uses an infinitely short interval (zero) between samples and a long sample period of, say, 10 or 15 s. The factor being measured is observed at its beginning (input) and end (output) points.

The path trace is based on a sample of individual vehicles using the study approach. Data on each vehicle sampled is recorded over the period of time the vehicle is within the study area (for example, the approach delay section). Using measurements on the sample group of vehicles, a statistical expansion of the data to represent all vehicles is made. This method is very similar to a traffic engineering spot-speed study.

The use of modeling in estimates of delay can include a wide range of field and analysis techniques. All methods that use one or more theoretical assumptions regarding arrival patterns, departure patterns, driver behavior in queues, and traffic signal operation are in this category.

How well each of the four basic procedures related to the objectives of the research was analyzed.

The point sample method has several advantages. First, the technique is self-correcting in that an error or omission in one sample will have almost no effect on the overall result, because each sample is independent of the previous one. Second, the technique is not dependent upon signal indications, except for the restraint of periodicity. This restraint refers to the need for a set of data points arrayed throughout the signal cycle and providing a representative sample of all traffic conditions in the cycle. A disadvantage of this method is that the accuracy of an observer's point sample might be considerably reduced if the count (say, of stopped vehicles) becomes quite high.

The input-output method suffers from one important disadvantage. The field data should be corrected for vehicles that enter or leave the study area between the point of input and the point of output. Correction factors should be applied at the beginning and end of the study period, and also at regular intervals throughout the study, to compensate for observer errors.

The path-trace method was considered simple to perform in the field and would yield, from a single study, all four measures of delay and stops described earlier. However, it was hypothesized that a very large sample of vehicles would be needed to provide an estimate of delay within reasonable levels of confidence.

Because modeling methods were considered too esoteric and difficult to apply to varying intersection conditions, we eliminated this category.

Selection of Methods to Be Tested

To provide a guide for selection of the most promising two or three delay measurement methods, a matrix of the four basic methods applied to the three basic measures of delay was developed to give 12 possible study methods. Each of these was then judged against a framework of eight criteria. For each, an overall utility score was obtained that served as an indication of the method's effectiveness in meeting the objectives of the research.

Factors such as field manpower requirements, need to establish base speed, and generalized application were among the eight selection criteria. One important criterion was whether a method depended upon continuous observance of traffic signal indications. If so, the method was considered less useful.

The procedures selected for testing were (a) point sample, stopped delay, (b) point sample, time-in-queue delay, (c) path trace, and (d) percentage of vehicles stopping.

Following selection of three delay procedures and a fourth related to percentage of vehicles stopping, a detailed study design was developed for each. It was decided that no electronic or complex mechanical devices would be used in applying the methods to the 20 film segments in the laboratory. In this way, a truly simple manual technique would be tested.

Description of Manual Methods Tested

The point sample, stopped delay technique was designed for intervals between samples of 15 or 13 s, the latter value being used at locations having pretimed signal controllers. For all but 2 of the 20 film segments studied, a two-person team was used to obtain the samples on a lane-by-lane basis. A third person, using a stopwatch, gave a cue at each sampling point. The team noted the number of stopped vehicles in each lane at each sampling point. The total stopped delay was computed by multiplying the interval between samples by the number of vehicles counted in all samples.

The point sample, time-in-queue delay technique was designed identically to the point sample, stopped delay method. Only the phenomenon being observed was different. Following a stop, a vehicle continued to be in queue until it crossed the STOP line.

The path-trace technique was performed by a three-person team for each of the 20 film segments. One person served as sample selector. This person counted vehicles crossing into the approach delay section at its upstream end. A previously fixed sample rate based on volume was used for each film, and the sample selector gave a cue to one of the two observers each time a selected vehicle approached the entry point of the section. The observer started a stopwatch when the vehicle entered the section and noted the elapsed time of all actions such as stop, start, change lane, and leave section at STOP line. This technique led to estimates of total volume, stopped delay, time-in-queue delay, approach delay, and percentage of vehicles stopping.

The fourth manual method was a count of all motorized vehicles crossing the STOP line. The count was categorized into stopping or not stopping. Any vehicle that stopped one or more times on the intersection approach was counted as one vehicle in the stopping category. The results from this method give estimates of both total volume and percentage of vehicles stopping.

DATA COLLECTION AND REDUCTION

Study Sites and Filming Procedures

Ten approaches to signalized intersections were used; each approach was filmed for two 50-min periods. Table 1 gives a list of the sites and basic characteristics of each. The study sites represented a broad range of geographic, geometric, and traffic conditions. The research methods were general in that they could be successfully used in varied situations and would not be dependent on specific features of an intersection.

A 16-mm camera with a crystal speed control was used to film 50 continuous min at exactly 16.0 frames/s. Color film was used and then studied in the laboratory in a real-time mode for the four manual methods described earlier.

An 8-mm camera with an intervalometer set at exactly 1.0 frame/s was run simultaneously at each site with the 16-mm camera. Color film was used and then studied in the laboratory in a time-lapse mode on a projector-analyzer.

Laboratory Work

Each of the four manual methods was performed on each of the 20 films, and arithmetic values for various delay measures were computed. All of the data were taken on a lane-by-lane basis, and the final summary of values for each method was given by lane and also for the total approach.

The time-lapse work involved studying each vehicle individually and determining the frame numbers at which the vehicle performed some action (i.e., stop, start, change lane, cross STOP line). Some 22 000 vehicles were included in the 20 films.

The data on each vehicle were punched onto standard cards, and a computer program developed during the research ran an error check on the data. Following corrections of the data base, the program calculated and summarized all possible measures related to stops and delay for each film. Recognition of vehicles that either entered or left the approach delay section at some intermediate point was given in the time-lapse work and was deemed important in computing true values for stops and delay. Also, all lane changes were noted from the time-lapse film, and this information was used to properly assign various types of delay to each lane.

During filming, data reduction, and analysis, tight control was maintained on all aspects of the work to ensure precise results. Such factors as calibration of stopwatches, camera speed, and projector speed were checked regularly.

DATA ANALYSES

Two general types of analysis were performed on the data. First, regressions of real-time values on the corresponding time-lapse values were derived and analyzed; second, regression analyses were made to compare the interrelationships among delay types. From the time-lapse film the following measures were obtained: (a) stopped time (equal to stopped delay), (b) time in queue (equal to time-in-queue delay), (c) approach delay, (d) percentage of vehicles stopping, and (e) volume estimate. From the real-time manual methods, the following measures were derived and are listed with their coded designation.

Measure	Coded Designation
Point sample, total stopped time	M1T
Point sample, stopped time per vehicle	M1PV
Point sample, time in queue	M2
Point sample, time in queue per vehicle	M2PV
Path trace, stopped time per vehicle	M3A
Path trace, time in queue per vehicle	M3B
Path trace, approach delay per vehicle	M3C
Path trace, percentage stopping	M3D
Path trace, volume estimate	M3E
Percentage of vehicles stopping	M4
Volume estimate from M4 study	FV

Ratio comparison was another type of analysis made. Several measures derived from the real-time studies were used to form a ratio with the true value of the corresponding measure derived from time-lapse film.

Statistical Terminology

The regression relationships discussed in the following sections are reported on the basis of the line of best fit in the form $Y = bX + a$, where Y is the variable plotted on the vertical axis, b is the slope of the regression line, X is the variable plotted on the horizontal axis, and a is the intercept of the best fit line at the vertical axis.

Also reported is the coefficient of determination (R^2) value for each regression line. This value relates to the amount of scatter of the data points about the regression line. A high value (greater than 0.90) indicates that the regression relationship is very strongly linear. The standard error is another statistic reported and is an indication of the range of values about the mean value that will encompass the true mean. Thus, if a value is reported as 1.10 ± 0.04, the indication is that 68 percent of the time the true mean value will lie between 1.06 and 1.14.

The term "significant" is also used in reporting the analyses. In this research, all tests were carried out by using Student's t-test at the 0.05 significance level, and all tests were of the two-tailed variety; i.e., significance would be declared if the statistic or value from the data set was either greater or less than the hypothesized value.

Manual Methods Compared With Time-Lapse Results

Table 2 summarizes the regression lines obtained when measures of delay, stops, and volume were compared with the corresponding measure derived from time-lapse photography.

Both of the point sample methods, M1 and M2, demonstrated a high level of precision: R^2 equals 0.99. This is an indication of a very strong linear relationship between the stopped time or time in queue derived from the field study and the true values. As noted in Table 2, the upward bias of the slope of the point sample lines is significantly different from 1.0, and in the case of total time in queue (M2T) the intercept is negative and significantly different from 0.0.

One definite conclusion reached from studying these relationships was that an upward bias existed in the point sample method, whether it was applied to stopped time or to time in queue. Thus, if the method had been applied in the field, the estimate of stopped time or time in queue would have been higher than the true value.

For the path-trace study, five measures were computed and each regressed against the true value of the measure obtained from the time-lapse work. Table 2 shows that except for the volume estimate all measures had a regression line slope not significantly different from 1.0. Thus, it can be said that the path-trace method is quite accurate but not as precise as the point sample methods, as is evidenced by lower values of R^2.

Interrelationships Among Delay Types

An important part of this research was the determination of the relationship between the true value for a measure and all other measures.

Table 3 lists the statistical qualities of the interrelationships. It is interesting to note that time in queue accounts for about 97 percent of approach delay and stopped time for about 76 percent of approach delay. The three linear relationships of delay with percentage stopping are not strong. The last relationship in Table 3 represents an attempt to develop a linear, as opposed to a curvilinear, relationship between percentage stopping and approach delay. By taking the log of the value for approach delay per vehicle, a strong linear relationship does result. This could be very significant because, of all field procedures, perhaps the easiest and least costly to perform is the study of percentage of vehicles stopping. If a good predictive relationship exists between percentage stopping and approach delay, the percentage stopping study might be used by jurisdictions lacking manpower to perform the more elaborate delay studies.

One potential disadvantage of using percentage stopping values for estimating delay is that, when conditions force all vehicles to stop, delay can increase drastically while percentage stopping remains constant at 100 percent. Because the final recommendations of this study did not include the use of percentage stopping values to estimate delay, this potential difficulty was not explored in detail.

Other Analyses

In addition to the analyses described above, several other factors were studied. First, each of the four basic measures was regressed against a ratio of volume to service volume at level of service C. The R^2 values for these plots ranged from 0.47 to 0.63, indicating that considerable scatter in the data points existed. One interesting fact was observed from these plots: All of the low delay locations were controlled by an interconnected signal system, while most of the high delay locations operated with isolated local control.

For the path-trace method, an analysis was made of the sample size necessary to achieve reasonable results. Depending on delay type, from 1200 to 2700 vehicles would be needed in a path-trace sample to obtain delay estimates at the 95 percent confidence level.

Finally, an analysis of arithmetic ratios was made. The value for a given measure taken from the real-time studies was divided by the true value for the same measure. Ratios were computed for the two methods that were finally recommended: point sample, stopped delay and percentage of vehicles stopping.

The ratio analyses can be summarized by the following observations. (a) There appears to be a strong upward bias in the estimate of stopped time from the point sample method when compared with the true value. It is interesting to note that only three of 20 ratios were below 1.000 and that all three occurred in peak-hour studies under heavy volume conditions. (b) For percentage stopping, the estimate from the manual method was always greater than the true value. (c) For volumes taken from the percentage stopping study, no correcting factor is necessary to achieve an accurate estimate of true volume.

Table 1. Study sites.

Film	Intersection	City, State	Direction Traffic Approached From	Type of Signal Control	No. of Moving Lanes on Approach	Signal System Operation	Exclusive Left-Turn Lane
1, 2	Pleasant St.* and Massachusetts Ave.	Arlington, Mass.	South	Pretimed	3	No	Yes
3, 4	Massachusetts Ave.* and Everett Ave.	Cambridge, Mass.	North	Pretimed	2	Yes	No
5, 6	Washington St.* and Madison St.	Alexandria, Va.	South	Pretimed	4	Yes	No
7, 8	Leesburg Pike* and Haycock Rd.	Falls Church, Va.	East	Semiactuated	2	No	No
9, 10	University Blvd.* and Viers Mill Rd.	Wheaton, Md.	East	Fully actuated	4	Yes	Yes
11, 12	Classen Blvd.* and N.W. 23rd St.	Oklahoma City, Okla.	North	Pretimed	3	No	Yes
13, 14	N.W. Expressway* and Pennsylvania Ave.	Oklahoma City, Okla.	West	Fully actuated	4	No	Yes
15, 16	Broadway Blvd.* and Craycroft Rd.	Tucson, Ariz.	East	Fully actuated	5	No	Yes
17, 18	Congress St.* and Granada Ave.	Tucson, Ariz.	West	Fully actuated	4	No	Yes
19, 20	Speedway Blvd.* and Mountain Ave.	Tucson, Ariz.	East	Semiactuated	3	Yes	Yes

*Street traffic approached on.

Table 2. Regressions of manual measures versus time-lapse values.

Method	Measure	Intercept[a]	Slope[a]	R^2
Point sample	M1T	-1350 ± 823	1.15[b] ± 0.03	0.99
	M1PV	-0.42 ± 0.64	1.10[b] ± 0.02	0.99
	M2T	-2010[b] ± 860	1.18[b] ± 0.02	0.99
	M2PV	-0.65 ± 0.79	1.12[b] ± 0.02	0.99
Path trace	M3A	-0.01 ± 2.54	0.99 ± 0.09	0.86
	M3B	-1.49 ± 3.09	1.04 ± 0.09	0.88
	M3C	-1.53 ± 2.38	1.04 ± 0.06	0.93
	M3D	-1.47 ± 3.83	0.98 ± 0.06	0.94
	M3E	90.07[b] ± 46.41	0.89[b] ± 0.04	0.96
Percentage stopping	M4	4.26[b] ± 1.21	0.96 ± 0.02	0.99
	FV	17.02[b] ± 5.70	0.98[b] ± 0.01	0.99

[a] Reported as the coefficient ± standard error.
[b] Intercept significantly different from 0.0 or slope different from 1.0 by a statistical test criterion: Student's t-test at 0.05 significance level.

Table 3. Interrelationships of time-lapse measures.

Y Axis	X Axis	Intercept[a]	Slope[a]	R^2
Stopped time per vehicle	Approach delay per vehicle	-0.99 ± 1.41	0.76 ± 0.04	0.96
Stopped time per vehicle	Time in queue per vehicle	0.49 ± 1.07	0.78 ± 0.03	0.97
Stopped time per vehicle	Percentage stopping	-9.54 ± 4.97	0.54 ± 0.08	0.72
Time in queue per vehicle	Approach delay per vehicle	-1.99 ± 0.88	0.97 ± 0.02	0.99
Time in queue per vehicle	Percentage stopping	-11.62 ± 6.60	0.67 ± 0.10	0.70
Percentage stopping	Approach delay per vehicle	26.89 ± 5.58	1.03 ± 0.15	0.72
Percentage stopping	Log_{10} approach delay per vehicle	-14.04 ± 4.62	54.97 ± 3.33	0.94

[a] Reported as the coefficient ± standard error

SUMMARY OF RESULTS

The following points provide a synopsis of both the performance of the manual methods and the statistical analyses of the data. M1 and M2 refer to point samples of stopped time and time in queue respectively, M3 refers to the path-trace method, and M4 refers to the percentage stopping method.

1. M1 was somewhat simpler to explain to field personnel and to perform than M2 or M3.
2. M2 was slightly more difficult to perform than M1 or M3 because observers must continuously study all approach traffic.
3. All four manual methods, M1, M2, M3, and M4, appear to be quite precise in predicting the true value of stops or delay. M3 does appear to be slightly less precise than the others, however.
4. M1 and M2 are somewhat less accurate (slope of regression line greater than 1.0) than M3 in predicting delay. At least two possible reasons for the overestimation of delay by point sample methods have been identified. First, there may be a tendency for observers to concentrate more on the upstream end of the queue and thus to add vehicles into their counts while delaying slightly the subtraction of vehicles that have actually departed from the front end of the queue prior to the sampling point. Second, there may be skew in the distribution of stopped time and time in queue such that the use of 15 s (or 13 s) as the average time stopped for each vehicle observed gives biased results.
5. M4 may provide a simple method of estimating approach delay by use of a logarithmic relationship.
6. The addition of various independent variables to the regression equation of delay obtained from M1 or M2 results regressed on the true value of delay does not significantly improve the predictive power of the equation.
7. Stopped time averages about 76 percent of approach delay, while time in queue averages about 97 percent of approach delay. M1, M2, and M3 all provide estimates of delay that can be used with considerable precision to estimate approach delay.
8. Mechanical aids accompanying manual methods may be useful. In M1 and M2, an audible cue from a cassette recorder eliminates the need to be constantly checking a stopwatch. For M3, the sample selector would benefit from a simple digital hand counter for vehicles crossing into the section. Likewise, observers performing the M4 study would find such a counter with at least two buttons useful for recording stopping and not stopping vehicles.

FIELD VALIDATION

Two methods, M1 and M4, were selected as best for general use by traffic engineers. The methods were found to give accurate estimates of stopped delay, percentage of vehicles stopping, and total volume. The principal reason for selecting M1 over M2 was that the time-in-queue procedure was considered by field personnel to be more difficult to perform than the stopped time procedure. M3 was eliminated because of the large sample size required to achieve good results.

The purpose of the field validation was simply to apply the selected methods to an actual field study and to compare the results with those derived from time-lapse

Table 4. Validation regression relationships of field versus time-lapse values.

Measure	Slope[a]	Intercept[a]	R^2
Stopped time	$1.04^{b} \pm 0.01$	350 ± 254	0.99
Percentage stopping	1.00 ± 0.04	5.46 ± 2.39	0.99
Volume	0.97 ± 0.01	32.02 ± 12.20	0.99
Stopped time per vehicle[c]	$1.05^{b} \pm 0.01$	0.31 ± 0.26	0.99

[a] Reported as the coefficient ± standard error.
[b] Intercept significantly different from 0.0 or slope different from 1.0 by a statistical test criterion: Student's t-test at 0.05 significance level.
[c] Using time-lapse volume for divisor of field stopped time.

photography run simultaneously with the fieldwork. In this manner, a check was made of how well the methods provided reliable measures of delay and stops.

The scope of the validation work was described as including three sites different from those used for the 20-film periods. The two methods selected for validation were to be applied for two 1-h periods at each site, one period during off-peak traffic conditions and the other during peak conditions. During the same periods, 8-mm films of the study approach were to be taken at 1 frame/s.

The three sites were in Tucson, Arizona, and a field crew was trained prior to performing the two manual techniques. A total of six data points resulted, two for each site. The sites represented different approach widths, number of lanes, signal operation and phasing, and traffic volumes.

During the validation work, the length of stopped queues varied from several vehicles to 25 or 30 vehicles/lane. Thus, it was concluded that a wide range of conditions was encountered by field personnel and that the results of the validation were sound.

Following the field studies and the simultaneous filming, all data were reduced. Comparisons were made between the true values (from time-lapse film) for stops, delay, and volume and the values derived from the manual methods.

Table 4 summarizes the regression relationships of the validation work. The upward bias in estimates from the point sample, stopped delay method was confirmed. Also, the slight upward bias in estimates of percentage of vehicles stopping was confirmed. The estimates of total volume as computed from the M4 study appear to be accurate; no bias occurred in either direction.

After the two field methods had been defined, analyzed, and field tested, a small, easy to use manual was prepared. The intent of this user's manual was to provide all basic information needed to successfully apply the two recommended field methods: point sample, stopped delay and percentage of vehicles stopping. The manual will be distributed by the Federal Highway Administration to engineering agencies.

CONCLUSIONS AND RECOMMENDATIONS

It is recommended that the point sample, stopped delay study be used for field measurement of delay and that the percentage of vehicles stopping study be retained as a practical and useful procedure.

For the point sample, stopped delay study, it is recommended that the value for stopped time from the field be corrected by applying a 0.92 multiplier to obtain a more accurate estimate of true stopped delay.

For the estimate of volume from the percentage of vehicles stopping study, it is recommended that the field value be used directly, with no correcting factor.

The value resulting from a field study of percentage of vehicles stopping should be corrected by a multiplier of 0.96 to achieve a more accurate estimate of the measure.

Once the recommended field data corrections have been made, stopped delay per vehicle multiplied by 1.3 will yield a good estimate of approach delay per vehicle.

It is concluded that all four of the manual methods tested can be relatively easily applied by typical traffic engineering agencies and can also yield fairly precise and accurate estimates of delay. Therefore, although two of the methods were recommended for inclusion in the user's manual, the other two methods (point sample, time in queue and path trace) might be considered in future work for special application.

One other conclusion that was reached after extensive study was that intersection delay studies should not, in most cases, be performed on an individual lane basis. Rather, an entire approach should be studied at one time. Although in theory it is possible to study several lanes individually on an approach, in practice there are numerous complicating factors that increase manpower requirements and reduce the reliability of the study results.

ACKNOWLEDGMENTS

We wish to acknowledge the support and assistance received during the research from Robert O. Kuehl, professor of statistics at the University of Arizona. Also, the cooperation of the Tucson Traffic Engineering Division is appreciated. S. L. Cohen of the Office of Research, Federal Highway Administration, served as contract manager for the study and contributed many useful ideas to the research and to the final report.

REFERENCES

1. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
2. W. R. Reilly, C. C. Gardner, and J. H. Kell. A Technique for Measurement of Delay at Intersections. JHK and Associates, Vols. 1, 2, and 3; Federal Highway Administration, Repts. FHWA-RD-76-135, 136, and 137, 1976.
3. B. D. Greenshields. A Photographic Method of Investigating Traffic Delays. Proc., Michigan Highway Conference, 1934.
4. I. Rivett. A Simple Method for Tabulating Traffic Delay. Traffic Engineering, Sept. 1940.
5. D. S. Berry. Field Measurement of Delay at Signalized Intersections. HRB, Proc., 1956, pp. 505-527.
6. D. Solomon. Accuracy of the Volume-Density Method of Measuring Travel Time. Traffic Engineering, March 1957.
7. A. D. May and D. Pratt. A Simulation Study of Load Factor at Signalized Intersections. Traffic Engineering, Feb. 1968.
8. G. S. Sagi and L. R. Campbell. Vehicle Delay at Signalized Intersections, Theory and Practice. Traffic Engineering, Feb. 1969.
9. M. G. Buehler, T. J. Hicks, and D. S. Berry. Measuring Delay by Sampling Queue Backup. TRB, Transportation Research Record 615, 1976, pp. 30-36.

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

Consistency of Maximum Flow Characteristics and Congestion Patterns on an Urban Freeway During Morning Peak Periods

Avishai Ceder, Road Safety Centre, Technion-Israel Institute of Technology, Israel
Adolf D. May, Institute of Transportation Studies, University of California, Berkeley

Based on criteria for selecting maximum flow characteristics, data for 420 cells were collected on the Santa Monica Freeway for the period 1972 to 1975. Each cell includes three observed values: maximum flow rate, corresponding percentage of occupancy, and corresponding time. For the analysis of morning congestion patterns, 6 d of data (5:00 to 11:00 a.m.) were collected. These data were selected for a randomized-block design with three independent variables: years, workdays, and locations. Analysis-of-variance and multiple-pairwise-contrast procedures showed that the most crucial independent variable causing inconsistency is locations, the next most crucial is years, and the least important is workdays. Among the various results, the maximum flow characteristics were found to be affected by the change in the overall conditions during 1975 (particularly, a different daylight savings time and perhaps the reduced speed limit and other energy crisis factors). The morning congestion patterns and the time-sequence pattern of the flow-concentration trajectories were found to be relatively undisturbed during the 3-year period. The results may be useful when applied to problems of desirable flow and occupancy (concentration) under high demand level, estimated freeway capacities, sensitivity of an estimated capacity (capacity buffer), and traffic behavior during the peak period.

Three papers (1, 2, 3) served as the stimulus for this research report. Analyses of traffic flow models performed in these papers emphasize the need for investigating the consistency of traffic flow variables, particularly the maximum flow characteristics.

We attempted to amplify traffic behavior under normal peak-period conditions (on an urban freeway) and to examine qualitatively and quantitatively the variability of maximum flow characteristics during the period 1972 to 1975. In addition, and as a by-product, these characteristics were investigated in view of the changes in overall conditions caused by the energy crisis [the speed limit was reduced from 105 to 88 km/h (65 to 55 mph)] and by the different daylight savings time instituted in 1974.

SAMPLING PROCEDURE

Available Data in Los Angeles

The data considered in this paper were collected by the Los Angeles Area Freeway Surveillance and Control Project (LAAFSCP). This project is located in the heart of urban Los Angeles on three of the most heavily traveled freeways in the world: the Santa Monica, San Diego, and Harbor freeways. The project is 68 km (42 miles) in length, encompassing 56 freeway interchanges within its boundaries. The purpose is to find methods to reduce delay and accidents, to relieve motorist frustration, and to provide motorist services. Therefore, the central computer of this project provided real-time information about traffic from all the 68-km (42-mile) loop detectors. The real-time information, given on a continuous basis for a 24-h day, ran from 1972 to 1975 and includes

1. Five-minute roadway occupancy and volume measurements;
2. Five-minute information about weather (wet or dry) and light conditions (darkness, daylight, or fog); and
3. Reports of (a) the time the computer algorithm detected an accident, (b) location of this incident, (c) source of incident verification (service patrol, tow truck, helicopter, radio), and (d) type of incident (accident, disabled vehicle, gawking, slow truck, spilled load, fire, construction, maintenance).

From this extensive information the data considered in this research were selected.

Selection of Maximum Flow Data During Morning Peak Period

Among the most important characteristics of traffic flow models are the maximum flow characteristics. Data about these characteristics may reveal (a) the consistency of the time of the maximum observed flow, (b) the consistency of the maximum observed flow values, and (c) the consistency of the observed percentage of occupancy (when the maximum flow is observed).

Based on information available in Los Angeles, five criteria were chosen for selecting maximum flow characteristics. These criteria, which required that morning peak-period data be obtained before maximum flow was observed, consist of the following:

1. A base month during the spring season;
2. Normal demand pattern, i.e., no significant incident occurrence, excluding weekends and holidays;
3. Good weather conditions (dry pavement, no fog);
4. No operational difficulties for data recording;
5. Randomized-block design such that observations can be classified according to the independent variables of years, workdays, and locations.

With regard to criterion 2, most of the incidents can be divided into congestion-causing incidents (primarily accidents) and non-congestion-causing incidents (primarily disabled vehicles on the shoulder lane). The normal demand pattern is more likely to be obtained in situations without congestion-causing incidents or short-term non-congestion-causing incidents.

By inspecting the data collected from February 1972 to May 1975 and by using the above criteria, the month of March was selected as base month. This inspection also revealed considerable operational difficulties for data recording during 1974. All situations and criteria here mentioned limited the various possibilities for data selection (3).

The observed maximum flow data were classified as follows: (a) Data are obtained during the 10-d periods during March 1972, 1973, and 1975; (b) each 10 d of data represents five groups of 2 d of data from each workday (Monday through Friday); and (c) each of these 30 d of data includes the maximum flow characteristics for 14 locations along the Santa Monica Freeway (SM-12 to SM-25). Consequently, data for a total of 420 cells were collected on the Santa Monica Freeway. Each cell included three observed values averaged on a per lane basis: maximum flow rate in vehicles per hour (q'_m), corresponding percentage of occupancy (% occ), and corresponding time in the morning (T'_0). In addition, two estimated values, corresponding to q'_m, were considered: corresponding concentration (k'_0) and corresponding speed in kilometers per hour (u'_0). These estimations are based on two assumptions: that $k'_0 = 3$ (% occ), and that $u'_0 = q'_0/k'_0$. The first assumption, suggested by Athol (4), and the second assumption are both supported by LAAFSCP traffic engineers, who frequently introduced a test automobile into the traffic stream to directly measure average traffic speed.

Selection of Data for Analyzing the Consistency of Morning Congestion Patterns

The previous section is concerned with one specific data point (associated with the observed maximum flow); this section describes the sampling procedure for selecting morning peak-period data points, obtained from 5:00 to 11:00 a.m. on specific days. Criteria 2, 3, and 4 and data days from those selected in the previous section are applied. In addition, the data days were classified according to two independent variables: years and workdays.

By inspecting the data from 5:00 to 11:00 a.m. in terms of the above-mentioned criteria, 6 d of data were identified as follows:

1. March 14 (Tuesday), 1972
2. March 24 (Friday), 1972
3. March 27 (Tuesday), 1973
4. March 9 (Friday), 1973
5. March 4 (Tuesday), 1975
6. March 14 (Friday), 1975

Each of the six data days includes the morning volume-occupancy data points for each of the 14 roadway locations.

A statistical analysis shows that the sample size of 6 d of data is appropriate for analyzing the consistency of congestion patterns; for a given 14 locations, the number of days, n_i ($i = 1, 2, \ldots, 14$), such that the estimation error is less than 1.5 percent occupancy, using $\alpha = 0.05$, is determined to be six.

CONSISTENCY OF OBSERVED MAXIMUM FLOW CHARACTERISTICS

This section attempts to answer several questions. Is there a consistency of maximum flow characteristics regarding the differences among all considered years, between each pair of years, among all workdays, and between each pair of workdays? What is the magnitude of such consistency (or inconsistency)? What is the priority ranking of the three independent variables (locations, years, and workdays) regarding the consistency of the maximum flow characteristics? What are the average values and standard deviations of the maximum flow characteristics for the 14.5-km (9-mile) Santa Monica Freeway; and, based on the answers given for the previous questions, on what basis are they recommended for consideration (a yearly basis or a workday basis or both)?

The maximum flow characteristics are obviously consistent with locations (primarily because of different freeway geometry and demand pattern). However, it is the extent of such consistency regarding the other two independent variables, years and workdays, that needs statistical determination.

Means and Standard Deviations

The means and standard deviations of the five maximum flow characteristics mentioned earlier are indicated in Table 1 for the two independent variables, years and locations. The value in each cell is an average of 10 values (data from 10 d for each year). As mentioned previously, because of the energy crisis, the speed limit was reduced from 105 to 88 km/h (65 to 55 mph), and a different daylight savings time was instituted in 1974. While the speed changes may not affect the maximum flow characteristics observed at speeds below 80 km/h (50 mph), the change in daylight savings time had some effect on the light conditions during the 5 min of observed maximum flow. In 1972 and 1973, full light was observed before 6:45 a.m.; during 1975 full light was observed after 7:15 a.m. This could be one of the reasons why the mean q'_m values for 1975 were lower than those for 1972 and 1973.

In addition, the corresponding observed % occ was higher in 1975 than in 1972 and 1973. Differences in the average calculated speeds support this observation: In 1972 and 1973 u'_0 values were higher than in 1975. The results of the standard deviations, however, emphasize the differences in u'_0 values rather than the differences in q'_m values for the data obtained in less daylight (1975). One can conclude that (for a given demand pattern) in less daylight lower q'_m, higher % occ, and lower u'_0 values will be observed than in full daylight.

The average values and standard deviations of q'_m, % occ, k'_0, u'_0, and T'_0 for the two independent variables, workdays and locations, are given in Table 2. The value in each cell in this table is an average of six values (data for 2 d and 3 years). The means for location are the same as those given in Table 1.

Significant Differences Among and Within the Maximum Flow Characteristics

One of the criteria listed previously for maximum flow characteristics was that the data should be selected for a randomized-block design (criterion 5); thus, the 420 measurement cells of each characteristic would be classified according to the three independent variables, years, workdays, and locations. Consequently, the analysis of variance procedure can be applied to this randomized-block design. In addition, a multiple-pairwise-contrast method can be used to find out if there is a significant difference between each pair of independent variables.

Data generated by multivariable experiments can be analyzed best by a variance procedure using an analysis-of-variance (ANOVA) table. The null hypothesis, "there is no difference in treatment means," is tested by the F-statistic. Four ANOVA tables are combined, for convenience, in Table 3; they indicate the variations in the four maximum flow characteristics: observed q'_m, observed % occ, calculated u'_0, and T'_0. Table 3 shows how portions of these variations are assigned to each of the three independent variables, locations, years, and workdays, and to their interactions. As expected, there is significant difference between the 14 locations (SM-12 to SM-25) at the 0.01 level for all the maximum flow characteristics. The independent variable "years" (1972, 1973, and 1975) differs significantly, however, as does "workdays" at the 0.001 level for q'_m and % occ and at the

0.01 level for T_0'. There is, however, no significant difference with respect to u_0'. These statistical results are somewhat surprising, since it is difficult to detect any visual differences among the average values in Table 2. These results will be clarified by testing each pair of the two independent variables, years and workdays.

The two known methods for multiple contrasts are Scheffe's and Tukey's. Scheffe's method is appropriate for all multiple contrasts, but Tukey's multiple comparisons method, efficient for pairwise contrasts with equal sample size (5, 6), is the one used here. Table 4 summarizes results of these multiple pairwise contrasts for differences in mean values with respect to years. This table is arranged so that when differences in the mean value are significant (at the 0.01 level), the magnitude is indicated. Two examples of how to read the indicated numerical values are given at the bottom of Table 4. The most consistent results are associated with observed % occ and calculated u_0' between the years 1972 and 1973 and the corresponding time T_0' between

Table 1. Average maximum flow characteristics with respect to years and locations.

Location	q_m'	% occ	k_0'	u_0'	T_0'	q_m'	% occ	k_0'	u_0'	T_0'	q_m'	% occ	k_0'	u_0'	T_0'
	1972					1973					1975				
SM-12	2053	14.7	27.3	73.6	22.5	1993	14.5	27.0	74.9	11.0	2000	15.3	28.5	71.8	22.0
SM-13	2223	16.6	30.9	72.1	23.5	2113	15.8	29.4	72.3	15.0	1951	15.3	28.5	68.6	19.5
SM-14	2229	17.3	32.2	69.6	22.0	2189	15.6	29.0	75.5	10.5	2086	16.3	30.3	70.0	18.5
SM-15	2116	16.1	30.0	70.9	22.5	2060	16.8	31.2	66.5	12.0	2004	17.5	32.6	62.0	20.5
SM-16	2133	14.6	27.2	78.6	21.5	2087	14.7	24.9	83.7	9.0	1999	16.5	30.7	71.0	20.0
SM-17	2055	15.9	29.6	69.9	18.5	2011	15.2	28.3	71.8	7.0	1906	17.9	33.3	60.2	18.5
SM-18	1903	18.2	33.9	57.8	21.0	1835	18.2	33.9	55.9	13.5	1778	19.7	36.6	49.7	21.0
SM-19	2155	17.2	32.0	68.1	17.5	2097	16.9	31.4	67.3	9.0	1970	19.8	36.8	53.8	20.5
SM-20	2166	18.2	33.9	64.6	13.5	2136	18.0	33.5	63.9	10.0	1998	19.4	36.1	56.4	18.0
SM-21	1983	16.5	30.7	65.5	19.5	1955	15.6	29.0	67.5	13.5	1938	16.9	31.4	61.8	21.0
SM-22	1974	17.8	33.1	60.0	22.0	1934	16.9	31.4	61.7	14.0	1985	18.5	34.4	58.0	22.0
SM-23	2098	16.7	31.1	67.8	20.5	2060	17.9	33.3	62.8	11.5	2046	18.9	35.2	58.8	22.5
SM-24	2096	17.0	31.6	66.3	21.0	2055	16.4	30.5	67.6	7.5	1978	17.1	24.0	62.3	22.5
SM-25	2210	15.9	29.6	74.9	17.0	2166	16.9	31.4	69.6	7.0	2057	17.4	32.4	63.8	22.0
	Mean					Standard Deviation									
SM-12	2016	14.9	27.6	73.4	19.7	32.8	0.42	0.8	1.5	6.50					
SM-13	2096	15.9	29.6	70.8	19.3	136.8	0.65	1.2	2.1	8.25					
SM-14	2173	16.4	30.5	71.8	17.0	73.8	0.85	1.6	3.3	5.89					
SM-15	2060	16.8	31.2	66.7	18.3	56.0	0.70	1.3	4.3	5.57					
SM-16	2073	14.8	27.6	75.8	16.8	68.1	1.06	2.0	6.4	6.82					
SM-17	1992	16.1	29.9	67.1	14.7	76.5	1.40	2.6	6.2	6.63					
SM-18	1839	18.7	34.8	54.6	18.5	62.6	0.87	1.6	4.2	4.33					
SM-19	2074	18.0	33.4	63.1	15.7	94.6	1.59	3.0	8.1	5.96					
SM-20	2100	18.5	34.5	61.7	13.8	89.6	0.76	1.4	4.6	4.01					
SM-21	1959	16.3	30.4	64.9	18.0	22.7	0.66	1.2	2.9	3.97					
SM-22	1964	17.7	34.4	59.9	19.3	26.8	0.80	1.5	1.9	4.62					
SM-23	2068	17.8	35.2	63.1	18.2	26.9	1.10	2.0	4.5	5.86					
SM-24	2043	16.8	31.8	65.4	17.0	59.9	0.38	0.7	2.8	8.26					
SM-25	2144	16.7	24.6	69.4	15.8	78.7	0.76	1.4	5.6	7.64					

Notes: 1 km/h = 0.62 mph; 1 vehicle/km = 1.61 vehicles/mile.
q_m' is in vehicles per hour, k_0' is in vehicles per kilometer, u_0' is in kilometers per hour, and T_0' is time in minutes after 6:45 a.m.

Table 2. Average maximum flow characteristics with respect to workdays and locations.

Location	q_m'	% occ	k_0'	u_0'	T_0'	q_m'	% occ	k_0'	u_0'	T_0'	q_m'	% occ	k_0'	u_0'	T_0'
	Monday					Tuesday					Wednesday				
SM-12	2025	16.2	30.1	68.4	20.8	2031	14.3	26.7	74.5	20.0	2029	15.5	28.2	73.0	19.2
SM-13	2064	16.3	30.4	68.1	19.2	2118	16.0	29.8	71.0	21.7	2130	15.3	28.5	73.2	18.7
SM-14	2167	16.0	29.8	73.0	14.2	2212	17.7	32.9	68.2	18.3	2170	16.5	30.7	71.3	18.3
SM-15	2060	17.5	32.6	67.3	16.7	2061	16.7	31.0	66.5	18.3	2083	16.8	31.3	66.6	19.2
SM-16	2091	14.8	27.6	76.1	14.2	2072	14.5	27.0	76.7	20.0	2111	15.2	28.2	75.8	17.5
SM-17	1982	15.3	28.5	69.5	11.7	2022	16.3	30.4	67.4	14.2	2009	17.8	33.2	61.3	16.7
SM-18	1824	15.9	29.6	53.7	16.7	1840	22.7	42.2	44.2	22.5	1844	19.3	36.0	51.6	20.0
SM-19	2059	17.5	32.6	63.7	11.7	2112	19.5	36.3	59.5	15.8	2068	17.5	32.6	64.7	14.2
SM-20	2066	17.7	32.9	63.1	8.3	2091	19.8	36.9	57.8	15.0	2133	19.2	35.7	60.2	14.2
SM-21	1944	17.0	31.6	62.6	16.7	1961	16.3	30.4	65.2	16.7	2009	16.2	30.1	67.1	14.2
SM-22	1981	18.2	33.5	59.1	17.5	1959	18.2	33.8	58.2	21.7	1963	18.2	33.8	58.2	16.7
SM-23	1989	17.7	32.9	61.0	17.5	2074	16.7	31.0	67.3	18.3	2087	19.5	36.3	58.4	16.7
SM-24	2022	16.8	31.3	64.7	15.8	2064	16.5	30.7	67.1	15.8	2059	17.3	32.2	53.3	16.7
SM-25	2110	17.0	31.6	66.9	16.7	2172	16.7	31.0	70.3	19.2	2150	18.0	33.5	66.1	12.5
	Thursday					Friday					Standard Deviation				
SM-12	1980	14.5	27.0	73.5	20.8	2012	14.0	26.0	77.2	17.5	21.2	0.92	1.7	3.2	1.38
SM-13	2058	15.8	29.5	70.2	19.2	2109	15.7	29.1	72.2	17.5	32.7	0.37	0.7	2.0	1.53
SM-14	2153	15.7	29.1	74.0	16.7	2162	16.2	30.1	71.4	17.5	22.8	0.77	1.4	2.7	1.70
SM-15	2057	16.5	30.7	67.6	17.5	2039	17.3	32.2	63.7	19.2	15.6	0.42	0.8	1.8	1.09
SM-16	2093	15.0	27.9	76.3	15.8	1998	14.7	27.3	75.5	16.7	44.1	0.27	0.5	3.5	2.15
SM-17	2004	14.8	27.6	72.7	14.2	1942	16.2	30.1	65.5	16.7	31.3	1.15	2.1	3.4	2.09
SM-18	1860	16.2	30.1	62.1	15.0	1824	16.7	31.0	60.5	18.3	15.1	2.87	5.3	4.7	2.90
SM-19	2108	18.3	34.1	63.4	16.7	2023	17.3	32.2	63.6	20.0	36.9	0.91	1.7	4.9	3.07
SM-20	2141	17.7	32.9	65.5	15.0	2089	18.3	34.1	61.3	16.7	31.8	0.93	1.7	5.2	3.22
SM-21	1955	16.7	31.0	62.3	22.5	1924	15.5	28.8	66.6	20.0	31.5	0.57	1.1	5.2	3.24
SM-22	1972	16.8	31.3	62.8	20.8	1958	17.2	31.9	60.8	21.7	9.8	0.67	1.2	3.9	2.40
SM-23	2067	17.8	33.2	62.9	21.7	2028	16.8	31.3	65.6	17.5	40.1	1.12	2.1	3.2	1.96
SM-24	2051	17.0	31.6	65.3	17.5	2017	16.5	30.7	66.1	19.2	21.7	0.34	0.6	2.3	1.42
SM-25	2171	15.8	29.5	74.2	11.7	2118	16.5	30.7	69.5	18.3	29.1	0.80	1.5	5.5	3.40

Notes: 1 km/h = 0.62 mph; 1 vehicle/km = 1.61 vehicles/mile.
q_m' is in vehicles per hour, k_0' is in vehicles per kilometer, u_0' is in kilometers per hour, and T_0' is time in minutes after 6:45 a.m.

Table 3. Results of the analysis of variances for the maximum flow characteristics.

Source	d.f.	Observed q_o' MS	Observed q_o' F	Observed % occ MS	Observed % occ F	Calculated u_o' MS	Calculated u_o' F	T_o' MS	T_o' F
Location	13	218 025	49.88[a]	43.69	12.27[a]	57.20	2.98[a]	96.43	2.90[a]
Years	2	508 108	116.25[a]	59.66	16.76[a]	858.16	44.67[a]	4030.23	121.14[a]
Workdays	4	28 771	6.58[a]	19.66	5.52[a]	61.13	3.18	118.84	3.57[b]
Location × years	26	14 667	3.36[a]	4.76	1.34	32.04	1.67	37.16	1.12
Location × workdays	52	2 940	<1.0	4.86	1.36	20.20	1.05	27.71	<1.0
Years × workdays	8	10 489	2.40	2.39	<1.0	9.31	<1.0	52.41	1.57
Location × years × workdays	104	2 280	<1.0	2.30	<1.0	11.10	<1.0	25.04	<1.0

Note: The significant differences are the same for % occ as for estimated k_o' because of the linear transformation between % occ and k_o'.
[a] Significant at $p < 0.001$. [b] Significant at $p < 0.01$

Table 4. Effect of years on the consistency of maximum flow characteristics (overall analysis of the 14 locations).

Variable	Year	1973	1975
Observed q_o', vehicles/h	1972	51	120
	1973	—	69
Observed % occ	1972	NS	-0.94
	1973	—	-1.27
Calculated u_o', km/h	1972	NS	6.82
	1973	—	6.98
T_o', min	1972	9.07	NS
	1973	—	-9.5

Notes: 1 km/h = 0.62 mph.
Value in cell indicates magnitude of difference between years at 0.01 level.

Table 5. Effect of workdays on the consistency of maximum flow characteristics (overall analysis of the 14 locations).

Variable	Weekday	Tues.	Wed.	Thur.	Fri.
Observed q_o', vehicles/h	Mon.	NS	NS	NS	NS
	Tues.	NS	NS	NS	40
	Wed.	—	—	NS	45
	Thur.	—	—	—	NS
Observed % occ	Mon.	NS	NS	NS	NS
	Tues.	NS	NS	1.00	0.95
	Wed.	—	—	0.99	0.93
	Thur.	—	—	—	NS
Calculated u_o', km/h	Mon.	NS	NS	NS	NS
	Tues.	NS	NS	-3.16	NS
	Wed.	—	—	NS	NS
	Thur.	—	—	—	NS
T_o', min	Mon.	-2.9	NS	NS	-2.8
	Tues.	NS	NS	NS	NS
	Wed.	—	—	NS	NS
	Thur.	—	—	—	NS

Notes: 1 km/h = 0.62 mph.
Value in cell indicates magnitude of difference between days at 0.01 level.

the years 1972 and 1975. Nevertheless, all the other results, although indicated by numerical values (significant differences exist), seem to have relatively low magnitudes, which for practical purposes are generally negligible (particularly between the years 1972 and 1973 when the overall conditions were alike). The maximum differences indicate that the effect of years on the maximum flow characteristic is marginal and that, if neither the 88-km/h (55-mph) speed limit nor the different daylight savings time is changed and if the same demand pattern and freeway geometry is retained, maximum flow characteristics observed in future years during the month of March can be expected to be similar to those observed during March 1975.

The results of the multiple pairwise contrasts for workdays, using Tukey's method, are summarized in Table 5. (This table is similar to Table 4 in that the magnitudes of the differences at the 0.01 level are indicated.) There is a statistically significant difference, particularly of mean q_m' and mean % occ values between Tuesday and Friday and between Wednesday and Friday. Surprisingly, both the mean q_m' and mean % occ values are smaller on Friday than on either Tuesday or Wednesday, although logically one would anticipate lower q_m' values when higher % occ values are observed (see differences in q_m' and % occ values between 1972 and 1975 and between 1973 and 1975 in Table 1). There is, then, almost no effect of workdays on the consistency of maximum flow characteristics; therefore, measurements from one workday can be applied to others.

CONSISTENCY OF MORNING TIME-SEQUENCE PATTERNS, FLOW-CONCENTRATION TRAJECTORIES, AND MORNING CONGESTION PATTERNS

This section considers the following questions. Is there a consistency in time-sequence patterns that illustrates flow-concentration trajectories? Is there consistency in morning congestion patterns, particularly regarding the obtained concentration contour maps? Are there any effects from reducing the speed limit from 105 to 88 km/h (65 to 55 mph) or from a different daylight savings time or from other energy crisis factors (perhaps, movements toward smaller automobiles and car pools), where these three possibilities are reflected in the 1975 data on both the morning time-sequence flow-concentration trajectories and the morning congestion patterns?

The answers developed below seem to be important to the analysis of traffic flow models, consistency of the demand pattern, and sensitivity of some traffic flow characteristics because of changes in overall conditions (speed limit, daylight conditions).

Consistency of Morning Time-Sequence Patterns and Flow-Concentration Trajectories

Based on the sampling procedure described earlier, 6 d of data were identified for the imposed criteria and a statistical analysis. Analyses of the traffic flow models performed in the overall study of flow models (3) reveal that five phases can be distinguished in the morning flow-concentration (q - k) trajectories observed on the Santa Monica Freeway. These phases can be summarized as follows and are shown in the examples in Figure 1:

1. The traffic stream proceeds from low concentration conditions to the "intersection concentration";
2. The traffic stream approaches the maximum flow value;
3. A breakpoint (discontinuity) occurs and results in a transition from free-flow to congested-flow conditions;
4. The traffic stream fluctuates under high concentration conditions; and
5. The traffic stream approaches free-flow conditions at the intersection concentration through a flow value lower than that indicated in phase 2.

The observed time-sequence pattern in which the q - k trajectories are portrayed is generally not influenced by

Figure 1. Comparison of morning flow-concentration trajectories in 1972 and 1975 at two locations.

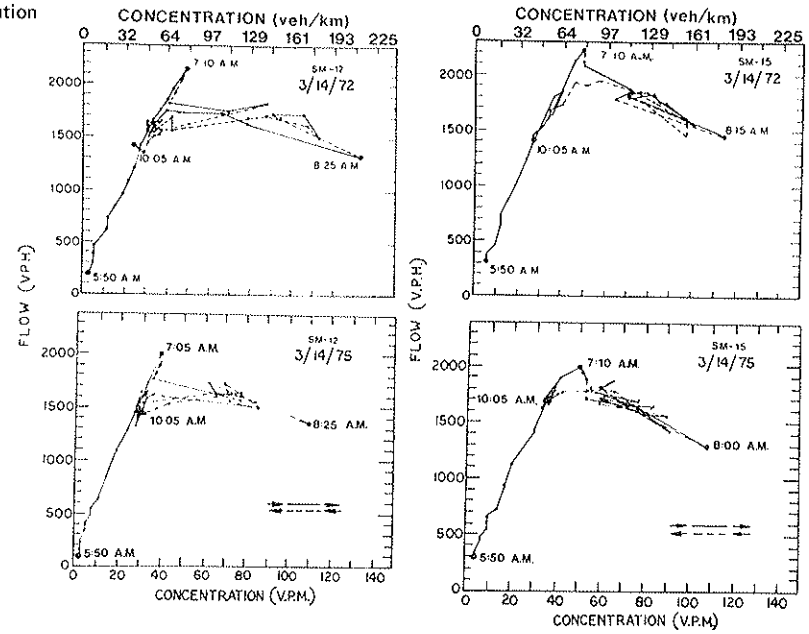
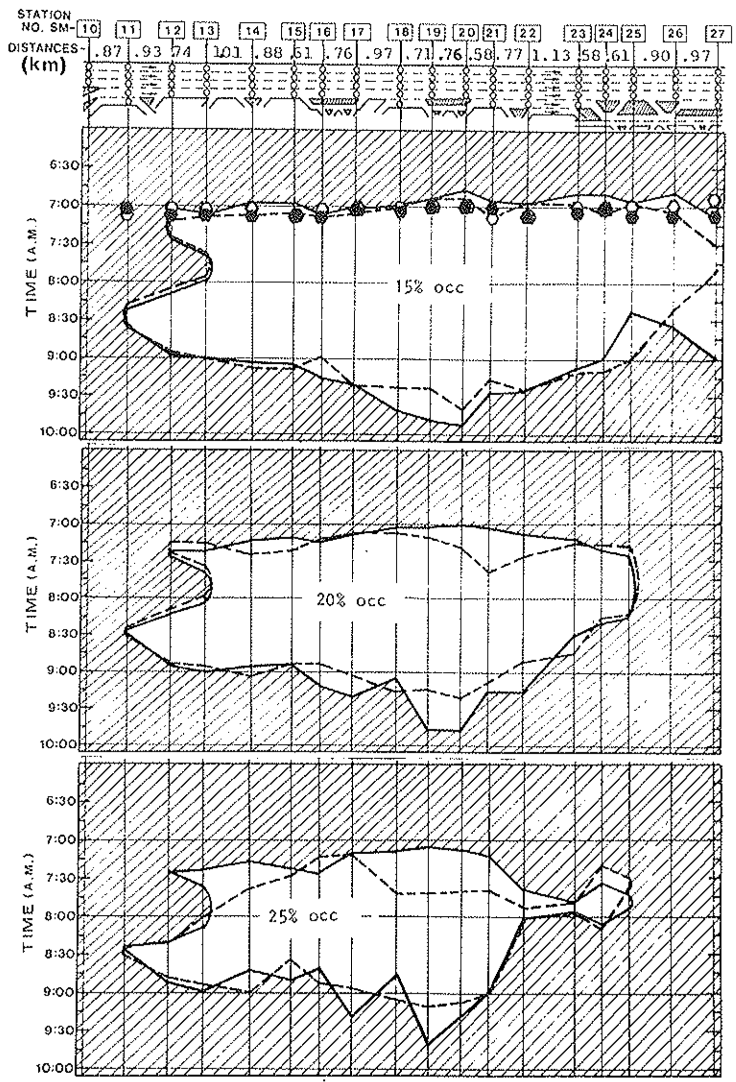


Figure 2. Comparison of three percentages of occupancy in 1972 and 1975.



—— 1972; ● -the corresponding time for the maximum observed flow in 1972.

---- 1975; ○ -the corresponding time for the maximum observed flow in 1975.

Note: 1 km = 0.62 mile

either years or workdays. To visually demonstrate the significant consistency of the dynamic q - k trajectories, a specific data day was selected; the day most unlike Tuesday, March 14, 1972, is Friday, March 14, 1975, because it combines dissimilarities in overall conditions (speed limit, daylight savings time, and other energy crisis factors) and in workdays.

The q - k trajectories of the data collected in 1972 and 1975 on March 14 are shown for 11 detector stations (SM-12 to SM-22) in Figure 1, which illustrates how well the consistency of the time-sequence pattern is preserved during a 3-year period.

Although the time-sequence pattern of the q - k trajectories is preserved, the flow magnitudes observed in 1975 are significantly different from those of 1972. A comparison suggests that, from the time the maximum flow was observed until the traffic stream approached free-flow conditions, the q - k trajectories of 1975 shifted toward flow values approximately 150 vehicles/lane·h lower than those of the 1972 trajectories. These differences are particularly evident between approximately k = 19 and k = 37 vehicles/km (30 to 60 vehicles/mile) (or between 10 and 20 percent occupancy), and the highest differences in the flow values are generally associated with q_m' values. For example, the differences between the q_m' values of the March 14, 1972, data and those of the March 14, 1975, data are 156, 372, 264, 153, 252, 264, 192, 270, 198, 108, 0 vehicles/lane·h for locations SM-12 to SM-22 respectively.

Certainly, one would expect to observe lower flow values in less daylight; it is, however, surprising that the time-sequence pattern of the q - k trajectories is preserved so well, probably the result of unchanged demand pattern and existing bottlenecks.

Consistency of Morning Congestion Patterns

The congestion patterns can be best represented by concentration contours or by occupancy contours. The 6-d data were compared primarily by occupancy contours at various congestion levels, and data of Tuesday, March 14, 1972, are compared in Figure 2 with those of Friday, March 14, 1975, at three congestion levels: 15, 20, and 25 percent occupancy [about 28, 37, and 47 vehicles/lane·km (45, 60, and 75 vehicles/lane·mile) respectively]. Occupancy contour maps that compare

Figure 3. Means of observed maximum flow rate, percentage of occupancy, and corresponding time in 1972, 1973, and 1975 at 14 locations.

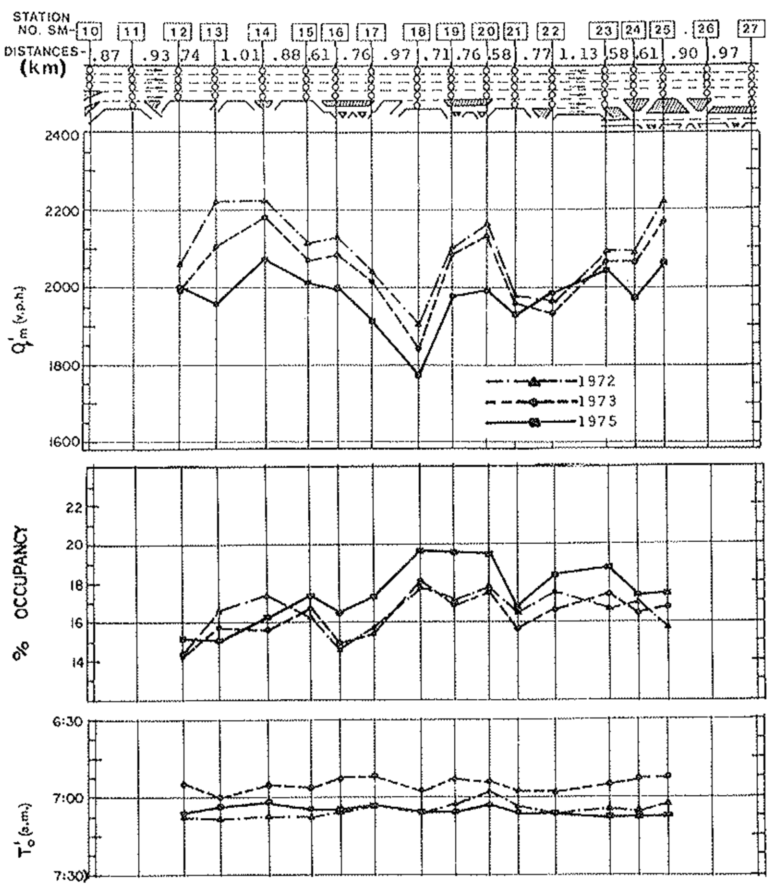


Table 6. Grand means and standard deviations of the maximum flow characteristics.

Year	Item	q_o' (vehicles/h)	% occ	Estimated k_o' (vehicles/km)	Calculated u_o' (km/h)	T_o'
1972, 1973	Mean	2075	16.45	30.6	68.5	7:00 a.m.
	Standard deviation	97.4	1.16	2.1	6.2	2.7 min
1975	Mean	1980	17.60	32.7	61.6	7:05 a.m.
	Standard deviation	74.5	1.50	2.8	6.6	1.5 min

Note: 1 km/h = 0.62 mph; 1 vehicle/km = 1.61 vehicles/mile.

the remaining 5-d data are shown elsewhere (3). In addition, Figure 2 indicates the corresponding time (T'_0) for the observed maximum flow. The consistency of the morning congestion patterns was clearly stable in the 3-year period (1972 to 1975). Figure 2 further emphasizes that these patterns were apparently unaffected by changes in the overall conditions.

To quantify this observed consistency, an analysis was made of the differences between the beginning and ending times of congestion and the maximum observed occupancy characteristics (3). This analysis strongly supports the observed consistency, which was lack of influence by either of the two independent variables, years or workdays.

SUMMARY AND CONCLUSIONS

Consistency of Maximum Flow Characteristics

Based on the statistical results indicated in Tables 1, 2, and 3, the most inconsistent independent variables are clearly locations and years. A graphic representation of the data in Table 1 is, therefore, given in Figure 3 for the observed variables q'_m, % occ, and T'_0. There is, as expected, relatively high inconsistency regarding location because of detector-station positioning and the demand pattern along the Santa Monica Freeway. For example, location SM-18 is characterized by a much lower q'_m value and a higher % occ value than those at other locations, perhaps because SM-18 is on a five-lane rather than on a four-lane section, and consequently there is a possible weaving effect. In addition, the effect of different daylight savings time can be seen in Figure 3 for the 1975 solid line. The most significant findings are listed below.

1. From a practical viewpoint, the effect of years on the consistency of the maximum flow characteristics is only marginal. If, however, one considers only 1972 and 1973 data, when no changes occurred in the overall conditions, very high consistency is observed. Statistically, nevertheless, there are significant differences among all the characteristics regarding years, apparently caused by the changes in the daylight conditions. It is therefore hypothesized that similar observed maximum data will be found in the near future if no changes occur in the overall conditions, and that, if the time of sunrise and resulting conditions are the same as those of 1972 and 1973, the observed maximum flow data will be similar (on the average) to those observed during 1972 and 1973.
2. Workday variations had almost no effect on the consistency of the maximum flow characteristics, and therefore, measurements from one workday can be applied to others.
3. Location was the most crucial independent variable causing inconsistency in the observed maximum flow characteristics (the second most crucial was years and the least important was workdays).
4. Grand mean and standard deviation values of the maximum flow characteristics for all 14 locations are listed in Table 6. It should be noted, however, that each detector location has its own demand and geometry characteristics, and therefore Table 6 cannot represent locations. Nonetheless, and for practical purposes, one can compare Table 6 and the values in the Highway Capacity Manual (7) (particularly q'_m values with the known 2000 vehicles/lane·h).

Consistency of Morning Congestion Patterns

From the analysis of the time-sequence patterns that illustrate q - k trajectories and morning congestion patterns, the following conclusions can be drawn.

1. The time-sequence pattern of the q - k trajectories is well preserved during the 3-year period 1972 to 1975 (Figure 1). This pattern emphasizes that the traffic stream first approaches maximum flow conditions, then congested conditions, and then, through a recovery process, backward to free-flow conditions.
2. There is high consistency in the morning congestion patterns during the 3-year period (Figure 2). This consistency is reflected by similar occupancy measurements (both in magnitude and duration) under morning peak-period conditions.
3. Changes in overall conditions [88 km/h (55-mph), speed limit, daylight savings time, and other energy crisis factors] have little apparent effect both on the time-sequence pattern in which the q - k trajectories are portrayed and on the morning congestion patterns. The exception to this finding concerns flow magnitudes, particularly under maximum flow conditions where the q'_m value was observed to be about 100 vehicles/lane·h less in 1975 than in 1972 and 1973.

The above findings and conclusions may be useful when applied to problems of desirable flow and occupancy (concentration) under high demand level, estimated freeway capacities, sensitivity of an estimated capacity (capacity buffer), and traffic behavior during the peak period.

ACKNOWLEDGMENTS

We wish to express our appreciation to District 7, California Department of Transportation, Los Angeles, for kindly supplying the real-world data.

REFERENCES

1. A. Ceder and A. D. May. Further Evaluation of Single- and Two-Regime Traffic Flow Models. TRB, Transportation Research Record 567, 1976, pp. 1-15.
2. A. Ceder. A Deterministic Traffic Flow Model for the Two-Regime Approach. TRB, Transportation Research Record 567, 1976, pp. 16-30.
3. A. Ceder. Investigation of Two-Regime Traffic Flow Models at the Micro- and Macroscopic Levels. Univ. of California, Berkeley, PhD dissertation, Nov. 1975.
4. P. Athol. Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream. HRB, Highway Research Record 72, 1965, pp. 58-87.
5. L. A. Marascuito. Statistical Methods for Behavioral Science Research. McGraw-Hill, New York, 1971.
6. K. A. Brownlee. Statistical Theory and Methodology in Science and Engineering. Wiley, New York, 1965.
7. Highway Capacity Manual. HRB, Special Rept. 87, 1965.

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

Procedure for Estimating Demand for Regional Fringe Parking Facilities

R. K. Mufti, L. S. Golfin, and C. D. Dougherty, Delaware Valley Regional Planning Commission, Philadelphia

The purpose of this study was to determine the best location and the optimum feasible quantity of additional parking spaces that would effectively serve potential demand for change-of-mode parking at the interface between highway and passenger rail systems. Selection criteria, such as available land, accessibility to highway system, current rail ridership, and current parking demand, were used to identify 20 potential fringe parking sites. Future demand for parking spaces at the selected sites was determined in four steps. The first step dealt with trip interchanges. All future trip makers who reside in the influence area of each of the potential sites, and whose trip destinations lie in the distribution service area of the passenger rail system, were identified and quantified. In the second step, the market share of each mode was calculated by using a disutility mode-choice model. Disutility rates for the automobile and rail modes were computed for each of the trip origin areas, and the percentage of passenger rail trips was derived from diversion curves. In the third step, the proportion of projected commuter rail patrons demanding parking spaces at each site was established by using a relationship between the distances patrons travel to the station and their access modes to the station. Finally, additional parking spaces over and above the number of spaces already existing or planned were calculated for each site.

Recent federal-aid highway acts provided for the use of Highway Trust Fund monies for the construction of regional fringe parking facilities at the interfaces between major highway routes and commuter railroad and transit lines. In February 1974 the Pennsylvania Department of Transportation (PennDOT) authorized the Delaware Valley Regional Planning Commission (DVRPC) to proceed with a study seeking the best location and optimum feasible quantity of additional parking to effectively serve future demand for change-of-mode parking at the interface between the highway and passenger rail systems within the five-county Pennsylvania portion of the Delaware Valley region. The underlying regional goals of this study were

1. To reduce highway congestion, particularly during the peak periods and in the region core;
2. To reduce projected demand for Philadelphia central business district (CBD) parking space and thus free land and airspace for more productive uses;
3. To provide incentives to attract trip makers to more efficient modes; and
4. To reduce air pollution levels in the CBD.

The four-phase study performed by the DVRPC encompassed site selection and interagency coordination, development of demand estimation methodology, analysis of demand estimates, and community impact analysis. Although all four phases are necessary to move regional fringe parking into the design and implementation phase, the intention of this paper is to show how a regional planning agency might respond to a request to provide design data for a project not normally considered in the long-range urban transportation planning process. Therefore, we have dealt with only the two phases concerning demand estimation.

SITE SELECTION, INTERAGENCY COORDINATION, AND COMMUNITY IMPACT

During the fall of 1973, DVRPC, in association with PennDOT, coordinated a multiagency task force that included representatives of county planning commissions and other concerned agencies. The task force was charged with the review and selection of candidate sites for a regional fringe parking program. The candidate sites would then be subjected to more detailed analyses under each of the study phases.

These agencies cooperated to select 20 potential regional fringe parking sites. These high-priority sites were selected on the basis of available land, compatibility of parking with adjacent land uses, placement within a high-density travel corridor, accessibility to the highway system, and minimization of disruptive impact on the local community. The full list of criteria against which the recommended sites were reviewed is given in Table 1. Recommended sites were not necessarily restricted to existing rail stations, and the recommendation to construct a new station or consolidate a number of stations was considered within the realm of the study.

After the demand estimation process was completed for each of the 20 sites, a preliminary impact analysis based on existing conditions was conducted regarding land use and community development, illegal street and off-street parking, construction or upgrading of access roads, and alleviation of traffic congestion on major highway facilities.

The future impact of additional peak-period traffic on the local access roads to the commuter stations was determined by a forecast that was made of the average annual daily traffic (AADT) on those roads in 1985 and that used growth factors based on trends and future land-use information. The additional parking space demand was equated with additional peak-period vehicles and was added to peak-period traffic volume. The sums represented the total future peak-period vehicle trips on the access roads. Finally, comparison was made with the access roadway capacities (vehicles per lane hour for level of service E), as developed by DVRPC staff, in order to calculate the volume to capacity (v/c) ratios used to determine the impact of the additional peak-period traffic.

DEVELOPMENT OF DEMAND ESTIMATION METHODOLOGY

Four tasks and procedures were required to establish the quantity of additional parking needed on the basis of future demand for change-of-mode parking at each of the selected sites.

Task 1: Relevant Trip Interchanges

We identified and quantified all future trip makers residing in the influence area of each preliminary site that has trip destinations in the distribution service areas of the passenger rail system. Task 1 was subdivided into three parts: (a) delineation of the area of trip origin, (b) delineation of the area of trip destination, and (c) tabulation of the number of trip makers wishing to travel between origin and destination areas for given years in the future.

Origin Area Delineation Procedure

The origin area for each station site was defined as the geographic area in which the patrons of the station reside. For purposes of forecasting future patrons, it was first necessary to delineate the potential future market area of the station. This future market area included the present influence area of the station plus an additional area that would be influenced by the increase in station access opportunities, which is a manifestation of the increased parking supply that would permit more potential patrons to enter the passenger rail system. In enlarging the influence area, expansion should logically occur along highway corridors and into residential areas accessible to these highways.

We chose two potential area sizes: a maximum and a minimum, both based on the core market area as defined for sites at existing station points in SEPACT II (Southeastern Pennsylvania Transportation Compact operations plan for 1975, which included market surveys and an analysis of 1966 operations of the commuter railroads serving metropolitan Philadelphia). This existing core was the area in which 67 percent of the station's patrons resided. The perimeter of this core area was expanded along highway routes that fed into the station and could be used by potential park-and-ride station patrons. Judgment was applied to this expansion process to account for how far (in terms of time and distance) people might actually drive before changing modes and the degree to which they would be willing to back travel (drive to the station in a direction opposite to that of their destinations). The maximum area assumed considerable access and back-traveling distances. The minimum area assumed distances marginally greater than those for the core area. Finally, the expanded perimeter was made to conform with the boundaries of the DVRPC transportation analysis zone.

For new station sites near existing stations, the core areas were merged and the above process continued. For new station sites in far outlying areas not covered by the SEPACT II market analysis, the maximum and minimum areas were determined by assuming a small core and using the expansion process as before.

In this procedure, each site was analyzed independently; that is, the maximum and minimum areas of any one site were not affected by the influence area of any other site. This independent analysis procedure permitted study of each site on its own merits and aided in determining the priority of each site.

Destination Area Delineation Procedure

Analysis of available data revealed that the vast majority of passenger rail trips are bound for the core area of Philadelphia. We decided to limit our study destinations to this city core area in order to make the demand estimation of future rail ridership systematic. Here, also, two sizes of destination areas were selected: The maximum area included all 46 CBD zones and 3 zones from the University City area; the minimum area excluded 11 of those zones that lie along the Delaware waterfront, in the southwest CBD residential area, and in other areas either without an employment base or poorly accessible to the city rail stations (Thirtieth Street, Penn Center, or Reading Terminal).

Since this procedure directly considered only those destinations in the city core, it was necessary to adjust the rail patronage projections to account for rail trips to all other destinations. This adjustment procedure is discussed under task 2.

Tabulation of Trip Interchange Volume Procedure

Once the areas of origin and destination had been defined for each site, the person-trip interchange data from existing DVRPC trip tables were compiled into a travel demand matrix for each combination of maximum and minimum sizes of trip-end areas. These travel demand matrices were then scaled to the project analysis years of 1976, 1980, and 1985 relative to projected trends of the primary transportation variables and actual trends of ground count and passenger ridership data.

Task 2: Modal and Submodal Split

The purpose of task 2 was to determine the proportion of the trip makers who were identified and quantified in task 1 and who would be likely to choose passenger rail as their primary mode (given certain specific assumptions about mode-choice behavior and transportation system attributes). This task was composed of two parts: (a) calibrating the model and (b) assembling model input and calculating passenger rail patronage.

Model Determination

A utilitarian mode-choice model was used to find the proportion of total trip makers on an interchange likely to use passenger rail. The basic formulation of this model was a set of stratified diversion curves relating the percentage of transit trips for any interchange of a given strata to the cost difference of travel by the transit mode and the private automobile mode. Cost in the model was defined for mode X as

$$\begin{aligned}\text{Cost (mode X)} = {} & K_1 \text{ (excess time mode X)} \\ & + K_2 \text{ (running time mode X)} \\ & + \{[\text{monetary cost (mode X)}] \\ & \div [K_3 \text{ (median income of trip makers)}]\} \qquad (1)\end{aligned}$$

K_1, K_2, and K_3 are calibration constants; excess time is out-of-vehicle time; running time is in-vehicle time; monetary cost is any fare, parking charges, tolls, or the like associated with the one-way trip; and median income is median total family income of the aggregated zones of residence in the origin area.

Calibration of the model with 1960 DVRPC survey data yielded the following equations:

$$\begin{aligned}\text{Transit cost} = {} & 1.67 \text{ (transit run time)} + 2.5 \text{ (excess time)} \\ & + \{[\text{fare} + (1\tfrac{1}{2}) \text{ parking charge}] \\ & \div (0.25 \text{ median income})\} \qquad (2)\end{aligned}$$

$$\begin{aligned}\text{Highway cost} = {} & 1.67 \text{ (highway run time)} + 2.5 \text{ (excess time)} \\ & + \{[(\text{cost/mile}) \text{ mileage} + (\tfrac{1}{2}) \text{ CBD parking charge}] \\ & \div (0.25 \text{ median income})\} \qquad (3)\end{aligned}$$

The cost difference or utile rate (U) of the competing modes was then defined as

$$U = \text{(cost transit mode)} - \text{(cost highway mode)} + 200 \qquad (4)$$

The diversion curves were stratified by area type of origin and destination, trip purpose, and principal transit submode. The diversion curve used in this analysis was stratified by origins and destinations in suburban, rural, and open rural areas to CBD areas; home-based work trip purpose; and passenger rail submode.

Table 1. Criteria for selecting potential fringe parking sites.

Selection Criteria	Factors
Geographic location	County, township or borough, land use surrounding the site, rail or transit line, distance from CBD (rail), distance from CBD (highway)
Relationship to adjacent highway	Highway adjacent to site, functional classification and funding, status of adjacent highways, traffic volume and existing volume/capacity ratios, projected volume and future volume/capacity ratios
Physical characteristics	Existing ridership and parking spaces available, availability of land for expansion and preliminary cost estimate for land acquisition, parking lot utilization, parking on adjacent streets, present use of land proposed for fringe parking lot
Travel time	Travel time by rail or transit to CBD, travel time by automobile to CBD, multimode travel time to CBD
Travel cost	Transit fare to CBD, total cost of rail trip (including parking, cost of first mode, and personal time worth), total cost of highway trip (including parking and personal time worth)
Access	Adequacy of highway access to site, relationship of parking site to pedestrian, type of intersection (at-grade, grade separated, signalized)
Relationship to community	Traffic flow on local streets from highway to parking site, compatibility with existing or proposed land uses
Rail line adequacy	Frequency of service, potential for increased service, type of cars and potential improvements, consideration of potential new station stop

Table 2. Ranges considered in input variables for determining rail patronage.

Input Variable	Variable Level High	Medium	Low
Transit parking cost, $	Free	0.16	0.25[a]
Destination area excess time, min	A[b]	A + 1.0	A + 2.0
Highway excess time, min	5	4.5	4.0
Highway running time, min	B[c]	0.95 B[c]	0.90 B[d]
Highway out-of-pocket cost per kilometer, $	0.034	0.027	0.023
CBD parking cost, $	2.40	2.10	1.80

Note: 1 km = 0.62 mile.
[a] All pay spaces.
[b] Original calculation A: weighted average time for egress from station in city core and walking to destination.
[c] Original calculation B: time for vehicle to go from origin to destination area.
[d] Considers higher speeds.

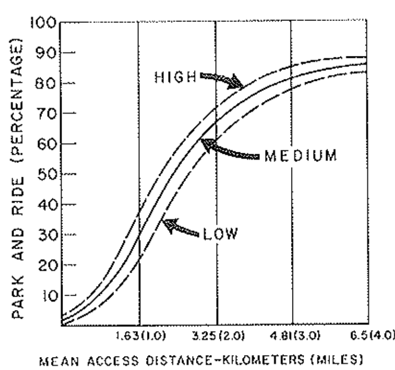


Figure 1. Park-and-ride percentage ranges as a function of origin area mean access distance to station.

Model Input and Passenger Rail Patronage

Excess time for passenger rail consisted of station entrance and waiting time (waiting time is equal to half the headway) but not exceeding 7.5 min plus egress and time spent walking to the destination. Excess time for highway was the average of the time spent parking and retrieving the automobile from a CBD parking space plus time spent walking to the destination. Entrance, egress, walking, and parking times were based on past experience and calculated by the staff.

Running time for passenger rail was obtained from published schedules. Running time for the highway was based on probable route selection and DVRPC data on speeds. The freeway network considered in route selection was the portion of the DVRPC freeway network expected to be completed by each of the project analysis years.

Monetary costs for transit included 1970 fares plus half of any station parking cost, and for the highway the average out-of-pocket costs per vehicle kilometer plus half the average CBD parking charge.

Median income for the origin area was based on the 1970 census data aggregated from the analysis zone to origin area level.

The analysis considered different levels of the input variables so as to provide the decision maker with a range of probable values. Table 2 presents three levels of input variables considered in the analysis. The combined effect of the changes made for the individual variables on the mode-choice model results in the range of percentages for each site.

Once the model inputs for a site had been assembled, the utile rate was calculated, and the percentage of transit trips was derived from the diversion curve. This percentage was then applied to the total number of persons making the trip interchange for each of the study years to determine projected rail patronage.

Because task 1, trip interchanges, considered only destinations in the city core, it was necessary to adjust the rail patronage projections calculated in task 2 to account for all probable destinations by rail from the trip origin area of the site. Our procedure was to multiply the core rail trips derived from the mode-choice process of task 2 by the ratio of total rail trip destinations to core area rail trip destinations of existing rail patrons. This ratio is 100 to 86 and has held relatively constant since 1965. This final value is the total rail trip demand for the trip origin area for each study year.

Task 3: Park-and-Ride Estimation

The purpose of task 3 was to establish, from task 2, the proportion of projected passenger rail patrons who will demand parking spaces at each site.

It was hypothesized that there is a relationship between the distances patrons travel to the station and their access modes to the station. This relationship was approximated by a plot of the mean radius of the core market area of each station with park-and-ride facilities against the percentage of rail patrons who park and ride at the

Table 3. Daily person trips between origin area and city core area.

Site Location	1976 Max	Avg	Min	1980 Max	Avg	Min	1985 Max	Avg	Min
Bensalem	3702	3261	2809	4141	3647	3153	5381	4746	4112
Baldwin-Crum Lynne	3708	3349	2990	3818	3448	3078	4057	3665	3273
Radnor	5019	4429	3746	5223	4610	3997	5678	5010	4342
Fort Washington	3583	3167	2752	3806	3364	2923	4320	3819	3318

stations. This plot is shown in Figure 1. The mean radius of the trip origin area of the site was then determined on the basis of the weighted average of the distances the zone centroids are from the site station. This was then used to enter the curve and derive the percentage of total patrons who will park and ride. The curve was structured to provide a range of percentages for each site. That percentage was used to determine the number of park-and-ride patrons. The number of those patrons was then divided by the average automobile occupancy rate to obtain the number of park-and-ride vehicles and thus the number of spaces demanded.

Task 4: Calculation of Parking Needs

The purpose of task 4 was to determine the number of parking spaces, over and above those already existing or planned under other programs, that will be needed to meet projected demand. The three parts of this task are (a) tabulation of all existing and proposed parking spaces for stations serving the trip origin area of the site, (b) allocation of these spaces to the trip origin area of the site, and (c) calculation of additional space needs.

Tabulation of Existing and Proposed Spaces

For each site, a listing was made of all stations whose market areas overlap the trip origin area of the site for both maximum and minimum area levels. The number of existing and planned parking spaces for each of these stations was tabulated. Proposed additional spaces were determined from an application by the Southeastern Pennsylvania Transportation Authority for a grant to improve passenger rail stations and from information solicited from county planning commissions.

Table 4. Percentage of daily person trips by rail with and without center city commuter connection.

Site Location	High		Medium		Low	
	With	Without	With	Without	With	Without
Bensalem	75.0	72.5	53.0	50.5	41.0	38.5
Baldwin-Crum Lynne	83.0	83.0	81.5	80.0	73.5	70.0
Radnor	78.0	77.5	64.0	60.5	53.0	48.5
Fort Washington	83.0	82.0	78.5	75.5	58.5	53.5

Table 5. Percentage of rail patrons who will park and ride as a function of mean access distance to station.

Site Location	Access Distance (km)	Rail Patrons (%)		
		High	Medium	Low
Bensalem	8.13	90.0	89.0	85.0
Baldwin-Crum Lynne	3.25	71.0	66.0	60.5
Radnor	5.41	86.5	83.5	80.5
Fort Washington	4.06	80.5	75.0	70.5

Note: 1 km = 0.62 mile.

Allocation of Parking Spaces to Site Area

Each site area was given a number of the existing and proposed parking spaces of the listed stations. This number was calculated according to the proportion of the market area of the listed station that overlapped the site area.

Calculation of Additional Need

Task 3 determined the future demand for park-and-ride spaces within the trip origin area of the site, and task 4 determined existing and proposed supply within that area. The difference between the projected demand (task 3) and the actual supply (task 4) was the additional parking supply required.

ANALYSIS OF DEMAND ESTIMATES

Interchange Volumes: Task 1 Results

Table 3 presents a sample of the output of task 1 by year for three levels of analysis. The figures represent the number of persons who will travel by all modes between a given trip origin area of a site station to the core area of the city of Philadelphia.

The maximum and minimum trip data were developed by tabulating all person trips from the trip origin area to the destination area at the maximum and minimum levels respectively. The average trip data are simply the average of the maximum and minimum levels.

Passenger Rail Trips: Task 2 Results

Table 4 presents a sample of the output of task 2 for the three levels of input variables. The table further presents the impact of the proposed center city commuter rail connection (CCCC) on the mode choice of trips bound for the city core. (The CCCC is a high priority project of the city of Philadelphia, approved by UMTA, to connect the Penn Central and Reading railroads via a four-track tunnel under the CBD. The connection will transform two stub-end networks into a fully integrated rail system.) The figures represent the percentage of the total trip makers in Table 3 who would choose, according to the mode-choice model, to take a passenger rail train as their primary mode to reach the city core area.

Park-and-Ride Patrons: Task 3 Results

Table 5 presents a sample of the output of task 3 for three levels of estimates of park-and-ride patrons derived as a function of mean access distance in the trip origin area. The mean access distance is based on the weighted average of zone centroid distances to the site station for the average trip origin area. The percentages were derived from SEPACT II data shown in Figure 1. The range of percentages for each mean distance indicates the possible variation in park-and-ride patrons as exhibited in the SEPACT II data, although this variation may be caused by differences in automobile ownership,

Table 6. Parking space supply and demand.

Site Location	Supply			Demand (medium case)			Additional Spaces Demanded		
	Max	Avg	Min	1976[a]	1980[b]	1985[b]	1976[a]	1980[b]	1985[b]
Bensalem	564	466	368	1466	1720	2236	1000	1254	1770
Baldwin-Crum Lynne	196	115	35	1821	1855	1971	1706	1740	1856
Radnor	768	650	532	2237	2464	2677	1587	1814	2027
Fort Washington	486	486	486	1793	1981	2248	1307	1495	1762

[a] Without center city commuter rail connection. [b] With center city commuter rail connection.

local feeder service, and income as well as by traditional preference.

Additional Space Demand: Task 4 Results

Table 6 presents a sample of the output of tasks 3 and 4 for the medium case. The parking space supply is the number of spaces, existing or planned, that are available to the patrons at each level. These figures do not include those spaces used by park-and-ride rail patrons but not designated as part of the station lots (shopping centers, schools, local streets). The demand is the number of vehicles demanding parking spaces in the study years, assuming average trip origin area, medium level modal and submodal split variables, and medium level park-and-ride response. The medium case is presented as the most reasonable projection of parking demand based on the underlying assumptions and the reasonableness of its output in terms of magnitude, impact, and ability to be implemented. Furthermore, the reasonableness of these projections is substantiated by the fact that all sites selected are at the interface between a rail line and an Interstate route or major arterial. A large fraction of projected demand comprises trips diverted from these highway facilities.

The additional spaces demanded are the differences between demand and supply. These figures represent the demand by future potential patrons who reside within the trip origin area for park-and-ride spaces that will not be satisfied by the existing parking supply. Future parking demands for all three levels of analysis were calculated on an unrestrained basis. The analysis assumes that land is available and that the rail system can provide the required level of service and capacity. Restraining the projection by any one of several factors (land, line capacity, frequency of service, or speed) would result in lower parking demands.

Based on the results of this study, four sites have been given priority for development. If and when developed, these sites will have a combined total increase in parking capacity of 6300 spaces by 1980 and 7400 spaces by 1985. These projects were placed in the Transportation Improvement Program for the Delaware Valley region and are now in the final design and detailed traffic impact analysis stage of development under PennDOT's direction.

ACKNOWLEDGMENTS

This paper was financed, in part, by the Federal Highway Administration and Urban Mass Transportation Administration of the U.S. Department of Transportation and the Pennsylvania Department of Transportation. The contents reflect our views and not necessarily those or the policies of the funding agents.

Publication of this paper sponsored by Committee on Parking and Terminals.

Designing a Parking Management Program

Ann B. Rappaport, Center for Transportation Studies, Massachusetts Institute of Technology

Parking management measures have received considerable attention as a means of controlling automobile use in urban areas, but relatively little attention has been given to the specifics of combining proposed parking strategies into a scheme to help an area realize particular transportation and planning goals. The goal of reducing vehicle kilometers traveled has been selected for the purpose of this discussion, although other goals including reducing peak-period congestion, improving traffic circulation, improving aesthetics, and stimulating retail business should be examined to ensure that the proposed parking strategies are consistent with these goals. This paper focuses on possible traveler responses to various parking control strategies and discusses the implications of these responses for program design. Control of both on- and off-street parking may be necessary in some areas to reduce automobile use. Because parking controls are often fragmented, the coordination of efforts by local and regional agencies is critical to the success of a parking management program.

Parking management is one of the most interesting transportation planning techniques, because it can be used to actually modify automobile-use patterns whereas other techniques are directed toward making alternatives to single-occupant automobile use more attractive. Parking management assumes that the amount, location, and price of parking can affect travel mode choice, trip frequency, and trip destination and that these choices can be modified to produce more desirable travel patterns.

