881

TRANSPORTATION
RESEARCH RECORD 881

Traffic Control Devices and Traffic Signal Systems

TRB
TRANSPORTATION RESEARCH BOARD
NATIONAL ACADEMY OF SCIENCES

The **Transportation Research Record** series consists of collections of papers in a given subject. Most of the papers in a Transportation Research Record were originally prepared for presentation at a TRB Annual Meeting. All papers (both Annual Meeting papers and those submitted solely for publication) have been reviewed and accepted for publication by TRB's peer review process according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The views expressed in these papers are those of the authors and do not necessarily reflect those of the sponsoring committee, the Transportation Research Board, the National Academy of Sciences, or the sponsors of TRB activities.

Transportation Research Records are issued irregularly; approximately 50 are released each year. Each is classified according to the modes and subject areas dealt with in the individual papers it contains. TRB publications are available on direct order from TRB, or they may be obtained on a regular basis through organizational or individual affiliation with TRB. Affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Academy of Sciences, 2101 Constitution Avenue, N.W., Washington, DC 20418.

TRANSPORTATION RESEARCH RECORD 881

Traffic Control Devices and Traffic Signal Systems

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1982

Transportation Research Record 881
Price $10.40
Edited for TRB by Susan Singer-Bart

mode
1 highway transportation

subject area
54 operations and traffic control

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Traffic control devices and traffic signal systems.

(Transportation research record; 881)
1. Traffic signs and signals—Addresses, essays, lectures.
I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 881 [TE228] 380.5s [388.3'122] 83-4227
ISBN 0-309-03468-X ISSN 0361-1981

Sponsorship of the Papers in This Transportation Research Record

GROUP 3–OPERATION AND MAINTENANCE OF TRANSPORTATION FACILITIES
Patricia F. Waller, University of North Carolina at Chapel Hill, chairman

Committee on Traffic Control Devices
Ken F. Kobetsky, West Virginia Department of Highways, chairman
H. Milton Heywood, Federal Highway Administration, secretary
Robert Dewar, Paul Fowler, Robert M. Garrett, Alan T. Gonseth, Robert L. Gordon, Robert David Henry, Gerhart F. King, Kay K. Krekorian, Hugh W. McGee, Zoltan A. Nemeth, Peter S. Parsonson, Harold C. Rhudy, H. Douglas Robertson, Edward F. Smith, Ronald E. Stemmler, James I. Taylor, James A. Thompson, Clinton A. Venable, Earl C. Williams, Jr., Walter P. Youngblood

Task Force on Traffic Signal Systems
Adolf D. May, Jr., University of California, chairman
William D. Berg, Kenneth G. Courage, J. R. Doughty, H. Milton Heywood, E. Vinson Hoddinott, Jack L. Kay, David Koski, Joseph K. Lam, C. John MacGowan, Peter S. Parsonson, Raymond S. Pusey, Philip J. Tarnoff, Frederick A. Wagner, Roy L. Wilshire

David K. Witheford, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units, officers, and members are as of December 31, 1981.

Contents

Authors of the Papers in This Record

Courage, Kenneth G., College of Engineering, University of Florida, Gainesville, FL 32611
Davis, Christian F., University of Connecticut, Storrs, CT 06268
Gartner, Nathan H., College of Engineering, University of Lowell, Lowell, MA 01854
Hurley, Jamie W. Jr., Department of Civil Engineering, Memphis State University, Memphis, TN 38152
Kahng, Suk June, Institute of Transportation Studies, University of California at Berkeley, 109 McLaughlin Hall, Berkeley, CA 94720
Lee, Clyde E., Department of Civil Engineering, University of Texas at Austin, Austin, TX 78712
Lin, Feng-Bor, Department of Civil and Environmental Engineering, Clarkson College, Potsdam, NY 13676
Lum, Harry S., Federal Highway Administration, U.S. Department of Transportation, Washington, DC 20590
Lyles, Richard W., Department of Civil and Sanitary Engineering, Michigan State University, East Lansing, MI 48824; formerly with the University of Maine at Orono
Machemehl, Randy B., Department of Civil Engineering, University of Texas at Austin, Austin, TX 78712
Mao, Andrew C.M., Texas State Department of Highways and Public Transportation, Highway Building 11th and Brazos, Austin, TX 78701
May, Adolf D., Institute of Transportation Studies, University of California at Berkeley, 109 McLaughlin Hall, Berkeley, CA 94720
Messer, Carroll J., Texas Transportation Institute, Texas A&M University, College Station, TX 77843
Mullowney, Wiliam L., Bureau of Transportation Systems Research, Division of Research and Demonstration, New Jersey Department of Transportation, 1035 Parkway Avenue, Trenton, NJ 08625
Parsonson, Peter S., School of Civil Engineering, Georgia Institute of Technology, 225 North Avenue NW, Atlanta, GA 30332
Radwan, A. Essam, Department of Civil Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061
Rogness, Ramey O., North Dakota State University of Agriculture and Applied Science, Fargo, ND 58105
Ryan, Timothy A., Kidde Consultants, Inc., 1020 Cornwell Bridge Road, Baltimore, MD 21204
Stockton, William R., Texas Transportation Institute, Texas A&M University, College Station, TX 77843
Todd, Kenneth, 1954 Columbia Road NW, Washington, DC 20009
Wallace, Charles E., College of Engineering, University of Florida, Gainesville, FL 32611
Williams, Jim, Texas State Department of Highways and Public Transportation, Highway Building 11th and Brazos, Austin, TX 78701
Wu, Ching-Shuenn, Department of Civil Engineering, University of Texas at Austin, Austin, TX 78712

Effects of Multiple-Point Detectors on Delay and Accidents

CHING-SHUENN WU, CLYDE E. LEE, RANDY B. MACHEMEHL, AND JIM WILLIAMS

The number and location of detectors on intersection approaches that have actuated signal controllers and high traffic approach speeds have been studied by various researchers. The relation of detector activity to yellow signal intervals and the presence of dilemma zones has also been investigated. Several procedures for locating multiple detectors on problematic intersection approaches have been proposed as solutions to dilemma zone and other traffic control problems. Four multiple-detector placement methods are compared through computer simulation in a relative evaluation of their effects on vehicular delay. Traffic performance statistics produced through computer simulation are compared with those obtained through field observation. Conventional single-point detection schemes are also compared with multiple detectors through before-and-after field tests at 10 typical field sites. The four methods for placing multiple detectors were not found to produce statistically significant differences in vehicular delay when compared with each other or with single-point detection. Multiple detectors were found to produce statistically significant reductions in accident experience for approach speeds of 50 mph or greater.

Actuated traffic signal controllers use current traffic information to vary signal timing in response to actual traffic demand. The required real-time traffic data are acquired by detectors that are designed and located to fit each particular geometric and traffic situation. The most widely used type of vehicle detector is the inductance loop. This detection system is highly adaptable in that the size and shape of the in-road sensing device can be designed to suit most traffic control needs. Conventional installations generally use a single loop on each inbound intersection approach.

This single-loop (or single-point) detection system has the potential to cause problems for drivers who must respond to the yellow signal indication at intersections where speeds of approaching traffic are greater than approximately 35 mph. Certain combinations of high approach speeds, detector location, and controller timing make it difficult for the driver to determine whether to stop or proceed through the intersection from certain locations in advance of the intersection after the appearance of the yellow indication. These locations constitute a dilemma zone or a zone of complex risk evaluation for the driver. Also, under moderate-to-light traffic conditions, controllers that use the single-point detection scheme and are timed for heavy traffic may allow frequent loss of green due to "gapping out," and thus present more yellow intervals and more opportunities for wrong driver decisions. Erratic signal controller operation associated with premature gapping out is frequently cited as an indication of inefficient operation. Such inefficiency might be responsible for unnecessary vehicular delay and increased accident potential.

Various detector placement and controller timing schemes have been proposed for solving these problems at intersections that have high approach speeds (1). Several detection schemes are mentioned in the following paragraphs, and one series, which is referred to as multiple-point detection, is examined in detail. Results of theoretical analysis, simulation, and field evaluation are presented as bases for evaluating four multiple-point detection methods and the potential that they might provide for reducing accidents and improving signal operating efficiency.

DILEMMA ZONE OR ZONE OF COMPLEX RISK EVALUATION

The principal justification for using multiple detectors and special controller timing on high-speed intersection approaches is to prevent, whenever possible, the yellow signal indication from being initiated when a vehicle is within what has been called the dilemma zone or zone of indecision (1). When the signal indication displayed to traffic approaching an intersection changes from green to yellow, drivers must decide immediately whether to stop before entering the intersection or to continue through the intersection without stopping. This requires that each driver evaluate a number of specific time, distance, velocity, and acceleration parameters during perception-reaction time as the vehicle continues toward the intersection. The driver must weigh the risks associated with stopping against those associated with continuing. The action decided on during perception-reaction time will presumably involve the least overall risk in the judgment of the driver.

For a given approach speed, the relative risk involved with a decision to stop or to continue varies with the distance from the intersection at which the approaching vehicle is located when the yellow indication begins as well as with the duration of the yellow indication. When the distance is large, a stop can be accomplished easily with a low rate of deceleration and therefore low risk of skidding or being hit from the rear, but continuing involves a long travel time to the intersection and a high risk of not being able to clear the intersection during the yellow. The time needed to stop is not important. At locations closer to the intersection, the risks related to stopping increase, and those associated with continuing decrease, thereby making the driver's task of risk evaluation more complex. When the vehicle is near the intersection at the onset of yellow, a decision to stop requires a high rate of deceleration with the associated high risks, but a decision to continue allows the vehicle to enter or clear the intersection during a yellow indication of normal duration with low risk. Outside the zone described by these bounding distances, the low-risk decision is obvious. Most drivers will be able to choose the proper action easily, but within the zone, the problem of choosing the low-risk alternative action is complicated. If the traffic engineering objective of eliminating the need for a driver decision under these difficult and complex circumstances is to be realized, the nature and extent of the zone must be defined in descriptive terms.

In analyzing driver response to the yellow signal, May (2) described the zone in which a vehicle could be located at the onset of yellow whereby it could neither stop safely nor clear the intersection during the yellow interval as a dilemma zone. This is the conventional use of the term dilemma--a situation involving choice between equally unsatisfactory alternatives (3). The term option zone was used to describe situations in which the yellow interval was long enough to allow vehicles to either stop safely or to clear the intersection during the yellow.

In a technical report concerning detector-controller configurations that use small-area detectors, Parsonson and others (4) coined an arbitrary definition of dilemma as a probability of stopping of more than 10 percent but less than 90 percent and described the dilemma zone as the range of distance from the intersection within which drivers are often indecisive. This terminology, which has been used in a number of technical papers since 1974, deviates from the conventional usage of the word dilemma and describes drivers who are forced to respond to a yellow signal indication as often being indecisive. There is little evidence to indicate that drivers are indecisive in this situation. When confronted with the same circumstances drivers may vary in which of the two available alternatives they choose, but every driver makes a decision. The concept of using probabilities to delineate the zone in which risk evaluation is evidently a complex task for drivers to perform when facing a yellow signal is commendable. Zegeer's experimental work (5) extends that of Parsonson and others (4) and interprets the observations of several others in support of the concept. Understanding of this concept would possibly be facilitated if nomenclature other than dilemma zone were used. Descriptive terminology such as zone of complex risk evaluation or zone of varying probability of stopping would be more cumbersome but would probably be more accurately interpreted by traffic engineers.

The zone of concern can be delineated adequately for detector placement purposes in various ways. The special detector placement and signal timing schemes that are evaluated in this report attempt to recognize the existence of areas on intersection approaches where driver decisionmaking is problematic and relieve the problem by controlling the timing of the onset of the yellow signal indication.

DETECTOR PLACEMENT METHODS

Three types of special detector placement methods have been developed in recent years and used at a number of locations around the country (1). Another innovative development is described by Parsonson and others (6), but this new system has not yet been used widely. The three techniques listed below use conventional hardware and have been installed at several sites. These methods include the following:

1. Green extension systems for semiactuated controllers,
2. Extended-call detection systems, and
3. Multiple-point detection systems for basic controllers, such as the Beirele method, Winston-Salem method, and SSITE method.

A comprehensive description of each of these methods is given by Sackman and others (1), and a flow chart to guide in selecting detector-controller configurations for specific situations is included. Detailed examples of calculating proper detector locations and controller settings for various high-approach-speed intersections are presented.

The first two detector placement methods ordinarily use two inductive loop detectors with extended-call timing features and do not directly allow for large variations in approach speed by sensor location. With these systems, higher speeds tend to lengthen the dilemma zone and make the effects of timing much more critical as approach speeds and traffic volumes vary.

A number of multiple-point detector installations have been made at intersections in Texas where approach speeds are high. The Beirele method was used to design most of the systems, but a modification was made to the basic method by the Texas State Department of Highways and Public Transportation (TSDHPT) for some locations. It was desirable to evaluate the relative effectiveness of these installations and compare them with similar multiple-detector methods. The four multiple-detector methods that were included in the study are described elsewhere (1) and are presented in outline form below.

Beirele Method

The Beirele method of multiple detector placement uses a 1-s vehicle interval setting on a basic controller operating in the locking detector memory mode. Each detector is located in advance of the intersection at a distance that is at least adequate for a driver who receives a yellow indication at that point to react and stop safely from an assumed speed. Safe stopping sight distances are based on a 1-s perception-reaction time plus braking distances that result from coefficients of friction between 0.41 and 0.54 for speeds between 55 and 20 mph. The outermost, or first, detector is placed at safe stopping sight distance from the intersection for full approach speed. The next detector is tentatively located at safe stopping sight distance from the intersection for a speed assumed to be 10 mph less than that used for locating the first detector. If the travel time for a passenger car between the two presence-mode loop detectors (6x6-ft size) is greater than 1 s, the downstream detector is relocated to allow the vehicle to reach it during the 1-s vehicle interval set on the controller. This location procedure is repeated for each successive detector until the last loop is 75 ft from the intersection. Minimum assured green time is set on the controller to allow vehicles stored between the last detector and the intersection to enter the intersection. Recommended locations of detectors for different speeds are shown in Table 1. Beirele suggests the addition of special speed detection features for approach speeds above 50 mph.

Winston-Salem Method

The Winston-Salem method was developed by Holloman in 1975. The principles used in the method are basically the same as those used by Beirele; however, the differences between the methods are as follows:

1. This method uses slightly different stopping distances,
2. This detector location procedure starts with placement of the innermost detector and works outward, and
3. This method is suggested for speeds up to 60 mph.

Detector locations for three speeds are shown in Table 1.

SSITE Method

The SSITE method of detector placement was described initially in a report by the Southern Section of the Institute of Traffic Engineers (ITE) in 1976 (7). Basically, this method uses both an iterative process and engineering judgment to locate the inductive loops. Detectors are connected in a series-parallel arrangement and operated in the presence mode with a nonlocking controller. Six detectors are used in an attempt to provide detection along the full length of the approach from the intersection to the outer limit of the dilemma zone as defined by Parsonson and others (4). The outer two detectors provide protection for high approach speeds, and the inner four detectors are positioned to allow for reduced speed nearer the intersection and to provide for queue discharge without premature gap-out. A vehicle interval is set in the con-

Table 1. Detector spacing.