In years past, efforts to manage parking were concentrated on providing an ample supply of spaces at a nominal rate so that retail business could flourish and commuters would find it convenient to drive to work. The Environmental Protection Agency's (EPA) 1973 transportation control plans (TCPs) for a number of cities, including Boston, Denver, and San Francisco, created widespread negative publicity for modifying demand and reducing vehicle use. Measures such as parking surcharges, elimination of on-street parking, and freeze or reduction of off-street parking supplies were proposed to reduce the amount of automobile use in polluted areas so that national air quality standards could be met.

In the December 10, 1973, version of the Energy Emergency Act, the House Committee on Interstate and Foreign Commerce attached a rider forbidding the EPA to impose parking surcharges without the consent of Congress. Surcharges had been included in transportation control plans for 10 areas in California, Massachusetts, New Jersey, and the District of Columbia. Although the Energy Emergency Act was not passed by Congress, the EPA administrator announced that congressional intent on the surcharge issue was clear (1). As a result, all surcharge regulations were withdrawn, and the review date for new parking facilities (to determine their impact on air pollution) was postponed until

January 1, 1975. This was also to be the effective date of parking facility review under the indirect source regulations.

In the August 22, 1974, Federal Register, the EPA administrator published proposed amendments to the parking management regulations. The appendix to the proposed regulations contains guidelines for parking management plan development that describe information requirements and concentrate on how transportation and land use relate to meet air quality standards (2).

Various aspects of parking were covered by EPA in three different ways: as measures in the transportation control plans (including various on- and off-street controls), under the parking management regulations (for new facilities in areas with TCPs), and under the indirect source regulation (for new facilities in any area in the country).

In January 1975 the parking management regulations were suspended on the grounds that proposed Clean Air Act amendments would include provisions on parking management. The parking management regulations were then suspended indefinitely on July 15, 1975, again on the basis of expected congressional guidance regarding parking programs. At that time, the EPA administrator stated, "In the absence of congressional action, EPA may finalize revised parking management regulations in order to complement other transportation control measures" (3). As of January 1977, neither EPA nor Congress had acted.

Because parking management had become associated in the public's mind with the most draconian EPA tactics, and because talk about restricting parking often resulted in making enemies of retail businesses, developers, and other influential community members, the July 1975 suspension of parking management regulations might well have been the last word on parking if the U.S. Department of Transportation (DOT) had not been simultaneously faced with increased demands on its limited funds. One way DOT made its transportation dollars go farther was to emphasize efficient use of the existing transportation system, and it issued the transportation system management (TSM) regulations on September 17, 1975. One of the items in the appendix to the now famous regulations is management and control of parking.

Although nothing could have brought parking back more quickly than its inclusion in a DOT regulation like TSM, some areas, e.g., Cambridge, Massachusetts, had begun to implement parking programs on their own. Reducing congestion on downtown streets, improving delivery of city services such as snowplowing and garbage pickup, making transit more competitive with the automobile, maximizing tax dollars by discouraging open-air parking lots and encouraging developments with higher assessments, and improving the quality of urban life have all been cited as reasons for the development of parking management programs.

In general, parking management may be described as any alteration of parking supply or parking rates that discourages or prevents parking in certain areas, at certain times, or by certain groups. A number of parking controls or strategies have been used or proposed in managing an area's parking supply. Brief descriptions of some are contained in this paper; more extensive descriptions may be found elsewhere (4, 5).

Work trips are the target of most parking controls because they usually occur at set times during the day on a regular basis and can thus be diverted from single-occupancy automobile to shared ride with fewer adverse impacts than can other trip types. An individual must continue making work trips to earn an income, regardless of the disincentives, and will switch from single-occupant automobile to other modes (including shared ride). The intent of the parking controls is not to reduce person trips but to encourage people to make trips in such a way that single-occupant vehicle use is reduced. One alternative to making any trip other than the work trip (in the face of disincentives) is to not make the trip at all. This possibility is usually considered to have disastrous economic implications (loss of income to retail businesses, hotels, motels, resorts; loss of sales of recreational equipment). Therefore, reeducation is aimed at the commuter, who is a captive trip maker.

Although parking management may be used to meet a wide variety of goals, this paper will concentrate on vehicle travel reduction, which, when not accompanied by a reduction in person trips, implies more efficient utilization of the transportation system. Parking strategies have been proposed as a means of reducing total vehicle travel, work-trip travel, peak-period travel, and travel within the core area. It is important to be aware of potential undesirable effects of various parking control measures, to examine the types of incentives and disincentives that each measure implies, and to consider the effect of each measure in terms of the area's goals for parking management. For example, a reduction in total vehicle travel might mean a decrease in person trips and therefore decreased mobility; a reduction in work-trip travel implies that automobiles left at home might increase nonwork travel; a reduction in the amount of peak-period travel may simply mean a redistribution of trips over time and no net change in vehicle travel; and a decrease in vehicle travel in the core area may be offset by an increase in travel elsewhere.

In general, parking control measures fall into two groups, rate controls and supply controls, each of which may be subdivided into on-street and off-street controls as shown below.

Control	On Street	Off Street
Rate	Erect meters or increase meter rates Add surcharge	Impose tax Add surcharge Change rate structure to discourage long-term parking
Supply	Issue permits or licenses Ban parking either totally or at specific times Erect meters or adjust meter times	Freeze, cut back, or restrict growth Use time and vacancy rate restrictions

In this paper I discuss measures under each category, impacts of parking strategy on travel behavior, and the potential of a parking scheme design to reduce vehicle travel.

RATE CONTROLS

Description of Rate Controls

Increased rates for both off-street and on-street spaces have been widely proposed to control parking. A reduction or restriction of off-street supply could force a rate increase; rates could be imposed if parking is currently free; or existing rates could be arbitrarily raised. Meters could be erected on streets where parking is now free to provide a cost-and-time disincentive, and rates on currently metered spaces could be raised to provide an additional disincentive.

Higher rates at private off-street facilities could result in a number of legal problems. Conversion of the rate structure from its current daily maximum to a flat rate, for example, requires that the commercial parking facilities be regulated by local government. To regulate rates requires one to show that parking is a business that affects the public interest and that the regulation is necessary in the interests of public health, safety, morals, and general welfare. There are no direct precedents for this type of parking regulation. Furthermore, the authority of local governments to regulate business must be delegated by the state; this transfer of authority is by no means automatic and may be a barrier to cities' regulatory efforts (4, 5).

One backdoor means of encouraging privately owned facilities to raise rates may be through raising rates at municipal garages (if municipal garages are significant competitors). In Boston, privately owned facilities try to keep their rates competitive with municipal rates, so requiring city facilities to increase parking charges might result in an overall increase.

Parking Tax

City taxes on nonresidential parking transactions have been proposed to discourage automobile use and to generate revenue. San Francisco had a 25 percent tax from October 1, 1970, to June 30, 1972, at which time the tax was lowered to 10 percent (6); Pittsburgh levied a 20 percent tax in 1969. The Supreme Court upheld the validity of the latter tax in Pittsburgh v. Alco Parking Corporation (Sup. Ct., slip opinion 73-582, June 11, 1974):

> By enacting the tax, the city insisted that those providing and utilizing the nonresidential parking facilities should pay more taxes to compensate the city for the problems incident to off-street parking. The city was constitutionally entitled to put the automobile parker to the choice of using other transportation or paying the increased tax.

Taxation is inherently a state power, and cities may levy taxes only with specific grants of authority from the state constitution or legislature. States have made different provisions for local taxation. Some local governments have been granted broad authority to establish local tax policy subject only to prohibition by or conflict with state law; other local governments are limited to certain types of taxes such as revenue-producing taxes on business (4, p. 118). Use of tax revenues by cities depends on the authority under which they were generated; proceeds from a revenue-producing tax will typically be put in a city's general fund.

Parking Surcharge

Most parking surcharges, flat fees on top of the existing ones, range from \$1.00 to \$5.00. They may be applied, for example, to all parking within a specified area such as a central business district (CBD), to all long-term parking (4 or more hours), to all parking in transit-adequate zones, or to all parking arriving in an area between 7:00 and 9:30 a.m. The scheme depends on which group is to bear the burden of the disincentive. Parking surcharges vary in magnitude and may cause changes in travel behavior.

Changed Rate Structure

Rate structure changes favoring short-term parking in business and commercial areas have been proposed to discourage commuter but not shopper and tourist parking. Cities, however, often lack adequate police power to directly control rate structures of private parking facilities. These structures can be designed to encourage short-term at the expense of long-term parking (most facilities currently favor long-term parking by charging the same for a 3-h as for an 8 to 9-h parker), or rates can be applied on a flat, per-hour basis, ostensibly favoring neither, but usually the commuters pay more.

Discussion of Rate Controls

Rates and Rate Structures

Many city officials believe that controlling parking rates is crucial to an effective parking management plan. Their reasons include ease of administration (relative to supply controls), ease of enforcement, and potential for increased revenue. It is often proposed that parking rates should favor short-term parkers (shoppers and tourists), so that retail sales in controlled areas do not suffer. Besides (the argument goes), commuters are usually in a better position to use transit or shared-ride modes. However, it is conceivable that, if short-term parkers are favored, short-term trips may increase. This behavior was observed in Philadelphia, where rate changes at individual garages produced short-term increases of 15 to 20 percent (7).

Flat per-hour parking rates seem more reasonable. They would increase commuter fees, since present rate structures usually favor the all-day occupant and, depending on the magnitude of the charge, could provide the type of disincentive desired to divert commuters to other modes. The short-term parker may or may not face increased rates, depending on the per-hour charge. Although such a policy will not encourage shopping trips, it will not discourage them either; the short-term parker will be better off, since shorter occupancy will cost less. The fact that a flat-rate system does not encourage short-term trips is important if the objective is to reduce total vehicle travel in an area. If commuters leave their automobiles at home and commute by transit or car pool, then automobiles will be available for those at home to make shopping and other nonworking trips; and if parking rates encourage shopping trips to the CBD, then it can be assumed that there will be some increase in nonwork vehicle travel in the core, which may or may not offset reductions in vehicle travel from commuter disincentives. A flat rate takes into account the fact that vehicle travel is vehicle travel regardless of who generates it, but at the same time places a heavier burden on the long-term parker, who is generally a commuter and more likely to be induced by a disincentive to seek new modes to work, than on the short-term parker, who is likely to be a shopper or a tourist and likely to do his or her retail spending elsewhere if the disincentive is sufficiently burdensome. A flat-rate approach might be a reasonable compromise between those interested in reducing vehicle travel and those interested in the economic effects of parking policies.

Rate Increases in General

Type of control relies on the notion that there are enough individuals (mostly commuters) who can be encouraged to make different mode-choice selections when parking rates are raised above a certain level. If the vast majority of individuals elected to pay increased rates, then the parking control measure, having achieved only minimum reductions in vehicle travel, would have failed.

Those developing parking price schemes must determine the level of price increase necessary to divert a given percentage of single-occupant automobile drivers to transit or shared-ride vehicles. The Philadelphia rate case does indicate that use of parking facilities changes with changes in rates, but it does not give any indication of behavior that can be expected when rates at all facilities are increased.

Although increased parking rates may be justified as a disincentive to automobile use, the equity of such increases deserves consideration, since they will be felt more by lower income drivers. In general, wealthy com-

Table 1. Impact of controls on parking rates and supply.

Control	Measure	Increases Cost to Commuter	Encourages Short-Term Use	Requires Enforcement	Encourages Flexible Work Hours
Rate	Tax	X			
	Surcharge	X			X
	Rate change	X	(X)		
	New meters or increased rates	(X)	X	X	
Supply	Permit or license	(X)	(X)	X	(X)
	New meters or adjusted times	(X)	X	X	X
	Reduced or limited growth	(X)			
	Time and vacancy rates	(X)	(X)	X	X
	On-street parking ban				
	Total	(X)		X	
	At specific times		X		X

Note: X = direct and (X) = indirect or optional.

muters are more likely to pay the additional money and continue driving while lower income commuters are more likely to seek alternative means of getting to work. Therefore, low-income automobile commuters who have no alternative to driving alone will be penalized.

Table 1 summarizes the impacts of parking rate controls.

SUPPLY CONTROLS

Off-Street Supply Controls

Description of Off-Street Controls

Off-street supply controls are designed to limit the amount of parking, and it is assumed that the uncertainty or difficulty of obtaining a parking space will cause some automobile users to divert to transit and shared-ride modes. This control, nevertheless, runs counter to many cities' goal of accommodating drivers by providing ample off-street parking.

Prohibiting the construction of additional parking capacity is probably feasible, but regulating the number of spaces that existing parking facilities utilize (retrospective application of controls) may prove quite difficult. Off-street facilities are privately and government owned, so that regulation of both may be subject to legal challenges. Private facilities may claim that a regulation is "taking without compensation." Furthermore, any classification scheme must be reasonable and must provide equal protection for all facilities. Government facilities are usually exempt from zoning regulations, and the ability of one level of government to regulate the facilities of another may be greatly complicated by sovereignty. These regulatory issues are explored in depth elsewhere (4).

Freeze on Parking Spaces

Freezing the number of parking spaces, as it is generally practiced, puts an upper limit on off-street parking spaces equal to those in existence on a certain date. The effectiveness of a freeze depends on the amount of parking on the freeze date compared with that needed for development over the next few years. If the number of existing spaces exceeds current demand, then the impact of the freeze will not be felt until some time in the future.

In Boston, as spaces are eliminated, they are put on a "freeze bank" and may be allocated to new or other developments within the freeze area. Developing a reasonable and equitable means of distributing the spaces in the bank to those who desire them has proved difficult; other areas may not wish to permit banking of spaces, particularly if capacity greatly exceeds demand.

Reduction of Off-Street Parking Spaces

A reduction in the number of off-street parking spaces has been proposed for areas with excess capacity. Although this approach could be effective—via voluntary space cutbacks by government and cutbacks of private space through a nonconforming use zoning approach—it will undoubtedly have to weather legal challenges. Those whose lots are phased out will probably level the accusation of "taking." In addition, which spaces to phase out may be even more difficult to decide than which to allocate in a freeze.

Restricted Parking Supply Growth

Restricted parking growth may be necessary in areas undergoing active development. Parking controls are intended to keep the supply of spaces below demand to encourage decreased drive-alone automobile use, but extreme constraints on parking supply may have an adverse effect on business location decisions. In rapidly developing areas, it may be necessary to permit some increases in supply while still keeping overall supply below overall demand for spaces. EPA has suggested this strategy for providing a gradual increase in parking supply (for instance, 100 spaces) each year and then recommends (2):

> Applicants for the limited number of new parking space permits could then be judged based on predetermined and published criteria. Such criteria could include such diverse elements as community need, proximity of mass transit, financial per space contribution to mass transit, VMT [vehicle miles of travel] impacts and efforts made to minimize VMT.

Other schemes might be based on measures of growth, such as one space for every two new jobs created, or on some floor area ratio for construction completed in the last year. The difficult problem of allocating the spaces equitably to a group of qualified applicants remains to be resolved at the local level and seems to be basically similar to allocating banked spaces in a freeze.

Time and Vacancy Rate Restrictions

Time and vacancy rate restrictions have been proposed as a means of favoring one automobile-driving group over another (for example, shoppers over commuters) or of putting pressure on drive-alone commuters, the group with the greatest opportunity for shared ride. There are two different ways of achieving this. The first is to require that certain facilities in the area open only after 9:30 a.m. (for example, making them primarily for shopper and other short-term use), and the second is to require that some percentage of the spaces in all facilities be available at 9:30 a.m. There must be some reasonable basis for distinguishing between regulated and unregulated facilities, if the measure is to be applied selectively. In

Boston, a 40 percent vacancy rate was proposed under the first transportation control plan and was to be applied to all facilities in the core area. In South Terminal Corporation v. EPA (Ct. App., 1st Cir., Sept. 24, 1974), the court ruled that such a measure did not constitute a "taking":

> The Government has not taken title to the spaces, and the decision about alternative uses of the space has been left to the owner.... The right to use is not extinguished entirely; nor is it transferred to anyone else. Indeed, the ingenuity of operators may result in fewer disadvantages than urged.... In any event, even a diminution of profits or a requirement that some loss be suffered is not enough, when all other accoutrements of ownership remain, to be a "taking."

Discussion of Off-Street Controls

Freezes, Cutbacks, and Restricted Growth

Measures to control the number of off-street parking spaces within an area, including freezes, cutbacks, and restrictions on growth, will probably have a similar impact on commuter parking and will be considered here as a single group.

Programs that establish freezes and provide for phasing out of parking spaces in underutilized, outdated facilities and for incorporating the spaces in new developments may actually be increasing the effective capacity of parking in the freeze area if the new spaces are used and the old ones were not. If supply currently meets all needs, then vehicle travel will not be reduced below its current level unless some of the available spaces are eliminated. Probably the easiest off-street supply restrictions to implement, parking freezes and restricted growth programs, alone might serve to keep vehicle travel at approximately current levels in the short run. In the long run (assuming that business and retail trends continue at approximately the same levels and that no mass exodus to the suburbs occurs), market forces will probably raise the price of off-street parking as demand begins to exceed supply. A rate increase caused by market forces or as part of the parking program will have the same kinds of impact on commuters as those discussed for rate controls. As demand for spaces exceeds supply, traveler behavior changes.

If off-street parking supply restrictions are to reduce vehicle travel, it will probably be necessary to create a situation in which (a) on-street parking is restricted and (b) demand for spaces exceeds supply at the present time. This may occur naturally in some areas and may have to be created artificially in others by reducing the number of spaces.

Assume for a moment that no price increase will accompany a supply restriction and that commuters will react only to the problem of space availability. One possible outcome is that to assure themselves of parking spaces, commuters will arrive at the CBD earlier; thus, either the morning peak period will occur somewhat earlier or the duration of the peak will be increased. Ideally, enough drivers will find parking so inconvenient they will choose other modes.

Vacancy Rate and Time Restrictions

If a vacancy restriction is applied uniformly to all off-street facilities in an area, so that, for example, 40 percent of all spaces must be available at 9:30 a.m., then the impact will probably be analogous to an on-street parking ban. That is, businesses may be encouraged to institute staggered work hours, and commuters arriving before 9:30 a.m. who are unable to secure an off-street space may park on-street until the additional 40 percent is available. If on-street parking is restricted, changing commuter mode choice depends on willingness of area businesses to have staggered or flexible work hours. Although such work hours may reduce peak-period volumes, relieve congestion, permit higher average speeds, and thereby reduce localized carbon monoxide concentrations, there is no reason to assume that they will also reduce vehicle travel.

The other method of imposing a vacancy restriction is to designate certain parking facilities that may not open before a certain time. This would be legally possible only if there were some reasonable basis for distinguishing between late and early opening facilities. Such a basis might be the location of parking facilities in predominantly retail shopping and tourist areas, as opposed to those in the commercial-office district. Used alone, a regulation that garages in shopping areas could not open until 9:30 a.m. would probably not have a discernible impact (because if the facilities had a great deal of commuter use, the distinction would be invalidated), particularly if stores do not open until 10:00 a.m. If, however, such a restriction were used in conjunction with other supply measures such as a freeze, people might be increasingly willing to walk or take transit from the shopping area to their places of work. In the limited supply case, the effect of this type of vacancy restriction on travel behavior will probably be similar to the effects of a vacancy requirement at all facilities, except that the number of individuals affected may be smaller, depending on the number of garages that open after 9:30 a.m.

On-Street Supply Controls

Jurisdiction over supply and regulation of on-street parking is usually held at the local level (with some state constraints), and controls of this type are not expected to encounter the legal difficulties that have been raised with regulation of private off-street facilities.

Parking Ban

A parking ban is intended to reduce the availability of parking (usually to commuters) and can be applied in a variety of ways. On-street parking can be totally eliminated in areas with sufficient off-street parking supply or transit access or both. Or, in areas with relatively small supplies of off-street parking, on-street parking could be banned between certain hours, for example, 7:00 to 10:00 a.m. This would favor shoppers and may be desirable or even necessary if some provision such as vacancy rate (for example, 40 percent of spaces at off-street facilities available at 10:00 a.m.) is not used, since commuters typically arrive at the CBD first and will have first chance at the spaces. An interesting variation of the ban is to make on-street spaces in nonretail areas available only to car pools; this is a much-needed incentive for areawide car-pool programs.

A ban on on-street parking is an attractive parking control measure because it can usually be implemented and enforced entirely by a city's existing departments; it needs no new grants of power; and it raises minimum legal challenges. The main costs to the city of such a ban are erecting signs and increasing enforcement.

Most parking bans aim at limiting on-street parking for work trips by banning parking between 7:00 and 10:00 a.m. or 7:00 and 9:30 a.m. This time restriction will theoretically work to the advantage of shoppers, since they arrive at the CBD later than commuters. This type of parking ban is also seen as a necessary compensating mechanism for areas that will have a limited number of off-street spaces. Again, because the commuter arrives earlier and will have the best chance of getting a space, the ban is seen as a way of reserving some spaces for shoppers in the absence of a vacancy rate provision at off-street facilities.

If a city's aim is to reduce peak-period travel, then such a ban may well contribute to this goal; if the aim is to reduce vehicle travel, then a ban may not work. A ban assumes that (a) once a commuter has arrived at work the vehicle will not be moved until it is needed for some "legitimate" purpose or until the commuter goes home and (b) businesses in the area will continue to operate within their preban work schedule.

When off-street capacity is limited and employers are willing to allow flexible or staggered work hours, the effectiveness of a ban may be limited to a reduction in peak-period travel; effects on total vehicle travel will be minimized. In mixed office and retail areas, staggering hours so that employees can arrive after the ban has been lifted may tend to have an adverse effect on availability of parking for retail customers.

Metering

Metering of on-street spaces in business and commercial areas can be used to achieve a number of goals. Meter rates may be adjusted to provide a cost disincentive (particularly in areas where parking has previously been free) and meter times can be used to encourage shopper parking over commuter parking by imposing a 1 or 2-h limit. However, effective use of meters for parking control depends on a rigorous enforcement program. This involves having fines large enough to present a disincentive and a systematic way to recover fines. Many areas will find these criteria difficult to meet. Maintaining an enforcement staff is costly; determining upper levels of fines for violations may be done by the legislature; and expediting court procedures for ticket processing may be costly and time consuming, particularly if it involves computerization of the recording and summons-issuing tasks.

In some areas, metering alone cannot change mode choice. A commuter's decision might involve: (a) cost of driving and parking at a meter versus cost of other modes, (b) time limits on the meter, (c) expected quality of enforcement (if enforcement is known to be good and tickets can be expected for time violations and for meter feeding, meters will be a greater disincentive for commuters than if enforcement is expected to be lax and a commuter feels that the risk of being ticketed is quite low), and (d) cost of a violation (if a violation is very costly and enforcement and recovery are known to be good, then the disincentive posed by metered spaces may be significant).

Metering may be most effective when used in conjunction with other control measures, such as the parking ban, and with off-street measures to prevent a diversion from off-street to on-street parking when off-street controls are introduced.

Area Licenses or Permits

Area parking licenses or permits have been proposed when one class of user may be legitimately distinguished from another. For example, residential permit programs have been established in mixed commercial and residential areas to give the city resident some sort of priority in on-street parking. In Cambridge, the system is set up so that on certain streets parking is restricted to vehicles with residential permits except on Sunday. In some areas residential permits could be used primarily to exempt city residents from a 7:00 to 10:00 a.m. on-street parking ban. Other programs would require permits for parking only during business hours, for example, from 7:00 a.m. to 6:00 p.m. Constitutional issues have been raised about some residential permit systems on the grounds that they restrict the right to work and travel. Although there are no causes in point, one study suggests (4, pp. 131-132) that such a program would probably be determined constitutional if the ordinance has

> . . . 1) a clear statement of public need for the ordinance, under the police power over public health, safety, and welfare; 2) reasonable provisions for parking by nonresidents to the maximum extent consistent with its purposes; 3) no unreasonable restrictions on commercial vehicles; 4) provisions for adequate notice to the public; 5) no provision for arbitrary exceptions; 6) the same effect on non-area city residents as it does on non-city residents; and 7) the cost of a residential permit is kept minimal.

The first provision could probably be met on the grounds that the ordinance is designed to reduce automobile traffic in residential areas to protect residents from the detrimental effects of high levels of automobile use: danger to children playing in or near streets, exposure of residents to high levels of air pollutants and traffic noise, and disruption of community life. The purpose of this provision is to clearly justify the use of an area's police power as the authority for such a program.

Most residential permit programs are designed to prevent commuters (noncity residents) who cannot find parking spaces in commercial and business areas from spilling over into adjacent residential areas and causing parking shortages, increased traffic, and congestion there. A residential permit program generally increases the urban resident's probability of finding a legal parking space near the home.

Singapore has implemented an area license scheme that restricts entry to the CBD before 10:15 a.m. to vehicles bearing a special permit. Permits may be purchased on a daily or monthly basis (S$4/d, S$80/month) from stations at the edge of the CBD. Some schemes also propose restricted parking but not entry (8, 9).

For the person working in or near a restricted area, such a program has an impact on work trip mode choice only if the worker is accustomed to driving to work and finding an on-street parking space in the residential area. If the commuter typically parks off-street in an employer-provided lot or in a commercial lot, then clearly the choices are not greatly changed, except that the commuter must expect increased competition for the off-street spaces. The effect on those accustomed to parking on-street greatly depends on the availability of alternative parking. If the major employers in such an area can be convinced to expand their employee parking facilities (either free or at a nominal charge), then the impact of the program on commuters may be minimum. On the other hand, if employer space cannot be expanded and commercial space is limited, then there will be a point at which the commuter must compare the cost of driving to work plus paying to park at the commercial facility with the cost of all other available modes to work (including shared ride).

Table 1 summarizes the impacts of parking supply controls.

INSTITUTIONAL CONSIDERATIONS

Various institutional constraints and peculiarities that may influence parking management have been mentioned here. The institutional arrangements that govern parking are often unfamiliar to transportation planners because parking controls are typically held at the local level and parking policy has traditionally been a city concern. As a result, parking policy may be a product of the interaction among a variety of city agencies and interests including the public works or traffic department, the planning department, the airport authority, the urban renewal authority, the zoning board, and the police department. Control will probably be fragmented. For example, an urban redevelopment authority will dictate off-street parking policy within an urban renewal area; the traffic department will control on-street parking; and another city agency, such as the Real Property Board in Boston, will run the city's off-street facilities.

Because of this fragmented control and multiplicity of actors and interests, development of a parking management plan is necessarily negotiation intensive. In that respect, it is similar to many other TSM measures, such as establishing preferential lanes or modifying bridge tolls to favor car pools. Identifying the various institutions involved in parking should be an integral part of the early development of any city's parking management plan.

DEVELOPMENT OF A PARKING MANAGEMENT PROGRAM

Because parking policy may be a product of many local interests, one should begin by finding out what the existing institutional arrangements governing parking are and perhaps by identifying particular instances in which legislative changes are desirable (for example, formation of a citywide parking authority). While acquiring information on the existing distribution of authority, one should be able to simultaneously acquire an understanding of the political climate as reflected in parking policy (for instance, encourage CBD development at all costs, discourage new construction, encourage renovation of existing buildings, increase transit use, continue to improve automobile accessibility) and to gain an understanding of the divergent interests influencing parking policy and the constraints on radical change in parking policy (for example, revenues from certain facilities may have been pledged as security for bonds).

Once one understands the existing parking situation, information on location, number of spaces, ownership, and current charges should be gathered. Then parking information should be fed into the metropolitan transportation planning process to develop parking strategies compatible with overall transportation goals. This process should clarify the role of parking management and make it possible to develop a statement of goals. Strategies to meet the goals should be developed with input from all interested (affected) city and regional agencies, and, where applicable, state and federal agencies, and the public. Involvement of the public is particularly important in helping people to understand planning motivation and the alternatives.

A strategy for implementation should be selected from the alternatives, weighing the impacts, costs, and practicality of each. Prior to implementation of a parking plan, all implementing, enforcing, and monitoring agencies should have agreed to carry out the responsibilities that fall to their agencies.

CONCLUSIONS AND RECOMMENDATIONS

There is every reason to expect that parking management can be used to modify travel patterns, but it is not clear at what level parking price increase or supply decrease will cause a particular mode change or vehicle travel reduction. Experiments with parking controls and other automobile restraints will continue to widen the data base, so that eventually a clear relationship between controls and responses can be established.

Parking management planning must become part of transportation system planning. Parking policies, transit policies, and highway policies must be coordinated to reinforce one another if the transportation system is to be used in the most efficient manner possible. For example, a city cannot expect to increase the transit trips while continuing to improve automobile accessibility and to provide ample parking.

Finding the appropriate group of measures to achieve the desired results and to "plug all of the leaks" in the parking system requires an assessment of the institutional arrangements governing parking as well as an understanding of parking supply and use characteristics. Restricting on-street parking will have a tendency to increase use of off-street parking, so that a policy to reduce automobile use should contain both on-street and off-street measures. To exercise some parking controls may require additional grants of authority from states to cities. In addition, it may be desirable to make institutional changes, such as consolidating parking authority in one city agency. Major changes of this type should be identified early in the parking management process, for they may take a long time to implement.

ACKNOWLEDGMENTS

This paper is based in part on research conducted for the U.S. Environmental Protection Agency. The opinions and conclusions expressed are my own and are not necessarily those of the Environmental Protection Agency or of the people who provided information and assistance in the research. The contributions of Elizabeth Deakin and Greig Harvey were of great value to this effort and are gratefully acknowledged.

REFERENCES

1. Revisions to Surcharge and Management of Parking Supply Regulations. Federal Register, Vol. 39, Jan. 15, 1974, p. 1949.
2. Proposed Amendments to Parking Management Regulations. Federal Register, Vol. 39, Aug. 22, 1974, p. 30440.
3. Indefinite Suspension of Parking Management Regulations. Federal Register, Vol. 40, July 15, 1975, p. 29713.
4. Parking Management Policies and Auto Control Zones. Metropolitan Washington Council of Governments and National Capital Region Transportation Planning Board; U.S. Department of Transportation, Draft Rept. DOT-OS-400045-1, May 1975.
5. D. J. Zaelke, Jr., and J. W. Russell, Jr. Energy Conservation Through Automobile Parking Management. Energy Conservation Project, Rept. 6, May 1976.
6. D. J. Kulash. Parking Taxes as Roadway Prices: A Case Study of the San Francisco Experience. Urban Institute, Working Paper 1212-9, Dec. 1973.
7. Philadelphia Central Business District Parking Rate Experiment. Delaware Valley Regional Planning Commission, Sept. 1974; NTIS, Springfield, Va., PB 238 168.
8. Business Impact Study: Stage 1. Transportation Research Division, International Bank for Reconstruction and Development, July 1975.
9. Business Impact Study: Stage 2. Transportation Research Division, International Bank for Reconstruction and Development, Jan. 1976.

Publication of this paper sponsored by Committee on Parking and Terminals.

Critical Lane Analysis for Intersection Design

Carroll J. Messer and Daniel B. Fambro, Texas Transportation Institute, Texas A&M University

This paper presents a new critical lane analysis as a guide for designing signalized intersections to serve rush-hour traffic demands. Physical design and signalization alternatives are identified, and methods for evaluation are provided. The procedures used to convert traffic volume data for the design year into equivalent turning movement volumes are described, and all volumes are then converted into equivalent through-automobile volumes. The critical lane analysis technique is applied to the proposed design and signalization plan. The resulting sum of critical lane volumes is then checked against established maximum values for each level of service (A, B, C, D, E) to determine the acceptability of the design. We provide guidelines, a sample problem, and operational performance characteristics to assist the engineer in determining satisfactory design alternatives for an intersection.

To provide an acceptable level of service to drivers operating along an urban arterial, the signalized intersections must keep the traffic moving. The ability of a signalized intersection to move traffic is determined by the physical features of the intersection, by the type of signalization used, and by the geometric design (1). Thus, total system design of a signalized intersection involves concurrent evaluation of the proposed geometric design and signalization as an operational system.

Designing a signalized intersection frequently involves making trade-offs between design variables (and their associated costs) and the resulting level of service. Level of service at an intersection describes the quality of traffic flow afforded motorists on a particular approach to the signalized intersection. The various levels of service may be characterized qualitatively as follows:

Level of Service	Description
A	Light traffic on approach, short stable queues during red
B	Moderate traffic on approach, stable queues, little additional delay
C	Moderately heavy traffic on approach, moderately long but stable queues during red, moderate but acceptable delay
D	Heavy traffic on approach, long unstable queues, sometimes excessive delays
E	Heavy flow (capacity) on approach, long queues, excessive delays
F	Heavily congested traffic conditions, more traffic demand than signal capacity

DESIGN PROCEDURE FOR SIGNALIZED INTERSECTIONS

The critical lane analysis technique is used in this procedure to determine if a proposed design will provide an acceptable level of service. Level C is the minimum during the peak 15-min period of the design hour; however, all operating conditions can be evaluated.

Basic design variables include the number of approach lanes provided, the possible length and use of left and right turn lanes, the combination of traffic movements using the lanes provided, and the type of signal phasing. Minimum design standards for the basic design variables of lane width [3.0 to 3.7 m (10 to 12 ft)] and curb return radii [4.6 to 9.1 m (15 to 30 ft)] will normally provide satisfactory operation during rush hours.

Volume Data Preparation

We will use a sample problem to illustrate the application of the procedure over a range of initially given volume data conditions. In practice, the given traffic data would dictate the appropriate step in the volume preparation procedure at which the designer should begin the analysis.

Step 1. Average Daily Traffic

In our sample problem we assumed that the given volume data are the forecast, design year, average daily traffic (ADT) volumes. These two-way ADT volumes are shown in step 1 of Figure 1. Our given traffic and operational conditions are as follows: Volumes are 1985 ADT; design hour factor (K) is 10 percent; directional distribution (D) is 67 percent; trucks and through buses (T) is 5 percent; and population for 1985 is 400 000.

Step 2. Design Hour Movement Volumes

We first converted the two-way ADT volumes into approach movement volumes for the design hour being analyzed. The morning peak hour is assumed in this example. The evening design hour could also be checked, because, if the given volumes are in ADT, the morning design hour volumes for left turns on one approach become the right turns on the departure leg during the evening peak hour.

The morning design hour volumes are shown in step 2 of Figure 1. The directional peak flows move from left to right and from bottom to top. Other factors being equal, the location and the orientation of the intersection in the metropolitan area dictate the peak directions of flow. The larger of the two design hour, directional, movement volumes flowing between legs a and b is calculated from

$$DHV_{ab} = ADT_{ab} \cdot K \cdot \bar{D} \qquad (1)$$

where DHV_{ab} is the design hour, peak direction movement volume between legs a and b, ADT_{ab} is the average daily traffic interchanging between legs a and b (step 1, Figure 1), and $\bar{D}$ is the average directional distribution split (decimal equivalent) between the approaches. $\bar{D}$ is either 0.67 or 0.50. The off-peak direction movement volume between legs a and b is calculated from

$$DHV_{ab} = ADT_{ab} \cdot K \cdot (1.00 - \bar{D}) \qquad (2)$$

Step 3. Design Period Volumes

We used the peak 15-min period of the design hour to evaluate the level of service. The traffic volume flow rates during this period consistently exceed the average for the design hour by approximately 20 to 30 percent. These peaking factors have been found to vary with the population of the city (1, 2) as follows. According to the Highway Capacity Manual (1) and others (2), these peaking factors have been found to vary with the population of the specific city in the following manner: For populations under 100 000 the peaking fac-

Figure 1. Steps in volume data preparation.

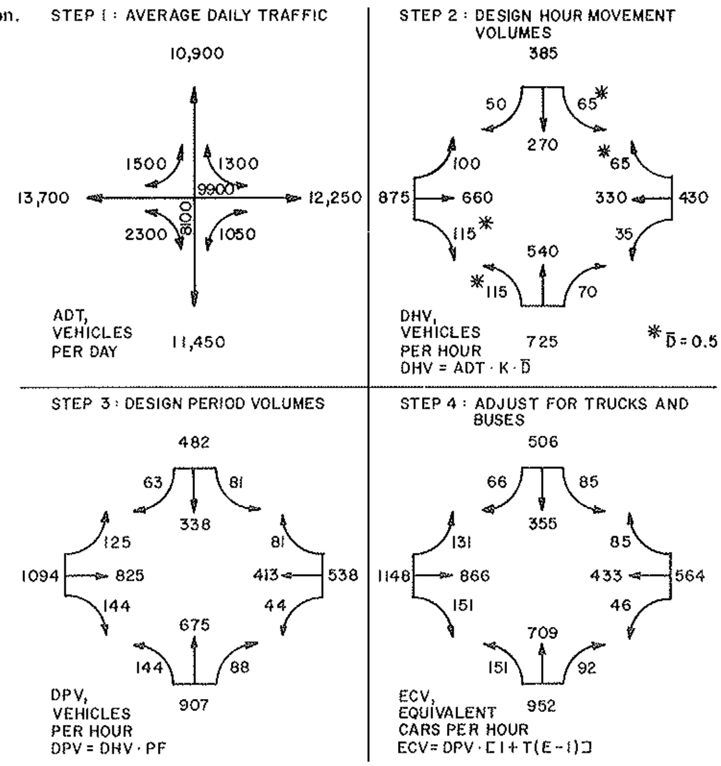
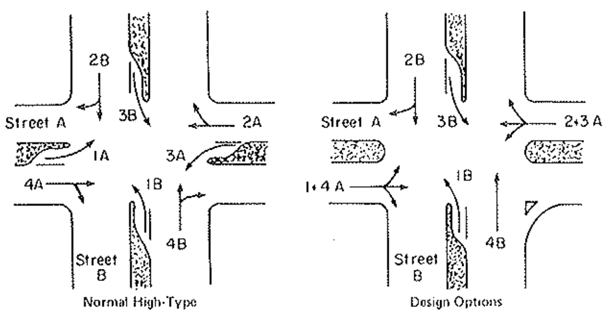


Figure 2. Definition of movements.



tor is 1.35; for those 100 000 to 300 000 it is 1.30; for those 300 000 to 500 000 it is 1.25; and for over 500 000 it is 1.18.

Design period flow rates for each movement are calculated from

$$DPV_{ab} = DHV_{ab} \cdot PF \tag{3}$$

where DPV_{ab} is the design period volume from leg a to b as given in step 3 of Figure 1; DHV_{ab} is the design hour volume from step 2 of Figure 1; and PF is the peaking factor given above, in this case 1.25 for a population of 400 000.

Step 4. Equivalent Passenger Automobile Volume

This design procedure converts all design period volumes of mixed traffic (5 percent trucks and through buses in this example) into an equivalent number of automobiles. One truck or through bus is equal to two automobiles (2). Thus, the equivalent automobile volumes (ECV) in step 4 of Figure 1 are calculated from the design period mixed traffic flow rates of step 3, by

$$ECV_{ab} = DPV_{ab}[1.0 + T(E_T - 1)] \tag{4}$$

where E_T is the automobile equivalent for trucks and through buses. Since E_T is assumed to be 2.0,

$$ECV_{ab} = DPV_{ab}(1.0 + T) \tag{5}$$

Geometric Movement Volumes

The next step in the design guide requires that the individual ECV turning movement volumes (step 4 of Figure 1 and Equation 5) be defined by the way they will be combined in the geometric design of the intersection. Eight basic geometric movement volumes would exist at an urban arterial, four-legged intersection having left turn bays on all approaches, as depicted on the left intersection of Figure 2. When a left turn bay or a separate left turn lane is provided on an approach, the left turn geometric movement volume is the same as its corresponding turning movement volume in ECV from Equation 5 ($GMV_{1A} = ECV_{lt}$). The adjacent through-right geometric movement volume would be calculated as $GMV_{4A} = ECV_{th} + ECV_{rt}$. If a channelized or free right turn lane is provided, we can dispense with the ECV right turning volume (ECV_{rt}).

When an approach does not have a left turn lane, the left turning movement volume (ECV_{lt}) is added to the appropriate through-right movement volume forming a combined left-through-right geometric movement volume [$GMV_{(1+4)A} = ECV_{lt} + ECV_{th} + ECV_{rt}$] for the eastbound approach of the right intersection shown in Figure 2.

Geometric Design Volumes

All of the geometric movement volumes are further adjusted to account for the (proposed) design and operational features of the intersection. These equivalent-effect volumes are calculated from

$$GDV_m = U \cdot W \cdot TF \cdot GMV_m \tag{6}$$

where GDV_m is the geometric design volume for movement (m) of Figure 2, vehicles per hour; U is the lane utilization factor (3); W is the lane width factor; TF is the turning movement factor (Equation 7); and GMV_m is the geometric movement volume of movement (m). The sum of one or more turning movement volumes (ECVs from Equation 5) is illustrated in Figure 2.

Lane Utilization

As the number of lanes serving a movement increases, there is an increasing tendency for one lane to be used more than the others, which is accounted for by the lane utilization adjustment factor (U). These factors are given below.

Number of Lanes	Factor
1	1.0
2	1.1
≥3	1.2

Lane Width

Lanes 3 m (10 ft) or more in width have little influence on rush-hour traffic flow rates as reflected by the lane width adjustment factor below (1 m = 3.3 ft). The width of a lane does not include any pavement used for or appreciably affected by parking. Through lanes less than 3.4 m (11 ft) wide may experience related safety problems.

Width (m)	Factor
2.7 to 3.0	1.1
≥3.0	1.0

Turns

The effects of turning vehicles on flow are given by the turning factor as

$$TF = 1.0 + L + R \tag{7}$$

where L, which adjusts for the effects of left turns and R for right turns, is L_1, L_2, or L_3 and is described as follows.

For an approach having no left turn bay, the left turn adjustment factor to be applied to a combined left-through-right movement volume is calculated from

$$L_1 = P_L(E - 1.0) \tag{8}$$

where P_L is the decimal fraction of the total approach volume turning left and E is the appropriate left turn equivalent factor from Table 1 (1).

For an approach having a left turn bay, the adjustment factor is

$$L_2 = (1700 - E)/S - 1.0 \tag{9}$$

where S is the saturation flow of the left turn bay obtained from Figure 3 for a given storage length and equivalent left turning ECV from step 4 of Figure 1. The left turn equivalent factor (E) is obtained from Table 1.

The desired minimum left turn bay storage length, which does not include either the taper section or any length of the bay beyond the usual stop line, for a given equivalent turning volume is presented at the top of Figure 3 (see the other paper by Messer and Fambro in this Record). Shorter bay storage length results in saturation flow rates less than 1700. For normal urban street conditions, a taper length of 21.3 to 30.5 m (70 to 100 ft) may be considered appropriate; for higher types of urban facilities and rural highways, it should be 45.7 to 91.4 m (150 to 300 ft).

When a left turn bay is provided, any blockage effects that left turns may cause the through-right movement are calculated from

$$L_3 = (1700 - S)/[1700(N - 1) + S] \tag{10}$$

where N is the number of lanes serving the adjacent through-right movement.

When a separate right turn lane is provided, neither right turning volume nor right turn lane is analyzed. In other cases, the analysis of right turns depends on accuracy requirements. From the viewpoint of practicality and simplicity, the adjustment factor (R) for most designs can be set at zero (R equals zero in Equation 7 if right turns on red will be permitted). If a detailed analysis is desired, the following approach may be used to estimate the effects of right turning vehicles (4):

$$R = (5 \cdot P_R)/c \tag{11}$$

where P_R is the decimal fraction of movement combination turning right and c is the related curb return radius (m). In addition, the estimated number of vehicles turning right on red is subtracted from the through-right geometric movement volume combination (GMV_m). This estimate, which should not exceed 0.5 of the right turning volume, is calculated from

$$ROR = 50[P_c/(1 - P_c)] < 100 \tag{12}$$

where ROR is the estimated right-on-red volume (vehicles/h), and P_c is the estimated decimal fraction of traffic in the curb lane turning right.

Capacity

It is assumed throughout this procedure that the capacity of a normal protected through lane is 1750 automobiles/green·h (5). This is equivalent to a minimum average headway of 2.06 s/automobile. In addition, the normal protected left turn capacity is 1700 automobiles/green·h (5). The type of signalization also affects the capacity of the intersection and will be reflected in the critical lane analysis technique to follow.

Critical Lane Volumes for Each Street

To begin the analysis of a design's acceptability, the geometric design volume for each movement, calculated in Equation 6, is divided by the number of usable lanes provided to serve the movement to obtain a design volume per lane (V_m) of

$$V_m = GDV_m/N \tag{13}$$

If there is a left turn bay on an approach, then N equals 1 for the separated left turning movement, and N equals 1, 2, or 3 for the through-right movement, depending on the number of through lanes provided in the design. If no left turn lane is provided, then N would be the total number of approach lanes, since GDV_m is the total approach movement volume.

Figure 2 defined the movements at a typical intersection to be considered in the critical lane analysis technique. For each street, these movements may be combined in different patterns according to the type of signalization used, as shown here.

Signal Phasing	Maximum Lane Movements
One phase with no left turn bay	1 + 4 or 2 + 3
One phase with left turn bays	1, 2, 3, or 4
Two phases (no overlap) with bays	
Dual lefts leading	1 or 3 + 2 or 4
Dual lefts lagging	2 or 4 + 1 or 3
Leading left	1 or 4 + 2 or 3
Lagging left	2 or 3 + 1 or 4
Three phases (overlap) with bays	
Dual lefts leading	1 + 2 or 3 + 4
Dual lefts lagging	1 + 2 or 3 + 4
Leading left	1 + 2 or 3 + 4
Lagging left	1 + 2 or 3 + 4

For any given type of signalization on a street, one movement or a sum of two movements will be larger (critical). Different types of signal phasings will usually result in different critical lane volumes.

Level of Service Evaluation for Intersection

To evaluate the acceptability of the design, it is necessary to total the critical lane volumes (ΣV) for the intersecting streets.

$$\Sigma V = \Sigma V_{max}(\text{street A}) + \Sigma V_{max}(\text{street B}) \quad (14)$$

This sum is then compared with maximum values established for a given level of service as presented in Table 2. Here, level C is recommended for the design of urban signalized intersections. The maximum service volumes vary slightly depending on the signalization used: Two-phase signalization has no protected left turning on either street; three-phase has protected turning on only one of the two streets; multiphase signalization has protected left turning on both streets. A detailed description of the selection of total critical lane volumes relative to level of service criteria will be presented later.

Design Sample Problem

We shall now present a design alternative for the projected traffic. Some assumptions made in the problem were selected to illustrate computational procedures of the critical lane analysis design technique rather than optimum design practice.

Our purpose was to determine the acceptability of the proposed design, given the following: Traffic volume data are presented in equivalent automobiles/h for the intersection (step 4 of Figure 1); all lanes are 3.7 m (12 ft) long; all left turn storage lengths are 30.5 m (100 ft) except for 1B [66 m (150 ft)]; all right turn effects are ignored (R = 0.0); and both A and B streets have three phases with dual lefts leading (Figure 4).

These are the left turn adjustment factors:

$$L_{1A} = [(1700 \times 1.03)/1640] - 1.0 = 0.07$$

$$L_{1B} = [(1700 \times 1.03)/1700] - 1.0 = 0.03$$

$$L_{3A} = [(1700 \times 1.03)/1700] - 1.0 = 0.03$$

$$L_{3B} = [(1700 \times 1.03)/1700] - 1.0 = 0.03$$

$$L_{4A} = \{(1700 - 1640)/[(1700 \times 2) + 1640]\} = 0.012$$

Now, given the volumes adjusted according to Equation 6,

$$GDV_{1A} = 1.0 \times 1.0 \times 1.07 \times 131 = 140$$

$$GDV_{2A} = 1.2 \times 1.0 \times 1.0 \times 518 = 622$$

$$GDV_{3A} = 1.0 \times 1.0 \times 1.03 \times 46 = 47$$

$$GDV_{4A} = 1.2 \times 1.0 \times 1.012 \times 1017 = 1235$$

$$GDV_{1B} = 1.0 \times 1.0 \times 1.03 \times 151 = 156$$

$$GDV_{2B} = 1.1 \times 1.0 \times 1.0 \times 421 = 463$$

$$GDV_{3B} = 1.0 \times 1.0 \times 1.03 \times 85 = 88$$

$$GDV_{4B} = 1.1 \times 1.0 \times 1.0 \times 709 = 780$$

and the design volume per lane for each movement found by Equation 13,

$$V_{1A} = 140 \qquad V_{1B} = 156$$

$$V_{2A} = 207 \qquad V_{2B} = 232$$

$$V_{3A} = 47 \qquad V_{3B} = 88$$

$$V_{4A} = 412 \qquad V_{4B} = 390$$

The sum of the critical lane volumes (Figure 5) for street A will be

$$\left.\begin{array}{l} V_{3A} = 47 \\ V_{4A} = 412 \end{array}\right\} (47 + 412 > 140 + 207)$$

and those for street B will be

$$\left.\begin{array}{l} V_{3B} = 88 \\ V_{4B} = 390 \end{array}\right\} (88 + 390 > 156 + 232)$$

for the critical lane volume

$$\Sigma V = 937 < 1100$$

Therefore, this is an acceptable design for a service level B.

CRITICAL LANE VOLUME DESIGN CRITERIA

This section describes the theory of the critical lane analysis technique and operational measures used as a basis for selecting the design criteria of Table 2.

Effective Green

The effective green time is defined as that portion of the signal phase when saturation (capacity) flow occurs. Its duration is

$$g = G + Y - L \quad (15)$$

where g is effective green time (s); G is actual green (s); Y is yellow clearance (s); and L is lost time (4 s). Saturation flow conditions are assumed to begin in about 2 s after the start of green and to end about 2 s before the yellow clearance time expires. The total lost time per phase is 4 s in this paper. The effective green and actual green times are about the same if the yellow clearance interval is established according to basic intersection approach speed and width criteria.

Saturation Ratio

The saturation ratio of the signal phase (X) serving a movement could more descriptively be called the volume-to-capacity ratio, since

$$X = \text{volume/capacity} = Q/[(g/C) \cdot S] = (Q \cdot C)/(g \cdot S) \quad (16)$$

Table 1. Left turning equivalent for opposing ECV.

Intersection and Signal Phasing	Traffic Movement	Number of Opposing Through Lanes	Factor by Opposing ECV[a] 200	400	600	800	1000
No protected turn phase							
Two-phase							
No bay on approach	Left and through	1	2.0	3.3	6.5	16.0[b]	16.0[b]
	Left and through	2	1.9	2.6	3.6	6.0	16.0[b]
	Left and through	3	1.8	2.5	3.4	4.5	6.0
With bay on approach	Left	1	1.7	2.6	4.7	10.4[b]	10.4[b]
	Left	2	1.6	2.2	2.9	4.1	6.2
	Left	3	1.6	2.1	2.8	3.6	4.8
Three-phase							
No bay on approach	Left and through	1	2.2	4.5	11.0[b]	11.0[b]	11.0[b]
	Left and through	2	2.0	3.1	4.7	11.0[b]	11.0[b]
	Left and through	3	2.0	2.9	4.2	6.0	11.0[b]
With bay on approach	Left	1	1.8	3.3	8.2[b]	8.2[b]	8.2[b]
	Left	2	1.7	2.4	3.6	5.9	8.2[b]
	Left	3	1.7	2.4	3.3	4.6	6.8
Protected turning							
No bay	Left	Any	1.2		1.2		1.2
With bay	Left	Any	1.03		1.03		1.03

[a] In automobiles per hour. Includes total through volume (from step 4, Figure 1) on the approach opposing the left turn being analyzed. The opposing volume also includes any turning volume (left or right or both) for which no separate turning lane is provided.
[b] Turning capacity only at end of phase. Not recommended for design. Add additional through lane, turning lane, or protected left turn phasing.

Figure 3. Saturation flow of left turn phase as a function of bay storage length and turning volume.

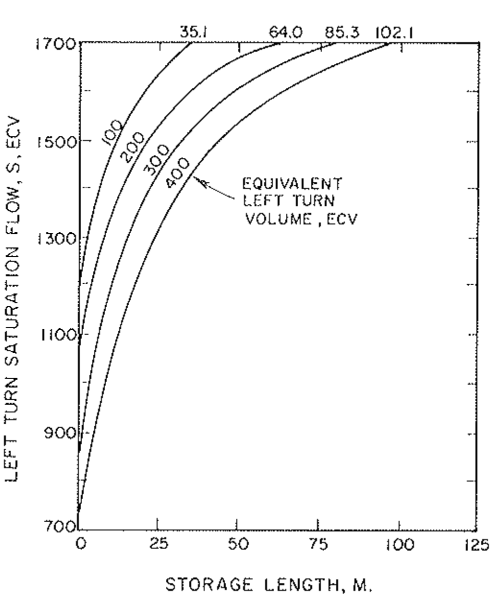


Figure 4. Volume data and proposed design of example.

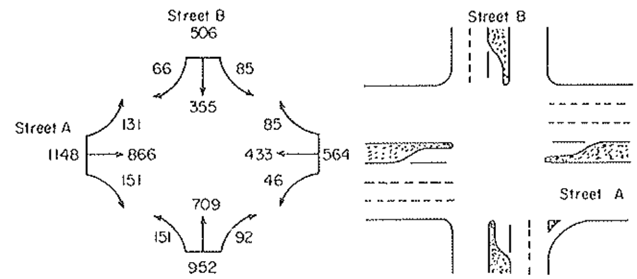


Figure 5. Design lane volumes for streets A and B.

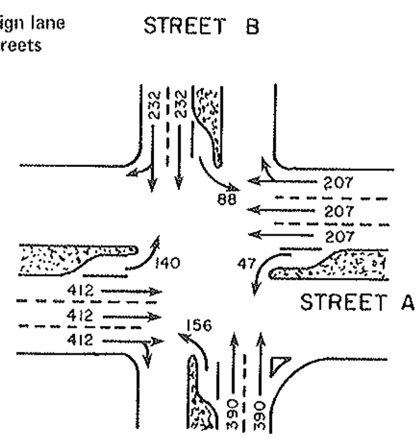


Table 2. Level of service and maximum sum of critical lane volumes at signalized intersections.

Level of Service	Traffic Flow Condition	Volume to Capacity Ratio	Critical Lane Volumes Two-Phase	Three-Phase	Multiphase
A	Stable	$\leq$0.6	900	855	825
B	Stable	$\leq$0.7	1050	1000	965
C	Stable	$\leq$0.8	1200	1140	1100
D	Unstable	$\leq$0.85[a]	1275	1200	1175
E	Capacity	$\leq$1.0	1500	1425	1375

[a] Modified (1).

where Q is approach movement volume (vehicles/h); C is cycle length (s); and S is saturation flow of approach (vehicles/green· h).

The saturation ratio is a good tool for describing quantitatively what traffic operating conditions will be like. When $X > 0.85$, vehicle delays on the approach become very long, and queues frequently fail to clear the approach at the end of the green phase.

Critical Lane Development

A normal, four-legged intersection having multiphased signalization will have four critical volume phases, as was shown in Figure 2. The sum of these four critical phases, including green plus yellow times, is one cycle. The equation for this requirement is

$$C = \phi_{A1} + \phi_{A2} + \phi_{B1} + \phi_{B2} \quad (17)$$

or

$$C = (G + Y)_{A1} + (G + Y)_{A2} + (G + Y)_{B1} + (G + Y)_{B2} \quad (18)$$

where ϕ_{A1} is the first critical phase on street A (s).

Substituting the equivalent effective green (g) and lost time (L) from Equation 15 into Equation 18 for each critical phase and rearranging terms yield

$$g_{A1} + g_{A2} + g_{B1} + g_{B2} = C - nL \quad (19)$$

where n is the number of critical phases per cycle. Solving for g in the saturation ratio formula (Equation 16), substituting in Equation 19, and dividing by C results in

$$(C - nL)/C = [Q/(X \cdot S_{A1})] + [Q/(X \cdot S_{A2})] + [Q/(X \cdot S_{B1})] + [Q/(X \cdot S_{B2})] \quad (20)$$

If the saturation ratios are selected to be the same for all critical phases, then

$$(C - nL)/C = (1/X)[(Q/S_{A1}) + (Q/S_{A2}) + (Q/S_{B1}) + (Q/S_{B2})] \quad (21)$$

or

$$[X \cdot (C - nL)/C] = Q/S_{A1} + Q/S_{A2} + Q/S_{B1} + Q/S_{B2} \quad (22)$$

Since the ratios Q to S are the same on a per lane basis as for the total approach movement and if $Q/N = V$ and $S/N = 1750$ according to the critical lane analysis procedure, then

$$X \cdot [(C - nL)/C] = (V_{A1}/1750) + (V_{A2}/1750) + (V_{B1}/1750) + (V_{B2}/1750) \quad (23)$$

where V_{A1} is the critical lane volume on phase A_1. Thus,

$$V_{A1} + V_{A2} + V_{B1} + V_{B2} = 1750 \cdot X \cdot [(C - nL)/C] \quad (24)$$

It follows that the sum of the critical lane volumes (ΣV) for a given level of service saturation ratio ($X_{L.O.S.}$) and design cycle length (C) is

$$\Sigma V = 1750 \cdot X_{L.O.S.} \cdot [(C - nL)/C] \quad (25)$$

The results of Equation 25 are plotted in Figure 6 for two X ratios: 1.0 (capacity) and 0.8. Also shown in Figure 6 are desirable design hour cycle lengths: 55 s for two-phase, 65 s for three-phase, and 75 s for four-phase. These lengths were used to determine the capacity values and other sums of critical lane service volumes in Table 2.

Minimum Delay Cycle Length

The average vehicle delay (d) in seconds per vehicle experienced by an approach movement having Poisson arrivals can be calculated by using Webster's rather lengthy formula:

$$d = \{[C(1 - g/C)^2]/[2(1 - Q/S)]\} + \{(1800 \cdot X^2)/[Q(1 - X)]\} - 0.65\,[C/(Q/3600)^2]^{1/3}\, X^{2+5(g/C)} \quad (26)$$

To illustrate average delay conditions at an intersection, assume that a four-phase signalized intersection has equal approach volumes and green times. The four critical lane volumes are also equal, but their magnitude can change. Figure 7 shows how the average delay varies with cycle length for each of the volume conditions, as given by the sum of the critical lane volumes. The sums of the critical lane volumes of 825, 1100, and 1175 correspond to levels of service A, C, and D respectively in Table 2 for multiphase (four) signals. The minimum delay cycle lengths are 55, 78, and 88 s for the sums of critical lane volumes of 825 (A), 1100 (C), and 1175 (D) respectively. Multiphase (four) intersections designed to level C could operate at almost any normal range of peak-hour cycle lengths (60 to 90 s). However, intersections having a sum of critical lane volumes of 1175 (D) will be heavily congested at cycle lengths less than 70 s.

Figure 8 summarizes the relationships between the design variables and operational measures of (a) sum of critical lane valumes, (b) saturation ratio, (c) minimum delay cycle length, (d) type of signal phasing, and (e) level C design criteria. The latter result in the minimum delay cycle lengths desirable for operating coordinated arterial signal systems during the peak hours.

Other Operational Variables

There are several other traffic operational measures besides delay that are descriptive of the traffic conditions existing on a movement at a signalized intersection. Load factor was used in the Highway Capacity Manual (2) to define level of service; traffic engineers have also used Poisson's probability of cycle failure (6); and Miller recently extended the probability of queue failure concept to include the effects of queue spillover from one cycle to the next as queued vehicles fail to clear during the green (7). A summary of these new models developed by Miller follows.

$$\text{Probability of queue failure} = e^{-1.58\phi} \quad (27)$$

$$\text{Load factor} = e^{-1.3\phi} \quad (28)$$

$$\text{Probability of queue clearing} = 1 - e^{-1.58\phi} \quad (29)$$

where $\phi = [(1 - X)/X]\,[(S \cdot g)/3600]^{1/2}$.

Figure 9 presents average approach values at an intersection having four critical phases for these operational measures together with Poisson's probability of failure and saturation ratio as a function of minimum delay cycle length. When 75 s is the minimum delay cycle length, the saturation ratio (X) is 0.78; the load factor (LF) is 0.36; Miller's probability of queue failure (P_f) is 0.30; and Poisson's probability of failure is 0.18 (18 percent).

Figure 10 illustrates operating characteristics on a single approach movement as volume conditions increase. The effective green time and cycle length were held constant at 35 and 70 s respectively. Few cycles fail to clear stopped queues at saturation ratios less than 0.7. Acceptable queue failure rates still exist at 0.8, but larger values result in unstable operation and increasingly long delays.

ACKNOWLEDGMENTS

This paper discusses one phase of the research project, Effects of Design on Operational Performance of Signal Systems, conducted by the Texas Transportation Institute and sponsored by the Texas State Department of Highways and Public Transportation in cooperation with the Federal Highway Administration, U.S. Department of Transportation. We wish to thank Harold D. Cooner and Herman E. Haenel of the Texas State Department of Highways and Public Transportation for their technical inputs and constructive suggestions throughout this project. The assistance of Donald A. Andersen, Don A. Ader, Donald R. Hatcher, and Murray A. Crutcher of the Texas Transportation Institute is also appreciated. We also wish to acknowledge the Australian Road Capacity Guide from which several of the concepts described in this paper were drawn. The contents of this paper reflect our views, and not the official views

Figure 6. Variation in sum of critical lane volumes with cycle length as a function of saturation ratio.

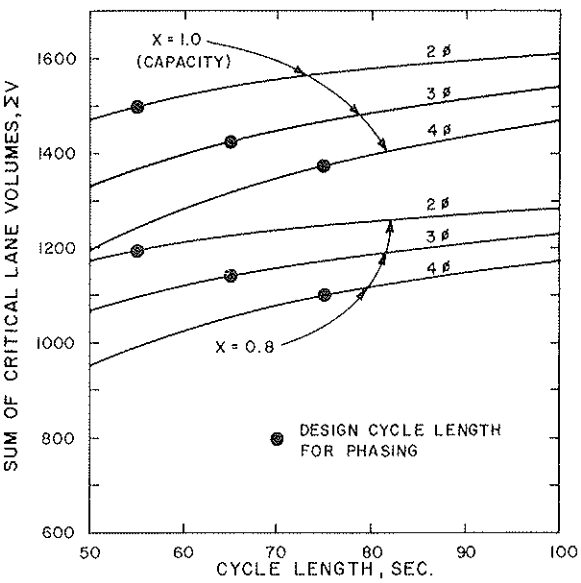


Figure 7. Delay versus cycle length as a function of sum of critical lane volumes.

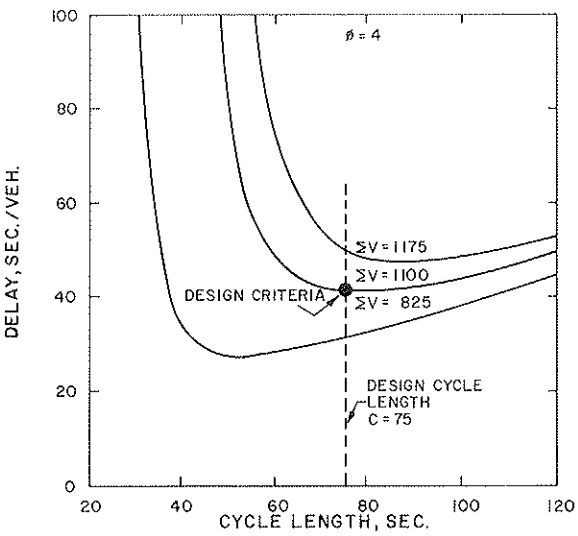


Figure 8. Relationship between design variables and operational measures.

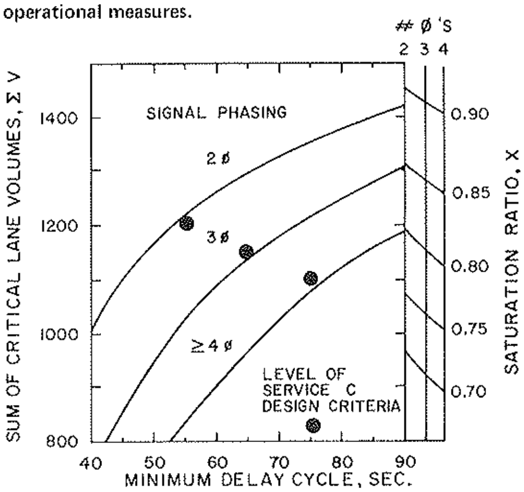


Figure 9. Relationship between operational measures of effectiveness and minimum delay cycle length.

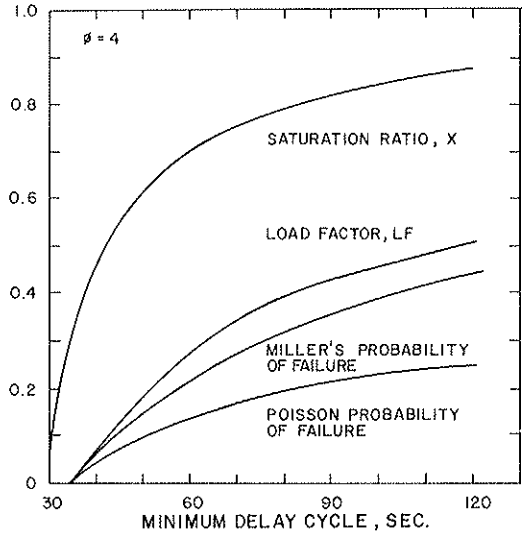
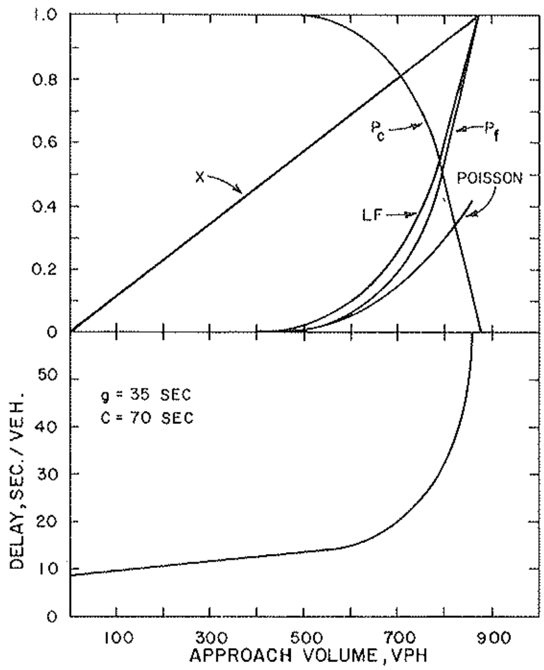


Figure 10. Relationship between operational measures of effectiveness and volume.



or policies of the Federal Highway Administration. We alone are responsible for the facts and accuracy of the data presented. This paper does not consitute a standard, specification, or regulation.

REFERENCES

1. C. J. Messer, D. B. Fambro, and D. A. Andersen. A Study of the Effects of Design and Operational Performance of Signal Systems. Texas Transportation Institute, Res. Rept. 203-2F, Aug. 1975.
2. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
3. L. J. Pignataro. Traffic Engineering Theory and Practice. Prentice-Hall, Englewood Cliffs, N.J., 1973.

4. A. D. May. Intersection Capacity 1974: An Overview. TRB, Special Rept. 153, 1975, pp. 50-59.
5. D. J. Berry. Capacity and Quality of Service of Arterial Street Intersections. Texas Transportation Institute, Res. Rept. 30-1, Aug. 1974.
6. J. E. Tidwell and J. B. Humphreys. Relation of Signalized Intersection Level of Service to Failure Rate and Delay. HRB, Highway Research Record 321, 1970, pp. 16-32.
7. A. J. Miller. On the Australian Road Capacity Guide. HRB, Highway Research Record 289, 1969, pp. 1-13.

Discussion

Stephen G. Petersen, Gaithersburg, Maryland

In the federal government, the phrase "adequate public facilities" is becoming commonplace. This phrase, when directed at builders and developers, tells them that their projects can only go forward if all the public facilities required to serve a development are in place and capable of meeting the needs of a specific project. Many facilities are included: fire protection, police protection, schools (if a residential project), water, sewer, and finally transportation but more specifically roads. The test of adequacy is generally easily met except for two facilities, sewers and roads. Usually both are impacted; but if sewers are not, then roads are almost certain to be a problem for any development lying in the developed urban area.

As a test of road adequacy, the two Maryland suburban counties of Washington, D.C., have generally adopted a technique known as critical movement analysis or critical lane analysis that was based on a paper of mine (8). The technique was selected as a planning tool because it is relatively simple to apply, avoids the correction factors common to the Highway Capacity Manual (HCM), and derives an answer relevant to the whole intersection in one step rather than as part of a multistep process.

A comparison of the techniques described in the authors' paper and in mine reveal the same basic elements. The authors, however, describe a more sophisticated set of correction factors for a more accurate result. It is encouraging to see this kind of effort, because, if we are going to place an expensive development at the mercy of one or two intersections near it, then we had better be sure that our techniques are reasonable representations of real-life conditions.

I had several questions and comments, which are listed below.

1. The source cited for volume adjustment for the lane utilization factor is work done in New Jersey in the early 1960s. However, it is understood that these values were confirmed by field data collected in a study of unprotected left turns at opposing flows of up to 1000 vehicles/h. What happens beyond this volume? Does lane utilization become essentially even at lower levels of service (higher approach volumes)? If so, does this factor need to become a variable? This is a question for the local area because a difference shows up between the data published in the paper and the technique as applied and can make a difference as great as one level of service.

2. Table 1 deals with opposing volumes of only 1000 vehicles/h. Presumably left turns in opposition to volumes in excess of this can only be accommodated by protected turning. However, this is not stated, and for opposing traffic in multiple lanes the factors have not reached turning capacity. Is the number of approach lanes independent of the opposing volume? Also, it is not clear which phasing is in effect when the term "three-phase" signal is used under the No Protected Turn Phase.

3. I would like to see a level of service comparison between this technique and that used in the HCM.

4. I attempted to work through this method with a more heavily loaded, less sophisticated intersection and found that one must be extremely careful to follow the phasing rules for critical lane volumes explicitly for each street or the result of the calculations will be totally erroneous. This need to carefully select signal phasing and to apply factors detracts from the use of the technique as a planning tool. I recognize that the intent of the research was for design purposes, but I hope it can be adapted with some simplification as a planning tool. In planning, the critical lane technique is a valuable communication tool among various professionals, but its effectiveness is lost when correction factors and selection of different phasing arrangements transform too many steps into variables.

5. The authors state, but I emphasize again, that the critical lane values in Table 2 are the result of deliberately selecting cycle lengths that produce less than maximum capacity even though Webster says they minimize delay. The effects of this show up when one tries to compare the level of service determined by field examination with that predicted by the method. In an example in which such a comparison was possible, the technique showed level of service E, whereas the description of the actual condition was closer to level D. This suggests that there is a need either to more carefully relate the description of level of service to the results of the analysis or, more likely, to quantify level of service to some measure such as delay in order to relate critical lane totals to specific measures of total intersection delay.

6. Also, in reviewing the design volume discussion, I notice the use of the peak 15 min for design purposes. It is possible that this factor has distorted comparisons between field observations that deal with a total hour of flow and critical lane totals that are keyed to a peak 15-min period.

In summary, I believe the authors have assimilated a considerable amount of useful information toward the development of improved intersection design techniques that deal with the whole intersection and the interaction within it. The technique now needs to be expanded to account for maximum practical cycle lengths, for other intersection configurations, and for situations controlled by STOP signs and not likely to be signalized. Particular emphasis needs to be place on these aspects in planning situations such as that described at the beginning of these comments.

REFERENCE

8. H. B. McInerney and S. G. Petersen. Intersection Capacity Measurement Through Critical Movement Summations: A Planning Tool. Traffic Engineering, Jan. 1971, pp. 45-51.

Frederick A. Wagner, Alan M. Voorhees and Associates, Inc.

The authors are to be commended for their excellect paper. It synthesizes a substantial range of knowledge on intersection capacity and level of service in a systematic manner. It will guide the traffic engineer through a step-by-step procedure for determining critical lane volumes and corresponding levels of service at an intersection, given the intersection design, traffic control, and traffic flow characteristics. Critical lane volume is inherently more appealing and more fundamentally valid than the established intersection capacity analysis techniques set forth in the Highway Capacity Manual.

This discussion addresses several points: (a) potential refinements in certain detailed aspects of the technique, (b) the need to develop a way to best use the technique for existing operational intersections (in addition to design intersections), (c) the possible invalidity of equations based on lost time and cycle length relationships, and (d) the need for systematic empirical validation of this technique and alternative methods of intersection capacity and level of service analysis.

POTENTIAL REFINEMENTS

One of the appealing attributes of the critical lane analysis technique is its relative simplicity. Certain aspects of the technique, however, border on oversimplification and result in the risk of losing precision. I shall give a few examples (not exhaustive) in which refinements might make the analysis technique more precise without significantly increasing its complexity.

Lane Width Adjustment

Messer and Fambro consider only two classes of lane width: < 3 and ≥ 3 m (10 ft). A volume adjustment factor of 1.10 is applied if lane width is less than 3 m (10 ft). This is perhaps their most serious oversimplification. All other major (American, Australian, and British) capacity analysis techniques utilize a more refined adjustment for lane width (or total approach width). The Australian factors for adjusting saturation flow are given below (1 m = 3.3 ft).

Lane Width (m)	Adjustment (%)	Lane Width (m)	Adjustment (%)
2.4	-12	4.0	+3
2.7	-7	4.3	+4.5
3.0 to 3.7	0	4.6	+6

The reciprocals of these factors would be used to adjust volumes. However, one unresolved debate concerns the use of number of lanes or the use of approach width. Based on a questionnaire survey in 1974 by the TRB Committee on Highway Capacity, approach width versus number of lanes was the number one priority research problem.

Right Turn Adjustment (Pedestrian and Parking)

The Messer-Fambro technique computes a right turn adjustment as a function of only two variables: fraction of geometric movement volume turning right and curb return radius. Two other factors seem to be extremely important. First, pedestrians on the intersection crosswalks create a major interference to right turning vehicles and in downtown areas especially can make right turns as difficult as left turns. Second, parking conditions on the approach will also significantly influence the effect of right turns. The authors do not consider parking except to say that "the width of a lane may not include any pavement used for or appreciably affected by parking." Parking prohibited for a short distance back from the stop lane provides extra approach space for storage and movement of right turn vehicles.

The British developed a method for determining effective loss of approach width as a function of clear distance from the stop line to the nearest parked automobile and green time. An analogous approach might be incorporated into the critical lane analysis technique.

Grade

Although the great majority of intersection approaches are relatively flat, those on even moderate grades have significantly altered capacity characteristics. A simple volume adjustment factor could be incorporated for positive and negative grades. Webster found that for each 1 percent of upgrade or downgrade saturation flow rate decreases or increases by 3 percent.

Cycle Length and Left Turns

On intersection approaches with heavy opposing flow, unprotected left turns are made difficult by few gaps. Most left turns are made at the end of the green phase (during yellow). Messer and Fambro give left turn equivalency factors that cover such situations, but they do not consider how drastically cycle length affects left turn capacity. For example, if virtually all left turns are made during yellow intervals, left turn capacity will be 50 percent higher with a 60-s cycle (60 yellows/h) than with a 90-s cycle (40 yellows/h). This fact should be reflected in the technique, especially considering the propensity among traffic engineers to use only long signal cycles. (More on this later.)

APPLICATION AT EXISTING INTERSECTIONS

The authors indicate that their paper is intended as a guide for designing signalized intersections. The techniques could also be adapted to evaluating capacity and level of service at existing intersections. After all, most traffic engineers devote far more time to the evaluation of operation changes at existing intersections than to the physical redesign of intersections or the design of entirely new ones.

We should emphasize the procedures that directly measure capacity flow rates associated with each geometric movement volume rather than rely on estimates based on generalized adjustment factors. Most intersections have at least a few measurable loaded cycles during peak periods. Direct measurement of a relatively small number of loaded cycles may provide a more precise estimate of intersection capacity than the various adjustment factors do. There is an opportunity here to develop a unified approach that incorporates both direct measurement, where possible, and utilization of generalized volume adjustment factors only where necessary. For example, a redesign of a problem intersection may only involve changing one or two intersection approaches. An analysis of the redesign could be based on direct empirical data for the unchanged approaches and could utilize the generalized adjustment factor technique only for those approaches having substantially changed designs.

POSSIBLE INVALIDITY OF BASIC EQUATIONS

The various equations set forth in the authors' paper for determining level of service, average delay, minimum delay cycle length, probability of queue failure, load factor, and probability of queue clearing, all evolve from Webster's original work. He divided the green-plus-yellow phase for a given intersection approach into effective green time (when traffic discharges at a constant saturation flow rate) and lost time (when zero flow occurs).

The authors' Equation 25 gives the sum of critical lane volumes (ΣV) as a function of basic saturation flow rate (1750 vehicles/h), volume to capacity ratio ($X_{L.O.S.}$), cycle length (C), lost time (L), and number of phases (n). The factor nL/C is a crucial one. It represents the fraction that total lost time on all phases is of cycle length. Implicitly, Equation 25 and others using nL/C show that conditions will improve with longer cycle length, because lost time will then represent a smaller fraction of total time.

The equations are theoretically correct, but are they valid empirically? In 1974 Moskowitz argued eloquently:

> Green plus yellow is a better predictor of volume during saturated phases than green alone (or green plus yellow minus lost).
>
> Observations at intersection approaches where cycle length was modified do not show that higher volumes can be served if longer cycle lengths are used.
>
> If the cycle length is long enough to provide optimum splits, given minimum green constraints, short cycles almost always result in better service (loss delay) than long cycles.
>
> The habit of using long cycles causes more unnecessary delay to Americans than any other single fallacy in the traffic engineering profession.
>
> Green plus yellow (with no correction for lost time) should be used in measuring saturation flow and in computing capacity values. This would cause traffic engineers to re-think the fallacious habit of using long cycles in the attempt to increase throughput.

Moskowitz explained that on multilane approaches it is almost impossible for saturated flow to last throughout a long green interval on all lanes of the approach. Messer and Fambro account for underutilized lanes by applying a correction factor for multilane approaches. However, their adjustment factors ignore cycle length. If total approach really is less productively used as green time increases, it might be possible to develop adjustment factors that consider both number of lanes and green phase duration.

A serious question needs to be answered: Are the many theoretical equations that depend on full saturation flow during the effective green period to explain traffic flow at signalized intersections valid when compared with actual traffic operation?

SYSTEMATIC VALIDATION NEEDED

Carefully designed experiments are needed to validate the basic theoretical equations of intersection operation and to compare current alternative methods for evaluating capacity and level of service (HCM method, British saturation flow method, Australian method, critical lane volume analysis, multiple regression equations developed in Canada).

These experiments should stress sensitivity tests in which conditions would be changed at field intersections and the resulting changes in capacity, if any, and other operating characteristics would be compared with changes predicted by the theoretical methods. Only through this type of scientific experimentation will it be possible to determine the validity of the various approaches and to isolate the weaknesses of current methods.

Authors' Closure

We would like to express our appreciation to Petersen and Wagner for their discussions. Their comments from the perspective of planning and operation engineers point out the different levels of accuracy required in transportation planning, design, and operations.

Petersen questions the values of our utilization factors and their relationship to approach volume. These values are representative of the moderate to high volume conditions common to intersection design. The user may wish to follow these lane utilization formulas for increased accuracy:

$$U_2 = 1.1 + 0.9\ e^{-0.18m} \quad (30)$$

for two lanes, and

$$U_3 = 1.2 + 1.8\ e^{-0.13m} \quad (31)$$

for three lanes, where m is average approach arrivals per cycle (3).

Table 1, very important but slightly complicated, was developed according to a cycle length of 70 s in all cases. Two- and three-phase equivalents were based on G/C ratios of 0.51 and 0.36 respectively. With these splits, left turning capacity is at or near zero at opposing volumes of 1000 vehicles/h except for three-lane approaches (4). The user may opt for an arbitrary maximum limit on the left turn equivalent factor (E) in Table 1 and use this value for all heavier volume cases (for instance, E = 11.0) (3, p. 234).

Petersen notes the need for care during the selection of signal phases. If an approach does not have a separate left turn bay (lane), then 1 + 4 or 2 + 3 must be used.

Wagner suggests general potential refinements for the analysis technique. In particular, the user may wish to use the lane width adjustments presented by Wagner with the following conversions: 1.14, 1.08, 1.00, 0.97, 0.96, 0.94. The effects of pedestrians, parking, and grades were not included in the design procedure. Users may also wish to include these.

Wagner expresses his personal concern about the overall concept of phase lost time and cites Moskowitz's paper, which we had also reviewed. We believe the equations provide practical operational solutions using a lost time of 4 s/phase.

Several research projects on capacity will be conducted in the National Cooperative Highway Research Program and by the Federal Highway Administration during the next few years. We hope this paper will aid in the development of practical intersection capacity analysis techniques for planning, design, and operation needs.

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

Review of Road Traffic Network Simulation Models

Paul Ross, Office of Research, and David Gibson, Office of Development,
Federal Highway Administration

Proposals for computer traffic simulation date from 1951, and the first actual documented simulation was performed in 1955 on an analog computer. Since that time, many simulation models have been developed as knowledge of traffic principles has increased and computer facilities have improved. Such traffic simulation models may be placed in three groups: single road, single intersection, and traffic network. This paper discusses only the last. Nineteen network simulation models are discussed. These are grouped into ten obsolete models, six traffic network models suitable for current computers, and the simulation portions of three signal optimization programs. A brief description of the operating principles and unique features of each model is given, and the level of modeling detail, model flexibility, and usefulness of the output are assessed. Validation efforts on the model are considered, and, where known, computer language, type of machine, core requirements, and speed of execution are given.

There are three classes of traffic simulation models: single road, single intersection, and network. Hsu and Munjal (2) have reviewed the single road freeway models, but rural road models, which reproduce the effects of sight distance restrictions, rolling terrain, and passing opportunities, are rare. Single intersection models have generally been built for a specific purpose and are not usually available as a consumer product. This paper discusses the currently available network simulation models and some now obsolete network models. We also summarize three signal optimization programs that include traffic network simulation.

Modern traffic simulation programs use digital computers and time-oriented bookkeeping; the event-oriented bookkeeping associated with most simulation languages seems inappropriate.

Some models (macroscopic) treat the traffic stream as a continuum and conceptualize the traffic stream as a fluid. Individual vehicles are not identified in macroscopic models; indeed, a network element may even contain fractional vehicles. Macroscopic models generally are very economical of computer storage and very fast in execution, because they do not have to keep track of each vehicle. However, they cannot represent the traffic stream in the detail that many traffic engineers want.

Platoon models are a half-step toward detailed realism. In these models, vehicles are grouped into platoons, and the program keeps track of the location, speed, acceleration, and size of the platoon. Platoon speed, usually a function only of the general density of vehicles in the vicinity of the platoon, avoids complicated car-following calculations.

Microscopic simulation models are the ultimate in detailed treatment. Each vehicle is identified and its position, speed, acceleration, and other characteristics are stored. Microscopic models deal with two separate areas: streets and intersections. On streets, vehicle behavior is usually approximated by controlling one vehicle's speed by the preceding vehicle (car following). Car-following rules can be quite complex. Queuing behavior and the delay in accelerating away from the head of the queue are usually treated in detail; there are indications that this can be a critical factor in the accuracy of simulation, at least for urban networks. Lane changing may or may not be allowed, and other details of street traffic can be represented. Buses and trucks can also be handled separately.

Intersection behavior is considerably more complex and difficult to model. Problems of pedestrian interference, turning radii, and collision avoidance all must be faced. Usually these problems are treated superficially; for example, random delay is assigned for pedestrian interference. Because one primary function of network models is to evaluate signal control strategies, traffic signals are usually modeled in considerable detail.

The various simulation models (see Table 1) discussed here should be assumed, unless specifically indicated otherwise, to have the following characteristics. They are fully microscopic, meaning that the model stores a specific location, speed, and acceleration for each individual vehicle. Moreover, the vehicles engage in car following; that is, the speed of a vehicle is determined in some reasonable, predetermined manner by the speed of any vehicle directly in front of it. If there is no vehicle in front, it travels at its "desired" speed. Vehicles obey traffic signals and stop and yield signs.

All variables in the system (vehicle locations, speeds, accelerations, and signal indications) are updated once each time step, which is a user-specified constant, often 1 s. The computer program is written in FORTRAN. The simulation starts collecting data only after some initial period, which starts with the system empty and ends when some test for equilibrium is satisfied.

Vehicle routes are determined from origin-destination tables (O-D) or the probability of turning left, through, or right at the end of each link (turning movements).

OBSOLETE MODELS

The *DISCRETE AND CONTINUOUS VARIABLE SIMULATION MODEL*, developed by Matthewson, Trautman, and Gerlough, was perhaps the first traffic simulation model to actually be constructed (3, 4). It represented a signalized intersection network with one lane in each direction and pedestrians who delayed turning traffic. Validation was primarily by reasonableness. The model was based on an analog rather than a digital computer and therefore had to be rebuilt for each new intersection or network.

The *STARK/NBS MODEL*, produced by Stark of the National Bureau of Standards for the IBM 704 computer (5), achieved realism through the use of fairly detailed traffic behavior rules. The model included traffic signals, stop signs, one- and two-way streets, and oblique intersections within the program structure. Documentation for this model is hard to obtain and would not be sufficient to recreate it for use on another computer. The model provided complete statistics on vehicle behavior. The most interesting output was a movie of the simulation, which was the only verification of the realism of the model.

A simulation model was developed by Francis and Lott for the *ROAD RESEARCH LABORATORY* to examine proposed changes in traffic signal patterns in central London (6). This program had a moderate level of detail but did not distinguish between different types of

Table 1. Characteristics of network traffic simulation programs.

Model	Type	Time Step	No. of Vehicle Types	Route Selection	Language, Machine	Time Compression	Remarks
Discrete and continuous variable	Analog	Continuous	1	—	Hardwired machine	1:1	Electronic elements: one-to-one correspondence with network
Stark/NBS	Micro	¼ s	3	Fixed during vehicle generation	Assembly, IBM 704	15:1 (9 nodes)	First to use cartoon movie graphics as output medium
Road Research Laboratory	Micro	1 s	1	Turn %	Assembly, Ferranti Pegasus	2:1 (36 links, 9 nodes)	0, 1 representation of vehicle location
Australian or Pak-Poy	Micro	—	1	Turn %	IBM 1620 and 7090	2000:1 (1 node, 4 links) IBM 7090	
VTS	Micro	Event oriented	1	Turn %	GPSS, IBM 7090	1:1 (10 nodes) 1:3.5 (25 nodes)	
TRANS	Short platoons	2 s	1	Turn %	Assembly, IBM 709	20:1 (36 links, 9 nodes)	Trucks diffused uniformly through traffic
Hartley	Hardware	⅓ s	1	Fixed	Atlas traffic simulator	NA	
Birmingham	Micro	—	4	Fixed routes	EGTRAN 3, ICL	4:1 (3 nodes)	
DYNET	Micro	1 s	2	Turn %	FORTRAN, IBM 360	5-10:1 (46 links, 21 nodes)	
Sakai and Nagao	Large platoons	4 s	—	Turn %	NA	Very fast	All vehicles in a 50-m segment act as platoon
VETRAS	Micro	Event oriented	1	Turn %	GPSS, IBM 360/370	Slow	
VPT	Micro	Variable (1 s)	3	Turn %	FORTRAN and COMPASS, CDC 7600	Very slow (1:1 or slower)	Very detailed
UTCS-1	Micro	1 s	3	Turn %	FORTRAN, IBM 360, CDC 7600	2.5:1 (26 links, 12 nodes)	
SCOT	Platoon	Variable (6 s)	3				
SIGNET	Micro	Variable (1-s multiples)	2	Turn %	FORTRAN IV, CDC 6500	6.5:1	Traffic signals only, no stop or yield
Microassignment	Macro	Static	1	O-D	FORTRAN and Assembly, IBM 360	5:1 (1334 links)	
CORQ	Macro	Static	1	O-D	FORTRAN, IBM 360	30:1 (125 links, 65 nodes)	Carries excess demand over to next assignment
TRANSYT	Macro	Static	5	Turn %	FORTRAN	Very fast	Single-pass calculation of pulse shapes
SIGOP II	Macro	Static	1	Turn %	FORTRAN	Function of number of iterations desired	User-selected, multiple-pass calculation of pulse shapes
CORQ1C	Macro	Static	2	Both	FORTRAN, CDC	Very fast	Can generate O-D

vehicles or follow them through the network. Each lane was divided into short sections that corresponded to a computer storage "bit." For each occupied section, a "one" would be registered in the storage location. Ones moved from bit to bit in the computer memory to represent vehicle motion. This model simulated only signalized intersections of three or four legs. However, separate specification of right- and left-turning volumes and signal offsets and phasing was possible. Output included total time simulated, volumes and delays by link, average delays and average queues by link, and maximum queues by link. The maximum network size was 30 intersections, 80 links, and 20 peripheral arms. The program was executed at approximately 18/n times real time, where n was the number of intersections.

The *AUSTRALIAN OR PAK-POY MODEL* (7) was written to analyze traffic signal controllers and intersection capacity and was reasonably flexible. Validation was limited and consisted primarily of comparison of model results with before-and-after field studies of intersections. Documentation is unavailable. The primary outputs of the model were delay, queue length, and degree of saturation.

VEHICLE TRAFFIC SIMULATOR (VTS) was constructed by A. H. Blum (8, 9) and used general purpose simulation system (GPSS) language intersection modules assembled to form a network. The level of detail in VTS was good; the model simulated traffic signals and special sources and sinks of traffic. Fixed time signals were assumed to control all intersections, and the car-following logic was "change lane or assume speed of leading vehicle." Vehicles were not identified by type, but different types could be approximated by assigning vehicle characteristics. No validation was performed on this model and documentation is limited to one article (8).

The *TRANS* (simply a name) *MODEL* developed by Gerlough, Wagner, Rudder, and Katz, was the first network simulation model to enjoy wide usage and validation. At least four versions were produced from 1962 to 1968 (10, 11, 12, 13, 14). Although TRANS is now obsolete, it led to several other models (see the DYNET, SIGNET, and UTCS-1 models below). The level of detail in the TRANS model was moderate. Automobiles were grouped into short platoons. Vehicles could switch lanes instantaneously if they were queued behind left-turning vehicles. Vehicle behavior at signals was quite detailed. Different models of TRANS used different time steps; version 4 had a user-specified updated period that could not be less than 2 s. The TRANS model was the first to be widely validated, and validation runs were made in Washington, D.C., Los Angeles, and Detroit. The documentation explained fairly clearly how to prepare input data, but documentation of program logic was scanty. Data preparation was tedious, because many parameters had to be supplied. Output gave the usual traffic parameters both link by link and networkwide.

HARTLEY, in England, developed a basic model utilizing a special purpose computer identified as the Atlas Digital Traffic Simulator. A diffuser was used to link intersections and replicate platoon behavior, while a random pulse generator created input traffic. Because of its fixed program, the model was relatively easy to run but had very limited flexibility. Validation was by sensitivity analysis and reasonableness. Documentation was limited to two articles and a doctoral dissertation (15, 16, 17).

The *BIRMINGHAM MODEL*, also developed in England, by Storey (18), handled four categories of vehicles. Car-following logic was limited, and vehicles were assigned to their lanes based on their turning movement at the subsequent intersection. An unusual feature of this model was the assignment, to avoid path looping, of complete vehicle routes from entry through the network. The model could represent both fixed-time and vehicle-actuated signals. Output was limited to printouts of vehicle location within the network and both journey times and delays for each route. Although the model was not validated, sensitivity testing indicated that it produced reasonable variations in output values based on input changes.

The *DYNAMIC NETWORK ANALYSIS OF URBAN TRAFFIC FLOW (DYNET) SIMULATION MODEL* was developed by Lieberman of General Applied Science Laboratories (GASL) in 1969 (19). It was based on the TRANS model, used many of its features, and made numerous specific improvements. Particular effort was expended on making the model fully microscopic and simplifying the input requirements. Trucks and automobiles were represented separately and were generated according to a shifted exponential distribution. A vehicle was assigned a lane when it first entered a link according to the turn the vehicle expected to make when it exited from that link. A vehicle could change lanes when it entered a queue if the queue in the adjoining lane were shorter and the turning movement permissible from the new lane; the vehicle would change lanes only if an acceptable gap were available. Left-turning vehicles examined the oncoming traffic and remained in queue until a suitable gap appeared. At stop signs and signals vehicles behaved very much as they did in the TRANS model. Its detailed microscopic character and the shorter time step generally used (1 s) made the DYNET model actually run more slowly than the TRANS model. If the time steps had been the same in both models, speed of execution would have been comparable.

SAKAI AND NAGAO of Japan developed a simple platoon macroscopic model (20). In this model, roadways were partitioned into 50-m segments. The computer kept track of the number of vehicles in these segments, and the average speed of these vehicles was a function of number of vehicles in the segment. Buses and trucks were replaced by their equivalent number of passenger automobiles.

CURRENT MODELS

VEHICLE TRAFFIC SIMULATION MODEL (VETRAS) (21) by IBM builds on Blum's earlier VTS model; it is somewhat more refined and detailed than models built from scratch. The logic includes provision for right turn on red, and car-following logic is "switch lanes or assume speed of leader car." The model allows easy modification of control logic, and its modular structure and input format make it easy to set up a variety of networks. Network flexibility is limited by a restriction to four-legged intersections. A variety of vehicle characteristics can be generated, but the addition of trucks and buses as distinct entities would require minor programming changes to produce separate statistics. Documentation for VETRAS is not extensive but is clear and does include flow charts. Validation was limited to sensitivity testing.

VEHICLE PERFORMANCE IN TRAFFIC (VPT) MODEL developed by the Aerospace Corporation is an exceptionally detailed, totally microscopic network model (22, 23, 24). Automobiles, trucks, and buses are generated according to a Poisson distribution. In addition to the individual vehicles, the characteristics of the drivers are generated stochastically and include desired speed, desired lane, gap acceptance characteristics, and a frustration factor that determines how long a driver will tolerate following a slower driver. These characteristics are correlated so that a reckless driver generally has the characteristics associated with that description. Vehicles follow each other according to a reasonable car-following law based on the apparent rate of change of the visual angle subtended by the leading vehicle. Lane changing is more complex in this model than in any other; a driver can change lanes simply because he is tired of following a slower driver. This is the only simulation program that includes "actual" accidents. When two vehicles merge into the same spot, they are considered disabled and remain parked in that spot throughout the simulation. Flexibility of this model is somewhat restricted because surface streets are assumed to intersect at right angles only. Also, there are some restrictions on traffic signal displays that make these representations less flexible than those found elsewhere. The model does not consider pedestrian interference. Validation of this model is poor; it has given reasonable results in one or two small tests. User documentation is not generally available. Input requirements are quite extensive because of the detailed microscopic nature of the model. The user may choose the desired output from a wide variety of traffic-related measures such as average speed, average delay, fuel consumption, and vehicle emissions.

URBAN TRAFFIC CONTROL SYSTEM (UTCS-1) AND SIMULATION OF CORRIDOR TRAFFIC (SCOT) are closely related models. UTCS-1 was developed by Peat, Marwick, Mitchell and Company and GASL, while SCOT was produced by GASL alone. UTCS-1 was designed to evaluate traffic signal systems (25, 26). It is based on the DYNET model and is fully microscopic; vehicle data are stored in packed words. The packing operations are done in the program where the information is used or produced. Initialization continues until the number of vehicles in the system appears constant. The time step is fixed at 1 s. SCOT is identical to UTCS-1 except that the Dynamic Analysis of Freeway Traffic (DAFT) model has been included (27). On freeways, vehicles are grouped into platoons; individual statistics are not kept for each vehicle. The update period on the freeway is variable but is usually about 6 s. Level of detail in the UTCS-1 model is excellent. Pedestrian interference is represented, and the lane-changing rules are reasonable. Oncoming traffic interferes with left turns as does traffic that is backed up from the preceding intersection. The freeway portion of SCOT is less realistic. The speed rule as a function of density is somewhat arbitrary, and platoons have an inherent lack of realism. Validation of these models is moderately good. UTCS-1 was extensively validated in Washington, D.C., and the freeway portion of SCOT was briefly evaluated in Dallas. UTCS-1 has been widely used and is generally accepted as giving reasonable results. Documentation is hard to follow but is complete and readily available. The detailed character of the required data makes input preparation difficult and time consuming. The SCOT freeway data are particularly awkward in that two parameters determined by a separate computer program are required. Output is full and complete both link by link and systemwide. A new version of UTCS-1 gives emissions and fuel consumption.

The *SIGNET* (a signal network optimization system) model (28, 29) was developed by Davies as a master's thesis at Purdue University. Input to SIGNET is fairly complex. Output statistics include total vehicle miles, total delay, average delay, delay standard deviation, and average speed. The program will accommodate up to 85 links without modification of dimension statements. SIGNET borrows heavily from the TRANS model but is fully microscopic and utilizes computer words to describe vehicles in a manner similar to UTCS-1. Stop and yield sign control is not represented. However, the model is fairly flexible in terms of intersection geometries and traffic controllers. Model validation was extremely limited and merely confirmed that the simulation outputs were reasonable and consistent with travel time studies. The documentation consists of a very thorough master's thesis project report that includes subroutine descriptions and flow charts.

MICROASSIGNMENT (30, 31, 32), by Creighton, Hamburg, Inc., is a traffic assignment tool developed for transportation planning purposes; it was not designed as a traffic simulation program. However, microassignment considers origins and destinations on a block-by-block basis and therefore has many of the features normally associated with a simulation model. Basically, it was designed to give the expected traffic on any street for a given O-D pattern, but it also outputs the average speed and delay on each link. Microassignment is very macroscopic (for a simulation model) and ignores many of the fine details critical to traffic engineering. Its treatment of traffic signals includes only the coarsest representation of phasing and ignores the effects of signal offsets altogether. The key to the operation of microassignment is in the novel link-node structure. Nodes are placed in the center of each block so that a left-turning vehicle travels on a totally different link from one traveling straight through the intersection. In this way, each link can be assigned a fairly accurate value for the delay associated with the corresponding movement. These delays reflect the character of the control at the intersection (for example, a left-turn link has a lower delay if there is a protected left-turn phase) and, possibly, the traffic on opposing links. For any given O-D pattern, microassignment computes the minimum path and then assigns all the traffic from a batch of origins to their destinations based on minimum time. A new minimum time is then computed, and another batch of trips is assigned. The batch size can be specified by the user; usually, to achieve maximum accuracy, only one path is loaded at a time. Because of its macroscopic character, microassignment can rapidly handle very large networks of even 1000 city blocks. The assignments are essentially static, and a time compression ratio is not directly pertinent. A network of 1334 links took from 108 to 365 s (depending on the batch size) to do one complete assignment. If new assignments are required every 15 min, this corresponds to a time compression of 2.5:1 to 9:1.

CORQ for traffic assignment with queuing in corridors) is a new traffic simulation model written by Yagar of the University of Waterloo (33). It has many features of microassignment. Given a set of O-D tables for, say, successive 15-min periods, it calculates the traffic assignments for one period. It uses a conventional link-node representation of the network and represents signals as capacity restrictions only. The unique feature of this model is that, when capacity cannot accommodate demand, the excess vehicles are stored on the link and added to the demand for the next period. One may visualize the flow as a fluid leaking through the network. User documentation is not yet available for CORQ.

SIGNAL OPTIMIZATION PROGRAMS

TRANSYT (simply a name) was developed by Robertson of the U.K. Transport and Road Research Laboratory as a signal optimization program (34, 35, 36, 37, 38) and, therefore, lies outside the primary emphasis of this paper. However, it contains, as an integral element, a simulation program that can be used without the optimization feature. The logic of the TRANSYT simulation program is deceptively simple. The program is totally macroscopic and completely deterministic; no random numbers are used at all. Uniform vehicular flow enters the upstream end of the farthest upstream link of the network. The flow arrives at the link's downstream end, where it accumulates during the red phase. When the signal turns green, vehicles discharge at the capacity rate of the signal until the queue is dissipated; thereafter the vehicles discharge at the rate at which they arrive. This emergent platoon of vehicles now has a specific shape and arrives at the next downstream stop line with a delay appropriate to the length of the link and to the speed of progression on the link. To enhance realism, the shape of the platoon is changed slightly to reflect dispersion. Again, the discharge from the signal is at intersection capacity until the queue is discharged. In this way, the shape of the platoons at any intersection reflects the effects of all the upstream intersections. Provisions are made for turning vehicles and for the arrival of vehicles at the stop line from the secondary flows that have turned onto the link. Thus TRANSYT is able to represent the performance of the network in a single pass without any initialization. Input preparation is comparatively simple and easy, but output, primarily delays and stops, is not as detailed as that of most models.

SIGNAL OPTIMIZATION (SIGOP II) (39, 40, 41), developed by Lieberman and Woo, is a descendant of TRANSYT and SIGOP I (SIGOP I is an optimization program now obsolete and contains nothing relevant to the subject of this paper) and, like them, is a signal optimization program. Like TRANSYT, it contains a macroscopic traffic flow model that can be used to evaluate the stops and delays of an existing signal system. The logic of SIGOP II utilizes dynamic programming techniques. SIGOP II orders the traffic flows into nine primary and secondary platoon combinations according to when the platoon departs from the first intersection and when it arrives at the second. Modification of platoon characteristics depends on the length of the link and the speed of progression on the link. Provision is made for turning movements and turn pockets. One evolutionary step in SIGOP II is its ability to model the effects of multiphase signals. Input preparation is fairly straightforward once the notation for describing multiphase signals has been mastered. Special coding sheets and a fully documented case study are included with the documentation. Output includes time-space plots of signal control along specified arterials as well as link-by-link statistics. SIGOP II is still being field tested by the Federal Highway Administration and, therefore, is currently available only on an experimental basis.

CORQ1C is a corridor optimization program for queuing (42, 43). As such, the range of networks that can be analyzed is restricted; the program explicitly assumes a one-directional freeway and a parallel arterial with two-way connecting streets. The freeway part of the program is FREQ3C (44), which was developed by May and his coworkers at the University of California. The surface street portion is the TRANSYT program discussed above. Both optimization programs contain simulation programs as subsets. If simple simulation is desired, the network geometry can be considerably more general than that assumed for optimization. Both models have been well validated.

REFERENCES

1. H. H. Goode. Simulation: Its Place in Systems Design. Proc., Institute of Radio Engineers, Vol. 39, No. 12, Dec. 1951, pp. 1501-1506.
2. Y. S. Hsu and P. K. Munjal. Freeway Digital Simulation Models. TRB, Transportation Research Record 509, 1974, pp. 29-41.
3. J. K. Matthewson, D. L. Trautman, and D. L. Gerlough. Study of Traffic Flow by Simulation. Proc., HRB, Vol. 34, 1955, pp. 522-529.
4. J. H. Matthewson and others. Analysis and Simulation of Vehicular Traffic Flow. ITTE, Univ. of California, Los Angeles, Rept. 20, Dec. 1954.
5. M. C. Stark. Computer Simulation of Street Traffic. National Bureau of Standards, Technical Note 119, Nov. 1961.
6. J. G. F. Francis and R. S. Lott. A Simulation Programme for Linked Traffic Signals. Proc., 2nd International Symposium on Theory of Road Traffic Flow, London, 1963, OECD, Paris, 1965, pp. 257-259.
7. M. G. Grace, R. W. J. Morris, and P. G. Pak-Poy. Some Aspects of Intersection Capacity and Traffic Signal Control by Computer Simulation. Proc., Australian Road Research Board, 1964, pp. 274-304.
8. A. M. Blum. A General Purpose Simulator and Examples of Its Application: Part III—Digital Simulation of Urban Traffic. IBM Systems Journal, Vol. 3, No. 1, Nov. 1, 1964, pp. 41-50.
9. A. M. Blum. Vehicle Traffic Simulator Program. SHARE General Program Library 7090/IBM/0027, International Business Machines Corp., Yorktown Heights, N.Y., Aug. 1965.
10. D. L. Gerlough, F. A. Wagner, Jr., J. B. Rudden, and J. H. Katz. A Traffic Simulation Program for a Portion of the Traffic System in the District of Columbia. District of Columbia Department of Highways and Traffic; U.S. Bureau of Public Roads, 1963.
11. F. A. Wagner, Jr., F. C. Barnes, D. P. Stirling, and D. L. Gerlough. Urban Arterial and Network Simulation. Planning Research Corp., Los Angeles, Rept. PRC R-926, Dec. 1966.
12. D. L. Gerlough and F. A. Wagner, Jr. Improved Criteria For Traffic Signals and Individual Intersections. NCHRP, Rept. 32, 1967.
13. F. A. Wagner, Jr., F. C. Barnes, and D. L. Gerlough. Refinement and Testing of Urban Arterial and Network Simulation. Planning Research Corp., Los Angeles, Rept. PRC R-1064, Nov. 1967.
14. F. A. Wagner, Jr., and F. C. Barnes. Traffic Simulation Case Study: City of Detroit New Center Area Network. Planning Research Corp., Los Angeles, Rept. PRC R-1064B, April 1968.
15. M. G. Hartley and S. I. Saleeb. Simulation of Traffic Behavior Through a Linked Pair of Intersections. Transportation Science, Vol. 2, 1968, pp. 51-61.
16. M. G. Hartley and D. H. Green. Study of Intersection Problems by Simulation on a Special Purpose Computer. Traffic Engineering and Control, Vol. 7, No. 3, July 1965, pp. 219-223.
17. S. I. Saleeb. Logical Design and Associated Computer Studies for the Hardware Simulation of Traffic Dispersion. Manchester Univ., England, PhD thesis, 1967.
18. C. E. Storey. Simulation of Traffic by Digital Computer. Traffic Engineering and Control, Vol. 2, No. 10, Feb. 1970, pp. 464-467.
19. E. Lieberman. DYNET: Dynamic Network Analysis of Urban Traffic Flow. General Applied Science Laboratories, Westbury, Conn., Rept. TR-718, April 1969.
20. T. Sakai and M. Nagao. Simulation of Traffic Flows in a Network. Communications of the ACM, Vol. 12, No. 6, June 1969, pp. 311-318.
21. IBM Traffic Signal Control Systems Simulators. Federal Systems Center, International Business Machines Corp., Gaithersburg, Md.
22. P. H. Young and H. S. Porjes. Surface Street Simulation Using the FREEWAY Program. Aerospace Corp., El Segundo, Calif., Rept. ATR-76 (9320)-2, Oct. 31, 1975.
23. H. S. Porjes, C. Speisman, and P. H. Young. An Advanced Computer Program for Determining Vehicle Emissions and Fuel Economy Under Road Traffic Conditions. Proc., 4th Transpo LA—Economic Leverage for Tomorrow, Western Periodicals Co., North Hollywood, Calif., Nov. 1975, pp. 242-251.
24. G. W. Harju, L. R. Bush, R. F. Kramer, and H. S. Porjes. An Advanced Computer Concept for Freeway Traffic Flow Modeling. Proc., Summer Computer Simulation Conference, Simulations Councils, La Jolla, Calif., June 1972.
25. E. B. Lieberman, R. D. Worrall, and J. M. Bruggeman. Logical Design and Demonstration of the UTCS-1 Network Simulation Model. HRB, Highway Research Record 409, 1972, pp. 46-56.
26. Network Flow Simulation for Urban Traffic Control System: Phase II. Peat, Marwick, Mitchell and Co., 1973, 5 vols.; NTIS, Springfield, Va., PB 230 760-230 764.
27. P. Ross, ed. SCOT Model, User's Manual and Program Documentation. Traffic Systems Division, Office of Research, Federal Highway Administration, Working Document, May 1975.
28. G. W. Davies, W. L. Grecco, and K. W. Heathington. A Generalized Street Network Simulation Model. TRB, Transportation Research Record 509, 1974, pp. 16-28.
29. G. W. Davies. Optimization of a Traffic Signal System Through Computer Simulation. Joint Highway Research Project, Engineering Experiment Station, Purdue Univ., Lafayette, Ind., Project C-36-17JJ, No. 8-4-36, June 22, 1972.
30. G. J. Brown and R. S. Scott. Micro-Assignment: A New Tool for Small Area Planning. HRB, Highway Research Record 322, 1970, pp. 149-161.
31. Micro-Assignment. Creighton, Hamburg, Inc.,

Delmar, N.Y., Final Rept., 1969; U.S. Bureau of Public Roads.
32. Adaption and Calibration of Micro-Assignment to the Golden Triangle Area of Pittsburgh. Creighton, Hamburg, Inc., Bethesda, Md., Feb. 1973.
33. S. Yagar. CORQ: A Model for Predicting Flows and Queues in a Road Corridor. TRB, Transportation Research Record 533, 1975, pp. 77-87.
34. F. A. Wagner, Jr. TRANSYT Method for Network Optimization. Proc., 42nd Annual Meeting of Institute of Traffic Engineers, New York, Sept. 26, 1972.
35. D. I. Robertson. TRANSYT: A Traffic Network Study Tool. British Road Research Laboratory, Crowthorne, England, Rept. BRRL 253, 1969.
36. D. I. Robertson. TRANSYT Method for Area Traffic Control. Traffic Engineering and Control, Vol. 2, No. 6, Oct. 1969, pp. 276-281.
37. D. I. Robertson. TRANSYT: Traffic Study Tool. Proc., 4th International Symposium on the Theory of Traffic Flow, Karlsruhe, Beiträge zur Theorie des Verkehrsflusses, Herausgegeben von Bundesminister fur Verkehr, ADT, Bonn, 1968, pp. 134-144.
38. D. I. Robertson. User Guide to TRANSYT Version 5. U.K. Transport and Road Research Laboratory, Crowthorne, England, Technical Note TN 813, March 1973.
39. F. A. Wagner, Jr. SIGOP/TRANSYT Evaluation: San Jose, California. Federal Highway Administration, Rept. FH-11-7822, July 1972.
40. SIGOP: Traffic Signal Optimization Program, User's Manual. Peat, Marwick, Livingstone and Co., 1968; U.S. Bureau of Public Roads; NTIS, Springfield, Va., PB 182 835.
41. SIGOP: Traffic Signal Optimization Program, Field Tests and Sensitivity Studies. Peat, Marwick, Livingstone and Co., 1968; U.S. Bureau of Public Roads; NTIS, Springfield, Va., PB 182 836.
42. M. P. Orthlieb and A. D. May. Freeway Operations Study: Phase IV—Freeway Corridor Control Strategies. ITTE, Univ. of California, Berkeley, Rept. 74-5, June 1975.
43. A. D. May and M. P. Orthlieb. An Urban Freeway Corridor Control Model. TRB, Transportation Research Record 601, abridgment, 1976, pp. 8-9.
44. Y. Makigami, L. Woodie, and A. D. May. Bay Area Freeway Operations Study: Part I—The Freeway Model. ITTE, Univ. of California, Berkeley, Final Rept., 1970.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Simulation of Bus Lane Operations in Downtown Areas

R. W. Bowes and J. van der Mark, De Leuw Cather, Canada, Ltd.

An investigation of potential operational capacities of downtown bus lanes was carried out in Ottawa, Canada. The results could improve coordination between bus priority treatments within downtown areas and those on the major transit corridors feeding the downtown areas. This paper describes the simulation model that was developed during the course of the study to simulate bus behavior in a downtown bus lane. Several operating strategies were tested and evaluated. Depending on the strategy selected, a bus flow rate of 150 to 170 buses/h can be achieved. This flow rate can accommodate 8000 to 9000 passengers/h with acceptable loading standards, operating speeds, and existing standard equipment. Further tests and research are suggested.

Bus priority systems, introduced in many cities, may utilize different measures from the groups listed below:

1. Freeway related measures: reserved lanes, separate roadways, exclusive ramps, queue jumping, metered ramps, and bypass of toll barriers;
2. Arterial related measures: reserved lanes with-flow or contra-flow, separate roadways (bus-only streets), special signalization, special turn permission, and pulling away priority at bus stops;
3. Busway measure: exclusive bus-only roadway either grade separated or with at-grade intersections; and
4. Terminal measure: off-street loading, unloading, and short-term storage of buses, usually connected to freeways or arterials with bus priority measures.

Coordination of the operational capabilities of these various measures could lead to an improved level of service and more efficient use of buses during peak periods. For example, if downtown bus lanes can accommodate the large number of buses arriving from reserved freeway lanes from suburban areas, overall travel times can be reduced.

Generally accepted minimum installation criteria for bus lanes are 30 to 90 buses/h (4); maximum operating capacities of downtown bus lanes are currently 90 to 120 buses/h. These are sometimes exceeded, as in Ottawa, where flows in excess of 140 buses/h occur in some bus lanes.

The success of bus transit in this context relies on how far into the future a municipality's improved bus-based system can meet peak-hour transit needs in major demand areas. The capacity limits of a bus-based system will probably be experienced in the downtown collection and distribution system rather than on reserved freeway lanes and busways between downtown and suburban areas. Because bus lane volumes in Ottawa are already as high as any that have ever occurred in the transit industry's experience, a study was commissioned to investigate the operating capacities of bus lanes in a downtown environment.

After briefly reviewing the literature and summarizing field experiments, this paper will describe the simulation model we developed during the course of the project. A number of bus operating strategies were tested and evaluated, and the results are reported and suggestions for further tests and research made.

LITERATURE REVIEW

An extensive literature review revealed the following findings with regard to bus lanes in downtown areas.

1. Operational capacities are currently in the range of 90 to 120 buses/h, although this maximum has been exceeded.
2. Extensive bus operation research has been carried out, but the potential capacity of downtown bus lanes has not been fully investigated.
3. Preemption of downtown traffic signals is not very beneficial, but fixed-time signal plans have offered buses some measure of priority over other traffic.
4. The maximum number of passengers at the most heavily patronized stop is a major capacity restriction.
5. The interaction between buses and right-turning vehicles in the section of the bus lane immediately before the traffic signal (hereafter called the common section) has been the subject of a number of investigations.
6. Bus lane simulation models are applicable to specific situations only. Most models do not allow for either overtaking by buses or interference by pedestrians, contrary to actual practice in Canadian cities.

FIELD EXPERIMENTS

The purpose of the field experiments was to find values for the different parameters to be used in the simulation model. These values were then compared and supplemented with those cited in the literature (2, 3, 4, 5) and those of the operational experience of Ottawa transit operators. The experiments were conducted on test sections of two downtown Ottawa streets during the afternoon peak period in July and August of 1975. These streets, Albert Street and Slater Street, are both one-way and have with-flow bus lanes and three other traffic lanes. Most observations were carried out with the aid of closed-circuit television. Experiments were conducted to determine the following:

1. Minimum distance between individual berths at a bus stop;
2. Total bus stop delay time, defined as the interval between the moment the bus stops and the moment it sets in motion again;
3. Total bus stop delay time in operating strategies, such as bus platooning, in which an advance passenger information system will be used (such a system would inform waiting passengers of which bus was coming next so that they could form a queue before it arrived);
4. Existing operational conditions on the test sections in downtown Ottawa (this included bus load checks, running times, number of right-turning vehicles, and lane distribution of traffic); and
5. Additional bus-flow data, such as acceleration rates, deceleration rates, cruising speeds, and headways when buses are operating as single units or in platoons.

SIMULATION MODEL

An analysis of the great number of factors influencing bus operations and the continuous nature of some factors required a simulation model. The bus lane operating strategy simulation model (BLOSSIM), developed during this project, simulates bus movements only; the effects of other traffic components are introduced by means of coefficients.

Model Description

The key ingredient of the model is a constraint representing the minimum distance (D_{min}) in meters between two buses. This parameter is expressed as a function of the speeds of the two buses only and varies for the different sections and lanes and is $13 + D_v$ for a bus lane outside the common section (that portion of the bus lane in advance of the intersection where automobiles may enter to turn right, Figure 1), $13 + D_v + D_f$ for a bus lane within the common section, and $19 + D_v$ for lane 2.

$$D_V = (12\,V_f^2) - (7\,V_l^2)/30 \qquad (1)$$

D_v is based on a maximum cruising speed of 11 m/s (35 ft/s) and bus headways of 3 s derived from observations. V_l is the speed of the lead bus and V_f is the speed of the following bus. The filtering distance (D_f) equals $R \times C$, where R is a random number between 0 and 6 and C represents headways between two vehicles; C = 6 m was adopted.

The equation shows that the absolute minimum separation distance (D_{min}) (measured from bus front to bus front) is 13 and 19 m (42 and 62 ft) for buses in the bus lane and in lane 2 respectively. The other two coefficients are delay caused by right-turning vehicles (C_r) and additional delay caused by pedestrians crossing the side street and so interfering with right-turning traffic (C_p).

At each time interval (1 s) model buses must either accelerate, if traveling at a speed less than the maximum cruising speed, or decelerate, depending on the bus in front, the state of the traffic lights, and the need to pick up passengers. A bus will stop for a red traffic signal, to load passengers, and for the bus in front.

The same separation criteria are applied to buses in front of and behind the bus that makes a lane change. A bus driver will try to change lanes if he or she experiences a delay. These delays may be caused by a bus ahead boarding passengers or by right-turning vehicles affected by pedestrian interference.

The model is programmed in FORTRAN IV so that the simulation can be implemented on a large IBM S/360/370 installation under Operating System. An important feature of the implementation is the separation of the simulation functions from the reporting function. Special care was taken to represent a typical downtown area, although most of the parameter values are related to conditions in downtown Ottawa.

Input Data

The road system simulated consists of six blocks, each 180 m (600 ft) long. A block is measured as shown in Figure 1.

There are three stops per block and these can be located midblock, near side, or far side. We prepared a table by route that gives the measurement from the zero point at block 1 along the total system to the point at which each stop is located. We can vary not only the location of the bus stops but also the allocation of buses per stop.

Spacing between bus stops is fixed at 26 m (86 ft) so one bus can pull in between two buses or the second bus can pull out, plus some extra space for model flexibility. Also, the far side location is set into the block at a sufficient distance to allow a bus to cross the stop line but not to enter the bus stop. The location of a bus stop from the zero point (1 m = 3.28 ft) is

Location	Spacing (m)
Midblock	54 - 80 - 106
Midblock (bus platooning only)	66 - 80 - 307
Far side	26 - 52 - 78

The delay time at a bus stop equals 2B + 7 s, where B is the number of passengers boarding. The delay time is reduced to 2B + 5 s when a passenger information system is used. All buses on each route must stop at all designated stops and incur a minimum penalty of 7 or 5 s, even if no one boards.

The model has 18 bus routes divided equally into three groups, A, B, and C. The A routes stop at all A bus stops, and so on. The frequency of the routes varies.

Passengers are generated for each route at each stop as a function of route frequency, which in actual practice is determined by demand. A random number (seed number) is selected to start up the generation of buses and passengers. The generation rate can vary for each route and bus stop and thus reflects a uniform or nonuniform profile along the test section. In the test program the generation rate is higher in blocks 1 and 6 than in blocks 2 and 5, and higher in 2 and 5 than in 3 and 4.

Average bus occupancy is also variable, but 50 to 55 is our system aim. Bus capacity varies but is set in the model at 80 passengers, which is reached by some buses in most runs. Even when full, buses stop at all stops and incur a minimum penalty of 2 s.

Each intersection is signal controlled, and each cycle length is 80 s with 40 s green and 4 s amber in the direction of travel. Variable offsets are set at 15, 25, 40, and 65 s in the model. Automobiles may enter and turn right in a common section of the bus lane for 45 m (150 ft) back from the stop line in blocks 2, 4, and 6.

The top speed and acceleration or deceleration rates are variable. In the model, a top speed of 11 m/s (35 ft/s) has been used. Acceleration and deceleration rates have been set at 1.2 and 1.5 m/s (4 and 5 ft/s) respectively.

Table 1 summarizes input characteristics and values used in the model runs.

Figure 1. Geometry of a block.

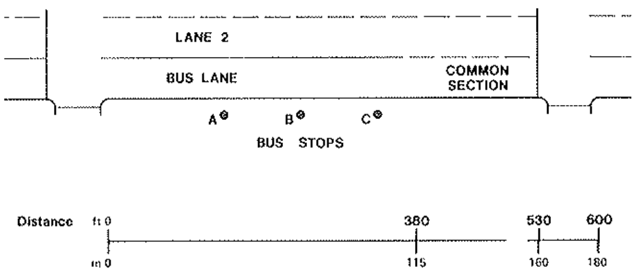


Table 1. Model input characteristics.

Characteristic	Fixed or Variable	Values Used
Number of blocks	F	6
Length of block excluding intersection, m	F	160
Length of intersection, m	F	20
Location of common section, block no.	V	2, 4, 6
Number of bus stops per block	F	3
Number of bus routes per stop	F	6
Location of each stop	V	Varies
Spacing between stops, m	V	14 or 26
Minimum delay per stop, s	V	5, 7
Added delay per passenger per stop, s	V	2
Cycle length for traffic signals, s	V	80
Length of green time, s	V	40
Length of amber time, s	V	4
Time offset between adjacent signals, s	V	15, 25, 40, 65
Bus capacity, seats and standees	V	80
Bus acceleration rate, m/s^2	V	1.2
Bus deceleration rate, m/s^2	V	1.5
Maximum speed, m/s	V	11
Bus flow generated, buses/h	V	150 to 185
Bus passenger flow generated, passengers/h	V	7500 to 9200

Note: 1 m = 3.28 ft.

Output Data

Each model run can produce the following three types of output data:

1. Terminal data including particulars for each bus, such as the time it enters and leaves each block, the number of passengers, mean speed, and so on, and the number of buses and the number of passengers per hour;
2. Photographs of events and status of queues during the run at selected time points; and
3. Comprehensive traces of individuals or groups of buses during a time period.

Table 2. Summary of simulation results.

	Operating Conditions							Model Results			
	Bus Stop		Automobile Right Turns From Common Section		Traffic Signal Offset	Buses Allowed to Overtake Buses					
Strategy	Midblock	Far Side	Yes	No	(s)	Yes	No	Buses/h	Passengers/h	Avg Bus Occupancy	Avg Bus Speed (m/s)
Alternating bus stop spacing	X		X		25	X		160	8570	54	2.85
Buses operating in platoons of three	X			X	65		X	174	9100	52	2.14
Bus overtaking allowed, no right turns by other vehicles	X			X	25	X		170	9290	55	1.71
Variations in traffic signal offset	X		X		Various	X		129 to 147	7615 to 8389	59 to 57	1.89 to 1.55
No bus overtaking allowed	X		X		25		X	128	7980	62	1.27

Note: 1 m = 3.28 ft.

Model Testing

Before examining different bus operating strategies, a detailed analysis for model realism was made of bus running times, interference coefficients, and bus operations in the second lane. Further, we needed to know how to measure the bus flow rate, which seed number (to start up the generation process) to use, and how long to extend each computer run. Testing with different seed numbers and run lengths indicated variations in flow rates comparable to those for vehicle counts. We decided to investigate a number of bus operating strategies using the same seed number and run length of approximately 60 min.

STRATEGY EVALUATION

The results of the computer runs indicate that with the described set of test conditions some strategies are more effective than others. Table 2 briefly summarizes the results of the simulation runs.

The most effective strategy we tested is one in which alternating bus stops are employed to increase the average bus stop spacing from 180 to 360 m (600 to 1200 ft). The buses on half of the routes stop in blocks 1, 3, and 5; those on the other routes stop in blocks 2, 4, and 6. This strategy yields a flow rate of approximately 160 buses/h and a good quality of service as measured by average bus occupancy and average operating speed through the six-block test section. This strategy was felt to be the most effective, because it achieved the best compromise among bus volume, speed, and occupancy. Other strategies did achieve higher bus volumes but only at a cost of lower average speeds. Even higher volumes are probably also possible with this strategy if right turns by automobiles are restricted.

If overtaking is not possible, platooning could be applied. This is a strategy whereby buses in a group of three or more are introduced into the bus lane and then proceed through the system as a group. Their order of arrival at a bus stop is automatically given in advance to waiting passengers by means of a scanning system, which allows passengers to form into lines and board more efficiently. In addition, platooning allows signal offsets to be adjusted to accommodate bus movements. This strategy yielded a high flow rate that was sensitive to variations in traffic signal offsets. The speed, however, was relatively low. The necessary advance bus arrival sign system may be expensive unless buses are already equipped with detecting devices, although the driver could operate the sign system by radio.

The third strategy, wherein buses are allowed to overtake but other vehicles are not allowed to make right turns from the common section, produces high throughput with acceptable bus occupancy but low speed. This points to banning right turns by automobiles from the common section. However, some right turns in the downtown area would be required, and the location of the common sections is thus important.

The influence of variations in traffic signal offset is inconclusive for existing operations with bus overtaking and right turns by automobiles because of the multiplicity of stops and bus routes in the system and the interaction among throughput, bus occupancy, and speed.

Although overtaking by buses is permitted in all major Canadian cities, it is not always possible because of traffic in adjacent lanes. This would influence bus operation considerably. Allowing no overtaking would reduce bus throughput and quality of service. The provision of bus bays would remedy this.

CONCLUSIONS AND RECOMMENDATIONS

BLOSSIM is a valuable tool in the analysis of bus operating strategies. Using the same seed number and run lengths of approximately 60 min, we were able to evaluate several strategies.

Bus flow rates of between 150 and 170 vehicles/lane·h can be achieved without having to resort to extraordinary and unconventional operating strategies. This flow rate can accommodate 8000 to 9000 passengers/h with acceptable loading standards, operating speeds, and existing standard equipment.

To refine the model and to extend the knowledge of operating strategies we recommend the following subjects for further research:

1. Model validation through additional runs to establish sensitivity to seed number and run length,
2. Model modification to incorporate variations in block length,
3. Investigation of additional strategies,
4. Additional field validation tests, and
5. Refinement of the algorithm that defines the distances between buses, particularly for bus movements in the second lane.

ACKNOWLEDGMENTS

The research described in this paper was part of a study carried out by De Leuw Cather, Canada, Ltd., under the direction of a working committee representing the Canada Transportation Development Agency of Transport, the Ontario Ministry of Transportation and Communications, the Ottawa-Carleton Regional Transit Commission and Transportation Department, and the National Capital Commission. The assistance of the committee members during the study is appreciated. Neil V. McEachern and Tony Zimmer of Group Five Consulting, Ltd., contributed greatly to the development of the model. The original study was carried out in imperial units.

REFERENCES

1. Bus Lane Operating Strategies in Downtown Areas. De Leuw Cather, Canada, Ltd., London, Ontario, 1976.
2. H. S. Levinson, C. L. Adams, and W. F. Hoey. Bus Use of Highways: Planning and Design Guidelines. NCHRP, Rept. 155, 1975.
3. W. F. Hoey and H. S. Levinson. Bus Capacity Analysis. TRB, Transportation Research Record 546, 1975, pp. 30-43.
4. Bus Rapid Transit Options for Densely Developed Areas. Wilbur Smith and Associates, New Haven, Conn., Feb. 1975.
5. J. E. Baerwald, ed. Transportation and Traffic Engineering Handbook. Prentice-Hall, Englewood Cliffs, N.J., 1976.

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

Microscopic Traffic Simulation Package for Isolated Intersections

Thomas W. Rioux and Clyde E. Lee, Center for Highway Research, University of Texas at Austin

The Center for Highway Research at the University of Texas at Austin has developed a new microscopic traffic simulation package, called the traffic experimental and analytical simulation (TEXAS) model, that can be used to evaluate existing or proposed intersection designs and to assess the effects on traffic operations of changes in roadway geometry, driver and vehicle characteristics, flow conditions, intersection control, lane control, and signal timing plans. A geometry processor calculates vehicle paths on the approaches and in the intersection, identifies points of conflict between intersection paths, and determines minimum available sight distance along each inbound approach. A driver-vehicle processor generates individual driver-vehicle units and describes their characteristics. An auxiliary headway distribution analysis processor helps select an appropriate headway distribution. A simulation processor simulates the movement of each driver-vehicle unit through the system and gathers performance statistics. Linear acceleration and deceleration models and a noninteger, microscopic, generalized car-following equation are used. Traffic signal simulators are included for pretimed, semi-actuated, and fully actuated controls. Other intersection control options include no control and yield, less-than-all-way stop, and all-way stop signs. New simulation techniques include lane change decision and geometry, sight distance restriction checking, intersection conflict checking, and storage management and logic processing methods. A new field device for recording validation data and for determining suitable model input is described. Input was designed to be user oriented and minimal; output is concise and functional. Documentation has been developed for both users and programmers.

Satisfactory solutions to traffic control problems at intersections involve evaluating capacity, efficiency, safety, environmental impact, and cost and, therefore, always require a detailed analysis of the expected individual driver-vehicle response to the complex geometry, traffic, and control conditions. Inadequate predicting and tracing methods have plagued transportation engineers through the years and have limited them to macroscopic estimates of traffic stream flow characteristics for their analyses and designs.

Advances in digital computer technology during the past decade, however, have allowed complex simulations that are quite sophisticated. Several traffic simulation programs have been developed, but none of these has been designed specifically to handle the single, multileg, multilane, mixed-traffic intersection operating either without control or with a conventional traffic control.

In 1971, development of such a program was undertaken as part of the Cooperative Research Program between the Texas State Department of Highways and Public Transportation and the Center for Highway Research at the University of Texas at Austin. The scope of the study was purposely restricted to simulation of traffic at a single intersection, since other models were being designed to handle primarily multi-intersection, signalized networks. Emphasis was placed on making the simulation package user oriented and on minimizing computational requirements.

The result of this research is a microscopic traffic simulation package called the traffic experimental and analytical simulation (TEXAS) model for intersection traffic (1, 2, 3, 4). In this model, each individually characterized driver-vehicle unit is examined separately. At selected time intervals the computer program provides the simulated driver with information such as desired speed, destination, current position, velocity, acceleration, relative position and velocity of adjacent vehicles in the system, critical distances to be maintained, sight restrictions, and location and status of traffic control devices. The simulated driver may maintain speed, accelerate, decelerate, or maneuver to change lanes. Response is a function of driver and vehicle characteristics, roadway geometry, traffic control, and actions of other driver-vehicle units in the system. The highest priority logical response of the driver-vehicle unit is determined on the premise that the driver wants to maintain a desired speed but that he or she will obey traffic laws and maintain safety and comfort.

Structured programing techniques were used in developing the simulation model and storage requirements kept to a minimum. FORTRAN IV was used on both CDC and IBM computers. The overall structure of the model was arranged in three separate processors for computational efficiency.

ORGANIZATION OF THE TEXAS MODEL

An essential function of the simulation process is the definition of the geometry of the intersection. A special geometry processor (GEOPRO) was developed to calculate and store all geometric details that would be held constant for each simulation run. Likewise, a driver-vehicle processor (DVPRO) was developed to characterize each driver-vehicle unit in the simulated traffic stream. Repetitious computations that are required for simulating the movement of each driver-vehicle unit through the intersection in response to the geometry, traffic, and control conditions were incorporated into a simulation processor called SIMPRO. The interrelation among these three processors is shown in Figure 1. Documentation for the model includes a programmer's guide, a user's manual, and numerous comment statements in each processor (1, 2, 3).

Geometry Processor

GEOPRO calculates the vehicle paths on the approaches and within the intersection, the points of conflict between intersection paths, and the minimum available sight distance between each inbound approach (5). This information, first processed so that all need for Cartesian coordinate information within the simulation processor is eliminated, is then written onto a tape for subsequent use by SIMPRO. The generalized input is the GEOPRO title, and information for the approach, the lane, the arc (for plotting), the line (for plotting), the sight distance restriction, and the GEOPRO options, which include the path "type" option, the plot option, the maximum radius for intersection paths, the definitions for straight and U-turn movements, and the minimum distance between two intersection paths for a

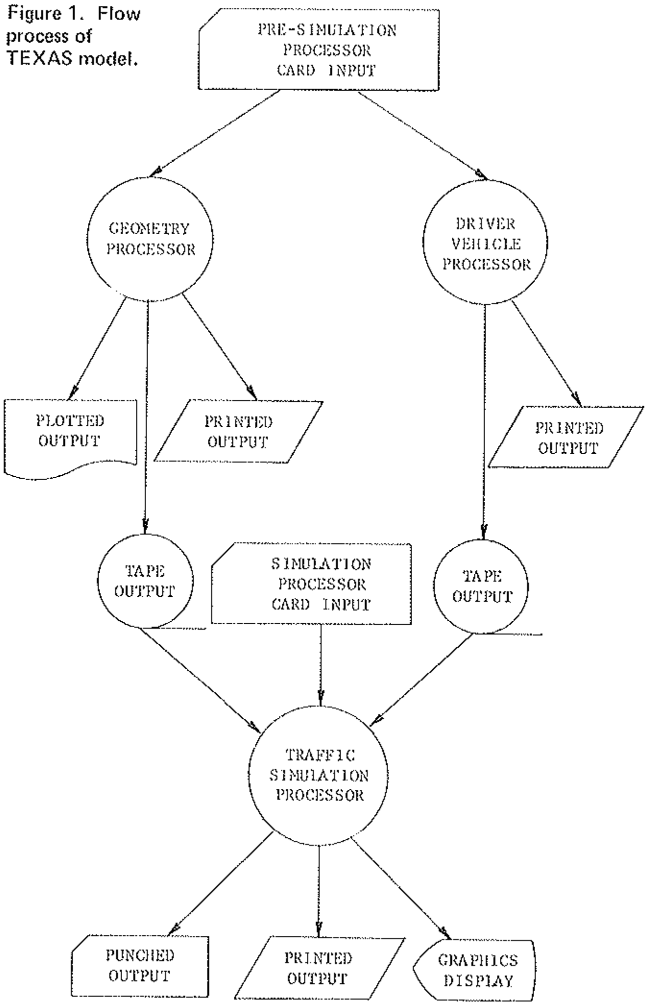


Figure 1. Flow process of TEXAS model.

conflict. Most of the information required will normally be available from a plan view diagram of the intersection.

GEOPRO uses straight-line segments and arcs of circles to describe paths that vehicles will follow in the intersection. Algebraic equations for the intersection of two lines, the intersection of an arc and a line, and the intersection of two arcs are used for calculating points of conflict between paths. Safe-side friction factors are used to compute maximum vehicle speed for negotiating an intersection path. The minimum available sight distance between inbound approaches is calculated for each 7.6-m (25-ft) increment along an inbound approach. Printed output from GEOPRO includes an echo print of the input, a listing of minimum available sight distances between inbound approaches, a listing of the intersection paths, a listing of conflicts between the intersection paths, and certain input or execution errors. Plot output is optional but may include a plot of the full length of all approaches and sight distance restriction points, a plot of an enlargement of the intersection area, and a plot showing all generated intersection paths. Tape output includes the title for GEOPRO and all geometry information needed as input for the SIMPRO. The proper functioning of GEOPRO was verified by analyzing debug prints and plots from various test data sets.

GEOPRO requires 26 900 words of storage on CDC 6600 computers and 176 000 bytes of storage on IBM 370-155 computers. Geometry computations in the central processor for an average intersection take 6.3 s on CDC computers and 9.2 s on IBM computers.

Driver-Vehicle Processor

DVPRO describes the characteristics of several driver and vehicle classes, generates individual driver-vehicle units, and orders these units sequentially by queue-in time. This information is stored on tape for later use in SIMPRO. The generalized input to DVPRO includes the DVPRO title; the approach information; the lane information; the driver-vehicle processor options, which include the number of minutes for generating traffic, the minimum time between two vehicles in the same lane, and a variety of options that would allow for an override of the standard program-supplied driver-vehicle mix; and special driver-vehicle units. Such information will usually be available from experience with similar intersections or from routine traffic studies. A common input card deck for DVPRO and GEOPRO is used because much of the information is the same.

In DVPRO, headways of vehicles that arrive on each inbound approach are generated as random variables of one of the following distributions: uniform, log normal, negative exponential, shifted negative exponential, gamma, Erlang, or constant. Mean headway for traffic on each inbound approach is calculated from the specified volume for that approach, and only one additional parameter, which indicates randomness, needs to be specified to describe these distributions. Driver and vehicle class, desired outbound approach, and inbound lane number are generated as random variables of empirical distributions (percentages specified by the user). Desired speed is generated as a random variable of the normal distribution; the user prescribes the mean and the 85th percentile speed.

Printed output includes an echo print of the input, statistics of generation, and certain input or execution errors. Tape output includes the DVPRO title and all information necessary for SIMPRO. DVPRO requires 14 500 words of storage and typically 2 s of central processor time for 12 min of traffic on CDC computers and 95 000 bytes of storage and 3 s of central processor time on IBM computers.

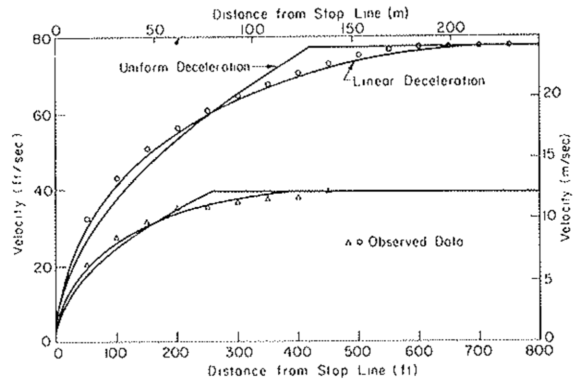
The proper performance of DVPRO was verified by analyzing debug prints of sample problems and by using a special distribution analysis processor called DISPRO. DISPRO is an ancillary processor that fits selected mathematical distributions to empirical headway data. Chi square is then calculated as a goodness of fit indicator for each distribution, and the maximum cumulative difference is found for a Kolmogorov-Smirnov one-sample test. As an aid to the user, a histogram of the input headway data and of each distribution fitted is plotted.

Simulation Processor

Input Requirements

The purpose of SIMPRO is to process each driver-vehicle unit through the intersection system and to gather and print performance statistics about the simulation. The input requirements for SIMPRO are the output tape produced by GEOPRO, the output tape produced by DVPRO, and card input to SIMPRO. The card input consists of SIMPRO title and options, which include start-up and simulation time; time-step increment for simulation; speed for delay below a set speed; maximum clear distance for being in a queue; lambda, mu, and alpha values for the generalized vehicle-following equation; type of intersection control; desired summary statistics; time for lead and lag zones for intersection

Figure 2. Velocity versus distance during normal deceleration to a stop.



conflict checking; and lane control for each lane. For a signalized intersection, additional card input is necessary. The signal indication information for each inbound lane consists of input that models the cam stack found in most signal controllers plus the timing scheme for displaying each interval. If the intersection is actuated, supplementary information about detector type and location is required. Highly sophisticated, multiphase controllers, except volume-density types and minicomputers, are modeled in detail, and all required input information can be obtained from conventional sequence patterns and interval timing plans that are familiar to transportation engineers.

Overview

SIMPRO uses a fixed-time increment in the ½ to 1 s range. It basically has three types of links on which to simulate driver-vehicle units: outbound approaches where there is no control mechanism at the end, intersection paths, and inbound approaches where there may be a control mechanism that regulates entry into the intersection. SIMPRO processes driver-vehicle units on outbound approaches, then on intersection paths, and then on inbound approaches; then new driver-vehicle units are added to the system; and finally signal status is processed.

Validation of SIMPRO proceeded in two stages: validation of its specific portions and validation of its performance statistics.

Acceleration

For the model to replicate real-world phenomena as accurately as possible, a thorough investigation of acceleration and deceleration models was undertaken. The uniform acceleration model frequently used does not match observed behavior accurately on a microscopic scale. A linear acceleration model that hypothesizes use of maximum acceleration when vehicle velocity was zero, zero acceleration at desired velocity, and a linear variation of acceleration over time was adopted. Comparisons of this model with observed data (6, Figure 2.10, p. 27) indicated excellent agreement. This model also compared favorably with the nonuniform acceleration theory (7, p. 9) used in describing the maximum available acceleration for the vehicle.

The parameter that is determined by driver desire in effecting the position, velocity, and acceleration of the vehicle is acceleration slope (jerk). Dramatic changes in acceleration in a short period of time are restricted in the model by limiting acceleration slope range.

In validating the acceleration models, position, velocity, and acceleration versus time plots were produced by SIMPRO and checked to ensure that the vehicles responded in a reasonable manner under varying conditions. Studies by Beakey (8) led to the development of the relationship between maximum initial acceleration and desired speed that is used in SIMPRO (9, p. 10).

Deceleration

The investigation of deceleration models revealed that the uniform deceleration model did not closely approximate actual vehicle behavior. We chose a linear deceleration model in which the vehicle has an initial deceleration of zero and reaches maximum deceleration at the instant the vehicle stops and in which deceleration rate varies linearly over time. Comparisons of this model with real-world data (6, Figure 2.14, p. 30) indicate that the model very accurately represents vehicles decelerating to a stop (Figure 2). Again, the parameter that affects the position, velocity, and deceleration of the vehicle is deceleration slope (jerk).

Validation of the deceleration models was the same as that for the acceleration models.

Car-Following

Several car-following techniques were investigated, but the noninteger, microscopic, generalized car-following equation (10) was selected because of its superiority and flexibility. A value of deceleration was computed from the equation, and then we chose a deceleration slope that brought the vehicle to the computed value of deceleration in one time increment. Studies by May (11) indicated acceptable ranges for lambda and mu in the car-following equation. Values used in the TEXAS model may range from lambda of 2.3 to 4.0 and mu of 0.6 to 1.0.

In validating the car-following model, position, velocity, and acceleration versus time plots were produced and checked by SIMPRO.

Acceleration and Deceleration Logic

To determine a unique response for each driver (accelerating to desired speed, accelerating to lead vehicle speed, following the vehicle ahead, remaining stopped, checking whether deceleration to a stop is necessary, or continuing deceleration to a stop), a logical binary network for acceleration and deceleration was developed and used in the model. The model maintains current values for all independent variables in the network and executes a special logic routine to determine the appropriate decision for the conditions.

Signalization

The signal indications displayed to traffic on each inbound lane are determined from information provided on card input and from the dynamic response of a simulated controller. For a pretimed controller, the specified sequence of indications is implemented, and the duration of each interval is referenced to a simulated real-time clock. In representing a traffic-actuated controller, all time intervals, including those affected by detector actuation, are simulated. During any particular green indication, the controller may extend the green, "max out," or "gap out" according to time limits, detector actuations, and presence or absence of demand for another phase. Appropriate amber and all-red clearance intervals are provided.

When the signal indication changes for the driver, the appropriate response (enter the intersection and do not check conflicts, enter the intersection and check conflicts, go on amber, stop on amber, or stop for red) is determined. The go or stop-on-amber decision is based upon whether a driver-vehicle unit can stop at the stop line without exceeding a maximum level of deceleration.

Validation was based on testing many cases of amber-go and amber-stop decisions for reasonableness and consistency. Delay to the first vehicle in the queue and "march-out" headways were compared with observed values (12). In this comparison, particular attention was given to the location of the screen line as suggested by Berry (13).

Intersection Logic

To define driver response to intersection control, a logical binary network was developed and used in the model. The model maintains correct values for all independent variables and, after executing a logic routine, sets dependent variables to proper values. This determines the conditions under which a driver-vehicle unit may enter the intersection.

Lane Change

The TEXAS model distinguishes between two types of lane changes: the forced lane change (the currently occupied lane does not provide a path to the desired outbound approach) and the optional lane change (less delay can be expected by changing to an adjacent lane that also connects to the desired outbound approach). The decision to change lanes is controlled by the availability of an acceptable gap. A cosine curve is used to represent the path followed during the lane-change maneuver.

Time-lapse photography served as the basis for the development of equations for acceptable lead and lag gaps for a lane-change maneuver (14). The time required to complete a simulated lane-change maneuver is approximately 3 to 4 s for any reasonable speed. Numerous test cases were conducted to ensure that lane changes, both forced and optional, were made in a reasonable and safe manner under various conditions.

Sight Distance Restriction

GEOPRO locates all sight obstructions from information included as input and calculates the distance that is visible along other inbound approaches for each 7.6-m (25-ft) segment of an inbound approach. SIMPRO then stores this information as a series of data arrays. The simulated driver utilizes the information in evaluating the effects of sight restrictions to ensure that the impending entry into the intersection will be legal and safe. He or she controls the velocity of the vehicle in such a way as to avoid a potential collision with a hypothetical vehicle that may be hidden from view by the obstruction.

Intersection Conflicts

In checking conflicts, the simulated driver forecasts time of arrival at points of potential conflict and compares this with the projected arrival time of other vehicles in the intersection and other vehicles already committed to entering the intersection. The decision to enter the intersection is then based on whether the driver's vehicle can pass safely through the point of conflict in front of or behind other vehicles that have the right-of-way. Conflict computations are executed only for those vehicles that may be required to yield to other approaching vehicles.

Storage Management and Logic Processing

SIMPRO uses a special storage management and logic processing program called COLEASE, which provides a mechanism for storing specified variables in a format that maximizes computer bit usage by disregarding normal word boundaries. COLEASE also allows efficient processing of logical binary networks (15) and reduces the main storage requirements on CDC computers to approximately one-seventh of that normally required.

Output

Output from SIMPRO consists of title from the GEOPRO tape, title from the DVPRO tape, title from the card input to SIMPRO, echo print of all card input to SIMPRO, information about each collision (if collisions occur), listing of driver-vehicle units eliminated (if any) from the simulation because of full entry lane, summary statistics, and certain input or execution errors. Summary statistics include total delay; queue delay; stopped delay; delay below a set speed; total and average vehicle travel distance; total and average travel time; number of vehicles processed and equivalent hourly volume; average desired speed, time-mean speed, and space-mean speed; average maximum uniform acceleration and deceleration; average and maximum number of vehicles in the queue for each lane; average ratio of entry speed to desired speed; and actuated signal performance indicators. Total delay is defined as the actual travel time through the system minus hypothetical desired travel time. Queue delay is that spent in a queue waiting to enter the intersection; it includes move-up time. Stopped delay is that spent stopped in queue waiting to enter the intersection; it does not include move-up time. Delay below a set speed is the amount of time traveled at a velocity less than or equal to that speed anywhere in the system.

Validation

Performance of the simulation processor was evaluated through independent testing of selected subprograms, analysis of position, velocity, and acceleration-versus-time plots, and review of interactive graphics displays of the movement of individual vehicles through the system. Figure 3 illustrates the position and velocity-versus-time plot for a test case in which the first vehicle entered the system at 9 m/s (30 ft/s) and 7 s later the second vehicle entered at 15 m/s (50 ft/s). After about 12 s the second vehicle overtook the first and decelerated until the speeds matched. The first vehicle later decelerated until the speeds matched. The first vehicle later decelerated to a stop at the intersection, and the second vehicle stopped behind the first. After the first vehicle entered the intersection, the second vehicle advanced to the stop line and, after checking potential conflicts, entered the intersection. The second vehicle again caught up with the first and trailed at a safe distance. The smooth trajectory of each vehicle indicates that all components of the model functioned properly for this test case.

SIMPRO was validated by comparing various performance statistics from the model with data from field observations. The primary basis for comparison was delay resulting from different traffic demands, types of control, and intersection configurations. Extensive in-

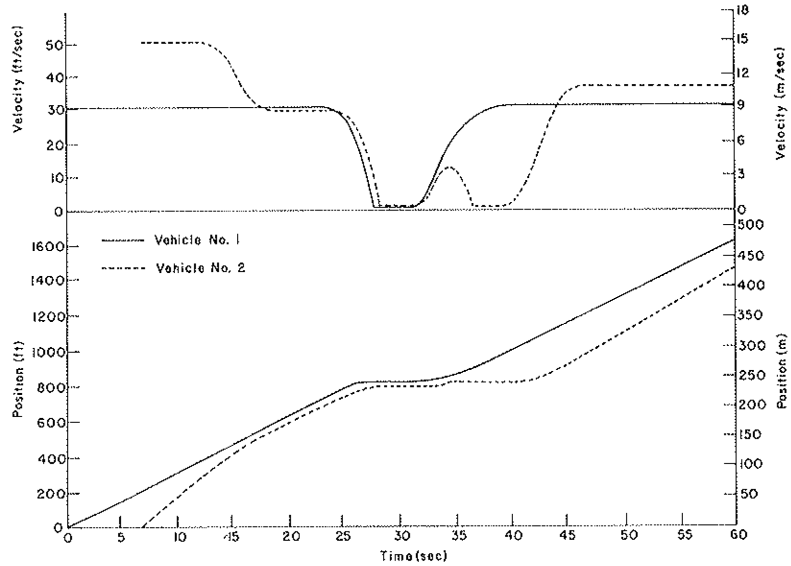


Figure 3. Position and velocity versus time trajectories for two vehicles.

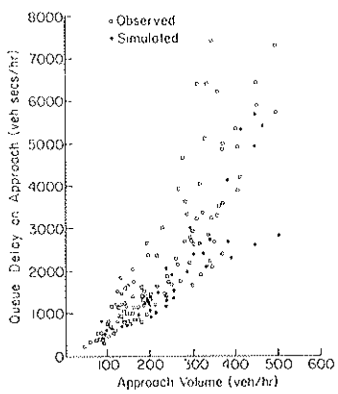


Figure 4. Delays at five intersections.

tersection delay and volume data were available from a previous research study (16), and special additional studies were conducted for validation of the TEXAS model. The earlier data were entered into an electromechanical recording device.

A new recording device was developed and used for the more recent field studies. This was superior to the older one in the following features: portability, independent dc power supply, solid-state electronic component reliability, accurate time reference synchronized for several units, display of the current counter reading to the observer, use of an inexpensive voice-grade cassette tape recorder for storing digital data in analog form, and economical construction and maintenance. Each second, the recording device interrogates two counters (which are kept current by an observer using an increment, decrement, and zero switch) and writes digital values, current time, and recording device identification number onto a cassette tape.

The field observation technique that was used to measure queue delay at nonsignalized intersections called for an observer to increment a counter each time a vehicle joined the queue waiting to enter the intersection; this counter was decremented each time a vehicle crossed the stop line and entered the intersection. Thus the current number of vehicles in the observed queue was indicated by the counters. Volume information was obtained by having another observer increment a counter each time a vehicle entered the intersection. At signalized intersections the observer used a similar technique to keep a current indication of the number of vehicles that were actually stopped while waiting in a queue to enter the intersection.

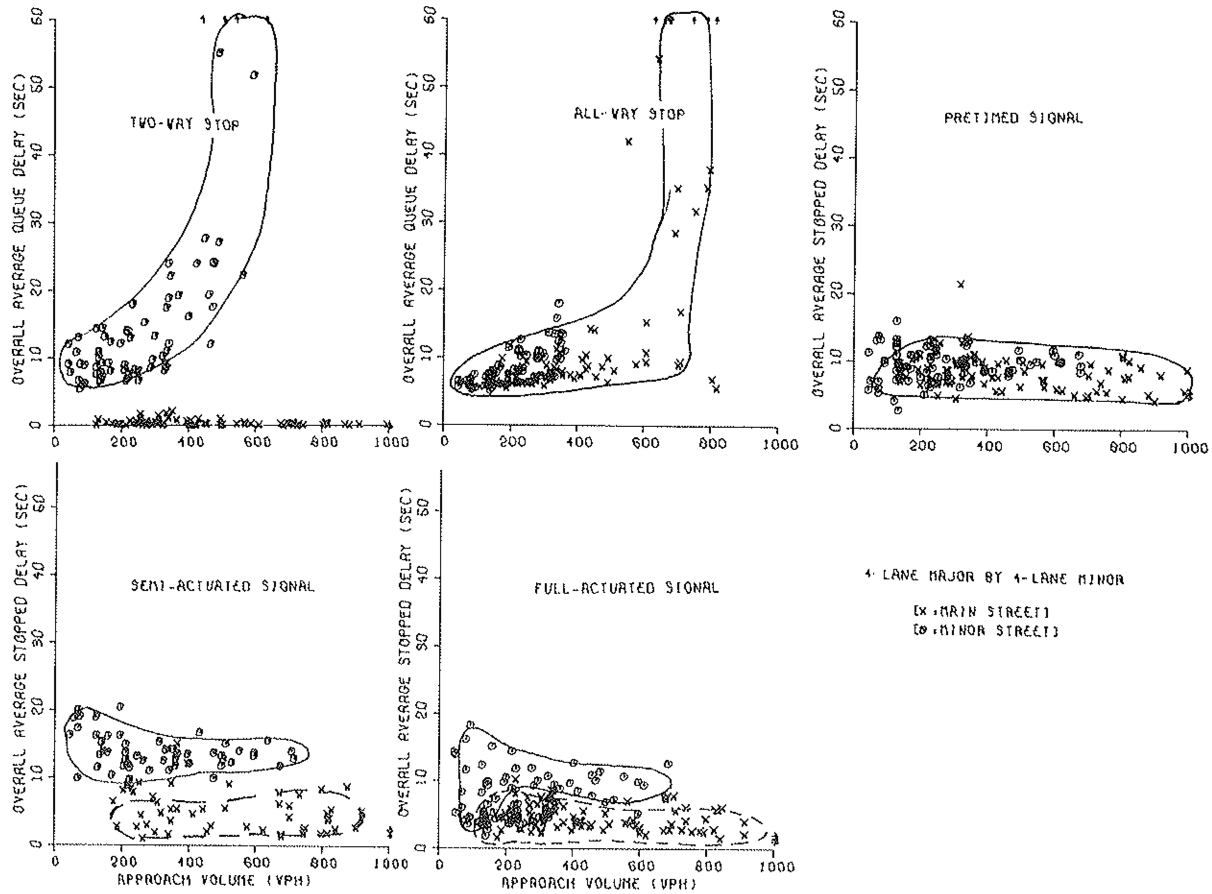
Data from the cassette tapes were retrieved in two steps: (a) the data stored in analog form on as many as six tapes were simultaneously converted to digital form by an analog-to-digital conversion processor (ADPRO) on a Hewlett Packard HP2115A computer system; and (b) the appropriate statistics were computed from the digital data by using a special delay, volume, and headway processor (DVHPRO).

Initial comparisons of simulated delays with observed delays indicated the need for adjusting certain components of the model in order to obtain satisfactory agreement over a wide range of traffic volumes. Modifications were made in the maximum velocity at which a vehicle first enters the queue and begins experiencing queue delay and stopped delay or both, the value of the time zones that the driver considers safe when checking intersection conflicts, the hesitation time for vehicles entering unsignalized intersections, the hesitation time for the first vehicle in a queue entering the intersection when the signal turns green, and the parameters for the generalized car-following equation (lambda, mu, and alpha). The resulting agreement between the observed and the simulated delay values, as shown in Figure 4, is one example that demonstrates the validity of the TEXAS model.

Computer Time and Storage Requirements

The computer time requirements for SIMPRO are difficult to reduce to a single value. As an indication of the efficiency of the model, a simulated time to computer time ratio for CDC computers has been calculated for each run of SIMPRO. This ratio varies with the type of intersection control, the lane length, the time-step increment, and the total number of vehicles processed. For signalized intersections, 180-m (600-ft) lanes, and a time step increment of 1 s, the lower limit of efficiency (worst) is in the general range from 30 at a total equivalent hourly volume of 1000 vehicles/h to 8 at a volume of 2000 vehicles/h. The upper limit of efficiency (best) is 45 and 15 respectively for the same volumes. For nonsignalized intersections, 180-m (600-ft) lanes, and a time step increment of 1 s, the lower limit of efficiency (worst) is in the general range from 40 at a volume of 750 vehicles/h to 8 at a volume of

Figure 5. Simulated delays for five types of controlled intersections.



1250 vehicles/h. SIMPRO uses 32 250 words of storage on CDC computers and 210 000 bytes of storage on IBM computers.

USES OF THE TEXAS MODEL

This model may be used to study traffic behavior at an intersection operating either without control or with any conventional control. Features of the model that make it particularly suitable include accommodation of 5 driver classes and 15 vehicle classes; 6 approaches with 6 lanes/approach; 305-m (1000-ft) lane lengths; sight restrictions; uncontrolled operation; sign-controlled operation; 8-phase signal control with skip-phase, dual-left, and parent-and-minor options; 2 detector types and 5 detectors/lane; 72 signal intervals; geometrically correct lane-changing maneuvers and paths through the intersection; and left-turn or right-turn-on-red options. The effects of changes in roadway geometry, driver and vehicle characteristics, flow conditions, intersection control, lane control, and signal control options can be readily evaluated.

More than 600 runs of the TEXAS model have been made, and the results have been evaluated in relation to capacity analysis of unsignalized intersections and warrants for various types of control (4, 17, 18). Figure 5 illustrates these results for comparable values of overall average delays (total delay divided by the total number of vehicles using an approach) for an intersection with four lanes on both major and minor streets.

ACKNOWLEDGMENTS

The research on which this paper is based was conducted for the Texas State Department of Highways and Public Transportation by the Center for Highway Research, the University of Texas at Austin, in cooperation with the Federal Highway Administration. The contents of this report reflect our views, and we only are responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

REFERENCES

1. The TEXAS Model for Intersection Traffic: Development. Center for Highway Research, Univ. of Texas at Austin, Research Rept. 184-1, in preparation.
2. The TEXAS Model for Intersection Traffic: Programmer's Guide. Center for Highway Research, Univ. of Texas at Austin, Research Rept. 184-2, in preparation.
3. The TEXAS Model for Intersection Traffic: User's Guide. Center for Highway Research, Univ. of Texas at Austin, Research Rept. 184-3, in preparation.
4. The TEXAS Model for Intersection Traffic: Analysis of Signal Warrants and Intersection Capacity. Center for Highway Research, Univ. of Texas at Austin, Research Rept. 184-4, in preparation.
5. T. W. Rioux. Simulation of Traffic Movements in

an Intersection. Master's thesis, Univ. of Texas at Austin, May 1973.
6. J. E. Baerwald, ed. Traffic Engineering Handbook, Institute of Traffic Engineers, Washington, D.C., 3rd Ed., 1965.
7. D. R. Drew. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
8. J. Beakey. Acceleration and Deceleration Characteristics of Private Passenger Vehicles. Proc., HRB, 1938, pp. 81-89.
9. W. P. Bulloch, Jr. Driver-Vehicle Characteristics in Traffic Simulation. Univ. of Texas at Austin, master's thesis, Aug. 1973.
10. D. C. Gazis, R. Herman, and R. W. Rothery. Nonlinear Follow-the-Leader Models of Traffic Flow. Operations Research, Vol. 9, No. 4, 1960, pp. 545-567.
11. A. D. May, Jr., and H. E. M. Keller. Non-Integer Car-Following Models. HRB, Highway Research Record 199, 1967, pp. 19-32.
12. B. R. Greenshields, R. Schapiro, and E. L. Ericksen. Traffic Performances at Urban Street Intersections. Eno Foundation for Highway Traffic Control, Saugatuck, Conn., 1947.
13. D. S. Berry. Discussion of Relationship of Signal Design to Discharge Headway, Approach Capacity, and Delay, by G. Z. King and M. Wilkinson, TRB, Transportation Research Record 615, 1976, pp. 42-43.
14. I. H. C. Fett. Simulation of Lane Change Maneuvers on Intersection Approaches. Univ. of Texas at Austin, master's thesis, May 1974.
15. R. S. Walker. A Model for Traffic Simulation and Control. Univ. of Texas at Austin, master's thesis, Jan. 1970.
16. C. E. Lee and W. C. Vodrazka. Evaluation of Traffic Control at Highway Intersections. Center for Highway Research, Univ. of Texas at Austin, Research Rept. 78-1F, Mar. 1970.
17. G. E. Grayson. Evaluation of Traffic Signal Warrants Using the TEXAS Simulation Model. Univ. of Texas at Austin, master's thesis, Dec. 1976.
18. V. S. Savur. Determining the Capacity of Unsignalized Intersections Using Computer Simulation. Univ. of Texas at Austin, dissertation in preparation.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Abridgment

Postoptimality Analysis Methodology for Freeway On-Ramp Control

Menahem Eldor, Transportation Research Institute, Technion—Israel Institute of Technology
Ilan Adler, Operations Research Center, University of California, Berkeley

Postoptimality analysis is concerned with changes in an optimum decision value caused by changes in the parameters (input data) of a decision model. It is one way of approaching issues of uncertainty when using deterministic techniques such as linear programing (LP). We applied the techniques to a northbound section of the Eastshore Freeway (I-80) in the San Francisco Bay Area. The LP technique bases its calculations on point estimates rather than on a range of values. Postoptimality analysis assists in determining the importance and effects of deviations from such estimates.

The superiority of postoptimality analysis associated with LP over other mathematical programing techniques lies in its simplicity and systematic procedures. Postoptimality analysis allows us to obtain from the final (optimum) LP tableau (in addition to the optimum solution) a wealth of information on a wide range of operations in the neighborhood of the optimum.

Previous studies have focused on the potential applications of postoptimality analysis (1, 2). However, no such analysis has been attempted in recent applications of LP to freeway on-ramp control (3, 4).

It should be clear that one way to analyze postoptimality is to formulate and resolve a modified problem. This modification could, for example, be a slight change in one of the model parameters. Still, to investigate the effects of this slight change, the analyst must put this change into the model and rerun the program. Such a procedure is clearly inefficient and time consuming. Substantial economy of time and analysis is often possible if the information available in the optimum solution to the original problem is fully utilized instead. We shall demonstrate this.

THE LP CONTROL MODEL

The LP control model used here is similar to Wattleworth's original formulation and can be regarded as a resource allocation model. The resources—freeway subsection capacities—are allocated to competing input-station demands in order to maximize a certain objective function (for example, total allowable input rate) that is subject to the constraint of no congestion on the freeway and other operational constraints. The allowable flow rates at each input station are our decision variables, which are typically characterized by the upper and lower bounds imposed on them. Eldor has presented the mathematical details of the model (5, 6).

POSTOPTIMALITY ANALYSIS METHODOLOGY

The type of postoptimality analysis that can be performed on variations of a parameter depends upon its role in the optimization problem. The LP technique allows the effects of some variations to be examined quite easily.

Table 1. Capacity constraints: expected flows and slack and dual variables.

Subsection No.	No. of Lanes	Capacity	Expected Flow	Excess Capacity	Value of Dual 1[a]
1	3	5750	5404	346	0
2	3	5806	5806	0	0.227
3	3	5728	5263	465	0
4	3	5806	5671	135	0
5	3	5520	5295	225	0
6	3	5950	5539	411	0
7	3	5806	5083	723	0
8	3	5880	5803	77	0
9	3	5950	5803	147	0
10	3	5950	5577	373	0
11	3	5728	5146	582	0
12	4	6850	6188	662	0
13	3	5800	5800	0	1.069
14	3	5806	4642	1164	0
15	3	5800	4950	850	0
16	3	5049	4431	618	0
17	3	4746	4431	315	0
18	3	4700	4651	49	0

[a] Dual 1 (l) equals the value of the dual variable associated with the l th capacity constraint; the value of the objective function equals 8748 vehicles/h.

A detailed discussion regarding analysis of model parameters is given by Eldor (5, 6), who was concerned with the sensitivity of an optimum decision to possible variations on the right and upper-bounding vectors only.

Changes in the Right Vector

The right vector (b_k) represents the capacities of freeway subsections. Because these capacities are the true resources in the optimization process, it is often interesting to investigate the effects of their variations on the objective function (or the measure of effectiveness).

The dual variable associated with the kth capacity constraint indicates the rate of change in the objective function due to a unit change in the capacity of the kth subsection. In addition, one should investigate the range (of capacity) of this dual variable (or shadow price) when all else remains constant. Such an investigation is of practical importance to the analyst because it will assist him or her in determining the importance of deviating from his or her initial (point) estimate of capacity that was used by the LP model.

Let b_k' be a new right vector defined as

$$b_k' = \begin{cases} b_k + \delta, & k = h \\ b_k, & k \neq h \end{cases} \qquad (1)$$

Thus, δ represents a change in only one of the right vector entries (δ can be either positive or negative).

The specific questions often asked in this type of analysis are what the range of δ is for which the optimum basic sequence is still optimum or the optimum solution to the dual problem is still optimum, and what the corresponding change in the value of the objective function would be. We then formulated answers by using Equation 1 as a starting point and the information automatically generated by the LP algorithm (in particular the final working tableau). The measures were developed specifically for the upper-bounding version of the simplex method.

If both the immediate shadow price and the range of applicability of changes in a given capacity constraint are known, the analyst has a great deal of information about the value of changing a single capacity estimate. This we shall demonstrate with a real-life application.

Changes in the Upper-Bounding Vector

The LP decision model we used incorporates two types of upper bounds on the decision variables: the demand at an input station and the maximum metering rate limit. Because the larger of these parameters is clearly redundant for the optimization process, the analyst may combine the two into one constraint. Consequently, the redundant constraint (or quantity) does not enter explicitly the problem. An investigation of the variations in the upper-bounding vector entries and their effects on the value of the objective function thus deals with either of these two parameters (demand or maximum metering rate limit) at each input station. We followed procedures similar to the ones discussed above to develop and computerize (6) our information and measures.

POSTOPTIMALITY ANALYSIS APPLICATION

Northbound I-80 was selected as the test system for demonstrating the postoptimality analysis methodology. Our roadway and traffic data included the capacity profile of the freeway, 15-min origin-destination tables, and metering rate limits. The analysis concerns one 15-min interval of the afternoon peak period (5).

An efficient upper-bounding LP algorithm and the methodology presented above were computerized and integrated into a new software system called freeway responsive control optimization techniques (FRESCOT) (6), which is an efficient ANS FORTRAN traffic-management package for freeway on-ramp control. Eldor (6) gives detailed discussions of FRESCOT and a program listing and user's guide.

Using the computerized package with the input data, we derived a control strategy coupled with postoptimality analysis measures. The strategy and related measures of effectiveness are presented by Eldor (5, 6). Samples of the measures for the capacity constraints are given in Tables 1, 2, and 3. Each computer run also generates similar measures for the upper-bounding vector.

Table 1 includes the value of the dual variables, the slack variables (excess capacity), the expected flow in each subsection, and the capacity values used as input for the control run. Excess capacity is simply the difference between the capacity and the expected flow and is exactly the value of the slack variable (at optimality) introduced into a capacity constraint in the process of converting the LP problem into a standard form (all constraints are converted into equalities). The slack variable measures the rate of capacity underutilization of a subsection. A positive slack means that the constraint is nonbinding. The slack variables are directly associated with the dual variables, which are equal to zero for all the nonbinding constraints.

The optimality ranges for changes in the right vector are given in Table 2. For each subsection, a range of applicability of the capacity estimate is given with the corresponding range of changes in the value of the objective function. For example, consider subsection 2. The initial estimate of capacity for this subsection was 5806 vehicles/h. The range of optimality for this capacity estimate, over which the associated dual variable is unchanged, is 5757 to 5852; the corresponding range for the objective function is 8737 to 8759. The results in Table 2 show the effects of deviating from the initial capacity estimate.

Table 3 provides the associated control strategy with each change (at the bounds) of the capacity vector as given in Table 2. The computer generates a strategy for each binding capacity constraint (6). The strategy for the nonbinding capacity constraints (within the range

Table 2. Ranges of optimality for vector changes.

	Optimality Range of Subsection Capacity				Optimality Range of Objective Function			
Subsection No.	Change of Lower Bound	Change of Upper Bound	Lower-Bound Value	Upper-Bound Value	Change of Lower Bound	Change of Upper Bound	Lower-Bound Value	Upper-Bound Value
1	-346		5404		0	0	8748	8748
2	-49	46	5757	5852	-11	10	8737	8759
3	-465		5263		0	0	8748	8748
4	-135		5671		0	0	8748	8748
5	-225		5295		0	0	8748	8748
6	-411		5539		0	0	8748	8748
7	-723		5083		0	0	8748	8748
8	-77		5803		0	0	8748	8748
9	-147		5803		0	0	8748	8748
10	-373		5577		0	0	8748	8748
11	-582		5146		0	0	8748	8748
12	-662		6188		0	0	8748	8748
13	-751	35	5049	5835	-802	38	7946	8786
14	-1164		4642		0	0	8748	8748
15	-850		4950		0	0	8748	8748
16	-618		4431		0	0	8748	8748
17	-315		4431		0	0	8748	8748
18	-49		4651		0	0	8748	8748

Table 3. Strategy for ranges of optimality for vector changes.