		Detector Spacing (ft)[a]					
Method	Speed (mph)	Stop Line-1st Detector	1st-2nd Detector	2nd-3rd Detector	3rd-4th Detector	4th-5th Detector	5th-6th Detector
Beirele	30	48	39				
	40	48	39	58			
	50	52	48	62	76		
Winston-Salem	30	86					
	40	86	61				
	50	86	61	69			
SSITE	30	0	15	31	43	74	
	40	0	15	25	74	106	
	50	0	15	25	45	105	124

[a]Detector spacing is measured upstream from stop line; loop size is 6x6 ft.

Table 2. Loop layout for TSDHPT modified Beirele method.

	Inductive Loop Layout (ft)[a]						
	Without Optional Detector			With Optional Detector			
Speed (mph)	Stop Line-1st	1st-2nd	2nd-3rd	Stop Line-1st	1st-2nd	2nd-3rd	3rd-4th
30	108			55	47		
40	108	64		55	47	64	
50	108	64	83	55	47	64	83

[a]Inductive loop layout is measured upstream from stop line; loop size is 6x6 ft.

troller to hold the green as vehicles pass between successive detectors. Spacing between the inductive loops is shown in Table 1.

TSDHPT Modified Beirele Method

In addition to the three multiple-point detector placement methods described above, a location technique developed and tested by the TSDHPT was also studied. The concept is similar to the Beirele method. A 1-s perception-reaction time is included in stopping distance calculations, but braking distance computations use American Association of State Highway and Transportation Officials (AASHTO) assumed speeds and coefficients of stopping friction (8). In the basic method the closest detector is located 114 ft from the intersection, but a further modification locates an optional detector 61 ft from the stop line. The addition of the optional detector has the effect of reducing the required initial interval (minimum assured green) and possibly improving operational efficiency. The loop layouts for speeds of 30, 40, and 50 mph by using the TSDHPT modified Beirele method are shown in Table 2.

Differences in Detector Placement Methods

Controller type, detector mode, applicable speed range, loop layouts, and allowable gap for each of the basic multiple-detector placement methods are summarized in Table 3. A look at the table indicates that the major differences among these methods are (a) number of inductive loops used and (b) inductive loop spacings.

The length of the zone of complex risk evaluation, or the dilemma zone, becomes larger as approach speed increases; therefore, more detectors are required to trace a vehicle through the zone. In addition, the longer the spacing between successive loops, the longer the required vehicle interval and the less efficient the controller is likely to be. So, in general, multiple-detection systems are more appropriate for signalized intersections that have high-speed traffic on the approaches.

COMPARISON OF MULTIPLE-POINT DETECTION METHODS BY SIMULATION

Although the theoretical potential of multiple-point detection systems for solving driver-decision problems is fairly clear, the actual effects on vehicular delay and accident experience are not well documented. Field observation is costly and time consuming. In order to study the relative differences in vehicular delay that can result from four detector placement methods, an experiment that used computer simulation was conducted. A factorial experiment design was developed to evaluate each placement method at three volume levels and three speed levels for both diamond interchange and four-leg intersection geometric configurations. A schematic representation of the factorial experiment design is presented in Table 4. The mathematical model to be employed in the analysis of variance is

$$Y_{ijK} = \mu + M_i + S_j + MS_{ij} + V_K + MV_{iK} + SV_{jK} + E_{ijK} \quad (1)$$

where

Y_{ijK} = predicted average delay;
μ = grand mean;
M_i = placement method i = 1, 2, 3;
S_j = approach speed j = 1, 2, 3;
V_K = lane volume K = 1, 2, 3;
MV_{iK} = interaction between M and V;
SV_{jK} = interaction between S and V;
MS_{ij} = interaction between M and S; and
E_{ijK} = error term.

No replication is provided in the basic experiment, therefore, the possible three-way interactions are confounded with the error term. Six slightly different types of vehicular delay were tested separately as the dependent variable in the basic experiment. They include the following:

1. Average total delay for all vehicles,
2. Average queue delay for all vehicles,
3. Average stopped delay for all vehicles,
4. Average total delay per delayed vehicle,
5. Average queue delay per vehicle incurring queue delay, and
6. Average stopped delay per vehicle incurring stopped delay.

Total delay is measured as the difference between actual travel time and the travel time required if the vehicle maintains a prespecified desired speed. Queue delay is accumulated only when a vehicle is part of a queue on the intersection approach. A vehicle is said to be in a queue when it is less

Table 3. Summary of detector placement methods.

Design	Green Extension Systems for Semiactuated Control	Extended Call Detector Systems for Basic Controller	Multiple Detection Systems		
			Beirele Method	Winston-Salem Method	SSITE Method
Controller type	Nonlocking type	Nonlocking type	Locking type	Locking type	Nonlocking type
Detector mode	Presence	Presence	Presence	Pulse	Presence
Speed range	V = 85th percentile speed	V = 85th percentile speed	$V \leq 50$	$V \leq 60$	$V \leq 60$
Loop layout					
Outermost loop[a]	$D = 1.47Vt + (V^2/30f)$	$D = 1.47t + (V^2/30f)$	Use stopping distance from intext driver testing	Use stopping distance from Traffic Engineering Handbook	Use SSITE Report
Innermost loop	$D_1 = 1.47V[(V/30) + 1]$	0	48 or 69 ft	86 ft	0 ft
Spacing between loops[b]	$(D - D_1)/V > 2$ s	$(D - 70)/V_{low\ limit} > 2$ s	1 s	1 s	2 s
No. of loops[c]	2 or 3	2	$(V/10) - 1$	$(V/10) - 2$	≤ 6
Allowable gap	5∿6 s	5∿6 s	2∿5 s	2∿5 s	5∿7 s

[a]Distance is measured from the stop line to the upstream end of the loop.
[b]$V_{low\ limit}$ = low speed limit, for example 15th percentile speed.
[c](V/10) represents the integer part of V/10, for example (3.5) = 3.

Table 4. Factorial design.

Method	Speed (mph)	Lane Volume [(vehicles/h)/lane]		
		300	500	700
Diamond-interchange and four-leg intersection				
Beirele	30	$A_1B_1C_1$[a]	$A_1B_1C_2$	$A_1B_1C_3$
	40	$A_1B_2C_1$	$A_1B_2C_2$	$A_1B_2C_3$
	50	$A_1B_3C_1$	$A_1B_3C_2$	$A_1B_3C_3$
Winston-Salem	30	$A_2B_1C_1$	$A_2B_1C_2$	$A_2B_1C_3$
	40	$A_2B_2C_1$	$A_2B_2C_2$	$A_2B_2C_3$
	50	$A_2B_3C_1$	$A_2B_3C_2$	$A_2B_3C_3$
SSITE	30	$A_3B_1C_1$	$A_3B_1C_2$	$A_3B_1C_3$
	40	$A_3B_2C_1$	$A_3B_2C_2$	$A_3B_2C_3$
	50	$A_3B_3C_1$	$A_3B_3C_2$	$A_3B_3C_3$
Study of optional detector in TSDHPT modified Beirele method				
Without optional detector	30	$A_1B_1C_1$[a]	$A_1B_1C_2$	$A_1B_1C_3$
	40	$A_1B_2C_1$	$A_1B_2C_2$	$A_1B_2C_3$
	50	$A_1B_3C_1$	$A_1B_3C_2$	$A_1B_3C_3$
With optional detector	30	$A_2B_1C_1$	$A_2B_1C_2$	$A_3B_1C_3$
	40	$A_2B_2C_1$	$A_2B_2C_2$	$A_2B_2C_3$
	50	$A_2B_3C_1$	$A_2B_3C_2$	$A_2B_3C_3$

[a]$A_1B_1C_1$ is method 1 (Beirele) when speed at first level (30 mph) and lane volume at first level [(300 vehicles/h)/lane].

than a specified distance (4-40 ft) from the stop line (for the first driver-vehicle unit in the lane) or from the driver-vehicle unit ahead and is traveling less than 3 ft/s. Stopped delay is accumulated when a driver-vehicle unit is stopped or traveling at a velocity less than 3 ft/s. Each of these types of delay is routinely calculated by the TEXAS computer simulation model.

Computer Simulation

The TEXAS model (9,10) which was developed at the Center of Highway Research, University of Texas at Austin, was selected as the most suitable traffic simulation model for this investigation. This model is comprised of three major component programs:

1. Presimulation geometry processor,
2. Presimulation driver-vehicle processor, and
3. Simulation processor.

Both the geometry processor and the driver-vehicle processor are supportive programs for the simulation processor. The outputs from these two programs serve as the input for the simulation processor. The input for the geometry processor includes a detailed description of intersection geometrics and the inputs for the driver-vehicle processor characterize the individual drivers and vehicles that operate in the traffic stream. The additional inputs for the simulation processor include (a) simulation time parameters, (b) car-following parameters, (c) signal timing parameters, and (d) detector location and operating mode information. In the simulation the program sequentially examines each driver-vehicle unit in the intersection system and allows each to respond to surrounding traffic and traffic control devices and predicts its position, speed, and acceleration in the next increment of simulation time. Each unit is thus stepped through the intersection in small time increments. Delay, speed, and volume statistics are accumulated through the simulation process and reported at the end of a selected time period.

Analysis of Simulation Results

Analysis of variance was used to evaluate the significance of effects on the dependent variables produced by each delay-related factor. The null hypothesis stated that the effect produced by each factor on the delay statistics was not significant at a 5 percent level of significance. If the probability associated with the calculated F-statistics was found to be less than 0.05, then the null hypothesis could be rejected. Table 5 summarizes the values of significance of F from 48 analyses of variance. From this table, the following statements can be made.

In both the diamond-interchange and the four-leg intersection study neither detector placement method nor approach speed has a significant effect on the average delay experience by all vehicles that use the intersection or on the average delay experienced by only the delayed vehicles at the 5 percent level of significance. Lane volume, however, has a significant effect on both types of average delay at a 5 percent level of significance.

Analysis of the TSDHPT modified Beirele method with and without an optional detector at a four-leg intersection produced a basis for the following conclusions:

1. The option does not produce significant effects on either type of average delay at a 5 percent level of significance;
2. Approach speed does not have significant effect on either type of average delay when all approaches are analyzed together; however, it produces significant effects at a 5 percent level of signifi-

cance when individual approaches are tested; and

3. Lane volume produces significant effects on both types of average delay at a 5 percent level of significance.

FIELD INVESTIGATIONS OF SINGLE-POINT AND MULTIPLE-POINT DETECTION

In addition to the simulation-based study of delay associated with various detector placement methods, a series of field observations were used to compare the effects of multiple-point with single-point detection. Ten test sites located in Texas that have actuated signal controllers and relatively high approach speeds were selected.

At each test site existing single detectors were replaced by multiple units on selected approaches that were deemed to be most problematic. The place-

Table 5. Significance of F.

		Intersection Geometry							
		Diamond Interchange		Four-Leg Intersections[b]			Optional Detector in TSDHPT Modified Beirele Method		
Delay Measure	Delay	All Approaches[a]	Major Approaches	All Approaches	Major Approaches 1	Major Approaches 2	All Approaches[a]	Major Approaches 1	Major Approaches 2
Avg total delay	Per approach vehicle								
	Detection method	0.103	0.133	0.091	0.251	0.353	0.828	0.199	0.949
	Speed	0.191	0.433	0.269	0.400	0.473	0.231	0.076	0.002
	Lane volume	0.007	0.001	0.007	0.001	0.001	0.001	0.002	0.001
	Per delayed vehicle								
	Detection method	0.090	0.170	0.08	0.252	0.363	0.799	0.176	0.764
	Speed	0.269	0.471	0.34	0.367	0.442	0.233	0.004	0.002
	Lane volume	0.001	0.001	0.01	0.001	0.001	0.001	0.001	0.001
Avg queue delay	Per approach vehicle								
	Detection method	0.119	0.145	0.080	0.209	0.614	0.219	0.108	0.324
	Speed	0.189	0.430	0.340	0.362	0.384	0.230	0.008	0.001
	Lane volume	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Per delayed vehicle								
	Detection method	0.208	0.427	0.070	0.275	0.259	0.997	0.159	0.945
	Speed	0.843	0.583	0.245	0.161	0.229	0.100	0.004	0.093
	Lane volume	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Avg stop delay	Per approach vehicle								
	Detection method	0.119	0.282	0.089	0.246	0.285	0.878	0.339	0.338
	Speed	0.598	0.459	0.220	0.363	0.353	0.715	0.023	0.001
	Lane volume	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Per delayed vehicle								
	Detection method	0.135	0.578	0.083	0.360	0.270	0.410	0.009	0.001
	Speed	0.968	0.593	0.163	0.195	0.237	0.410	0.009	0.001
	Lane volume	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

[a]Only major street approaches receive multiple detectors. [b]Simulation conducted only for four-leg intersection.

Table 6. Location and spacing of multiple detectors.

	Detector Spacing (ft)			
Intersection Approach	Stop Line-1st Detector	1st-2nd Detector	2nd-3rd Detector	3rd-4th Detector
SH-183 and Roaring Springs				
Northbound SH-183	144	74	91	
Southbound SH-183	108	64	83	
SH-174 and FM-917				
Northbound SH-174	108	64	83	
Southbound SH-174	108	64	83	
FM-1220 and Boat Club Road				
Westbound FM-1220	144	74	91	
Southbound Boat Club	144	74	91	
SH-199 and Fire Hall Drive				
Northbound SH-199	141	73		
Southbound SH-199	141	73		
SH-199 and Roberts cut off				
Westbound SH-361	108	64	83	
Eastbound SH-199	108	64	83	
FM-361 and FM-1069				
Westbound SH-361	108	64	83	
Eastbound SH-361	108	64	83	
US-84 and SH-317				
Westbound US-84	141	73		
Eastbound US-84	141	73		
Southbound US-84	141	73		
US-290 and FM-1960				
Westbound US-290	108	64	83	
Eastbound US-290	108	64	83	
Northbound SH-6	108	64		
Southbound FM-1960	108	64	83	
SH-6 and Jackson				
Westbound SH-6	144	74	91	
Eastbound SH-6	144	74	91	
SH-146 and Crest Lane				
Northbound SH-146	108	64	83	97
Southbound SH-146	144	74	91	

Note: Detector spacing is measured upstream from the stop line; all loop detectors are configured 6x6-ft.

ment method used to locate the multiple detectors was the TSDHPT modified Beirele method. Spacings of multiple detectors on the respective approaches are given in Table 6.

Traffic volume and stopped time delay data were observed and recorded at each field site both before and after the existing single-loop detectors were replaced with multiple detectors on selected approaches. Comparisons of the before and after stopped-time delay and traffic data were used as a means of evaluating the effect of two multiple-detector placement systems.

Field Data Collection

Stopped-time delay and traffic volume data were collected at each test site by using procedures specified by Reilly and others (11). At each location data were acquired during both peak and off-peak traffic volume conditions, with and without multiple-detection systems. Data collection for multiple-detector systems was conducted a minimum of one year after system installations, thus providing time for driver familiarization and signal timing fine tuning.

Data Analyses

The field data analysis process was designed to test a general hypothesis that multiple-detection systems affect stopped delay when compared with conventional single-detector systems. Conventional parametric analysis of variance testing was used to examine this hypothesis.

By using stopped-time vehicular delay as the dependent variable, three-way analysis of variance testing was applied independently to data from each test site. In order to normalize differences in before-and-after vehicular delay data due to variations in traffic volume, all delay statistics were divided by appropriate intersection approach traffic volume totals. Therefore, the dependent variable was actually mean stopped-time delay per vehicle passing through the intersection approach.