	Subsection 2		Subsection 13	
Origin	Lower-Bound Capacity	Upper-Bound Capacity	Lower-Bound Capacity	Upper-Bound Capacity
1	5404	5404	5404	5404
2	353	448	402	402
3	408	408	408	408
4	244	244	244	244
5	720	720	720	720
6	1080	1007	240	1080
7	308	308	308	308
8	220	220	220	220

of optimality) remains unchanged. Thus, four additional, meaningful control strategies that are associated with possible changes in the initial capacity estimates of the bottleneck subsections are provided.

SUMMARY AND CONCLUSIONS

We present postoptimality analysis methodology for applying LP to freeway on-ramp control. In addition to the optimum control strategy for the original (or initial) data set, the analyst is provided with much valuable information concerning the deviations from the initial capacity and upper-bound estimates. Not only are ranges of optimality given with their associated changes in the value of the objective function, but the corresponding control strategies are also provided. The expense of generating this information is, practically speaking, negligible. Only simple calculations are required (6), and the computerization of these calculations can be considered as a one-time effort.

The methodology developed in this study is also applicable to priority-entry LP on-ramp control. Expansion of the FRESCOT software system to account for priority entry schemes has been initiated and will be reported at another time.

REFERENCES

1. J. A. Wattleworth and D. S. Berry. Peak-Period Control of a Freeway System—Some Theoretical Investigations. HRB, Highway Research Record 89, 1965, pp. 1-25.
2. A. D. May. Discussion of Peak-Period Analysis and Control of a Freeway System, by J. A. Wattleworth, HRB, Highway Research Record 157, 1967, pp. 15-21.
3. J. J. Wang and A. D. May. Computer Model for Optimal Freeway On-Ramp Control. HRB, Highway Research Record 469, 1973, pp. 16-25.
4. K. Ovaici, R. F. Teal, J. K. Ray, and A. D. May. Developing Freeway Priority-Entry-Control Strategies. TRB, Transportation Research Record 533, 1975, pp. 122-137.
5. M. Eldor and I. Adler. Postoptimality-Analysis Methodology for Freeway On-Ramp Control. Transportation Research Institute, Technion—Israel Institute of Technology, Haifa, Publication 77-8, 1977.
6. M. Eldor. Integrated Real-Time Freeway On-Ramp Control Strategies. Institute of Transportation Studies, Univ. of California, Berkeley, PhD dissertation, 1976, 625 pp.

Publication of this paper sponsored by Committee on Freeway Operations.

Abridgment

Areawide Impact of Traffic Control Devices

Gerard M. Ebbecke, Philadelphia Streets Department
James J. Schuster, Institute for Transportation Studies, Villanova University

The Philadelphia traffic signal demonstrations were a particularly effective manifestation of the power-to-the-people mood of the mid-1960s. At such gatherings, neighborhood residents would band together to blockade traffic from a public roadway; their purpose was to secure a traffic signal at a particular intersection. Initially, these demonstrations were spontaneously generated by a traffic accident. However, as they received growing media coverage, demonstrators lost their spontaneity and began to increase in numbers and to create community unrest. In the latter stages of this period, they would often refuse to clear an intersection until installation of a control had actually commenced. Nor was it uncommon to have simultaneous demonstrations at several adjoining intersections.

The city administration was afraid that arresting demonstrators would escalate the relatively minor disturbances into a major confrontation of the 1964 Philadelphia type that resulted in a $4 million loss of private property. The consequence of this policy of acceding to demonstrator's demands was that between 1964 and 1969 two intersections a week were being signalized, three times the normal rate. Of the 364 signals installed during this era, 65 percent were initiated by demonstrations. It became evident after several years that this situation was likely to continue, and therefore a generally acceptable alternative to the traffic signal became imperative. Thus it was decided in late 1967 to use the four-way stop in Philadelphia.

THE FOUR-WAY STOP PROGRAM

Judicious use of four-way stops had proved effective at several problem locations since the early 1960s. However, city traffic engineers were reluctant to use the device on a large scale because it lacked substantive evaluation as a safety device and because by impeding the flow of traffic it conflicted with a basic traffic engineering principle.

In addition, recommended guidelines governing the use of four-way stops needed modification. The 1971 Manual on Uniform Traffic Control Devices (MUTCD) (1) suggested its use at locations with combined approach traffic volumes averaging "at least 500 vehicles per hour for 8 hours of an average day." This would have limited its use in Philadelphia to only those intersections handling more than 9500 vehicles/d. Low-volume intersections, despite the fact that they may have had the most acute accident problems, were therefore excluded from consideration. It should be noted that 60 percent of the traffic signals installed from 1964 to 1969 did not meet the guidelines. Intersection 1 in the following example is ineligible for four-way stop control, even though its accident rate (accidents/10 million vehicles) is twice that of intersection 2.

Intersection	Average Daily Traffic	Accidents/Year	Accident Rate
1	4 000	4	28.4
2	10 000	5	14.2

Furthermore, the MUTCD suggested that four-way stops be used at locations where accident problems were "indicated by five or more reported accidents of a type susceptible of correction by a 4-way stop in a 12-month period." However, although in some states any damage at all constitutes a legal accident, in others (including Pennsylvania) damage must exceed several hundred dollars. Thus a four-way stop (or traffic signal for that matter) is more easily warranted in those states having a more liberal accident definition. These criteria, then, were deemed too restrictive for practical application.

A 1967 study of nearly 300 Philadelphia intersections designated a dangerous intersection as one where the yearly accidents exceed half of the average daily traffic (ADT) in thousands (2). This was the basic parameter, together with good engineering judgment, that was used to establish the need for a four-way stop. A 1970 before-and-after analysis of those initial 57 intersections converted from two-way to four-way stop control indicated that accidents decreased by 87 percent, and personal injuries decreased by 92 percent (3). The four-way stop program has proved successful in reducing the number of yearly new traffic signal installations by 65 percent. Unfortunately, the four-way stops have proliferated over the past 8 years, until by the end of 1976 they controlled some 1800 of Philadelphia's 20 000 intersections.

Although these two studies served to establish the need and effectiveness of the initial four-way stop installations, there remained a critical need for an objective examination of the wide-range effects of the citywide traffic control changes that averaged 5/week over the past 9 years. Initially four-way stop installations, because they were uncommon, may have commanded more respect than the multitudes of such installations do today.

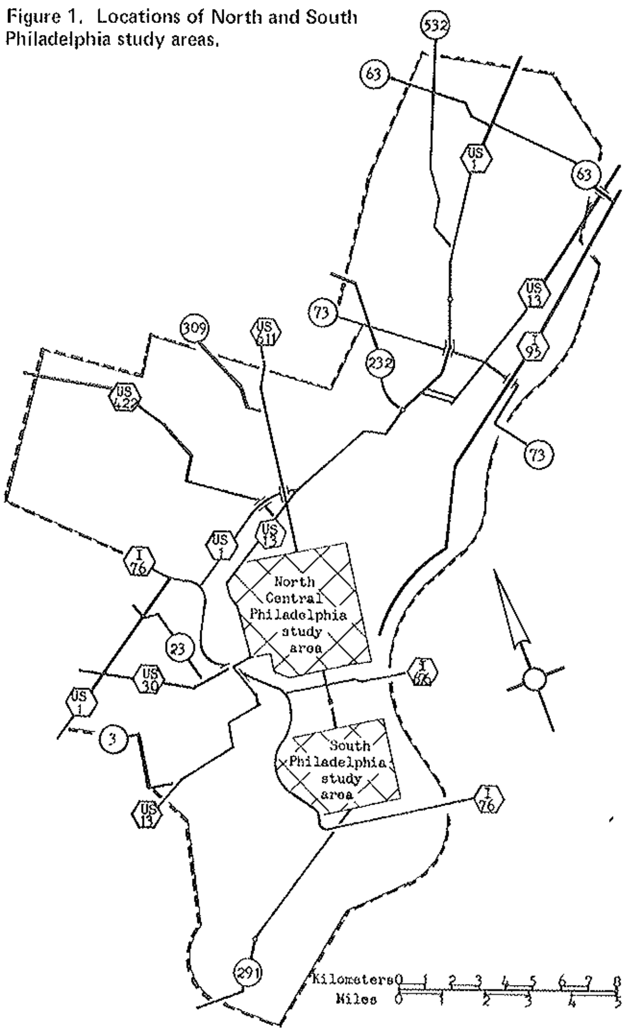
SELECTING THE STUDY AREA

The city of Philadelphia encompasses 336 km^2 (130 miles^2). Detailed accident data were available for 1968 through 1976, and our object was to select those areas in which the traffic control device itself was the only variable. Remote sections developing relatively rapidly, and hence experiencing changing traffic patterns, were rejected. Other areas proved undesirable because of the probable influence of varying street widths and vehicle speeds.

Two geographically distinct but geometrically similar areas of the city were finally selected: south Philadelphia and north central Philadelphia (Figure 1). A grid network of cartways 8 m (26 ft) wide with one-lane, one-way street intersections predominates, so all intersections were included in the study. This totaled 449 locations in the southern area and 444 in the northern area. Even though both combined constitute only 10 percent of the total city area, they contain 20 percent of the city's 3300 traffic signals and 1800 four-way stops today. Nine years ago, 40 percent of the study intersections were controlled by traffic signals, and the remainder were two-way stop controlled. Today, 40 percent are still signalized and 20 percent remain two-way stop controlled; 40 percent are now being four-way stop controlled. During the 9 years, 33 study intersections were newly signalized, and 352 intersections were converted from two-way to four-way stop control.

It must be emphasized that none of the 385 traffic signals studied was either interconnected or coordinated with another signal. All comprise two signal heads (green, amber, and red) post-mounted on each of the four intersection corners, thus facilitating two farside indications for each traffic approach and pedestrian indications at each end of the four intersection crosswalks. None of the study installations has pedestrian signals. Stop-sign controlled approaches have both right and left stop signs posted, the right stop carrying a four-way rider for all those intersection approaches where applicable.

Figure 1. Locations of North and South Philadelphia study areas.



TRAFFIC VOLUMES

In the spring and fall of 1973, 1500 manual, vehicle classification, turning movement, intersection counts and one thousand 7-d, 24-h automatic traffic recorder (ATR) counts were conducted in both study areas. Results indicated that

1. Ninety percent of the study intersections have an ADT of 1000 to 9000 vehicles/d;
2. ADT of the 384 traffic signals of both areas is 6700;
3. ADT of the 154 two-way stopped intersections is 4400;
4. ADT of the 198 south Philadelphia four-way stops is 6200;
5. ADT of the 157 north Philadelphia four-way stops is 3600; and
6. At more than 90 percent of the locations studied the minor street accounted for 20 percent more of the total intersection ADT.

A comparison of the 1973 south Philadelphia intersection volumes with the 1960 area traffic study confirmed no area traffic growth. A similar north Philadelphia intersection study revealed that the 1973 volumes were almost identical to those of the earliest recorded citywide traffic study done in 1938! Thus, the traffic volumes in both study areas were indeed static throughout the 9-year study period.

TRAFFIC ACCIDENTS

A total of 19 492 Philadelphia Police Department accident records for 1969 to 1976 were analyzed for the 893 study locations. Sheer magnitude precluded summarization of all but the essential items: date of occurrence, severity (property damage, occupant injury, pedestrian, fatality), type (right angle, rear end, fixed object, sideswipe, pedestrian), and time of day (daylight, darkness).

Accident Severity, Type, and Time

Table 1 summarizes the accident severity, type, and time of day data for each of the three modes of intersection traffic control for both study areas. Figures 2 and 3 illustrate relative impact, and the results indicate the following:

1. One of every 8 intersection accidents involves a pedestrian regardless of the mode of traffic control;
2. Whereas 1 of every 5 accidents at traffic signals and two-way stops in these areas results in an occupant injury, only 1 of every 10 accidents at four-way stops results in an occupant injury;
3. Accidents involving two occupied vehicles (right angle, rear end, and sideswipe) constitute 61 percent of the signal accidents, 67 percent of the two-way stop accidents, and only 46 percent of the four-way stop accidents;

Table 1. Percentages of accidents for each traffic control mode.

		Severity				Type					Time		
Location	Control	Property Damage	Personal Injury	Pedestrian	Fatality	Right Angle	Rear End	Fixed Object	Sideswipe	Pedestrian	Day	Night	AADT
South Philadelphia	Traffic signals	70	18	12	0	26	24	31	7	12	66	34	6700
	Two-way stops	68	20	12	0	51	12	20	5	12	68	32	4400
	Four-way stops	77	11	12	0	23	17	40	8	12	67	33	6200
North Philadelphia	Traffic signals	68	20	12	0	33	23	22	10	12	67	33	6900
	Two-way stops	69	19	12	0	50	10	22	6	12	74	26	4400
	Four-way stops	79	8	13	0	17	17	44	9	13	67	33	3600

4. Right-angle, rear-end, and fixed-object accidents occur in comparable numbers at signals; half of the two-way stop accidents are right angle, but fixed-object accidents predominate at four-way stops by default; and

5. Night driving proves 50 percent more hazardous than day, since a third of the accidents occur at night, which has a fourth of the ADT.

Accident Rates

Table 2 summarizes the accident rates over the 9-year study period for the three modes of intersection traffic control for both study areas (see Figures 4 and 5 for the trends), and the results indicate the following:

1. The two-way stop rates have increased linearly throughout the period, doubling in only 6 years. Reasoning backward from this growth leads to the obviously incorrect conclusion that these rates were 0 in 1960 (some external source has affected the two-way stop accident rates in both areas);

2. The south Philadelphia two-way stop rate is

Figure 2. Accident severity.

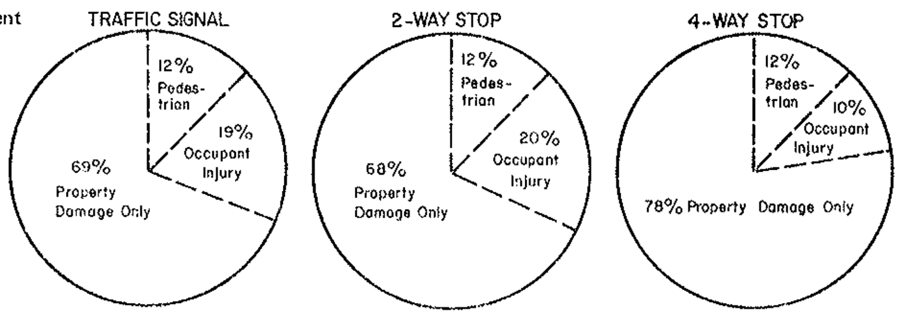


Figure 3. Accident type.

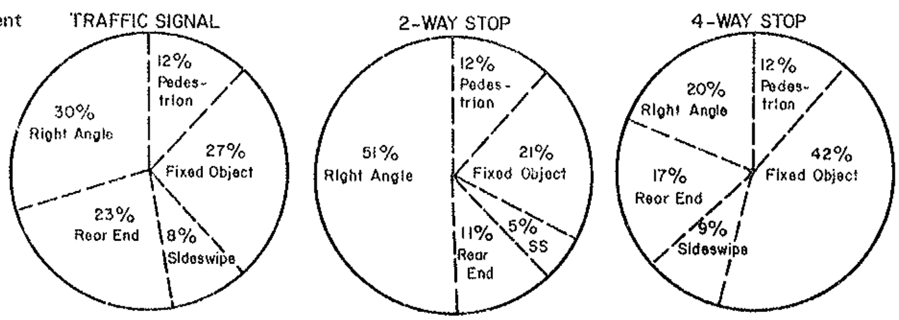


Figure 4. Accident rate for South Philadelphia.

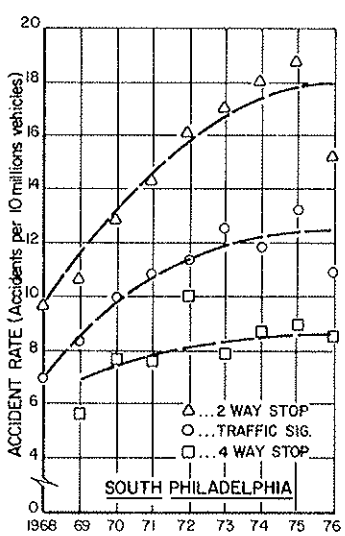


Figure 5. Accident rate for North Philadelphia.

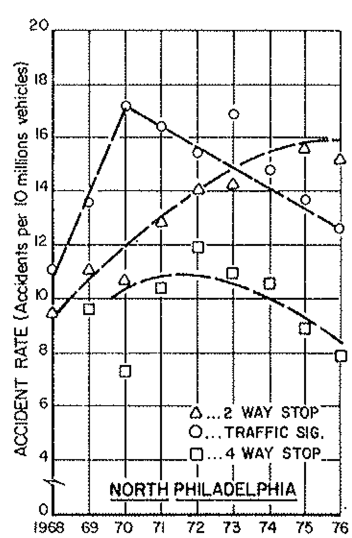


Table 2. Annual accident rate for each intersection control mode.

Location	Control	1968	1969	1970	1971	1972	1973	1974	1975	1976
South Philadelphia	Traffic signals	6.9	8.4	9.9	10.8	11.4	12.6	11.9	13.3	11.0
	Two-way stops	9.7	10.6	12.9	14.4	16.1	17.1	18.1	18.8	15.3
	Four-way stops	—	5.6	7.7	7.6	10.0	7.9	8.7	9.0	8.5
North Philadelphia	Traffic signals	11.2	13.7	17.3	16.5	15.5	16.9	14.8	13.8	12.6
	Two-way stops	9.5	11.1	10.7	12.9	14.1	14.3	16.5	15.6	15.3
	Four-way stops	—	9.7	7.3	10.4	11.9	11.0	10.6	8.9	7.9

Note: Rate = accidents/10 000 000 vehicles.

plateauing at about 18;

3. The south Philadelphia signal rate is plateauing at about 12.5;

4. The south Philadelphia four-way stop rate is plateauing at about 18.5, which is lower today than the other two but higher than they were in 1967;

5. The north Philadelphia two-way stop rate is plateauing at about 16; and

6. For several years the north Philadelphia signal and four-way rates have been on the decline, the latter always 30 percent lower than the former.

BEFORE-AND-AFTER ANALYSIS OF ALL TRAFFIC CONTROL CHANGES

To ensure a definitive evaluation of the before-and-after statistics at all locations where the control mode changed during the study period, we decided to compare the 2-year period both immediately before and immediately after the change. Thus all changes from 1970 to 1974 were evaluated; however, the seven newly signalized locations were insufficient for definitive conclusions.

Results for the 222 two-way to four-way conversions, which the Traffic Engineering Handbook (4) criteria labeled dangerous locations (that is, having an accident rate of nine), indicate the following:

1. Accident reduction conversion results in both study areas were similar;
2. In general, three of every four conversions from two-way to four-way stop control improved conditions, regardless of the before accident rate;
3. Half of the safe two-way conversions to four-way increased accidents;
4. Six of seven dangerous two-way conversions reduced accidents;
5. Total accidents decreased by 55 percent after conversion to four-way stop;
6. Occupant personal injury accidents decreased by 81 percent after conversion;
7. Pedestrian injury accidents decreased by 83 percent after conversion;
8. Right-angle accidents decreased by 83 percent after conversion; and
9. Rear-end, fixed-object, and sideswipe accidents were unaffected.

REFERENCES

1. Manual on Uniform Traffic Control Devices. Federal Highway Administration, 1971.
2. J. Heany. How to Identify Dangerous Intersections. Drexel Univ., 1967.
3. J. Heany. The Four-Way Stop:—Effective Safety Device. Public Works, 101, 1970.
4. Traffic Engineering Handbook. Institute of Traffic Engineers, 3rd Ed., 1965.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Approach to Real-Time Diversion of Freeway Traffic for Special Events

Graeme D. Weaver, Conrad L. Dudek, Donald R. Hatcher, and William R. Stockton, Texas Transportation Institute, Texas A&M University

In Dallas, Texas, on July 4, 1976, freeway traffic bound for a fireworks display was diverted to an alternate arterial route. The object was to validate primary candidate messages and displays resulting from extensive laboratory studies of human factors. Two primary candidate messages were displayed at alternate times on matrix signs located on the Central Expressway. The first message caused 56.2 percent of the traffic to divert and the second 43.8 percent.

Special events at places such as stadiums generate large volumes of traffic and congestion at the site and on adjacent freeways. Practically every driver on the way to a ball game at a major stadium has experienced considerable delay in lengthy queues.

Less congested alternate routes are often available, however, and, if some of the approaching freeway traffic can be diverted to alternate routes, congestion can be reduced. This depends on several factors: (a) An acceptable alternate route must be available; (b) drivers must be made aware of the alternate route; and (c) guidance must be provided along the alternate route so that drivers, once diverted from the primary route (usually a freeway), can progress easily and confidently along the alternate. Research findings in real-time driver information transfer techniques that have been developed from human factors engineering principles are of critical importance to effective route diversion.

The purpose of this study is to develop effective information displays for real-time incident management and route diversion. Extensive human factors laboratory studies are being conducted to develop primary candidate messages and displays that will then be field validated. One phase of the research effort was directed toward route diversion messages and displays for special events. This paper discusses one such study conducted in Dallas, Texas, during a Fourth of July fireworks display at the Fair Park in 1976.

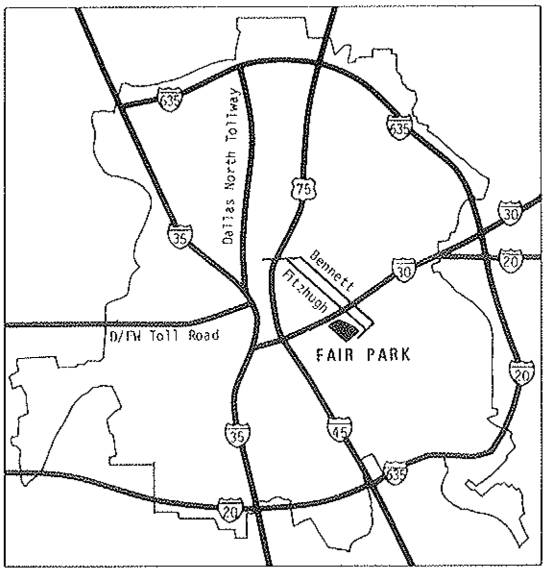
FAIR PARK CHARACTERISTICS

Fair Park, a 97-hm^2 (240-acre) area in south central Dallas, houses the Cotton Bowl and permanent buildings, facilities, and midway for the annual Texas State Fair. In addition it has several cultural buildings such as a music hall, museums, and the Texas Hall of State. Many college and professional football games are played in the Cotton Bowl (seating capacity of 73 000), which is often used for other events and exhibitions such as Fourth of July celebrations.

The location of Fair Park relative to the freeway system is shown in Figure 1. The primary access route to Fair Park and its parking facilities from the north, west, and south is provided by two exit ramps (Second Avenue and Haskell Avenue exits) from I-30. Traffic from US-75 (Central Expressway), I-35E, and the Dallas-Ft. Worth Turnpike must connect with I-30 and then exit via either Second Avenue or Haskell Avenue. Queues of exiting traffic to Fair Park will often extend back for 3.2 km (2 miles) during peak demand conditions and will affect I-30 and even US-75 and I-35E. This occurs regularly before the Cotton Bowl game on New Year's Day or the Oklahoma-Texas University game in October. Each of these games attracts more than 73 000 people. The October game also coincides with the Texas State Fair, which attracts an additional 225 000 people to Fair Park on the same day.

Fourth of July celebrations are held annually at the Cotton Bowl. The study year, that of the bicentennial celebration, attracted more than 65 000 people for various events, 42 000 of whom attended an elaborate fireworks display that began at 9:15 p.m. in the Cotton Bowl. This special event was selected as the first of three to evaluate how well the matrix sign messages accomplished route diversion.

Figure 1. Dallas Fair Park.



STUDY SITE

We chose to study the southbound Central Expressway (US-75), because a suitable arterial street was available as an alternate route to Fair Park. The primary route, Central Expressway, and the alternate route, Fitzhugh Avenue, are shown in Figure 2. The Expressway from Fitzhugh Avenue to I-30 is a six-lane facility with a direct two-lane ramp connection to eastbound I-30. Fitzhugh Avenue is a one-way arterial with eight traffic signals between Central Expressway and Fair Park. The street narrows from three to two lanes approximately 1.6 km (1 mile) east of the expressway. Alignment is straight. Traffic signals at the Fitzhugh-Central Expressway diamond interchange were timed to favor the expected left-turn demand, and the signals along Fitzhugh Avenue were timed and coordinated to provide steady progression. Parking was prohibited along Fitzhugh Avenue to provide a minimum of two operating lanes throughout the complete alternate route.

SIGNS

Matrix Sign Location

Two trailer-mounted lamp matrix signs were positioned on the overcrossing structures above the southbound

Figure 2. Primary and alternate routes.

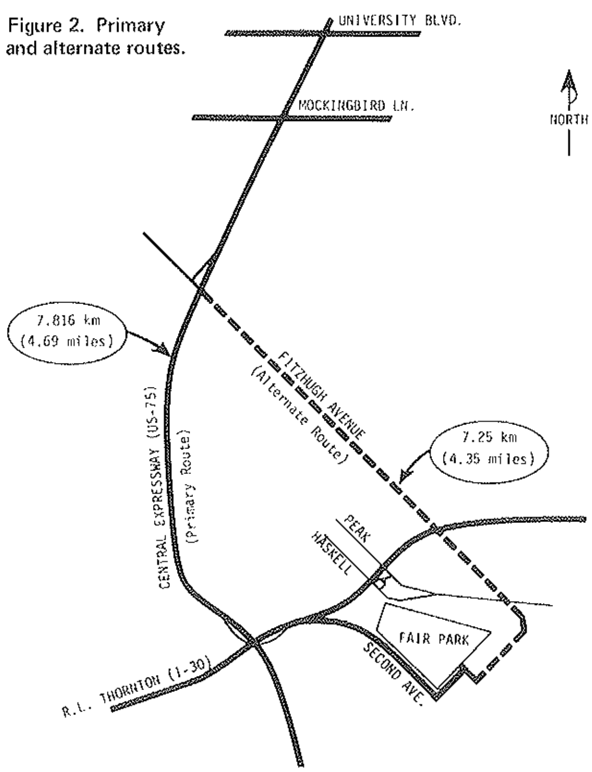
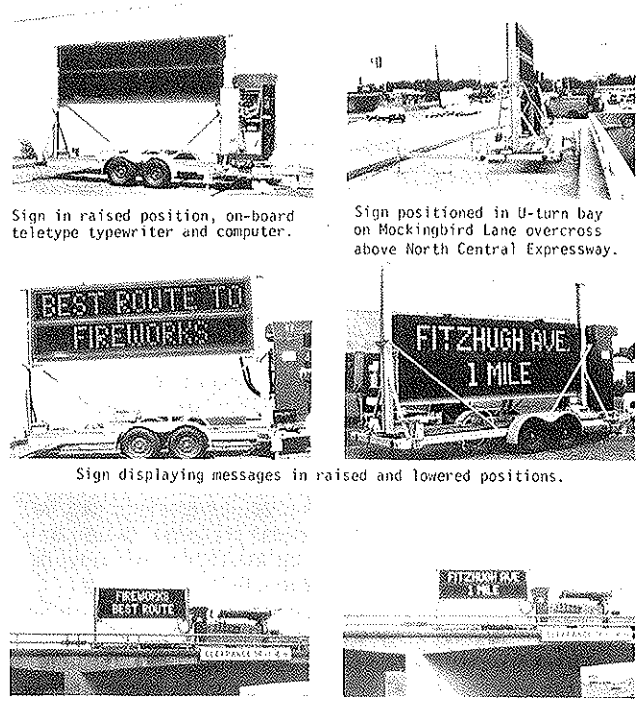


Figure 3. Trailer-mounted matrix sign.



Sign in raised position, on-board teletype typewriter and computer.

Sign positioned in U-turn bay on Mockingbird Lane overcross above North Central Expressway.

Sign displaying messages in raised and lowered positions.

Sign displaying messages during study.

lanes of Central Expressway. One sign was located on the University Avenue overcross, approximately 3.2 km (2 miles) north of the Fitzhugh Avenue exit. The second sign, on the Mockingbird Lane overcross, was approximately 2.4 km (1.5 miles) north of the Fitzhugh Avenue exit. Unobstructed sight distance to southbound traffic was greater than 366 m (1200 ft) at each location.

Matrix Sign Characteristics

The portable matrix signs are illustrated in Figure 3. The signs can display messages on two lines using 46-cm (18-in) characters. A computer located on the front side of the trailer provides almost unlimited message selections and displays. Messages can be displayed in a stationary mode or flashed or alternated with other messages. Message displays are commanded via a teletype on each sign trailer. Each sign system can be either hooked up to regular line power or connected to a generator.

Matrix Sign Messages

The two primary candidate messages shown below were selected for evaluation. These messages were the results of extensive human factors laboratory studies.

Sign	Message 1	Message 2
1 (University Avenue)	BEST ROUTE TO FIREWORKS	ROUTE TO FIREWORKS
	USE FITZHUGH AVE	INFORMATION AHEAD
2 (Mockingbird Lane)	FIREWORKS BEST ROUTE	FIREWORKS BEST ROUTE
	FITZHUGH AVE 1 MILE	FITZHUGH AVE 1 MILE

Message 1 was designed so that the two sign messages were so redundant that even if a driver saw only one sign the information would still be conveyed. Message 2 simulated a signing system where the first sign (most likely static) would alert drivers that information concerning the best route would be given downstream. The intent was to evaluate the need for redundancy.

Human factors laboratory studies indicated that care must be exercised in selecting the destination name used on the sign display. Although the Cotton Bowl-Fairgrounds complex in Dallas is locally called Fair Park, this name was ruled out for matrix signs because it might not have been understood by nonlocal drivers. Highway maps designate the area as Cotton Bowl-State Fairgrounds. Since the major event on the Fourth of July was the elaborate fireworks display, the decision was made to use FIREWORKS. COTTON BOWL would be the preferred choice for football games, whereas FAIRGROUNDS would be appropriate for the state fair.

Human factors laboratory studies also suggested that it would be better to break long messages into chunks than to display the entire message at once. In addition, the message chunks should each be a complete phrase. Chunks or phrases can be displayed alternately on the sign to form the complete message. Sequencing the message phrases has the added advantage of attracting the attention of the drivers.

Trailblazer Signs

Human factors laboratory studies indicated a great need for route guidance along the alternate route for drivers who diverted from their primary routes. Results of the studies also indicated that symbolic or logo signs helped provide trailblazing information to drivers along an unfamiliar route if these were proper transitions between the primary message on the freeway and the logo trailblazers.

The Dallas traffic operations personnel requested consideration of a unique Fair Park logo and word-message trailblazer they were developing. The trailblazer signs adopted cooperatively by the city of Dallas and the Texas Transportation Institute (TTI) research staffs for the study are shown in Figure 4.

The most difficult task in using trailblazers, particularly logo signs, occurs in the transition region where the driver first encounters the sign. To reduce the driving task load as much as possible during exit from a freeway and subsequent left-turn maneuver (an already loaded state), a series of transition signs was used to guide drivers through the signalized intersection and properly orient them along Fitzhugh Avenue.

Immediately after gaining correct alignment on the southbound service road, the driver was presented sign 1 in Figure 4. Sign 2 was displayed at the signal. After turning left at the Fitzhugh Avenue intersection and crossing the Expressway the driver could easily see sign 3 on the narrow separator median immediately beyond the northbound service road. Two subsequent signs of sign 4 style were located several blocks down Fitzhugh Avenue to guide the driver in the transition from the Fireworks supplementary panel to the logo trailblazer shown in signs 5 or 6. Signs with appropriate straight, advance turn, or turn arrows were located along Fitzhugh Avenue to near the entrances to Fair Park parking lots. In this vicinity, Public Parking signs with a P logo and directional arrows (signs 7 and 8) were added to the basic trailblazer sign for guidance to parking areas.

All trailblazers were fabricated of high-intensity reflective sheeting on aluminum panels 61 × 61 cm (24 × 24 in) in size. The supplementary Fireworks panels in signs 1, 2, 3, and 4 were 15.2 × 61.0 cm (6 × 24 in). The Fair Park sign was a brown panel supporting a yellow and orange logo, white Fair Park legend, and white arrow. Parking signs were green on white.

After the July 4 studies, the upper panels on signs 1, 2, and 3 were replaced by one saying COTTON BOWL and FAIRGROUNDS on a two-line panel. Likewise, the FIREWORKS legend on sign 4 was replaced with COTTON BOWL-FAIRGROUNDS. These more generalized signs then remained in position until subsequent special events.

FIELD STUDIES

Matrix Sign Operation

The intended plan was to display each of the two alternative messages on the University Avenue and Mockingbird Lane overcrosses according to the following time schedule:

Time (p.m.)	Message	Time (p.m.)	Message
6:30 to 7:00	2	7:50 to 8:05	2
7:00 to 7:10	blank	8:05 to 8:15	blank
7:10 to 7:40	1	8:15 to 8:30	1
7:40 to 7:50	blank		

Unfortunately, during the study the Mockingbird sign malfunctioned between 6:30 and 6:45 p.m. and was inoperative from 6:45 to 7:00 p.m. This required field adjustments to the display schedule.

The original intent was to display messages 1 and 2 twice each during the study period. Although message 1 was successfully replicated, the Mockingbird sign

Figure 4. Trailblazer signs.

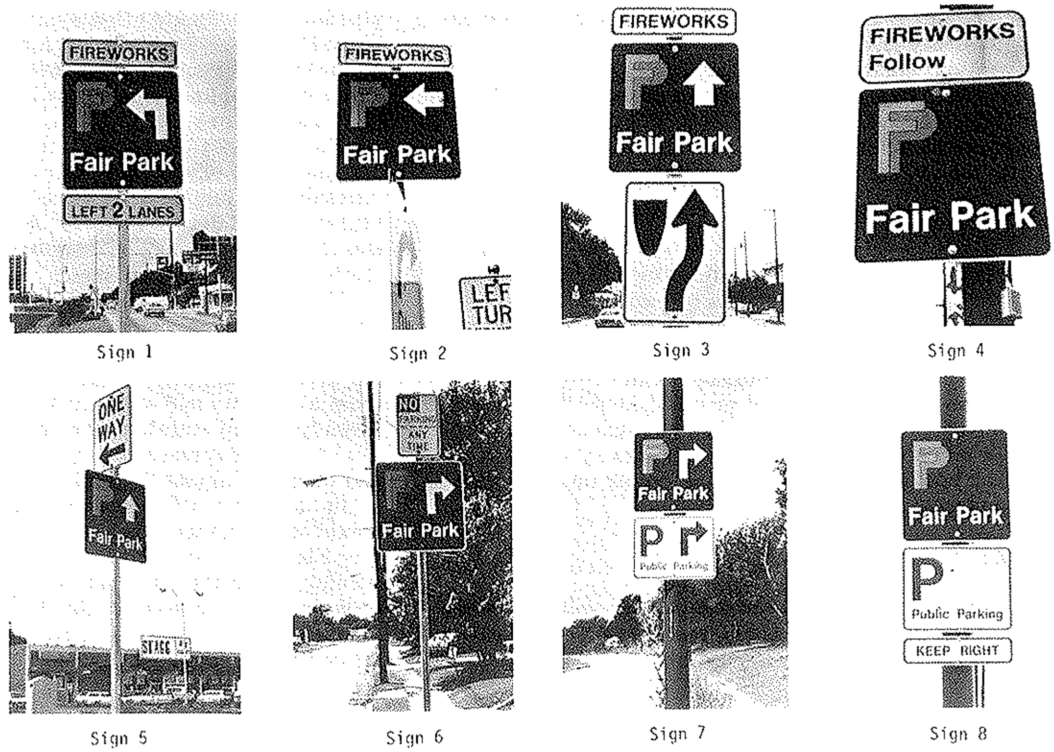
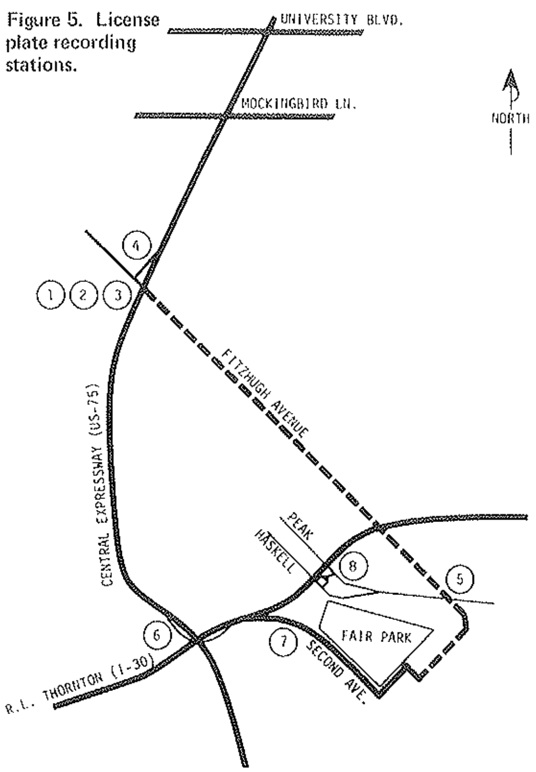


Figure 5. License plate recording stations.



failed to display the proper sequence of message 2 during the initial phase of the study. In addition, the field crew was trying to put up the correct message, which was therefore not displayed continuously throughout the 15-min period.

The actual message display schedule was as follows:

Time (p.m.)	Message	Time (p.m.)	Message
6:30 to 6:45	2a (malfunction)	7:55 to 8:10	2
6:45 to 6:55	blank	8:10 to 8:20	blank
6:55 to 7:00	1 (University only)	8:20 to 8:25	1
7:00 to 7:35	1	8:25 to 8:30	blank
7:35 to 7:55	blank		

Data Collection

Comprehensive origin-destination data were collected throughout the study. License plates of almost every vehicle passing the eight locations shown in Figure 5 were recorded continually throughout the 2 h 15 min study. The locations were selected to permit as accurately as possible a determination of travel patterns on both primary and alternate routes and to include all vehicles that passed the matrix signs on the way to Fair Park.

License plates were recorded on cassette tape for each location. The heavy volume of main lane traffic passing the Fitzhugh Avenue overcross necessitated individual lane surveillance. License plates were read from the overcross sidewalk through high-powered binoculars mounted on tripods. Recorders rotated every 10-min and provided better than a 95 percent count when compared to the main-lane volume data from the permanent detectors. Observers also recorded the time to ensure proper coordination with the sign messages.

License plate information was transcribed from the cassette tapes, punched on computer cards, and processed to identify matches at various recording locations. This allowed us to trace travel patterns of almost all the vehicles without spot-sampling and interviewing drivers at selected locations.

RESULTS

A comparison of the route choices by southbound Central Expressway drivers destined to Fair Park is presented in Table 1 and Figure 6. The data are presented by

Table 1. Route choices by Central Expressway drivers.

		Drivers to Fair Park			
		Using Fitzhugh Avenue			
Message	Time Period (p.m.)	No.	Percent	Using US-75/I-30	Total
Blank	6:17 to 6:32	8	10.1	71	79
2a	6:32 to 6:47	30	28.6	75	105
Blank	6:47 to 6:57	15	15.5	82	97
1	6:57 to 7:37	251	68.6	115	366
Blank	7:37 to 7:57	13	10.2	114	127
2	7:57 to 8:12	101	56.4	78	179
Blank	8:12 to 8:22	10	16.1	52	62
1	8:22 to 8:27	33	70.2	14	47

Table 2. Message effects on route diversion from Central Expressway.

	Drivers to Fair Park				
	Using Fitzhugh Avenue				
Message	No.	Percent	Using US-75/I-30	Total	Drivers Diverted to Fitzhugh Avenue (%)
Message 1	284	68.8	129	413	56.2
Message 2	101	56.4	78	179	43.8
Message 2a[a]	30	28.6	75	105	16.0
No message	46	12.6	319	365	–
Before study (May 22, 1976)	22	9.8	202	224	–

[a]Sign malfunction.

Figure 6. Percentage of drivers using Fitzhugh Avenue to Fair Park.

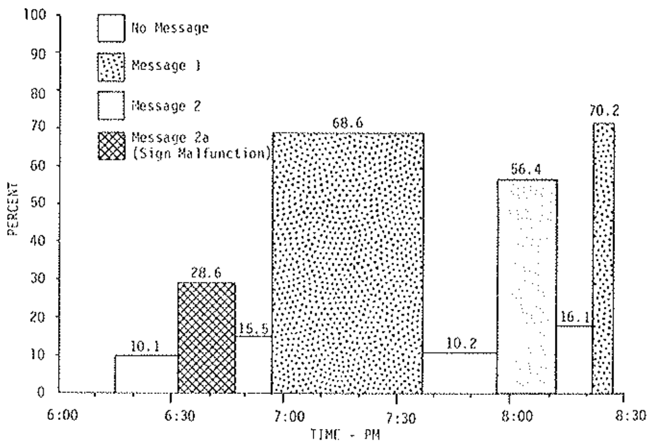
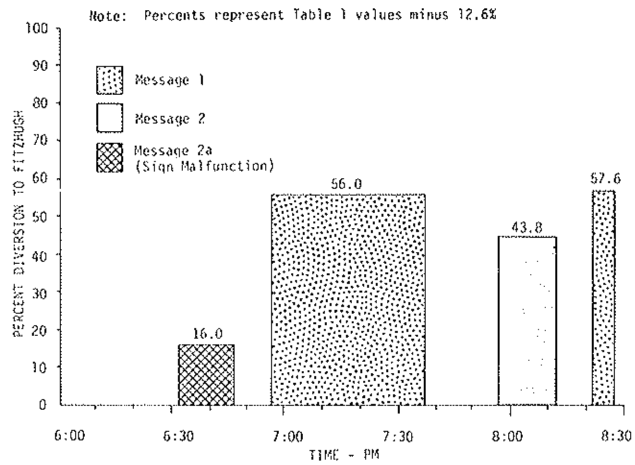


Figure 7. Effects of matrix sign messages on route diversion.



time periods in relation to times when matrix sign messages were displayed. Since the matrix sign nearest to the Fitzhugh exit ramp was about 2.4 km (1.5 miles) upstream, the effects of the messages would not be noticed at the Fitzhugh ramp until about 1.6 min after a sign message was displayed [assuming 88 km/h (55 mph) average speed], the time periods shown in Table 1 and Figure 6 have therefore been adjusted by 2 min.

The results reveal a very pronounced and positive effect on the number of drivers using the Fitzhugh Avenue alternate route when matrix sign messages 1 and 2 were displayed. The percentage of drivers using Fitzhugh when no messages were displayed ranged between 10.0 and 16.1 percent. In contrast, between 56.4 and 70.2 percent took the alternate route when messages 2 and 1 were displayed (these data do not include percentages for message 2a when the Mockingbird sign malfunctioned).

Table 2 and Figure 7 are presented to illustrate the amount of diversion resulting from the matrix sign messages. During the July 4 study an average of 12.6 percent of the drivers routinely chose the Fitzhugh route (no message displayed). This compares favorably with earlier studies conducted by TTI on Sunday, May 23, 1976, when 9.8 percent of the drivers used the Fitzhugh route (the midway is open every weekend).

Subtracting the 12.6 percent from the percentage using Fitzhugh Avenue when specific messages were displayed yields the following results: message 1 with its redundant sign messages influenced 56.2 percent (weighted average, Table 2) of the drivers to divert to the alternate route; message 2 and its advanced warning and route information resulted in 43.8 percent diversion. On the average, 52.4 percent of the drivers diverted when a message was displayed.

ACKNOWLEDGMENTS

This paper discusses one phase of a research project entitled Human Factors Requirements for Real-Time Motorist Information Displays, conducted by the Texas Transportation Institute and sponsored by the Federal Highway Administration. The contents of this paper reflect our views, and we alone 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. The paper does not constitute a standard, specification, or regulation.

Publication of this paper sponsored by Committee on Freeway Operations.

Abridgment

Vehicle Platoon Parameters: Methodology for Traffic Control

Jose L. Rodriguez, Consultores Tecnicos Asociados, Hato Rey, Puerto Rico
Roy C. Loutzenheiser,* Greater Southwest Regional Planning Commission, Garden City, Kansas

An underwater tunnel for vehicles often becomes a restricted facility creating congestion if demand exceeds capacity in one or more of its sections. These sections become bottlenecks from which slow-moving queues (platoons) emanate, especially during the peak period. Not only do these high concentration areas decrease velocities, but they also reduce the average flow rate.

It has been found on some facilities, however, that bottleneck situations can be alleviated by traffic controls (1). The usual problem has been to decide which control strategy (type of system or plan of operation or both) will achieve optimum operation through a system with one or more restricted points.

In 1971 the Department of Civil Engineering of the University of Maryland initiated a three-phase study of traffic flow in the Baltimore Harbor Tunnel (2, 3). During phases 1 and 2, it was observed that as traffic increased vehicles tended to form platoons regardless of the control alternatives tested. However, the degree, length, and frequency of platoon formation did vary with each alternative. These observations suggested that a methodology utilizing platoon flow characteristics might be developed to evaluate the control alternatives and to determine the best control strategy (2). This paper, undertaken as part of phase 3, evaluates traffic flow in terms of these characteristics.

LITERATURE REVIEW

Studies of the effects of traffic behavior in platoons on traffic flow have not been conclusive except to show that platoon behavior is a major concern in the application of traffic flow theory. One of the first traffic studies involving platoon behavior was conducted on the Pasadena Freeway by Forbes (4) in 1951. Forbes reported that platoon behavior was not adequately described by the behavior of the overall traffic stream. In 1959, the Port of New York Authority (5) conducted a series of experiments to evaluate platoon behavior and measure road capacity in the south tube of the Holland Tunnel. In another analysis of the Holland Tunnel data, Greenberg and Daou (6) observed the tendency of vehicles to have a higher flow rate when they follow a gap in the traffic stream.

In Greenshields' study (7), the minimum spacing distribution is random and extends from about 9 to 61 m (30 to 200 ft). Evidently there are few, if any, spacings below 9 m. Beyond gaps of 61 m there is another random distribution different from that below 61 m. This may be interpreted to mean that under 61 m the distribution varies according to the reaction-perception time of the driver and his judgment of what constitutes a safe distance. Beyond 61 m, spacing may be judged according to the chance placement of the vehicles within the system.

This leads to a simple criterion for platoon definition: Use a minimum spacing of 61 m to separate successive platoons and to separate platoons and noninteracting vehicles (one-vehicle platoons) (1, 2, 3, 8).

STUDY APPROACH

The Baltimore Harbor Tunnel consists of two tunnel tubes, each with two traffic lanes 3 m (10 ft) wide with no shoulders. The Baltimore Harbor Tunnel Thruway regulations specifically prohibit lane changing within the tunnel, and trucks are restricted to the right lane, used also by passenger vehicles. The left lane is nearly 100 percent passenger vehicles. This unbalanced demand caused many shock waves in the left lane.

This research concerns the evaluation of control alternatives with respect to increased flow, and we therefore decided to concentrate on the left lane.

Initial data collection for the research project began in February 1973. Data were collected only on days when the pavement was dry and demand was sufficiently high to cause traffic flow problems. Included in these data were peak-period flows (3:30 to 6:30 p.m.) on Tuesday, Thursday, and Friday (February 13, 15, and 16 respectively) and Tuesday (March 13). This produced a total of 12 h of traffic characteristics for the uncontrolled situation.

Greenshields' criterion was applied to the individual vehicle flow data, which were obtained via a computer program that identified a platoon leader as a vehicle having a space headway between it and the vehicle in front of it greater than 61 m. A vehicle is defined as a single vehicle platoon if its space headway and that of the following vehicle are both greater than 61 m. This single vehicle was not considered an interacting vehicle and was not utilized in this analysis. The following basic values were obtained from each platoon: number of vehicles, average vehicle velocity defined as platoon velocity, and average space headway of the vehicles within the platoon.

After many delays a pretimed metering system was installed in October 1973, and we began data collection in December. Two traffic signal heads with 30-cm (12-in) signal lenses were installed adjacent to the northbound lanes, one for each lane, upstream of the tunnel entrance. Based on subjective analysis (3), cycle lengths of 120, 160, 180, and 240 s were chosen. For example, the 120-s cycle consisted of a 7.2-s red, 3.6-s amber, and 109.2-s green.

We included seven vehicle detection stations in each lane, each station consisting of two photocell detectors slightly over 4.1 m (13.5 ft) apart and capable of sensing many flow characteristics. Time headway, space headway, and velocities are obtainable, and individual vehicles may be identified and traced through the tunnel. Further details of this data collection system, data storage and manipulation, and derivation of the traffic flow characteristics are discussed by Carter and Palaniswamy (2).

DATA ANALYSIS AND RESULTS

A simple method of evaluating the control alternative is to inspect the average platoon velocities (APV) of each control alternative. Figure 1 shows profiles of platoon velocity change from stations 1 to 3 to 5. Station 1 is approximately at the tunnel entrance (downhill); station 3 is approximately at the midpoint of the level section;

and station 5 is at the beginning of the uphill section. It is evident that the no-control alternatives tend to be high in stations 1 [17 m/s (55 ft/s)] and 3 [18 m/s (60 ft/s)] but drop below the 120-s alternative [17 m/s (55 ft/s)] at station 5 [less than 17 m/s (55 ft/s)].

At station 5, the APV of the 240-s alternative [9 m/s (29 ft/s)] is only 60 percent of the next highest alternative, the 160-s alternative [15 m/s (48 ft/s)]. There is only a 1.7-m/s (5.6-ft/s) difference between the remaining 4 alternatives at station 5; only the no-control alternative has an APV lower than that at station 1 (as is also the case for the 240-s alternative).

Figure 1. Platoon velocity profiles.

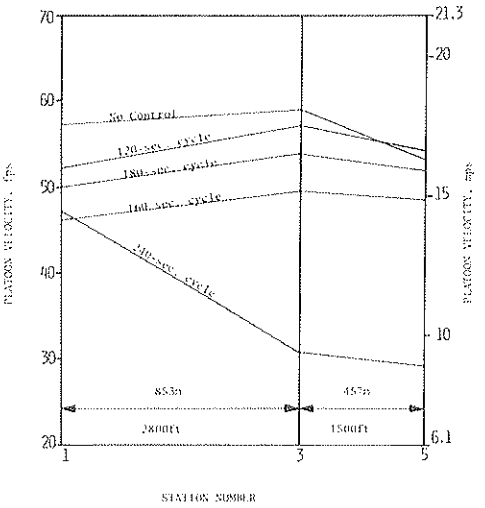
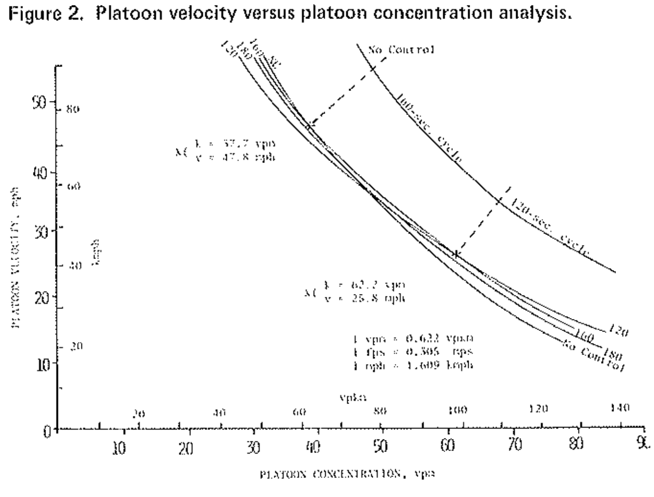


Table 1. Platoon concentration and average number of vehicles.

Velocity Increment (m/s)	Control Alternative							
	No Control		120 s		160 s		180 s	
	PC[a]	AVP[b]	PC	AVP	PC	AVP	PC	AVP
0 to 6	46.83	3[c]	—	—	—	—	—	—
6 to 8	73.03	74[c]	—	—	—	—	78.67	83[c]
8 to 9	71.61	123	72.68	123	72.79	228	70.66	137
9 to 11	63.02	73	65.70	83	66.82	242	65.37	141
11 to 12	55.30	32	60.74	47	58.50	55	59.94	43
12 to 14	50.01	16	57.68	34	61.15	37	50.83	22
14 to 15	43.75	8	52.19	18	54.47	18	53.17	19
15 to 17	48.42	9	48.07	13	50.04	12	48.12	12
17 to 18	44.56	8	46.50	9	44.97	9	44.87	8
18 to 20	46.13	9	41.44	6	41.08	6	42.86	6
20 to 21	40.73	5	38.82	5	37.91	5	38.04	4

Note: 1 m/s = 3.28 ft/s.
[a] Platoon concentration.
[b] Average number of vehicles per platoon.
[c] Less than six platoons in sample.

Figure 2. Platoon velocity versus platoon concentration analysis.



Since the 240-s alternative obviously had the worst APV, no further analysis was made. This poor showing of the 240-s alternative agrees with earlier research at the Baltimore Harbor Tunnel (3). Further evaluation should be primarily concerned with the station 5 bottleneck, because any improvement at this point will benefit flow upstream from the bottleneck.

Concentration is defined as the number of vehicles in a given length of roadway. In the case of a platoon, concentration (PC) may be thought of as the average number of vehicles within the platoon (AVP), separated by an average space headway plus the platoon definition, that is, the total distance from the front of the first vehicle to the rear of the last vehicle plus the platoon criterion, in this case Greenshields' 61 m.

Table 1 shows PC and AVP obtained at station 5 for each control alternative. Values for velocities greater than 21 m/s (70 ft/s) were not obtained because the system would be operating over the maximum posted velocity [22 m/s (73 ft/s)].

The PC values for control alternatives vary slightly at velocities below 11 m/s (35 ft/s). The maximum difference between the values for each control alternative is approximately 2.5 vehicles/km (4 vehicles/mile). It is interesting to note that although the PC values do not show any significant differences at velocities below 11 m/s there are significant differences in the AVP. At the 9 to 11 m/s (25 to 30 ft/s) increment, the AVP value for the 160-s alternative is almost double that of the other alternatives, almost twice that of the 180-s alternative, and almost three times that of the 120-s and no-control alternatives.

Within the velocity range of 11 to 17 m/s (35 to 55 ft/s), except for the no-control alternative, there is not much difference in the AVP among the control alternatives. The AVP for the no-control is almost half of those for the other alternatives. Above 17 m/s (55 ft/s), the AVP for all control alternatives is approximately the same.

The relationship between concentration and velocity is essential to understanding and evaluating how well a system is operating. To investigate the platoon concentration and platoon velocity relationship, a least squares fit of the data was attempted by using Greenberg's exponential model. The results for each control alternative at station 5 are plotted in Figure 2. The optimum platoon velocities are high, 17 m/s (56 ft/s), while the optimum platoon concentrations (k_m) lie within the range of 22 to 28 vehicles/km (35 to 45 vehicles/mile). The jam concentration (k_j), in the case of platoons, may be regarded as the maximum platoon concentration the control alternative can handle without traffic coming to a complete standstill (flow = 0).

From Figure 2, one can see where each control alternative is dominant by finding the highest points of intersection of the curves. It was found that the 120-s, the 160-s, and the no-control alternatives are dominant. The results form the basis for a control strategy utilizing platoon concentration and velocities. An optimum control policy can be utilized for on-line computer control for a facility such as the Baltimore Harbor Tunnel.

FINDINGS

A basic framework for the evaluation of traffic control alternatives has been formulated in this research. The conclusions we drew were that

1. The platoon parameters (APV, PC, and distribution of platoon vehicles) provide a simple and effective

methodology for the evaluation of control alternatives;

2. The parameter APV established that the no-control, 120-s, 160-s, and 180-s alternatives yielded high and almost constant platoon velocities through the tunnel; and

3. The platoon flow values established that in some PC and velocity ranges one alternative predominated.

A control policy utilizing the no-control, the 120-s, and the 160-s alternatives was proposed.

ACKNOWLEDGMENTS

We wish to express our thanks to the project staff at the Department of Civil Engineering, University of Maryland, for their help and to the Maryland State Highway Administration and to the Federal Highway Administration for sponsorship of the project and help during the project. The opinions, findings, and conclusions expressed in this report are our own and not necessarily those of either the Maryland State Highway Administration or the Federal Highway Administration.

REFERENCES

1. P. F. Everall. Urban Freeway Surveillance and Control. Federal Highway Administration, U.S. Department of Transportation, 1972.
2. E. C. Carter and S. Palaniswamy. Study of Traffic Flow on a Restricted Facility: Interim Report—Phase I. Univ. of Maryland, 1973.
3. E. C. Carter, R. C. Loutzenheiser, and S. Palaniswamy. Study of Traffic Flow on a Restricted Facility: Interim Report—Phase II Summary. Univ. of Maryland, July 1974.
4. T. W. Forbes. Speed, Headway, and Volume Relationships on a Freeway. Proc., 22nd Annual Meeting, ITE, Arlington, Va., 1951, pp. 103-126.
5. L. C. Edie and R. S. Foote. Effect of Shock Waves on Tunnel Traffic. Proc., HRB, Vol. 39, 1960, pp. 492-505.
6. H. Greenberg and A. Daou. The Control of Traffic Flow to Increase the Flow. Operations Research, Vol. 8, 1960, pp. 524-532.
7. B. D. Greenshields. A Study in Highway Capacity. Proc., HRB, Vol. 14, 1934, pp. 448-477.
8. J. L. Rodriguez. An Evaluation of Traffic Control Alternatives Utilizing Vehicular Platoon Parameters. Univ. of Maryland, master's thesis, 1974.

Publication of this paper sponsored by Committee on Traffic Control Devices.

**Mr. Loutzenheiser was with the Department of Civil Engineering of the University of Maryland when this research was performed.*

Abridgment

Comparison of Two Types of Left-Turn Channelization

Judy C. Chang, Judson S. Matthias, and Mary R. Anderson,
Department of Civil Engineering, Arizona State University

As traffic volumes increase, turn controls at intersections can eliminate or reduce conflicts, decrease accident hazard, reduce delay, and increase intersection capacity. One of the most useful left-turn controls is a separate lane, called slot, reservoir, pocket, or bay.

In Tempe, Arizona, separate left-turn lanes have been installed at almost every major intersection with an adequate street width. The function of the left-turn bays is to guide vehicles out of the way of through traffic and to prevent rear-end collisions.

A typical left-turn channelization is shown in the Manual on Uniform Traffic Control Devices (MUTCD)(1) with the recommended design of letter markings and arrows applicable to lane-use control. This is the type A marking adopted by the Tempe Traffic Engineering Department (Figure 1).

Another type of marking, type B, not illustrated or recommended by the MUTCD but used in other cities, is shown in Figure 2. Two solid yellow lines in the form of parallel reverse curves are used in the type B marking to identify the path a left-turning vehicle should follow; the type A example shows only an open entry space for access to the left-turn lane.

PURPOSE OF THE STUDY

The purpose of this study was to investigate driver performance at selected sites under types A and B left-turn markings and to determine which type produced better driver performance. After the data for type A markings had been collected, the markings were changed to type B at each location by the Tempe Traffic Engineering Department.

PROCEDURES

Intersections Studied

Of the 12 type A intersections indicated by the Tempe traffic engineer, 4 were selected for this study. Two were chosen for observing the turning movement from an arterial to a collector street (AC I and AC II) and 2 from an arterial to an arterial street (AA I and AA II). Only 1 approach on each intersection was observed in this study. The last 2 intersections have many left turns into areas surrounded by commercial and cultural activities; the first 2 have less volume and lead to residential areas.

Data Collection

Driver performance at each location was classified as proper or improper, depending on whether the vehicle entered the left-turn bay in the proper entry zone or not (Figure 1).

To avoid bias caused by hourly and daily variations in traffic characteristics, data were collected for the same hour and day for each observation of each before-and-after case. Also, to avoid the influence of driver unfamiliarity, we did not gather data for the type B striping for a period of at least 3 weeks after the striping was altered from type A to type B.

A statistical data analysis using a standard test of proportions indicated that there was no significant difference between driver performances; that is, the proportion of drivers who executed a proper entry into the left-turn storage lane on weekdays was at a 0.05 level of significance.

Five different times were selected for data collection at each location: weekday daytime, peak hour, and nighttime; and weekend daytime and nighttime. Data were collected for 1 h at each of the four locations for each of the five time periods.

PRESENTATION AND ANALYSES OF DATA

Table 1 gives the proportion of proper use of each type of left-turn bay marking during the study hour (1.0 would indicate 100 percent compliance). In general, the data show that the proportion of proper use of the type B marking is greater than that of type A.

The results of the field study showed that, with a standard deviation of 0.111, the average proportions of the proper uses were as follows:

Location	Type A	Type B
AC I	0.429	0.531
AC II	0.606	0.658
AA I	0.565	0.728
AA II	0.541	0.791
Overall average	0.543	0.677

A statistical analysis of the data shows that there is no linear relationship between the proportion of proper use and the total approach volume at the study locations.

Driver Performance Versus Time of Day

To analyze the sample data obtained for five different time-of-day combinations at each location, a chi-square test was conducted at a 0.05 level of significance to determine whether that driver performance varied as the time changed. The test results on the four locations show that only for the type A marking at the AA I location could the null hypothesis be rejected in the time-independency test. The results of further analysis of the observed data at this location are shown below.

Type	Day Versus Night	Weekday Versus Weekend
A	Accepted	Rejected
B	Accepted	Accepted

Statistical Analysis

Driver performance in various time groups and at the four study locations was tested at a 0.05 level.

Results for the first test at the four locations are shown in the following table.

Time	AC I	AC II	AA I	AA II
Weekday day	Accepted	Accepted	Rejected	Rejected
Weekday peak hour	Accepted	Accepted	Rejected	Rejected
Weekday night	Accepted	Accepted	Rejected	Rejected
Weekend day	Accepted	Accepted	Accepted	Rejected
Weekend night	Accepted	Accepted	Accepted	Rejected

The hypothesis could not be rejected on the AC I and AC II locations, where the turning movement was from an arterial street to a collector street. Therefore, we concluded that there was no significant difference in driver performance between the before-and-after study periods in such locations. The hypothesis was rejected on AA II, where the turning movement was from an arterial street to another arterial. Therefore, we concluded that there were significant differences in

Figure 1. Typical multilane, two-way marking with single lane, two-way, left-turn channelization (type A).

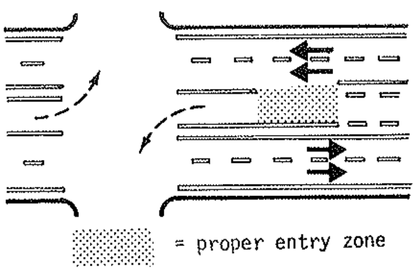


Figure 2. Typical multilane, two-way marking with single lane, two-way, left-turn channelization (type B).

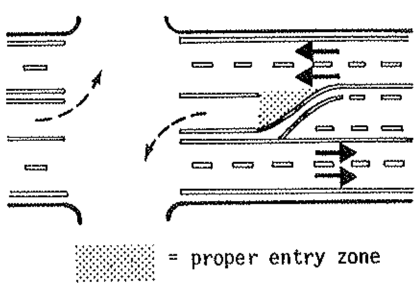


Table 1. Proportion of proper use of lane markings at each location during five time groups.

	Weekday						Weekend			
	Day		Peak Hour		Night		Day		Night	
Location	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B
AC I	0.457	0.561	0.430	0.500	0.343	0.500	0.516	0.536	0.400	0.556
AC II	0.647	0.618	0.571	0.658	0.529	0.632	0.720	0.714	0.563	0.667
AA I	0.606	0.772	0.455	0.765	0.529	0.699	0.645	0.703	0.592	0.702
AA II	0.630	0.826	0.562	0.723	0.400	0.767	0.659	0.837	0.452	0.800

driver performance for type A versus type B markings at this location. At AA I the hypothesis was rejected on weekdays, but could not be rejected on weekends. Hence, at this location, we concluded that there was significant difference in driver performance on the weekdays, but no significant difference on the weekends for type A and B markings.

Next we needed to determine which type of marking produced the better driver performance. According to the function of the street the vehicle turned into, the four locations were put into two corresponding categories (locations AC I and AC II were categorized as the collector streets and locations AA I and AA II as arterial streets). Comparison of daytime and nighttime driver performance for the two categories was examined.

The tabulation below shows that the new marking produces significantly better performance during the night; during the day, it produces significantly better performance only for category II.

Street	Day	Night
Collector	No difference	Type B better
Arterial	Type B better	Type B better

In order to draw an overall conclusion from the selected study locations, a chi-square test was conducted for the number of successes or failures (driver performance as proper or improper). The data consisted of the number of observations falling into either the collector street category or the arterial street category.

We concluded that driver performance with type A markings is independent of the location but that driver performance with type B markings is dependent on the location (i.e., whether the turn is made from an arterial street to another arterial street or to a collector street).

CONCLUSIONS AND RECOMMENDATIONS

After analysis of the data obtained in this study, the following primary conclusions were made:

1. Driver performance did not differ significantly between the two types of markings for locations where turns were made into collector streets;
2. Driver performance differed significantly between the two types of markings on weekdays but not on weekends at AA I;
3. Driver performance differed significantly between the two types of markings at the AA II location;
4. When turning movements were executed into a collector street, driver performance during the day did not differ significantly between the two types of markings;
5. When turning movements were executed into an arterial street, driver performance differed significantly between the two types of markings, and type B marking produced better driver performance;
6. Driver performance differed significantly between the two types of markings at night, and type B marking produced better driver performance; and
7. Driver performance with type A markings were independent of the location chosen.

In general, the results of field observation show that type B marking produces better driver performance and is therefore recommended for use over type A.

Although no accidents occurred during the study, we believe that the type B marking conveys an immediate understanding of the situation that will provide for more uniform traffic flow, will reduce the potential accident rate, and will add to roadway safety in the long run.

REFERENCE

1. Manual on Uniform Traffic Control Devices. Federal Highway Administration, U.S. Department of Transportation, 1971.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Accident Experience With Right Turn on Red

Hugh W. McGee, BioTechnology, Inc., Falls Church, Virginia

The right-turn-on-red traffic signal, once used only in the western states, is now permitted in some form in all but one state and the District of Columbia. However, its adoption was slow primarily because of the concern over its safety aspects. As part of a comprehensive study for the Federal Highway Administration, six separate studies on accidents associated with right turn on red were conducted in Virginia and Colorado and in the cities of Denver, Chicago, Dallas, and Los Angeles. In Virginia and Chicago before-and-after studies were performed; in the other locations records were analyzed to determine both the number of accidents and the causes. From the results of the accident analyses, it appears that the accidents related to right turn on red are very infrequent compared with all intersection accidents (0.4 percent versus 3.3 percent). The Chicago and Virginia studies do not reveal a statistically significant increase in intersection accidents, nor do accidents related to right turn on red appear to be less severe than the average intersection accident; no fatalities were found in the entire accident data base. The general conclusion is that right turn on red does not significantly degrade the safety of signalized intersection traffic operation.

Right-turn-on-red (RTOR) signals, previously used only out west, are now permitted in some form in all but one state and the District of Columbia. Use of this control came slowly, primarily because of the concern over potential for causing accidents.

Presumably no collisions should occur if the motorist makes the RTOR maneuver safely by stopping and yielding to the appropriate vehicles and pedestrians in the intersection. However, not all drivers drive safely all the time, and accidents do happen as a result of a host

Table 1. Virginia intersection accidents.

Accident Category	Before RTOR	After RTOR	Increase or Decrease
Fatal	0	0	0
Injury	96	86	-10
Property damage	396	392	-4
Rear end	135	127	-8
Right turn	19	24	+5
RTOG	19	8	-11
Pedestrian	—	1	+1
RTOR	—	16	+16
RTOR injury	—	1	+1
RTOR pedestrian	—	0	0
Total	492	478	-14

of factors involving the driver, the environment, the road, and the vehicle. Furthermore, vehicles negotiating right turns on green (RTOG) also have collisions. RTOR maneuvers are not, a priori, substantially more hazardous than RTOG maneuvers. A key question is whether permitting RTOR significantly degrades the safety of a signalized intersection.

Several studies on RTOR-related accidents (1, 2, 3, 4, 5, 6, 7, 8, 9) tend to support the claim that there are relatively few RTOR accidents compared with all intersection accidents. But these studies were not sufficiently comprehensive or detailed to identify the magnitude of the RTOR accident problem, types of RTOR accidents, or possible causes. Therefore, additional accident experience data were needed to develop a national policy and implementation guidelines.

This accident analysis consisted of six separate studies conducted in Virginia and Colorado and the cities of Denver, Dallas, Chicago, and Los Angeles. Each of the studies was of a different design dictated by the availability of data. For example, in Virginia and Chicago before-and-after RTOR studies were actually performed; in other locations only records of RTOR and intersection accidents could be analyzed. Procedures and results of each accident analysis will be discussed in detail.

VIRGINIA

Since 1972, Virginia has been following the sign-permissive rule for RTOR, but few signalized intersections are signed for RTOR. In fact, as of September 1975 only 8.6 percent of all possible intersections were so signed.

A before-and-after accident analysis was conducted to determine if these RTOR installations have had any effect on safety. Virginia maintains an inventory of RTOR sign installations, so it was possible to identify specific intersections and the dates of installation. From this master list, 29 intersections were selected for a 1-year-before and 1-year-after installation analysis; no significant changes that would have affected accident frequency were found in either the geometrics or the operating conditions.

For each of the 29 intersections, copies of all accident report forms were reviewed for the 2-year study period (1972 to 1974). The following data were extracted:

1. Total number of accidents within 61 m (200 ft) of the intersection,
2. Total number of right-turn accidents for each approach,
3. Total number of rear-end accidents for each approach,
4. Total number of RTOR accidents (by type) for each approach, and
5. Amount of property damage.

Throughout the study, Virginia required all accidents involving a fatality, an injury, or damage assessed at $100 or more to be reported.

All the information on the form, including the verbal description, the diagram, and other pertinent data, was considered. Also, accident rate statistics were determined by average daily traffic (ADT) volumes obtained from the Virginia Department of Highways and Transportation. However, because the traffic volume data were not complete for all the roads or did not correspond directly with the two study periods, valid comparisons of accident rates before and after institution of RTOR could not be made.

The results of the accident data are shown in Table 1. There was a total reduction of 14 accidents (statistically significant at 95 percent confidence using t-test) for the after case. Assuming that all other conditions did not change, this result would indicate that the RTOR signs actually brought about an improvement in safety. However, the reduction in accidents is more likely attributable to some other circumstances, because no RTOR accidents were identified.

As indicated earlier, the number of rear-end accidents was used for both cases to test whether RTOR would reduce or increase this type of accident. In the before case, there were 135 rear-end accidents (27.4 percent of all accidents), whereas, in the after case, there were 127 (26.5 percent of all accidents). This 0.9 percent difference is not statistically significant (chi-square test), and RTOR had no effect on this type of accident.

Another accident type examined was any collision involving a right-turning vehicle. For the before case, there were only 19 such accidents (3.9 percent of all accidents). However, in the after case there were 24 right-turn accidents (5 percent of all accidents). This increase in accidents is not statistically significant, and therefore it cannot be concluded that this increase in right-turn accidents is attributable to RTOR.

Further analysis of the right-turn accidents showed that there were 16 RTOR-related accidents (3.34 percent of all accidents) in the after case. Of these RTOR accidents, none involved pedestrians, only one resulted in an injury, and none resulted in a fatality. The average property damage for the RTOR accidents was $245 as compared to an average property damage of $554 for all accidents and $537 for the before-and-after cases.

The 16 RTOR accidents occurred at only 8 of the 29 intersections, but 1 intersection had 5 accidents, 2 had 3, and 5 had only 1.

COLORADO

The statute allowing an RTOR unless a sign prohibits it has been in effect in Colorado since July 1, 1969. The Colorado Department of Highways has been compiling RTOR accident statistics since 1970. Table 2 shows RTOR-related accident data provided by the state for 1970 to 1975 and includes the yearly figures for all accidents, for two-vehicle accidents at all intersections (signalized or not), and for injury accidents at intersections. In 1973, Colorado began using a short accident report form for property-damage accidents. As a result, we can no longer distinguish between intersection accidents and RTOR-related accidents (however, the short form is not used in the city and county of Denver). Therefore, the reduction in accidents during 1973 and 1974 for intersection accidents is mostly attributable to the fact that intersection accidents involving property damage are not included. It is also likely that this new form has let a few RTOR accidents go unreported.

For the 4 years 1970 to 1973, all accidents increased, but RTOR accidents decreased. However, this trend reversed itself in 1974, when all accidents decreased and the number of RTOR accidents increased to 90 (plus the unknown number not reported on the short form). For 1975, the number of RTOR accidents decreased to 73.

Although we cannot state conclusively that RTOR accidents have decreased over the years, RTOR accidents are obviously only a very small percentage of all accidents or even all intersection accidents. Statewide, RTOR accidents account for less than 0.1 percent of all accidents and about 0.3 percent of all intersection accidents. Also, RTOR pedestrian accidents account for only 0.4 percent (average of 6 years) of all intersection pedestrian accidents. For the years 1970 to 1972 (1973 to 1975 excluded because of the short form), the percentage of RTOR injury accidents was less than the percentage of all intersection injury accidents.

The Colorado Department of Highways also released a breakdown of RTOR accident types and some data on economic losses. A list of the 426 RTOR accidents reported for the years 1970 to 1974 follows.

Accident	Number	Percentage
Pedestrian	12	2.2
Rear end	69	13.6
Broadside	203	40.7
Sideswipe (same direction)	181	36.2
Sideswipe (other direction)	9	1.8
Overtaking turn	8	1.6
Fixed object	3	0.6
Parked car	1	0.2
Bicycle	9	1.8
Ran off road	3	0.6
Overturned	1	0.7
Total	499	100.0

Broadsides and sidewipes same direction accounted for 76.9 percent of all RTOR accidents. These are the accidents that usually occur with RTOR. Although the exact percentage is not known, many of these accident types involved drivers turning right on a red light and failing to yield the right-of-way to opposing traffic turning left on a green arrow (Figure 1b). In 1974, 28 percent of all RTOR accidents were of this type.

The next most frequent type was the rear end, which was 13.8 percent of RTOR accidents. Also, it should be noted that 12 accidents involved pedestrians and 9 involved bicycles. Together, these account for only 4.2 percent of all RTOR accidents.

The economic loss resulting from these 499 RTOR accidents was estimated at $303 440; average vehicle property damage was $218; and the average estimated economic loss per accident was $608. No fatalities attributed to RTOR were reported during the 6 years.

DENVER

In 1974, the city of Denver initiated its computerized accident and traffic data records system and used the 1974 accident data to conduct a citywide RTOR accident analysis. There are 1137 signalized intersections in Denver. At 78 of these RTOR is prohibited on one or more of the approaches, most of which are located downtown where all-pedestrian signal phasing is employed and all vehicle signals are red when all pedestrian signals display WALK.

Denver accident analyses are displayed in Table 3. During 1974, there were 7431 reported accidents at the 1137 signalized intersections, of which only 50 (or less than 1 percent) involved RTOR vehicles. (An intersection accident is defined as any accident occurring within the prolongation of the curb lines or edge of pavement.) Of the 50 RTOR accidents only 3 resulted in injuries, and none involved pedestrians or resulted in a fatality. The average property damage per accident was $275. Seven of the RTOR accidents occurred at locations where the movement is prohibited, and 6 of these occurred at one intersection that will be discussed later.

The statistics displayed on the lower half of Table 3 shed more light on the RTOR accident problem. For each accident statistic described in the first column, the appropriate percentages were calculated for all 1137 signalized intersections, for all 1059 intersections where RTOR is permitted at all approaches, and for the 78 intersections where RTOR is prohibited. Although statistics are shown for all three sets of intersections, we cannot draw any conclusions based on the comparison across the columns, especially for the last two. This is because characteristics of the 78 intersections where RTOR is prohibited differ from those of the remaining intersections. Also, there is a large disparity in the sample sizes (1059 versus 78), which increases the chance of a wrong conclusion. Some of the more important findings are as follows:

1. The 50 RTOR accidents represent only 0.67 percent of all signalized intersection accidents, whereas RTOG accidents accounted for 8.6 percent;
2. RTOR injury accidents were only 0.19 percent of all injury accidents, whereas RTOG injury accidents comprised 4.0 percent; and
3. RTOG pedestrian accidents accounted for 23.2 percent of all pedestrian intersection accidents, but there were no RTOR pedestrian accidents.

The statistics must also be viewed in connection with the exposure factor or, in this case, the percentage of RTOR vehicles. Based on RTOR usage data collected at eight intersections in Colorado, two of which were in Denver, the average percentage of RTOR vehicles to all right turns is about 20 percent. If RTOR were just as hazardous as RTOG, one would expect RTOR accidents to be 20 percent of all right-turn accidents. This is not the case; RTOR accidents account for only 7.7 percent of all right-turn accidents (50 of 690) and only 4.8 percent of all right-turn injury accidents (3 of 62).

It would appear from these statistics that RTOR does not pose a significant safety problem for Denver. All relevant statistics show that RTOR accidents are a small percentage of all accidents, even of right-turn accidents. Furthermore, it is not as hazardous to pedestrians as the RTOG movement.

In an effort to learn more about what causes RTOR accidents, we reviewed each of the 50 RTOR accident reports and field checked some of the sites. Of the 43 RTOR accidents at intersections where the movement is permitted, 3 occurred at one intersection, 2 occurred at each of five intersections, and the remaining 30 accidents occurred at different intersections. At the intersection where 3 occurred, each accident involved an RTOR vehicle using a different approach: 1 resulted from a conflict with an opposing left-turning vehicle, and the other 2 were collisions with cross-street vehicles (see Figure 1).

Each of the 50 accidents was categorized as to conflict type. The most frequent collision (70 percent of the total) was with vehicles moving across the intersection on green. Another accident type occurred frequently (18 percent) when an RTOR vehicle collided with a vehicle making a left turn from the opposite direction on a left-turn phase. In each situation, there was more than one lane on the intersection exit leg. In some cases the RTOR vehicle made a wide turn into the inside lane, while in others the left-turn vehicle made a wide turn into the curb lane.

The geometrics of the intersections did not show any common characteristic that could be considered a contributing factor. However, as indicated above, the presence of a left-turn signal phase for opposite direction traffic adds another conflict potential for the RTOR vehicle and could contribute to RTOR accidents.

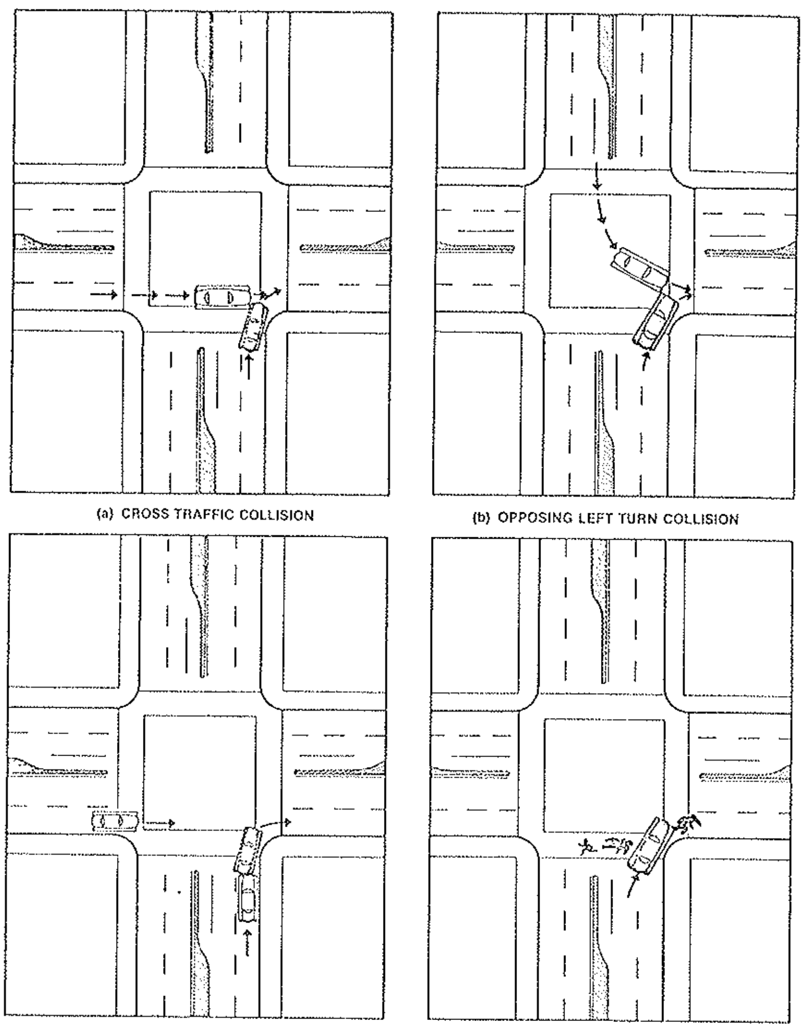
The problem of sign visibility appeared to be a significant factor for at least one location. Of the seven accidents where RTOR was prohibited, six occurred at one location on the same approach. This particular intersection happened to have six legs. The no-RTOR signs were post-mounted on the right curb on the near side and the far side of the intersection; traffic signals were both overhead and post-mounted. For each of the six accident reports, the RTOR motorist indicated that he or she was not aware of RTOR prohibition because he or she did not see the sign. An overhead sign placed by the signal head might very well have eliminated the sign detection problem.

Table 2. Colorado statewide RTOR accident data (1970-1975).

Accident Category	1970	1971	1972	1973	1974	1975
Intersection[a]	29 176	31 510	35 257	25 058[b]	20 652	20 374[b]
Intersection injury	6 835	7 085	7 870	7 038	6 142	6 567
Intersection pedestrian	463	482	455	518	435	458
RTOR	93	93	80	70	90	73
RTOR injury	10	16	11	7	8	10
RTOR fatality	0	0	0	0	0	0
RTOR pedestrian	2	2	1	2	1	3
Total	89 599	95 908	110 541	111 425	104 528	110 773
RTOR/total, %	0.10	0.10	0.07	0.06	0.09	0.07
RTOR/intersection, %	0.32	0.30	0.23	0.28	0.44	0.36
RTOR/intersection injury, %	1.36	1.31	1.02	0.99	1.46	1.11
RTOR injury/intersection injury, %	0.15	0.23	0.14	0.10	0.13	0.15
RTOR pedestrian/intersection pedestrian, %	0.43	0.41	0.22	0.38	0.23	0.66

[a] Both signalized and nonsignalized intersections. [b] Introduction of a short form in 1973 accounts for reduction of these categories.

Figure 1. RTOR accident types.



(a) CROSS TRAFFIC COLLISION

(b) OPPOSING LEFT TURN COLLISION

(c) REAR END COLLISION

(d) PEDESTRIAN CONFLICTS

Table 3. Denver accidents in 1974.

Accident Category	Signalized Intersections	RTOR Intersections	Non-RTOR Intersections[a]
RTOG	640	561	77
RTOR	50	43	7
Injury	1555	1454	101
Pedestrian	125	113	12
RTOG injury	59	53	6
RTOR injury	3	2	1
RTOG pedestrian	29	28	1
RTOR pedestrian	0	0	0
Total	7431	6676	755
RTOR/total, %	0.67	0.69	0.93
RT/total, %	9.30	9.10	11.10
RTOG/total, %	8.60	8.40	10.20
Injury/total, %	20.90	21.80	13.40
RTOG injury/total injury, %	4.00	3.80	6.90
RTOR injury/total, %	0.04	0.03	0.01
RTOR injury/total injury, %	0.19	0.14	1.00
Pedestrian/total, %	1.70	1.70	1.60
RTOG pedestrian/total pedestrian, %	23.20	24.80	8.40

[a]Of the 1137 signalized intersections in the city, 78 have RTOR prohibitions. Of these, 54 are located downtown where an all-pedestrian signal phasing is employed.

DALLAS

The city of Dallas has had the RTOR rule since August 1973 when Texas adopted the law. Because of this change, it was possible to conduct a 1-year before-and-after accident analysis in the course of this study. However, because of the difficulty of extracting the accident data and then defining RTOR accidents, the analysis was limited to pedestrian accidents. At the request of the project staff, Dallas provided a summary listing of pedestrian accidents that occurred at intersections throughout the city.

Included on the summary listing was a column that indicated the appropriate traffic control (signal or stop sign) and a column indicating the direction analysis of the vehicle (right, straight, left, or unknown). These summary listings were provided for 1972 to 1974, allowing us to extract pedestrian accidents at signalized intersections during the 2-year study period. The summary listing did not reveal whether or not the accident involved an RTOR vehicle, but from the police accident reports for the after period this determination was possible.

Dallas has approximately 1000 signalized intersections, of which 86 have no-RTOR controls at one or more approaches. Some, but not all, of these restrictions are because of pedestrian conflicts.

The numbers of pedestrian accidents that occurred at all intersections for 1972, 1973, and 1974 were 50, 55, and 53 respectively, and at all signal-controlled intersections 23, 17, and 17. As shown by the data, intersection pedestrian accidents over the 3-year period have been fairly consistent.

Table 4 shows the comparative results for 1 year, both before and after the RTOR rule became effective. The data revealed a slight increase in pedestrian accidents at all intersections as well as at signalized intersections, but these increases are not statistically significant at the 95 percent confidence level.

Of the 18 pedestrian accidents that occurred at signalized intersections during the after period, only 1 involved a right-turning vehicle (the same as for the before period). The narrative portions of the police accident report forms indicated that only 1 pedestrian accident (that noted as a right-turning accident) could have involved an RTOR vehicle. The data in the tables suggest that RTOR has not caused any increase in pedestrian accidents in Dallas.

CHICAGO

On January 1, 1974, the RTOR law became effective for the state of Illinois. Before then the state had a sign-permissive rule. This change afforded the opportunity to analyze accidents at numerous intersections in Chicago under the two basic RTOR rules as compared to RTOG only.

The accident analysis was built on a study partially completed by the Chicago Traffic Engineering Department. The city selected 97 intersections to determine if the sign RTOR rule causes accidents. These intersections were a geographic sampling distributed throughout the city. Later, two intersections were eliminated because the permissive RTOR signs were removed. These study intersections represent about 4 percent of nearly 1460 signalized intersections.

Chicago collected accident data for a 9-month period (April through December 1972) during which RTOR was not allowed and for the same 9 months of 1973 after RTOR signs had been installed. Because the current generally permissive RTOR rule became valid on January 1, 1974, the city could not obtain a full year of data; therefore, it completed its analysis at that point.

Accident data were collected for the same 95 intersections for a full year under the generally permissive rule. Unfortunately at 17 of the experimental sites no-RTOR signs were installed where RTOR had previously been allowed, thereby reducing to 78 the total number of intersections that could be compared.

In addition to the accident data collection, RTOR usage counts were made at a sample of the intersections in order to determine an average RTOR exposure factor. Later the city did this at all 97 intersections for the analysis and at 10 of the intersections for the expanded analysis under the generally permissive rule.

Table 5 shows the summary accident statistics for the 95 intersections for 9 months in 1972 without RTOR and 9 months with RTOR by sign in 1973. The before-and-after accident data show that there was a statistically significant increase of 110 accidents (13.3 percent). However, for the same two periods, there was a 10.5 percent increase (21 315 to 23 595 or 2280) in total signalized intersection accidents for all 2460 intersections. Therefore, this increase in total accidents cannot be attributed solely to the RTOR feature.

There was a very significant increase, 52 percent, in right-turn accidents. In the before case the 91 right-turn accidents accounted for 11 percent of all accidents; in the after case the 138 right-turn accidents were 14.7 percent. On a citywide basis, there was a 16.1 percent increase (1792 to 2080) in right-turn accidents at all signalized intersections, 27 instances of which involved RTOR. These 27 RTOR accidents represented only 2.9 percent of all the intersection accidents but nearly 20 percent of all right-turn accidents.

Pedestrian accidents increased by 42.5 percent (40 to 57). Furthermore, it was found that of the 11 right-turn vehicle-pedestrian accidents after the RTOR signs were installed 5 involved an RTOR maneuver.

While the RTOR signs were in place, the city conducted a 1-h RTOR count at each intersection. The degree of RTOR usage varied from a low of 3.6 percent of all right turns to a high of 40.6 percent; the average was 14.8 percent. As noted before, the RTOR accidents accounted for 20 percent of all right-turn accidents. Therefore, with RTOR averaging about 15 percent of the total right-turn volumes, the accident statistics for these intersections indicate that RTOR had a slightly higher accident rate than RTOG.

Table 6 shows similar accident data for 78 of the 95 intersections under three RTOR conditions [RTOR prohibited (1972), RTOR allowed with a sign (1973), and RTOR allowed without a sign (1974)]. The sample size was reduced for this three-way analysis because at the remaining 17 intersections no-RTOR signs were installed at various times during the 1974 9-month study period. The lower half of the table shows various accident statistics derived from the data in the upper half.

In comparing the 1972 no RTOR with the 1973 RTOR with sign, we found the same results for the 78 as for the 95 intersections. The percentage of RTOR accidents of all accidents (3.4 percent) is slightly higher for the smaller sample set than the larger set (2.9 percent), because the right-turn pedestrian accidents (including RTOR pedestrian accidents) were all located at the 78-intersection subset.

The accident distribution was quite similar for the last period (1974) with generally permissive RTOR when compared to 1973 with sign-permissive RTOR. There were slight decreases in total accidents and RTOR accidents but neither was statistically significant. The only significant difference was a decrease in total pedestrian accidents in 1974, but the levels of right-turn and RTOR pedestrian accidents were comparable for the two periods.

To determine how the 1974 RTOR accident statistics might have been affected by RTOR usage, 1-h counts were made at 10 of the intersections. Each of these counts was made at the same hour and day of the week as the 1973 count under the RTOR sign. For these 10 intersections, the RTOR usage (RTOR volume of total right turns) varied from a low of 9.2 percent to a high of 35.8 percent; the average was 24.4 percent. For the same 10 intersections, the average RTOR usage was 19.9 percent under the sign rule. Although statistically inconclusive, it appears that there were fewer RTOR-related accidents under the general rule than under the sign rule, even with a higher percentage of RTOR maneuvers. This reduction in RTOR accident frequency was probably due more to driver experience with the maneuver than to the degree of safety under either rule.

Table 4. Dallas pedestrian accidents before and after RTOR.

Accident Category	Year Before RTOR	Year After RTOR	Increase	Percentage Change
All intersections	48	55	7	14.6
Signalized intersections	16	18	2	12.5
Signalized intersections involving right-turning vehicles	1	1	0	0.0

Table 5. Accidents at 96 Chicago intersections with and without RTOR.

Accident Category	No RTOR 1972	RTOR With Sign 1973	Increase or Decrease	Percentage Change
RT	91	138	+47	51.6[a]
RTOG	91	111	+20	22.0[a]
RTOR	—	27	+27	—
Pedestrian	40	57	+17	42.5
RTOG pedestrian	7	6	-1	-14.3
RTOR pedestrian	—	5	+5	—
Total	826	936	+110	13.3[a]

[a]Statistically significant at 95 percent confidence level.

Table 6. Accidents at 78 Chicago intersections.

Accident Category	No RTOR 1972	RTOR With Sign 1973	RTOR Without Sign 1974
RT	65	103	101
RTOG	65	79	80
RTOR	—	24	21
Pedestrian	29	47	24
RTOG pedestrian	5	6	3
RTOR pedestrian	—	5	6
Total	618	709	694
RT/total, %	10.5	14.5	14.6
RTOR/total, %	—	3.4	3.0
RTOR/RT, %	—	23.3	20.8
Pedestrian/total, %	4.7	6.6	3.5
RTOG pedestrian/total pedestrian, %	17.2	12.8	12.5
RTOR pedestrian/total pedestrian, %	—	10.6	25.0

LOS ANGELES

Los Angeles has allowed RTOR at all intersections unless otherwise posted since 1947. RTOR is rarely prohibited in Los Angeles; in fact, out of approximately 3300 signalized intersections where right turns can be made, only 33 (1 percent) have a no-RTOR sign on one or more of the approach legs.

For several years, the city has had a computerized traffic accident information system that has a wide range of accident-analyzing capabilities. Included in the system are not only accident data but also geometric, signal control, and traffic volume data. As a matter of procedure, RTOR accidents are coded into the computerized record system from the accident report. With this information it was possible to conduct a finer analysis of RTOR accidents throughout the city. This analysis was based on accident files for 1973 and 1974. All RTOR accidents were identified by a computer sorting routine in the Los Angeles Traffic Engineering Department. The accuracy of RTOR accident reports would therefore depend on the reliability of police officers making the report and the clerk coding the accident into the computer file.

For 1973 and 1974 a total of 287 accidents were identified as being RTOR related. This figure represents only 0.69 percent of all 41 316 accidents that occurred at 3235 signalized intersections where RTOR can legally be made.

Table 7 shows the comparison of the RTOR accidents with all signalized intersection accidents for various classifications. With respect to severity, it was found that there were no fatal accidents involving an RTOR vehicle. However, nearly 50 percent of the reported RTOR accidents involved an injury, as is the case for all intersection accidents. The percentage of injury accidents is high because as many as two-thirds of all property-damage-only (PDO) accidents go unreported (the California vehicle code specifies that all accidents valued at $100 or more shall be reported, but this requirement is not commonly known or heeded). There were no doubt more RTOR accidents than the 287 identified in this analysis. However, it is likely that RTOR accidents as a percentage of total accidents would be nearly the same as 0.69 percent, since PDO accidents at other intersections also go unreported.

The accidents were categorized into seven types as noted in Table 7. It was found that, of the 287 RTOR accidents, 54 (18.8 percent) involved a pedestrian. When compared to the total of 1487 pedestrian accidents occurring within a 30-m (100-ft) distance of the signalized

Table 7. Los Angeles RTOR accidents versus all signalized intersection accidents (1973-1974).

Classification	Total Intersection Accidents		RTOR Accidents	
	Number	Percentage	Number	Percentage
Accident severity				
Fatal	154	0.4	0	0.0
Property damage	19 796	46.6	144	50.2
Injury	22 474	53.0	143	49.8
Accident type				
Pedestrian	1 487	3.5	54	18.8
Run off road, overturned, parked vehicle, fixed object, other noncollision	5 728	13.5	23	8.0
Right turn all types	407	1.0	94	32.8
Left turn all types	9 969	23.5	4	1.4
Right angle, rear end, sideswipe, head on	21 798	51.3	63	22.0
U-turn or other multivehicle	2 147	5.1	49	17.0
Alley	886	2.1	0	0.0
Lighting conditions				
Day	28 374	66.9	227	79.1
Dusk or dawn	754	1.8	1	0.3
Dark	13 296	31.3	59	20.6
Street type				
Local-local	413	1.0	4	1.4
Collector-local	49	0.1	1	0.3
Collector-collector	116	0.2	1	0.3
Minor arterial-local	433	1.0	2	0.7
Minor arterial-collector	818	1.9	3	1.0
Minor arterial-minor arterial	1 124	2.6	7	2.4
Principal arterial-local	4 876	11.5	25	8.7
Principal arterial-collector	5 626	13.4	35	12.2
Principal arterial-minor arterial	10 708	25.3	77	26.8
Principal arterial-principal arterial	18 224	43.0	132	46.0
Number of legs				
3 T or Y	3 651	8.8	21	7.7
4	32 707	79.2	220	80.3
>4	4 963	12.0	33[a]	12.0
Signal operation				
Two phase	37 338	88.0	243	84.7
Three phase, leading left	2 393	5.6	24	8.4
Three phase, lagging left	724	1.8	3	1.0
Multiphase	1 943	4.6	17	5.9

[a] Does not total 287.

intersections, the RTOR pedestrian accidents account for only 3.6 percent. However, the 18.8 percent is significantly greater than the 3.5 percent for all pedestrian accidents compared to all intersection accidents.

As expected, the remaining distribution of RTOR accident types is dissimilar to that of all intersection accidents, since an RTOR accident is commonly coded as a right-turn accident. In this case, nearly 33 percent of all the RTOR accidents were classified as right-turn accidents. However, the true percentage was probably higher, because of the many RTOR accidents identified in the other categories. It is interesting to note that, of 407 accidents classified as right turn all types, 94 (23 percent) involved an RTOR accident. RTOR volume data were not available for Los Angeles, but in Berkeley Ray found an average of 18 percent of the right-turn volumes were RTOR maneuvers (4). If we assume that this percentage is applicable to Los Angeles, then it would appear that RTOR accidents as a percentage of all right-turn accidents (23 percent), while slightly higher, is similar to the RTOR volume percentage.

The accidents were classified into three lighting condition periods: day, dusk or dawn, and dark (street lights and no street lights). The results show a statistically significantly higher percentage of RTOR accidents during daylight than of all accidents, meaning that RTOR is no more dangerous at night than in day, and, therefore, there would be no reason to prohibit the movement during the night.

Table 7 also gives the accident distribution by street type (local, collector, minor arterial, and principal arterial). RTOR accidents occurred at intersections of different street types in nearly the same proportion as total intersection accidents. The greatest spread in percentages was for the principal arterial-principal arterial (43 percent for all versus 46 percent for RTOR accidents), but this difference is not statistically significant (chi-square, one-way classification). The greatest number of RTOR accidents occurred at the intersections of two principal arterial roads, which is probably a direct function of the traffic volumes.

Another breakdown of accidents was by the number of approach legs to the intersection (three T or Y, four, and more than four). Of course, the vast majority of both RTOR and all accidents occurred at four-legged intersections, but, somewhat surprisingly, the percentage distributions are nearly identical. However, note that 12 percent of the RTOR accidents occurred at intersections with more than four approaches. Many of the no-RTOR signs are installed at these types of approaches, but as a general rule the city does not prohibit RTOR at this type of intersection.

Still another classification was by signal operation mode. The many modes were grouped into four types: two phase, three phase with leading left, three phase with lagging left, and multiphase. The purpose of this classification was to determine if complex signal phasing had any effect on RTOR accident frequency. Once again, the two percentage distributions are comparable and not statistically different. The majority of RTOR accidents, nearly 85 percent, occurred where there was a simple two-phase signal operation.

The final comparative accident distribution analysis

Table 8. RTOR accidents at several locations.

RTOR Rule	Location	Study Year	Signalized Inter-sections	Inter-section Accidents	RTOR Accidents		Pedes-trian Accidents	RTOR Pedestrian Accidents		
					Number	Percentage		Number	Percentage of Pedestrian Accidents	Percentage of RTOR Accidents
Generally permissive	Los Angeles	1973-1974	3235	41 316	287	0.70	1487	54	3.6	18.8
	Denver	1974	1059	7 431	50	0.70	125	0	0	0
	Dallas	1973-1974	1000	—[a]	—[a]	—[a]	18	0	0	0
	Chicago	1974	78	694	21	3.0	24	6	25.0	28.6
	San Francisco	1953-1955	75	3 328	12	0.4	14	4	29.0	33.0
	Portland	—[a]	—[a]	52 677	253	0.5	—[a]	20	—[a]	7.9
	Jacksonville	—[a]	405	1 756	13	0.7	—[a]	2	—[a]	15.4
	Dade County	—[a]	29	700	9	1.3	—[a]	0	0	0
	Omaha	1971-1972	26	497	11	2.2	—[a]	0	0	0
	Salt Lake City	—[a]	24	600	8	1.3	—[a]	0	0	0
Sign permissive	Chicago	1973	95	936	27	2.9	57	5	8.8	18.5
	Columbus	1973	12	415	11	2.7	—[a]	0	0	0
	Virginia	1973	29	478	16	3.3	1	0	0	0

[a] Unknown.

Table 9. RTOR injury accidents versus all intersection injury accidents.

Location	Total	Injury		RTOR	RTOR Injury	
		Number	Percentage		Number	Percentage
Colorado	35 510	7 870	22	80	11	14
Denver	7 431	1 555	21	50	2	4
Los Angeles	42 424	22 474	53	287	143	50
Virginia	478	86	18	16	1	6

was a classification of pedestrian accidents by actions, that is, whether the pedestrian was crossing the intersection legally with the signal or illegally against the signal. Below we show the number and percentage of pedestrian accidents occurring as pedestrians crossed with or against the signal.

Accident Category	Crossing With Signal		Crossing Against Signal	
	Number	Percentage	Number	Percentage
All	896	83	187	17
RTOR	53	95	3	5

The total number of accidents does not equal that noted in the table because some accidents could not be classified in either category. In nearly all the RTOR pedestrian accidents, the pedestrian was crossing the intersection with the signal, presumably in the crosswalk immediately in front of the RTOR vehicle. Only three accidents involved an RTOR motorist hitting a pedestrian crossing against the signal, in this case presumably on the crosswalk. By contrast, for all the pedestrian accidents there were more involving the pedestrian crossing against the signal.

The next analysis was to examine more closely those intersections where RTOR accidents occurred. The 287 identified RTOR accidents for the 2-year period occurred at 267 intersections throughout the Los Angeles area. These intersections represent only 8 percent of the 3235 signalized intersections where RTOR can be made. Only 1 intersection had 3 RTOR accidents during the 2-year period and 18 had 2 accidents, with the remaining 248 accidents occurring at different intersections.

It was not possible to make an on-site investigation of each site, but those intersections where more than one RTOR accident occurred were field checked for any particular feature that may have caused these intersections to experience more than one RTOR accident. Factors considered were number of approach lanes, number of cross-street lanes, average speeds of cross-street traffic, cross-street volume, sight distance, pedestrian activity, and parking on cross-street approach.

If these intersections were particularly hazardous for RTOR movements, then logically they should have had some common geometric or operational feature. However, all the factors were found to vary widely. Approach lanes varied from one to five and cross-street through lanes from one to four. Surprisingly, none of the intersections had cross-street speeds greater than 72 km/h (45 mph), and most were between 40 and 64 km/h (25 and 40 mph). Cross-street volumes expressed in 24-h ADT also ranged considerably from 2000 to 35 000. Although sight distance measurements were purely subjective, only two intersections were noted as having restricted (poor) sight distance.

SUMMARY OF RTOR ACCIDENT EXPERIENCE

The preceding sections discussed RTOR accidents at six different locations. As noted in the introductory remarks, each study was different in scope and methodology. Nonetheless, the results of the separate analyses can be synthesized to provide some general observations regarding the accident problem associated with RTOR.

RTOR Accident Frequency

From the results of our accident analysis and those reported by others, it appears that RTOR accidents are very infrequent. Shown in Table 8 are the overall RTOR accident statistics for 13 different locations using data developed in this study and reported by other researchers. The percentage of RTOR accidents (all types) to all accidents occurring at the specified number of signalized intersections for the generally permissive rule range from a low of 0.4 in San Francisco in a 1953 to 1955 study by Ray to a high of 3.0 for the 78 test intersections in Chicago; the weighted average is 0.61 percent (4).

For the three locations with the sign-permissive rule, there was less variability in the RTOR percentage statistics; the lowest figure was 2.7 percent for Columbus, Ohio, and the highest was 3.3 percent for the 29 locations in Virginia. The weighted average is 2.95 percent, which differs considerably from the 0.61 percent for the generally permissive rule states. This disparity might indicate that the sign-permissive rule is more dangerous than the generally permissive rule. However, this conclusion is not statistically correct, because the sample size was so much smaller for the sign-permissive rule than for the generally permissive. Also, in each of the three sign-permissive rule locations, the RTOR accident statistics represent initial experience with the maneuver. One might expect that RTOR accidents would decrease in frequency as motorists became more familiar with its operation.

Also included in Table 8 are the RTOR pedestrian accidents shown as a percentage of all pedestrian accidents and as a percentage of all RTOR accidents. For the first of the two statistics, there was a wide variability with a low of 0.0 percent found in five different locations with the generally permissive rule to a high of 29 percent found in Ray's study in San Francisco. The weighted average is 3.75 percent, which was strongly influenced by the Los Angeles data. For the second statistic, it was found that where there were some pedestrian RTOR accidents the percentage ranged from 7.9 to 33.0 with a weighted average of 14.7.

For the sign-permissive rule, only Chicago provided any significant data showing that 8.8 percent of all intersection pedestrian accidents are related to RTOR, and 18.5 percent of all RTOR accidents involve pedestrians.

Before-and-After RTOR Comparisons

It was feasible to make only limited comparisons of accident statistics before and after RTOR, and the results obtained were mixed.

In Chicago, total accidents, right-turn-related accidents, and pedestrian accidents all rose at the 95 study intersections after sign-permissive RTOR was instituted, but there were only 27 accidents involving RTOR vehicles (or about 3 percent) out of a total of 936. Moreover, total intersection accidents citywide in Chicago rose that same year by nearly the same percentage as the 95 study intersections where RTOR had been instituted, so it is difficult to attribute much of the increase at the study locations to the change in RTOR rule. Chicago then switched to the generally permissive RTOR rule, and at 78 intersections that previously had sign-permissive RTOR not much change occurred in overall accident frequency. However, pedestrian accidents dropped sharply, more than offsetting the rise in accidents in the previous year under the sign-permissive RTOR rule.

A similar study was conducted at 29 intersections in Virginia, where a change was made from no RTOR to the sign-permissive RTOR. Yearly total accidents decreased slightly from 492 to 478 (statistically significant), but this decrease cannot be attributed to the change in the RTOR rule. Right-turn accidents rose insignificantly from 19 to 24, 16 of which involved RTOR maneuvers. Only one of the RTOR collisions resulted in an injury, and none involved pedestrians.

In Dallas the generally permissive RTOR rule was adopted in August 1973 as part of the statewide change, but it was feasible to study only the effects on pedestrian accidents. Comparing the 1-year periods before and after the RTOR rule change, pedestrian accidents increased slightly (but statistically insignificantly) from 16 to 18. During both the year before and the year after the RTOR rule change only one pedestrian accident involving a right-turning vehicle occurred, and there was no indication of whether RTOR maneuvers were involved.

RTOR Accident Severity

Based on the results of the accident analyses, RTOR-related accidents are less severe than the average intersection accident. For the six locations studied, there were no fatalities as a result of RTOR accidents.

Table 9 shows the percentage of RTOR injury accidents compared to all injury accidents at signalized intersections (10). In each case the percentage of RTOR injury accidents was smaller than that of all injury accidents. The high percentage for Los Angeles (50 percent) is explained by the many PDO accidents that go unreported.

RTOR accidents also tend to have less property damage compared to all intersection accidents. For example, in Virginia the average amount of property damage (as reported) was $538 for all accidents compared to only $229 for RTOR accidents. This latter figure compares favorably with the average property damage of $218 for RTOR accidents reported in Colorado.

RTOR Accident Types

Several distinct types of accidents appear to be associated with RTOR. Illustrated in Figure 1 are four prominent types of RTOR conflicts: rear end, opposing left turn, cross street, and pedestrian.

The most common RTOR accident type occurs when an RTOR vehicle collides with a vehicle moving on green on the cross street (Figure 1a). Sixty-five percent of the reported RTOR accidents in the studies could be categorized under this type, which usually occurs when an RTOR motorist either fails to see the approaching vehicle or perceives a wider gap than is required to make a safe maneuver. These accidents are usually caused by driver error, although limited sight distance can be a contributing factor.

Another frequent type of RTOR accident involves an RTOR vehicle colliding with a vehicle making a left turn from the opposite approach on a left-turn phase (Figure 1b). Of all RTOR accidents, 18 percent were of this type. In this situation, an RTOR motorist looking to his or her left for oncoming cross-street traffic may not be aware of a conflict with left-turning vehicles moving on a separate phase. This conflict can be more serious if there is only one lane to turn into. Where there are multiple departure lanes, the turning vehicles can avoid collision. The accident forms, however, indicated that these types of accidents occurred with multiple lanes as well as with single departure lanes. Prohibiting RTOR at all locations with a left-turn phase should preclude these accident types from occurring. However, because of the randomness and infrequency of these accidents, it would not be practical to prohibit the movement at all such locations.

The third type of RTOR accident is the rear end (Figure 1c). These occur when a vehicle in the process of making an RTOR stops abruptly and is hit in the rear by the following vehicle. This type of accident accounted for 5 percent of all identified RTOR accidents. Because these typically result from driver error, there does not appear to be any way to eliminate them.

The fourth major type of accident is an RTOR vehicle hitting a pedestrian crossing the intersection. As shown in Table 8, the percentage varied widely from 0 to 33 percent. In most cases, the pedestrian was hit while crossing on green in the crosswalk immediately in front of the turning vehicle. However, in a few instances a pedestrian was hit while crossing on a red signal (which is illegal in most states) in the crosswalk on the lane the vehicle is turning into.

In addition to these four major types of RTOR accidents, there were two others that were very infrequent.

1. Two RTOR vehicles sideswipe. Sometimes two vehicles collided while making an RTOR simultaneously when there were double right-turn lanes, or when one of the vehicles used the shoulder.
2. The RTOR vehicle induces an accident. In two cases the RTOR vehicle, although not involved in the accident, created a situation that resulted in an accident. Once the cross-street vehicle collided with another cross-street vehicle to avoid hitting the RTOR vehicle. In another case (which happened in Ohio and was brought to our attention), the RTOR vehicle apparently induced a following vehicle to cross the intersection on red resulting in an accident with a cross-street vehicle coming from the opposite direction. This latter accident resulted in a fatality, the only one during the whole course of study.

ACKNOWLEDGMENTS

This paper is based on a study for the Federal Highway Administration, U.S. Department of Transportation, performed while I was associated with Alan M. Voorhees and Associates of McLean, Virginia. The opinions, findings, and conclusions are my own and not necessarily those of FHWA.

REFERENCES

1. W. Marconi. Right Turn on Red—One View. Traffic Engineering, May 1973, p. 7.
2. R. L. May. RTOR: Warrants and Benefits. Purdue Univ., Joint Highway Research Project 74-14, Aug. 1974.
3. M. B. Meyer. Right Turn on Red Signal Study. Cedar Rapids Traffic Engineering Department, Iowa, Aug. 1970.
4. J. C. Ray. The Effect of Right-Turn-on-Red on Traffic Performance and Accidents at Signalized Intersections. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, student research paper, May 1956.
5. Review—Right Turn on Red. Wichita Traffic Engineering Division, Kansas, Aug. 1971.
6. Right Turn on Red. Committee 3M(65), ITE, May 1968.
7. Right Turn on Red Accident Study. Office of Traffic Engineering, Minnesota Highway Department, Final Rept., 1971.
8. M. R. Parker, R. F. Jordan, Jr., J. A. Spencer, M. D. Beale, and L. M. Goodall. Right Turn on Red. Virginia Highway and Transportation Research Council, Report to the Governor and General Assembly of Virginia, Sept. 1975.
9. Some Right-Turns-on-Red Facts. Oklahoma City Department of Traffic Control, Nov. 1971.
10. H. W. McGee. Right-Turn-on-Red: Current Practices and State-of-the-Art. Federal Highway Administration, Rept. FHWA-RD-75-5, Oct. 1974.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Abridgment

Driver Behavior During the Yellow Interval

William L. Williams, Office of Development, Federal Highway Administration

Every driver has experienced the anxiety of approaching an intersection as the signal turns yellow. The driver must then decide quickly whether to stop or to go through before the signal turns red. The change period is one of the most important and least studied intervals of the signal cycle.

This investigation has a threefold purpose: to provide an understanding of driver characteristics during the yellow interval, to determine the ability of drivers to stop in time, and to present a method for determining the length of the clearance interval for urban intersections. The data collected in this study of one intersection help answer the following questions: What do drivers do when the signal changes to yellow? How fast a deceleration rate will drivers accept when the signal changes to yellow? How long should the clearance period be to satisfy the drivers' needs?

RESULTS

Probability of Stopping as a Function of Distance

At the intersection studied, 816 close-decision vehicles were recorded. The probability of stopping was plotted against the cube root of the distance from the stop line at the instant the signal turned yellow, and the results are shown in Figure 1.

Probability of Stopping as a Function of Approach Velocity

The total distribution was stratified in order to obtain distributions of the probability of stopping for vehicle speeds of 16.1, 24.1, 32.2, and 40.2 km/h (10, 15, 20, and 25 mph). From these distributions the probability of stopping given distance from the stop line was deduced.

The resulting distributions are shown in Figure 2, which gives a series of distance values that corresponds to a certain probability level for the speeds. The log of the velocities was plotted against the corresponding distances by using a constant probability of stopping. The results are shown in Figure 3, which reveals a linear relationship between distance from the stop line and the log of approach velocity. The theoretical relationships denoted by the straight lines in Figure 3 were replotted on rectangular coordinate graph paper. The result, shown in Figure 4, reveals the curvilinear relationship between distance and approach velocity when plotted as a function of the probability of stopping. The curves give results when 15, 50, 80, and 95 percent of stops are successful.

Probability of Stopping as a Function of Potential Time

Potential time is that required for a vehicle to reach the stop line after the signal turns yellow, assuming a constant velocity. Figure 5 shows the probability of stopping as a function of potential time.

Deceleration Rates

In order to obtain the maximum deceleration rates acceptable to drivers having a choice of either proceeding through or stopping, the sample of stopping vehicles was stratified to represent only the vehicles that stopped most quickly. Several stopping vehicles were observed during each cycle, but the sample was stratified to include only the vehicle that stopped first. At the intersection studied, 166 first-to-stop, close-decision vehicles were recorded. The resulting distribution is presented in Figure 6.

Ability to Stop

Figure 7, a curve representing the probability of stopping as a function of deceleration, shows that half the drivers chose not to accept a deceleration rate faster than 2.8 m/s^2 (9.1 ft/s^2), a fifth did not accept a rate of 2.1 m/s^2 (7.0 ft/s^2), but a tenth did accept a rate over 4.9 m/s^2 (16.0 ft/s^2).

Method of Determining Length of Yellow Interval

There are two basic conditions that must be considered when the yellow interval is timed: when a driver will choose to stop and when a driver will choose to go through an intersection. It can be shown that the required length of the yellow interval can be computed by evaluating the terms of the following general equation (derivation not included).

$$Y = R + (V/2a^-) + [(W + L)/V] - [K + (2d/a^+)^{1/2}] \quad (1)$$

where

- R = driver decision and reaction time (1.1/s) (1);
- V = 85 percentile approach speed;
- a^- = deceleration accepted 85 percent of the time [2.0 m/s^2 (6.5 ft/s^2) from Figure 7];
- W = distance from stop line to the line where the vehicle is shadowed;
- L = length of vehicle [5 m (17 ft) for automobiles];
- K = reaction time of cross-flow traffic (0.4 s) (2, p. 23);
- d = distance between vehicles and cross-flow traffic; and
- a^+ = maximum acceleration of cross-flow traffic [4.9 m/s^2 (16.0 ft/s^2)].

Or, Y equals decision time plus deceleration time plus clearance time minus cross-flow acceleration time. The parameters of this equation for the intersection studies are shown in Figure 8.

INTERPRETATIONS

Probability of Stopping as a Function of Distance

An approximate mathematical model for the probability of stopping (P_s) that adequately describes driver behavior during the yellow interval is as follows:

$$P_s = (1720 \times 10^{-4} D^2) - (257.4 \times 10^{-5} D^3) \quad (2)$$

where D is how far back from the stop line a vehicle is at the instant the signal turned yellow. P_s is a percentage. This equation can be used in conjunction with computer simulation.

Probability of Stopping as a Function of Velocity

Figure 3 shows that for a given probability the relationship between approach velocity and distance from the stop line is logarithmic. The equation for the straight lines in Figure 3 relating velocity and distance for a constant probability of stopping was found to be

$$D = m \log V - b \quad (3)$$

where D is distance from stop line, V is velocity, m is constant, and b is constant.

In Figure 4 the equation for the curves that relate velocity and distance was found to be

$$V = (10^{D/m})(10^{b/m}) \quad (4)$$

Probability of Stopping as a Function of Potential Time

Figure 5 shows that, when the signal turns yellow, out of 100 vehicles having a potential time of 1.9 s, 15 stopped. Vehicles having a potential time of 2.9 s have a fifty-fifty chance of stopping or going through. At a potential time of 5.5 s, 5 out of 100 went through, and 95 stopped. Figure 5 also shows that 37 percent of the drivers entered the intersection after the yellow interval ended (potential time of 3.2 s). Fifteen percent entered 1 s after the red interval began (potential time of 4.2 s), and 5 percent entered 2.2 s after the red. The observation that 37 percent of the vehicles entered and crossed the intersection after the beginning of the red interval makes it clear that the existing yellow interval at the intersection of 3.2 s is inadequate.

CONCLUSIONS

Driver Characteristics

At the intersection studied, 37 percent of the vehicles entered and crossed the intersection after the 3.2-s yellow interval. Eighty-five percent of the close-decision drivers chose both to stop at the intersection when they were farther than 30.5 m (100 ft) from the stop line at the instant the signal turned yellow and to go through the intersection when they were within 13.1 m (43 ft) of the stop line. At a distance of approximately 20.4 m (67 ft) from the stop line, 50 percent of the close-decision

Figure 1. Probability of stopping during yellow interval.

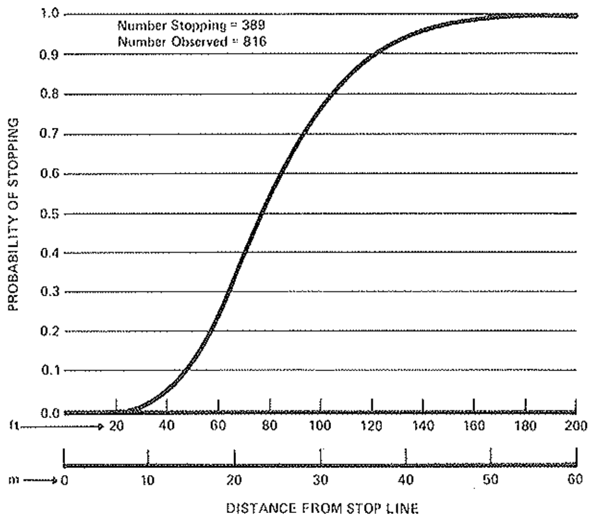


Figure 2. Probability of stopping as a function of velocity and distance from stop line.

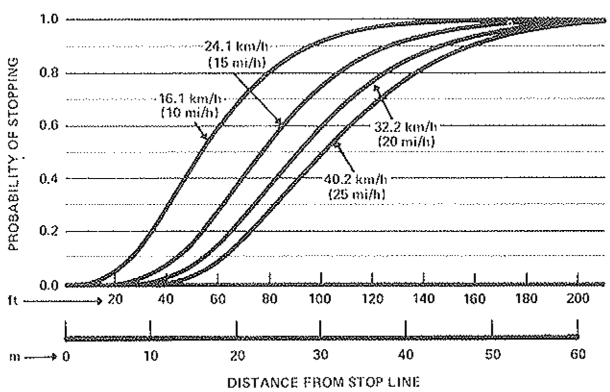
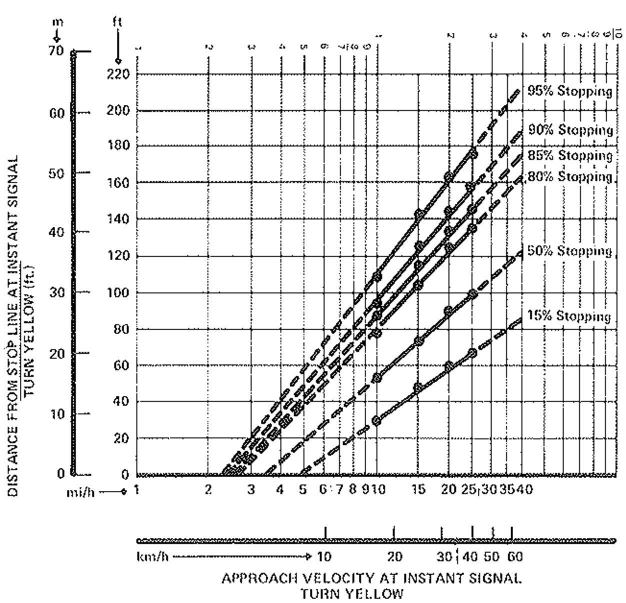


Figure 3. Log of probability of stopping as a function of distance and velocity.



drivers stopped, and 50 percent went through.

Ability to Decelerate

The average maximum deceleration rate accepted by

Figure 4. Distance of vehicle from stop line and approach velocity as a function of percentage stopping.

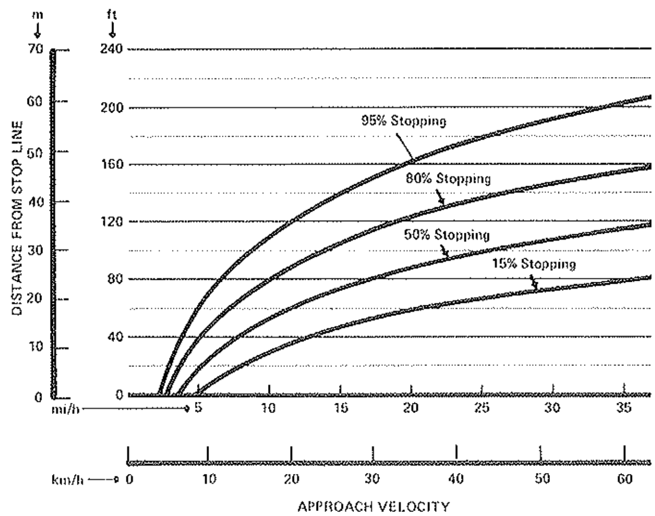


Figure 5. Probability of stopping during yellow interval versus potential time to stop line.

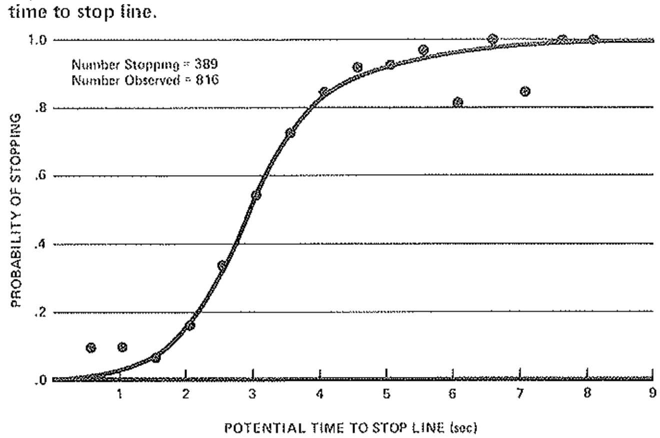


Figure 6. Cumulative frequency distribution of average deceleration rates for 166 stopping vehicles.

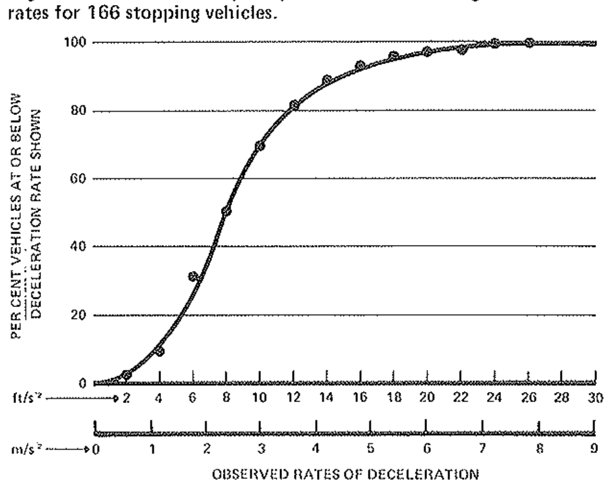


Figure 7. Probability of stopping during yellow interval versus accepted deceleration rate.

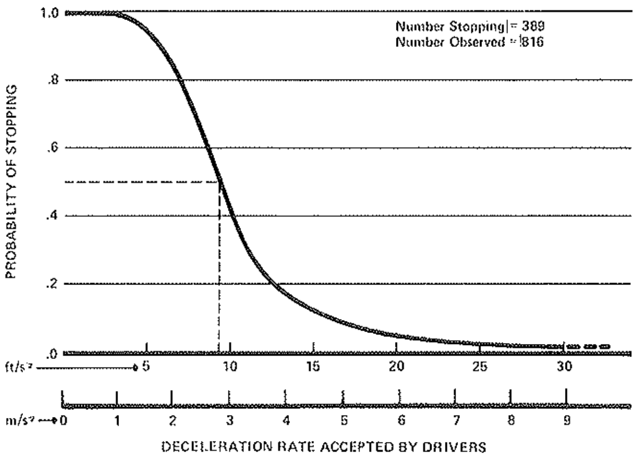
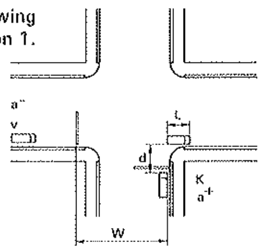


Figure 8. Intersection showing parameters used in Equation 1.



the vehicles stopping most quickly is 2.0 m/s^2 (9.7 ft/s^2). Drivers confronted with a close decision during the yellow interval will accept a deceleration rate of 2.0 m/s^2 (6.5 ft/s^2) 85 percent of the time.

Method of Determining the Length of the Clearance Interval

The minimum length of the clearance interval can be calculated to adequately serve drivers' needs and to meet law enforcement purposes by using Equation 1.

It should be noted that the time calculated from this equation is that needed for clearance. This can be provided by the yellow interval in combination with an all-red interval. Using this technique, a city or county can standardize the length of the yellow interval (say, 3.6 s for 60-s cycle phase) and provide additional clearance time with the all-red interval. If this equation is to be correctly applied, engineers should conduct field studies in their own locations to determine local values for the unknown parameters.

The reader should note that the terms of Equation 1 are not new and that various permutations of them have been recorded in the literature since 1929 (3). The value of Equation 1 is that it is theoretically correct and includes all parameters involved in the clearance decision. Engineers should develop probability of stopping versus time charts similar to Figures 1-7 for their own cities. In this way the decision time $[R + (V/2a^-)]$ can be computed by how drivers actually behave in the area being studied. The time deduction for cross-flow acceleration needs to be applied with caution, and a value of zero should be used if light jumping is possible (yellow interval visible to waiting traffic).

REFERENCES

1. P. L. Olson and R. Rothery. Drivers' Response to the Amber Phase of Traffic Signals. General Motors Research Laboratories, Warren, Mich., 1960.
2. W. S. Smith and F. W. Hurd. Traffic Engineering. McGraw-Hill, New York, 1955.
3. T. M. Matson. The Principles of Traffic Signal Timing. Trans., 18th Annual Safety Congress, National Safety Council, Vol. 3, 1929, pp. 109-139.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Optimization of Pretimed Signalized Diamond Interchanges

Carroll J. Messer, Daniel B. Fambro, and Stephen H. Richards,
Texas Transportation Institute, Texas A&M University

This paper describes a computer program that can determine the best strategy for a pretimed signalized diamond interchange to minimize the average delay per vehicle. This program, PASSER III, is one of a series of signalization programs developed for the Texas State Department of Highways and Public Transportation. All basic interchange signal phasing sequences, including all possible patterns from lead-lead, lag-lead, lead-lag, and lag-lag phasings, are evaluated by the program. Interchange performance is evaluated by using average vehicle delay; exterior delay is calculated by Webster's delay equation; and interior delay is determined from deterministic delay-offset techniques. Minimum delay analyses of 18 sample problems were made. Many signalization phasing patterns were found to provide optimum operation over the set of problems evaluated. While four-phase overlap and three-phase timing plans were normally found to provide good operation, other signalization patterns may produce even better operation.

The signalized diamond interchange is a critical facility for providing high performance levels along urban freeway corridors. Efficient movement of traffic through the interchange and the quality of service provided motorists depend to a large measure on the type of signalization used. However, there seem to be differences of opinion regarding the best way to signalize a diamond interchange. This is probably because no efficient methodology for analyzing the problem has been proposed, although numerous researchers have made significant attempts at providing guidelines for improving diamond interchange signalization.

This paper describes a computer program that can determine the best strategy for a pretimed signalized diamond interchange to minimize the average delay per vehicle. This program, named PASSER III, is one of a series of signalization programs (1, 2) developed for the Texas State Department of Highways and Public Transportation.

Munjal (3) presented a systematic discussion of diamond interchange signalization, and he and Fitzgerald (4) reported on diamond interchange simulation programs. Much discussion in the literature has addressed the relative merits of four-phase overlap signalization compared with other types of phasing patterns. One paper (5) has contributed to this discussion and, perhaps, to the confusion. We hope that our paper will provide a tool that can be used by traffic engineers to accurately analyze their interchange problems and that can eliminate the need to rely on debatable guidelines.

The basic problem is how to determine, for a given set of traffic demands, the best pattern for pretimed signals at a diamond interchange. If the traffic volumes are those presented in Figure 1, Poisson arrivals are assumed for exterior traffic flow. All arterial through movements and frontage roads have two-lane approaches, whereas the two interior left-turn volumes are serviced by one-lane left-turn bays of adequate storage capacity. All geometrics and volume assumptions are arbitrary.

SIGNAL PHASING

Let us now look at the left intersection of a diamond interchange shown in Figure 2 and see how many different signal phases with no conflict among movements this intersection can have. Phase A is when the off-ramp and the left-turn traffic from the arterial are stopped and the straight through traffic is moving. Phase B results when the traffic from the off-ramp is given a green signal; all other movements at this intersection must be stopped at this time. Phase C occurs when the outbound arterial left-turn traffic is given a green signal, and all the incoming conflicting traffic feeding the diamond at this intersection stops. There are no additional phases at this intersection. There are only three similar phases on the right intersection of the interchange; these form the basis for all the possible phasing patterns. Any phases for pedestrians, as well as the amber phases for motorists, have been excluded from these and from all phasing patterns discussed in this paper.

Munjal (3) has shown that the phase order of left and right intersections can be ABC or ACB independently of one another. Order ABC was called leading left turns and order ACB lagging left turns. Thus there are only four possible basic interchange phasing codes (sequences) that can be generated (Figure 3). Munjal's equivalent descriptions are as follows:

Phase Code	Left Phase Order	Right Phase Order	Munjal's Description
1	ABC	ABC	Lead-lead
2	ACB	ABC	Lag-lead
3	ABC	ACB	Lead-lag
4	ACB	ACB	Lag-lag

All of the possible signal phasing patterns that an engineer might devise can be developed by using these basic phase codes and then varying the offset between the two intersections from zero to one cycle length. In this paper, the offset is defined as the time difference in seconds between the start of left basic phase A and the end of right basic phase B.

An example of how an interchange signal phasing pattern results from a given interchange phasing code and offset is presented in Figure 4. Phase code 1 and an arbitrary offset have been selected.

SIGNAL TIMING

Signal green times are usually calculated independently of the interchange phase code selected by PASSER III as if the two intersections were also independent of one another. It is possible to override this basic operating procedure and force selected movements to have equal green times, although this reduces the efficiency of the interchange. One other exception is permitted, the four-phase with overlap signal calculation requirements (5).

The green times of phases A, B, and C of Figure 2 are calculated, in the independent mode of operation, using Webster's formula (6):

$$G = (y/Y) \cdot (C - L') + L \tag{1}$$

where

- G = phase green on approach (s);
- y = q/s;
- q = approach volume (vehicles/s);
- s = approach saturation flow (vehicles/s green);
- Y = sum of all y at intersection;
- C = cycle length (s);
- L = intersection lost time; and
- L' = sum of intersection phase lost times (s).

Messer and Berry (5) have shown that a formula similar to Equation 1 should be used to calculate green times for four-phase overlap signalization (a special case of interchange phase code 1). In this case, green times on the four external approaches to the interchange are calculated from

$$G = (y/Y) \cdot (C + \phi - L') + L \tag{2}$$

where G is the green phase on exterior approaches in seconds and ϕ is the sum of interchange overlap (travel times) in seconds.

EXTERIOR DELAY

The performance of a diamond interchange is evaluated primarily on the basis of average delay for all vehicles using the interchange. At the beginning of the analysis procedure, delays on the four exterior approaches to the interchange are first calculated by Webster's delay equation (6)

$$d = [C(1-\lambda)^2]/[2(1-\lambda x)] + \{x^2/[2q(1-x)]\} - 0.65(C/q^2)^{1/3}\, x^{(2+5\lambda)} \tag{3}$$

where

- d = average vehicle delay for exterior approach movement (s/vehicle);
- q = approach movement flow rate (vehicles/s);
- λ = proportion of cycle green for approach movement; and
- x = signal saturation ratio, qC/gs; $g = G - L$ (s).

A total of 14 separate exterior movements are analyzed for delay, 1 for each identifiable turning movement from the exterior approaches. The two arterial approaches have 3 movements: right turn, through on arterial, and through then left turn within the interchange. The two ramps (frontage roads) have 4: right turn, through, left turn then through on the arterial, and left turn then left turn within the interchange (a ramp U-turn).

Figure 1. Signalization and approach volumes at a diamond interchange.

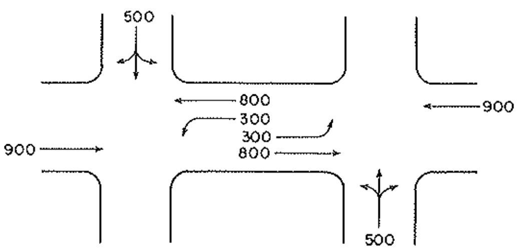


Figure 2. Three basic phases at left intersection of interchange.

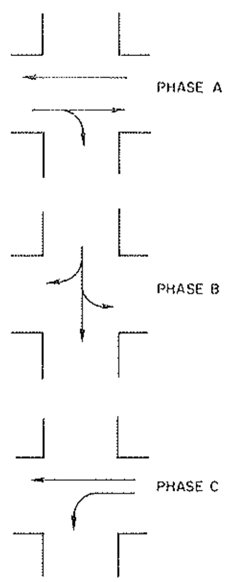
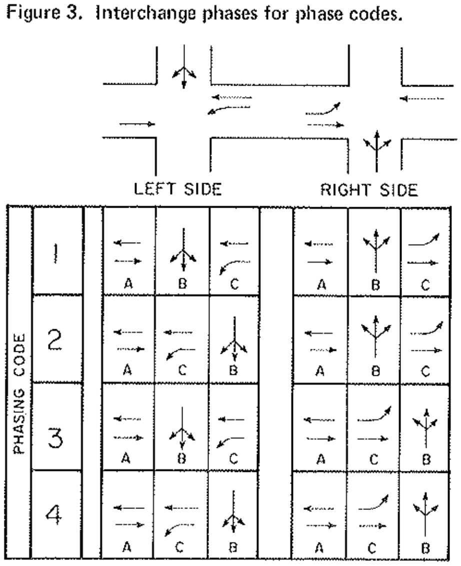


Figure 3. Interchange phases for phase codes.



INTERIOR DELAY

Vehicle delays that occur within the interchange are calculated by a version of the deterministic delay-offset technique (7). Several excellent papers have described applications of this technique to signalized intersections (8, 9, 10). Documentation and validation of PASSER III is likewise available (2).

Figure 5 (9, p. 16) shows how Gartner defined a traffic link as a section of street carrying a traffic flow movement in one direction between two signalized intersections. Delay is incurred at the downstream signal of the link where traffic exits from the link. The offset across any link may be defined as the time difference between the starting point of green phase A at the upstream signal of the link and the starting point of the next green phase at the downstream signal. A link is a directional quantity, assuming the direction of traffic flow along it. Gartner described the flow of traffic through the link's exit signal and developed the computational procedure for obtaining a delay-offset relationship (9, pp. 13-15).

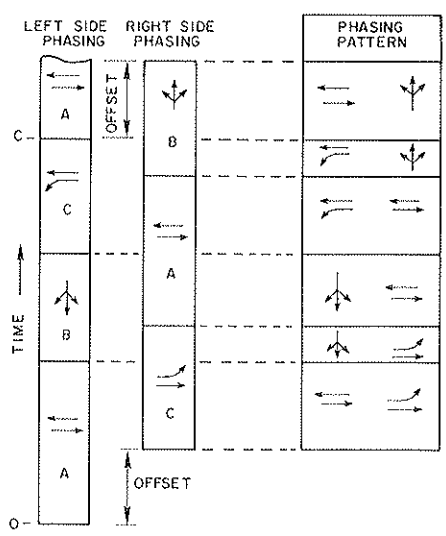
INTERCHANGE DELAY

Average delay per vehicle using the diamond interchange is calculated by combining the effects of exterior and interior interchange delays. For an otherwise given set of geometric, volume, and signalization inputs, interchange delay changes only as the offset between the two intersections is varied, as illustrated in Figure 6.

Figure 6 was developed from the volume data in Figure 1 with an assumed U-turn volume of 150 vehicles/h on both ramps, a 70-s cycle length and a 14-s travel time between the two intersections. Interchange delays were calculated by interchange phase code 1 (ABC:ABC). Delay is observed to drop to a minimum delay value at a 14-s offset and then to begin to rise beyond this minimum delay offset. Also shown in Figure 6 is the component of interchange delay occurring within the interchange. External delay remained constant.

Figure 7 shows the variation in maximum queue lengths that would occur on the interior left-turn and through lanes for the left-to-right (eastbound) arterial as a function of offset. Queue storage capacities, although unlimited in all our analyses, are important input constraints on the PASSER III program.

Figure 4. Development of diamond interchange phasing pattern from phasing ABC:ABC and offset.



INTERCHANGE PHASING ANALYSES

In addition to the four basic interchange phasing codes, a fifth one was studied. This code, 1A, represents a special case of the normal code 1 (lead-lead) interchange phase. The phase sequences are the same for code 1 (ABC:ABC), but the four external green times are calculated (see Equation 2) to total

$$C + 2 \cdot \text{travel time} \tag{4}$$

The popular four-phase with overlap signal phasing results if the offset between the two intersection signals is selected to be the same as the travel (overlap) time (3).

Figure 5. Traffic links connecting pair of adjacent signalized intersections.

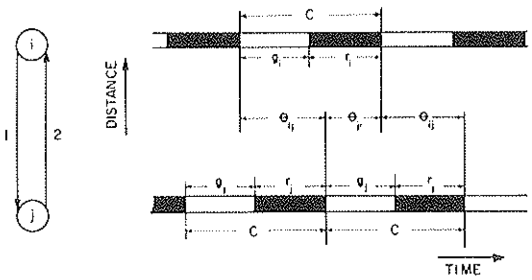


Figure 6. Variation in interchange delay with offset for phase code 1.

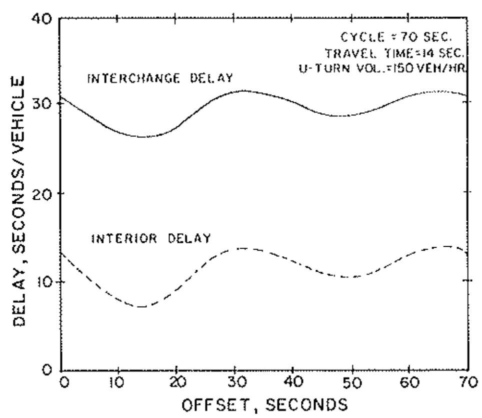
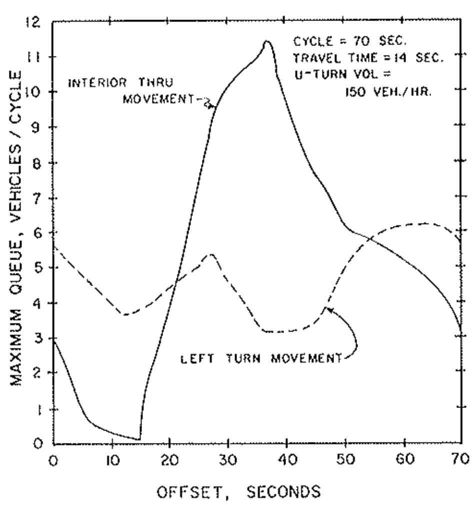


Figure 7. Variation in maximum queue length with offset for phase code 1.



The performance of four-phase with overlap can be determined in Figure 8 from the delay curve of interchange phase code 1A at an offset of 14 s, which, as might be expected, results in the minimum delay for this set of conditions. Other offsets increase the average interchange delay. It should be noted that the normal unimpeded travel time between the two intersections is assumed to remain constant at 14 s, regardless of the offset selected. In the real world, motorists may adjust their travel time slightly depending on the offset. If a queue forms on a movement in the interior of the interchange, a queue start-up delay or signal lost time is also assumed to occur (11).

A comparison of the performances of two different types of interchange phasing arrangements, 1A (ABC: ABC) and 4 (ACB:ACB), can be made from Figure 8. Minimum delay for code 4 occurs at 0- and 70-s offsets, which are the same because the cycle length is also 70 s. A 0-s offset for 4 results in a three-phase, lag-lag interchange signal phasing pattern. In his subjective review of diamond interchange signal phasing arrangements Munjal (3) concluded that there are two preferred sets of phasing patterns: four-phase with overlap and three-phase, lag-lag patterns. For this sample problem, the PASSER III outputs indicate that these two patterns would operate well. More important, however, is the fact that the phasing patterns that give minimum delay can be determined.

Although interchange phase codes 1 or 1A and 4 may be able to generate good operating conditions if the proper offset for each is selected, there are other basic interchange phasing arrangements that might provide even better results. Until all of these phase codes have been considered, the best interchange phasing pattern cannot be selected.

An example of the performance of all five interchange phasing codes is presented in Figure 9. For this problem, codes 1, 1A, and 4 will provide relatively good operation at their respective minimum delay offsets. Codes 2 and 3 do not perform as well as the others, and their performance curves, in the middle range of delay values, are not as responsive to differences in offset. A total of 350 different interchange timing plans were analyzed to generate the results presented in Figure 9. A manual analysis would not be practical, and a detailed microscopic simulation may not be economical.

These delay results tend to support the previously discussed general guidelines that the four-phase with overlap (lead-lead) and the lag-lag are the generally preferred signalization strategies. This general guide-

Figure 8. Performance of phase codes 1A and 4.

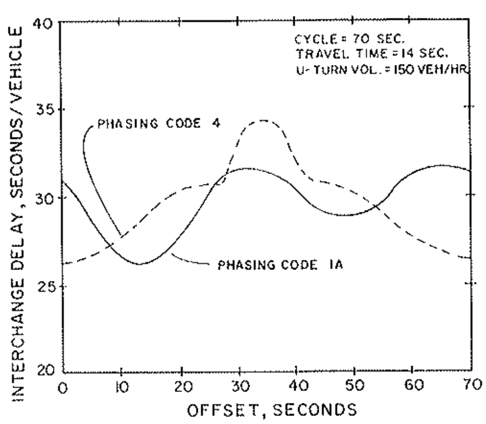


Figure 9. Variation in interchange delay with offset for five phase codes.

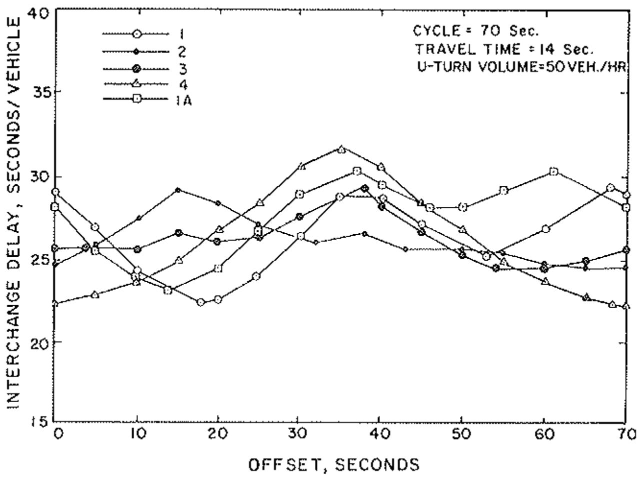


Table 1. Minimum delay for interchange and phase codes 1 and 1A for 50 vehicles/h U-turn volume.

Travel Time (s)	Cycle Length (s)	Minimum Interchange Delay (s/vehicle) Optimum Phasing	Code 1	Code 1A
6	60	20.85	21.89	23.95
6	70	23.17	25.33	27.52
6	80	25.68	27.85	31.03
10	60	19.92	20.30	21.28
10	70	22.92	23.61	24.81
10	80	25.95	26.93	28.31
14	60	19.35	19.35	20.00
14	70	22.05	22.22	23.07
14	80	24.80	25.42	26.31

Table 2. Minimum delay for interchange and phase codes 1 and 1A for 150 vehicles/h U-turn volume.

Travel Time (s)	Cycle Length (s)	Minimum Interchange Delay (s/vehicle) Optimum Phasing	Code 1	Code 1A
6	60	22.75	25.32	26.65
6	70	25.02	28.87	29.97
6	80	27.71	31.81	33.75
10	60	23.93	23.54	24.10
10	70	26.99	27.24	28.04
10	80	28.72	30.85	31.94
14	60	22.57	23.40	22.82
14	70	26.28	26.28	26.28
14	80	29.35	29.35	29.83

Table 3. Minimum delay phase codes for 18 interchange signalization problems.

Travel Time (s)	Cycle Length (s)	Phase Codes by U-Turn Volume 50 Vehicles/h	150 Vehicles/h
6	60	2, 3	2, 3
6	70	2, 3	2, 3
6	80	2, 3	2, 3
10	60	4	1
10	70	4	2, 3
10	80	4	2, 3
14	60	1	4
14	70	4	1, 1A
14	80	4	1

line may be useful, but it does not indicate which of the two is the better. As the following study results will show, the other phasing codes may operate better under a different set of conditions.

MINIMUM DELAY STUDIES

A number of geometric, signalization, and traffic flow studies will be presented to demonstrate PASSER III program features and to illustrate the need for a thorough investigation of available performance of design and signalization options. Delay performance curves similar to those in Figure 9 were developed for 18 basic signalization problems.

Throughout the studies, we used the interchange external volumes in Figure 1 and held them constant. These volumes result in exterior volume-capacity ratios of about 0.8. The turning movement variations within the interchange allowed ramp (frontage road) U-turn volumes of 50 and 150 vehicles/h. A U-turn volume in excess of 100 may be considered large (12). The three interchange spacings selected for study gave 6, 10, and 14-s running travel times between the two intersections. We thought this range would include most signalized diamond interchanges; we have successfully tested the program calculations against field data from an interchange having slightly higher travel times (2). Last, cycle lengths of 60, 70, and 80 s were analyzed. Five interchange phasing codes (1, 1A, 2, 3, and 4) were analyzed for all possible offsets in 1-s increments, and a minimum delay was then selected for each of the 18 problem sets.

Results

Minimum delay results for the 50 vehicles/h U-turn volume problems are presented in Table 1. Table 2 contains the minimum delay results for U-turn volumes of 150 vehicles/h and shows minimum interchange delay and minimum delays for phase codes 1 and 1A. The minimum delay offsets (not shown) for phase codes 1 and 1A in all cases would provide signal phasings in the four-phase with overlap family.

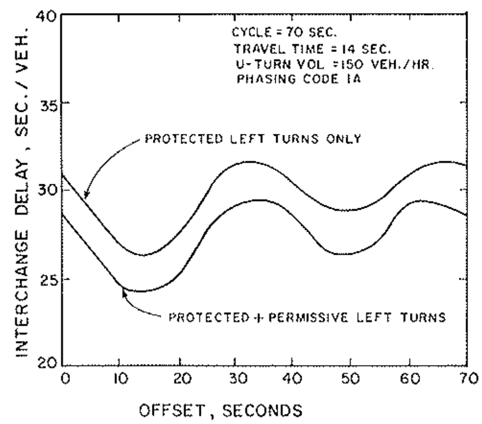
The minimum delay interchange phasing codes for all 18 signal problems studied are given in Table 3. Our most important finding was that every one of the possible interchange phasing codes produced a minimum delay solution in at least one of the 18 problems studied (determined from Table 3). As travel time increases (the distance between the ramps increases) from 6 to 14 s, the interchange phase code that provides minimum delay also changes.

Discussion

If the results of this study are as descriptive of the real world as we believe them to be, the varying opinions traffic engineers express concerning the relative merits of different diamond interchange phasing schemes seem to have been justified. For example, a four-phase can be better than a three-phase scheme in some cases; in other cases three phases are better than four. However, another phasing pattern may be better than either three or four.

We believe PASSER III removes the guesswork from selecting the best minimum delay signal phasing pattern at a pretimed diamond interchange. A total of 6300 interchange phasing options were analyzed to find the minimum delay phasing codes (shown in Table 3) and their respective interchange phasing patterns. This analysis was done at a total computer cost of $25 on the local university computer system running on the lowest

Figure 10. Reduction in interchange delay from addition of permissive left turns.



computer job priority level. A higher priority run (8 a.m. to 5 p.m.) would have cost only \$100. The PASSER III program will automatically select the best interchange phasing pattern.

Some of the literature might be interpreted to suggest that four-phase with overlap signalization has unusual advantages over other types. It is true that it does have some good features, for example, arterial progression, but no diamond interchange signal phasing pattern has mystical powers, not even four-phase with overlap. It is simply a lead-lead phasing arrangement that is timed to perfect progression for the front of the two arterial through platoons. As intersection spacing and travel time increase, the green times on the external movements at both intersections must be increased to maintain the perfect progression of the arterial through movements. This increase will result in an obvious increase in external signal capacity. Increasing external green times reduces the signal capacity and green times of the interior left-turn phases. In the standard lead-lead phasing arrangement (code 1), greens are split at the two intersections in proportion to the volumes at each intersection. Increasing the spacing does not change the green split, but progression may not be as good. As the previous results show, it is difficult to estimate what net effects these features will have on total average interchange delay.

PERMISSIVE LEFT TURNS

A number of states have begun using the protected left-turn phase at signalized intersections (left-turn arrow) followed by a permissive left-turn phase (left turn legal on circular green if clear) in order to increase high-volume intersection capacity. Texas has also begun using the protected-plus-permissive left-turn phasing at some critical diamond interchanges. This type of control effectively provides some left-turn capacity on the arterial through phase (Figure 2, phase A). This type of signalization may completely change preferred phasing patterns engineers are accustomed to using and may also change the minimum delay interchange phasing patterns for a given signalization problem (Table 3).

The PASSER III computer program can analyze the protected-plus-permissive phasing concept in either the leading or lagging phase sequence. The effects of opposing queues and traffic flow are considered. A mathematical model of this process has been developed by Fambro, Messer, and Anderson in a paper in this Record.

An example of the reduction in overall interchange delay that would occur if a permissive left-turn phase were added to phase A at both ramp intersections can be determined from Figure 10. In this case, an overall reduction in delay of approximately 2 s/vehicle would result. A much higher reduction in delay for the interior left-turn vehicles, where the capacity is increased, occurs when maximum queue lengths are shortened.

SUMMARY

The results of this study show that the best minimum delay, pretimed diamond interchange signal phasing pattern can be estimated using PASSER III. While signalization guidelines and preferred signal phasing patterns are helpful, their usefulness is limited and their performance uncertain. A detailed analysis of all pretimed signalization options can now be performed efficiently. We hope that at least some of the issues that have clouded diamond interchange signalization can now be analyzed.

ACKNOWLEDGMENTS

The research reported in this paper was performed during a project on the development of a frontage road level of service evaluation program (PASSER III) conducted by the Texas Transportation Institute and sponsored by the Texas State Department of Highways and Public Transportation in cooperation with the Federal Highway Administration, U.S. Department of Transportation. We wish to thank Herman E. Haenel and Elmer A. Koeppe of the Texas State Department of Highways and Public Transportation for their technical inputs and constructive suggestions throughout the project. The contents of this paper reflect our views, and we only are responsible for the facts and accuracy of the data presented. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration. This paper does not constitute a standard, specification, or regulation.

REFERENCES

1. C. J. Messer, H. E. Haenel, and E. A. Koeppe. A Report on the User's Manual for Progression Analysis and Signal System Evaluation Routine—PASSER II. Texas Transportation Institute, Research Rept. 165-14, Aug. 1974.
2. D. B. Fambro, D. L. Putnam, H. E. Haenel, L. W. Cervenka, and C. J. Messer. A Report on the User's Manual for Diamond Interchange Signalization—PASSER III. Texas Transportation Institute, Research Rept. 178-1, Aug. 1976.
3. P. K. Munjal. An Analysis of Diamond Interchange Signalization. HRB, Highway Research Record 349, 1971, pp. 47-64.
4. P. K. Munjal and J. W. Fitzgerald. A Simulation Model for the Evaluation of Real-Time Computer Control of Diamond Interchanges. Systems Development Corp., Document TM-4601/004/00, 1971.
5. C. J. Messer and D. J. Berry. Effects of Design Alternatives on Quality of Service at Signalized Diamond Interchanges. TRB, Transportation Research Record 538, 1975, pp. 20-31.
6. F. V. Webster. Traffic Signal Settings. British Road Research Laboratory, Crowthorne, England, Technical Paper 39, 1958.
7. F. A. Wagner, D. L. Gerlough, and F. C. Barnes. Improved Criteria for Traffic Signal Systems on

Urban Arterials. NCHRP, Rept. 73, 1969.
8. F. A. Wagner, F. C. Barnes, and D. L. Gerlough. Improved Criteria for Traffic Signal Systems in Urban Networks. NCHRP, Rept. 124, 1971.
9. N. Gartner. Microscopic Analysis of Traffic Flow Patterns for Minimizing Delay on Signal-Controlled Links. HRB, Highway Research Record 445, 1973, pp. 12-23.
10. L. Rach, J. K. Lam, D. C. Kaufman, and D. B. Richardson. Evaluation of Off-Line Traffic-Signal Optimization Techniques. TRB, Transportation Research Record 538, 1975, pp. 48-58.
11. S. Spitz. Signalization of Diamond Interchanges. Proc., Western Section, ITE, Anaheim, Calif., 1963.
12. D. L. Woods. Limitations of Phase Overlap Signalization for Two-Level Diamond Interchanges. Traffic Engineering, Sept. 1969, pp. 38-41.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Virginia's Crash Program to Reduce Wrong-Way Driving

N. K. Vaswani, Virginia Highway and Transportation Research Council, Charlottesville

Over a 4-year period beginning in 1970, wrong-way incidents and accidents were reduced on Virginia's interstate highways by 50 percent and on noninterstate four-lane divided highways by 70 percent. However, since 1975 an upward trend has been observed on interstate roads, while the downward trend has continued on noninterstate roads. This paper discusses the following engineering measures taken to reduce wrong-way driving: using reflectorized pavement arrows on ramps, eliminating pavement flares, providing stop lines across exit ramps near junctions with crossroads, continuing the pavement edge line across exit ramps, continuing double yellow lines on two-lane divided crossroads opposite exit ramps, reducing crossover width across exit ramps, adding guidance to local drivers on new interchanges, informing the driver of the geometry of the intersection before he or she enters it, and providing guidance for drivers at T-intersections without a crossover.

A survey of wrong-way incidents by the Virginia Department of Highways and Transportation and the Virginia State Police was initiated in June 1970 and has continued since, except for December 1970 to June 1971. The data collected show that until June 30, 1976, a total of 114 wrong-way accidents involving 54 deaths and 120 injuries had occurred on interstate highways. In the 167 accidents reported on other four-lane divided highways during the same period, 33 were killed and 173 injured.

Fatalities and injuries caused by wrong-way driving on interstate highways and four-lane divided highways in Virginia were compared with total accident fatalities and injuries on major highways in the state during 1970 to 1976. This comparison showed that although wrong-way accidents were relatively few compared to the total number of accidents they were exceptionally severe. The data showed that the fatality rate per wrong-way accident was 31 times greater than that for other types of accidents on interstate highways and 10 times that for other types on four-lane divided highways. The data also showed that the injury rate was 2.9 times that for other types of accidents on interstate highways and 2.3 times the rate for these on four-lane divided highways.

However, as shown below, the wrong-way incidents and accidents could not be related to the total accidents on a statewide basis for interstate and other four-lane divided highways. Table 1 gives the vehicle kilometers, total accidents, wrong-way incidents, and wrong-way accidents for each calendar year since 1970 for interstate, arterial, and primary highways.

These data show that on interstates the total number of accidents during 1972 was 1947 billion vehicle kilometers (1515/billion vehicle/miles) of travel. In 1973 the total dropped to 868 (1389), a decrease of 8.3 percent from 1972, which was possibly accomplished by the legislation effective in June 1972 that reduced the blood alcohol content (BAC) level from 0.15 percent to 0.10 percent for a presumption of drunk driving and stipulated a mandatory revocation of the driver's license for a period of 6 months for persons convicted of driving while intoxicated (DWI). Later, in December 1972, breath tests were introduced to make conviction for drunk driving easier.

On interstate roads in 1973 enforcement of these regulations sharply cut the total anticipated accident rate. In 1974 on the interstate highways the total number of accidents decreased to 639 billion vehicle kilometers (1022/billion vehicle/miles), a reduction of 26.4 percent from 1973. This shift might have been caused largely by the energy crisis of 1973 to 1974 and its accompanying reduction in the speed limit to 88.5 km/h (55 mph).

As shown in Figure 1, there was a considerable dip in wrong-way incidents and accidents at the beginning of 1973, possibly because of fear of DWI conviction. Later in 1973 the trend reversed and did not seem to be affected by the new legislation and reduced speed limit. The 26.4 percent reduction in total accidents during 1974 was apparently not reflected in the figures for wrong-way incidents and accidents (Figure 1 and Table 1).

From 1970 to 1973, when the total travel and accidents were increasing, incidents and accidents either remained constant or decreased. Since 1974, when total travel and accidents again increased, wrong-way incidents and accidents also tended to increase, a reversal of the relationship between total accidents and wrong-way incidents and accidents. Therefore there is no apparent relationship between vehicle kilometers of travel or total accidents and wrong-way incidents or accidents on interstate highways.

On arterial and primary highways total travel increased until 1973. In 1974 it decreased, probably be-

Table 1. Traffic and wrong-way driving and accident summary.

Year	Interstate				Arterial and Primary Highways			
	Annual Vehicle Kilometers of Travel (millions)	Total Accidents	Wrong-Way Incidents	Wrong-Way Accidents	Annual Vehicle Kilometers of Travel (millions)	Total Accidents	Wrong-Way Incidents	Wrong-Way Accidents
1970	7 532	6729	38[c]	12[c]	16 198	35 617	83[c]	18[c]
1971	8 393	8133	38[c]	13[c]	16 807	37 195	113[c]	26[c]
1972	9 568	9005	64	25	17 794	40 366	135	39
1973	10 515	9076	41	19	18 081	39 929	100	24
1974	10 197[a]	6474[a]	37	19	18 286[a]	35 125[a]	60	18
1975	10 916	6617[b]	42	17	18 902	30 747[b]	56	31
1976	—	—	28[d]	12	—	—	22	9

Note: 1 km = 0.62 mile.

[a] The energy crisis during 1973 and 1974 reduced the total travel. A reduction in the speed limit to 88 km/h (55 mph) also accounted for fewer total accidents.
[b] Effective January 1, 1975, the accident reporting threshold changed from $100 to $250.
[c] June through December.
[d] January through June.

cause of the energy crunch, and began a continuing increase in 1975. The total accident rate on the arterial and primary highways increased from 1970 through 1972 (as shown in Table 1), and since then has decreased (see Figure 2). This reduction might be attributed to the 1972 change in DWI legislation. For four-lane divided noninterstate highways, no statistical relationship seems to exist between total travel or total accidents and wrong-way incidents or accidents.

From Figure 1 we see that wrong-way incidents decreased, during 1970 through 1974, from 38 during the first 7 months to 18 during the second 6-month period of 1974. For the same periods, wrong-way accidents on the interstate system fell from 12 to 6, about a 50 percent reduction.

Figure 2 shows that wrong-way incidents from 1970 to 1976 decreased from 83 during the first 7 months to 22 during the last 6-month period. Wrong-way accidents on the four-lane divided noninterstate highways simultaneously fell from 18 to 9, reductions of more than 70 percent in wrong-way incidents and approximately 50 percent in the accident rate during the 4-year period.

Figures 1 and 2 also show the proportional decrease in day and night accidents from wrong-way entries. These data for 1974 to 1976 show very low rates both of increase in wrong-way incidents and the accident rate on interstate highways and of reduction on four-lane divided noninterstate highways.

The data collected from 1970 to 1976 show that about 50 percent of the wrong-way entries originate from interchanges; about 15 percent originate at crossovers and rest stops or are associated with U-turns and median crossings. The origins of the remaining 35 percent are unknown.

On noninterstate four-lane divided highways, about 40 percent of the drivers making wrong-way entries emerge from intersections with crossroads or exit ramps connecting with interstate roads; about 25 percent originate from business establishments such as gas stations and motels; and about 20 percent originate from residential areas, crossovers, beginnings of divided sections, and construction sites or are associated with U-turns and median crossings. The origins of the remaining 15 percent are unknown.

In order, therefore, to reduce wrong-way driving, improvements have been focused on interchanges, intersections, and exits from business areas. In addition, some crossovers on interstates have been closed. U-turns, which are more common on interstates than on other four-lane divided highways, do cause wrong-way driving, but no preventive measures for them have been devised.

The drunk driver compounds the problem of wrong-way driving. A survey showed that, out of the 287 wrong-way drivers spotted on interstate highways, 152 were drunk, 85 were not drunk, and the condition of the others was not stated. On the noninterstate four-lane divided highways, 188 out of 569 wrong-way drivers were drunk, 302 were not drunk, and the condition of the remainder was not stated. Among the 302 sober drivers on noninterstate highways, many could have made intentional wrong-way maneuvers to cut driving distance (Figure 3). Also among the sober wrong-way drivers were quite a few whose responses to stimuli were impaired because of age, sickness, medication, or other factors.

The 1972 Virginia legislation on drunk driving does not seem to have reduced wrong-way driving by drunk drivers. Figure 4 shows that although the total number of wrong-way incidents on interstates decreased after June 1972 it was the elimination of faulty maneuvering by sober drivers rather than improved drunk driving that caused the reduction. Figure 5, however, shows that on noninterstate highways the decrease in the number of wrong-way incidents was attributable to improvements by both drunk and sober drivers. The data do not, therefore, confirm that the change in the law has contributed to the reduction of wrong-way incidents. The reduction probably resulted from other factors such as the engineering measures and increased public awareness of wrong-way driving.

ENGINEERING MEASURES TO REDUCE WRONG-WAY DRIVING

The most significant of the many engineering measures instituted in Virginia during the program to reduce wrong-way driving will be discussed in what follows.

Reflectorized Pavement Arrows on a Ramp

All interstate exit and entry ramps in Virginia have been provided with 5.6-m (18.8-ft) reflectorized direction arrow indicators. Two arrows are placed on the exit ramp and one on the entrance ramp, and the distance from the tip of the arrow to the stop line on the exit ramp varies from 1.5 to more than 7.5 m (5 to 25 ft).

Observations have shown that arrows placed close to the intersection of the crossroad and the exit ramp are visible to the potential wrong-way driver making either a right or a left turn onto the exit ramp as shown in Figure 6; those placed beyond 4 to 5 m (13 to 16.5 ft) from the intersection are not visible. Therefore, the first arrow on the exit ramp should be very close, say within 1.5 m, to the junction of the exit ramp with the crossroad. This arrow can then be seen by the potential wrong-way driver during the day and at night under low-beam headlights. The second arrow on the exit ramp should be placed approximately 30 m (100 ft) from the stop line (1) as a second warning to the wrong-way driver.

Similarly, the arrow on the entrance ramp can guide the driver from the crossroad into the correct direction only if the driver can see it from the crossroad. An arrow far from the junction of the crossroad and entrance ramp will not be visible to the driver from the crossroad and will fail to perform its function, except to reassure the the driver after he or she is already on the entrance ramp.

Elimination of Flares

On almost all interchanges on which wrong-way entries have been made, either into an exit ramp or from an exit ramp onto a crossroad, the left edge of the left lane of the exit ramp has been flared into the right pavement edge of the crossroad. An example of such a flare, which probably misleads the driver, is shown in Figure 7.

Figure 1. Incident and accident data for Virginia interstate system.

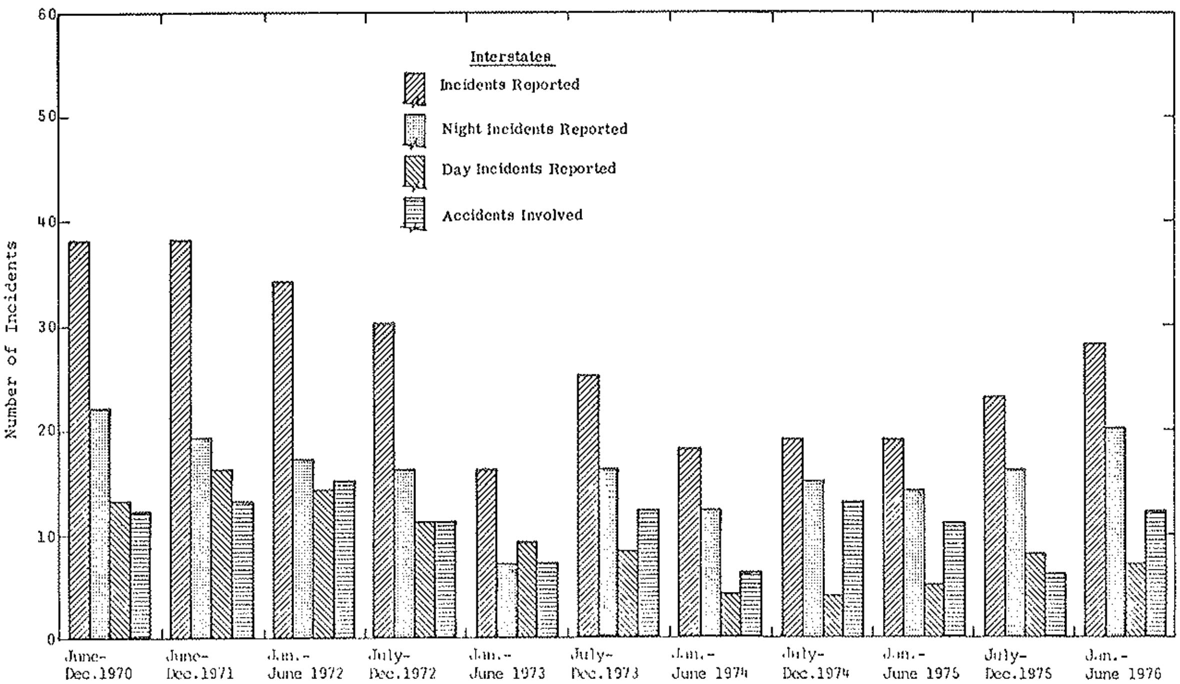


Figure 2. Incident and accident data for noninterstate Virginia highways.