A summary of the analysis of variance testing is presented in Table 7. Probability values that indicate the likelihood that observed effects could be due to chance alone are presented. For example, the probability that the observed effects (presumably) due to the detector scheme at SH-174 and FM-917 could have occurred due to chance alone is almost zero (0.004). On the other hand, the probability is extremely large (0.429) that the observed effects of multiple detectors at FM-1220 and Boat Club are

Table 7. Summary of significance of F-ratio form analysis of variance for single- versus multiple-detector installations.

	Source of Variation[a]			
Test Site	Main Effects	Before versus After[b]	Intersection Approach	Time
SH-174 and FM-917	0.003	0.004	0.007	0.031
FM-1220 and Boat Club Drive	0.214	0.429	0.073	0.128
SH-183 and Roaring Springs	0.062	0.236	0.865	0.015
SH-361 and FM-1069	0.001	0.052	0.001	0.661
SH-6 and Jackson Street	0.001	0.001	0.003	0.001
SH-146 and Crest Lane	0.001	0.084	0.276	0.001
US-290 and FM-1960	0.260	0.222	0.122	0.364
US-84 and SH-317	0.556	0.892	0.326	0.468
SH-199 and Fire Hall Drive	0.001	0.002	0.001	0.034
SH-199 and Roberts cut off	0.042	0.405	0.019	0.089

[a]Numbers in each cell can be interpreted as the proportion of all possible chances that differences of the size observed could have occurred due to chance alone. Minimum cell value is 0.0 and maximum is 1.0.
[b]Single versus multiple-point detection.

Table 8. Overview of field comparisons of multiple- versus single-point detection.

	Arithmetic Mean Stopped Vehicular Delay[a]			
Test Site	Before, Single Point	After, Multiple Point	Statistically Significant[b]	Detector Configuration Producing Least Delay
SH-174 and FM-917	5.80	9.12	Yes	Single
FM-1220 and Boat Club Road	16.00	14.42	No	Multiple
SH-183 and Roaring Springs	7.56	5.98	No	Multiple
SH-361 and FM-1069	5.70	5.16	No	Multiple
SH-6 and Jackson Street	16.32	8.14	Yes	Multiple
SH-146 and Crest Lane	11.44	13.71	No	Single
US-290 and FM-1960	19.61	29.24	No	Single
US-84 and SH-317	5.05	4.98	No	Multiple
SH-199 and Fire Hall Drive	16.95	10.43	Yes	Multiple
SH-199 and Roberts cut off	14.58	18.52	No	Single

[a]Includes all approaches. [b]Significant at $\alpha = 0.05$.

Table 9. Accident analysis parameters before and after multiple-detector installation.

	Annual Total Traffic Volume (000 000s vehicles)		Annual Accidents		Annual Accident Rate per million vehicles		
Intersection	Before	After	Before	After	Before	After	Approach Speed (mph)
SH-199 and Firehall Drive	10.2	10.2	12	7	1.18	0.69	45
FM-1220 and Boat Club Road	3.93	4.28	6	3	1.53	0.70	45
SH-199 and Roberts cut off	11.5	11.3	20	27	1.74	2.39	45
SH-183 and Roaring Springs	12.0	11.9	29	22	2.42	1.85	40
SH-174 and FM-917	6.04	7.10	7	7	1.16	0.99	55
US-220 and SH-6 and FM-1960	11.9	16.1	27	31	2.27	1.93	55
SH-6 and Jackson Street	4.94	4.49	5	2	1.01	0.45	55
SH-146 and Crest Lane (Barbours Cut)	8.46	11.6	6	0	0.71	0	55
SH-361 and FM-1069	3.75	4.44	4	6	1.07	1.35	40
US-84 and SH-317	2.74	2.81	13	14	4.80	4.98	45

indeed due to chance alone. Also included are analogous assessments for effects due to intersection approach and times of observation as well as all main effects taken together. A probability value of 0.05 is frequently assumed to be small enough to guarantee acceptable confidence that effects are not chance occurrences. If this policv is adopted, statistically significant differences in stopped-time delay due to detector scheme were observed at 3 of the 10 test sites. This statement does not imply, however, that in all three of these significant cases multiple detectors reduced delay. In fact, Table 8 demonstrates that in one of these three cases, there was a significant increase in vehicular delay under multiple detection.

Another view of the comparison between single and multiple-point detection is presented in Table 8. Arithmetic mean values of stopped-time delay per vehicle, including all observations, both for single- and multiple-point detection schemes are presented. The statistical significance of differences between before and after observations at a confidence level of 0.05 and an indication of which detector scheme produced smaller delay values are included. As already noted, effects attributable to detection scheme were significant in only three cases, and of these only two indicated greater efficiency under multiple-point detection.

A generalized comparison of the before-and-after data means indicates that 6 of the 10 test sites had at least marginal decreases in delay under multiple-detection schemes. Conversely, 4 of 10 performed more efficiently under the original single-point detection schemes. A conventional T-test was performed to evaluate the hypothesis that all means of before-and-after conditions drawn from the same population are equivalent. This test produced a T-statistic of 0.65 with 18 degrees of freedom, which when compared with a table value of 2.10 (for a 0.05 confidence level) is obviously not significant. In fact, this value is not significant at a 0.50 confidence level. Therefore, if stopped-time delay is taken as a measure of operational efficiency, data gathered at these 10 test sites do not demonstrate any significant difference in operational efficiency for the single- and multiple-point detection systems that were studied.

SIMULATION OF FIELD SITE CONDITIONS

In the previous sections, a field experiment that compares multiple-point and single-point detection has been described. Although the TEXAS model for intersection traffic that was used in the simulation study of four multiple-point detector systems described earlier has been previously verified through field studies, additional verification was deemed desirable.

Therefore, a typical field test site was selected for comparing delay statistics produced by the simulation model with those observed under field conditions. The intersection of SH-174 and FM-917 was selected for this experiment, and known geometry, signal timing, detector placement, and traffic characteristics were input to the simulation model. Conditions both before and after installation of multiple detectors were simulated and both peak and off-peak traffic volumes were used.

A factorial experiment was designed to test for statistically significant differences among treatment effects. Three main effects were studied; these included time (either peak or off-peak), intersection approach, and data source (field versus simulation).

Differences among simulation and field delay statistics are not statistically significant at an alpha level of 0.05. Differences due to the other two main effects are also not significant at the corresponding alpha level. Although the results of this limited experiment cannot be completely generalized, the assumption that the simulation technique does a reasonable job of representing real-world delay information is supported.

COMPARISONS OF SINGLE-POINT AND MULTIPLE-POINT DETECTION SCHEMES

A field test of single-point and multiple-point detection schemes has been presented. The test compared the two detection methods in a before versus after format with stopped-time vehicular delay as the response variable. A limited comparison of vehicular delay statistics produced by TEXAS simulation model and those collected through field measurement was also presented.

Based on these data and analyses, the following statements can be made:

1. A statistically significant difference in vehicular delay due to single-point versus multiple-point detection was not found and
2. Differences among vehicular delay data predicted by the TEXAS simulation model and that measured in field tests were not found to be statistically significant at a confidence level of 0.05.

ACCIDENT ANALYSIS

In order to assess the significance of multiple detectors on accident experience, data were acquired for each of the test sites. In all cases accident data were compiled for at least one year following installation of multiple detectors. Data for one to three years preceding installation was used as a basis for comparison.

Traffic volumes were used to convert numbers of accidents to accident rates and thereby produce statistics that were somewhat more comparable. Before and after traffic volumes, accidents, and accident rates are presented in Table 9.

Statistical significance of changes in numbers of accidents and rates was evaluated by using both a Poisson and a chi-square test (12). Two tests were used as a means of bounding possible results since the chi-square test is deemed rather conservative and the Poisson test somewhat liberal. The intersections of US-290 and SH-6 and US-84 and SH-317 were deleted from the analysis because of changes in the traffic environment during the data collection period that could not be controlled and would likely bias results.

The remaining eight intersections were grouped by approach speeds into a 40-45 mph class and a 50-55 mph class. Poisson and chi-square tests were applied to each of the two groups and to the aggregate.

Tests by both procedures indicated that changes in accidents and rates were statistically significant at a 95 percent confidence level for the high approach speed (50-55 mph) group. Changes in numbers of accidents or rates for the low approach speed (40-45 mph) group and the aggregate of all eight intersections did not indicate statistical significance at the 95 percent confidence level.

CONCLUSIONS

A comparative evaluation of vehicle detector systems for use in actuated signal control has been presented. The evaluation has compared vehicular delay and accident experience that result from detection systems by using single and multiple detectors and has compared four techniques for locating multiple detectors. In addition, vehicular delay statistics predicted by the TEXAS simulation model have been compared with those observed at a field site.

Based on these analyses the following conclusions may be stated:

1. Simulation studies indicated that there was no generally significant difference in the vehicular delay that resulted from applying the Beirele, Winston-Salem, or SSITE multiple-detector placement methods,
2. Vehicular delay predicted by the TEXAS traffic simulation model was not shown to be significantly different from that observed at a selected field test site,
3. Comparison of single- and multiple-detector installations at 10 test sites indicated no significant difference in stopped-time vehicular delay, and
4. Statistically significant reductions in accident experience were identified at intersections that had multiple detectors and high approach speeds (50-55 mph). Changes in accident experience attributable to multiple detectors at intersections that had approach speeds less than 50 mph were not statistically significant.

ACKNOWLEDGMENT

This paper was prepared 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 herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration. The paper does not constitute a standard, specification, or regulation.

Discussion

Peter S. Parsonson

The yellow interval of a traffic signal is such a familiar part of our everyday lives that most people would assume that its design has been well settled and agreed on by traffic engineers for several decades. On the contrary, the topic remains highly controversial to this day. There is still much discussion of when the yellow should start and for how long it should be shown. Multiple-detector schemes are used to control the beginning of the yellow. The authors of this paper are to be commended for being the first to apply a computer simulation model to multiple-point detection.

Multiple-detector strategies have been offered as solutions to a problem related to a zone of indecision (1,6). If the yellow comes on while a high-speed vehicle is in this zone the driver may have difficulty in deciding whether to stop or to go through, although the yellow is long enough to allow either decision. A safety problem can occur if the driver changes his or her mind. A last-second decision to stop abruptly may result in a rear-end collision or a swerve that produces a side-swipe accident. The zone of indecision has been well defined by Zegeer (5).

The multiple-point detectorization schemes examined by the authors were intended by their originators to help solve the problem of the zone of indecision.

Any multiple-point scheme should be compared with the single-point design that uses a density controller. Density designs were in use prior to any of the multiple-point schemes, which were devised primarily because many traffic engineers and technicians prefer a basic actuated controller to the more complex density machine. The density design is the defender in any discussion of schemes to alleviate the problem of the zone of indecision, because the gap-reduction adjustment permits the allowable gap to be as low as 2.0 s (14).

The authors, with good reason, prefer schemes that are effective over a range of approach speeds. Therefore, they discarded green extension systems and extended-call (EC) detector systems in the belief that designs with only two detectors are not effective over a range of speeds. Actually, the extended-call design as well as the density design can be adapted to a range of speeds by increasing the controller's unit extension (or minimum gap), or the detector's extension timing. This sacrifices allowable gap in favor of speed range. This tradeoff allows the engineer to make the design effective over a wider range of speeds at locations where light traffic poses no threat of extending the green to the maximum interval set on the controller.

The authors state that, at present, the Beirele, Winston-Salem, and SSITE methods are recognized as the most-common multiple-point detector-placement methods. The Beirele method keeps the allowable gap reasonably short by placing the first detector only 261 ft from the intersection for a 50-mph design speed (1). This distance is entirely inadequate, as the zone of indecision begins 350 ft from the intersection at this speed (5). The vehicle must be detected before entering that zone.

The Winston-Salem design has the same defect, as the upstream detector at 246 ft falls short by more than 100 ft. Again, effectiveness in eliminating driver indecision is sacrificed in order to keep the allowable gap reasonably short.

The SSITE method remedies this particular problem by placing the first detector 350 ft from the intersection. However, the allowable gap is so long, at 7 s, that this design would be useful only under the lightest of traffic conditions. The paper that originally presented the design explained that it requires undesirably long allowable gaps, stated specifically to be 7 s (15), "The controller's ability to detect gaps in traffic is substantially impaired.... As a result, moderate traffic will routinely extend the green to the maximum setting--an undesirable situation" because on max-out a vehicle may be caught in the zone of indecision. Sackman and others (1) stated that "This allowable gap is so long as to virtually disqualify the design from further consideration" and added that "the SSITE method will rarely be the design of choice."

It appears, then, that this project applied computer simulation techniques to designs that are unsatisfactory. In many respects they compare unfavorably with the traditional density design.

In 1979, perhaps too recently to have figured in this project, a novel EC-delayed call (DC) detection scheme was proposed (6) in response to a perceived need for a design offering loop-occupancy features; a basic, actuated controller with nonlocking memory; a short allowable gap; and effectiveness over a wide range of speeds. A 6-x25-ft loop at the stopline automatically switches from EC to DC operation at the strategic moment during the green interval. (A new detector unit now on the market makes the switch without any of the external relay logic described in the paper.) Normal-calling small loops 254 and 384 ft from the intersection provide protection against driver indecision at speeds from 40 to 60 mph, and 35 mph or lower (but not from 36 to 39 mph). The allowable gap is 4 s.

The authors found at their field sites that multiple detectors did not significantly reduce accident rates at intersections that have approach speeds less than 50 mph. A test installation of the EC-DC design in the Atlanta area has a median approach speed of 48 mph. The EC-DC scheme was found to reduce abrupt stops from 5 to none over the observation period (6). "Brake before clearing" maneuvers were reduced from 8 to none. Total conflicts were reduced 69 percent from 29 to 9. Overall, the authors rated the design superior to either the density scheme or the extended-call detector system. These observed reductions in erratic maneuvers suggest the potential for a reduction in accidents.

Zegeer (5) gathered accident data in addition to conflict rates. He found that his green-extension systems brought about a 54 percent reduction in total accidents, and rear-end collisions were reduced by 75 percent. At least two of the three locations studied appear to have average speeds of only 45 mph.

The authors found that the Beirele, Winston-Salem, and SSITE detector placement methods produced about the same delay in their computer simulations. This is an unexpected finding, as the allowable gaps vary over a wide range (4-7 s) and it is well established (16,17) that delay is sensitive to the allowable gap. Morris and Pak-Poy (16) found by computer simulation that delay can increase by as much as 45-105 percent if the allowable gap is increased from 4 to 7 s. Similarly, Tarnoff and Parsonson (17, p. 14) found by computer simulation that delay can increase by as much as 50 to 70 percent if the allowable gap is increased from 4 to 7 s.

Possibly the authors used a low setting of the maximum interval, thereby causing the green to change to yellow before gap-out could take place.

It is worth emphasizing that a long allowable gap is objectionable not primarily because of delay but because only moderate volumes can extend the green to the maximum interval. In that case, a vehicle may well be caught in the zone of indecision. A computer simulation could be very helpful in the preparation of guidelines for the maximum volumes that can be tolerated by designs of various allowable gaps.