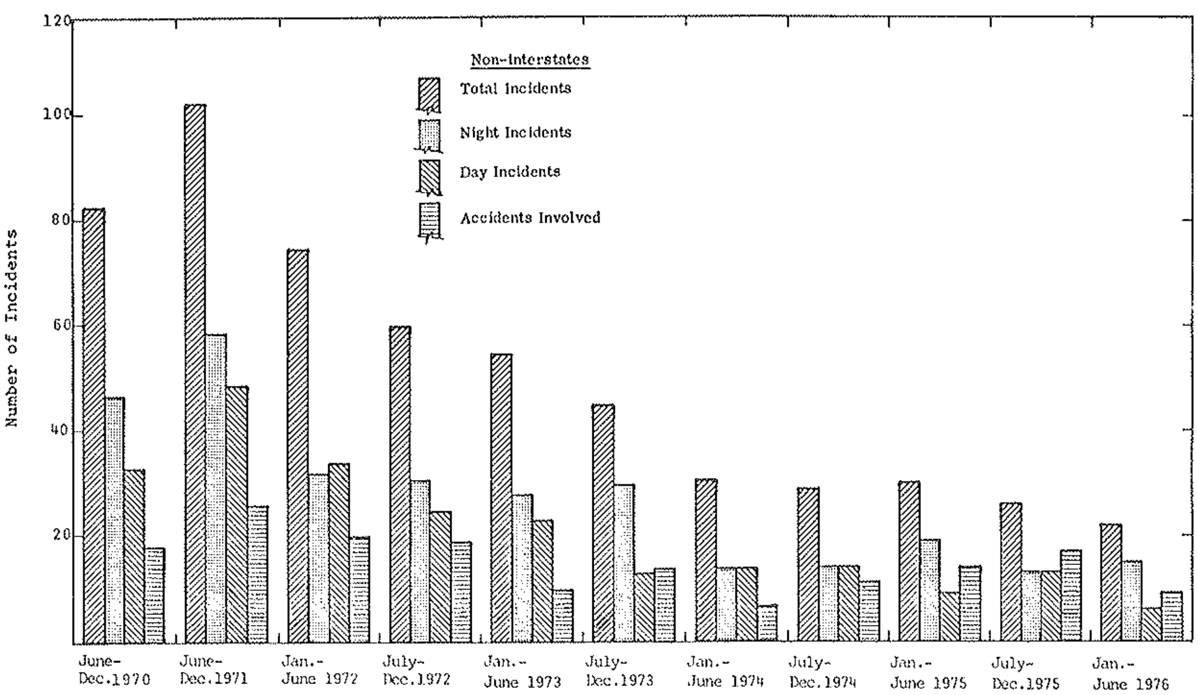
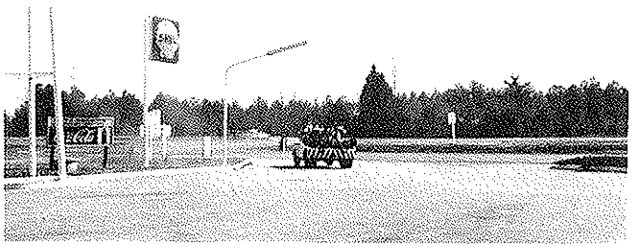


Figure 3. An intentional wrong-way exit from a gas station.



The results of removing these flares have been very encouraging, and the following example shows how wrong-way entries can be completely eliminated.

This flare at the intersection of an interstate exit ramp and a primary highway experienced six wrong-way entries (the highest in Virginia), all by sober drivers during the first 2 years of the survey period. All approaching drivers would stop on the stop line. In 1974 I suggested that the flare be eliminated by installing two right-angled white lines (2), and over the next 30 months no wrong-way incidents were reported. The marking

Figure 4. Wrong-way incidents and drunk and nondrunk drivers on the Virginia interstate system.

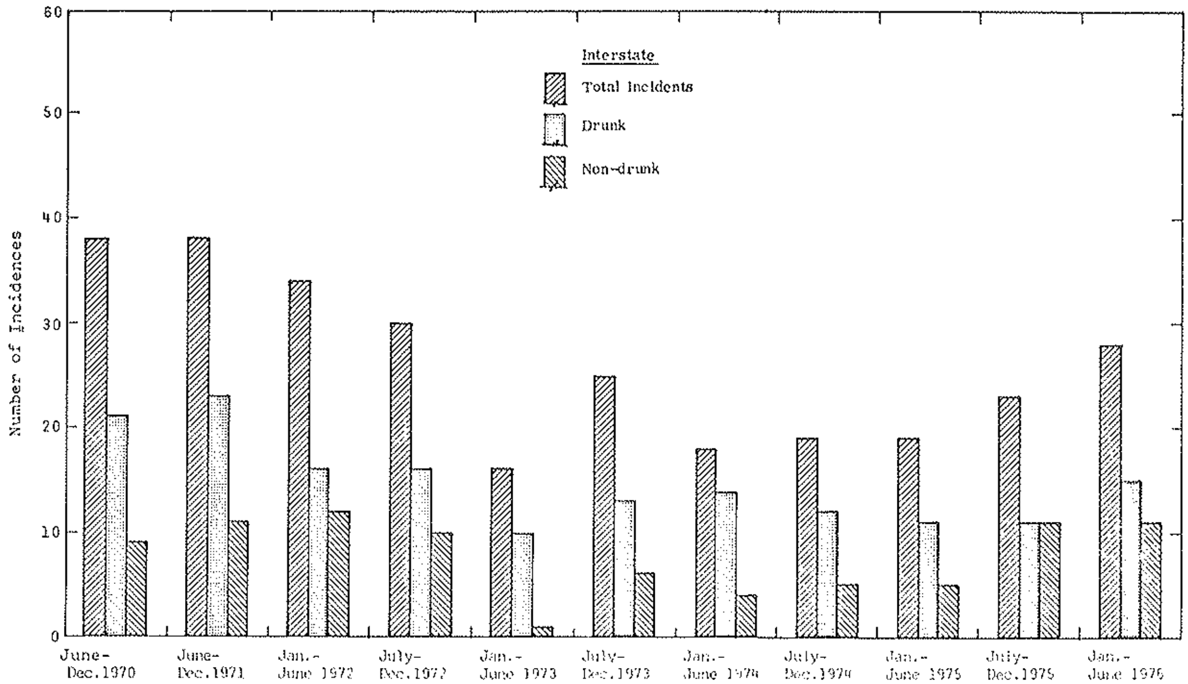
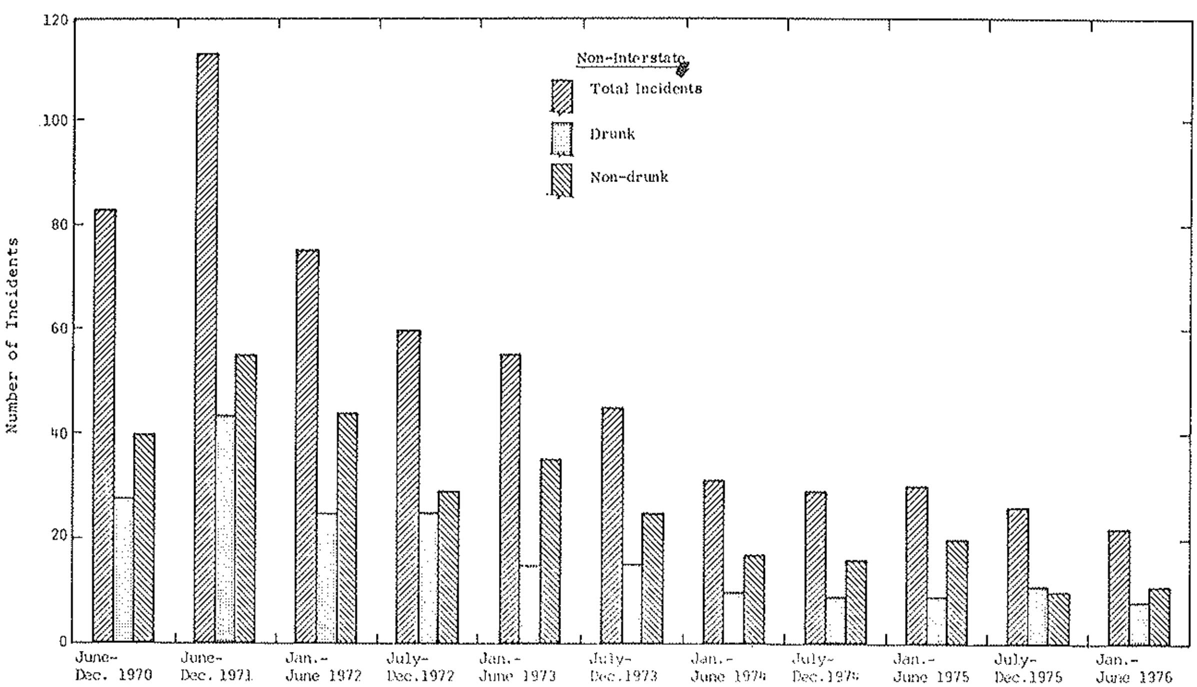


Figure 5. Wrong-way incidents and drunk and nondrunk drivers on noninterstate four-lane divided Virginia highways.



(Figure 8) apparently discourages drivers from quickly turning left onto the wrong side of the median and also increases visibility on the primary highway. A driver who needs more visibility distance crosses the stop line and comes to a stop at the corner of the flare marking.

Other examples of removing flares are given elsewhere

Figure 6. Pavement arrow markings near stop line (no incidents).

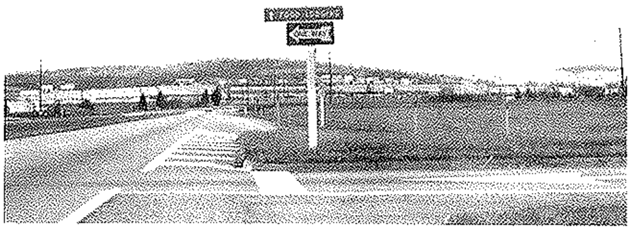


Figure 7. Junction of exit ramp and primary highway before marking left flare (six incidents).

Figure 8. Junction of exit ramp and primary highway after marking left flare (note that driver ignores stop line to get better view of crossroad; no incidents after marking).

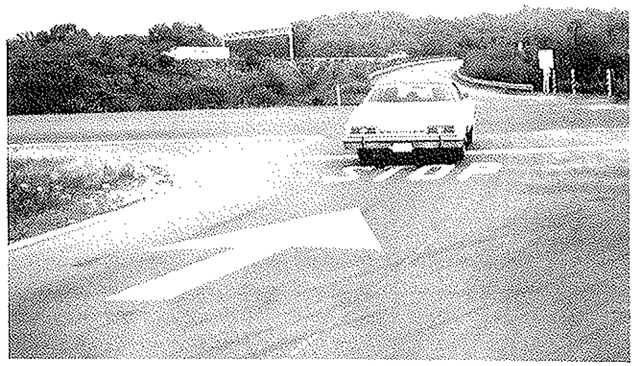
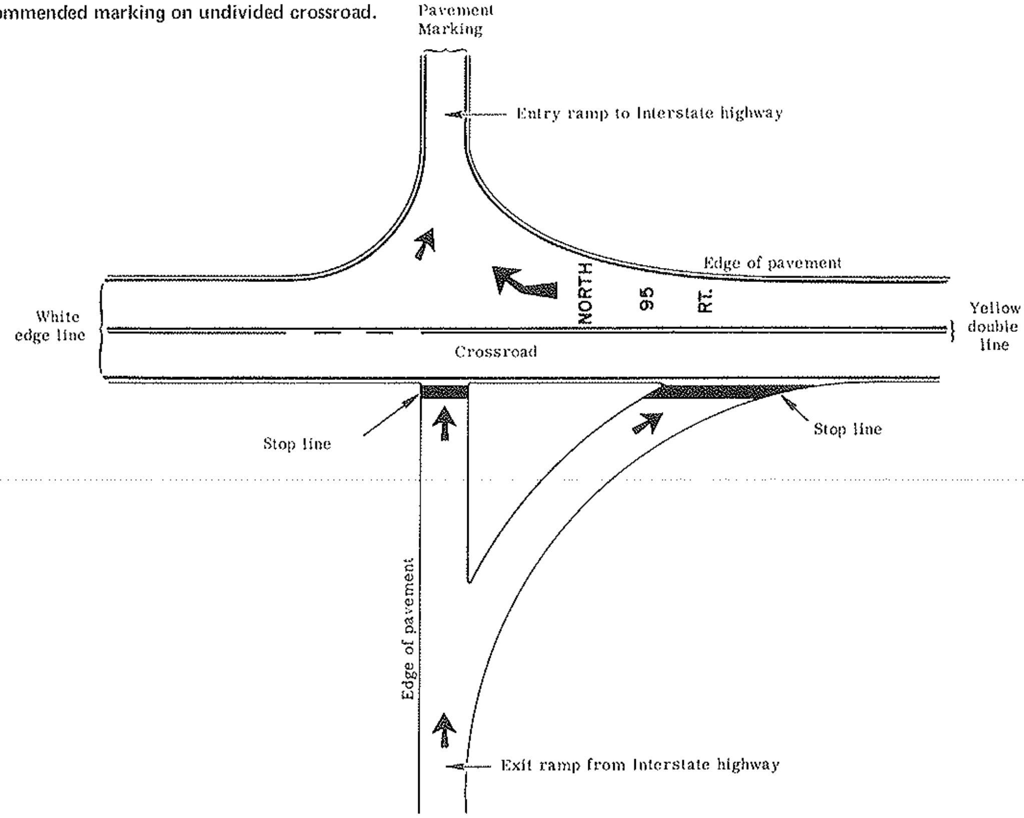


Figure 9. Visibility of stop line from a considerable distance on crossroad (no incidents after marking).

Figure 10. Recommended marking on undivided crossroad.



(3), and Scrifes has also recommended flare removal to reduce wrong-way driving (4).

Stop Line

Traffic on one-way exit ramps of all partial interchanges must stop at the ramp's junction with the crossroad. During field investigations it was observed that many exit ramps that had been involved in wrong-way entries onto the crossroad or onto the interstate highway did not have stop lines.

The stop line probably has the following two advantages:

1. More drivers tend to stop for a stop line than for a stop sign only, and
2. The stop line probably discourages the driver on the crossroad from entering the exit ramp.

During our investigations we found that at a few locations the stop line was very close to the crossroad and exit ramp intersection. Such lines were found to be very clearly visible from considerable distances ahead of the intersection while driving along the crossroad during the day and at night under low-beam headlights.

In Figure 9 a stop line placed very close, say, within 1.5 m of the junction of the crossroad and the exit ramp, is visible to the driver on the crossroad day and night and will be of immense help in preventing wrong-way entries into the exit ramp. Figure 10 shows the placement of a stop line at an intersection.

Continuation of Pavement Edge Line

Drivers are now so accustomed to the pavement edge line that they seem to use it as a guide. Drunk drivers or drivers with poor responses probably depend on this line because of their restricted ranges of vision. If this line were continued across the exit ramp (see Figure 10) it would make the ramp inconspicuous to such drivers on the crossroad, especially at night. In fact, not continuing the pavement edge line is very dangerous, because in Virginia it has been observed that the line has usually been continued into the exit ramp.

One intersection where there was a wrong-way entry into the exit ramp at night had a line continued into the exit ramp that encouraged a driver with a low response to enter the exit ramp. An exit ramp with no stop line visible to the driver is even worse.

Double Yellow Lines on Two-Lane Undivided Crossroads

Undivided crossroads at interchanges are provided with double yellow lines with very wide gaps opposite exit ramps. In one case a wrong-way drunk driver entering this gap caused an accident in which six were killed (2). There is a strong possibility that a gap in the lines opposite the exit ramp encourages wrong-way entry through the gap onto the exit ramp. I therefore recommend that no gap be provided in the double yellow lines and that double yellow lines be continued opposite the exit ramp as shown in Figure 10.

Some undivided crossroads at interchanges provided with solid double yellow lines have had no reported wrong-way entries. It has also been observed that continuous double yellow lines do not cause any inconvenience to drivers who cross these lines to negotiate an interchange. Providing continuous double yellow lines is not expensive, and I am of the opinion that they prevent some wrong-way entries without interfering with normal traffic.

Added Guidance at Unfamiliar Interchanges

Wrong-way incidents are more common at interchanges during the first year or two after construction than in later years. It has been found that it is local drivers who are involved in such wrong-way incidents, probably because they have had no previous experience with interchanges. More facilities for guiding drivers need to be provided during the first two years after construction.

The most economical measure would be pavement markings (on crossroads) that would fade after a year or two. Instructions and an arrow, as shown in Figure 10, are likely to be helpful to a driver who is confused by a new interchange.

Reduced Width of Crossover Opposite the Exit Ramp at Interchanges

The width of the crossover opposite the exit ramps is often many times more than that needed by a vehicle making a left turn around the nose of the crossroad median from the exit ramp. This excess has contributed to a number of wrong-way maneuvers onto and from the exit ramps. Virginia's traffic engineers realize the need for channelizing crossovers, and the state is encouraging such programs.

I investigated locations where the widths of the crossovers were reduced and found that even after extension of the nose of the median there was often an ample gap between the nose and the largest truck.

A photograph showing this gap at one of the interchanges is given in Figure 11. The left median in this photograph was recently extended to reduce the width of the crossover. A simple method has been developed (5) to help the traffic technician locate the nose for a channelized crossover. By this method the shape and location of the two noses on either side of the crossover can be determined, and the crossover can be properly channelized.

Optical Illusions at Night

Wrong-way entries can be caused by drivers experiencing optical illusions. I have observed two such wrong-way entries, one of which is shown in Figure 12, a photograph of the intersection. On inquiry, the driver (local, sober) said that he did not see the lane on the far side of the median. This is a level intersection, and unless a person is very observant he or she is unable to see these lanes.

A divided highway sign on the left before a left turn will inform the driver of the geometry of the intersection. An additional turn sign at the left nose of the median would provide further guidance (Figure 13). These signs have been installed on an experimental basis at 72 intersections over a 92-km (57-mile) stretch of a primary highway in Virginia.

Our divided highway sign is the same as that used in Delaware to reduce wrong-way entries. This sign has been used in Delaware for several years, and engineers there say that it has been very effective. They claim that in 1967 wrong-way incidents were reduced to five accidents, or 0.004/million (0.006/million vehicle/miles) vehicle kilometers traveled. This number was far below the national average.

T-Intersections Without a Crossover

Small business areas, such as gas stations, clubs, restaurants, and motels, and small residential areas are not provided with exits through the medians of divided highways. At such locations an exiting driver

Figure 11. Extension of median nose at wrong-way entry site (note ample gap between nose and truck).

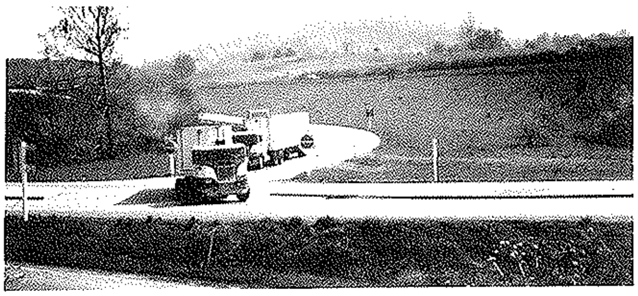


Figure 12. Poor visibility of lanes across median at wrong-way entry site.

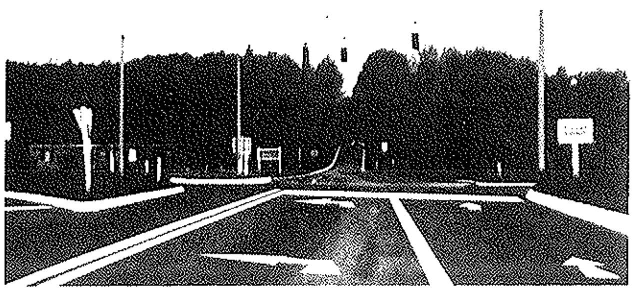
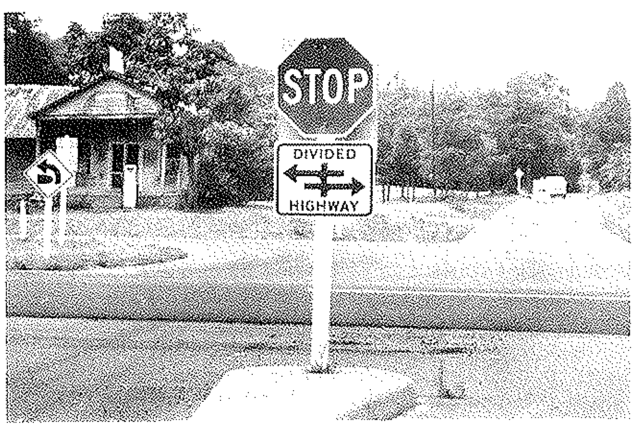


Figure 13. Location of divided highway and turn signs.

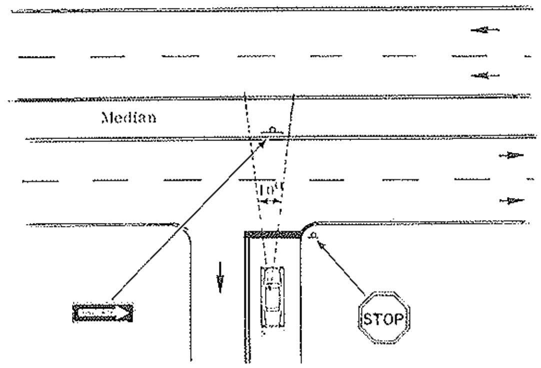


assumes the road to be a two-lane highway and sometimes makes a wrong-way entry. In most cases, there are one-way arrow signs opposite these exits. A survey has shown that there is no definite pattern in the location of the signs, and in many cases they are not visible with low-beam headlights. The best use of this sign might be obtained by placing it opposite the vehicle coming out of a business or residential area as shown in Figure 14.

Raised Pavement Markers

In addition to the investigations leading to the above improved signs and pavement markings, one of the other noteworthy recommendations has been to provide raised pavement markers spread over the width of the pavement on a short section of the exit ramp. The feasibility of such markers has been determined (6), and they have now been placed at two exit ramps for further study.

Figure 14. Recommended sign placement in small residential or business area.



CONCLUSIONS

1. Data show that in Virginia incidents of wrong-way driving on interstate highways decreased by more than 50 percent in the 4 years following 1970; on noninterstate four-lane divided highways the reduction was more than 70 percent.
2. The reduction in wrong-way incidents was probably the result of engineering measures and increased public awareness of the wrong-way driving problem.

ACKNOWLEDGMENTS

This study was conducted under the general supervision of J. H. Dillard, head of the Virginia Highway and Transportation Research Council, and was financed with state research funds. The contents of this paper reflect my views, and I am responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official views or policy of the Virginia Department of Highways and Transportation or the Research Council.

REFERENCES

1. Pavement Arrows. Traffic and Safety Division, Virginia Department of Highways and Transportation, Memorandum T&S 99 on Pavement Markings, June 29, 1973.
2. N. K. Vaswani. Case Studies of Wrong-Way Entries at Highway Interchanges in Virginia. TRB, Transportation Research Record 514, 1974, pp. 16-28.
3. N. K. Vaswani. Measures for Preventing Wrong-Way Entries on Highways. Virginia Highway and Transportation Research Council, 72-R41, June 1973.
4. P. N. Scrifes. Wrong-Way Movement on Divided Highways. Purdue Univ., Project C-36-5900, Feb. 1974.
5. N. K. Vaswani. Engineering Measures for Reducing Wrong-Way Driving. Virginia Highway and Transportation Research Council, 76-R8, Sept. 1975.
6. F. D. Shepard. Installation of Raised Pavement Markers for Reducing Incidents of Wrong-Way Driving. Virginia Highway and Transportation Research Council, 76-WP27, Feb. 1976.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Effect of Flashing Beacons on Intersection Performance

R. B. Goldblatt, KLD Associates, Inc., Huntington Station, New York

This paper presents the results of a study on the operational effects of various types of continuously and vehicle-actuated flashing traffic control devices performed at the Federal Highway Administration's Maine facility. Both electronic and manual data collection techniques were used. Five intersection and three advance warning device configurations were tested at the intersection of US-2 and Me-152. The use of continuously flashing intersection beacons along stopped approaches encourages speeds consistently lower than those achieved by STOP signs or vehicle-actuated intersection beacons. Certain vehicle-actuated advisory warning devices helped to reduce speed variance on major (nonstopped) approaches. A vehicle-actuated STOP AHEAD beacon caused drivers to begin braking sooner than they would without a beacon. Reduced speed variance was also noted when the advance warning beacon was used. These effects disappeared if there was a beacon at the downstream intersection.

The use of flashing beacons at intersections is authorized by the Manual on Uniform Traffic Control Devices (1). These beacons supplement existing stop signs at intersections where traffic flow conditions do not justify the installation of a traffic signal but where a special hazard exists.

Although there is evidence (2, 3) of reduced accident rates accompanying the installation of intersection beacons, little work has been done to determine their effects on traffic flow. This paper presents the results of a study of the effects of both continuously and vehicle-actuated flashing beacons on a two-way stop controlled intersection. Both intersection beacons and STOP AHEAD beacons were tested. The work was performed as part of a larger study the Federal Highway Administration (FHWA) sponsored (4) to determine guidelines for the installation of various types of flashing beacons.

DESCRIPTION OF STUDY SITE

The performance studies were conducted at the FHWA Maine facility with the cooperation of FHWA, the Transportation Systems Center, and the Maine Department of Transportation. The Maine facility (5) is a computer controlled data acquisition system designed to track and record the passage of vehicles along an instrumented two-lane rural road. This system is installed along US-2 between Newport and Canaan, Maine, a distance of 24 km (15 miles).

Instrumentation consists of a configuration of four sensor loops that indicate vehicle location and direction. Each set of loops is defined as a node. Nodes are embedded at 61-m (200-ft) intervals along US-2 and on the important side-road approaches. Sensor nodes are connected to roadside electronic packages that process detector data and transmit it to the central computer. Although not all the nodes have a corresponding electronic package, the facility has a number of portable electronic packages that can be installed for individual experiments.

This experiment was performed at the intersection of US-2 and Me-152. The through traffic approaches studied were the eastbound and westbound approaches along US-2; the stopped approach was northbound along Me-152. Although the intersection has four legs, our equipment limitations allowed only one stopped approach to be studied. Figure 1 presents an intersection schematic.

EXPERIMENTAL DESIGN

Five intersection and three advance warning configurations were tested parallel to each other in order to study possible interaction. The configuration of the intersection devices is given below.

1. The existing conditions consisted of two-way stop without supplementary beacons. The major approaches along US-2 were uncontrolled.
2. A standard crosswire-mounted intersection beacon (1, 4E-3, p. 5) was added. The beacon flashed red to the approaches along Me-152 and yellow to the approaches along US-2.
3. The beacon was actuated by the approach of a vehicle along a minor (Me-152) approach. The detector used to actuate the intersection beacon was placed 114 m (373 ft) upstream of the intersection.
4. A set of vehicle-actuated WHEN FLASHING—VEHICLE CROSSING advisory warning signs and beacons were installed upstream of the intersection on US-2. The beacon flashed yellow to the nonstopped approaches along US-2 and was actuated by the same detector as the intersection beacon. Two advisory warning signs were tested (Figure 2); the dimensions corresponding to the letters in Figure 2 are A = 91 cm (36 in), B = 9 cm (3.5 in), C = 3.8 cm (1.5 in), D = 2.2 cm (0.9 in). The first, used in Charlotte, North Carolina (6), employed nonstandard colors; the second utilized standard colors (1, 2C-2).
5. In addition to configuration 4, a vehicle-actuated STOP sign beacon was mounted atop the STOP sign. The beacon flashed red to the approach along Me-152 and was actuated in the same manner as the other beacons tested.

The configuration of the advance warning devices is given below.

1. The existing condition was a STOP AHEAD pavement marking (1, 3B-18) and a standard STOP AHEAD warning sign (1, 2C-14).
2. A standard continuously flashing 30-cm (12-in) yellow beacon (1, 4E-1) was mounted above the standard STOP AHEAD sign.
3. The yellow beacon was actuated by the approach of a vehicle along the northbound lane of Me-152. The detector used to actuate the beacon was placed 91 m (300 ft) upstream of the STOP AHEAD sign.

Figure 3 is an intersection schematic showing the location of all control devices and sensor nodes.

DATA COLLECTION AND REDUCTION

Data Collection

Data were collected both electronically and manually. Data were collected electronically with the facility's computer, which polled each node 16 times/s as shown on Figure 3. Data include time of arrival at each node; vehicle length, direction, and speed; and the amount of time each vehicle spends at the stop line on the Me-152 northbound approach waiting for a gap.

For a sample of individual vehicles, three types of manual data were collected: the state in which the vehicle was licensed, whether the vehicle violated the stop line or made a rolling stop on Me-152, and the point on the approach where a vehicle's brake lights were observed for the first time.

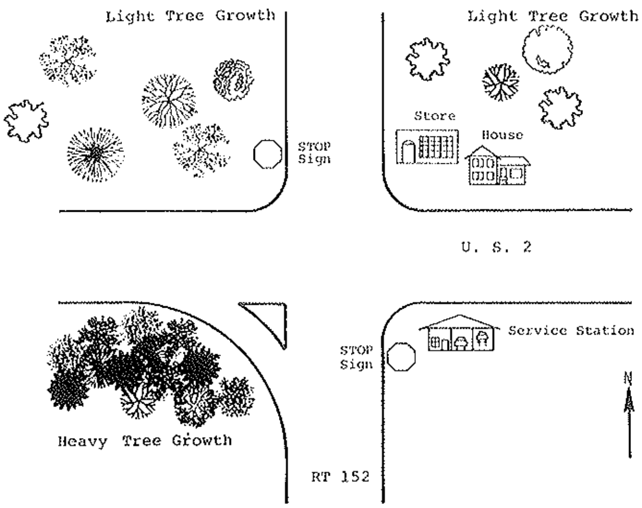
License plates were noted in an attempt to discern whether drivers familiar with the study site reacted any differently to the devices tested than those unfamiliar with the location. Drivers of vehicles with out-of-state license plates were considered unfamiliar.

The collection of stop line violation data was designed to test whether changes in the type of intersection control affected compliance. Unfortunately, no usable data were obtained because the manual data collection crew assigned to this task were not well trained.

Brake light data were collected to see whether the point at which brakes were applied was affected either by intersection control or by type of advance warning. These data were put directly into the roadside electronics package by a manual input channel.

Data were collected in Maine between February 27, 1975, and September 30, 1975. The first three treatments were each run for a week. However, because of the light traffic on Me-152 and the need for additional time for acclimatization, the length of each treatment was extended to 2 weeks. A total of 24 h of manual data were collected for each treatment. Manual data were collected during the day under weather conditions no worse than cloudy skies and bare, wet pavement.

Figure 1. Intersection schematic of US-2 and Me-152.



Data Reduction

A set of computer programs was written to decode and process the data, which, because of the large amount, were reduced and placed in a shorter format. The information from each run (there might be up to 40 runs per treatment) was processed into brakelight distance in meters and speed in meters per second for nodes 21-0, 19-2, and 19-1 northbound on Me-152, nodes 9 and 15 eastbound on US-2, and nodes 27 and 23 westbound on US-2.

If the analysis of data was to be meaningful, it was necessary to ascertain the expected measurement errors inherent at the Maine facility. Figure 4 presents the

Figure 2. Advisory warning sign for major approaches.

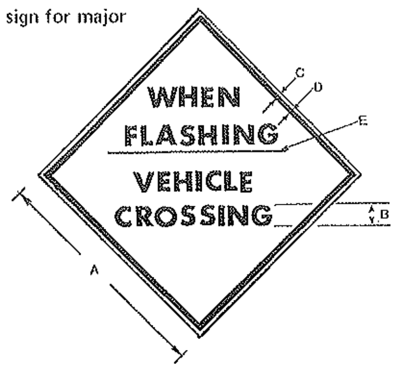
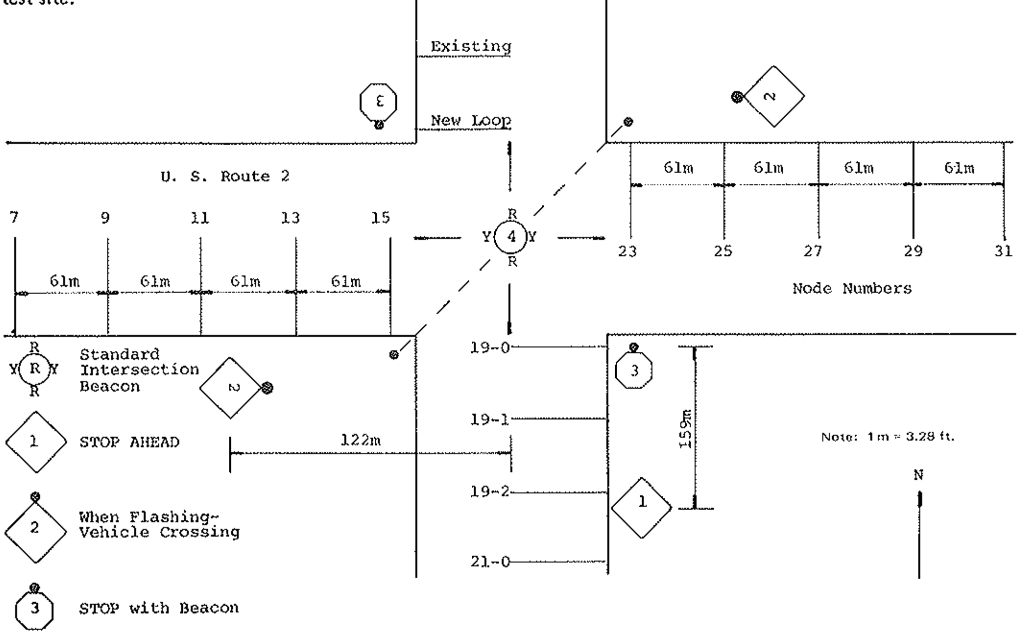


Figure 3. Schematic of Maine facility test site.



estimated error for both speed and brake light data.

A trap is defined as the distance between two successive sensor nodes. Our standard trap length was 61 m (200 ft). However, for the purposes of this experiment, trap lengths between 38 and 76 m (123 and 250 ft) were used.

The speed error was caused by the fact that the computer polled each node 16 times/s, which gave a time of arrival error on the order of $\pm\frac{1}{16}$ s. The error in the brake light measurements was caused by the computer time of arrival error and the observer reaction time error, estimated at $\frac{1}{2}$ s.

RESULTS

Intersection Control Devices

The effect of traffic control on speed along the main road (US-2) can be seen in Figure 5. Mean speeds and 85-percentile speeds are shown for the eastbound and westbound approaches to the intersection. Two positions are specified, the near position at the intersection and the far position approximately 122 m (400 ft) upstream of the intersection along US-2.

The five control types shown in Figure 5 are

1. Uncontrolled,
2. Standard with continuous intersection beacon,
3. Intersection beacon actuated by side street vehicles,
4. Intersection beacon with an advisory warning sign and beacon (Figure 2) placed along the major approaches and actuated by side street vehicles, and
5. Same as 4 except for a different advisory warning sign.

The t-test was used to determine whether the mean speeds associated with the control type differed significantly from the inherent measurement error. If the observed difference in mean speeds was less than the expected measurement error, the results could not be significant. Significance was determined by a two-tailed test at the 0.05 level.

The results showed that control type did little to explain changes in speed along the eastbound approach. However, along the westbound approach the use of the advisory warning signs and beacons seems to be of great importance in reducing speeds along the westbound approach, and significant differences are apparent for all devices except the uncontrolled case at the far location and for all devices at the near location.

The effects demonstrated for mean speeds are accentuated when 85-percentile speeds are used as a measure. Reductions in 85-percentile speeds are proportionally greater than the reduction in mean speeds. This fact indicates that the devices have greatest effects on faster vehicles.

The difference between the eastbound and westbound approaches can be explained by the following intersection geometry:

Direction	Grade (%)	Sight Distance (m)	Sight Time (s)
Eastbound	-3	305	14
Westbound	+3	152	8

The difference in grade explains the difference in approach speed, whereas the increased effectiveness of the westbound control devices might be related to restricted sight distance.

Figure 6 presents composite speed profiles along the approach to the stop sign on Me-152 for various intersection controls. The advance warning was a simple STOP AHEAD sign without beacons. The results in-

Figure 4. Maine facility measurement error.

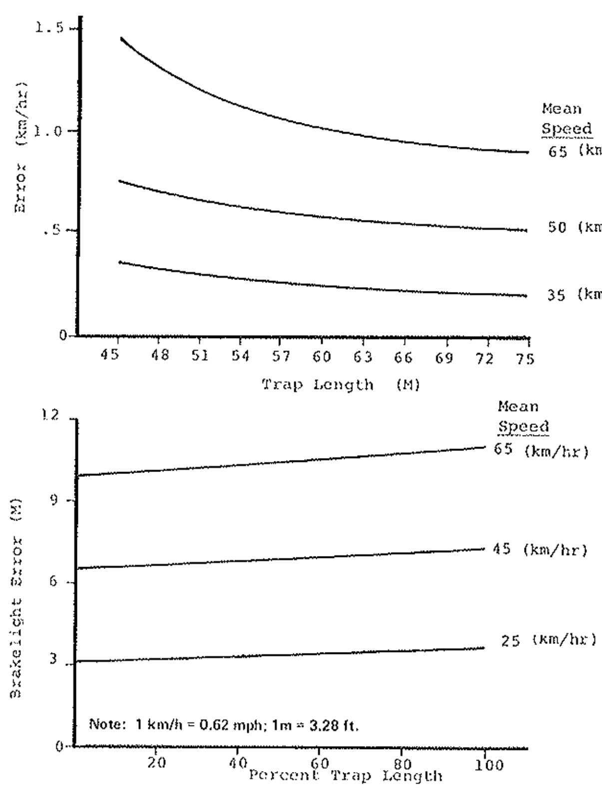
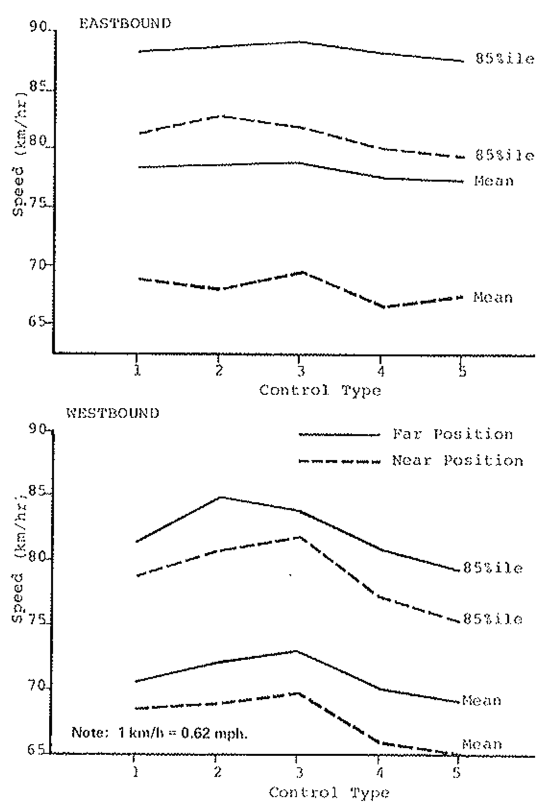


Figure 5. US-2 speed data.



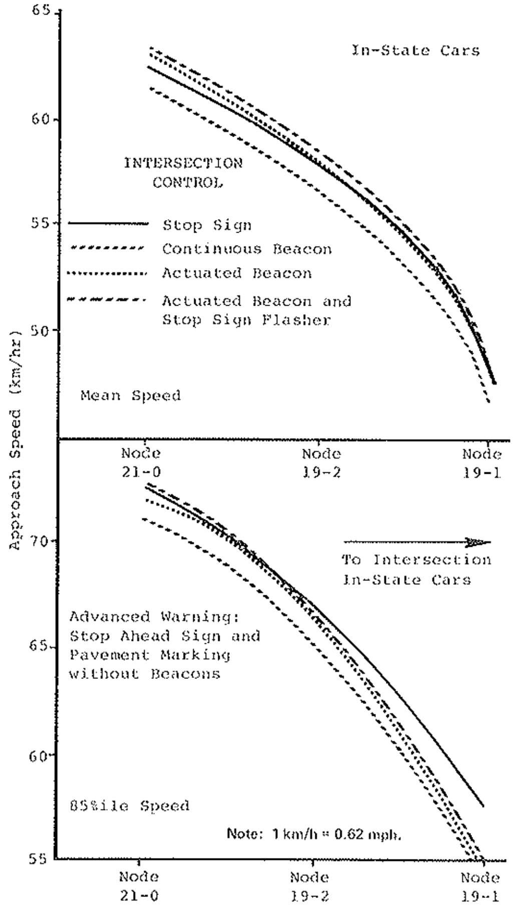


Figure 6. Me-152 speed profiles.

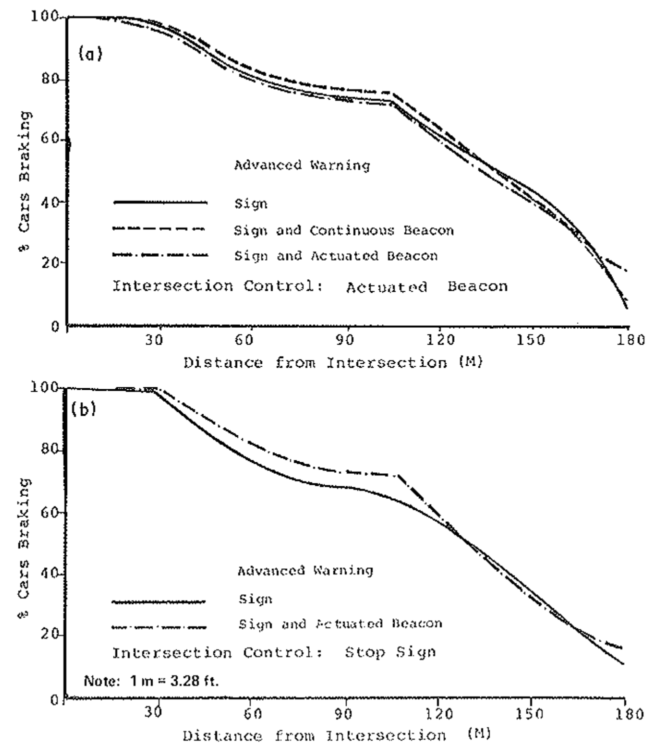


Figure 7. Effect of advance warning device on brakelight applications.

dicated that a simple flashing beacon encouraged significantly lower speeds at all locations. The use of actuated beacons and the addition of actuated stop sign beacons did not result in speed profiles significantly different from ones caused by the stop sign alone.

Day and Night Effects

The effects of a number of devices were tested both at night and during the day. The only difference noted was a slight decrease in mean speed at night that was found to be statistically insignificant.

Seasonal Differences

A number of identical control conditions were run during winter and summer. The results of these tests were not unexpected: summer conditions showed slightly higher mean speeds and lower variances than corresponding winter conditions. These results were tested statistically and were shown to be marginally significant.

Effect of Vehicle Population

Three types of vehicles were identified: in-state automobiles, out-of-state automobiles, and trucks. Analysis of traffic flow data indicates no significant changes that can be attributed to vehicle population.

Advance Warning Devices

Analysis of the speed data indicates that no significant changes in mean speed occur when a beacon is added to the STOP AHEAD sign. This result is independent of whether beacons are used at the intersection downstream or not.

Although mean speeds show no significant differences, there is a significant reduction in speed variance. It is interesting to note that the reduction in speed variance that occurs when a yellow beacon is added to the STOP AHEAD sign is significant only when the intersection control downstream is a simple stop sign. When any type of beacon control is present downstream, speed variance is unchanged or slightly increased when a continuously or actuated flashing beacon is used with an advance warning sign.

Figure 7 presents the cumulative brake light application distributions for vehicles approaching the stop sign, and the distributions for the three types of advance warning configurations tested (Figure 7a). The control at the intersection was an actuated intersection beacon. Note that there are no significant differences between the distributions. On the other hand, Figure 7b compares a STOP AHEAD sign alone and one with a vehicle-actuated beacon. The sign with beacon was 160 m (523 ft) from the stop line. In both cases, the downstream intersection was controlled by a stop sign without beacons. A Kolmogorov-Smirnov test was used to determine significance. The use of an actuated beacon significantly increased the percentage of vehicles braking at a point downstream of the STOP AHEAD sign.

CONCLUSIONS

A number of conclusions were drawn from our fieldwork in Maine regarding the use of beacons at stop sign controlled intersections. These conclusions were

1. The installation of continuously flashing intersection beacons without WHEN FLASHING—VEHICLES CROSSING advisory warning signs has little or no effect

on the main street speed distribution;

2. A continuously flashing beacon encourages lower vehicle speeds along the stopped approach, but not if the beacon is actuated; and

3. The use of the actuated WHEN FLASHING—VEHICLE CROSSING signs and beacons along the main street approaches causes a reduction in speed dispersion along the approach, which is more pronounced on the approach with poor sight distance.

The use of advance warning beacons in conjunction with a STOP AHEAD sign was found to reduce speed variance. In addition, vehicles begin the braking maneuver farther from the intersection. However, these results become less significant when any beacon is used at the downstream intersection, probably because the intersection beacon flashes red while the STOP AHEAD beacon flashes yellow. This presents the driver with conflicting indications and negates any positive benefits. There does not seem to be any operational advantage to actuating an advance warning beacon.

ACKNOWLEDGMENT

The work presented here was sponsored by the Federal Highway Administration, U.S. Department of Transportation. The contents of this paper reflect the views of KLD Associates, Inc., who are responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official views or policy of the Department of Transportation.

REFERENCES

1. Manual on Uniform Traffic Control Devices for Streets and Highway. U.S. Department of Transportation, 1970.
2. C. G. Hammer and T. N. Tamburri. Evaluation of Minor Improvements—Part 1, Flashing Beacons. California Division of Highways, May 1967.
3. J. W. Bennet, P. D. Cribbins, and C. M. Walton. Evaluation of Traffic Control Improvements on Low Volume Highways. Highway Research Program, North Carolina State Univ., June 1969.
4. R. B. Goldblatt. Guidelines for Flashing Traffic Control Devices. KLD Associates, Inc.; Federal Highway Administration, Rept. FHWA-DOT-FH-11-8532, July 1976.
5. S. R. Byington and M. J. Rosenbaum. The Maine Facility. Public Roads, Vol. 37, No. 7, Dec. 1973.
6. H. J. Hoose. Stop Signs Using Actuated Flashers, Speed Actuated Flashers, Speed Actuated Curve Flashers, and Green Extension Units. City of Charlotte, N.C., 1973.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Effects of Signal Phasing and Length of Left-Turn Bay on Capacity

Carroll J. Messer and Daniel B. Fambro, Texas Transportation Institute, Texas A&M University

A periodic scan computer simulation program was developed to investigate the effects of signal phasing and length of left-turn bay on capacity. After the simulation program was tested, inputs (phase sequence, volume, cycle length, and length of left-turn lane) were varied to evaluate their interrelationships under a range of conditions. Additional analysis was conducted by using a modified Poisson approach. The results show that, for a left-turn bay, traffic delay increases and signal capacity decreases when traffic interactions and flow blockages occur between left-turning and through vehicles. High left-turn volumes and short bay storage lengths experience the most severe reduction in capacity. We developed mathematical relationships between reductions in left-turn capacity and geometric and traffic conditions and provide design guidelines to minimize capacity reductions. Judicious selection of signal phasing reduces the loss in capacity to some extent, although all phasings can experience large losses under some geometric conditions.

Field observations of rush-hour traffic flow at signalized intersections having a protected left-turn bay suggest that the capacity of left-turn phases can be reduced by vehicles that block the entry of other vehicles into the left-turn bay. The left-turn bay may be blocked during the red phase of the signal so that the bay cannot fill, or vehicles may even be blocked from entering on a portion of the left-turn green phase. As traffic blockages begin to occur, the left turners may also begin to impede through vehicles, and capacity problems and intersection congestion are compounded.

Reductions in left-turn capacity generally occur as average traffic demands increase beyond the storage length of the left-turn bay and the cycle length of the signal. Shorter left-turn bays and longer cycles are more susceptible to such reductions. A shorter left-turn bay means that fewer vehicles can be stored before a blockage occurs; a longer cycle requires more vehicles to be stored for a given volume level before a green.

Some signal phasing sequences that improve traffic flow and left-turn capacity have been implemented, but primarily by trial and error methods. Little information that describes improvements made by increasing the left-turn bay length or by changing phasing sequence is readily available.

Basic design criteria for the length of the left-turn bay have been previously related to the Poisson approach (1, pp. 688-690), but design trade-off relationships are not provided. Operational corrective treatments for an existing situation are also limited and not emphasized.

The mathematical analysis of the movement of through and left-turning vehicles at an intersection under various traffic conditions, design configurations, and signal phasing sequences is extremely complicated, which is probably the reason for the lack of pertinent design and operations information.

SIMULATION APPROACH

The periodic scan computer simulation approach was selected to investigate the left-turn capacity problem. The many variables and project time and budget constraints meant that this study could not be completely exhaustive and that some questions would undoubtedly remain unanswered. Answers were sought, however, to basic cause-and-effect relationships and trends among (a) capacity, (b) demand volume, (c) signal phasing, and (d) length of left-turn bay.

Traffic operations were simulated on only one intersection approach, which included a protected left-turn lane and an adjacent through lane. A schematic of the approach model is depicted in Figure 1. The junction of the left-turn and through lanes is the first single storage position upstream of the left-turn bay and can be varied in the simulation program. Arriving automobiles (trucks and buses each equal two automobiles) progress through the left-turn and adjacent through approaches by moving from one storage position to the next in discrete movements according to a defined strategy. These queue positions were defined to represent an average storage length of an automobile stopped on red.

QUEUE CHARACTERISTICS

Field studies were conducted in College Station, Texas, to determine average automobile storage spacing characteristics. Stations every 7.6 m (25 ft) were marked along the median of the divided approaches, and distances to the end of each queue and the number of automobiles in the queue to the recorded point were manually estimated for each cycle. Queue lengths up to 131 m (429 ft) long were measured. There were no significant grades on the approaches to the intersection and few trucks. These average storage lengths are presented in the following table (1 m = 3.3 ft).

Study Location	Left-Turn Lane (m/automobile)	Through Lane (m/automobile)
University Avenue at South College Avenue	7.3	7.7
Texas Avenue at University Drive	7.1	7.3

We used a slightly conservative value of 7.6 m/automobile (25 ft) in the simulation program (2, p. 432). Left-turn and through storage lengths were assumed to be the same.

Queue movement characteristics were also important inputs to the simulation model. An automobile approaching the end of a queue was assumed to stop instantaneously when it reached the last unoccupied storage position. The stopped automobile remained at that position until a specific time after the signal turned green. At this time, the automobile began to move immediately at a speed that would result in crossing the effective stop line at the front of the queue at the correct clearance time for the given automobile position in the queue.

Studies of queue movement and clearance characteristics were conducted at three busy intersections in College Station. The results are summarized in Figure 2. Also shown are the two following representative equations for describing the data:

$$T_f = 2.0 + 1.0N_p \tag{1}$$

and

$$T_c = 2.0 + 2.0N_p \tag{2}$$

where

- T_f = time after start of green for the automobile in queue storage position number N_p to begin moving forward (s);
- T_c = time after start of green for the automobile in queue storage position number N_p to clear the stop line on the approach (s); and
- N_p = queue storage position number (Figure 1) for either left-turning or through automobile.

These equations were selected specifically to expedite the simulation process. They are obviously descriptive of the measured characteristics but were not determined by a formal optimization process such as linear regression. The simulation process was greatly simplified by assuming that all the coefficients of the previous two equations had integer values.

SIMULATION INPUTS

The following variables are inputs to the intersection approach simulation program:

1. Total lane approach volume (automobile/h),
2. Percentage of total approach volume turning left,
3. Cycle length of signal (s),
4. Length of left-turn bay storage (automobiles),
5. Green time of left-turn signal (s),
6. Green time of through movement signal (s), and
7. Leading or lagging left turns (single or dual) shown in Figure 3.

SIMULATION MODEL

The following is a brief outline of the simulation model in statement format.

1. The left-turn and adjacent through lanes are divisible into discrete automobile length storage positions, as illustrated in Figure 1.
2. The length of the left-turn lane is defined by the first upstream single storage position or the junction.
3. The simulation scans the system every second in the periodic scan mode, updating from front to back all storage positions that should be changed. Operational measures of effectiveness are recorded.
4. Automobiles arrive according to the Poisson distribution and are put into the system at storage position 26.
5. Automobiles were not permitted to enter the system at headways less than 2 s.
6. Every input automobile is tagged as being a left turner or a through automobile in a random manner at the desired average rate of left turners.
7. Every storage position can have only one of three states: (a) empty, (b) moving (M), or (c) stopped or queued (Q).
8. Moving automobiles (M) can move forward only into an empty position.
9. Where possible, all M move forward into the next position every 1-s scan period.
10. When an M cannot move forward into the next position, its status and storage position are changed to a queued automobile (Q), and it is delayed 1 s.
11. When a Q occupies the position immediately behind another Q for the scan period being analyzed, the first Q remains queued and is delayed 1 s.
12. When the signal is red, position zero acts like a Q so that no one can leave position 1 and enter the intersection.

Figure 1. Simulation model of actual intersection approach.

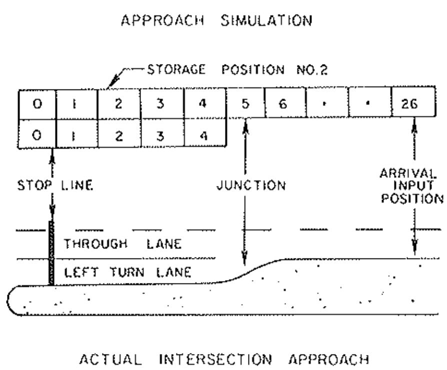


Figure 2. Movement characteristics of left-turning vehicles at busy intersections.

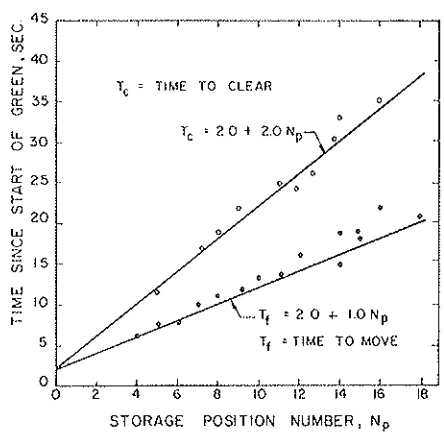


Figure 3. Signal phase sequences.

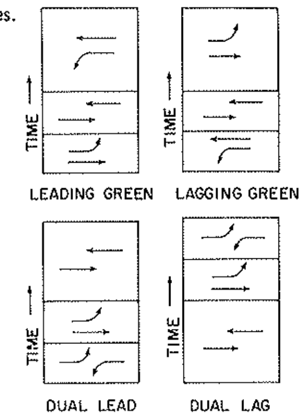


Figure 4. Queue movement in one lane.

Signal	Queue Storage Position, N_p									
0	1	2	3	4	5	6	7	8	9	10
R	.	.	.	.	.	.	.	.	.	.
R	.	.	.	.	.	.	.	.	.	M
R	.	.	.	.	.	.	.	.	M	.
R	.	.	.	.	.	.	.	M	.	M
R	.	.	.	.	.	.	M	.	M	.
R	.	.	.	.	.	M	.	M	.	M
R	.	.	.	.	M	.	M	.	M	.
R	.	.	.	M	.	M	.	M	.	M
R	.	.	M	.	M	.	M	.	M	.
R	.	M	.	M	.	M	.	M	.	.
R	M	.	M	.	M	.	M	.	.	.
R	O	M	.	M	.	M	.	.	.	.
R	Q	Q	M	.	M	.	.	.	.	.
R	Q	Q	Q	M	.	.	.	.	.	.
G	Q	Q	Q	Q	.	.	.	.	.	.
G	Q	Q	Q	Q	.	.	.	.	.	.
G	M	O	Q	Q	.	.	.	.	.	.
G	.	M	Q	Q	.	.	.	.	.	.
G	M	.	M	Q	.	.	.	.	.	.
G	.	M	.	M	.	.	.	.	.	.
G	M	.	M	.	.	.	.	.	.	.
G	.	M	.	.	.	.	.	.	.	.
G	M	.	.	.	.	.	.	.	.	.
G	.	.	.	.	.	.	.	.	.	.

Time

Periodic Scan
One Second Time Intervals

Direction of Movement

13. When the signal turns green, position zero is immediately set to the moving state. Two scanning periods later, the Q in position 1 is changed to M.

14. When a Q is behind an M or an empty space, its status is changed to an M, but it does not move forward until the next scan period. It is therefore delayed 1 s.

The execution of these queue behavior rules is illustrated in Figure 4. The movement and clearance times of the queues obey Equations 1 and 2, as required to simulate the actual traffic conditions.

15. Automobiles at the junction position can be either left turners or through automobiles. Left turners obey the status of the next lower position in the left-turn lane; through automobiles obey the status of the next lower position in the through lane. If a through automobile is queued in the junction position, then no left-turning automobile can enter the left-turn bay until the through automobile has cleared the junction, and vice versa. Through automobiles can block left turners and left turners can block throughs.

SIMULATION OUTPUTS

Several traffic flow measures of effectiveness are calculated by the simulation program. These are (a) output volume for each movement (automobile/h), (b) delay/automobile for each movement (s/automobile), and (c) frequency plots of queue length and individual delay/automobile.

PROGRAM TESTING

The computer program was written in a combination of FORTRAN IV and Assembly and was tested in two ways. First, computer printouts were made of the simulated movement of automobiles on the approaches as the signals changed from green to red over several cycles. Movements of individual automobiles were observed for realism and obedience to the simulation rules for movement, blockage, and stoppage. Second, unimpeded delays calculated from the simulation program were found to be consistent with the results obtained from Webster's theoretical delay equation. In addition, subsequent simulated delay calculations followed expected trends as queue interactions and blockages occurred.

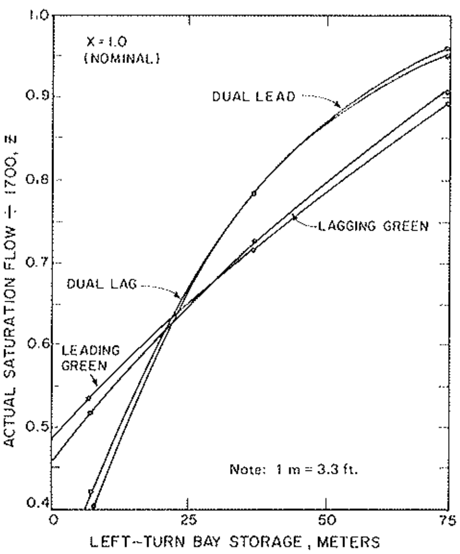
SIMULATION RESULTS

The results were most encouraging and revealed consistent trends and realistic outcomes. Many of the results were determined over 300 simulated cycles of operation for each data point. No fewer than 60 cycles were ever used. Five cycles were used to initialize the simulation model before we simulated the analysis cycles from which average values of the measures of effectiveness were calculated.

Table 1. Simulated average delay per vehicle movement.

Saturation Ratio	Left-Turn Demand (APH)	Through Demand (APH)	Left-Turn Bay Length (automobiles)	Left-Turn Delay (s/automobile)	Through Delay (s/automobile)
0.21	80	120	1	22	16
			5	22	16
			10	21	16
			20	21	16
0.42	160	240	1	39	26
			5	24	17
			10	24	17
			20	23	17
0.64	240	360	1	133	112
			5	39	29
			10	28	18
			20	28	18
0.85	320	480	1	121	106
			5	90	82
			10	56	45
			15	39	32
			20	35	30
0.95	360	540	1	137	117
			5	100	83
			10	94	57
			20	81	35

Figure 5. Reduction in left-turn saturation flow by phasing.

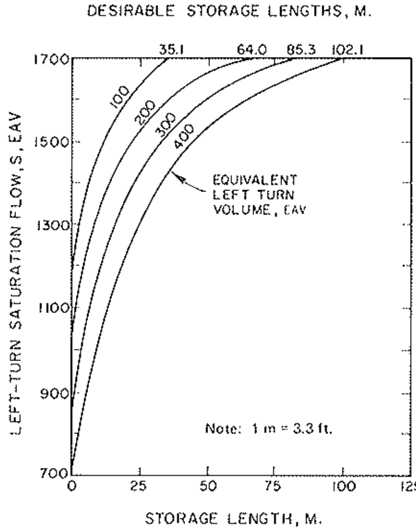


Delay

The initial analysis phase of the simulation study focused primarily on evaluating the effects of left-turn bay length and signal phasing on average automobile delay. Two signal phasing arrangements were studied: the leading left-turn and the lagging left-turn phase sequences. Cycle lengths of 60 and 80 s were studied. Approximately equal nominal volume-capacity (saturation) ratios were simulated for both left-turn and through movements. A nominal saturation ratio is defined as the normal demand on the movement divided by the phase's capacity when the left-turn bay is long enough to prevent blockages or interactions between the left turners and the throughs. In other words, the left-turn saturation flow is assumed to be 1700 automobiles/h of green (APHG), the nominal value for long bay lengths (3).

Simulation results of one of the delay studies is presented in Table 1. In this study green times were proportioned to yield uniform demand-capacity ratios for a 60-s cycle leading left. Delay increases with in-

Figure 6. Left-turn saturation flow and desirable storage lengths.



creasing volume, nominal saturation ratio, and cycle length. Delay also increases as the length of the left-turn bay shortens. Lagging green resulted in a slight reduction in delay for the conditions studied. Nominal saturation ratios of about 0.6 to 0.8 appear to be critical for bay lengths of 5 to 10 automobiles insofar as experiencing increased blockages and delay are concerned. These results indicate that the actual saturation ratio for the shorter bay lengths must have been considerably higher than the nominal value and that the saturation flow (and capacity) must have been correspondingly less than 1700 APHG.

Left-Turn Capacity

Left-turn capacity and saturation flow studies were conducted in view of the previous findings. Most of these subsequent simulation runs were made at nominal saturation ratios of about 1.0. During these capacity studies, two additional phase sequences of left turners first (dual lefts leading) and through movements first (dual lefts lagging) were added. Average results of these simulation studies are depicted in Figure 5. For the conditions evaluated, we observed some differences in saturation flow with lagging and leading left-turn green phasing slightly better for extremely short bay lengths; dual lefts leading or lagging performed better at bay lengths of 5 to 10 automobiles.

It is important to note, however, that all of the phasing arrangements experienced reductions in capacity for these conditions, a nominal saturation ratio of 1.0. A left-turn bay length of 5 automobiles experienced a 20 to 30 percent reduction in capacity. General reductions in capacity were observed in most of the 90 simulation runs, and greater reductions in capacity occurred at higher volumes. Similar reductions in capacity were experienced by the adjacent through lane. Reductions in capacity also varied with the percentage of traffic turning left and the green split between the two movements in an apparently complex manner. No overall mathematical model that included all the identified variables was developed.

To aid design and operations engineers in estimating a reasonable capacity and saturation flow for a given left-turn bay storage length, the combined simulation results of all 90 runs were pooled, from which the following multiple regression model (statistical R-squared value 0.80) was developed:

$$Z = 0.98 - 0.14 \times V - 0.19 \times X_L \times V + 0.24 \times X_T \times V \tag{3}$$

where

- Z = actual left-turn saturation flow divided by the nominal saturation flow ($Z = S/1700$);
- X_L = nominal left-turn saturation ratio;
- X_T = nominal through movement saturation ratio; and
- V = $X_L \times X_T \times K$, where K is the average number of left turners arriving for each cycle divided by the storage length of the bay.

This equation was used to develop the saturation flow and storage design curves shown in Figure 6. Inputs selected for design were X_L is 0.8; X_T is 0.8; nominal saturation flow is 1700 APHG; assumed storage requirement is 7.6 m/automobile; and cycle length is 75 s. The saturation flow (S) for left turns in Figure 6 was equal to 1700Z. Volumes are equivalent automobile volumes (EAV) in automobiles per hour.

At the top of Figure 6 are the left-turn bay storage lengths that will result in practically no reduction in capacity for the intercept left-turn volume level. These storage lengths can be used as practical design storage lengths. Interpolated storage lengths can be calculated for intermediate left-turn volumes. These storage lengths compared favorably as design values for 12 queue distributions of automobile storage available from the simulation runs. Computer plotting costs allowed only 12 plots of queue distributions.

A special set of simulation runs was made to test and illustrate the capacity results of Figure 6. An intersection was assumed to have a left-turn bay of 7.6 m and a leading left-turn signal phasing sequence. It was also assumed that the left-turn volume was 320 EAV and that the through movement volume was 480 EAV. Corresponding (effective) green times were 14 and 20 s. Nominal saturation ratios of about 0.8 existed on both movements. According to Figure 6, however, the 7.6-m bay length combined with a 320 left-turn volume should result in a large reduction in left-turn capacity and saturation flow from 1700 APHG to an actual flow of about 1060 APHG. If this reduction in capacity does exist, then the given conditions are overloaded and large delays should result. The actual saturation ratios, X, would be about 1.30 on both movements.

Table 2 illustrates the consequences of the short bay and reduced capacity. The first row of Table 2 contains initially given conditions and results. Low flows and excessively long delays occurred. As movement green times are lengthened, flows climb to the volume levels being simulated, while delays drop to acceptable levels. In order to compensate for the 38 percent reduction in saturation flow estimated from Figure 6 (640/1700 = 0.38) and to provide actual saturation ratios of about 0.8, large increases in green are required. Green times of 22 s for the left turners and 32 s for the throughs provide the needed 57 to 60 percent increase. It would appear for this one extreme example that the reduction in capacity is slightly larger than estimated by Figure 6, although delay variations are very sensitive in the region being analyzed. However, the general trend and practical magnitude of expected left-turn saturation flows given in Figure 6 are supported.

LEFT-TURN BAY LENGTH—MODIFIED POISSON APPROACH

The previous simulation studies of the capacity and desirable length of left-turn bays were an outgrowth from an earlier project analysis that utilized a simpler approach, which was an extension of the Poisson procedure frequently used by traffic engineers. The Poisson approach forms the basis for storage length recommendation given by the American Association of State Highway Officials' red book (1), which states that

> At signalized intersections, the required storage length depends on the cycle length, the signal phasing arrangement, and rate of arrivals and departures of left-turning vehicles. The storage length should be based on 1.5 to 2 times the average number of vehicles that would store per cycle, predicated on the design volume.

The modified Poisson approach we shall present subsequently provides guidance in determining the relationship between the multiplier (1.5 to 2 times) and the design left-turn volumes. In addition, these results will support the previously recommended storage bay lengths given in Figure 6. Other important interrelationships will be presented between design and operational variables.

In the following equation, which we adapted with some minor changes in notation from Miller (4), we estimate the average number of automobiles remaining in the queue at a pretimed signal at the end of the green phase:

$$A = \exp^{-1.3}[(1 - X/X)(qC/X)^{1/2}]/2(1 - X) \tag{4}$$

where

- A = average number of automobiles in the left-turn bay at end of green;
- q = left-turn flow rate (automobile/s);
- C = cycle length (s);
- X = left-turn saturation ratio (qC/gs);
- g = left-turn effective green (s); and
- s = left-turn saturation flow (automobile/s·green).

The number of left-turning automobiles, in addition to A, arriving during the effective red that must be stored in the left-turn bay is

$$B = q \times R \tag{5}$$

where

- B = number of left-turning automobiles arriving on red;
- q = left-turn flow rate (automobile/s); and
- R = left-turn effective red time (s).

After the left-turn signal turns green, additional left-turning automobiles are joining the rear of the stopped left-turn queue for a time (T) until it is time for the automobile in queue position N_p to begin moving forward (see Equation 1). If T_f is set equal to the arrival time of automobile N_p after the start of green, then

$$T_f = 2 + 1 \times N_p = (N_p - A - B + 2 \times q)/q \tag{6}$$

and

$$N_p = (A + B)/(1 - q) \tag{7}$$

The left-turn flow rate (q) should be higher than the average left-turn flow rate to account for the short-term peak flows that occur cycle by cycle during random (Poisson) flow. The flow rate was selected so that the average number of cycle failures during the peak 15-min period of the design hour would equal 0.50. That is

$$\Sigma P_q \times 3600/C \times 1/4 = 0.50 \tag{8}$$

where ΣP_q is the cumulative Poission probability of exceeding flow rate q (5), and C is cycle length (s). Letting the design storage capacity of the bay be N_p, which in turn is calculated from q, then the above probability of overflow criterion can be expressed in design level of performance terms as follows: The odds are 50/50 that the left-turn storage demands will exceed capacity only once during a peak 15-min period of the design hour. Table 3 summarizes input values used to develop modified Poisson left-turn bay storage requirements from Equation 7.

Results of this approach are presented in Figures 7, 8, and 9. Figure 7 shows that the length of storage required increases with left-turn volume and with the signal phase's saturation ratio (X). This latter fact is important for several reasons. The normal Poisson approach to left-turn bay storage design (1) does not account for the signal's operating saturation ratio. If the saturation ratio exceeds 0.85, the length of storage needed to reduce the likelihood of interaction and blockage increases dramatically. As was shown in the earlier section on simulation of left turns, blockages cause a reduction in saturation flow. A maximum saturation ratio of 0.8 seems practical for use in design, although 0.85 would be more conservative.

Figure 8 presents the length of storage required as a function of cycle length and left-turn volume for the assumed design saturation ratio of 0.8. The storage length increases with increasing cycle length, but the rate of increase is only about 40 percent as large as suggested by the normal Poisson approach. This is explained by the fact that while longer cycle lengths require more automobiles to be stored per cycle, there are fewer cycles that have the opportunity to "fail" during the peak 15-min period of the design hour. This reduction is not accounted for in the normal Poisson approach.

Figure 9 presents comparative results between the design guidelines (1) previously noted and results obtained from the modified Poisson approach using a saturation ratio of 0.8 and a cycle length of 75 s. The variable m in Figure 9 is the normal Poisson parameter, i.e., average number of left turns per cycle. The guidelines of 1.5 to 2 m bound the modified Poisson curve up to left-turn volumes of 350 automobiles/h. The left-turn bay length required in Figure 9 is within 10 percent of those storage lengths, shown at the top of Figure 6, that were developed from the simulation analyses. In general, cycle lengths in excess of 80 s in Figure 8 result in slightly longer storage requirements than those given in Figure 6.

SUMMARY AND RECOMMENDATIONS

The results of this study show that traffic delay increases and signal capacity decreases for a left-turn bay when traffic interactions and flow blockages occur between left-turning and through automobiles. High left-turn volumes and short bay storage lengths experience the most severe reductions in capacity. Delay begins to occur when the signal saturation ratio reaches 0.6 to 0.8 for bay storage lengths of 5 to 10 automobiles, respectively.

The operational quality of service provided by a left-turn bay design was shown to depend to a significant degree on how well the traffic engineer signalized the intersection. A design can fail simply because the signal saturation ratio approaches 1.0. In addition, the signal phase sequence was found to affect operational performance. The leading and lagging phase sequences performed slightly better for short bay lengths, and the dual lead and dual lag sequences were superior for bay storage lengths over 22.9 m (75 ft). However, all four signal phase sequences experienced considerable reductions in capacity at high saturation ratios and short bay lengths.

The following left-turn bay storage design recommendations are offered on the basis of supporting results from two different study approaches. However, if a higher saturation ratio, 0.85 for example, is anticipated at an intersection, Figure 7 could be used to scale recommended distances up to this higher level. On the basis of the study results using a saturation ratio of 0.8, the length of storage for automobiles in left-turn bays at signalized intersections should not be less than the recommended values shown below (1 m = 3.3 ft).

Left-Turn Volume (equivalent automobiles/h)	Storage Length (m)
100	35.1
200	64.0
300	85.3
400	102.1

Automobile storage is assumed to be 7.6 m/automobile and does not include any distance provided in advance of the stop line or within the transition section into the bay. Truck and bus volumes should be converted into equivalent automobile volumes at a rate of two automobiles per truck or bus. Figures 6 and 8 or both may also be used to determine bay storage requirements.

The phase timing of left-turn signals at pretimed signalized intersections should account for the reduction in saturation flow that may occur during rush-hour traffic conditions as illustrated in Figure 6.

Table 2. Simulation of reduced saturation flow effects.

Green (s)		Flow (EAV)		Green Increase (%)		Delay (s/automobile)	
Left	Through	Left[a]	Through[b]	Left	Through	Left	Through
14	20	220	320	0	0	121	107
18	26	265	402	28	30	87	74
22	30	315	459	57	50	60	51
22	32	316	467	57	60	53	44
24	32	319	481	71	60	34	27
26	32	317	480	85	70	23	18

[a] Left-turn simulated volume = 320 automobiles/h (EAV).
[b] Through simulated volume = 480 automobiles/h (EAV).

Table 3. Input values of left-turn bay storage requirements for modified Poisson approach.

Cycle Length (s)	Cumulative Poisson Probability	Input Left-Turn Volume by Peak 15-Min Volumes					
		50 EAV	100 EAV	150 EAV	200 EAV	300 EAV	400 EAV
60	0.033	132	234	312	396	540	672
70	0.039	129	216	295	365	509	643
80	0.044	122	207	279	347	486	617
90	0.050	116	200	268	336	468	596
100	0.055	108	190	259	324	454	576

Figure 7. Left-turn bay storage versus saturation ratio.

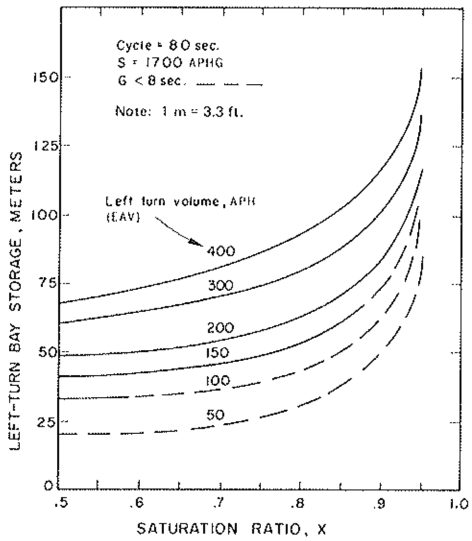
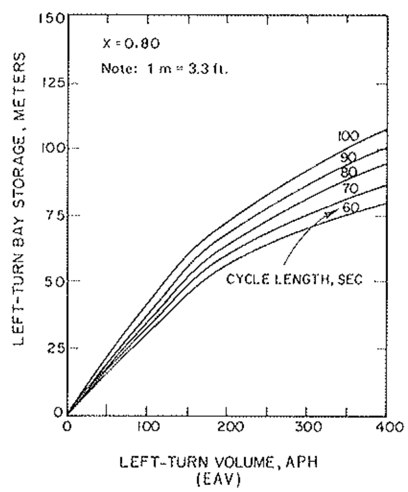


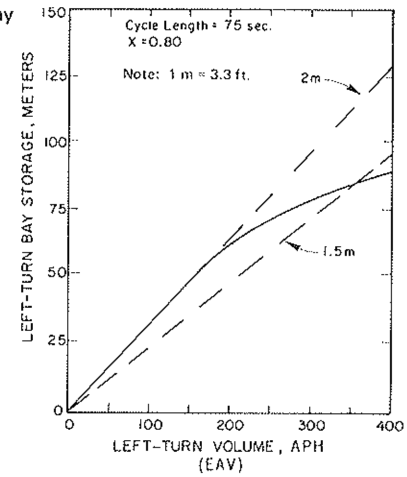
Figure 8. Left-turn bay storage versus turning volumes for various cycle lengths.



ACKNOWLEDGMENTS

We wish to express our appreciation to Don Ader of the Texas Transportation Institute for developing the computer simulation program described in this paper. A special debt of gratitude is also owed to Herman Haenel, Harold Cooner, and Elmer Koeppe of the Texas State Department of Highways and Public Transportation for

Figure 9. Left-turn bay storage versus turning volume.



their timely and constructive inputs to the overall research effort. This paper discusses one phase of a research project conducted by the Texas Transportation Institute and the Texas State Department of Highways and Public Transportation in cooperation with the Federal Highway Administration. The contents of this paper reflect our views, and we are responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration. This paper does not constitute a standard, specification, or regulation.

REFERENCES

1. A Policy on Design of Urban Highways and Arterial Streets. American Association of State Highway Officials, Washington, D.C., 1973.
2. A Policy on Geometric Design of Rural Highways—1965. American Association of State Highway Officials, Washington, D.C., 1965.
3. D. J. Berry. Capacity and Quality of Service of Arterial Street Intersections. Texas Transportation Institute, Research Rept. 30-1, Aug., 1974.
4. A. J. Miller. The Capacity of Signalized Intersections in Australia. Australian Road Research Board, Bulletin 3, Mar. 1968.
5. D. R. Drew. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Optimizing Settings for Pedestrian-Actuated Signal Control Systems

Feng-Bor Lin, Department of Civil and Environmental Engineering, Clarkson College of Technology, Potsdam, New York

A pedestrian-actuated traffic signal control system has been implemented at intersections where most traffic conflicts are between vehicle flows and pedestrians. When properly implemented, such a system can process traffic flows more efficiently than a pretimed control system. It is not yet clear, however, how a pedestrian-actuated system can be best utilized, and misuse is not uncommon. This study provides a basis for proper use of a pedestrian-actuated system by characterizing the performance of the system at isolated intersections in terms of traffic delay. Three optimization problems of signal setting—the minimization of (a) average vehicle delay, (b) total delay of pedestrians and vehicle riders, and (c) the difference between average pedestrian delay and average vehicle delay—are discussed.

The operation of a pedestrian-actuated control system is fairly straightforward (Figure 1). When the first pedestrian arrives at an intersection at time t_1 and pushes a button to change the signal, it takes the control a response time of T s to switch the signal from red to green for the pedestrian. Response time can be considered as equal to vehicle amber time. Subsequent arrivals of pedestrians between time t_1 and time t_2 have no influence on the control.

The pedestrian green, totaling G_p s, consists of a WALK duration of E s and a DON'T WALK duration of S s. After the end (t_4) of the pedestrian green, the signal changes to vehicle green for a minimum of M s.

Amber time is treated as green time in this study and is included in minimum vehicle green duration M. If there are pedestrian arrivals between the start of DON'T WALK (t_3) and the end of the minimum green (t_5), the next pedestrian green would begin at time t_5. Otherwise, the first pedestrian arriving after t_5, for instance at t_6, would induce another pedestrian green at time t_7 and cause an overall vehicle green duration of G_v s. Thus, vehicle green duration is a function of pedestrian arrivals.

A pedestrian-actuated control system can be evaluated by measures of performance, perhaps the most pertinent of which is traffic delay. For pedestrians delay is a function of pedestrian flow patterns and signal settings, while for vehicles it is a function of vehicle flow patterns, signal settings, and the sequence of pedestrian green intervals.

Assuming that traffic arrival is random, I derived the following formulas for estimating traffic delays. These formulas are then applied to three signal setting optimization problems: (a) minimizing average vehicle delay, (b) minimizing total delay of pedestrians and vehicle riders, and (c) minimizing the difference between average pedestrian and average vehicle delays.

SYSTEM PERFORMANCE

Frequency of Pedestrian Green Interval

The performance of a pedestrian-actuated signal control system is governed largely by the frequency of pedestrian green intervals. For randomly arriving pedestrians the frequency measured in terms of the number of pedestrian green intervals per hour can be determined analytically as

$$N = 3600/\left\{G_p + M + h_p \exp[-(S + M - T)\lambda_p]\right\} \quad (1)$$

where

- N = the number of pedestrian green intervals per hour;
- G_p = pedestrian green duration (s);
- M = minimum vehicle green duration (s);
- T = pedestrian signal response time (s);
- h_p = pedestrian headway (s);
- S = pedestrian DON'T WALK duration (s); and
- λ_p = pedestrian flow rate (persons/s).

In Equation 1 I also assumed that pedestrians arriving in DON'T WALK durations would wait for the next green interval. This equation indicates that as λ_p increases N approaches a fixed value of $3600/(G_p + M)$, at which an actuated signal pattern becomes the same as a pretimed pattern with a cycle length equal to $G_p + M$.

Average Pedestrian Delay

Average pedestrian delay with random arrivals can be determined from

$$D_p = N\{T(3600/N - G_p - M) + (S + M)^2/2\}3600 \quad (2)$$

where D_p is the average pedestrian delay in seconds per pedestrian. Delay suffered by a pedestrian is measured from the time of arrival at an intersection to the start of the first pedestrian green.

Results of a recent simulation analysis of pedestrian-actuated signal control systems (1) reveal that the 80th percentile pedestrian delay is approximately twice as long as the average delay for a given combination of flow and signal settings.

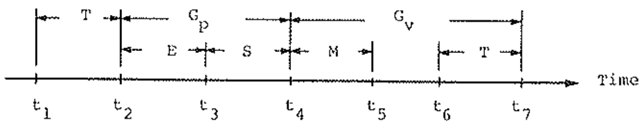
Compared to a pretimed signal system with a pedestrian DON'T WALK duration of S s, a pedestrian green of G_p s, and a vehicle green plus amber duration of M s, average delay has an expected value of $(S + M)^2/2(G_p + M)$. Therefore, it can be concluded from Equations 1 and 2 that the average delay incurred by a pedestrian-actuated signal is always less than that expected from a pretimed signal.

Average Vehicle Delay

Vehicles are delayed by signal controls because queues form at red lights. To relate delays to traffic flows and signal settings, let

- C = 3600/N, average cycle length (s);
- G_e = $C - G_p - 3.7$, average vehicle effective green duration (s);

Figure 1. Timing of pedestrian-actuated signal control.



Q_v = single lane vehicle flow (vehicles/h);
Q_s = vehicle saturation flow (vehicles/h), taken as 1800 vehicles/h for automobiles;
x = G_e/C, proportion of the average cycle length that is effectively green;
y = Q_vC/G_eQ_s, saturation rate with respect to Q_s; and
D_v = average vehicle delay (s).

Based on the same simulation analysis I found that with the above definitions Webster's formula (2)

$$D_v = [C(1-x)^2]/[2(1-xy)] + 3600y^2/[2Q_v(1-y)] - 0.65[C/(Q_v/3600)^2]^{1/3}\, y^{(2+5x)} \quad (3)$$

can provide reasonable estimates of average vehicle delays whether signal controls are pretimed or pedestrian actuated. It can be concluded from Equation 3, however, that a pedestrian-actuated control system will always result in fewer delays than a pretimed system. As in the case of pedestrian delays, the 80th percentile vehicle delay is approximately equal to $2D_v$.

Equation 3 should be used with discretion, because it does not reveal the actual performance of a control system under all circumstances. This requires further explanation regarding the interaction between vehicle flows and the vehicle-processing capacity of an intersection. This capacity, measured in terms of the maximum number of vehicles that can pass through the intersection in an hour, is a function of signal settings and characteristics of vehicle discharging headways at the intersection. Greenshields' distribution of discharging headways (3) is shown in the following table.

Position of Vehicle in Queue	Discharging Headway (s)	Position of Vehicle in Queue	Discharging Headway (s)
First	3.8	Fourth	2.4
Second	3.1	Fifth	2.2
Third	2.7	Sixth and up	2.1

Based on those headways, such a capacity (Q_{max}) can be estimated as

$$Q_{max} = 5N - (3600 - NG_p - 14.2N)/2.1 \quad (4)$$

where 14.2 represents the total time in seconds needed for the first five queuing vehicles to enter the intersection.

The average saturation rate (y_a) for an average cycle can be approximated by

$$y_a = Q_v/Q_{max} \quad (5)$$

From the simulation analysis mentioned earlier, I found that when y_a has a value greater than about 0.8 average delay has a tendency to become a function not only of flow rate but also of the sequence of arriving headways. The performance of a control system may therefore be unstable. In other words, for a given flow rate, estimates of average delays based on different field samples either have a large variation or may become time dependent (increase with respect to time). Then Equation 3 may not provide a good estimate. Whenever possible, signal settings should be chosen in such a way as to avoid a y_a value greater than 0.8.

The system performance represented by Equations 1, 2, and 3 is characterized by several features that dictate the requirements of optimum signal settings. First, as depicted in Figure 2, average vehicle delay for a given G_p decreases at a decreasing rate with respect to M, while average pedestrian delay increases at a more or less constant rate with respect to M. Second, for a given M, average vehicle delay increases while average pedestrian delay decreases with respect to G_p. And, finally, when M is smaller than G_p, average vehicle delay increases rapidly when M is reduced.

Validity of the Assumption of Random Arrivals

To test the validity of the assumption that traffic arrives randomly, data related to pedestrian and vehicle flow patterns were collected at several locations in Potsdam, New York (see Figures 3 and 4).

Figure 3a depicts a vehicle arrival pattern identified at a distance 42.7 m (140 ft) upstream from an intersection. The rate of arrival is light and affected only slightly by an upstream intersection controlled by a pedestrian-actuated signal. Figure 3b presents a vehicle arrival pattern of moderate flow rate at a location 76.2 m (250 ft) upstream from an intersection. This arrival pattern is not noticeably affected by a remote intersection upstream. On the other hand, the vehicle arrival patterns shown in Figures 3c and 3d respectively were observed 213.4 m (700 ft) downstream from an intersection controlled by a semiactuated vehicle signal. The flow rates at the time of data collection were moderately heavy, and interactions between vehicles were substantial because of the build-up of queues downstream.

Regardless of the dissimilarities in flow conditions, chi-square tests show that all the four arrival patterns conform to Poisson distributions at the 5 percent significance level, as does the pedestrian arrival pattern shown in Figure 4. The relatively high chi-square value of 8.14 compared to the critical value of 9.48 may be attributable to the grouping of pedestrians when friends or relatives arrive together.

From the above observations, traffic arrival patterns at an isolated intersection can be reasonably assumed to be random. When they are not random, such as when a heavy vehicle flow or a forced flow prevails, Equations 1, 2, and 3, because they demand minimum data input in the form of flow rates, are still good for finding optimum signal settings.

OPTIMUM SETTINGS

Optimum settings for a traffic signal control system will be dictated by the purpose of the control, which may vary from one locale to another. Nevertheless, the following are probably the most important purposes:

1. Minimizing average vehicle delay,
2. Minimizing the differences between average delay of vehicle riders and that of pedestrians, and
3. Minimizing total rider and pedestrian delays.

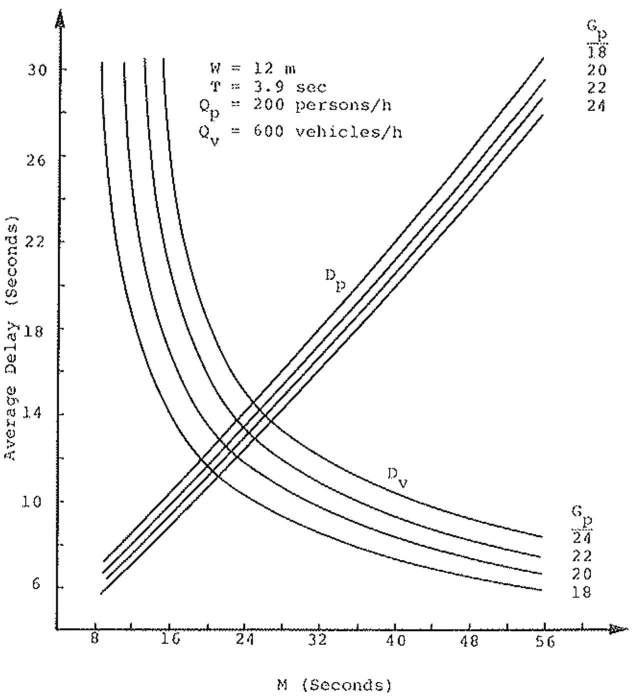
The signal controls based on these will be referred to as vehicle priority operation, equity operation, and minimum total delay operation respectively.

Vehicle Priority Operation

The rationale behind minimizing average vehicle delay hinges on the fact that they waste time, increase vehicle operating costs, and pollute the air. Therefore, as long as a control does not bring about excessive pedestrian delays, average vehicle delay may be minimized.

Several factors should be taken into account. First, in setting minimum vehicle green, M should be long enough to allow those vehicles stopped by a pedestrian green to cross the intersection during their green. For a pedestrian green of G_p s and a vehicle flow of λ_v vehicles/s, average number of queuing vehicles per pedestrian green is approximately $G_p\lambda_v$. The green time required for these vehicles to enter an intersection can be estimated from Greenshields' data on vehicle discharging headways, which on the average decrease from a high of 3.8 s for only one queuing vehicle to a stabilized value of about 2.1 s for a queue of considerable length. As a conservative measure, an average of 4 s/vehicle is allowed in the following analysis. Thus, green time required for $G_p\lambda_v$ vehicles to enter an intersection is considered as $4(G_p\lambda_v)$. An amber duration equal to pedestrian signal response time should be added to $4(G_p\lambda_v)$ to form a lower bound of minimum M.

Figure 2. Average delay as a function of G_p and M.



It is not clear what the maximum delay most pedestrians are willing to tolerate is. To maintain the spirit of a pedestrian-actuated control system, however, an upper bound should be imposed on M. This upper bound is denoted as M_{max} and is limited to 60 s.

In addition, there are several concerns in setting a pedestrian green duration G_p. For WALK duration, the Manual on Uniform Traffic Control Devices (4) set a minimum of 7 s to allow waiting pedestrians to enter an intersection. Abrams and Smith (5) have recently indicated that, except when pedestrian queues are longer than 15 people (an unlikely event for intersections controlled by a pedestrian-actuated system), 7 s is sufficient. This should be considered a lower bound because too short a WALK duration tends to create confusion and stress on pedestrians.

On the other hand, DON'T WALK duration should permit pedestrians to clear an intersection before vehicles are released. Thus, if W denotes the maximum width of an intersection and V_p is a design speed for pedestrians, then minimum DON'T WALK duration should be at least equal to W/V_p. In consequence, the minimum requirement of G_p is $7 + (W/V_p)$. When a G_p longer than the minimum is preferred, it would be desirable to set

Figure 3. Vehicle arrival patterns.

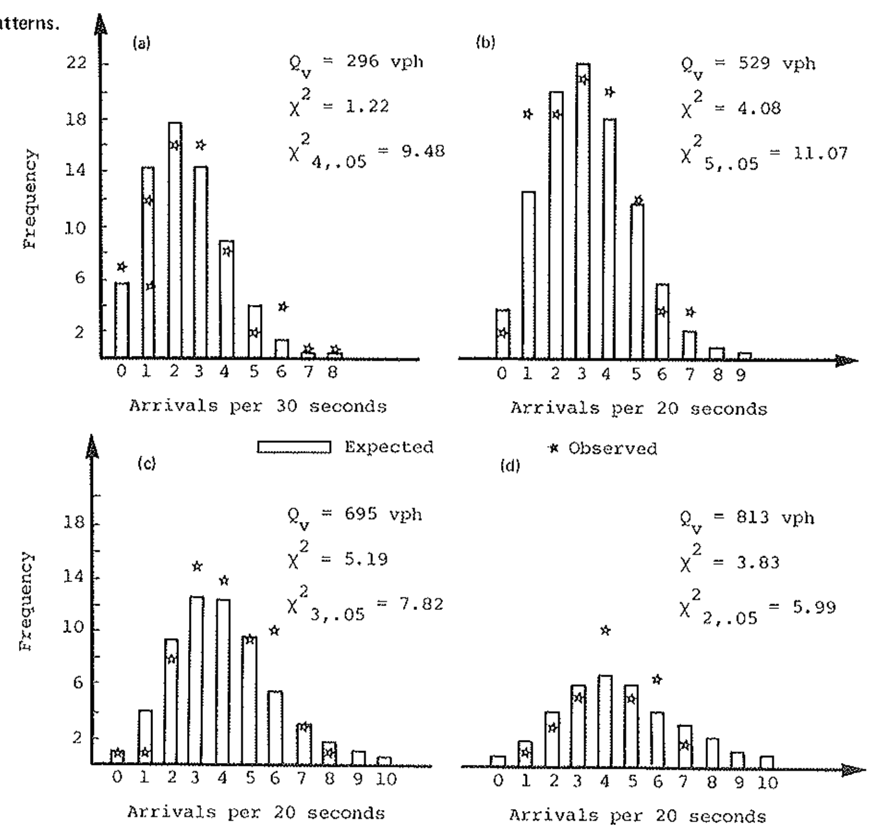
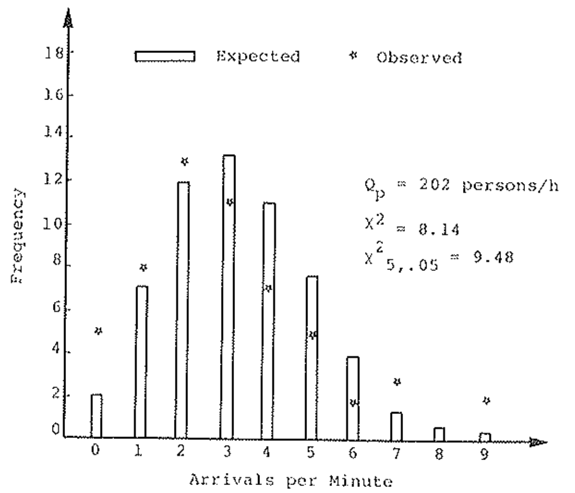


Figure 4. Pedestrian arrival pattern.



the DON'T WALK duration to its minimum of W/V_p and to allocate the rest of G_p to WALK.

Finally, pedestrian signal response time T, which may be regarded as vehicle amber time, should allow vehicles to clear an intersection before the green is turned over to pedestrians. According to the Traffic Engineering Handbook (6), T can be estimated from

$$T = r + V/(2a) + (W + L)/V \tag{6}$$

where

- r = perception-reaction time (1 s);
- V = approach speed of clearing vehicle;
- a = deceleration rate of clearing vehicle [4.57 m/s^2 (15 ft/s^2)];
- W = intersection width; and
- L = length of vehicle [6.1 m (20 ft)].

The problem of determining optimum settings for vehicle priority operation can be formulated as Min D_{vm}, which is subject to $4(G_p\lambda_v) + T \leq M \leq M_{max}$, $G_p = E + S$, $S = W/V_p$, and $E \geq 7$, where D_{vm} represents the average delay of the heaviest single-lane vehicle traffic flow that pedestrians interfere with. Its value is given by Equation 3.

The solution to this optimization problem is straightforward, because, as shown in Figure 2, average vehicle delay increases and decreases with respect to G_p and M. Therefore, the best settings for any combination of pedestrian and vehicle flows call for the use of a minimum acceptable G_p and a maximum allowable M: $E = 7$, $G_p = 7 + (W/V_p)$, and $M = M_{max}$.

Table 1 shows an example of vehicle priority operation based on $V_p = 1.1$ m/s (3.5 ft/s), V = 40 km/h (25 mph), and $M_{max} = 60$ s. The operations are characterized by the following features:

1. Average pedestrian delay (D_p) is overwhelmingly greater than the average vehicle delay (D_v);
2. Average pedestrian delay of about 30 s for all combinations of intersection widths and traffic flows implies an 80th percentile delay of about 60 s; and
3. Values of y_a for the ranges of intersection widths and traffic flows considered are well below 0.8. Thus, the signal settings would result in stable system performances with respect to vehicle delay, meaning that average delay is unlikely to grow with respect to time.

If the average pedestrian delay and the corresponding 80th percentile delays are excessive, then a smaller value of M_{max} should be used. For example, with an M_{max} of 40 s, the average pedestrian delay shown in Table 1 could be reduced by about 10 s, but average vehicle delay would not be increased by more than 4 s. The corresponding 80th percentile delay would have a more acceptable value of about 40 s. Use of a smaller M_{max}, however, has the danger of leading to an unstable system performance with respect to vehicle delays when both pedestrian and vehicle flows are heavy.

Equity Operation

If a control system is intended to permit fair use of an intersection by pedestrians and motorists, then signals can be set to minimize the discrepancy between average pedestrian delay and average vehicle delay. For vehicle flows, the needs of the heaviest traffic flow stream are the most important concern. Lighter streams affected by pedestrians may be ignored.

The problem of finding optimum settings for equity operations can be formulated as Min $(D_p - D_v)$, which is subject to $4(G_p\lambda_v) + T \leq M \leq M_{max}$, $G_p = E + S$, $S = W/V_p$, and $E = 7$, where $(D_p - D_v)$ denotes the absolute value of the difference between D_p and D_v.

In this case, G_p has to be set at its minimum requirement to ensure that a unique solution exists and that combined delay of pedestrians and vehicles be kept as low as possible. This constraint is essential because, as can be seen from Figure 2, for any G_p it is always possible to find an M such that D_v equals D_p. In the figure these combinations of G_p and M at which D_p equals D_v are represented by the intersects of D_p and D_v curves. It is clear that larger G_p induces longer delays.

The approximate solutions to the above optimization problem with different combinations of intersection widths and traffic flows are shown in Table 2. The following observations can be made by comparing the equity operations with the vehicle priority operations under the same conditions.

1. Equity operations bring about substantial reduction in average pedestrian delays. They result in less total delay at light vehicle flows and greater total delay at heavy vehicle flows.
2. As indicated by the larger values of y_a, the equity operations are relatively unfavorable to vehicle flows. Values of y_a greater than 0.8 at heavy vehicle flows suggest that an equity operation here is undesirable.
3. For a given combination of intersection width and vehicle flow, the optimum settings of M are not sensitive to pedestrian flow rate.
4. For a given combination of Q_p and Q_v, the optimum settings of M increase by no more than 2.5 s for every additional 3.1 m (10 ft) of intersection width.
5. The 80th percentile delays range about 7 to 40 s for the various combinations of intersection widths and traffic flows. Therefore, from the viewpoint of pedestrians, equity operations are preferred to vehicle priority operations.

Minimum Total Delay Operation

When travel time is the dominant concern in the operation of a signal control system, it would be desirable to minimize total pedestrian and vehicle delay. Assume that vehicle occupancy rate is 1.5 riders/vehicle; the best settings can be identified by finding the solution to the problem

$$\text{Min}\left(D_pQ_p + 1.5\sum_{i=1}^{n} Q_{vi}D_{vi}\right) \quad (7)$$

where n denotes the total number of vehicle traffic flow streams in conflict with pedestrians, and Q_{vi} and D_{vi} represent the flow rate and the average delay of the ith vehicle stream respectively. D_p is given in Equation 2, and D_{vi} can be determined from Equation 3.

The objective function given above should be subjected to the same constraints as those described in the case of vehicle priority operation.

A complete treatment of this optimization problem is impossible because optimum settings are partially dictated by the number of vehicle traffic streams that have to be taken into account. Table 3 provides an insight into the nature of this type of operation and shows the best settings and their implications for only one vehicle traffic stream. The operations have the following characteristics:

1. G_p should be set at its minimum requirement;
2. Optimum settings of M increase with respect to intersection width and vehicle flow rate for a given pedestrian flow rate;
3. Values of y_a are all under 0.8 for the ranges of intersection width and traffic flows considered, and stable performance of the control systems with respect to vehicle delays can be expected; and
4. Optimum settings are comparable to those of the vehicle priority operations at heavy vehicle flows. And, at light vehicle and pedestrian flows, they are similar

Table 1. Optimum settings for vehicle priority operations.

			W = 6 m, G_p = 13 s, T = 3.3 s			W = 9 m, G_p = 16 s, T = 3.6 s			W = 12 m, G_p = 19 s, T = 3.9 s			W = 15 m, G_p = 22 s, T = 4.1 s		
Q_p	Q_v	M (s)	D_p (s)	D_v (s)	Y_a	D_p (s)	D_v (s)	Y_a	D_p (s)	D_v (s)	Y_a	D_p (s)	D_v (s)	Y_a
100	200	60	27.5	2.2	0.14	30.0	2.9	0.15	30.4	3.7	0.15	30.8	4.5	0.16
	400			2.7	0.28		3.6	0.29		4.5	0.30		5.5	0.33
	600			3.5	0.42		4.5	0.44		5.6	0.46		6.8	0.50
	800			4.5	0.56		5.8	0.59		7.1	0.61		8.6	0.66
200	200	60	29.4	2.3	0.14	30.8	3.1	0.15	32.1	3.9	0.16	33.4	4.8	0.16
	400			3.0	0.29		3.8	0.30		4.8	0.31		5.8	0.34
	600			3.8	0.43		4.8	0.45		6.0	0.47		7.2	0.50
	800			4.9	0.57		6.2	0.60		7.6	0.62		9.2	0.68
400	200	60	29.6	2.4	0.14	30.9	3.1	0.15	32.3	3.9	0.16	33.5	4.8	0.16
	400			3.0	0.29		3.8	0.30		4.8	0.31		5.8	0.34
	600			3.8	0.43		4.8	0.45		6.0	0.47		7.2	0.50
	800			4.9	0.57		6.2	0.60		7.6	0.62		9.2	0.68

Note: 1 m = 3.3 ft.

Table 2. Optimum settings for equity operations.

		W = 6 m, G_p = 13 s, T = 3.3 s				W = 9 m, G_p = 16 s, T = 3.6 s				W = 12 m, G_p = 19 s, T = 3.9 s				W = 15 m, G_p = 22 s, T = 4.1 s			
Q_p	Q_v	M (s)	D_p (s)	D_v (s)	Y_a	M (s)	D_p (s)	D_v (s)	Y_a	M (s)	D_p (s)	D_v (s)	Y_a	M (s)	D_p (s)	D_v (s)	Y_a
100	200	7.0	3.6	3.6	0.17	9.0	4.9	4.8	0.18	10.0	6.2	6.1	0.20	11.0	7.6	7.5	0.22
	400	10.5	4.4	4.4	0.34	12.0	5.8	5.8	0.36	13.5	7.4	7.3	0.39	14.5	9.0	8.9	0.42
	600	14.5	5.6	5.5	0.50	16.0	7.2	7.1	0.54	17.5	9.0	8.9	0.57	19.0	10.9	10.7	0.61
	800	18.5	7.0	7.0	0.66	20.5	9.0	9.0	0.70	22.5	11.1	11.3	0.74	25.0	13.5	13.8	0.81
200	200	11.5	5.4	5.4	0.23	12.5	7.0	6.9	0.25	13.0	8.5	8.6	0.28	14.0	10.3	10.1	0.30
	400	14.5	6.6	6.5	0.43	15.5	8.3	8.2	0.47	16.5	10.2	10.1	0.51	17.5	12.0	11.9	0.55
	600	17.5	7.9	8.0	0.62	19.0	10.0	10.1	0.66	20.5	12.1	12.3	0.70	22.5	14.5	14.3	0.73
	800	22.5	10.3	10.2	0.76	25.0	13.0	12.9	0.80	27.5	15.6	15.7	0.83	30.0	18.3	18.6	0.85
400	200	13.0	6.6	6.6	0.28	13.5	8.1	8.2	0.31	14.0	9.7	9.8	0.34	14.5	11.2	11.5	0.36
	400	15.5	7.7	7.9	0.51	16.5	9.5	9.6	0.54	17.5	11.4	11.4	0.57	18.5	13.2	13.1	0.60
	600	19.0	9.4	9.4	0.69	20.5	11.5	11.5	0.71	22.0	13.5	13.6	0.74	23.5	15.6	15.7	0.76
	800	24.0	11.8	11.9	0.80	26.5	14.4	14.4	0.82	29.0	17.0	16.7	0.84	31.5	19.6	19.1	0.85

Note: 1 m = 3.3 ft.

Table 3. Optimum settings for minimum total delay operations.

		W = 6 m, G_p = 13 s, T = 3.3 s				W = 9 m, G_p = 16 s, T = 3.6 s				W = 12 m, G_p = 19 s, T = 3.9 s				W = 15 m, G_p = 22 s, T = 4.1 s			
Q_p	Q_v	M (s)	D_p (s)	D_v (s)	Y_a	M (s)	D_p (s)	D_v (s)	Y_a	M (s)	D_p (s)	D_v (s)	Y_a	M (s)	D_p (s)	D_v (s)	Y_a
100	200	6	3.4	3.6	0.17	7	4.4	4.8	0.19	8	5.6	6.2	0.20	9	6.9	7.7	0.22
	400	6	3.4	4.5	0.34	11	5.5	5.9	0.37	8	5.6	7.7	0.40	9	6.9	9.4	0.44
	600	13	5.1	5.6	0.50	20	8.8	6.8	0.52	33	10.2	7.5	0.52	44	23.4	8.0	0.51
	800	36	14.8	5.8	0.61	51	24.7	6.3	0.60	60	30.4	7.1	0.61	60	31.9	7.6	0.63
200	200	6	3.5	5.9	0.25	7	4.8	7.7	0.29	8	6.3	9.5	0.32	9	8.0	11.4	0.36
	400	6	3.5	7.7	0.50	12	6.8	8.9	0.51	17	10.4	9.9	0.50	21	13.8	10.9	0.50
	600	22	10.1	7.1	0.58	28	14.5	8.0	0.57	34	19.0	8.8	0.57	38	22.4	9.9	0.57
	800	37	17.7	6.9	0.65	45	23.2	7.7	0.65	52	28.1	8.5	0.65	58	32.5	9.4	0.65
400	200	6	3.6	9.1	0.43	7	5.1	11.5	0.50	9	7.3	12.6	0.49	16	9.1	14.2	0.50
	400	13	6.5	8.9	0.56	16	9.3	9.8	0.55	18	11.5	11.1	0.56	19	13.4	12.8	0.59
	600	22	10.8	8.2	0.63	25	13.7	9.5	0.64	28	16.5	10.6	0.64	30	18.8	12.1	0.66
	800	32	15.7	7.3	0.70	37	19.6	9.3	0.71	41	22.9	10.5	0.71	45	26.2	11.7	0.72

Note: 1 m = 3.3 ft.

to those of the equity operations.

In general, minimum total delay operation is more desirable than vehicle priority operation and equity operation because it avoids long pedestrian delays that can result from vehicle priority operation at light vehicle flows and vehicle delays similar to those of vehicle priority operation. And, finally, M_{max} can be chosen to produce acceptable average pedestrian delays at heavy vehicle flows and to alleviate the risk of unstable system performance.

SUMMARY

This paper characterizes the performance of a pedestrian-actuated signal control system in terms of traffic delays, describes the three types of signal operations, and discusses the validity of the assumption that traffic arrives randomly. This was verified by data collected for isolated intersections at Potsdam.

Vehicle priority operation may result in excessive pedestrian delays when long minimum vehicle green durations are chosen. Shorter vehicle greens, on the other hand, risk unstable system performances with respect to vehicle delay. Equity operation is preferable from the viewpoint of pedestrians, but it may bring about unstable system performance when heavy vehicle flows are present. Minimum total delay operation reduces the negative features of the other two operations and is thus more desirable.

Regardless of the types of operation, pedestrian green duration should always be set at its minimum requirement, which is 7 s plus the time needed for a pedestrian to cross the intersection. The setting of minimum vehicle green duration, on the other hand, depends on intersection width and traffic flow. In general, broader intersections and heavier vehicle flows require longer minimum vehicle green duration.

ACKNOWLEDGMENT

This study was supported in part by a grant from the National Science Foundation.

REFERENCES

1. F. B. Lin. A Simulation Analysis of Pedestrian Actuated Traffic Signal Control Systems. Transportation Research, 1978, in press.
2. F. V. Webster. Traffic Signal Settings. Her Majesty's Stationery Office, London, Road Research Technical Paper 39, 1958.
3. B. D. Greenshields, P. Shapiro, and E. L. Ericksen. Traffic Performance at Urban Street Intersections. Yale Bureau of Highway Traffic, New Haven, Conn., Technical Rept. 1, 1947.
4. Manual on Uniform Traffic Control Devices. Federal Highway Administration, U.S. Department of Transportation, 1971.
5. C. M. Abrams and S. A. Smith. Signal Timing for the Pedestrian. TRB, Transportation Research Record 629, 1977, pp. 1-6.
6. Traffic Engineering Handbook. Institute of Traffic Engineers, Arlington, Va., 1965.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Level of Service at Signalized Intersections

T. C. Sutaria, Rady and Associates, Inc., Fort Worth, Texas
J. J. Haynes, Department of Civil Engineering, University of Texas at Arlington

In the 1965 Highway Capacity Manual levels of service at signalized intersections are related to load factor, which was intuitively judged the best level of service indicator available. Load factor has, however, presented such problems as its insensitivity to low service volume, absence of any rational basis for defining breakpoints, and difficulty in identifying loaded cycle. A rational method for quantifying the different levels of service at signalized intersections was therefore needed. In our research we used a road-user opinion survey that involved depicting and rating different traffic situations at a carefully selected single signalized intersection. Over 300 drivers rated randomly arranged film sequences of two types—a driver's view (microview) and an overall view (macroview) of an intersection—and evaluated these films, segment by segment, in terms of appropriate levels of service. Field studies and the attitude survey provided data for the development of two psychophysical models. Statistical analysis indicated that average individual delay correlated better with level of service rating than with measured load factor and encompassed all levels of service. Of all parameters affecting levels of service, load factor was rated highest by road users.

In the 1965 Highway Capacity Manual (HCM) (1) the concept of level of service was introduced for both uninterrupted flow conditions and street intersections with signalized control. Level of service is "a qualitative measure of the effect of a number of factors, which include speed, travel time, traffic interruptions, freedom to maneuver, safety, driving comfort, convenience and operating costs." There are stops at intersections, so speed cannot be the appropriate measure of level of service; an operational index called load factor (LF), then, was used to determine the various levels of service at signalized intersections. HCM defined this index as a "ratio of the total number of green signal intervals that are fully utilized by traffic during the peak hour to the total number of green intervals, for that approach during the same period." The different levels of service are identified alphabetically from A (free flow) to F (forced flow or stop-and-go conditions), based on the value of LF as follows:

Level of Service	Load Factor	Level of Service	Load Factor
A	0.0	D	0.3 to 0.7
B	0.0 to 0.1	E	0.7 to 1.0
C	0.1 to 0.3		

LITERATURE REVIEW

Although LF intuitively seemed to be the best measure of level of service, subsequent problems (2, 3, 4, 5, 6, 7, 8) did arise. May (3, 8) noted the absence of any rational basis for finding breakpoints in LF to define various levels of service. Furthermore, Pontier, Miller, and Kraft (4) and Reilly, Miller, and Jagannath (5) observed that the identification of loaded cycles required considerable judgment on the part of field observers.

Instead of an LF, the former authors (4) used a saturation factor, which they defined as "the number of saturated cycles divided by the total number of cycles during a specified time period." A saturated cycle is any cycle during which the number of vehicles stopped on red is greater than the number of vehicles passing through on the following green.

May and Pratt (3) and May and Gyamfi (6) used LF to correlate levels of service to average individual delay (AID). They took equal delay intervals as a possible basis for the choice of LF values to define various levels of service.

Tidwell and Humphreys (7) presented an alternative procedure that utilized a cycle failure rate to indicate level of service for pretimed signalized intersections. These authors defined a cycle failure as "any cycle during which approach arrivals exceed the capacity of departures." The cycle failure rate is predicted as a probability of arrivals exceeding departure capacity by using a cumulative Poisson distribution. The same authors investigated the feasibility of using AID as an index of level of service offered by a signalized intersection. They argued that, if speed is the criterion for uninterrupted flow conditions, a delay index can be used for intersection operation.

Sagi and Campbell (2) also observed that vehicle delay is a more satisfactory and inherently more useful criterion for assessing levels of service at signalized intersections. The recent Intersection Capacity Workshop (9) too voiced concern regarding the use of LF.

USER-RATING CONCEPT

The above review indicates that future research must

1. Establish, if possible, a rational basis for relating level of service and LF,
2. Check the validity of a hypothesis that LF is a better predictor of level of service than AID, and
3. Establish a relationship between AID and LF at a given level of service.

Our premise is that the quality of flow at any intersection should reflect the attitudes of the road users. There are different levels of satisfaction regarding intersection operation, and it was felt that the opinions of a group of road users could be used to establish a rational relationship between LF and levels of service. This would involve a compilation of all the road users' inclinations, feelings, and degrees of satisfaction about the quality of service at an intersection. We also contend that the service an intersection has been designed to provide might reflect both objective and perceived attributes or dimensions and that there exists a function by which the one can be related to the other. Thus drivers' subjective ratings of quality of flow at a signalized intersection would represent their attitutes or opinions, and LF, AID, or saturation factor would represent an objective rating.

We conducted a literature survey to determine the feasibility of using a user-response or attitude approach, and decided that, because such an approach had been successfully utilized (10, 11, 12) previously, it could provide meaningful results (13, 14).

Our research objectives were (a) to develop the necessary procedure for applying this user-rating concept, (b) to check the validity of the hypothesis that LF is a better predictor of level of service than AID, and (c) to establish a relationship between AID and LF at a given level of service for signalized intersections. These objectives are discussed in detail in the following pages.

APPLICATION OF THE CONCEPT

The steps we took in developing the concepts in this study are described in the following.

Step 1. Selection of a Typical Signalized Intersection

Because the 1965 HCM (1, pp. 111-159) deals primarily with fixed-time or pretimed signal controls, we studied 30 isolated fixed-time intersections for feasibility in the Dallas-Fort Worth area. It was disappointing, therefore, that not one of the 30 could either satisfy the criteria for a full range of load factors or offer a vantage point for an unobstructed overall view for film coverage. As an alternative, a fully actuated signalized intersection located at Lemmon and Oaklawn avenues in Dallas was selected as the best alternative.

Step 2. Conducting Field Studies

Before filming various traffic situations representing levels of service, a number of field studies were carried out during several selected hours and days on the westbound Oaklawn Avenue approach to the intersection. These field studies included:

1. Automatic traffic volume counts for a full week;
2. Manual traffic volume counts to determine directional distribution, vehicle composition, and peak-hour factor (PHF) (1);
3. Delay studies to estimate AID during selected hours [the Sagi and Campbell method (2) was employed because of its sound theoretical approach, its suitability for a fully actuated signal control, and its applicability to all levels of traffic volume, including oversaturation], and the resulting values shown in Table 1 for the selected westbound approach; and
4. LF measurements determined by the number of loaded cycles during the selected hours [a green phase was considered loaded when the following conditions (1) existed: (a) vehicles in all lanes were ready to enter when the signal turned green; and (b) vehicles in all lanes continued to be available to enter, with no long spaces between them during the entire green].

The measured LF [LF(M)] and AID values are tabulated in Table 1. The interpolated LF for the various volumes and the volume to capacity ratios (V/C) were obtained by referring to Figure 6.8 of the 1965 HCM (1, p. 154). The intersection of lines projected from the scale of the unadjusted service volume in vehicles per hour of green (VPHG) for standard or average conditions (1), and from the scale of approach width, identified the interpolated LF, from which the prevailing levels of service were specified. The observed variation between measured and interpolated LF may have been caused by uncertainty in the identification of loaded cycles and by the insensitivity of LF to low service volumes.

Step 3. Preparation of Film for Rating Survey

For the purpose of this research, a microview was defined as a filmed scene showing a traffic situation that an individual driver seated in his automobile would experience while driving through the intersection. A macroview was defined as a filmed scene showing the overall traffic situation on the given approach of the intersection from high above.

Film scenes or segments representing various specific levels of service were contemplated, for a duration of one (or two) signal cycles. It was thus necessary to know the within-the-hour variation in LF values and queue lengths, so that the total collection of filmed scenes would include a full range of known levels of service. Therefore, LF values and maximum queue lengths (Q_{max}) (based on a 15-min time interval) were plotted. By observing Q_{max} on the approach, an LF estimate could be made simply by referring to this plot. Thus the "wait 'n watch" strategy during the hour helped us to decide when to film microviews and macroviews for the various levels of service. The films were taken with a 16-mm camera at 16 frames/s. The final film was prepared from one or two representative film segments for each level of service. The final film included both types of view, and the order of the segments in each type was determined purely on a random basis. The final film contains the segments listed in Table 2.

Table 1. Summary of results.

Time	Day	AID[a] (s)	LF(M)
8:00 to 9:00 a.m.	Sunday 12/8/74	23.39	0.000
7:00 to 8:00 a.m.	Tuesday 11/26/74	44.36	0.353
10:00 to 11:00 a.m.	Tuesday 11/26/74	35.89	0.143
10:00 to 11:00 a.m.	Tuesday 5/21/75	45.00	0.357
12:15 to 1:15 p.m.	Wednesday 11/27/74	64.73	0.885
12:00 n. to 1:00 p.m.	Wednesday 5/21/74	61.33	0.875
2:00 to 3:00 p.m.	Wednesday 11/27/74	52.68	0.536
2:00 to 3:00 p.m.	Friday 5/16/75	57.29	0.560
4:00 to 5:00 p.m.	Friday 5/16/75	62.98	0.905
4:00 to 5:00 p.m.	Friday 12/6/74	62.31	0.730
8:00 to 9:00 a.m.	Saturday 5/10/75	27.05	0.070

[a]Sagi and Campbell method (2).

Table 2. Duration of film segments.

View	Segment No.	Duration of Film at 16 Frames/s (s)	Assigned LF	Level of Service
Micro	1	52	0.070	B
	2	61	0.143	C
	3	70	0.070	B
	4	42	0.000	A
	5	52	0.143	C
	6	108	0.353	D
	7	193	0.885	E
Macro	8	53	0.000	A
	9	62	0.536	D
	10	95	0.143	C
	11	121	0.905	E
	12	109	0.070	B
	13	126	0.143	C
	14	69	0.070	B

Step 4. Conducting Rating Sessions for Group Attitude Survey

A questionnaire was developed for the group attitude survey. In all, 310 respondents participated in the survey at several sessions. After they answered the first part of the questionnaire, which inquired about sex, age, and driving experience, they were requested to indicate their opinions regarding the importance, to quality of service, of five factors: delay, number of stops, traffic congestion, number of trucks and buses, and difficulty of lane changing. Then the first part of the microview film was presented to them segment by segment. Each segment consisted essentially of a driver's view of approaching, waiting, then finally passing through the intersection. After each segment the group were requested to rate traffic operations on two different opinion scales (Figure 1) with regard to level of service provided. After rating each of the seven microviews, they rated the macroviews. These macroviews presented a bird's-eye view of the approach during about one complete cycle of operation.

After the presentation of the film, they were requested to again rate the five factors regarding quality of flow at the intersection. This was done in order to

Figure 1. Scale A and scale B of the attitude survey questionnaire.

SEGMENT:
QUALITY OF SERVICE RATING FORM

A

5 Excellent
4 Very Good
3 Good
2 Fair
1 Poor
0 Very Poor

B

Place a check in the most appropriate box.

I would describe the traffic situation presented in this film segment as a condition of:

(a) Free flow or as "free flowing" as can be expected if there is a traffic signal at the intersection under study, ☐

OR

(b) Tolerable delay, and nearly as good as could be expected at a signalized intersection, ☐

OR

(c) Considerable delay, but typical of a lot of ordinary signalized intersections during busy times, ☐

OR

(d) Unacceptable delay, and typical of only the busiest signalized intersections during the rush hours, ☐

OR

(e) Intolerable delay and typical only of the worst few signalized intersections I have seen. ☐

find out if the films influenced their first opinions.

Step 5. Analyzing the Data

From the field studies of the intersection approach, the four functional relationships established by using linear and curvilinear regression analysis (15, 16) are case 1, V/C ratio versus AID; case 2, AID versus LF(M); case 3, Q_{max} versus LF(M); and case 4, V/C ratio versus LF(M). The statistical analysis indicated that only case 2 was significantly curvilinear, suggesting that a second degree polynominal had a slightly better fit than a straight line (17). For the remaining cases, linear regression fit was adequate (17).

The analysis of the first part of the survey questionnaire is summarized in percentages as follows:

1. Sex: male = 72.9, female = 27.1;
2. Age group: under 25 years = 46.5, 25 to 50 years = 41.6, 51 to 65 years = 3.9, over 65 years = 8.1;
3. Driving experience: under 4 years = 35.8, 5 to 10 years = 36.5, over 10 years = 27.7;
4. Education: high school only = 33.2, some college = 52.3, college graduate = 10.3, college postgraduate = 4.2; and
5. Driving experience (the three percentages refer to little, average, and much experience respectively): downtown = 50.0, 37.4, 12.6; residential areas = 18.4,

Table 3. Summary of ratings for importance of factors to quality of flow.

	Before Rating Film Segments				After Rating Film Segments			
Factor	Mean	Standard Deviation[a]	Median	Rank	Mean	Standard Deviation[a]	Median	Rank
Delay	1.990	1.041	1.792	1	1.923	0.938	1.783	1
Number of stops	2.323	0.978	2.340	2	2.432	1.034	2.439	3
Congestion	2.426	1.106	2.332	3	2.226	0.921	2.154	2
Trucks and buses	3.568	1.220	3.524	5	3.742	1.190	3.726	5
Lane changing	2.703	1.232	2.633	4	2.887	1.090	2.910	4

[a]Very much = 1; above average = 2; average = 3; below average = 4; very little = 5; and no = 6.

Table 4. Level of service rating on continuous scale A.

View	Segment No.	Mean	Standard Deviation[a]	Median	Mode	Kurtosis	Skewness
Micro	1	3.066	0.883	3.049	3.000	0.161	-0.080
	2	2.497	0.771	2.591	3.000	-0.014	-0.178
	3	2.622	1.078	2.550	2.000	-0.483	0.021
	4	3.588	0.957	3.886	4.000	0.584	-0.576
	5	2.651	0.959	2.897	3.000	-0.119	-0.100
	6	1.937	0.977	1.971	2.000	-0.215	0.159
	7	0.893	0.841	0.946	1.000	1.013	0.937
Macro	8	3.440	0.969	3.472	3.000	-0.281	-0.445
	9	2.425	0.979	2.125	2.000	-0.069	-0.046
	10	2.299	0.854	2.125	2.000	0.106	-0.063
	11	1.259	0.849	1.020	1.000	-0.505	0.231
	12	2.989	0.898	2.999	3.000	0.145	-0.246
	13	2.936	0.947	2.979	3.000	0.416	-0.915
	14	3.405	0.867	3.396	3.000	-0.016	-0.220

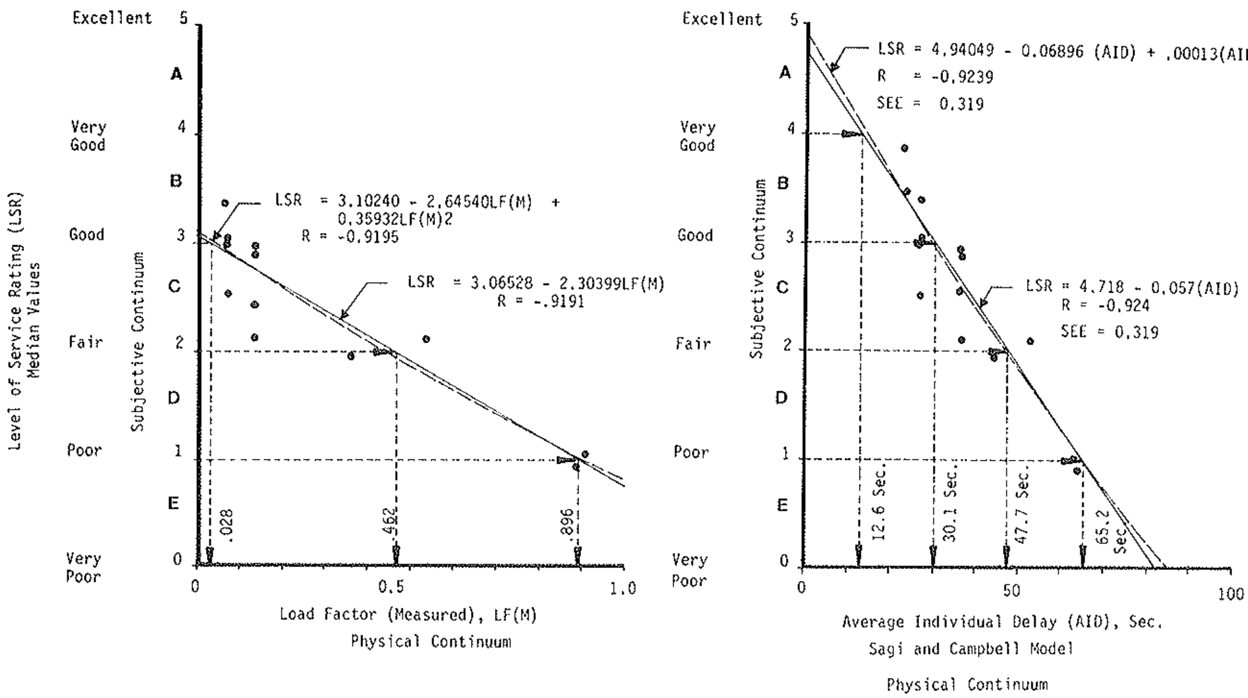
[a]Very poor = 0; poor = 1; fair = 2; good = 3; very good = 4; and excellent = 5.

Table 5. Level of service rating on point estimate scale B.

View	Segment No.	Mean	Standard Deviation	Median	Mode	Kurtosis	Skewness
Micro	1	1.981 3.774[a]	0.784	1.947	2.000	0.126	0.518
	2	2.590 3.013[a]	0.794	2.594	3.000	0.376	0.284
	3	2.539 3.076[a]	1.041	2.500	2.000	-0.487	0.284
	4	1.574 4.283[a]	0.796	0.000	1.000	2.826	1.568
	5	2.468 3.165[a]	0.919	2.403	2.000	-0.355	0.270
	6	3.229 2.214[a]	0.946	3.219	3.000	-0.251	-0.126
	7	4.174 1.033[a]	0.814	4.236	4.000	1.518	-1.050
Macro	8	1.545 4.319[a]	0.761	0.000	1.000	3.656	1.680
	9	2.606 2.993[a]	1.018	2.582	3.000	-0.323	0.291
	10	2.865 2.669[a]	0.871	2.872	3.000	-0.024	0.118
	11	3.868 1.415[a]	0.847	3.939	4.000	0.460	-0.643
	12	2.048 3.690[a]	0.904	1.975	2.000	0.626	0.749
	13	2.168 3.540[a]	0.916	2.084	2.000	0.083	0.599
	14	1.713 4.109[a]	0.758	1.650	2.000	2.417	1.200

[a]Transformed mean comparable to scale A [scale A = 6.25 to 1.25 (scale B)].

Figure 2. Two psychophysical models.



44.2, 37.4; urban freeways = 29.7, 43.9, 26.5; rural highways and freeways = 33.9, 43.2, 22.9.

The second part of the questionnaire had three subdivisions. The analysis of subdivisions 1 and 3 is presented in Table 3. Examining the mean or median ratings, we found that the delay was considered the most important factor both before and after viewing the film segments. The analysis of subdivision 2 is presented in Tables 4 and 5.

To determine the inconsistency in the interpretations of scales A and B, scale B ratings were transformed to equivalent scale A ratings as shown in the tables. The average level of service ratings of scale A were plotted against those of transformed scale B to obtain the deviation of the fitted line from a slope of 1 representing an inconsistency factor. In this particular case, the value was 0.1667.

TESTING THE HYPOTHESIS

For the purpose of testing the hypothesis that LF is a better predictor of level of service than AID, two psychophysical models, level of service rating versus LF(M) and level of service rating versus AID, were formulated as shown in Figure 2. The median values were preferred over the mean values mainly because the former are much less likely to be displaced than the latter in the case of skewed distribution. Both models were developed by a simple linear and curvilinear regression analysis, and a significance test of curvilinearity revealed that the linear regression was a better fit for both.

A hypothesis about the models fitting the data points was tested (17). It was concluded from this analysis that neither model suffered from lack of fit. (Both models fit reasonably well.) Examining the psychophysical model (a) in Figure 2, it is clear that the LF range of 0 to 1 does not encompass the full range of level of service rating from excellent to very poor. It may be that the LF could in some instances have had a negative value, which in its current definition (1) has little or no meaning. In other words, zero LF as defined in the 1965 HCM is insensitive to low service volumes. This can be explained by a plot of LF versus V/C ratio obtained from Figure 6.8 of the 1965 HCM for 6.1, 9.1, and 12.2 m (20, 30, and 40 ft) approach widths. This plot reveals that the zero LF cannot discriminate between V/C ratios below 0.68 (17).

After the above discussion of these two models, the hypothesis that LF is a better predictor of level of service than AID cannot be said to hold. This strongly supports the argument (2, 3, 6, 8) that the definition of level of service as a function of LF (1) is not adequate. These results have once again raised the same question May (8) raised: Can a load factor of less than 0.0 be obtained?

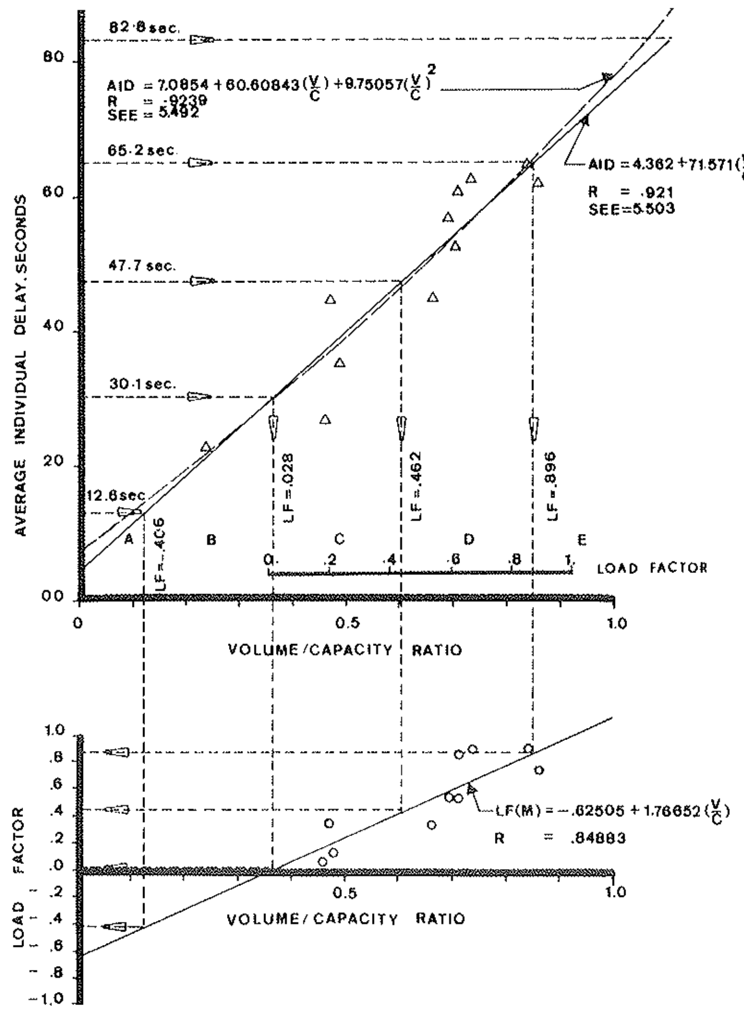
NEW RELATIONSHIPS AMONG LF, AID, V/C RATIO, AND LEVEL OF SERVICE

The two psychophysical models of Figure 2 provide a basis for establishing the breakpoints for LF and AID for defining various levels of service, based on attitudes of road users. These breakpoint values are obtained as follows (Figure 2): (a) draw horizontal lines (representing different levels of service) from the level of service rating scale to regression curve, and (b) from the point of intersections obtained in (a), project vertical lines down to LF or delay scale to obtain breakpoints. A plot (17) of the breakpoints of LF versus the breakpoints of AID would therefore depict the relationship between LF and AID at a given level of service as

$$LF(M) = -0.717\,33 + 0.024\,729\ AID \qquad (1)$$

The breakpoints as shown in Figure 3 and the previously developed functional relationships of AID versus V/C ratio and LF(M) versus V/C ratio helped in the construction of the nomograph illustrated in Figure 3. This nomograph presents the interrelationships among AID, LF, V/C ratio, and the perceived or rated level of ser-

Figure 3. New relationship among LF, AID, V/C ratio, and level of service.



vice. The LF scale of 0 to 1 superimposed on this figure is for comparison. Note that the relationship between levels of service and LF is different from that given in the 1965 HCM.

One could make practical use of these results to estimate the level of service at a signalized intersection approach by first obtaining the AID. The Sagi and Campbell method is recommended for measuring delay in the field. However, other methods such as the test-car (19), the time-lapse photography (17, 19), the Berger-Robertson (18), or the sampling (19) may be used by applying appropriate corrections for the variation between these methods and Sagi and Campbell's. The problem of wide variation in delay results has been addressed to some degree (17). AID could also be estimated by utilizing a known or anticipated V/C ratio. After obtaining AID (or V/C ratio), the nomograph can be used to determine the corresponding level of service.

It should be noted that the nomograph of Figure 3 was developed from the only study of the single, not ideal, signalized intersection. However, the methodology presented could be used to develop similar nomographs for any fixed-time signalized intersections. Applied to a sufficient number of different types of signalized intersections, it would provide a rational basis for establishing intersection level of service and would overcome the problems associated with (a) the present troublesome definition of the LF and its accompanying measurement difficulties and (b) the arbitrarily or intuitively chosen breakpoints for the various levels of service that do not now adequately encompass the full range of traffic volumes.

CONCLUSIONS AND RECOMMENDATIONS

The following conclusions resulted from this research.

1. The hypothesis that LF is a better predictor of level of service than AID was tested and was rejected.
2. The entire LF range of 0 to 1 failed to encompass all levels of service A through F (5-1) and indicated that LF values less than zero were required for ratings 5 and 4 (levels of service A and B). In other words, LF is insensitive to low service volumes.
3. AID, on the other hand, did correlate slightly better to levels of service rating than LF. It also encompassed all categories of levels of service rating. In addition, road users rated delay highest of the various parameters influencing levels of service at signalized intersections.
4. A new relationship was developed among AID, LF, V/C ratio, and level of service based on attitudes or perception. These relationships seemed to overcome the problems associated with using LF as the criterion for establishing the level of service. The resulting nomograph affords flexibility in the sense that one could predict level of service not only from LF, but also from AID and V/C ratio.

The following recommendations are made in view of the results of this study and the experience gained by it:

1. AID not LF should be used for predicting level of service, because AID provides a rational, meaningful, and inherently more useful predictor;
2. A fully actuated signalized intersection had to be used for this study, and similar studies should be carried out on a number of different kinds of signalized intersections; and
3. A simultaneous filming and field traffic studies procedure should be used because it eliminates the assignment of approximated LF values to film segments.

REFERENCES

1. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
2. G. S. Sagi and L. R. Campbell. Vehicle Delay at Signalized Intersections—Theory and Practice. Traffic Engineering, Feb. 1969, pp. 32-40.
3. A. D. May and D. Pratt. A Simulation Study of Load Factor at Signalized Intersections. Traffic Engineering, Feb. 1968, pp. 44-49.
4. W. E. Pontier, P. W. Miller, and W. H. Kraft. Optimizing Flow on Existing Street Networks. NCHRP, Rept. 113, 1971.
5. E. Reilly, I. Dommasch, and M. Jagannath. Capacity of Signalized Intersections. TRB, Transportation Research Record 538, 1975, pp. 32-47.
6. A. D. May and P. Gyamfi. Extension and Preliminary Validation of a Simulation of Load Factor at Signalized Intersections. Traffic Engineering, Oct. 1969, pp. 46-52.
7. J. E. Tidwell and J. B. Humphreys. Relation of Signalized Intersection Level of Service to Failure Rate and Average Individual Delay. HRB, Highway Research Record 321, 1970, pp. 16-32.
8. A. D. May. Initial Experiences With the Highway Capacity Manual. Traffic Engineering, Dec. 1967, pp. 46-49.
9. Intersection Capacity Workshop Notes. TRB, Transportation Research Circular 161, Dec. 1974.
10. A. Burg and S. F. Hulbert. Predicting the Effectiveness of Highway Signs. HRB, Bulletin 324, 1962, pp. 1-11.
11. J. N. Fox, J. J. Haynes, and B. T. Williams. An Evaluation and Analysis of Regional Public Attitudes Towards Existing and Proposed Transportation Operations. Univ. of Texas at Arlington, Feb. 1975; Urban Mass Transportation Administration.
12. W. N. Carey, Jr., and P. E. Irick. The Pavement Serviceability-Performance Concept. HRB, Bulletin 250, 1960, pp. 40-58.
13. W. S. Torgerson. Theory and Methods of Scaling. Wiley, New York, 1958.
14. C. H. Coombs. A Theory of Data. Wiley, New York, 1967.
15. N. Draper and H. Smith. Applied Regression Analysis. Wiley, New York, 1966.
16. G. W. Snedecor and W. G. Cochran. Statistical Methods. Iowa State Univ. Press, Ames, 1968.
17. T. C. Sutaria. A Study of Level of Service at Signalized Intersections. Univ. of Texas at Arlington, PhD dissertation, Aug. 1976.
18. H. D. Robertson and W. G. Berger. Berger-Robertson Method for Measuring Intersection Delay. TRB, Transportation Research Record 615, Abrigment, 1976, pp. 45-46.
19. J. P. Louice. Traffic Engineering: Theory and Practice. Prentice-Hall, Englewood Cliffs, N.J., 1973.

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

Estimation of Unprotected Left-Turn Capacity at Signalized Intersections

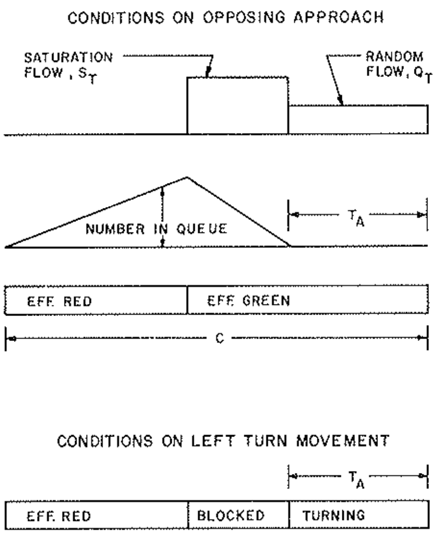
Daniel B. Fambro, Carroll J. Messer, and Donald A. Andersen, Texas Transportation Institute, Texas A&M University

A mathematical model was developed to calculate the unprotected left-turn capacity of a pretimed signalized intersection. The capacity depends primarily on the volume of traffic opposing the left-turn movement and the percentage of the cycle available for this maneuver. Parameters were determined from field studies conducted in several Texas cities. The model was used to estimate the unprotected left-turn capacity for approaches both with and without exclusive turn lanes. Opposing volumes ranged from 200 to 1000 automobiles/h in one, two, or three lanes. Green splits from 30 to 70 percent of the cycle were analyzed. Unprotected left-turn capacities as predicted by the model and by the Highway Capacity Manual were compared. General agreement was found at a 50 percent green split; significant differences existed at other green splits.

The Highway Capacity Manual (1, pp. 139-140) states that the service volume (capacity) of an unprotected movement with a left-turn bay of adequate length is equal to the "difference between 1,200 vehicles and the total opposing traffic volume in terms of passenger cars per hour of green, but not less than two vehicles per signal cycle." If a left-turn lane is not provided, an adjustment factor based on the percentage of traffic turning left is used to determine the capacity of that approach. This adjustment varies with street width and available parking. However, as demonstrated by a study at Northwestern University (2), this factor does not reflect the effects of different levels of opposing traffic flow.

The Australian method (3, 4) of estimating capacity utilizes a left-turn equivalency factor for opposed left turners based on opposing vehicle volume. The Metropolitan Toronto Roads and Traffic Department (5) uses gap acceptance criteria to estimate the left-turn capacity of two- or three-lane streets having an exclusive turn lane without an exclusive phase for the turning move-

Figure 1. Left-turn capacity conditions.



ment when opposing flows are greater than 600 vehicles/h. May (6) reported that from a list of 22 areas of needed research on intersection capacity, left-turn capacity was ranked as having the second highest priority.

In this paper the left-turn capacity of an approach having no protected signal phasing is related to opposing traffic volumes and intersections of different geometric design. Approaches with and without left-turn lanes are considered. A mathematical model was developed to calculate the unprotected left-turn capacity of signalized intersections. The parameters were determined from field studies conducted in several Texas cities, and practical applications have been documented in other reports (7, 8).

MODEL DEVELOPMENT

A review of the literature and the experience of the research team indicated that the left-turn capacity of an intersection was related to the amount of traffic opposing the left-turn movement. Unless there is an exclusive turn phase, left-turning vehicles must turn across the intersection through gaps in the opposing stream. For higher opposing flow rates, fewer gaps occur. At an intersection having no exclusive phase for turning, the left-turn movement is blocked for a time by the dissipation of the opposing queue that builds up during the red phase. Therefore, the left-turn capacity of an intersection having no exclusive turn phase is a function not only of the probability of gaps occurring in the opposing traffic stream but also of the length of the gaps. An equation to express the above concept can be written

$$Q_L = (T_A/C)Q_{LH} \qquad (1)$$

where

Q_L = left-turn capacity of an approach (automobiles/h);
T_A = available time per cycle during which turning may occur (s);
C = cycle length (s); and
Q_{LH} = left-turn capacity of an approach across free-flow, random traffic (automobiles/h of available green time).

Time Available for Turning

At intersections having no exclusive turn phase, the events shown in Figure 1 occur on the approach opposite a left-turn movement. At a point during the yellow time, left turns and opposing through traffic are stopped and a queue starts to build on the opposing approach at the average arrival rate of Q_T per 3600 vehicles/s. The portion of the yellow time not used by through traffic can be thought of as lost time. The queue continues to build on the opposing approach at the same rate during the red period of the signal. At the beginning of the green a second lost time occurs because of the time it takes for vehicles in the queue to start moving. At this time the queue has reached its maximum size.

After the lost time at the beginning of the green interval, the queue begins to clear at the rate of the saturation flow (S_T) minus the average arrival rate (Q_T) converted to vehicles per second. After the queue has cleared, normal flow resumes at the average arrival rate for the remainder of the green time plus a portion of the yellow time. During this time interval (T_A), left-turning vehicles may cross the opposing traffic movement as acceptable gaps occur in the opposing flow, which is assumed to be random.

If the average number of arrivals on each lane of the opposing approach were equal on a cycle-by-cycle basis, the queues in each lane would clear simultaneously. However, as this is not the case, the time available for turning should be based on the time required for the lane with the longest queue to clear the intersection.

Bellis (9) estimated that under capacity conditions the percentage distribution of traffic among lanes was 55 to 45 on a two-lane approach and 40 to 35 to 25 on a three-lane approach. If only one vehicle arrived during a cycle, the percentage distribution would be 100 to 0 or 100 to 0 to 0. By using these two boundary conditions, we estimated the percentage distributions of traffic in the highest volume lane for various volume conditions as $P = 0.55 + 0.45e^{-0.18m}$ for a two-lane approach and $P = 0.40 + 0.60e^{-0.13m}$ for a three-lane approach, where P = proportion of the traffic in highest volume lane (expressed as a decimal); m = average number of arrivals per cycle [$(Q_T \times C)/3600$]; Q_T = total opposing volume (automobiles/h); and C = cycle length (s).

As previously discussed, the left-turn movement cannot begin until the longest opposing queue has cleared the intersection. The time required to clear the longest opposing queue in a lane is

$$T_Q = [P \times Q_T(L_1 + R + L_2)]/(S_T - P \times Q_T) \qquad (2)$$

where

T_Q = time for longest opposing traffic queue to clear (s);
L_1 = portion of yellow time not used by through traffic (s);
R = length of red phase of cycle (s);
L_2 = initial lost time at the beginning of the green interval (s); and
S_T = saturation flow of longest opposing queue [1750 automobiles/lane·h (10)].

Therefore, the time available for left turning per cycle (T_A) is

$$T_A = G + Y - L_1 - L_2 - T_Q \qquad (3)$$

where

G = length of green phase of cycle (s) and

Y = length of yellow phase of cycle (s).

Free-Flow Turn Capacity

Once the longest queue of opposing traffic has dissipated, free-flowing vehicles continue to approach the intersection in a random manner and to form gaps of various sizes. Left turners wait at the signal until one of these gaps of adequate length arrives. An acceptable gap is one equal to or larger than the critical gap, which is that gap for which an equal percentage of turning traffic will accept a smaller gap as will reject a larger one (11).

More than one vehicle may turn through a gap if it is long enough, and the time between consecutive vehicles turning through the same gap is defined as the turn headway (H). If a stationary arrival rate can be expected during periods of free flow, the negative exponential distribution can be used to represent the probability of gap occurrence. Based on this concept, Drew (11) presents the following equation, which can be used to determine the left-turn capacity of an intersection during free-flow conditions:

$$Q_{LH} = Q_T(\{\exp(-q_T T_c)\}/\{1 - \{\exp(-g_T H)\}\}) \tag{4}$$

where

Q_T = total opposing through and right traffic (automobiles/h);
q_T = total opposing through and right traffic (automobiles/s);
T_c = critical gap (s); and
H = turn headway (s).

Figure 2. Field data reduction.

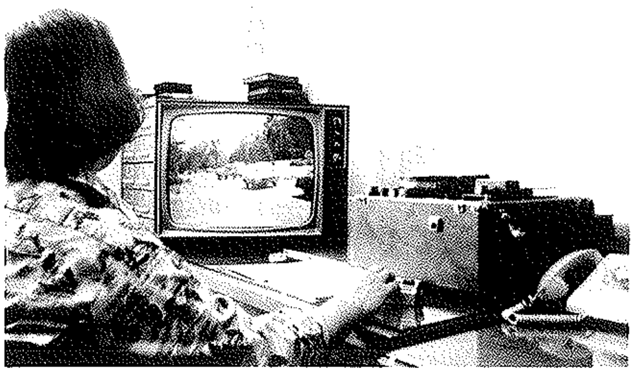


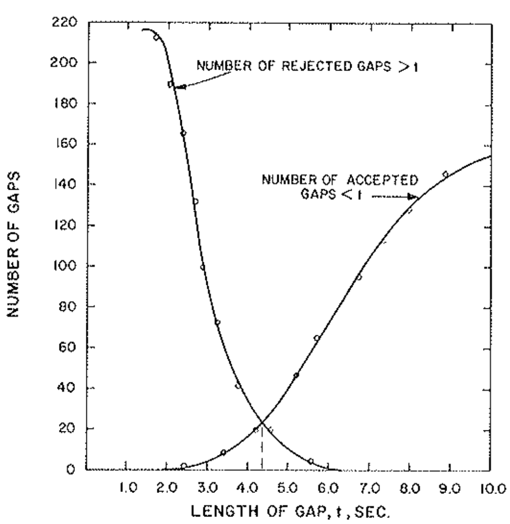
Table 1. Intersections with separate left-turn lanes.

Intersection	Location	Date	No. of Through Lanes	No. of Cycles Recorded
Fifteenth at Congress	Austin	10/10/74	2	22
Twenty-sixth at San Jacinto	Austin	12/4/74	3	13
Texas Avenue at Coulter	Bryan	1/9/75	2	19
Hammerly at Gessner	Houston	3/26/75	2	27
Montrose at Alabama	Houston	3/27/75	2	20
Montrose at Richmond	Houston	4/15/75	2	44

Table 2. Intersections without left-turn lanes.

Intersection	Location	Date	No. of Through Lanes	No. of Cycles Recorded
First at Oltorf	Austin	12/3/74	2	20
College at Sulphur Springs	Bryan	1/7/75	2	19
College at Dodge	Bryan	1/8/75	2	35
College at Dodge	Bryan	1/15/75	2	28
SH 21 at Nineteenth	Bryan	1/16/75	1	16
College at Carson	Bryan	1/27/75	2	50
Thirty-eighth at Lamar	Austin	1/28/75	2	31

Figure 3. Pairs of accepted and rejected gaps for three intersections with left-turn lanes.



SAMPLE PROBLEM

A sample problem illustrating the calculation of the left-turn capacity of an unprotected left-turn movement follows.

Given a two-lane approach with adequate left-turn bay length, two-phase signal timing, and two lanes of opposing flow,

Q_T = 600 vehicles/h,
C = 70 s,
G = 28 s,
Y = 3 s, and
R = 39 s.

Assume that

$L_1 + L_2$ = 4 s,
S_T = 1750 vehicles/h,
T_c = 4.5 s, and
H = 2.5 s.

Then determine the left-turn capacity of the approach in vehicles per hour and solve the lane distribution for a two-lane approach:

$$P = 0.55 + 0.45e^{(-0.18 \times m)} \tag{5}$$

where

$$m = (Q_T \times C)/3600 = (600 \times 70)/3600 = 11.7 \tag{6}$$

Therefore,

$$P = 0.55 + 0.45e^{(-0.18 \times 11.7)} = 0.605 \tag{7}$$

Calculate the time for queue to clear:

$$T_Q = [P \times Q_T \times (R + L_1 + L_2)]/[S_T - (P \times Q_T)] \tag{8}$$

where

Table 3. Average proportion of vehicles in higher volume lane cycle by cycle.

Location	Direction of Flow	Hourly Volume	Proportion in Higher Volume Lane
SH 6 at SH 21	Southbound	312	0.705
Montrose at Alabama[a]	Northbound	473	0.594
SH 6 at SH 21	Northbound	506	0.651
Hammerly at Gessner	Westbound	511	0.608
Montrose at Alabama[a]	Southbound	529	0.628
Hammerly at Gessner	Eastbound	600	0.614
Montrose at Alabama[b]	Southbound	820	0.562
Montrose at Richmond	Northbound	925	0.562
	Southbound	975	0.559
Montrose at Alabama[b]	Northbound	1002	0.543

[a]On 3/27/75. [b]On 3/15/75.

$$T_Q = [0.605 \times 600 \times (39 + 4)]/[1750 - (0.605 \times 600)] \quad (9)$$

Therefore

$$T_Q = 11.25 \text{ s} \quad (10)$$

Determine the total time available for left turns:

$$T_A = G + Y - (L_1 + L_2) - T_Q \quad (11)$$

where

$$T_A = 28 + 3 - 4 - 11.25 \quad (12)$$

Therefore

$$T_A = 15.75 \text{ s} \quad (13)$$

Calculate the left-turn capacity of the approach across free-flow, random traffic:

$$Q_{LH} = Q_T([\exp(-q_T T_c)]/\{1 - [\exp(-q_T H)]\}) \quad (14)$$

where

$$Q_{LH} = 600\{\exp[-(600/3600)](4.5)\} \div (1 - \{\exp[-(600/3600)](2.5)\}) \quad (15)$$

Therefore

$$Q_{LH} = 832 \text{ vehicles/h} \quad (16)$$

Determine the left-turn capacity of approach per hour of signal operation:

$$Q_L = (Q_{LH} \times T_A)/C \quad (17)$$

where

$$Q_L = (832 \times 15.75)/70 \quad (18)$$

Therefore

$$Q_L = 187 \text{ vehicles/h} \quad (19)$$

Compare this with the minimum left-turn capacity (1.6 left turns per cycle):

$$Q_{L_{min}} = (1.6 \times 3600)C \quad (20)$$

where

$$Q_{L_{min}} = (1.6 \times 3600)/70 \quad (21)$$

and where

$$Q_{L_{min}} = 82 < 187 \quad (22)$$

Therefore the left-turn capacity is

$$Q_L = 187 \text{ vehicles/h} \quad (23)$$

PARAMETER STUDIES

Data were collected at several intersections to determine the model parameters by using a portable videotape recording system. This permitted us to record and analyze more traffic measures than we could have with our limited number of data collectors. Two members of the research team were required to operate the system in the field. The playback unit and monitor shown in Figure 2 were used to replay the videotapes. A stopwatch was used to time vehicle movements and signal intervals. Twelve intersections were filmed during the course of the study, six with and six without separate left-turn lanes. Tables 1 and 2 summarize the study locations.

Results of Field Studies

Observations in the field and from the videotapes resulted in the following data and conclusions about vehicles turning left at intersections with no exclusive turn phase.

Critical Gap

Of the six intersections that had left-turn lanes, usable critical gap data were collected for three. The largest gap rejected and the smallest gap accepted for each left-turning vehicle were recorded. From these data, graphs of the cumulative totals for rejected and accepted gaps intersect at a value approximating the critical gap (11). The critical gap data for the three intersections were plotted as shown in Figure 3. Based on these results, a value for the critical gap of 4.5 s was selected as a reasonable value for use in the left-turn capacity equation.

Turning Vehicle Headway

Headways between left-turning vehicles were found by measuring the time between completion of the turn movement of successive vehicles turning through the same gap. Only those cycles during which more than one vehicle turned through the same gap resulted in usable data. The following table is a summary of the headways for vehicles turning from left-turn lanes.

Location	Number of Headways	Average Headway (s)
Fifteenth at Congress	15	2.71
Texas Avenue at Coulter	29	2.79
Hammerly at Gessner	111	2.60
Montrose at Alabama	10	2.44
Montrose at Richmond	146	2.31
Total	311	2.48

For intersections without left-turn lanes, the average turn headways were slightly higher as shown below.

Location	Number of Headways	Average Headway (s)
First at Oltorf	10	2.56
College at Dodge	20	2.72
College at Carson	5	2.32
Total	35	2.62

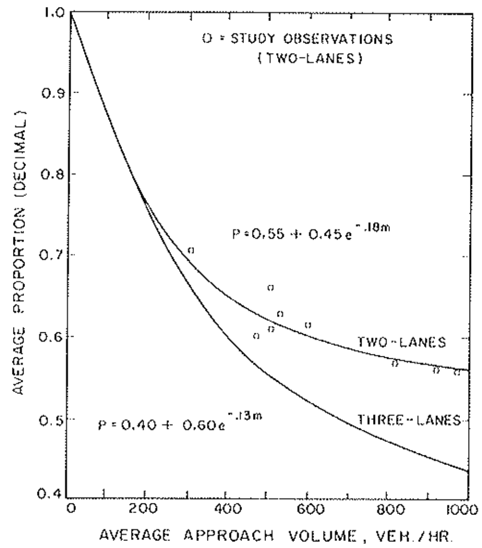
The results of this analysis indicate that values of 2.5 s for headways at intersections with left-turn lanes and 2.6 s for headways at intersections without left-turn lanes would be appropriate for use in the left-turn capacity equation.

Lane Distribution

As the data collection progressed, we observed a variation in the number of vehicles using each through lane on a cycle-by-cycle basis, which occurred even when the volumes in each lane were approximately equal over a period of time. Lane distribution is significant in the queue clearance portion of the model because left turners cannot begin to turn until the longest queue has cleared the intersection.

To measure this characteristic, vehicle volumes per lane were recorded cycle-by-cycle from the videotape. For each cycle the higher volume was divided by the total volume on the approach to get the proportion of the vehicles in the longest queue, expressed as a decimal. These values were averaged over the number of cycles recorded and are shown in Table 3 with corresponding expanded hourly volumes. This summary contains data from an extra hand count made at the intersection of Tex-6 and Tex-21 in Bryan for lower volume levels. As expected, the percentage of volume in the longest queued lane decreased as approach volume increased. The observed and modeled lane distributions are illustrated in Figure 4. For a no left-turn lane and two lanes of opposing flow, the equations presented in this paper slightly underestimated the length of the longest lane queue at high volumes. A complex mathematical equation was used to determine the lane distribution in three instances.

Figure 4. Average proportion of traffic in higher volume lane on one approach.



Observations From Field Studies

Before data collection, two questions were raised: How frequently do vehicles turn left before the opposing queue begins to move? and How many vehicles turn left on the yellow and start of red intervals each cycle?

Turning Before the Opposing Queue Starts

In 1966 Dart (12) observed 220 signal cycles containing a left turner at the head of the queue. Based on these observations, he concluded that when the lead vehicle

Table 4. Vehicles on approaches with left-turn lanes turning left before opposing queue starts.

Location	No. of Cycles With Left Turners at Head of Queue	No. of Vehicles Turning Before Opposing Queue Enters	Rate of Occurrence
Fifteenth at Congress	27	0	0
Twenty-sixth at San Jacinto	11	0	0
Texas Avenue at Coulter	30	0	0
Hammerly at Gessner	52	1	0.019
Montrose at Alabama	20	0	0
Montrose at Richmond	84	4	0.048

Table 5. Vehicles on approaches without left-turn lanes turning left before opposing queue starts.

Location	No. of Cycles With Left Turners at Head of Queue	No. of Vehicles Turning Before Opposing Queue Enters	Rate of Occurrence
First at Oltorf	15	1	0.067
College at Sulphur Springs	20	0	0.000
College at Dodge[a]	26	4	0.154
College at Dodge[b]	38	4	0.105
College at Carson	44	3	0.068
Thirty-eighth at Lamar	35	0	0.000

[a]On 1/8/75. [b]On 1/15/75.

Table 6. Turns on yellow and red at intersections with left-turn lanes.

Location	Direction of Flow	Opposing Volume	No. of Cycles: Lefts on Yellow	Lefts on Red	Lefts on Both	Total Vehicles Cleared
Montrose at Richmond	Northbound	975	3	11	3	29
	Southbound	925	3	12	4	26
Montrose at Alabama	Northbound	475	0	1	0	1
	Southbound	448	3	0	0	3
Hammerly at Gessner	Northbound	511	6	1	3	13
	Southbound	600	8	1	4	21
Texas at Coulter	Northbound	94	2	9	0	2
	Southbound	198	0	1	1	3
Fifteenth at Congress	Eastbound	655	6	8	3	23
Twenty-sixth at San Jacinto	Eastbound	295	0	11	0	13

Table 7. Turns on yellow and red at intersections without left-turn lanes.

Location	Direction of Flow	Opposing Volume	No. of Cycles: Lefts on Yellow	Lefts on Red	Lefts on Both	Total Vehicles Cleared
College at Carson	Northbound	525	0	0	0	0
	Southbound	437	5	0	0	5
College at Dodge	Northbound	451	3	0	0	3
	Southbound	432	4	1	1	7
First at Oltorf	Southbound	225	1	1	0	2
Thirty-eighth at Lamar	Eastbound	569	0	11	0	11
	Westbound	463	0	8	0	8

Table 8. Estimated left-turn capacity of an unprotected left-turn lane of adequate length.

Proportion of Green to Cycle[a]	No. of Opposing Through Lanes	Combined Opposing Through and Right-Turn Volumes 200 Automobiles/h	400 Automobiles/h	600 Automobiles/h	800 Automobiles/h	1000 Automobiles/h
0.3	1	232	82	82	82	82
	2	260	159	82	82	82
	3	261	168	105	82	82
0.4	1	368	204	82	82	82
	2	392	276	187	114	82
	3	393	285	207	147	100
0.5	1	503	333	181	82	82
	2	524	394	292	208	138
	3	525	401	309	236	177
0.6	1	639	463	307	164	82
	2	655	512	397	302	223
	3	656	518	410	324	254
0.7	1	775	593	434	291	153
	2	787	630	502	396	307
	3	788	634	512	413	331

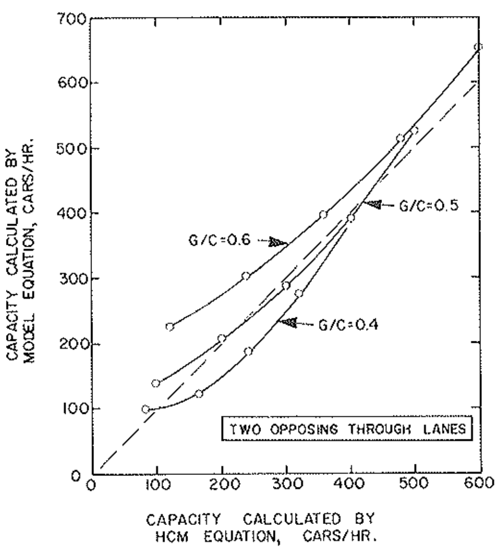
[a]Cycle length is 70 s.

Table 9. Estimated left-turn capacity of a lane with no protected signal phase or left-turn lane.

Proportion of Green to Cycle[a]	No. of Opposing Through Lanes	Total Opposing Volumes[b] 200 Automobiles/h	400 Automobiles/h	600 Automobiles/h	800 Automobiles/h	1000 Automobiles/h
0.3	1	132	50	50	50	50
	2	148	95	53	50	50
	3	149	101	66	50	50
0.4	1	209	122	50	50	50
	2	223	166	119	50	50
	3	223	171	131	97	50
0.5	1	286	200	114	50	50
	2	298	237	185	122	50
	3	298	241	195	156	122
0.6	1	363	278	194	108	50
	2	373	307	251	187	122
	3	373	311	259	214	175
0.7	1	441	356	274	192	105
	2	447	378	317	252	188
	3	448	381	323	273	228

[a]Cycle length is 70 s.
[b]Includes right turns and throughs for one opposing lane and includes right and left turns and throughs for two and three opposing lanes.

Figure 5. Comparison of Highway Capacity Manual and model left-turn capacities.



was a left turner the probability of its "jumping the gun" (turning in front of the opposing queue) was about 0.145. A much lower rate was found to occur during the course of our data collection, as shown in Tables 4 and 5. The relatively few vehicles jumping the gun do not appear to significantly increase the left-turn capacity of an intersection. Therefore, no adjustment is made for this phenomenon in the final capacity analysis.

Left Turns on Yellow and Red

For relatively light opposing traffic there is usually enough time for left turns during the green interval. As opposing traffic increases, green time diminishes for left turners until the queue fails to clear the intersection during the green. However, our observations indicated that at this point the lead left turner in the queue will usually be waiting in the intersection and will choose to turn either on the yellow or on the beginning of the red. In fact, for high opposing volumes, this appears to be the major source of left-turn capacity. The Highway Capacity Manual reflects this concept in giving left-turn capacity as "no less than two vehicles per cycle." The Australians (3, 4) indicate that at least 1.5 vehicles/cycle can turn left during this time period. The Texas State Department of Highways and Public Transportation (13) has been using a value of 1.6 left turners/cycle as a minimum in some capacity analyses.

To be certain that these turns were indeed occurring, we gathered the data summarized in Tables 6 and 7 from the videotapes. These data indicate that if the queue lead driver is not given an opportunity to turn left during the green interval, he or she will turn on the yellow or the red. As the opposing volumes rise to near capacity, the only left-turn capacity possibilities are those on yellow or red. Results of this study indicate that the left-turn capacity of an intersection averaged 1.41 vehicles/cycle turning on yellow and red when a left-turn lane was present and 1.03 vehicles/cycle turning on yellow and red when no turn lane was provided. As this study was rather limited, no change from the currently used value of 1.6 vehicles/cycle was made where a left-turn lane was provided. However, a minimum turn volume per cycle of 1.0 was selected where no left-turn lane was provided. It should be noted that these values are based on the characteristics of Texas drivers.

MODEL RESULTS

The capacity of a turning movement from a left-turn lane without protected signal phasing as estimated by the model is given in Table 8. The capacity values are the maximum possible sustained flow rates that could occur during the peak 15-min period of the design hour, by automobiles only. Trucks in the opposing volume and effective left-turn demand must be converted into equivalent automobiles. Any left-turn bay must be sufficiently long to prevent blockages between the left-turn queue and through vehicles.

An attempt was made to estimate the capacity of a left-turn movement without a protected signal phasing or a left-turn lane. The field studies had indicated that the previous model could be applied with some modifications and simplifying assumptions. Changes in the critical gap (T_c = 4.5 s) were neither observed nor assumed. The minimum headways between consecutive left turns from unprotected lanes were observed to be slightly longer (H = 2.6 s) than those from left-turn lanes. To simplify the capacity and equivalence calculations, it was assumed that 50 percent of the traffic in a median lane from which left turns can be made were left turners. This assumption is more representative of heavy volume conditions than light flow operations.

The minimum effective left-turn headway across long gaps becomes H_{eff} = 2.06 + 2.60 s, because we assumed that every other vehicle was going through at a minimum headway of 2.06 s (10). Using T_o = 4.5 s and H_{eff} = 4.66 s in the model, we calculated the left-turn capacities of an approach having no left-turn lane, no protected left-turn signal phasing, and 50 percent of median lane traffic turning left, as presented in Table 9.

A comparison of the unprotected left-turn capacity of a sufficently long separate turn lane was calculated by the model and the expression from the Highway Capacity Manual as shown in Figure 5. At a green time to cycle length (G/C) ratio of 0.5, the model and the Highway Capacity Manual expression yield approximately the same results. For G/C ratios less than 0.5, the model estimates left-turn capacity to be less than that predicted by the Highway Capacity Manual. At G/C ratios greater than 0.5, the model predicts a greater capacity than the Manual does. This indicates that the model is comparable to Highway Capacity Manual estimates under average intersection conditions; however, the model is more sensitive than the Manual to changes in these average conditions.

CONCLUSIONS

The left-turn capacity of an intersection approach that has no exclusive turn phase should not be expected to exceed those values given in Tables 8 and 9 unless field studies at the intersection so indicate. The left-turn capacities given are based on the model developed previously and on the characteristics of Texas drivers.

If a left-turn lane is not provided and the left-turn volume exceeds 80 percent of the given practical capacity in Table 9, a channelized or otherwise designated left-turn lane may be considered. If the left-turn demand is heavy when opposing approach volumes are also heavy, then a separate protected left-turn signal may also be required to reduce the delays suffered by the left turners.

ACKNOWLEDGMENTS

This paper discusses one phase of a research project on the effects of design on the operational performance of signal systems conducted by the Texas Transportation Institute and sponsored by the Texas State Department of Highways and Public Transportation in cooperation with the Federal Highway Administration, U.S. Department of Transportation. We wish to thank Harold D. Cooner, Herman E. Haenel, and Elmer A. Koeppe of the Texas State Department of Highways and Public Transportation for their technical inputs and constructive suggestions throughout the project. The assistance provided by Don A. Ader, Donald R. Hatcher, Murray A. Crutcher, and the secretarial staff of the Urban Transportation Systems Division of the Texas Transportation Institute is also gratefully acknowledged. The contents of this paper reflect our views, and we only are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration. This paper does not constitute a standard, specification, or regulation.

REFERENCES

1. Highway Capacity Manual—1965. HRB, Special Rept. 87, 1965.
2. E. Fellinghauer and D. S. Berry. Effects of Opposing Flow on Left-Turn Reduction Factors at Two-Phase Signalized Intersections. TRB, Transportation Research Record 489, 1974, pp. 13-18.
3. Australian Road Capacity Guide: Provisional Introduction and Signalized Intersections. Australian Road Research Board, Bulletin 4, June 1968.
4. A. J. Miller. The Capacity of Signalized Intersections in Australia. Australian Road Research Board, Bulletin 3, Mar. 1968.
5. Notes From the Intersection Capacity Workshop. TRB, Circular 161, Dec. 1974.
6. A. D. May. Intersection Capacity 1974: An Overview. TRB, Special Rept. 153, 1975, pp. 50-59.
7. C. J. Messer and D. B. Fambro. A Guide for Designing and Operating Signalized Intersections in Texas. Texas Transportation Institute, Research Rept. 203-1, Aug. 1975.
8. C. J. Messer, D. B. Fambro, and D. A. Andersen. A Study of the Effects of Design and Operational Performance of Signal Systems. Texas Transportation Institute, Research Rept. 203-2F, Aug. 1975.
9. W. R. Bellis. Capacity of Traffic Signals and Traffic Signal Timings. HRB, Bulletin 271, 1960, pp. 45-67.
10. D. J. Berry. Capacity and Quality of Service of Arterial Street Intersections. Texas Transportation Institute, Research Rept. 30-1, Aug. 1974.
11. D. R. Drew. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
12. O. K. Dart. Development of Factual Warrants for Left-Turn Channelization Through Digital Computer Simulation. Texas A&M Univ., College Station, PhD thesis, 1966.
13. Traffic Signal Warrant for a Separate Left-Turn Phase. Texas State Department of Highways and Public Transportation, File D-18T, Apr. 5, 1974.

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

Abridgment

Effect of Dotted Extended Lane Lines on Single Deceleration Lanes

W. D. Keck, Division of Economic and Environmental Analysis, and
A. W. Roberts, Bureau of Operations Research, New Jersey Department of Transportation

In the Manual on Uniform Traffic Control Devices (MUTCD) there is a provision for an optional dotted extension of the right edge line for parallel and tapered deceleration lanes. According to MUTCD, the dotted extension is usually made up of short segments 0.6 m (2 ft) long and longer gaps 1.2 m (4 ft) or more.

The purpose of this study was to determine the effectiveness of the optional dotted extension of the right edge line for deceleration lanes.

Tapered and parallel deceleration lanes are currently used with various interchange geometries in New Jersey. Different exit ramp and approach geometries and extended dotted lane lines may interact to cause variations in driver performance. In the summer of 1975, a study was made of the effect of dotted lines at both tapered and parallel right, single deceleration lane exits with varying approach and exit ramp geometries on New Jersey's Interstate system.

The data collection procedures included the following tasks: performing before studies, installing extended dotted lane lines, and performing after studies.

SITE LOCATIONS

Traffic was observed at two different locations in the state: Rt-440 and I-287 in the northeast, and in the south I-295 in Camden County, which served as the location of many sites. Sixteen sites were selected in all, and 12 were used for actual study. The remaining 4 were upstream sites where extended dotted lane lines were installed to help eliminate any possible uniqueness effect at the other 12 sites. Of these 12, 7 were categorized as tapered deceleration types and 5 as parallel deceleration types. Because other characteristics were similar, tapered and parallel lanes were differentiated by length, tapered lanes being those shorter than 122 m (400 ft). Two sites were selected as controls, where no dotted lane lines were applied.