REFERENCES

1. H. Sackman, P.S. Parsonson, B. Monahan, and A.F. Trevino. Vehicle Detector Placement for High-Speed, Isolated Traffic-Actuated Intersection Control; Vol. 2, Manual of Theory and Practice. Federal Highway Administration, Rept. FHWA-RD-77-32, May 1977, 178 pp.
2. A.D. May. Clearance Interval at Traffic Signals. HRB, Highway Research Record 221, 1968, pp. 41-71.
3. Webster's New Collegiate Dictionary. G. & C. Merriam Company, Springfield, MA, 1974.
4. P.S. Parsonson and others. Small-Area Detection at Intersection Approaches. Traffic Engineering, Feb. 1974, pp. 8-17.
5. C.V. Zegeer. Effectiveness of Green-Extension Systems at Highway-Speed Intersections. Division of Research, Bureau of Highways, Kentucky Department of Transportation, Lexington, Res. Rept. 472, May 1977.
6. P.S. Parsonson, R.A. Day, J.A. Gawlas, and G.W. Black, Jr. Use of EC-DC Detector for Signalization of High-Speed Intersections. TRB, Transportation Research Record 737, 1979, pp. 17-23.
7. P.S. Parsonson and others. Large-Area Detection at Intersection Approaches. Traffic Engineering, June 1976, pp. 28-37.
8. A Policy on Geometric Design of Rural Highways. AASHO, 1965, p. 136.
9. C.E. Lee, T.W. Rious, and C.R. Copeland, Jr. The TEXAS Model for Intersection Traffic Development. Center for Highway Research, Univ. of Texas at Austin, Res. Rept. 184-1, Dec. 1977.
10. C.E. Lee, G.E. Grayson, C.R. Copeland, Jr., J.W. Miller, T.W. Rioux, and V.S. Savur. The TEXAS Model for Intersection Traffic--User's Guide. Center for Highway Research, Univ. of Texas at Austin, Res. Rept. 184-3, Dec. 1977.
11. W.R. Reilly, C.C. Gardner, and J.H. Kell. A Technique for Measurement of Delay at Intersections; Vol. 3, User's Manual. Federal Highway Administration, Rept. FHWA-RD-76-137, Sept. 1976, 31 pp.
12. R.M. Michaels. Two Simple Techniques for Determining the Significance of Accident Reducing Measures. Traffic Engineering, Sept. 1966.
13. D. Gazis, R. Herman, and A. Maradudin. The Problem of the Amber Signal Light in Traffic Flow. Traffic Engineering, Vol. 30, No. 10, July 1960, pp. 19-26.
14. P. Parsonson. Signalization of High-Speed, Isolated Intersections. TRB, Transportation Research Record 681, 1978, pp. 34-42.
15. Southern Section, ITE. Large-Area Detection at Intersection Approaches. Traffic Engineering, June 1976, pp. 28-37.
16. R.W.J. Morris and P.G. Pak-Poy. Intersection Control by Vehicle-Actuated Signals. Traffic Engineering and Control, Oct. 1967, pp. 288-293.
17. P.J. Tarnoff and P.S. Parsonson. Selecting Traffic Signal Control at Individual Intersections. NCHRP, Rept. 233, 1981, 133 pp.

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

Driver Use of All-Red Signal Interval

TIMOTHY A. RYAN AND CHRISTIAN F. DAVIS

The purpose of this study was to investigate the theory that drivers use the red signal interval more frequently at intersections that have all-red intervals (i.e., all approaches to an intersection have a red indication) than at intersections that do not have all-red intervals. Data were collected at 10 intersections in four New England cities, during both peak and off-peak periods. Some 2764 signal cycles were observed, during which 1115 vehicles entered the intersection after the start of the red interval. The data were subjected to statistical analyses that yielded the following conclusions: (a) more drivers ran the red light at intersections that had all-red intervals than at intersections that had no all-red intervals; (b) the length of the all-red interval appeared to be cor-

related with driver use of the red interval (there were, however, four other predictor variables, each of which was more closely correlated with driver use of the red interval than was the length of the all-red interval); and (c) apparently drivers did not run the red light longer after the start of the red interval at intersections that had all-red intervals than at intersections that had no all-red intervals. The results of this study must be viewed with caution because the number of observed locations was relatively small, and some factors that influence driver use of the red interval may not have been studied during this project. However, the results indicate that a problem may exist and that more research should be done in this area.

The purpose of this study was to investigate driver use of the red signal interval, especially the all-red signal interval. An all-red signal interval is a period of time during a signal cycle during which all approaches to an intersection have a red indication, and all pedestrian signals show a steady DON'T WALK indication. All-red intervals (ARIs) have been in use for at least 20 years, and quite possibly longer. However, no uniform criteria may be applied in the decision to use or not to use an ARI in a particular situation. The Manual on Uniform Traffic Control Devices for Streets and Highways (MUTCD) (1) states, "The yellow vehicle change interval may be followed by a short all-way red clearance interval, of sufficient duration to permit the intersection to clear before cross traffic is released."

In addition, the Institute of Traffic Engineers' Transportation and Traffic Engineering Handbook (2) states, "When y_2 (the nondilemma yellow interval) exceeds the value selected for the yellow interval and when hazardous conflict is likely, an all-red clearance interval could be used for 2-3 s between the yellow interval and the start of green for opposing traffic."

Thus, an ARI is called for in a situation when more clearance time is needed but when it is not desirable to extend the amber interval.

Many drivers use the first few seconds of the red interval as an extension of the amber interval. The presence of an ARI might serve to encourage such behavior. Little information on this topic is available in the literature. Thus, research was undertaken to determine whether or not more drivers ran the red light at intersections that had ARIs than at intersections that had no ARIs.

Specifically, this research was directed toward answering three questions:

1. Do more drivers run the red light at intersections that have ARIs than at intersections that have no ARIs?
2. Is the length of the ARI correlated with driver use of the red interval? and
3. Do drivers run the red light longer after the start of the red interval at intersections that have ARIs than at intersections that have no ARIs?

FIELD METHODOLOGY

Decision Zone

When a vehicle approaches an intersection and the signal indication changes from green to amber, the vehicle must be in one of the following positions:

1. The vehicle is so close to the intersection that it is virtually impossible to stop before entering the intersection,
2. The vehicle is so far away from the intersection that it is impossible to enter the intersection until long after the start of the red interval, or
3. The vehicle is somewhere in between the two positions described above.

A vehicle in this in-between zone when the signal turns amber is said to be in the decision zone. The driver of the vehicle is forced to make a decision between stopping before entering the intersection or continuing through the intersection.

In order for a vehicle to stop safely before entering the intersection, it must be far enough upstream to allow the driver to react to the amber light and decelerate safely to a stop. As shown by Gazis, Herman, and Maradudin (3), this minimum distance is given by

$$x_c = v_o t_2 + (v_o^2/2d_m) \quad (1)$$

where

- x_c = minimum distance from the front of the vehicle to the stop line when the amber interval starts,
- v_o = initial velocity of the vehicle,
- t_2 = perception-reaction time of the driver for braking, and
- d_m = maximum deceleration rate.

Thus, for a given approach speed, the downstream boundary of the decision zone is defined by Equation 1 and may be found by using appropriate values for t_2 and d_m. By using the mean value of t_2 found by Gazis, Herman, and Maradudin (3) and an assumed maximum deceleration of 16 ft/s², Equation 1 reduces to

$$x_d = 1.14v_o + (v_o^2/32) \quad (2)$$

where x_d is the distance upstream from the intersection line (the curb line extended) at which the downstream boundary of the decision zone is located (ft) and v_o is measured in feet per second.

The upstream boundary of the decision zone is given by

$$x_u = v_o t_m + [a(t_m)^2/2] \quad (3)$$

where

- x_u = distance upstream from the intersection line at which the upstream boundary of the decision zone is located,
- t_m = longest time after the start of the amber interval that a driver will enter the intersection, and
- a = acceleration rate.

The longest amber interval at the observed intersections was 3.5 s, and 99 percent of the drivers who ran the red light entered the intersection 5.0 s or less after the start of the red interval. Thus, a reasonable value of t_m for these data is 8.5 s If we assume a maximum value for a of 5.0 ft/s², Equation 3 reduces to

$$x_u = 8.5v_o + 184.9 \quad (4)$$

where x_u is measured in feet and v_o is measured in feet per second. Thus, if a vehicle is located x ft upstream of the intersection line when the signal changes to amber and $x_d < x < x_u$, the vehicle is in the decision zone.

As the foregoing discussion indicates, the size and location of the decision zone for an individual vehicle are primarily a function of the approach velocity. Note, however, that t_m and d_m vary from driver to driver, and probably vary for an individual driver, depending on mood, influence of drugs or alcohol, or similar human factors. Therefore, the size and location of the decision zone will also be affected by individual driver characteristics.

Only one observer studied the intersections; therefore, it was not possible to observe the speeds of the individual vehicles and determine if they were in the decision zone when the signal turned amber. Instead, the observer was required to make a subjective judgment as to whether or not any vehicles were in the decision zone when the signal changed to amber.

Signal cycles during which at least one vehicle was in the decision zone when the signal turned amber were regarded as decision cycles. Signal cycles during which no vehicles were in the decision zone when the signal turned amber were regarded as nondecision cycles. Whenever there was any doubt as to whether a cycle was a decision cycle or a nondecision cycle, the cycle was regarded as being a decision cycle.

Data Collection

Each of the intersections chosen for this study was a four-legged, signal-controlled intersection in an urban area. At each intersection, each approach was perpendicular to the adjacent approaches or as close to perpendicular as possible. Nine of the 10 intersections were located in the central business district (CBD) of the city, and the 10th was in an urbanized section of the city a short distance from the CBD. Four of the intersections had ARIs; six did not.

At nine of the 10 intersections, data were collected on two days. At the 10th intersection data were collected on only one day. Each day, both the afternoon off-peak and the evening peak periods were observed. Afternoon off-peak data were collected between 12:30 and 3:00 p.m. The evening peak data were collected between 3:45 and 5:45 p.m. The observation periods were either 1.5- or 2-h long. Thus, each of the intersections observed on two days was observed for a minimum of 6 h and a maximum of 8 h. All data were collected on Monday, Tuesday, Wednesday, or Thursday.

At each intersection, the observer recorded the volume on the approach under consideration. The observer also recorded the direction taken by each of the vehicles as it left the intersection.

When the signal indication turned from amber to red for the approach under consideration, the observer started two hand-held stopwatches. If a vehicle crossed the intersection line (curb line extended) after the start of the red interval, one of the stopwatches was stopped when the vehicle's frontmost axle crossed the intersection line. If a second vehicle ran the red light, the second stopwatch was stopped by using the same criterion. The direction taken by each runner as he or she left the intersection was also recorded.

Note that the intersection line, rather than the stop line, was used as a reference line. This was done in order to maintain a constant reference line at all of the intersections. The distance from the stop line to the intersection line varied from 0 ft at one intersection to 40 ft at another, and drivers frequently stopped in the region between the two lines. Thus, many vehicles crossed the stop line after the start of the red interval but did not go through the intersection until the following green interval. However, virtually every time a vehicle crossed the intersection line during the red interval, it continued through the intersection. Thus, the intersection line was a better reference line.

Every 30 min the volume on the approach during the preceding 30 min was recorded, and the time was marked on the data-collection sheets. This made it possible to analyze all of the data in half-hour intervals. Table 1 shows the number of half-hour intervals observed and the number of runners observed at each intersection.

DATA ANALYSIS

All observed signal cycles were divided into two groups: decision cycles and nondecision cycles. A decision cycle is a signal cycle during which at least one vehicle is in the decision zone at the start of the amber interval. A nondecision cycle is a signal cycle during which no vehicles are in the decision zone at the start of the amber interval. During a nondecision cycle no driver has a chance to decide whether to stop or to continue through the intersection when the signal turns amber. Thus, a nondecision cycle is useless in an examination of how often drivers decide to use the red interval. For this reason, the nondecision cycles were eliminated from the data base before any analysis was performed.

In order to compare driver use of the red interval at the various intersections, the following equation was used:

$$P_r = (C_r/C_d) \times 100 \quad (5)$$

where

P_r = percentage of decision cycles during which at least one vehicle crossed the intersection line after the start of the red interval,
C_r = number of cycles during which at least one vehicle crossed the intersection line after the start of the red interval, and
C_d = number of decision cycles.

The values of C_d, C_r, and P_r for each intersection may be found in Table 2.

t-Test: P_r at Intersections That Have ARIs Versus P_r at Intersections That Do Not Have ARIs

The concept that the presence of an ARI is correlated with increased driver use of the red interval was tested. The observed P_r values, in 30-min intervals, were divided into two groups. The first group was comprised of the observations made at intersections that had ARIs, and the second group was comprised of the observations made at intersections that had no ARIs. The mean P_r values of the two groups were then compared by means of the t-test.

The variances of the two samples are significantly different at $\alpha = 0.01$, so it was necessary to use the modified t-test. The results of the modified test, given in the table below, reveal that the difference is significant at $\alpha = 0.01$ between the two groups. Thus, the presence of an ARI is correlated with increased driver use of the red interval.

Intersections	No. of Half-Hour Intervals	Mean P_r	SD
Have ARIs	47	37.2	22.53
Do not have ARIs	84	27.6	13.76

Regression Analysis

A regression analysis was performed to determine the factors that are important in driver use of the red interval. Only geometric features of the intersections and traffic control factors were considered;

Table 1. Intersections observed and data collected.

City	Intersection	Approach	No. of Half-Hour Intervals Observed: Off-Peak	Peak	Total	No. of Runners
Hartford, CT	Farmington-Sigourney	Farmington eastbound	8	8	16	122
	Main-Gold	Main northbound[a]	8	7	15	94
	Ann-Asylum	Asylum westbound[a]	4	4	8	74
New Haven, CT	College-Elm	Elm eastbound	7	8	15	149
	Church-George	George eastbound[a]	6	6	12	53
	Church-Elm	Elm eastbound	8	6	14	128
	Chapel-Church	Church northbound[a]	6	6	12	216
Providence, RI	Dorrance-Weybosset	Weybosset eastbound	6	6	12	87
	Empire-Washington	Washington westbound	6	6	12	118
Worcester, MA	Main-Pearl-Mechanic	Main southbound	8	7	15	74
Total			67	64	131	1115

[a]Approach has ARI.

Table 2. Driver use of the red interval.

City	Intersection	Approach	No. of Decision Cycles, C_d	No. of Cycles Run, C_r	Percentage of Decision Cycles, P_r
Hartford, CT	Farmington-Sigourney	Farmington eastbound	330	98	29.7
	Main-Gold	Main northbound[a]	181	57	31.5
	Ann-Asylum	Asylum westbound[a]	147	62	42.2
New Haven, CT	College-Elm	Elm eastbound	289	106	36.7
	Church-George	George eastbound[a]	210	46	21.9
	Church-Elm	Elm eastbound	255	97	38.0
	Chapel-Church	Church northbound[a]	246	155	63.0
Providence, RI	Dorrance-Weybosset	Weybosset eastbound	368	81	22.0
	Empire-Washington	Washington westbound	413	105	25.4
Worcester, MA	Main-Pearl-Mechanic	Main southbound	325	62	19.1
Total			2764	869	31.4

[a]Approach has ARI.

factors that vary from vehicle to vehicle, such as vehicle speed, number of occupants, and sex of the driver were not considered.

A priori, seven variables were selected that could have some significant effect on driver use of the red interval:

1. Length of the ARI,
2. Volume on the approach under consideration,
3. Width of the approach under consideration,
4. Width of the crossing roadway,
5. Volume-to-capacity ratio for the approach under consideration,
6. Percentage of signals in that particular city that have an ARI, and
7. Distance from the stop line to the intersection line.