PREPARATION OF SITES

The deceleration lanes were standardized according to MUTCD guidelines prior to the before study. Extended dotted lane lines were installed from the gore point on tapered lanes and from the ends of skip-lines on parallel lanes, and upstream to the point where the edge line begins to taper at the beginning of the exit lane (see Figure 1). The dotted lane lines were made up of 0.6-m (2-ft) strips placed every 8 m (27 ft) on center. The best spacing was determined subjectively by observing several gap lengths on deceleration lanes.

At each study site, a reference point was painted to help define the zones for observers, who classified each vehicle entering the deceleration lane according to zone. The reference point separated zones 1 and 2 where there was an even chance of exiting maneuvers from each, as determined by prior observation of traffic. It was felt that an equal distribution of zone 1 and zone 2 exiting maneuvers would provide the most sensitive measure of a change in exiting behavior. We counted another type of maneuver, labeled 4, which entailed crossing back from the deceleration lane into the through lanes.

DATA COLLECTION

Data were collected at the 12 sites by two teams of three hidden observers. One observer counted two-axle exiting traffic and all through traffic. The other two observers independently counted two-axle exiting vehicles by maneuver type.

Observations were conducted between 10:00 a.m. and 12:10 p.m. and between 1:00 and 3:10 p.m., with a 10-min break at 11:00 a.m. and 2:00 p.m. A total of 4 h of traffic data were recorded Monday through Friday, and for each day these before data were matched with data collected three weeks later for the same time and day of week.

The two control sites, 7 and 12, are both tapered, and each is located at one of the two locations near the experimental sites. Traffic was observed at each control site for the first 3 d of before studies, for 3 d between the before and the after studies, and for the last 3 d of after studies.

We divided each study week into high-volume sites on Monday and Tuesday and low-volume sites on Wednesday, Thursday, and Friday. This gave the lower volume sites more matched data periods than the higher volume sites, but the latter had larger samples in each data period.

METHOD OF ANALYSIS

Before and after rates of exit maneuvers in 10-min periods were matched by time of day and day of the week in order to see if the rates changed with statistical significance at the 95 percent confidence level. A standard deviation 1.64 or more is a significant difference at the 95 percent level, using a one-tailed test.

Changes between rates before and after were tested by using the conservative, nonparametric, Wilcoxon matched-pairs, signed-ranks test.

RESULTS AND CONCLUSIONS

The comparison of total exit maneuver rates by lane type for each study site in Table 1 shows a significant increase in zone 1 exit maneuvers (except for site 1) after dotted lane lines were installed; zone 2 exits decreased proportionately. This is a beneficial effect for orienting exiting traffic into the deceleration lane. No significant change was noted in before and after rates for zone 3 exit maneuvers on parallel deceleration lanes. Type 4 maneuvers, less than 1 percent of the total volume, are not shown.

Site 8 experienced a high increase in through volume (40 percent), but exit volume decreased by 22 percent for the period between the before and the after studies. This large change may be indicative of a bias effect on exit maneuver rates. The rest of the sites experienced less than a 15 percent change of volume.

Table 1 also presents a summary of exit maneuver rates for the two control sites. Significant changes in

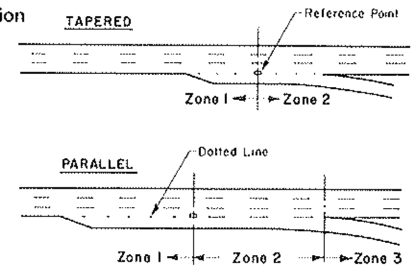


Figure 1. Deceleration lanes and zones.

Table 1. Comparison of exit rates in zone 1 before, between, and after dotted line marking.

Lane Type	Site No.	Before	Between	After	Statistical Significance (Z value)
Tapered	1	48	–	47	0.207, n = 41[a]
	2	45	–	69	7.07, n = 69
	5	41	–	64	5.78, n = 44
	8	58	–	60	2.33, n = 41
	9	60	–	77	4.99, n = 33
Parallel	3	54	–	66	5.85, n = 67
	4	31	–	50	5.36, n = 38
	10	24	–	37	6.14, n = 65
	11	70	–	75	3.59, n = 67
	13	45	–	51	2.34, n = 39
Control	7	55	52	60	–[b]
	12	50	58	56	–[b]

[a]Not significant. [b]Significant.

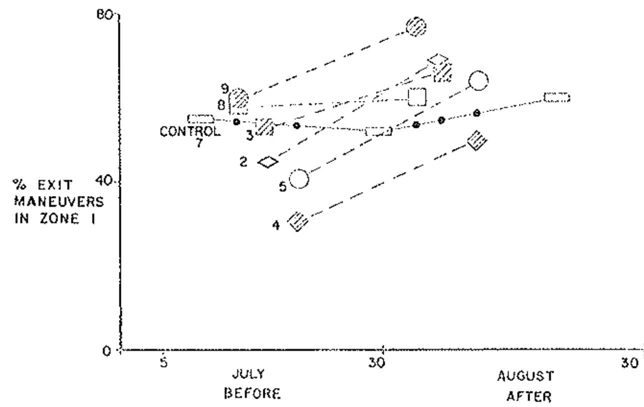


Figure 2. Percentages of exit maneuvers at I-295 sites.

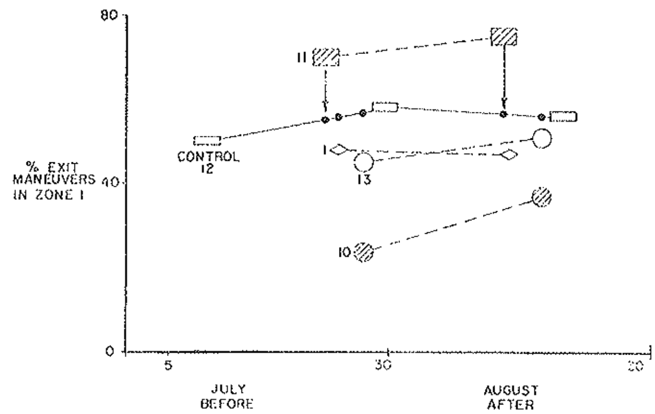


Figure 3. Percentages of exit maneuvers at I-287 and Rt-440 sites.

these rates in zone 1 occurred for each combination of data collection periods.

The graphic comparison of experimental and control site results in Figures 2 and 3 of zone 1 shows the negligible effect that control site percentages have. The control sites were not identical to the experimental sites they represent, so we do not know what the precise corrections should be. Correction accuracy is also limited by our assuming linear relationships between the percentage of exit maneuvers in zone 1 and the passage of time as determined by only three points representing control site variation.

The following major conclusions were drawn from the results of the Wilcoxon test of matched pairs and the subsequent analysis:

1. The dotted extension of a right edge line was more effective in orienting exiting traffic sooner into the deceleration lane, single tapered and parallel deceleration lanes and
2. No significant change was noted in zone 3 exits (crossing the painted gore regions) at parallel deceleration lanes.

It was also observed that exiting vehicles use the shorter deceleration lanes with less variation, because there is less room to maneuver. As a result, the use of dotted lane lines did not have as marked an effect on orienting vehicles into these shorter lanes.

ACKNOWLEDGMENTS

We wish to give special thanks to Jeffrey J. Gertler, who executed the project and supervised the collection of data. The innovative inputs and administrative support of Richard L. Hollinger are also gratefully acknowledged, as is the assistance provided by the Bureaus of Operations Research and Maintenance. The project was requested by Robert Nolan and funded by the Federal Highway Administration and the state of New Jersey.

Publication of this paper sponsored by Committee on Traffic Control Devices.

Abridgment

Evaluating Location Effectiveness of Freeway Directional and Diversion Signs

Moshe Levin, Chicago Area Expressway Surveillance Project, Illinois Department of Transportation

The location of freeway directional signs or signs displaying real-time messages informing drivers of downstream exit or diversion points is important to providing safe and efficient traffic operation at these points.

The costs and the operating and maintenance problems of real-time information signs usually allow only a single sign before a diversion point. Although the visibility of this kind of sign is hardly a problem, its location is critical and should be determined analytically.

For directional exit signs, the spacing and number specified by current practice generally provide adequate maneuvering distance. However, in certain situations a different sign arrangement should be quantified to determine its better potential effectiveness. This can be done by calculating the theoretical probability of completing a lane change, under various traffic flow conditions, within a certain distance determined by the size and location of a sign.

The following model uses gap acceptance concepts and considerations for safe manuvering to calculate the above probability and to evaluate the effectiveness of various sign locations.

MODEL DESCRIPTION

The Lane-Change Process

A vehicle is traveling on lane i + 1 at time mean speed U_{i+1} next to a traffic substream on lane i that is also traveling at U_{i+1} (in this case $U_{i+1} > U_i$). The probability density function of headways on the ith lane may be denoted by $\theta_i(t)$, where t is a certain time needed for a lane change (to the right adjacent lane) established by the driver. Here the lane-change process begins.

The driver changing looks at adjacent lane i and considers some or all of the following:

1. The speed of the approaching (lagging) vehicle on lane i,
2. The relative position of the lead vehicle on lane i,
3. His or her own speed and operational characteristics, and
4. His or her own gap acceptance characteristics.

At time t + T, where T is the driver's decision time, the driver either accepts the gap and changes lanes or rejects it. Rejection means that the driver's critical gap (the point at which anything shorter is unacceptable) has been exceeded. When the gap is accepted, the lane-change maneuver begins the moment a safe maneuver can be accomplished. If the gap is rejected, the driver increases speed, begins evaluating the next gap, and reaches a decision concerning that gap.

The lane-change process may be full or partial. The full process occurs in either of the following situations: (a) the gap encountered immediately after the need for a lane change has been established is rejected because it is smaller than the driver's critical gap; (b) the gap is larger than the critical gap, but the driver's position relative to the lead vehicle in it makes a lane change seem hazardous, and the gap is rejected.

The full process consists of the following three phases: phase 1 is waiting for an acceptable gap or lag; phase 2 is bringing the vehicle to such a position relative to the accepted gap that the maneuver can begin; and phase 3 is the actual lane-change maneuver.

A partial process will occur if the first encountered gap is accepted. This takes place in one of the following two forms. The driver needs to adjust position relative to the accepted gap before a safe maneuver can be initiated, or the dirver's relative position allows immediate initiation of the lane-change maneuver. Each of the forms consists of phases 2 and 3.

Each phase in the various forms of the lane-change process has its own distribution function with respect to distance. These functions are themselves functions of the actual traffic conditions. The expression $f_i(x_i, \bar{V}, \bar{U})$ denotes the distribution functions of the distance x required to complete phase 1 under volume condition $\bar{V}$ and speed condition $\bar{U}$ for a particular type or form of the process. $\bar{V}$ and $\bar{U}$ are vectors that represent traffic flow rate and time mean speed, respectively, on the lanes of a one-way freeway section during the time of the process.

Any function representing any phase of the process is considered independent of those characterizing other phases for a given set of speed and volume conditions. The distribution functions (f) of the distance required to complete the various lane-change processes were considered as the convolutions of the individual distribution functions (F, A, and B) describing each phase (1, 2, and 3) in the appropriate form of the process and can be presented as follows.

The full process is

$$f_F(x,\bar{V},\bar{U}) = f_{1F}(x_1,\bar{V},\bar{U}) \times f_{2F}(x_2,\bar{V},\bar{U}) \times f_3(x_3,\bar{V},\bar{U}) \tag{1}$$

The partial process, form A, is

$$f_A(x,\bar{V},\bar{U}) = f_{2A}(x_2,\bar{V},\bar{U}) \times f_3(x_3,\bar{V},\bar{U}) \tag{2}$$

and the partial process, form B, is

$$f_B(x,\bar{V},\bar{U}) = f_{2B}(x_2,\bar{V},\bar{U}) \times f_3(x_3,\bar{V},\bar{U}) \tag{3}$$

The composite distribution representing the combination of the various lane-change processes from one lane to the adjacent lane for a given set of volume and speed conditions is as follows:

$$f(x,\bar{V},\bar{U}) = a_F f_F(x,\bar{V},\bar{U}) + a_A f_A(x,\bar{V},\bar{U}) + a_B f_B(x,\bar{V},\bar{U}) \tag{4}$$

where a_F, a_A, and a_B are the probabilities of occurrence of the three types of the process. A full development of the model is given elsewhere (1). A more complicated expression could be developed for lane change in a three-lane section.

Figure 1. Lane-change distance-cumulative probability relationship for four-lane, two-way freeway.

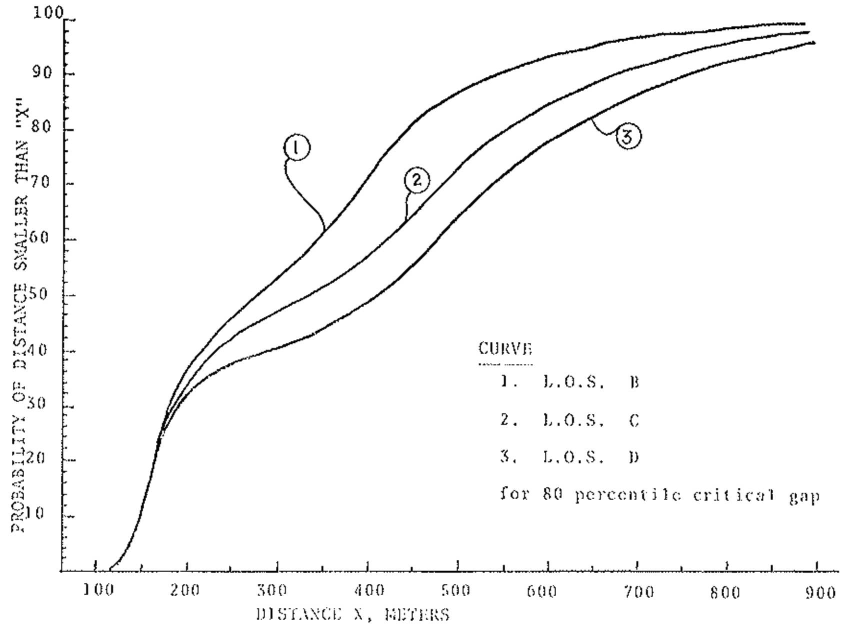


Table 1. Distance in advance of signs at which lane-change decisions are made.

Level of Service	Speed on Origin Lane (km/h)	Distance (m)		
		Real-Time Sign	Exit Area Sign	1.6 and 3.2-km Signs[a]
B	91	340	180	106
C	82	345	189	111
D	75	349	192	115

Note: 1 m = 3.3 ft; 1 km = 0.62 mile; 1 km/h = 0.62 mph.
[a]1 and 2-mile signs.

Table 2. Effectiveness of location of directional and diversion signs.

Sign Type	Distance (km)	Level of Service[a]		
		B	C	D
Directional	From exit	0.26	0.25	0.25
	1.6 before	0.80	0.80	0.80
	3.2 before	0.80	0.80	0.80
Diversion	From diversion point	0.48	0.40	0.35
	0.4 before	0.78	0.75	0.70
	0.8 before	0.80	0.80	0.80

Note: 1 km = 0.62 mile.
[a]Based on the 80th percentile critical gap.

Data Collection and Analysis

A three-lane section of the Gulf Freeway in Houston, Texas, was selected for this study. Data on traffic stream speeds and lane flow rates were collected by the Freeway Control Center during a dry 2-week period. An instrumented vehicle driven by a test driver was used to obtain data on the following: delay in making a lane change to the next lane once an instruction for a change was given and the distance traversed during the lane-change process.

Nearly 500 lane changes were performed, of which approximately 450 occurred at freeway levels of service B, C, and D. The critical gap characteristics of the test driver were determined by field measurements of the angular velocity, and the characteristics of the headway distribution function $\theta_i(t)$ on the freeway lanes were assumed to be of an Erlang nature. The value of the parameters varied according to flow rates given by Drew, Buhr, and Whitson (2).

How well the collected data fit the developed model was determined by using the Kolmogorov-Smirnov test (3) and was found to lie at the 1 percent level of significance.

DEVELOPMENT OF LANE-CHANGE DISTANCE-PROBABILITY RELATIONSHIP

As discussed above, the characteristics of the lane-change distance-probability relationship are partly a function of traffic flow rates and speeds on the origin and destination lanes and driver's gap acceptance characteristics.

The relationship between traffic speeds and flow rates on a four-lane freeway was investigated by Webb and Moskovitz (4). Lane values for speed and flow rate for levels of service B, C, and D were derived from this relationship. And the relationship between the driver's critical gap and his or her threshold angular velocity in the gap acceptance process, developed by Michaels and Weingarten (5), is shown in the following equation:

$$T = [(w/\psi U_i^2)(U_{i+1} - U_i)]^{1/2} \quad (5)$$

where W is car width and ψ is the driver's threshold angular velocity. It was Rock (6), in his studies of driver's threshold angular velocity, who established the cumulative distribution of this parameter.

The inverse relationship between threshold angular velocity and critical gap suggests that the P percentile angular velocity corresponds to the (i - P) percentile critical gap. The lane-change distance-probability curves were developed for the 80th percentile critical gap of 27×10^4 radians/s, meaning that 80 percent of drivers changing lanes are expected to complete their maneuvers within a certain distance with a certain probability determined by the appropriate curve.

Figure 1 shows the developed lane-change distance-cumulative probability curves for a four-lane freeway for levels of service B, C, and D for the 80th percentile critical gap.

EFFECTIVENESS OF LOCATION OF DIVERSION AND DIRECTIONAL SIGNS

A sign's legibility distance will determine when a driver will decide to change lanes. According to current practice, the desired legibility distance of such signs is 19.5 m (65 ft) per 2.5 cm (1 in) of letter height for daylight conditions, assuming that drivers have 20/20 vision.

For real-time information signs with 50-cm (20-in) letter height and a perception-reaction time of 2 s, a need for a lane change could arise 390 - (1.4 × U m) [1300 - (2.94 × U ft)], where U is the driver's speed in kilometers per hour in advance of the sign.

The Manual on Uniform Traffic Control Devices says that the letter height for the exit area sign is 30 cm (12 in) and is 20 cm (8 in) for both the 1.6-km (1-mile) and 3.2-km (2-mile) signs. With such a sign arrangement and a perception-reaction time of 2 s, a need for a lane change could arise 234 - (1.4 × U m) [780 - (2.94 × U ft)] in advance of the exit sign and 156 - (1.4 × U m) [520 - (2.94 × U ft)] in advance of each one of the other two signs.

Data from the volume-speed relationships developed by Webb and Moskovitz (4) give the speeds on the origin lane shown in Table 1. For real-time information and directional signs these speeds correspond to the distances in advance of the sign in the table.

The probability or effectiveness values for the specified locations of the directional and real-time information signs with respect to diversion points were derived from Figure 1. These values apply to only 80 percent of the driver population, but they were adjusted to represent the effectiveness of the location of signs for the total driver population. The adjusted values are presented in Table 2.

The above analysis considers each sign as if it were the only one there and ignores the effects of similar upstream signs. From Table 2 it can be seen that there is not much difference in the effectiveness of the directional sign at the exit area for the various levels of service, and both the 1.6-km and 3.2-km directional signs are equally effective.

As to the real-time information sign of 50-cm letter size, the variations in the effectiveness of a sign located at the diversion point are quite considerable for the three levels of service, as illustrated in the curves in Figure 1.

The effectiveness values shown in Table 2 indicate the percentages of drivers who will accomplish the lane-change maneuver with reasonable safety and comfort.

CONCLUSIONS

It seems that the probability curves for the distance involved in a lane-change maneuver on a four-lane freeway provide a reasonable measure of effectiveness of a location of a sign. This measure relates to the probability of accomplishing the maneuver under reasonably safe and comfortable conditions.

An evaluation of the current practice of locating directional signs reveals that both the 1.6-km and 3.2-km signs are approximately the same in effectiveness. Except for improving the visibility of the message, the contribution of the 3.2-km sign to the success of the lane-change maneuver can be considered rather small. Improving these signs or changing their locations with respect to exit points could reduce the number of signs needed.

As to real-time diversion signs of 50-cm letter height on a four-lane freeway, the gain in effectiveness from locating a sign more than 0.4 km (0.25 mile) in advance of the diversion point is relatively small.

ACKNOWLEDGMENTS

The contents of this paper reflect my own views, not those of the Illinois Department of Transportation, and I alone am responsible for the facts and accuracy of the material presented.

REFERENCES

1. M. Levin. Some Investigations of the Freeway Lane Changing Process. Texas A&M Univ., College Station, PhD thesis, 1970.
2. D. R. Drew, J. J. Buhr, and R. H. Whitson. The Determination of Merging Capacity and Its Application to Freeway Design and Control. Texas Transportation Institute, College Station, Research Rept. 430-4, 1967.
3. B. Ostle. Statistics in Research. Iowa State Univ. Press, Ames, 1963.
4. R. M. Webb and K. Moskovitz. California Freeway Capacity Study—1956. Proc., HRB, Vol. 36, 1957, pp. 587-641.
5. R. M. Michaels and H. Weingarten. Driver Judgement in Gap Acceptance. U.S. Bureau of Public Roads, 1965.
6. M. L. Rock. Visual Performance as a Function of Low Photopic Brightness Levels. Journal of Applied Psychology, No. 37, 1953, pp. 449-454.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Influence of Control Measures on Traffic Equilibrium in an Urban Highway Network

Nathan H. Gartner, Federal Highway Administration, U.S. Department of Transportation

Traffic control measures play a critical role in determining equilibrium between demand and supply in an urban highway network. Present techniques attempt to optimize network performance assuming a fixed demand pattern. Experience indicates that control measures can be used to influence demand patterns in such a way that total network performance is significantly improved. On the basis of the interdependence of control measures and resulting traffic flow patterns, this paper develops an analytical framework for a systematic optimization of the operation of a highway system. Two practical examples demonstrate such an optimization on a limited scale, and a model and a mathematical program for achieving it are presented.

Across the nation close to 150 cities and municipalities have turned to computerized traffic control systems to relieve congestion, to save time and energy, and to reduce pollution (1). We expect that within the next decade every urban area of 100 000 or more will have a computer system of one kind or another controlling traffic flow on its streets and freeways (2, 3). These systems are bound to have a profound impact on the performance of transportation networks. However, the electronic hardware seems to be far more advanced than practicing traffic engineers are yet capable of utilizing. It is the purpose of this paper to describe a framework for systematically optimizing the operation of such systems within the broader context of traffic equilibrium in a network.

The effects of various traffic engineering measures on traffic flow in urban areas has been the subject of many investigations, several of them sponsored by the National Cooperative Highway Research Program, Area 3: Traffic—Operations and Control (4).

One particular project on optimizing flow on existing street networks (5) stated its objective as developing "a practical method of measuring the degree of change in network traffic flow resulting from various system modifications." This objective was based on the recognized need to better understand the effects that many commonly used transportation system modifications have on traffic flow. The measures studied included directional control and lane use, curb lane controls, channelization, signal controls, and bus operation. The conclusions of this study suggested that, in improving a downtown area, those elements that involve the functional use of streets, such as one-way patterns, reversible lane operations, and major parking prohibitions, should be developed first. Then all the minor influences that create friction in the traffic stream, such as turning movements, truck loading, pedestrian interference, proper allocation of signal time to the various approaches of an intersection, and other traffic problems of the area, should be analyzed and corrected. Only when local friction has been adequately reduced can traffic be properly platooned to implement signal progressions. After this has been successfully accomplished, the best locations for bus stops may be considered, and bus movement in the progressive system may be developed.

There is no doubt that traffic flow can significantly benefit from such operational improvements in an urban street network. In most cases, the benefits to the traveling public will considerably outweigh the costs of analysis, engineering, and construction of the improvement. However, most of these improvements tend to reinforce the existing traffic patterns rather than to explore alternatives to them (6). This is most evident in calculating and updating signal control settings, which is based upon the existing or projected pattern of traffic. The result is that signal controls are regularly set to favor existing heavy flows of traffic over light flows. This is particularly evident when arterial progression schemes rather than areawide optimization schemes are being implemented. Consequently, the routes that drivers already use and the destinations they most frequently select are made more attractive, and the ones they usually reject are made even less attractive. Thus, the choices that created existing traffic flow patterns receive constant reinforcement.

Many traffic management schemes indicate that it may be desirable, for one reason or another, to reduce the amount of traffic using parts of the road network (7). It is quite conceivable that, in concert with other control measures, this could be achieved by means of traffic signals. Signals would be installed at the points of entry to the area concerned, and the settings would be used to regulate flow into the area. But even without restraining vehicle movements or forcing changes of destination, a change in the traffic patterns may be beneficial simply because drivers select a better route.

It is well known that individual route choices do not generally lead to the least total travel costs for all users of the system (8). This is so because drivers do not choose routes to enhance the common good but to minimize their travel costs. The resulting flow pattern (termed "user-optimal") is not necessarily the best for the community (termed "system-optimal"). Therefore we must decide whether traffic controls can be used to influence individual route choice in order to achieve a traffic flow pattern that benefits all. Two examples described in this paper indicate that this is indeed possible.

TRAFFIC EQUILIBRIUM

The core of any transportation analysis problem is predicting flows that will use all the segments of the system. The pattern of traffic flows in a road network is the result of choices subject to various constraints made by a large number of individuals. Such constraints may be of an economic nature, such as the income level of the individual, or technological, such as the performance characteristics of the available modes of transportation. Each individual, given the particular constraints that apply, chooses whether to travel at all or if so to which destination, at which time, by which mode, and by which route. The collective choices made by all individuals in a transportation system will determine a set of flows and levels of service that can be viewed as an equilibrium condition between the demand for and the supply of transportation services. The basic hypothesis underlying this statement is that there is a market for transportation that can be separated from other markets (9).

The basic paradigm of equilibrium in a transportation network is illustrated in Figure 1A. Let us define

- L = level of service (such as trip time) on a particular facility;
- V = volume of flow on this facility;
- S = supply;
- D = demand;
- T = specification of the transportation system in terms of its technological characteristics (including operating policies); and
- A = specification of the socioeconomic activity system.

Then the supply function $L = S(T, V)$ shows an increase in the level of service as volume increases, and the demand function $V = D(A, L)$ a decrease in volume as the level of service increases (in the negative sense). The resulting equilibrium point $E_o(L_o, V_o)$ occurs at the intersection of the two curves.

In computing traffic equilibrium for an urban street network with signalized intersections, one usually assumes that the origin-destination demand pattern is an inelastic function of the level of service. This is a reasonable assumption when estimating short-run equilibrium, that is, when no trip distribution changes are expected during the anlysis period. The interdependence between signal settings and flow patterns can also be explained with reference to the equilibrium paradigm shown in Figure 1B. A change in signal settings causes the entire supply curve to shift. On the other hand, a change in flow corresponds to a movement along the curve. For example, a change from curve S_1 to S_2 would result in a decrease in the level of service measure from L_1 to L_2 for the same flow V_1, whereas an increase in flow from V_1 to V_3, perhaps from increased supply, would cause an increase in the level of service measure from L_2 to L_3. The following sections examine the interaction of control measures and traffic volumes in determining equilibrium conditions in an urban road network.

TRAFFIC CONTROL

Intersections are an important source of delay in urban areas, and many are controlled by traffic signals. Therefore, analysis of signalized intersections is of primary importance in determining traffic equilibrium in an urban network. The average delay (d) per vehicle on a signalized approach can be represented, according to Webster (10), as the sum of two components

$$d = d_s + d_d = 0.9\Big(\big\{[C(1-g)^2]/[2(1-q/s)]\big\} + \big\{(3600x^2)/[2q(1-x)]\big\}\Big) \tag{1}$$

where

- g = G/C, proportion of green time (G) to cycle time (C);
- q = arrival flow (vehicles/h);
- s = saturation flow (vehicles/h);
- K = gs, capacity of the approach (vehicles/h);
- x = q/K, degree of saturation;
- d_d = delay if flow is deterministically uniform; and
- d_s = additional delay caused by stochastic nature of traffic flow.

The way delay varies with traffic flow (or, equivalently, with the degree of saturation) is illustrated in Figure 2. The characteristics closely resemble the transportation supply functions shown in Figure 1. Both green time and cycle time must be determined, and all flows must be considered before we can derive the level of service (delay) at which traffic through the intersection will be served. Analyses of single intersection controls can be found in the literature (10, 11).

When two or more intersections are in close proximity, some form of linking is necessary to reduce delays and to prevent frequent stopping. Because signalized intersections have a platooning effect on leaving traffic, it is advantageous to synchronize signals to operate within a common cycle. It also becomes necessary to coordinate the signals, that is, to establish an offset between the signals so that loss to traffic is minimized. Networks, like single intersections, also require that all demands be considered simultaneously when control variables are being determined.

Several computer methods have been developed for optimizing signal settings in networks (12, 13). One method particularly suitable for the application discussed in this paper is the mixed-integer traffic optimization method (MITROP) (14). In this, for each link (i, j) in the network a link performance function comprising a deterministic and a stochastic component is constructed. The deterministic component represents the average delay incurred per vehicle in a periodic flow through signal j. The stochastic component arises from variations in driving speeds, marginal friction, and turns and is expressed by the occurrence of overflow queues at intersection stop lines. Thus, we can also consider the total delay in the network to be composed of two components, D_d and D_s. The network signal setting problem can then be stated in a general form as

1. Determining offsets ϕ_{ij}, splits G_{ij}, cycle time C;
2. Minimizing total delay $D = D_d + D_s$; and
3. Subjecting them to loop offset constraints $\sum_{\text{loop}} \phi_{ij} = n_\ell C$ (n_ℓ - integer), to link split contraints $G_{ij} + R_{ij} = C$, and to signal capacity constraints $q_{ij}C \le s_{ij}G_{ij}$.

This is a nonlinear mathematical program with integer decision variables and can be solved by appropriate computer optimization packages.

TRAFFIC CONTROL AND ROUTE CHOICE

Traditionally traffic engineers have chosen signal settings that minimize travel costs by using any of the methods mentioned above, given a fixed traffic flow pattern. It is assumed that all route choices are predetermined and result in constant flows on each link, irrespective of the controls imposed on that link and, hence, irrespective of the level of service offered by the link. This assumption is, in most cases, incorrect. As shown in Figure 3, there exists an interdependence between traffic controls and link volumes; one determines the other and vice versa. We must therefore develop a model that incorporates both traffic controls and route choice (traffic assignment) and thus provides a tool for establishing mutally consistent signal settings and traffic flow patterns. The following two examples are given to illustrate quantitatively the object of such a model.

Example 1

Figure 4A shows the intersection of Commonwealth

Figure 1. Basic equilibrium paradigms in a transportation network.

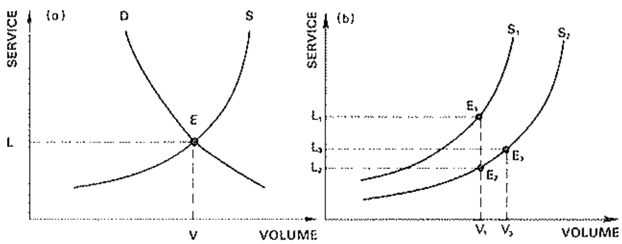


Figure 2. Average delay to traffic on a signal controlled intersection approach.

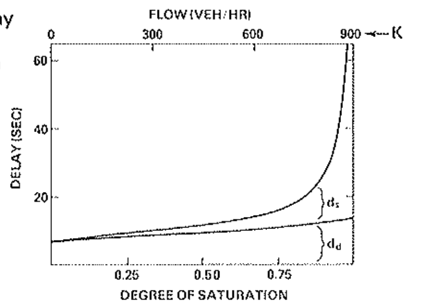
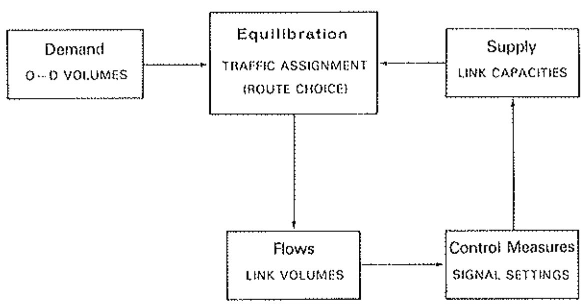


Figure 3. Interdependence of link volumes and traffic controls.



Avenue at Boston University Bridge when it has an advanced green phase allowing west-to-north left turns. The average delay for the left turners is 63.0 s, and the rate of total delay for all traffic passing through the intersection is 41.6 vehicles/s^2 (Table 1).

Turning to Figure 4B, it is seen that west-to-north movements can also be accomplished by traveling an extra 75 s through route $A_1 \rightarrow E \rightarrow F \rightarrow B_2$. The signal can now be operated on a shorter cycle time, because the left-turn movement is prohibited. Total expected travel time on the extended route is 22.0 + 75.0 + 16.7 = 103.7 s, an increase of 64 percent (Table 2). However, the total rate of delay for all users of the intersection is now only 32.3 vehicles/s^2, a reduction of 22.3 percent with respect to the previous figure.

It should be noted that despite the fact that a left-turn arrow is no longer shown, it is physically possible but illegal to make the turn against the opposing traffic. Occasionally vehicles do indeed make the turn. But, taking into account the good chance of getting a costly ticket, which can be regarded as a very long delay, most drivers will choose the longer route, thus achieving system optimization at this location.

Example 2

This example is based on results reported by Maher and Akcelik (15). Signal settings and traffic flows were calculated iteratively by the following procedure:

1. Optimizing signal settings for the current flow patterns;
2. Reassigning origin-destination demands corresponding to the prevailing signal settings by using a capacity restraint technique; and
3. Iterating among steps 1 and 2 until no significant changes occur in successive link flow patterns and signal settings.

The particular network studied is described in Figure 5 and this origin-destination matrix:

From	To X	To Y
A	600	600
B	0	600
C	0	2000

Two flow patterns are calculated, following the two principles enunciated by Wardrop. One is the following user-optimized flow pattern, representing the way traffic will distribute itself among alternative routes.

Flow	Route 1 to 3	Route 2 to 3
A to Y	0	600
B to Y	180	420

Total network travel time in this user-optimizing pattern is 60.8 vehicles/h^2. The second is this system-optimized flow pattern, representing the way traffic should distribute itself in order to minimize travel time for all users of the system.

Flow	Route 1 to 3	Route 2 to 3
A to Y	0	600
B to Y	600	0

Total network travel time in this system-optimizing pattern is 53.7 vehicles/h^2. It is seen that the second flow pattern produced a reduction of 12 percent in travel time and could be achieved simply by banning left-turns for flow B to Y at intersection 1.

Discussion

These examples clearly show the advantages of combining traffic control and route choice. The problem is how to incorporate route choice as part of the traffic control optimizing program. One approach is based on the general formulation given by $D = D_d + D_s$ and $q_{ij}C \leq s_{ij}G_{ij}$. Only origin-destination trip demands are given, and all link flows q_{ij} are decision variables. The objective function is modified to also include the travel time on the links T, in addition to the delay at the signals,

$$T = \sum_{i,j} q_{ij} t_{ij} \tag{2}$$

where t_{ij} is the travel time on link i, j. A new set of constraint equations is added to represent the origin-destination demands and continuity of flow by nodes. In matrix notation it takes the following form

$$\underline{A} \times \underline{q} = \underline{p} \tag{3}$$

where $\underline{A}$ is the node-link incidence matrix of the network; $\underline{q}$ is the link flow vector; and $\underline{p}$ is the origin-destination demand vector.

The solution to this program represents a system-optimized flow pattern and control program. By modify-

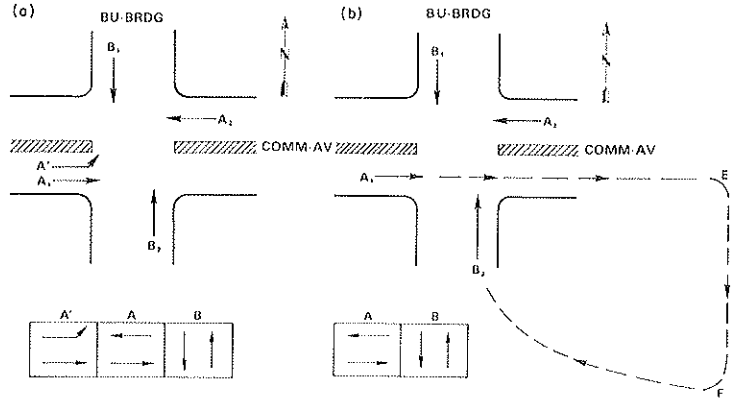


Figure 4. Signalization schemes for example 1.

Table 1. Delays for traffic at three-phase approach intersection.

Approach	Arrival Flow (vehicles/h)	Saturation Flow (vehicles/h)	Green Time/ Cycle Time[a] (cycles)	Degree of Saturation	Average Delay (s)	Total Rate of Delay[b] (vehicles/s²)
A′	240	1500	0.18	0.880	63.0	4.2
A_1	1500	3000	0.54	0.925	30.7	12.8
A_2	1200	4500	0.36	0.750	28.8	9.6
B_1	1050	3000	0.38	0.920	42.0	12.2
B_2	500	3000	0.38	0.440	20.4	2.8

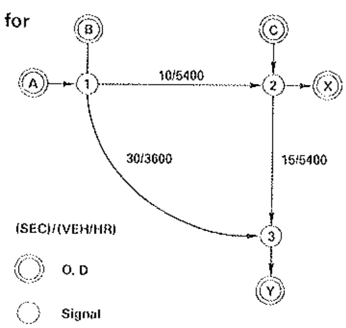
[a]Cycle = 100 s. [b]Level of service = 9 s.

Table 2. Delays for traffic at two-phase approach intersection.

Approach	Arrival Flow (vehicles/h)	Saturation Flow (vehicles/h)	Green Time/ Cycle Time[a] (cycles)	Degree of Saturation	Average Delay (s)	Total Rate of Delay[b] (vehicles/s²)
A_1	1740	4500	0.46	0.85	22.0	10.6
A_2	1200	4500	0.46	0.59	16.0	6.8
B_1	1050	3000	0.43	0.815	22.6	6.5
B_2	740	3000	0.43	0.58	16.7	3.4

[a]Cycle = 100 s. [b]Level of service = 9 s.

Figure 5. Test network for example 2.



ing the objective function, the same program can be used to determine a user-optimized flow pattern. As shown in example 2, comparisons of the two patterns should prove instructive to the traffic engineer in deciding which control procedures and management techniques are helpful in improving traffic flow. Methods of solving such a program can be developed by using modern traffic assignment techniques, such as those based on Frank-Wolfe decomposition (16).

CONCLUSIONS

Traffic equilibrium in an urban road network is the result both of control measures taken by the traffic engineer and of choices made by the individual driver. The examples described in this paper demonstrate that determining control measures consistent with the traffic flow patterns is preferable to calculating control programs and flow patterns independently of each other. The nonlinear optimization program, based on the MITROP optimization model, was presented to achieve this objective on a networkwide basis.

REFERENCES

1. Computerized Traffic Control Systems. Traffic Engineering, Vol. 45, No. 4, Apr. 1975.
2. J. F. Schlaefli. Street Traffic Control by Computer—A State-of-the-Art Discussion. Workshop on Computer Traffic Control, Univ. of Minnesota, Mar. 1974.
3. N. R. Kleinfield. Computerized Traffic Control. The Wall Street Journal, Aug. 13, 1974.
4. Summary of Progress through 1976. NCHRP, 1976.
5. W. E. Pontier, P. W. Miller, and W. H. Kraft. Optimizing Flow on Existing Street Networks. NCHRP, Rept. 113, 1971.
6. R. E. Allsop. Use of Traffic Signals to Influence the Amount and Routing of Traffic. Conference on Getting the Most From Our Transport Facilities—The Role of Traffic Engineering. U.K. Transport and Road Research Laboratory, Crowthorne, England, Apr. 1975.
7. W. D. Berg and P. J. Tarnoff. Research in Urban Traffic Management. TRB, Transportation Research Record 603, 1976, pp. 55-56.
8. R. B. Potts and R. M. Oliver. Flows in Trans-

portation Networks. Academic Press, New York, 1972.
9. M. L. Manheim. Fundamentals of Transportation Systems Analysis. Cambridge, Mass., in press.
10. F. V. Webster. Traffic Signal Settings. British Road Research Laboratory, Her Majesty's Stationery Office, London, BRRL Rept. 39, 1961.
11. N. Gartner. Area Traffic Control and Network Equilibrium. In International Symposium on Traffic Equilibrium Methods (M. Florian, ed.), Springer-Verlag, New York, 1976.
12. F. A. Wagner, F. C. Barnes, and D. L. Gerlough. Improved Criteria for Traffic Signal Systems in Urban Networks. NCHRP, Rept. 124, 1971.
13. L. Rach and others. Traffic Signal Control Strategies—A State-of-the-Art. Toronto Department of Roads and Traffic, Mar. 1974.
14. N. H. Gartner, J. D. C. Little, and H. Gabbay. Simultaneous Optimization of Offsets, Splits and Cycle Time. TRB, Transportation Research Record 596, 1976, pp. 6-15.
15. M. J. Maher and R. Akcelik. The Redistributional Effects of an Area Traffic Control Policy. Traffic Engineering and Control, Sept. 1975.
16. N. H. Gartner. Analysis and Control of Transportation Networks by Frank-Wolfe Decomposition. Proc., 7th International Symposium on Transportation and Traffic Theory, Kyoto, Japan, Aug. 1977.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Abridgment

Optimum Control of Traffic Signals at Congested Intersections

Panos G. Michalopoulos, Rensselaer Polytechnic Institute, Troy, New York

Optimum control of congested intersections continues to be one of the major concerns of traffic engineers. Several theories have been proposed for and successfully applied to optimum control of uncongested intersections, but relatively little theoretical work has been applied to congested or oversaturated conditions. Thus, at present we must rely on practical evidence to time signals at oversaturated intersections.

To be sure, the applicability of existing theories is severely limited because they require complex instrumentation and extensive computations or because the assumptions made in deriving the policy are oversimplified. Further, none of the proposed policies incorporates both of the two necessary criteria for optimum control of oversaturated intersections: (a) minimizing total intersection delay and (b) subjecting (a) to queue length constraints. I therefore conclude that the dependence of current practice on empirical considerations is not entirely unjustified. It can also be shown (1) that some of the practical control policies are in fact best applied under specific conditions.

In order to promote a better understanding of optimum control strategy, I shall first present an isolated congested intersection. Then, I shall extend the theory to a system of two or more intersections.

ISOLATED INTERSECTIONS

Before developing the stragegy, I need to further clarify the first criterion for optimum operation. When minimizing intersection delay at congested intersections, one should consider total delay for the entire oversaturation period rather than delay per cycle. This is so because a per cycle optimum does not, in general, correspond to the optimum policy for the entire peak period. Wrong conclusions, therefore, might be drawn.

The optimum control policy for minimizing total intersection delay was first developed by Gazis and Potts and D'Ans (2, 3, 4, 5). However, the second and in some instances most important objective of the control was not incorporated in the process. Thus, an analytical solution to the problem when queue length constraints are present is still lacking. However, I have adopted Gazis' approach here because it provides a reasonable basis for complete formulation of and solution to the problem.

The oversaturated intersection can be formulated and treated as a control problem in which: (a) the system is the intersection, (b) minimization of the aggregate delay subject to queue length constraints is the objective of the control, (c) the queues on each approach to the intersection describe the state of the system, and (d) cycle length and splits are bounded control variables.

Derivation of the optimum control policy is obtained by first considering the simplest case: an intersection of two one-way streets. It should be noted here that extending this method to more complex situations is straightforward and should not present any problems. The demands on each approach vary with time as in reality and are known for the entire control period. If we assume that demand becomes higher than capacity at roughly the same time on both approaches, that the cycle is constant, and that variable green times are bounded, we can formulate the problem by inspecting the cumulative input-output curves on each approach. The area between these two curves represents the total delay that must be minimized, and the problem is formally stated as follows: Minimize the delay function

$$\text{Min } D = X_0 = \int_0^T (x_1 + x_2)dt$$

and subject it to

$$f_0 = dx_0/dt = x_1 + x_2$$

$$f_1 = dx_1/dt = q_1(t) - u$$

$$f_2 = dx_2/dt = q_2(t) - s_2(1 - L/c) + (s_2/s_1)u$$

$$0 \leqslant i \leqslant \alpha_i$$

with the boundary conditions

$$x_i(0) = x_i(T) = 0$$

$$x_0(0) = 0$$

and the admissible control domain

$$s_1 g_{min}^{(1)}/c = u_{min} \leqslant u = s_1 g_1/c \leqslant s_1 g_{max}^{(1)}/c = u_{max}$$

where D represents total delay, T is the end of the oversaturation period, and c is the cycle length. The remaining parameters are detailed as follows:

- $x_i(t)$ = the queue length on approach i (and i = 1, 2);
- $q_i(t)$ = the input flow rate on approach i;
- $u(t)$ = the control variable = $[s_1 g_1(t)]/c$;
- α_i = the upper bound placed on queue i;
- s_i = the saturation flow on approach i, which is assumed constant;
- $g_i(t)$ = the effective green time on approach i as defined by Webster (6);
- $g_{max}^{(i)}(g_{max}^{(i)})$ = the maximum (minimum) green time allowed on approach i; and
- L = the total lost time per phase.

Space limitations preclude presentation of the details here, but the major results are summarized below (1).

1. When no queue length constraints are imposed, the optimum control policy is essentially the bang-bang control suggested by Gazis (7, p. 213), which, in more concrete traffic terms, implies that the best policy comprises two stages. During the first stage maximum green time should be given to the approach with the highest saturation flow. During the second stage, the operation should be switched so that minimum green time is given to the approach with the maximum saturation flow, maximum green time to the other approach.

2. When queue length constraints are imposed, several switch points may exist. Then the control is not necessarily bang-bang, because it can be shown (1) that the control variable u (and therefore the green times) may receive intermediate values. More specifically, the optimum control is bang-bang as long as none of the queues reaches its upper or lower bound. If this does happen, the optimum control is to switch the signals (to change u) as follows:

a. When the queue on the first approach reaches its upper or lower bound, the optimum control u* is

$$u^*(t) = q_1(t) \qquad (1)$$

b. When the second queue reaches either of its bounds, the control is given by

$$u^*(t) = [s_2(1 - L/c) - q_2]s_1/s_2 \qquad (2)$$

3. Examination of the above relationships leads to the conclusion that, when both queues reach their upper or lower bounds during overlapping time intervals, cycle length must then be free to vary according to the relationship

$$c = L/1 - [(q_1/s_1) + (q_2/s_2)] \qquad (3)$$

4. This implies that when two constraints are imposed, the optimum solution requires a variable cycle in addition to variable splits. If, however, only one constraint is imposed, the cycle can remain constant. Thus an upper and a lower bound should be placed on the cycle for optimum control with queue length constraints.

5. We can now see that when the intersection is not saturated the optimum control policy is to switch the signals as soon as the queue receiving green is exhausted. This policy has, in fact, been proposed by Dunne and Potts (8) and is known as the saturation flow algorithm. To be sure, the operation of traffic-actuated signals is based on this principle. It has been well established from experience that traffic-actuated control results in minimum delays at isolated intersections.

The major problem associated with the optimum control strategy is determining the switch points (if they exist), the final time, and the resulting minimum total delay. Because of the complexity of the problem, one must be content with a numerical solution generated by a digital computer when nonlinearities are introduced (nonlinear time-varying demands). It should be noted that a closed form solution requires analytical expressions describing the history of arrivals that vary from intersection to intersection; this makes the numerical approach even more attractive. I developed an eight-step algorithm allowing the determination of the optimum control policy for a given situation and present examples demonstrating the applicability and the effectiveness of the theory (1).

SYSTEMS OF INTERSECTIONS

In extending the theory to a system of two or more oversaturated intersections, we must take travel times and queuing storage between intersections as well as turning movements into account. In order to simplify the demonstration of this theory, consider a system of two oversaturated intersections of three one-way streets. Making assumptions similar to those for the single intersection and taking into consideration the average travel time d_1 between the two intersections, one can formulate the problem following the guidelines as before. However, the resulting optimum control problem is not trivial because there are two control variables, and the one associated with the upstream intersection appears in the process both without delay and with delay d_1. Space limitations preclude presentation of the details here, but a summary of the optimum control policy to minimize total system delay subject to queue length constraints shows:

1. The entire control period (O, T) should be divided into two intervals (O, T - d_1) and (T - d_1, T), where T represents the end of the control period (or the final time) defined as the time at which the last queue is gone.

2. During both intervals, the resulting optimum control is bang-bang with at most three switchover points per intersection, assuming that none of the queues reaches its upper or lower bound.

3. As soon as the queue between the two intersections reaches its upper or lower bound, the control action taken at the downstream intersection becomes an explicit function of the control action taken at the upstream intersection d_1 time units earlier. This simply means that for optimum control the two intersections must be coordinated. Coordination is more clearly demonstrated for the special case in which turns at the upstream intersection are prohibited. Thus, if $u_1(t)$ and

$u_2(t)$ represent the control action taken at the upstream and the downstream intersections respectively, it can be shown (1) that for optimum control $u_2(t) = u_1)(t - d_1)$, which indicates that in this special case the control action at the downstream intersection should be identical to the control action taken at the upstream intersection d_1 time units earlier.

4. For the remaining queues results similar to those presented for the single intersection can be obtained, because the former are not affected by the time delay.
5. The cycle at each intersection can remain constant as long as no more than one queue per intersection exceeds its bounds at a time. Otherwise, as expected, optimum control becomes more complex, involving variable cycles and splits as the single intersection does.

Extension of the theory to more than two intersections does not present any problem because it is a straightforward application of the pair of intersections. It should be noted, however, that, as the number of intersections increases or when a large number of queue length constraints are imposed, the optimum control policy becomes more complex. In the final analysis, then, it would be too difficult to apply it to a pretimed system. I therefore suggest that, as system and performance requirements increase, closed loop control (computer control) should be considered.

CONCLUDING REMARKS

The theory developed here clearly provides an analytical solution to the problem of optimum control of traffic signals at congested intersections with queue length constraints. The theory can also be employed for computerized control, because initial and final conditions (queues) can be other than zero. It should be pointed out, however, that in this case adaptive prediction algorithms are needed in order to change the control policy as new information concerning the state of the system and the predicted demands arrives. Such algorithms have been developed and used by computer control schemes, but it is generally agreed (9) that further research in this area is needed.

Examples demonstrating the applicability and effectiveness of the theory to real-life situations can be found in another of my works (1), where it is clearly shown that the optimum control, compared with conventional control techniques, can result in substantial delay savings. In these examples, as expected, queue length constraints resulted in a more complex optimum control policy. Further, the fixed time policy was still more effective than conventional control even for demands substantially different from the predicted levels. It is perhaps of interest to note that determining the optimum control strategy requires only knowledge of conventional traffic parameters such as saturation flow; 5-, 10-, or 15-min counts; and maximum allowable green intervals.

To repeat, extending the theory not only should not present any problem, it can also include a large-scale traffic signal network with both critical and uncongested intersections. By using principles of graph and control theory along with eigenvalue analysis, determination of "dominant" intersections (not necessarily oversaturated), defined as those requiring immediate control action at a given instant, can be made. Subsequently, optimum control action will be taken at these intersections alone, eliminating extensive data manipulation and computer computations. This approach should result in instantaneous and efficient decisions made by the computer and time-varying rather than fixed critical intersections. Dominant intersections can be identified by eigenvalue analysis because they are functions of the demand and the control decisions made in the past.

Other problems associated with oversaturated intersections, such as optimum cycle length determination and bus priorities at such intersections, have been studied, and results will be presented in the near future.

REFERENCES

1. P. G. Michalopoulos. Optimal Control of Oversaturated Signal Systems: Some Theoretical Considerations. Univ. of Florida, Gainesville, PhD thesis, 1975.
2. D. C. Gazis and R. B. Potts. The Oversaturated Intersection. Proc. 2nd International Symposium on the Theory of Road Traffic Flow, Organization for Economic Co-operation and Development, Paris, 1965, pp. 221-237.
3. D. C. Gazis. Optimum Control of a System of Oversaturated Intersections. Operations Research, No. 12, 1964, pp. 815-831.
4. D. C. Gazis. Modelling and Optimal Control of Congested Transportation Systems. Networks, No. 4, 1974, pp. 113-124.
5. G. C. D'Ans and D. C. Gazis. Optimal Control of Oversaturated Store-and-Forward Transportation Networks. Transportation Science, Vol. 4, No. 1, 1976, pp. 1-19.
6. F. V. Webster. Traffic Signal Settings. British Road Research Laboratory, Her Majesty's Stationery Office, London, BRRL Rept. 39, 1958.
7. D. C. Gazis. Traffic Control—Theory and Application. In Traffic Science (D.C. Gazis, ed.), Wiley, New York, 1974, pp. 175-237.
8. M. C. Dunne and R. B. Potts. Algorithm for Traffic Control. Operations Research, No. 12, 1964, pp. 870-881.
9. J. B. Kreer. Factors Affecting the Relative Performance of Traffic Responsive and Time of Day Traffic Signal Control. Transportation Research, No. 10, 1976, pp. 75-81.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Simulation and Control of Traffic on a Diamond Interchange

C. H. Knapp and D. Ghosh, Department of Electrical Engineering, University of Connecticut
R. Jain, Springfield Department of Traffic Engineering, Massachusetts

Signaling strategies that control left-turning traffic on a diamond interchange are evaluated by simulation. Although the simulation and control strategies were designed for a particular interchange, each may be adapted to the intersection of other one-way pairs of arterials. The simulation method is briefly discussed, and four related control strategies are presented. Evaluation of these strategies based on average and maximum transit times through the interchange indicates that choosing signal features can significantly profit from the use of short-term average volume information. A flexible strategy based only on queue lengths gives good but not entirely satisfactory performance.

One of the more challenging problems in designing controlled access highways is providing for the interchange of traffic on two or more intersecting facilities. The diamond interchange is now being used for this purpose because it requires less area than the traditional cloverleaf. A major disadvantage of the diamond interchange, however, is that vehicles desiring to make a left turn will, in general, be required to stop. It is therefore important that suitable signaling methods be developed to optimize interchange performance when traffic volumes vary (1).

The object of our research was to simulate the I-291 and Rt-15 diamond interchange in Newington, Connecticut, and to use this simulation to evaluate several methods of controlling left-turning traffic. This interchange, shown to scale in Figure 1, has three levels, one for each throughway and a third for traffic interchange. Simulation has been limited to the interchange level and then only to that portion of traffic turning left. Although the discussion and results presented here apply primarily to this particular interchange, it will be clear that the simulation and signaling strategies can be adapted to the intersections of other pairs of one-way arterials.

Figure 2 shows, schematically, the simulated portion of the interchange and the lane designations used. Each lane is approximately 134 m (440 ft) in length including the input lanes denoted North 01, North 02, West 01, West 02, and so forth. The progress of each vehicle is simulated from "birth" time until it exits at the final intersection. Thus, a vehicle arriving from the south, for example, is assigned to the shorter of queues South 01 and South 02. If South 01 has the shorter queue, the vehicle proceeds successively through lanes South 01, South 2, East 1, and exits at the north intersection.

SIMULATION METHODS

Definition of Terms

For convenience in discussing the material, several definitions of our terms are given as follows.

Interior or circulating lanes—those lanes denoted 1, 2, and 3 in Figure 2 and used by vehicles to accomplish a left turn,

Input lanes—all lanes denoted 01 and 02 used by vehicles to enter the interior lanes,

Occupancy lanes—lanes denoted 11 and 12 (not shown in Figure 2) [these lanes are a portion of input lanes 01 and 02 extending about 92 m (300 ft) from the intersection; occupancy lane queues are used in certain signaling schemes to determine the starting time and duration of the next green interval],

Phases A and B—light sequences experienced by vehicles on the circulating lanes and on the input lanes, respectively,

Intersections north, south, east, or west—intersections in Figure 2 at which vehicles enter respectively on input lanes designated north, south, east, and west,

Transit time for a designated lane—total time spent in the lane, including travel time, waiting time at a red light, if any, and time required to cross the subsequent intersection,

Total transit time for vehicles entering north, south, east, or west—total time spent in the interchange, starting 134 m from the first intersection, terminating when vehicle discharges at the last intersection [thus the transit time for a vehicle entering north is time spent in North 01 (North 02) plus time spent in North 2 (North 3) plus time spent in West 1 (West 2)],

Differential—time interval by which phase B green at one intersection precedes phase B green at the diagonally opposite intersection,

Maximum contents of a designated lane—maximum number of vehicles encountered in the given lane during the course of the simulation, and

Maximum time to leave system—longest total transit time encountered by any vehicle entering from a particular direction.

Traffic Flow Simulation

Two simulations of traffic flow within the interchange have been developed using the general purpose system simulation (GPSS) package (2). Program A maintains the position and velocity of every vehicle within the interchange area of Figure 2. This information is updated every second on the basis of a simple car-following model in the form of a nonlinear equation (3, 4, 5). Vehicles in program B, on the other hand, move between intersections in a given or computed travel time. This avoids the necessity of storing and updating position and velocity and saves substantial computer time. It is more difficult in this case, however, to realistically simulate large variations in traffic volume. All lanes, except occupancy lanes, are 134 m long. In program B, the travel time over each 134-m lane is a random number for each vehicle, uniformly distributed over 12.5 to 15.5 s.

Vehicles moving through an intersection in program A do so according to the car-following model. If a vehicle is required to stop, it will start up again at a maximum specified acceleration. The delay in achieving a steady flow is automatically provided by the model. Start-up delays must be built into program B, however. This has been accomplished by assigning the time in seconds required for each vehicle to pass through an intersection according to the following table.

Figure 1. Scale plan of Rt-15 and I-291 interchange.

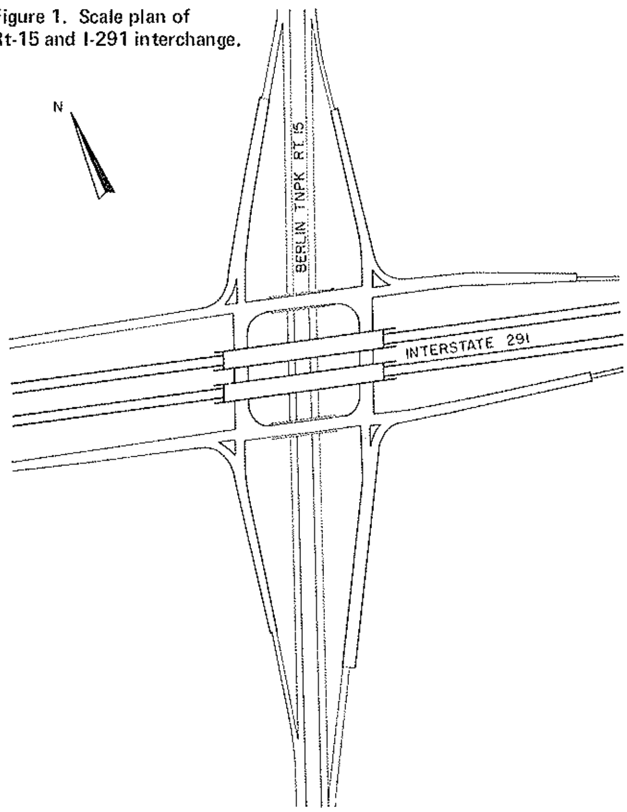
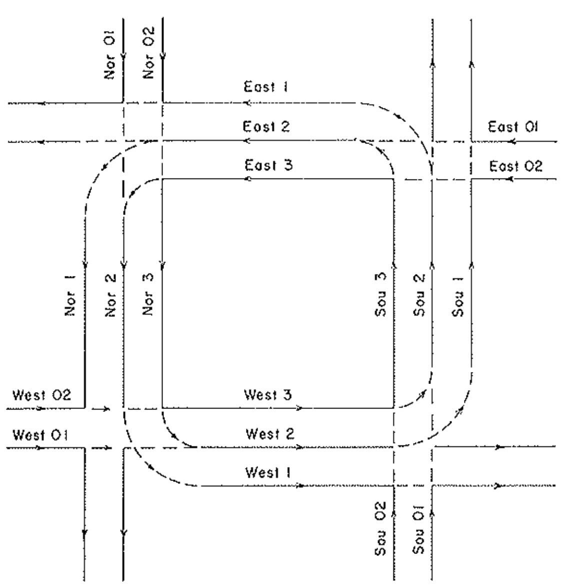


Figure 2. Interchange schematic with lane designations.



Time From Start of Green Phase (s)	Time Required to Cross Intersection (s)
0 to 3	4
3 to 6	3
>6	2

If, for example, three or more vehicles are stopped, the first vehicle will cross the intersection 4 s after the start of the green phase; the second vehicle will follow 3 s later; and subsequent vehicles waiting will follow at 2-s intervals. If, however, the first vehicle in the lane arrives at the intersection 3 to 6 s after the light turns green, it will require only 3 s to cross the intersection, and so on.

Each program maintains records and calculates, at the conclusion of the simulation, the average of the total transit times through the interchange, average transit time through each lane, distribution of transit times, maximum and average contents of each lane, and other aspects. This information is useful for evaluating the performance of different signaling strategies.

Vehicles are generated in both programs by using random numbers to establish the birth time of each subsequent vehicle. The interval between generations can be given any probability distribution desired. A truncated exponential distribution corresponding to a Poisson process with a 1-s mean has been used in all simulation studies thus far. An adjustable multiplier that provides the average time between arrivals necessary to produce any desired traffic volume is then used for each direction.

The car-following model in program A has been adjusted so that statistics for the program correspond well with those of actual traffic on a short span of two-lane highway that includes one intersection. Because program A is less efficient, however, it has been used primarily to check the validity of program B. In general, the two programs are in good agreement except when traffic volumes are unusually heavy, which makes program A give somewhat longer transit times. Because of this agreement, results will be provided for program B only.

Signal Control

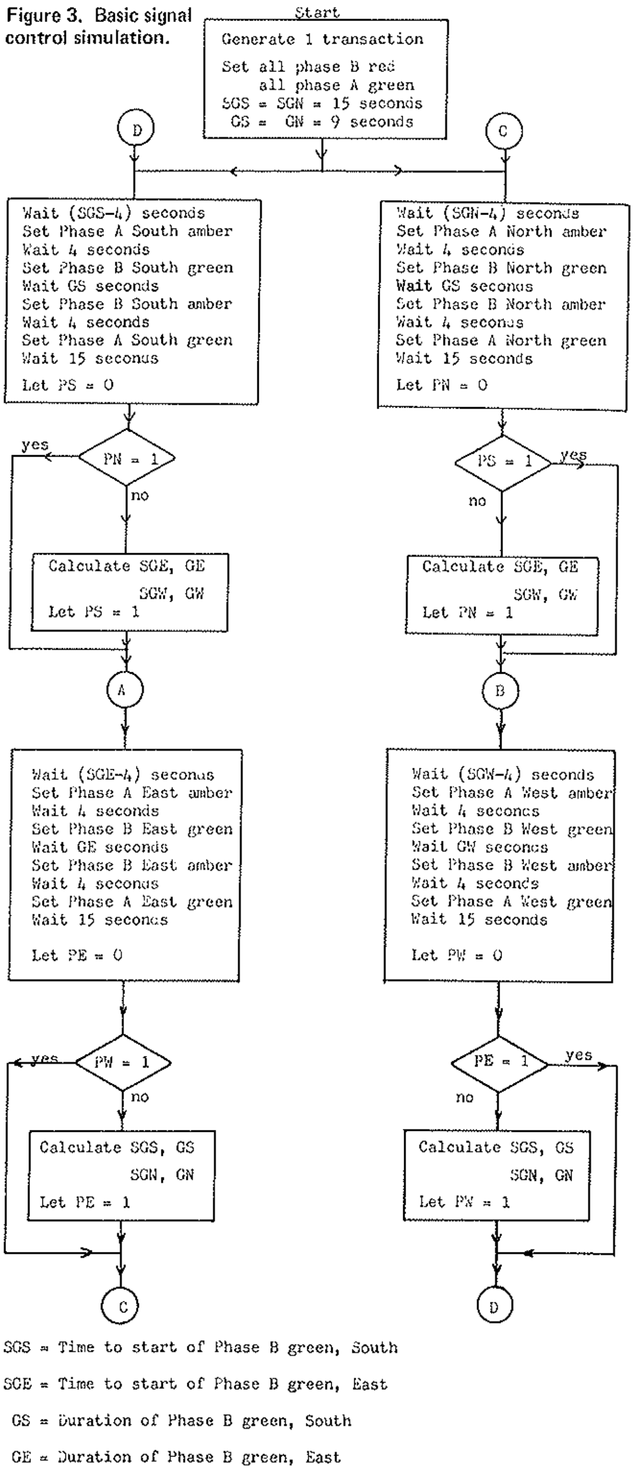
In the GPSS simulation, signals at intersections designated north, south, east, and west are controlled by two transactions that circulate in a loop of instructions. Variations in the signaling strategy are introduced by changing these instructions.

In general, all signal control strategies are designed to allow each vehicle, once it has entered the circulating lanes of Figure 2, to proceed without stopping. Assuming a nominal 15-s travel time between each intersection, this means, for example, that the east phase A green must commence within 15 s after the start of south phase B green. Similarly, north phase A green must begin within 30 s of the start of south phase B green (and vice versa). These requirements place strict constraints on the signal sequence.

Figure 3 is a simplified flow diagram indicating the manner in which traffic signals are controlled in the simulation. At the start of the program, signals are all set at phase A green. Fifteen seconds later, lights at north and south are set at phase B green for 9 s. Following the start of phase A green, north and south, there is a 15-s pause, during which platoons entering from north and south move past the west and east intersections, respectively. Then a decision is made concerning the starting times and durations of the phase B green at east and west. Fifteen seconds after the end of phase B green east or west, whichever occurs first, the starting times and durations of the north and south phase B green are calculated. The process then continues in this manner.

Signaling strategies ranging from a fixed cycle to a flexible traffic responsive strategy can be incorporated by changing the procedure by which starting times and durations of the phase B green intervals are calculated. Although Figure 3 is a simulation flow diagram, it also shows how a special digital controller might function when supplied with information about queue lengths and short-term average traffic volumes measured at the input to each intersection.

Figure 3. Basic signal control simulation.



Omitted from Figure 3, for simplicity, is the task of setting the intersection crossing times according to the previous table, as required in program B. (This is required in the simulation but not in actual implementation of the signaling strategy.)

SIGNALING STRATEGIES

Four signaling strategies have been evaluated, and each is designed to virtually eliminate stopping within the interior lanes. Method A is equivalent to a procedure suggested by the Traffic Engineering Division of the Connecticut Department of Transportation (DOT). The reader will find it helpful to refer to Figure 2 in following the description of these signaling methods.

Method C

This is a simple strategy in which signals at north and south are synchronous, and signals at east and west are synchronous. Phase B green starts simultaneously at north and south and terminates when occupancy lanes North 11 and South 11 are both empty, or after 26 s, whichever occurs first. Allowing for a 4-s amber period, this means that phase A green will commence a minimum of 30 s after the start of phase B green. Fifteen seconds after the start of north and south phase A green, phase B green starts at east and west. A typical phase B sequence is shown in Figure 4 where the 15-s nominal transit time over interior lanes is assumed.

Method B

Method B adds flexibility to method C above by permitting north and south (and east and west) to use different interval lengths for phase B green. Fifteen seconds after phase A is green at both east and west, the number of vehicles in occupancy lanes North 11 and South 11 are observed. If both occupancy lanes are full (13 vehicles or more) or neither is full, phase B north and south act synchronously as in method C. If one lane is full and the other is not, then a differential (Δ) is calculated according to the formula discussed below. Phase B green then starts Δ s early for the longer input queue. In Figure 5, for example, South 11 is full; North 11 is not; and Δ is calculated as 7 s. Phase B green south starts $\Delta = 7$ s before phase B green north and has a duration of $25 + \Delta$ s. Phase B green north lasts for $25 - \Delta$ s. This ensures that phase A green north will begin 29 s (that is, 25 s plus a 4-s amber) after the start of phase B green south, and vice versa.

Fifteen seconds after both north and south have switched to phase A green, a similar calculation is made for the east and west intersections to determine a differential Δ. If $\Delta \neq 0$, then east or west phase B green is delayed Δ s, and so on.

The differential Δ has been calculated in several ways. A formula suggested by the Connecticut DOT is

$$\Delta = \begin{cases} 16 - (15N/10), \text{ when } N < 10 \\ 1, \text{ when } N > 10 \end{cases} \qquad (1)$$

where N is the number of vehicles in the smaller queue. Another expression is obtained by noting that if $N = 2$ or more in the smaller queue and the time required to dissipate an N-vehicle queue is $2N + 3$ s, then it is desirable that

$$\text{green time, full queue/green time, shorter queue}$$
$$= (25 + \Delta)/(25 - \Delta) = [2(13) + 3]/(2N + 3) \text{ or}$$
$$\Delta = [(26 - 2N)/(32 + 2N)] \times 25, \text{ when } N > 2 \qquad (2)$$

Simulation has been carried out using Equations 1 and 2 and similar relations. There seems to be no great sensitivity in the choice of Δ, and results presented are for Equation 1.

Method A

An early start feature may be added to method B to provide improvement in some situations. Suppose, for example, that the queue in lane West 11 is depleted E s before East 11, and that as a consequence phase B green

Figure 4. Typical phase B signal sequence where differential (Δ) is not required.

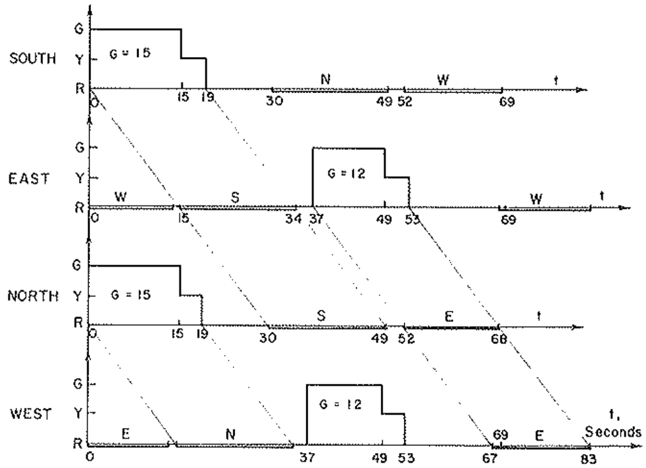
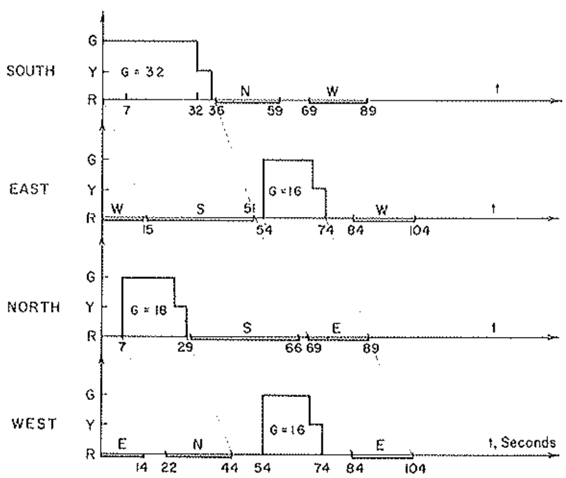


Figure 5. Typical phase B signal sequence where differential of 7 s is required for south.



west is terminated E seconds before east is. Fifteen seconds after the termination of phase B green west, the north-south differential is calculated. If $\Delta > E$ and favors south, then phase B green south is started immediately. If $\Delta < E$ and favors south, then phase B green south starts $E - \Delta$ s later. If $\Delta = 0$ or favors north, then no early start can be used in this case. Thus, when differential $\Delta \neq 0$, the direction favored may be able to start earlier than in method B and thereby to make more efficient use of the interchange. If $\Delta = 0$, no early start is employed.

Method D

Method D is a simple fixed sequence designed to accommodate the given volumes in a more or less optimum fashion. If one examines Figure 5, where $\Delta = 7$ s is used, it will be noted that west phase B green can start at 44 s, rather than wait until 54 s to coincide with phase B green east. Vehicles entering the interior lanes from west at 44 s or after will encounter phase A green 30 s later at east and may therefore exit without stopping. This observation has been used in Figure 6, which illustrates a sequence similar to that of Figure 5 with phase B green west advanced for volumes as in Table 1. The phase B green durations approximate proportionately the volumes of 500, 900, 400, and 300 vehicles/h for north, south, east, and west, respectively.

Observe that the total phase B green is about the same in both figures, but Figure 6 achieves this in 80 s, while Figure 5 requires 89 s. Thus the scheme in Figure 6 is about 10 percent more efficient for uniform traffic flow. Traffic does not arrive uniformly, even in the simulation, so it is questionable whether the fixed sequence of Figure 6 will, in fact, outperform the more flexible strategy of method C even when traffic volumes average those for which the signal sequence of Figure 6 was designed.

SIMULATION RESULTS

Tables 1 through 3 summarize performance for three different volume conditions. All simulations are continued until 1000 vehicles exit from the system. This requires about 30 min (real time) for the volumes of Table 1 and 25 min for Tables 2 and 3. Input volumes in Table 1 correspond to the estimated evening peak volumes for the year 1990 at the I-291 and Rt-15 interchange. Volumes of Tables 2 and 3 are increased over those of Table 1 in directions east and west or both.

Tables 2 and 3, particularly, show the benefits of differential (method B) over purely synchronous signaling (method C). Dramatic reductions in average transit time and maximum time to leave the system result for the high volume directions at small expense in performance to the low volume directions. Some improvement in the longer transit times also results from the early start feature (method A). The improvement afforded by the early start feature is somewhat marginal, however.

Methods A and D of Table 1 provide an interesting comparison. The fixed cycle of method D (based on traffic volume entering in each direction) is clearly superior to that of method A, which uses differential and early start. In fact, the average transit time for all vehicles is about 84 s with method A and 76 s with method D, a reduction of almost 10 percent. Maximum time to leave the system is also superior for the fixed cycle. The implication of this comparison is not that a fixed cycle is best, since the cycle features must be modified in some way as the input volumes vary. Rather, these results suggest that

1. An early start feature can be helpful (as in Figure 6) even when volumes are too low to require differential and
2. There appears to be a significant benefit in terms of average transit times in using (short-term) average volumes for calculating the features of phase B green. If these features are changed in response to existing queue lengths only, control is less effective.

There are at least two ways of employing short-term average volume information. One is to use an algorithm to design a fixed cycle strategy for any given set of volumes; every 5 to 10 min the cycle parameters may be recomputed on the basis of a new set of average volumes. An easier technique to implement is to modify method A so that it incorporates a more elaborate early start procedure that uses average volume information. For example, suppose phase B green west terminates before phase B green east as in Figure 6. Fifteen seconds after the termination of phase B green west the following calculations and decisions can be made:

1. Determine whether the queue South 01 can be depleted within 30 s after the earliest start of phase B green north (this will require knowledge of queue length

and volume of traffic arriving from the south); if not proceed to step 2; if so start south phase B green immediately and continue for the calculated time required to deplete the queue; start phase B green north 15 s after start of phase A green east; and

2. Use current queue lengths and volumes to determine if phase B green south should be longer than north; if so use the early start feature as in method A; if not use the differential feature of method B without early start.

This set of calculations can be expanded. For example, it may be advantageous to start phase B green south immediately, even if the input queue will not be depleted within 30 s after the earliest start of phase B north. That is, it may be better in the long run, depending on average volumes, to terminate phase B green south with vehicles still in the input lanes rather than to delay the start of phase B green south as may be required in step 2.

In order to implement an algorithm of this type, fairly complex decisions and calculations must be made. Also, short-term average volumes must be estimated and stored. Implementation of the algorithm will, therefore, be most readily accomplished by using a special digital or hybrid computer.

The behavior of vehicles in the circulating and input lanes was about as expected. The average transit times in all circulating lanes is about 16 to 17 s (including delay in crossing the intersection at the end of the lane). The distributions of transit times, with volumes as in Table 1, in lanes South 1, South 2, and South 3 are shown in the following table.

Figure 6. Phase B signal sequence for fixed 80-s cycle designed for volumes of Table 1.

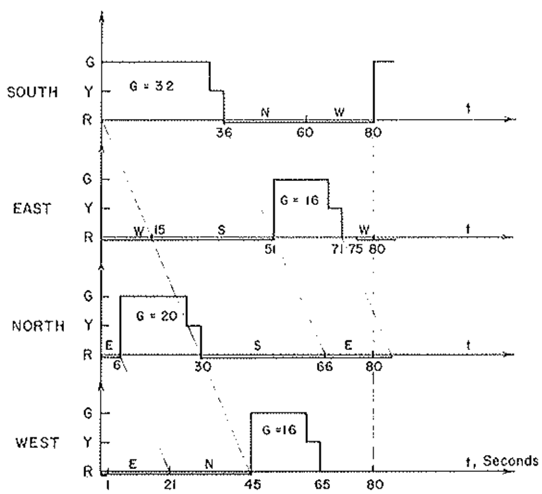


Table 1. Summary of simulation results for projected evening peak input volumes for the year 1990.

		Direction			
Simulation Item	Signal Method	North (avg 500 veh/h)	South (avg 900 veh/h)	East (avg 400 veh/h)	West (avg 300 veh/h)
Average total transit time, s	A	79.9	90.2	81.8	77.0
	B	79.2	93.8	81.7	76.9
	C	76.1	100.8	76.2	74
	D	80.4	73.3	76.4	75.3
Maximum time to leave system, s	A	130	140	140	140
	B	140	150	130	130
	C	120	170	130	120
	D	130	110	120	130
Maximum contents of input lanes, vehicles	A	10	22	12	6
	B	10	19	10	8
	C	10	19	8	6
	D	11	14	7	6
Percentage vehicles leaving in 110 s	A	89.4	77.8	82.8	87.9
	B	91.8	72.9	87.4	88.8
	C	97.0	59.9	86.4	93.1
	D	92.0	100.0	95.0	93.0
Percentage vehicles leaving in 120 s	A	98.4	92.0	97.1	95.4
	B	96.1	83.7	97.9	98.6
	C	100.0	72.7	98.5	100.0
	D	99.0	100.0	100.0	99.0
Percentage vehicles leaving in 130 s	A	100.0	96.7	99.5	99.0
	B	99.5	93.4	100.0	99.2
	C	100.0	83.4	100.0	100.0
	D	100.0	100.0	100.0	100.0

Table 2. Summary of simulation results for input volume set two.

		Direction			
Simulation Item	Signal Method	North (avg 500 veh/h)	South (avg 900 veh/h)	East (avg 400 veh/h	West (avg 720 veh/h)
Average total transit time, s	A	81.6	93.3	81.9	88.5
	B	79.2	99.8	80.9	96.0
	C	73.0	105.8	76.3	157.1
Maximum time to leave system, s	A	130	160	130	150
	B	140	170	140	160
	C	120	180	120	260
Maximum contents of input lanes, vehicles	A	11	25	8	14
	B	11	19	9	17
	C	11	19	8	25
Percentage vehicles leaving in 110 s	A	88.9	71.7	86.6	73.8
	B	87.5	58.2	81.0	62.7
	C	95.5	55.2	93.0	12.9
Percentage vehicles leaving in 120 s	A	97.6	79.7	98.8	90.2
	B	95.6	72.3	94.3	78.5
	C	100.0	63.5	100.0	18.6
Percentage vehicles leaving in 130 s	A	100.0	87.4	100.0	95.5
	B	98.5	82.8	99.3	90.7
	C	100.0	76.5	100.0	22.2
Percentage vehicles leaving in 140 s	A	100.0	93.1	100.0	98.8
	B	100.0	92.8	100.0	94.8
	C	100.0	89.7	100.0	27.2

Table 3. Summary of simulation results for input volume set three.

		Direction			
Simulation Item	Signal Method	North (avg 500 veh/h)	South (avg 900 veh/h)	East (avg 720 veh/h	West (avg 400 veh/h)
Average total transit time, s	A	83.8	94.4	94.5	81.2
	B	82.7	96.2	97.4	88.4
	C	78.9	130.6	132.6	81.0
Maximum time to leave system, s	A	160	160	160	150
	B	140	160	160	140
	C	120	220	220	130
Maximum contents of input lanes, vehicles	A	15	18	18	10
	B	10	18	17	9
	C	11	24	23	8
Percentage vehicles leaving in 110 s	A	83.9	66.8	64.2	80.6
	B	85.3	66.9	69.1	69.2
	C	96.8	21.7	20.2	90.6
Percentage vehicles leaving in 120 s	A	91.9	82.6	81.6	92.7
	B	94.6	78.0	79.7	88.1
	C	100.0	30.8	31.7	99.2
Percentage vehicles leaving in 130 s	A	96.6	91.1	88.7	98.7
	B	99.5	88.8	88.6	97.2
	C	100.0	42.9	39.3	100.0
Percentage vehicles leaving in 140 s	A	97.6	95.4	94.2	99.3
	B	100.0	97.4	94.8	100.0
	C	100.0	55.0	46.3	100.0

Time (s)	Cumulative Percentage (probability distribution x 100)		
	South 1	South 2	South 3
10	0.0	0.0	0.0
20	98.5	95.1	98.1
30	100.0	97.2	100.0
40	100.0	97.2	100.0
50	100.0	100.0	100.0

Thus 95 percent or more of the vehicles did not stop on these lanes (time was less than 20 s). A small number—1 in 70 in South 1, 15 in 290 in South 2, and 4 in 219 in South 3—were required to stop. In all other circulating lanes, the transit time was 30 s or less. Thus, only the slow driver who just manages to enter a circulating lane on an amber light is likely to be caught at the next intersection.

The two tables below indicate behavior in the input lanes for the volumes in Figure 3 or Table 1.

Lane	Average Transit Time (s)	
	Method A	Method D
North 01	48.3	46.6
North 02	46.2	43.7
West 01	48.4	42.8
West 02	42.4	39.7
South 01	58.4	39.4
South 02	56.8	37.2
East 01	51.7	42.5
East 02	47.2	38.6

The first table shows the average transit time for signal methods A and D. Comparison with Table 1 indicates that the improvement in method D over method A in south and east transit times is caused primarily by the reduction in transit times in the input lanes. This, of course, is expected, because circulating lane behavior was essentially the same for all methods.

Time (s)	Cumulative Percentage	
	Signal Method A	Signal Method D
30	8.7	33.4
40	21.3	48.2
50	35.3	66.9
60	48.0	86.0
70	68.9	99.1
80	83.8	100.0
90	92.5	
100	96.5	
110	100.0	

The second table indicates that the maximum time in the input lane South 01 was 80 s for method D, while about 16 percent of the vehicles required longer than 80 s with method A.

SUMMARY

Simulation of left-turning traffic in the I-291 and Rt-15 diamond interchange indicates that the interchange will efficiently service the peak evening traffic volumes projected for the year 1990. Average transit times through the interchange are typically 80 s, with about 50 s required for travel time and 30 s for waiting at the initial intersection. In most cases, vehicles do not need to stop once the circulating lanes are entered. Even the higher volumes of Tables 2 and 3 do not produce serious delays.

The superior performance of method D, which requires average volume data, has suggested further refinements in signaling scheme A, which uses queue length information only. Although implementation of signal control algorithms based on such refinements will likely require a digital or hybrid computer, this is consistent with the trend in signal control at major intersections.

ACKNOWLEDGMENT

This study was supported by the Connecticut Research Foundation and the Connecticut Department of Transportation.

REFERENCES

1. S. Spitz. Signalization of Diamond Interchanges. Traffic Engineering, Vol. 34, No. 7, 1964, pp. 15-17.
2. General Purpose Simulation System/360, Introductory User's Manual. IBM Application Program GH20-0304-4, 5th Ed., Nov. 1969.
3. G. Newell. Nonlinear Effects in the Dynamics of Car Following. Operations Research, Vol. 9, 1961, pp. 209-229.
4. D. C. Gazis, R. Herman, and R. W. Rothery. Nonlinear Follow-the-Leader Models of Traffic Flow. Operations Research, Vol. 9, 1961, pp. 545-567.
5. D. C. Gazis and L. C. Edie. Traffic Flow Theory. Proc., IEEE, Vol. 56, No. 4, 1968, pp. 458-471.

Publication of this paper sponsored by Committee on Traffic Control Devices.