In order to determine the correlation between each of the predictor variables and P_r, and also to determine the correlations among the supposedly independent predictor variables, a correlation matrix was constructed. The matrix is shown in Table 3. Examination of the matrix reveals that the correlation between P_r and the width of the crossing roadway is statistically insignificant at α = 0.05, and so is the correlation between P_r and the percentage of signals in the city that have an ARI. The correlation between P_r and the length of the ARI is significant at α = 0.05 but not at α = 0.01. The remainder of the correlations between P_r and the predictor variables are significant at α = 0.01.

The best linear representation of the data is given by

$$P_r = -18.3 + 0.56W_a + (38.02V/C) \qquad (6)$$

where W_a is the width of the approach under consideration (ft) and V/C is the volume-to-capacity ratio on the approach under consideration. An analysis of the partial regression coefficients is given below.

Variable	Partial Regression Coefficient	Confidence Interval at α = 0.05
W_a	0.56	0.228
V/C	38.02	22.695

Equation 6 leaves much of the variance unexplained and also has a considerable standard error. Thus, the equation should not be expected to accurately predict driver use of the red interval at a specific intersection. In addition, note that the data collected covered the ranges of the predictor variables as given in Table 4 and that extension of the equation beyond these limits is of questionable value.

Conspicuous by its absence from Equation 6 is the length of the ARI. As stated earlier, there is a significant correlation at α = 0.05 but not at α = 0.01 between P_r and the length of the ARI. This is solid, but not overpowering, evidence to suggest that the length of the ARI influences driver use of the red interval. However, in terms of predictive ability, the approach width and the volume-to-capacity ratio seem to be far superior to the length of the ARI.

In light of the results of the multiple linear regression, we thought that a nonlinear equation might give a better indication of the relation

Table 3. Correlation matrix.

Item	A	B	C	D	E	F	G	H
A	1.000	−0.1334	0.3394	0.3116	0.3509	0.2283	0.2550	0.1912
B		1.000	0.1090	0.5348	−0.0637	0.5244	−0.4967	0.4303
C			1.000	0.5152	−0.0245	−0.0702	−0.5010	0.0275
D				1.000	0.0756	0.3187	−0.5288	0.4676
E					1.000	0.4367	0.6243	0.2695
F						1.000	0.0727	0.3847
G							1.000	−0.0507
H								1.000

Notes: A = length of ARIs, B = volume on the approach under consideration (vehicles/h), C = width of the crossing roadway (ft), D = width of the approach under consideration (ft), E = distance from the stop line to the intersection line (ft), F = volume-to-capacity ratio, G = percentage of signals in the city that have an ARI, and H = P_r.

Table 4. Range of observed values of independent variables.

Variable	Range	Variable	Range
Length of ARI (s)	0.0-3.0	Volume-to-capacity ratio	0.43-0.98
Volume on approach (vehicles/h)	428-1738	Percentage of signals in the city that have an ARI	0-50
Width of approach (ft)	19-59		
Width of crossing roadway (ft)	34-62	Distance from stop line to intersection line (ft)	0-40

between P_r and the predictor variables. The following equation was used to test this concept:

$$P_r = a_0 x_1^{a_1} x_2^{a_2} x_3^{a_3} \ldots x_n^{a_n} \quad (7)$$

Equation 7 was then transformed by taking the logarithm of each side of the equation. Thus,

$$\log_{10}P_r = \log_{10}a_0 + a_1\log_{10}x_1 + \ldots + a_n\log_{10}x_n \quad (8)$$

The multiple linear regression on the transformed data resulted in a poorer representation of the data than was found in the regression described by Equation 6. Therefore, the results of the regression on the transformed data are not included here.

t-Test: Off-Peak P_r Versus Evening Peak P_r

As mentioned earlier, each intersection was observed during off-peak conditions and also during evening peak conditions. The purpose of this part of the data analysis was to determine whether time of day affects the frequency of driver use of the red interval at a given intersection. The concept was tested by comparing the off-peak mean P_r at a given intersection to the evening peak mean P_r at the same intersection. The data, as before, were examined in 30-min intervals. The null hypothesis was tested by means of the t-test. The results of the test may be found in Table 5.

Examination of Table 5 reveals that, for each intersection, the variances are homogeneous at α = 0.05. At 7 of the 10 intersections, there is no significant difference between means at α = 0.05. At the remaining three intersections, the difference is significant at α = 0.05 but not at α = 0.01. Interestingly, at two of these three intersections, more drivers used the red interval during the peak period than during the off-peak period; however, at the third intersection, more drivers used the red interval during the off-peak period than during the peak period.

t-Test: Mean Timing at Intersections That Have ARIs Versus Mean Timing at Intersections That Do Not Have ARIs

A t-test was conducted to determine if the presence of an ARI coincides with drivers entering the intersection at a later time than they would if there was no ARI. Because driver use of the red interval was greater at intersections that had ARIs, we thought that the mean timing of the runners might be greater at intersections that had ARIs than at intersections that did not have ARIs. This concept was tested by means of the t-test.

The results of the test may be found in the table below. Examination of the table reveals that the variances are significantly different at α = 0.01, and that the modified t-test value is insignificant at α = 0.05.

Intersections	No. of Runners	Mean Timing (s)	SD
Have ARIs	357	1.20	0.71
Do not have ARIs	604	1.24	0.85

Thus, despite the evidence that the presence of an ARI coincides with increased driver use of the red interval, it apparently does not coincide with drivers entering the intersection at a later time after the start of the red interval.

t-Test: Comparison of Mean Timing by Direction and Conflict Level

The mean timing by direction and conflict level was tested to determine if there is a significant difference in mean timings between directions and between two classifications of conflict for left turns. The entire set of 961 runners was divided into four groups. One of the groups was comprised of the right-turn runners and a second group was comprised of the straight-through runners. Left-turn runners were stratified into two groups; the left turn without conflict group was comprised of runners who turned left from a one-way street onto a one-way street (and thus encountered no more conflict than did a runner making an ordinary right turn) and the left turn with conflict group was comprised of the remainder of the left-turn runners.

The mean timing for each of these groups was then compared with the mean timing of the other three groups. These comparisons were made by means of the t-test.

The results of the tests may be found in Table 6. Table 6 reveals that there is a significant difference at α = 0.01 between the mean timing for the through runners and the mean timing for each of the other groups. The difference between the mean timing for left runners without conflict and the mean timing for right runners is significant at α = 0.05 but not at α = 0.01. The difference between the mean timings for the left runners with conflict and the mean timing for right runners is not significant at α = 0.05.

Further examination of Table 6 reveals that through runners have the lowest mean timing and are

Table 5. Variation in P_r with time of day.

Intersection	Period	No.	Mean P_r	SD	SE	F	t-Test	Result
Farmington-Sigourney	Off-peak	8	17.2	11.27	3.98	2.31	2.37	Significant at $\alpha = 0.05$
	Peak	8	34.4	17.13	6.06			
Main-Gold	Off-peak	8	26.8	14.05	4.97	2.82	0.05	Insignificant at $\alpha = 0.05$
	Peak	7	26.3	23.62	8.93			
Ann-Asylum	Off-peak	4	34.1	17.95	8.97	2.98	1.66	Insignificant at $\alpha = 0.05$
	Peak	4	51.3	10.40	5.20			
College-Elm	Off-peak	7	45.2	15.84	5.99	3.76	2.86	Significant at $\alpha = 0.05$
	Peak	8	27.0	8.17	2.89			
Church-George	Off-peak	6	23.8	12.75	5.20	1.17	0.71	Insignificant at $\alpha = 0.05$
	Peak	6	18.8	11.80	4.82			
Church-Elm	Off-peak	8	36.6	8.26	2.92	1.39	0.25	Insignificant at $\alpha = 0.05$
	Peak	6	37.8	9.73	3.97			
Chapel-Church	Off-peak	6	61.4	18.34	7.49	2.71	0.32	Insignificant at $\alpha = 0.05$
	Peak	6	64.2	11.14	4.55			
Dorrance-Weybosset	Off-peak	6	24.4	5.98	2.44	1.97	1.19	Insignificant at $\alpha = 0.05$
	Peak	6	19.4	8.39	3.42			
Empire-Washington	Off-peak	6	22.8	12.16	4.96	3.38	0.88	Insignificant at $\alpha = 0.05$
	Peak	6	27.8	6.61	2.70			
Main-Pearl-Mechanic	Off-peak	8	13.4	8.30	2.93	2.30	2.23	Significant at $\alpha = 0.05$
	Peak	7	25.5	12.60	4.75			

Table 6. Comparison of mean timings by direction and conflict level.

Direction Conflict Level	No.	Mean Timing (s)	SD	SE	F	t-Test	Modified t-Test θ (°)	Modified t-Test d	Result
Left turn without conflict	163	1.31	0.789	0.062	1.81		50	0.19	Insignificant at
Left turn with conflict	146	1.33	1.061	0.088					$\alpha = 0.05$
Left turn without conflict	163	1.31	0.789	0.062	1.56		56	3.36	Significant at
Straight through	487	1.08	0.645	0.029					$\alpha = 0.01$
Left turn without conflict	163	1.31	0.789	0.062	1.22	2.15			Significant at
Right turn	165	1.50	0.871	0.068					$\alpha = 0.05$
Left turn with conflict	146	1.33	1.061	0.088	2.71		60	2.70	Significant at
Straight through	487	1.08	0.645	0.029					$\alpha = 0.05$
Left turn with conflict	146	1.33	1.061	0.088	1.48		49	1.53	Insignificant at
Right turn	165	1.50	0.871	0.068					$\alpha = 0.05$
Straight through	487	1.08	0.645	0.029	1.83		57	5.68	Significant at
Right turn	165	1.50	0.871	0.068					$\alpha = 0.01$

followed in ascending order by left runners without conflict, left runners with conflict, and right runners.

CONCLUSIONS

The results of the analyses of the field data may be summarized as follows. In answer to the three questions posed earlier,

1. More drivers ran the red light at intersections that had ARIs than at intersections that had no ARIs. At the intersections studied, the difference is significant in the frequency of driver use of the red interval between intersections that had ARIs and intersections that had no ARIs. This difference is found to be significant, by means of the modified t-test, at $\alpha = 0.01$.
2. The length of the ARI appeared to be correlated with driver use of the red interval. The simple correlation between the length of the ARI and the frequency of driver use of the red interval is significant at $\alpha = 0.05$ but not at $\alpha = 0.01$. (The simple correlation between frequency of driver use of the red interval and each of four other predictor variables, however, was significant at $\alpha = 0.01$).
3. Apparently drivers did not run the red light longer after the start of the red interval at intersections that had ARIs than at intersections that have no ARIs. At the intersections studied, the difference in the mean timings between intersections that had ARIs and intersections that had no ARIs is not significant at $\alpha = 0.05$. In fact, the mean timing at intersections that had no ARIs is slightly higher than the mean timing at intersections that had ARIs.

These conclusions must be viewed with caution, however, for the following reasons. Only 10 intersections in four different cities were studied. Some factors that influence driver's use of the red interval might not have been studied during this project. For example, the type of phase following the observed phase (such as a pedestrian phase or the crossing street green phase) may very well influence driver use of the red interval. Also, at intersections that had ARIs, drivers may have used the red interval as frequently before the implementation of the ARI as afterward. In fact, it is possible that use of the red interval by drivers was the reason that the ARI was implemented in the first place.

These results raise some interesting questions. If it is indeed true that people treat the ARI merely as an extension of the amber, two potentially serious problems present themselves:

1. If a driver gets accustomed to having ARIs, and to taking advantage of them, he or she may try to take advantage of the ARI at an intersection where no ARI exists.

2. If a driver gets accustomed to being able to safely enter an intersection after the start of the red interval, at least a portion of the traffic signal's authority is diminished for that driver.

The implications of the second potential problem are somewhat less tangible but are just as serious as the implications of the first potential problem. Each driver on a street risks his or her life on the assumptions that all other drivers obey the signs and traffic signals used to control traffic and that all other drivers accept those signs and signals as the absolute authority over their travel.

If the authority of these signs and signals is reduced in some way (such as suggested above), and some drivers begin to treat the signs and signals contemptuously, the law-abiding driver takes a bigger risk each time he or she enters an intersection.

The effects of this potential problem would be difficult, if not impossible, to quantify, and the use of the ARI may do more good than harm. Indeed, many traffic engineers can quickly cite instances in which use of an ARI at a particular intersection has led to reductions in accident levels. Nonetheless, the potential problems listed here should be considered.

In conclusion, these results should not indicate to the reader that the use of ARIs should be abolished or even limited. The number of intersections studied is simply not large enough to make such a sweeping conclusion. However, these results indicate that a problem may exist and that more research should be done in this area.

In addition to the major findings of this report, the following results were also obtained from the data analyses. There is no evidence to suggest that driver use of the red interval is greater during peak hours than during off-peak hours. In fact, at 4 of the 10 intersections studied, driver use of the red interval was greater during off-peak hours than during peak hours. Right-turn runners had the highest mean timing. They were followed, in descending order, by left-turn-with-conflict runners, left-turn-without-conflict runners, and straight-through runners.

ACKNOWLEDGMENT

We wish to thank J.E. Stephens and D. Heffley for their assistance with this research. We also wish to thank H. Ridgeway and Nancy M. Kuntz for their assistance with the data analysis.

Discussion

Peter S. Parsonson

Worldwide concern is mounting that drivers may be becoming increasingly disobedient of the red signal (1). The authors have addressed an important and timely topic. The first conclusion of the paper is that more drivers ran the red light at intersections that had ARIs than at intersections that had no ARIs. The authors concluded that, if their 10 intersections are indicative of all signalized locations, then a potentially serious problem presents itself.

One difficulty with these results is that the paper does not report the length of the yellow interval at any of the intersections; neither are the approach speeds stated. So, for all the reader is told, the yellow intervals could have been shorter at those intersections that use all-red, and that could be the reason that more drivers ran the red signal at those intersections.

Ryan has said in response to queries that he did not know whether the yellow intervals were identical at the 10 intersections and in fact had not measured them. However, the paper states that "the longest amber interval at the observed intersections was 3.6 s," so it appears that the yellow intervals are in fact known.

As the paper stands, it is hard to see how the interested reader can find any assistance whatever with a problem in timing the intergreen period, as the British call it. It would be helpful if the authors would state the yellow times and compare them with the minimum calculated as

$$y = t + (V/2a \pm 64.4g) \quad (9)$$

where

- y = minimum yellow time (s);
- t = perception-reaction time of driver (s); the standard value is 1 s;
- v = approach speed (ft/s);
- a = deceleration rate (ft/s²), currently taken to be 10 (2,3); and
- g = percent of grade divided by 100 (added for upgrade and subtracted for downgrade).

This equation was proposed by the discussant previously (2) as the minimum yellow to ensure that drivers need not enter on the red.

Authors' Closure

In this discussion, Parsonson raises several valid points. Certainly, the length of the amber interval at each intersection is important, and a computation of the adequacy of the amber interval would also be of interest.

The length of the amber interval was observed at each intersection. Table 7 shows the length of the amber interval, the length of the ARI, and P_r for each intersection. The length of the amber interval and the length of the ARI are average values; some variation in the lengths of these intervals was observed at several of the intersections.

Casual examination of Table 7 does not appear to indicate a relation between P_r and length of the amber interval. Statistical analysis might, of course, yield a different result.

Further examination of Table 7 reveals that the longest amber interval observed was 3.5 s, instead of the 3.6 s reported earlier. The text has been corrected in regard to this point. Since this value was not used in any of the statistical analyses, none of the results are affected by this correction.

Unfortunately, we could not obtain speed data along with the driver behavior data, because only one observer studied each intersection. Because of this, we cannot use the equation suggested by Parsonson to check the adequacy of the amber interval.

Parsonson is correct when he states that the paper does nothing to assist the reader with timing of the intergreen period. However, it was not the intent of this research to investigate the timing of that period. Rather, it was to investigate driver use of the ARI, and specifically to investigate the three questions stated in the paper's introduction.

Table 7. Driver use of the red intervals.

City	Intersection	Approach	Length of Amber Interval (s)	Length of ARI (s)	Percentage of Driver Use, P_r
Hartford, CT	Farmington-Sigourney	Farmington eastbound	3.0	0.0	29.7
	Main-Gold	Main northbound	2.6	1.6	31.5
	Ann-Asylum	Asylum westbound	3.0	3.0	42.2
New Haven, CT	College-Elm	Elm eastbound	3.4	0.0	36.7
	Church-George	George eastbound	3.5	2.8	21.9
	Church-Elm	Elm eastbound	3.4	0.0	38.0
	Chapel-Church	Church northbound	3.2	1.8	63.0
Providence, RI	Dorrance-Weybosset	Weybosset eastbound	2.7	0.0	22.0
	Empire-Washington	Washington westbound	3.0	0.0	25.4
Worcester, MA	Main-Pearl-Mechanic	Main southbound	3.4	0.0	19.1

REFERENCES

1. Manual on Uniform Traffic Control Devices for Streets and Highways. Federal Highway Administration, 1971.
2. Institute of Transportation Engineers. Transportation and Traffic Engineering Handbook. Prentice-Hall, Englewood Cliffs, NJ, 1976.
3. D. Gazis, R. Herman, and A. Maradudin. The Problem of the Amber Signal Light in Traffic Flow. Operations Research, Vol. 8, No. 1, Jan.-Feb. 1960, pp. 112-132.

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

Comparison of Signs and Markings for Passing and No-Passing Zones

RICHARD W. LYLES

An experiment was undertaken to examine the relative effectiveness of five pavement marking and signing sequences for informing motorists of passing and no-passing zones on rural two-lane, two-way rural roads. Treatments included (a) standard pavement markings, (b) pavement markings plus standard regulatory signing, (c) pavement markings plus no-passing pennants, and (d) and (e) two combinations of regulatory signs and pennants. Data were collected on overtaking and passing vehicles by two observers in a staged vehicle that traveled over a measured length of roadway. The principal findings were that the addition of any sign sequence to pavement markings resulted in motorists being appreciably more observant of the passing and no-passing zones and spending less time in the passing (opposing) lane. Less conclusive evidence was presented in support of the more emphatic and informative sequences that resulted in progressively more compliance with the marked zones.

Overtaking and passing maneuvers are two of the most common sources of conflict between two or more vehicles on two-lane, two-way rural roads. Numerous possibilities exist for collision, including rear-end, sideswipe, and, most dangerous, head-on. Drivers, in overtaking and passing another vehicle, depend on a number of visual cues to ascertain whether such maneuvers can be completed safely. In addition to checking for oncoming traffic and gauging the speed of both any oncoming vehicles within sight and the vehicle to be overtaken and passed, the driver also uses the information provided by pavement markings and roadside signs to ascertain the advisability of the maneuver--Is he or she in a marked passing zone, how much of the passing zone remains, and so forth. Signs and marking can clearly provide considerable guidance to the motorist in making judgments about the relative safety of passing maneuvers. Despite the presumed importance of the signs and markings for passing and no-passing zones, there appears to be a considerable range in how such devices are, or should be, used in practice [see, for example, Nickerson, (1) and Weaver and others (2)].

In the context described above, the basic objective of the research described herein was to evaluate several alternatives for roadside signing, relative to traditional pavement markings, for indicating passing and no-passing zones.

STUDY METHODOLOGY

Many traffic situations lend themselves to straightforward examination; for example, vehicles approaching a specified intersection or other potentially hazardous situation can be observed or tracked by using sensors on the road surface and appropriate electronic equipment (3) with the acquired data being used to calculate vehicle speeds at certain points on the approach to the hazard. The result is that fairly extensive sets of data can be obtained in a relatively short time, even in low-volume situations. By contrast, overtaking and passing maneuvers are dynamic in nature and, hence, more difficult to document relative to where certain events took place. Alternative methods for documenting such maneuvers include the use of film or videotape, isolation of one specific passing or no-passing zone, or use of some sort of mobile data-collection device.

For a variety of reasons, including equipment availability and the explicit capability to use several different zones, a mobile data-collection device was selected in this instance. The basic approach was to have a staged vehicle (Jeep Wagoneer) traverse a specified section of road while traveling at an explicit (and constant) speed. The driver and observer in that vehicle would make observations on any other vehicle that overtook them and passed, or attempted to pass, as well as on approaching vehicles and certain location information (e.g., Was the staged vehicle in a passing zone?).

A technical description of the data collection is provided by Wyman and Lyles (4) and also by Lanman (3). In general, the system was based on a modified vehicle speed measurement device (i.e., the Traffic Analyzer System manufactured by Leupold and Stevens, Inc.) and allowed specific observed events to be recorded by the observers on a machine-readable cassette tape that, when analyzed, provided both an indication of the event occurrence and its associated time.

The site was an approximately 6-mile segment of US-2 just east of Canaan, Maine. US-2 in this vicinity traverses rolling terrain and is two lane over the entire distance with a maximum grade of approximately 7 percent. Passing zones (which were, for the most part, established by the Maine Department of Transportation (DOT) independent of the experiment) ranged in length from 600 to 3200 ft and no-passing zones ranged from 400 to 9800 ft. The longest no-passing zone occurred on the hill that had the maximum grade, although a passing zone had recently existed in that area. The speed limit over the entire road segment was 50 mph (although it had, at one time, been 60 mph).

Both directions of travel over the segment were used for data collection. The observers in the staged vehicle would begin at one end of the segment, start the data-collection equipment, drive through to the other end of the segment while collecting data on any maneuvers and opposing traffic, and then reset the data-collection device and return to the original starting point over the same segment in the opposite direction.

In order that there be public familiarity (for motorists on whom data were collected) with the signs being tested, whenever the sign condition was changed (three of the five treatments are not typically used in Maine) the new condition was erected (used) not only on the actual test segment but also for about 2 miles in advance of the test segment in each direction. Hence, regardless of direction of travel, the average motorist encountered the first of the test signs about 2 miles prior to the actual test segment.

In general, the procedure during the experiment was to (a) deploy a given treatment on the test segment of road (plus the advance sections), (b) have the staged vehicle operate over the segment for up to two weeks (dependent on the amount of data gathered), and (c) change the treatment condition and collect more data. In addition to the data collected by the two people in the staged vehicle (on overtaking and passing vehicles and opposing traffic), data were also collected on weather conditions, time-of-day, and treatment condition deployed. When the data were coded for the analysis, certain other information (e.g., approximate sight distance at any point) was also calculated and recorded.

The five treatment conditions that were evaluated are shown in Figure 1. They included the standard pavement markings, a regulatory DO NOT PASS sequence, a warning NO PASSING ZONE sequence, and two combinations of the DO NOT PASS and NO PASSING ZONE treatments. The final combination (treatment condition 5) consisted of both the DO NOT PASS sign (on the right-hand side) and the NO PASSING ZONE (on the left-hand side) at the beginning of each no-passing zone; the latter was repeated at intervals throughout the zone. The intervals in condition 5 varied such that a motorist would have either just passed a sign, be able to see one ahead indicating the no-passing zone, or be able to see the PASS WITH CARE sign at the beginning of the next passing zone. Hence, the actual intervals between sequential NO PASSING ZONE signs in the final treatment were variable, depending on topography and sight distance.

The staged vehicle operated at 35 mph throughout the experiment, although a speed of 45 mph was also tried. The latter speed, given prevailing traffic volumes and a normal mean speed of just over 50 mph, resulted in a very low number of overtaking and passings of the staged vehicle and the few data that were obtained were not used.

FINDINGS

The raw data (the events and their associated times of occurrence) that were collected were organized so that a number of variables could be calculated from any overtaking of passing vehicle's time of event record (e.g., the time spent entirely within the passing lane during a passing maneuver). Events recorded by the two observers in the staged vehicle included the following.

The passenger observed

1. Vehicle overtaking staged vehicle,
2. Vehicle making maneuver from queue behind staged vehicle,
3. Left wheels (of overtaking vehicle) over centerline,
4. Right wheels over centerline,
5. Passing vehicle adjacent to staged vehicle,
6. Right wheels recross centerline,
7. Left wheels recross centerline,
8. Pass completed when passing vehicle had Maine license,
9. Pass completed when passing vehicle had non-Maine license,
10. Recreation vehicle,
11. Automobile and trailer,
12. Truck,
13. Abort (an error was made by observer), and
14. East or west (direction of staged vehicle, indicated at start of data run).

The driver observed the following events:

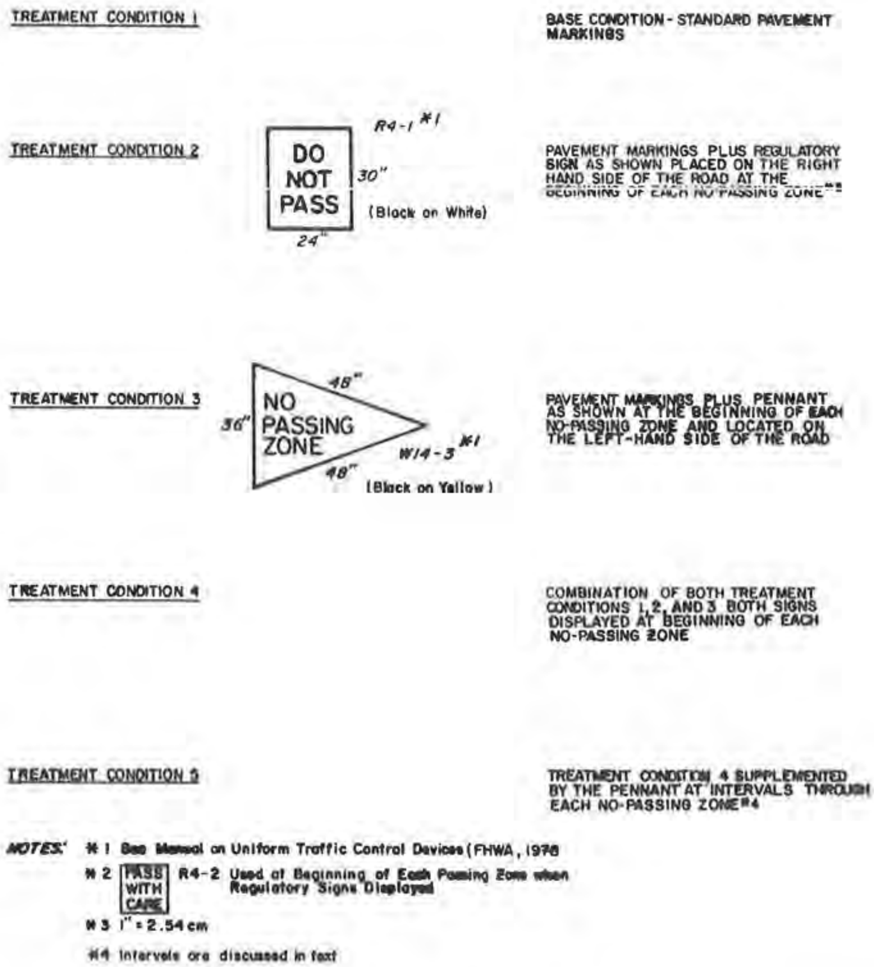
1. Staged vehicle entering no-passing zone,
2. Staged vehicle entering passing zone,
3. Opposing vehicles approaching (would be in sight of passing vehicle),
4. Opposing vehicle adjacent to staged vehicle (repeated if more than one), and
5. No opposition vehicles in sight.

Note that the data record for each overtaking or passing vehicle was a sequence of events with their associated times of occurrence.

Individual vehicles were also classified as to whether a given maneuver was a completed pass with no opposing traffic in view or one of several other categories (e.g., a quick-look where the overtaking vehicle's left wheels crossed the centerline although it pulled back in and did not pass), and whether the vehicle was a repeat (more than one maneuver was made) or not. In most instances, the data were analyzed with others in the same category unless analysis showed that there were no differences between the types being considered.

The data on the various maneuvers were of two basic types: concerning the context of the maneuver (e.g., in which passing or no-passing zone did it

Figure 1. Treatment conditions.



take place) or concerning the characteristics of the execution of the maneuver (e.g., how long was the passing vehicle completely in the opposing lane). Hence, analysis of the data was also of two general types: examination of the distribution of the maneuvers (e.g., the incidence of passes by passing and no-passing zone) and analysis of variance performed on some of the execution characteristics (e.g., time in the opposing lane).

The results reported here are not exhaustive but, rather, representative of the outcome of the overall analysis that was relatively extensive. Although not all aspects of the overtaking and passing maneuver and those independent factors that affect it are discussed, several were dealt with in the study. Some of the factors addressed in the complete report (6) included the impact of differences in topography and other characteristics of individual zones, the differences between familiar and nonfamiliar drivers (the latter were assumed, in the end, to be primarily represented in the data), speed of the vehicle being overtaken, and type of vehicle performing the maneuver. In some instances the factors were dealt with directly (e.g., length of passing zones) and, in others, indirectly (e.g., only automobiles were considered due to a lack of data on other types).

The distribution of the total number of observed maneuvers by type (although the last two types listed actually represent the absence of some action) is given below. Included in the vehicle types is ignored opportunity, which was defined as where the overtaking vehicle was unopposed in a passing zone with adequate sight distance and had an adequate amount of the passing zone remaining (i.e., the vehicle could have passed safely but did not make any maneuver).

Type of Maneuver	No. of Observations
Unopposed pass	485
Opposed pass	58
Partly opposed pass	132
Quick-look, unopposed	195
Quick-look, opposed	96
Lane change--no passing, unopposed	3
Lane change--no passing, opposed	6
Total	975
Never pass, opportunities ignored	71
Ignored opportunities	103
Total	1149

The distribution of maneuvers by vehicle status was as follows:

Vehicle Status	Observations (%)
New vehicles, i.e., those making first maneuver	57
Repeat vehicles, i.e., those that made more than one maneuver	27
Queue vehicles, i.e., those that were relatively close behind staged vehicle when maneuver of	13

Vehicle Status	Observations (%)
preceding was completed	
Pack vehicle, i.e., those that passed from a position other than immediately behind the staged vehicle	3

In general, the results of the experiment were not overwhelming relative to the desirability or effectiveness of one treatment over another, although there was a clear and definite break between the use of signs (of any sort) and the use of only pavement markings. Several positive results are presented below.

Maneuvers (both passes and quick-looks) tended to take longer if only the pavement markings were used (treatment 1) versus use of the pavement markings in conjunction with any type of sign condition (treatments 2-5). It can be hypothesized that this was due to the fact that motorists were more aware that they were in a passing zone and where the next no-passing zone started.

The number of clips (where a motorist was actually in the next no-passing zone before a pass was completed) of the next no-passing zone appeared to be unrelated to the marking or sign condition that was displayed. Approximate sight distances for passing maneuvers were somewhat lower when treatment 1 (pavement markings only) was displayed as opposed to the other four.

When opposing traffic was present, the acceptable time gap (termed passing gap) for the passing maneuver was about 14-16 s, whereas quick-looks were done when passing gaps averaged 10-12 s. However, no differences among the treatments were noted.

Comparison of the observations obtained in this experiment with similar values from other work for several key variables showed that there was basic agreement insofar as the structure of the passing maneuver was concerned. For example, the exposure time (time spent in the opposing lane) was in the same range as earlier reported figures (7,8). This finding lends credibility to both the experimental approach that was taken (i.e., using observers in a staged vehicle traveling at a set speed) as well as to the results obtained.

The most compelling result concerned the incidence of use of passing zones versus no-passing zones for any maneuvers (see Tables 1 and 2). As the treatment conditions became more emphatic and informative, the percentage of maneuvers done in passing zones continued to increase. The compliance statistic increased by 47-57 percent for all maneuvers and by 17-39 percent for unopposed passes between treatment 1 and treatment 5. The best rate of compliance was 92.9 percent for unopposed passes in the eastbound direction whereas the best for the pavement markings only was 72.6 percent.

Based on these results, no-passing signs (DO NOT PASS, NO PASSING ZONE, or some combination) used in conjunction with pavement markings seemed to increase not only compliance with the desired behavior (as indicated by the compliance of maneuvers in marked zones) but also more conservative, and presumably safer, passing behavior (as indicated, for example, by the drivers' requiring longer sight distance). There was also some evidence that more emphatic and informative treatments tended to have incrementally more effect. This last result was not, however, demonstrated conclusively.

These results are not substantially different from what might be expected intuitively, but they do provide an empirical foundation for using roadside signs in conjunction with pavement markings when better compliance with passing and no-passing zones is desirable. Presumably, use of such signs would be even more effective when visibility is somewhat restricted or when the pavement markings are not visible (e.g., when the road is snow covered). The positive increment of compliance could, however, be lessened if the signs were universally used for all passing and no-passing zones due to the potential for motorist disdain of oversigning. Although there is little question of the increased safety to be achieved through use of the signs in some situations, the results fall short of providing strong and conclusive support for selecting one sign treatment over another.

Table 1. Distribution of all maneuvers by passing and no-passing zones.

		Eastbound (%)		Westbound (%)	
No.	Treatment Condition	In Passing Zone	In No-Passing Zone	In Passing Zone	In No-Passing Zone
1	Pavement markings only	52.3	47.7	50.3	49.7
2	Do not pass regulatory signs	69.2	30.8	66.3	33.7
3	No-passing pennants	53.5	46.5	60.3	39.7
4	Regulatory signs and pennants	71.2	28.8	61.4	38.6
5	Treatment 4 plus supplemental pennants	82.0	18.0	73.8	26.2

Notes: A total of 975 observations were made, which does not include never pass and ignored opportunities maneuvers.
In order of emphasis and information, treatment 2 is considered to be more emphatic than treatment 3.

Table 2. Distribution of unopposed pass maneuvers by passing and no-passing zones.

		Eastbound (%)		Westbound (%)	
No.	Treatment Condition	In Passing Zone	In No-Passing Zone	In Passing Zone	In No-Passing Zone
1	Pavement markings only	66.7	33.3	72.6	27.4
2	Do not pass regulatory signs	80.0	20.0	82.0	18.0
3	No-passing pennants	80.0	20.0	82.0	18.0
4	Regulatory signs and pennants	88.9	11.1	75.0	25.0
5	Treatment 4 plus supplemental pennants	92.9	7.1	85.1	14.9

Notes: A total of 485 observations were made.
In order of emphasis and information, treatment 2 is considered to be more emphatic than treatment 3.

ACKNOWLEDGMENT

Considerably more detail on the research reported on herein can be found in the final report, submitted to the Federal Highway Administration (FHWA) (6). The research was funded by FHWA and undertaken at the Maine facility with staffing from FHWA, the Maine DOT, and the Social Science Research Institute of the University of Maine at Orono. The contributions of that staff notwithstanding, the responsibility for the conclusions presented here and in the final reports and any errors therein rest with me and are not necessarily those of FHWA, Maine DOT, or the University of Maine.

REFERENCES

1. M.H. Nickerson. No-Passing Zones State of the Art. Presented at 1979 AASHTO Traffic Committee Meeting at Colorado Springs, CO; Department of Transportation and Public Facilities, Juneau, AK, Nov. 1979.
2. G.D. Weaver, D.L. Woods, A. Huber, and S. Richards. Passing and No-Passing Zones: Signs, Markings, and Warrants. Texas Transportation Institute, College Station, Interim Rept. to FHWA, Nov. 1977.
3. M.H. Lanman, III. The Maine Facility: Capability for Rural Road Research. ITE Journal, Arlington, VA, Dec. 1978.
4. J.H. Wyman and R.W. Lyles. Mobile Traffic Data Collection System. Maine Department of Transportation and Social Science Research Institute, Univ. of Maine, Orono, Final Rept. to Office of Research, FHWA, Sept. 1979.
5. Manual of Uniform Traffic Control Devices. FHWA, 1978.
6. R.W. Lyles. Comparison of Signs and Markings for Passing/No-Passing Zones. Social Science Research Institute, Univ. of Maine, Orono; Maine Department of Transportation, Final Rept. to FHWA Office of Research, Rept. FHWA/RD-81/094 and FHWA/RD-81/095, 1981.
7. G.D. Weaver and J.C. Glennon. Passing Performance Measurements Related to Sight Distance Design. Texas Transportation Institute, College Station, Rept. 134-6, July 1971.
8. A Policy on Geometric Design of Rural Highways. AASHO, Washington, DC, 1965.

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

Effect of Raised Pavement Markers on Traffic Performance

WILLIAM L. MULLOWNEY

This project measured and documented the effect that snowplowable raised pavement markers (SRPMs) have on the behavior of traffic at certain geometric highway conditions. Two-lane rural curves, highway exits with deceleration lanes, and highway bifurcations were studied. Measures of performance selected to study the effects of the markers included erratic maneuvers such as cutting through painted gores, lane changes or encroachments, center and edgeline encroachments, point of entrance into deceleration lanes, and mean speeds and speed variance at curves. All erratic maneuvers studied were reduced significantly at various sites for traffic volumes per lane of up to 500 vehicles/h. At volumes per lane of between 900 and 1000 vehicles/h the markers had no effect on traffic. Raised markers were not successful in causing motorists to enter deceleration lanes at exits earlier. As far as speeds, the markers seem to have caused a smoother speed profile through the two curves studied, which resulted in less abrupt speed changes. The effect of SRPMs on speed variance was inconclusive. The markers were effective in reducing erratic maneuvers at sites with and without overhead lighting. At one site a significantly higher rate of erratic maneuvers during rain conditions before the markers were placed was not only severely reduced but the wet condition erratic maneuver rate approached the quality of the dry condition rate when markers were present.

This study was undertaken to determine whether snowplowable raised pavement markers (SRPMs) can reduce the variable behavior of traffic with regard to lane placement, choice of exit pathway, and speed to the extent that traffic conflicts and erratic maneuvers are reduced. The general belief is that the delineation provided by SRPMs would increase the driver's view of road and exit geometry and assist him or her in choosing a safe and efficient pathway.

OBJECTIVES

The study was designed to achieve the following objectives:

1. To measure the effect of SRPMs on centerline and edgeline encroachments on both lit and unlit curved sections of highway;
2. To measure the effect of SRPMs on speeds and speed variances on lit and unlit curves;
3. To measure the effect of SRPMs on the incidence of drivers encroaching on painted gores, both at exits and at highway bifurcations; and
4. To see whether SRPMs would cause motorists to enter the deceleration lanes at exits more consistently.

INSTALLATION PROCEDURE

Eight hundred raised pavement markers were installed at 11 sites in central and southern New Jersey. Amerace Corporation was contracted to provide the markers, concrete saw, epoxy dispensing machine and epoxy, and two machine operators. The New Jersey Department of Transportation provided the safety operation, a water truck, and sufficient workers to assist in placing the markers.

STUDY DESIGN

Potential sites were selected on the basis of the following criteria.

1. Existance of higher than normal rates of run-off-the-road accidents for a short section of highway;
2. Existance of a traffic performance problem such as encroachments, variability in exiting path, and weaving;
3. Subjective determination of the problem-

solving potential with the use of SRPMs;

4. Suitability of observation points for manual data collection;

5. Suitability of data collection by mechanical and photographic techniques;

6. Sufficient traffic volumes after dark to collect enough data for statistical analysis;

7. Distance from the research office, a concern for collection of data under rain conditions;

8. Lack of potential vandalism of markers and mechanical counting devices based on the accessibility of the site to pedestrians and whether the site is located in a developed area; and

9. Existence or lack of street lighting.

Pilot Studies at Potential Study Sites

A night time pilot study was performed at each site under consideration to determine what traffic characteristics should be studied at each location. The measures selected are listed in Table 1. The traffic maneuvers were defined as follows:

Centerline encroachment--any wheel of the vehicle crossed over both yellow lines and encroached on the opposing lane of travel;

Edgeline encroachment--any wheel of the vehicle crossed over the white edgeline and encroached on the shoulder;

Gore encroachment--any wheel of the vehicle touched any part of the painted gore at an exit or highway bifurcation;

Longitudinal exit placement--deceleration lanes at exits were divided into two zones; the first zone started at the beginning of the deceleration lane and ended at a point halfway to the painted gore, where the lane line extending from the gore began; the second zone ran from this point up to the physical gore; if any wheel of an exiting vehicle touched zone 1, it was considered a zone 1 exit;

Lane changes and encroachments--vehicles either completely changed lanes or encroached on the second exiting lane; and

Vehicle speed--spot speeds were collected at select locations.

Estimates of the frequency of each type of maneuver and traffic volumes were collected during the pilot studies and used to estimate the duration of data collection needed to gather enough samples for statistical analysis. The final locations for data collectors to position themselves were decided during the pilot studies.

Data Collection Method

Most of the data were collected manually by observers at each site. Observation points that allowed the observer to be raised up (preferably over the roadway) and hidden from view were used. Such points were commonly on overpasses and railroad bridges. Where these did not exist, observers were stationed on the side of the road on an embankment. Where this was not available, pneumatic traffic counters were used to collect data. At exits the observers counted total traffic, total exiting traffic, erratic maneuvers, and place of entry into the deceleration lane. The deceleration lane was divided into two zones, with the division at half the total length of the lane.

At curves, centerline and edgeline encroachments were gathered by visual observation at one site and by a combination of visual observation and an audio signal from a pneumatic traffic counter at another. Speeds at curves were collected with a hand-held radar unit. At bifurcations, gore encroachments were counted by using a visual and audio technique and traffic volumes were counted manually.

The audio technique involved running hoses from pneumatic traffic counters to the centerlines and edgelines of the curves studied and to the tip of the painted gore for the highway bifurcations. A car that encroached on the centerlines, edgelines, or gore would trip the counter to cause an audio signal that an observer stationed at the side of the road would record as an erratic maneuver (Figure 1).

Vehicles were classified into two-axle and three or more axle categories since it was believed that three or more axle vehicles would not react to the markers in the same manner as would two-axle vehicles.

Statistical Analysis

From the pilot studies, estimates of the time of data collection needed to collect sufficient sample for statistical analysis were generated. The number of erratic maneuvers aimed at for each site was 30, and the number of free-flowing spot speed samples was 100. However, at some sites these numbers were not reached but the sampling requirements of the statistical tests used still allowed the analysis to be performed. The specific tests used for each type of maneuver are described as follows.

Test of Proportions

The equation for the test of proportions is as follows (1, pp. 176-178):

$$Z = (p_1 - p_2)/\sqrt{(pq)[(1/N_1) + (1/N_2)]} \tag{1}$$

where

$p_1 = n_1/N_2$,
$p_2 = n_2/N_2$,
$p = (n_1 + n_2)/(N_1 + N_2)$,
$q = 1 - p$,
n_1 = number of before erratic maneuvers,
n_2 = number of after erratic maneuvers,
N_1 = traffic volume before, and
N_2 = traffic volume after.

This test was used to analyze the effect of SRPMs on gore encroachments, longitudinal exit placement, lane weaves, and centerline and edgeline encroachments. The test was applied (1, p. 117), "if the smaller value of p or q multiplied by the smaller value of N exceeds five." From the values of Z calculated by this test, the level of significance for the change in erratic maneuver rates was taken from a normal curve. For purposes of the decision, conclusions, and recommendations, a level of 95 percent or greater was considered significantly different for all statistical tests used.

t-Test

The t-test is calculated as follows (2, p. 200):

$$t = (\bar{X}_1 - \bar{X}_2)/\sqrt{(S_1^2/N_1) + (S_2^2/N_2)} \tag{2}$$

$\bar{X}_1$ = mean of before sample,
$\bar{X}_2$ = mean of after sample,
S_1 = standard deviation of before sample,
S_2 = standard deviation of after sample,
N_1 = before samples, and
N_2 = after samples.

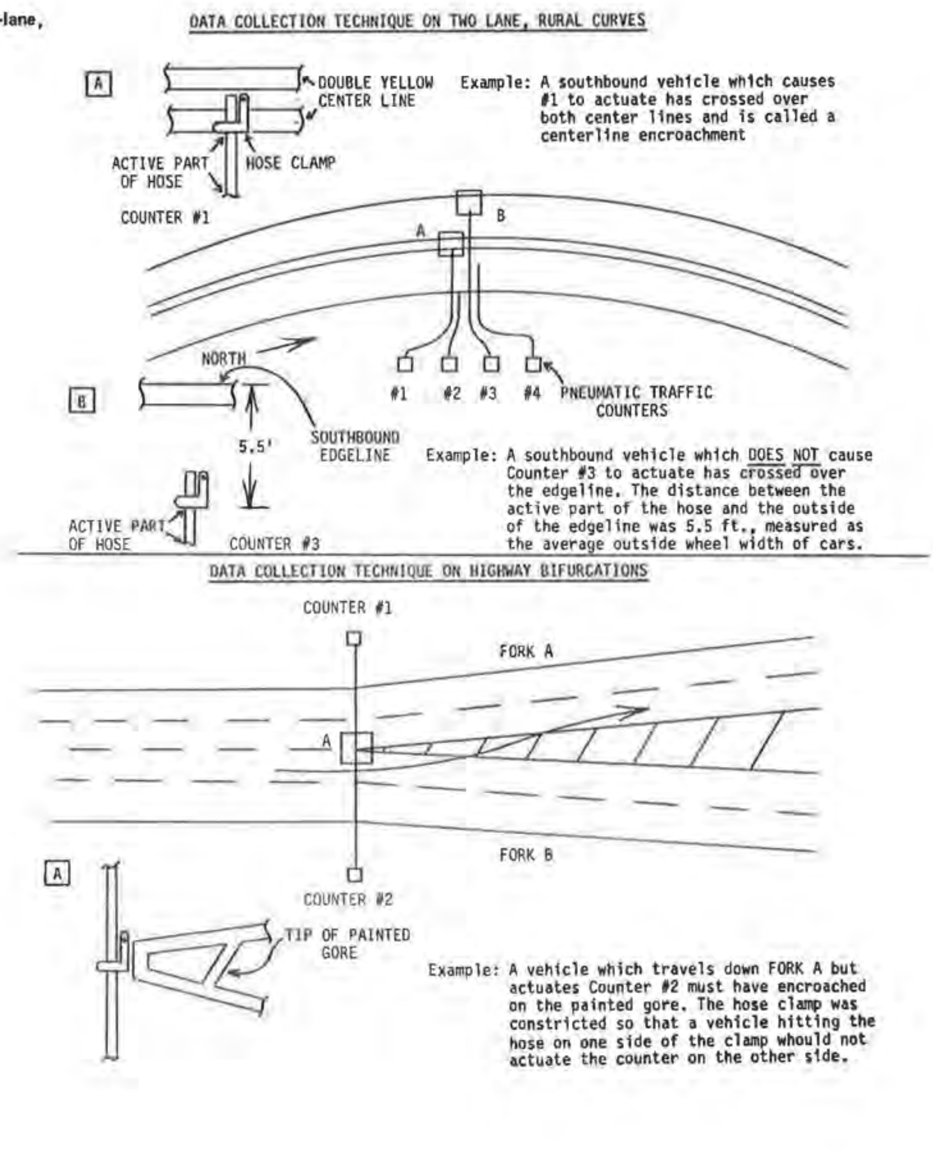
This test was used to analyze the differences in mean speeds attributable to the installation of the markers.

Table 1. Site descriptions and traffic performance measures studied.

Location	No. of Lanes	Lane Width (ft)	Shoulder Width (ft)	Gore Length (ft)	Decelera-tion Lane Length (ft)	Degree of Curve	Speed Limit (mph)	Lighting	Measures Studied[a]
Curves									
NJ-35	4 with 10-ft painted median	11	None			6	35	Yes	1
NJ-29	2	10	4			8	45	No	1,2,3
US-206	2	10	None			32	50	Yes	2
Bifurcations									
US-1 and US-1A	Right fork 2, left fork 2	12	10	400			55	Yes	4
I-287	Right fork 1, left fork 2	12	12	500			55	No	4
Exits									
NJ-440 and Garden State Parkway	3 thru, 1 right exit	12	10	142	410		55	Yes	4,6
US-1 and I-95	2 thru, 1 right exit	12	10	80	650		55	Yes	4,6
I-295 and NJ-168	3 thru, 1 right exit	12	None	170	830		55	Yes	4,6
I-295 and NJ-38	3 thru, 1 right exit	12	12	140	700		55	Yes	4,6
NJ-29 and Market St	3 thru, 2 left exits	13	None	88	730		50	Yes	4,5
I-287 and US-78	2 thru, 1 right exit	12	12	160	580		55	No	4,6
Control sites									
NJ-29	2	10	4			5	45	No	1,2,3
NJ-440 and US-9	3 thru, 1 right exit	12	10	134	480		55	Yes	4,6
I-295 and NJ-561	3 thru, 1 right exit	12	12	140	480		55	Yes	4,6

[a] 1 = speeds, 2 = centerline encroachments, 3 = edgeline encroachments, 4 = gore encroachments, 5 = lane changes or encroachments, and 6 = longitudinal exit placement.

Figure 1. Data collection technique on two-lane, rural curves and on highway bifurcations.



F-Test

The equation for the F-test is as follows (3, p. 131):

$$F = \sigma_1^2/\sigma_2^2 \qquad F_1 = N_1 - 1 \text{ df} \quad F_2 = N_2 - 1 \text{ df} \qquad (3)$$

where

σ_1^2 = variance of before or after sample,
σ_2^2 = variance of after or before sample,
N_1 = sample size used to compute σ_1, and
N_2 = sample size used to compute σ_2.

The larger variance is designated σ_1^2 and is used as the numerator whether it is the before or after sample.

This test was used to analyze differences in the variance between the before and after speed samples.

RESULTS: TWO-AXLE VEHICLES

Effect of SRPMs on Erratic Maneuvers Through Painted Gore at Exits

Six of nine sites experienced statistically significant reductions in the percentage of cars that cut through the painted gore; the two control sites did not change significantly (Table 2). Two sites (NJ-29 and NJ-168 during the earlier data collection period) did not change significantly and these sites, when studied under rain conditions, experienced an increase in the percentage of erratic maneuvers.

NJ-29 was the only left-side exit studied and the incidence of gore maneuvers was very small in the before studies, so the lack of a significant change is not surprising. This site was studied because it had two exit lanes between which a considerable amount of lane changing took place. The effect of the markers on this maneuver is discussed later.

That the NJ-168 site had an insignificant change during dry and wet conditions is somewhat perplexing. However, when the same site was studied later in the evening, a significant reduction in erratic maneuvers occurred. There was a large difference in the traffic volume per lane for the two different times of data collection--950 vehicles/h in the earlier period and 400 vehicles/h in the latter. If the traffic was spaced evenly over the three lanes for each condition, the average spacing between vehicles would be about 300 ft for the earlier time and about 750 ft for the later period. The closer average spacing in the first condition may have diminished the ability of the motorist to view enough of the exit markers in order to recognize the pattern. This may account for the lack of response to the markers under the higher-volume condition.

Effect of SRPMs on Choice of Exiting Path

Data were collected at four study sites and two control sites to see whether the SRPMs would cause more drivers to exit earlier in the deceleration lane (Figure 2, treatments A and B). The percentage of exiting vehicles that exited in zone 1 was collected before and after the installation of the markers.

Although the percentage of exiting vehicles in zone 1 changed significantly for all study sites except US-1 and I-95 (Table 3), the fact that the control sites experienced significant changes of a similar magnitude prohibits assigning of responsibility for the changes to the application of SRPMs.

The results of the study on the addition of edgelines and their effect on choice of exiting path are also listed in Tables 3 and 4. One trend exists for the data. When the gore and lane line were marked (Table 4, treatment B), all three sites had an increase in the percentage of zone 1 exits. When edgeline markers were added (treatment C), all three sites had a decrease in the percentage of zone 1 exits when compared with treatment B.

Effect of SRPMs on Lane Changing or Encroachment Between Two Exit Lanes

The incidence of lane changing or encroachment between two exit lanes on a left exit was significantly reduced with the application of SRPMs in both wet and dry conditions. In the rain, when the maneuvers were more prevalent than in dry condition, the reduction was greater and the percentage of erratic vehicles in the rain (44 percent) approached the percentage of erratic vehicles in the dry condition (38 percent) when SRPMs were present (see table below).

NJ-29 Condition	Total Exits	Erratic Maneuvers	Percentage	Change (%)	Level of Significance (%)
Dry					
Before	939	528	56.2		
After	941	365	38.8	-17.4	>99
Rain					
Before	139	100	71.9		
After	308	134	43.5	-28.4	>99

Effect of SRPMs on Gore Encroachments at Highway Bifurcations

The percentage of vehicles that cut across the painted gore at bifurcations was drastically reduced both for a lit (US-1 and 1A, 400-ft gore) and unlit site (I-287, 500-ft gore). No control site was studied for comparison; however, the magnitude of the change given in the table below is a telling statistic.

Route	Total Vehicles	Gore Encroachments	Percentage	Change (%)	Level of Significance (%)
US-1 and US-1A					
Before	3674	135	3.67		
After	3446	60	1.74	-1.93	>99
I-287					
Before	3983	96	2.41		
After	3544	22	0.62	-1.79	>99

Effect of SRPMs on Encroachments at Two-Lane, Rural Curves

Centerline and edgeline encroachments were reduced at the study sites by significant amounts but at a control site encroachments changed by nonsignificant amounts (Table 5). Unaccountably, the change at US-206, which has a good deal of street lighting, was larger than at NJ-29, which has no lighting.

The importance of minimizing centerline encroachments is easily apparent. The reduction of edgeline encroachments might not seem as important because conflict with other vehicles is not likely to occur. However, on a road like NJ-29, which is dark, with trees and telephone poles within a couple of feet of the edgeline, reduction of this type of erratic maneuver may be considered beneficial.

RESULTS: THREE OR MORE-AXLE VEHICLES

As previously stated, three or more-axle vehicles were differentiated from two-axle vehicles during data collection for the following reasons:

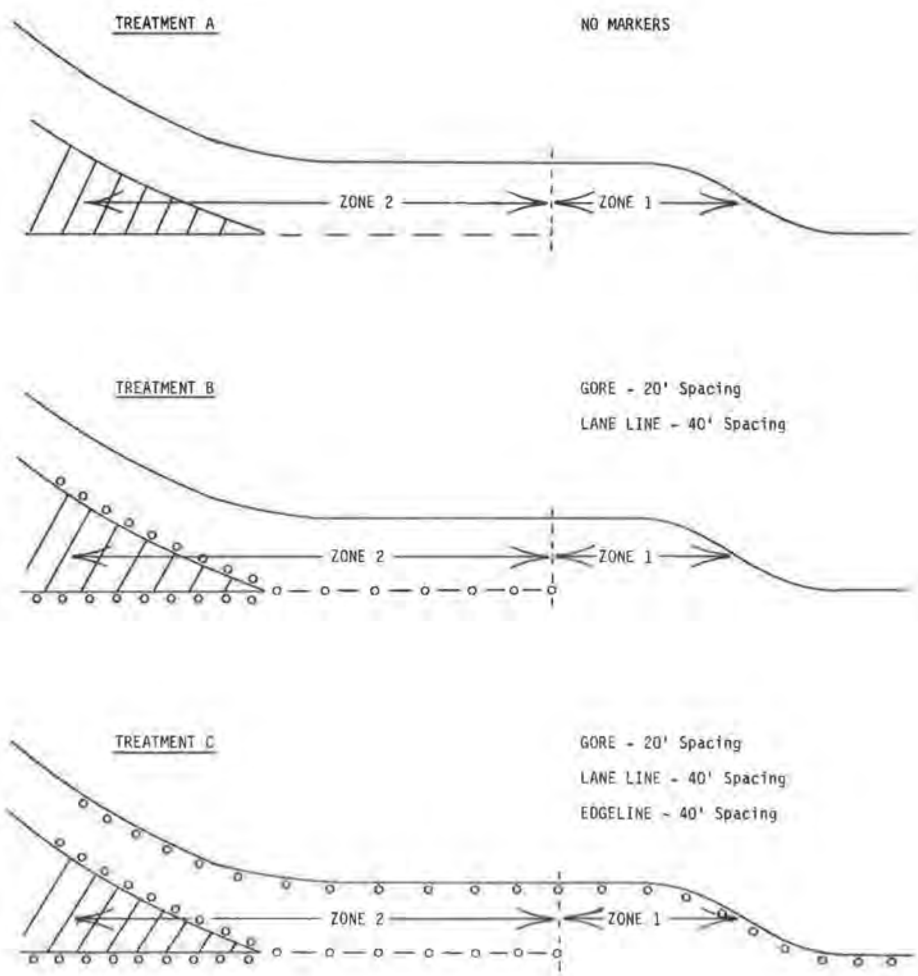
1. The greater vertical separation between the driver and the headlights may affect the visibility of the retroreflective devices and
2. Three or more axle vehicles have reduced maneuverability, which may inhibit their ability to react to SRPMs.

Table 2. Effect of SRPMs on gore encroachments.

	Before			After					
Site	Total Vehicles	Gore Encroach-ments	Percent	Total Vehicles	Gore Encroach-ments	Percent	Change (%)	Level of of Signifi-cance (%)	Vehicles per Hour per Lane
NJ-29[a]	2383	18	0.76	1880	13	0.69	-0.07	<50	250
NJ-29, rain[a]	310	1	0.32	725	8	1.10	+0.78	-[b]	250
US-1 and I-95[a]	1691	59	3.49	1883	13	0.69	-2.80	>99	450
I-295 and NJ-38[a]	3935	52	1.32	3586	12	0.33	-0.99	>99	450
NJ-440 and Garden State Parkway[a]	4039	42	1.04	4082	17	0.42	-0.62	>99	500
I-295 and NJ-168[a]	8077	27	0.33	7445	21	0.28	-0.05	<50	950
I-295 and NJ-168, rain[a]	2738	15	0.55	2397	14	0.58	+0.03	<50	950
NJ-440 and US-9[a,c]	5034	46	0.91	5251	39	0.74	-0.17	65	650
I-295 and NJ-561[a,c]	5271	27	0.51	7508	51	0.68	+0.17	79	900
NJ-440 and Garden State Parkway[d]	2781	23	0.83	3957	15	0.38	-0.45	98	250
I-295 and NJ-168[d]	4721	38	0.80	7872	19	0.24	-0.56	>99	400
I-287 and US-78, no lighting[d]	2785	14	0.50	5665	13	0.23	-0.27	>99	200

[a]Data were collected between 5:30 and 7:00 p.m.
[b]Insufficient data to apply statistical tests.
[c]Control site.
[d]Data were collected between 8:00 and 10:00 p.m.

Figure 2. Marker layouts for longitudinal exit placement studies.



Sufficient data were collected at nine sites to analyze the change from the before to the after condition for statistical significance. As previously outlined, the test for difference in proportions and the rule of thumb for determining whether sufficient data exist for applying the test were used in this analysis.

Table 6 shows the results of this analysis. Only one site, I-295 and NJ-38, experienced a change with a level of significance greater than 95 percent. Therefore, the general conclusion that SRPMs do not affect the traffic performance of three or more-axle vehicles with respect to the types of maneuvers studied can be reached.

RESULTS: EFFECT OF SRPMS ON VEHICLE SPEEDS AT CURVES

NJ-29, Hopewell

Speeds were collected at four locations in both the northbound and southbound directions. Location 1 was at the beginning of the south end of the installation, location 2 was at the apex of the curve, location 3 was at the north end of the installation,

Table 3. Percentage of two-axle exiting vehicles that enter first half (zone 1) of deceleration lane between 5:30 and 7:00 p.m.

	Before Treatment A			After Treatment C				
Sites	Exiting Vehicles	Exits in Zone 1	Percent	Exiting Vehicles	Exits in Zone 1	Percent	Change (%)	Level of Significance (%)
I-295 and NJ-168[a]	1876	1354	72.2	1823	1160	63.6	-8.6	>99
I-295 and NJ-38	1026	953	92.9	993	897	90.3	-2.6	96
US-1 and I-95	96	83	86.5	108	98	90.7	+4.2	65
NJ-440 and Garden State Parkway	1735	206	11.9	1749	344	19.7	+7.8	>99
NJ-440 and US-9[b]	955	714	74.8	1137	723	63.6	-11.2	>99
I-295 and NJ-561[b]	1154	1086	94.1	1594	1461	91.7	-2.4	98

[a]Data were compiled in axles, not vehicles. [b]Control site.

Table 4. Percentage of two-axle exiting vehicles that enter first half (zone 1) of deceleration lane between 8:00 and 10:00 p.m.

	Before Treatment A			Middle Treatment B			After Treatment C						
Site	Exits	Zone 1	Percent	Exits	Zone 1	Percent	Exits	Zone 1	Percent	Change, A to B (%)	Level of Significance	Change, B to C (%)	Level of Significance
NJ-440 and Garden State Parkway	989	79	8.0	985	234	23.8	378	68	18.0	+15.8	>99	-5.8	97
I-295 and NJ-168[a]	1342	856	63.8	1142	886	77.6	1218	730	59.9	+13.8	>99	-17.7	>99
I-287 and US-78[b]	216	161	74.5	236	212	89.8	232	197	84.9	+15.3	>99	-4.9	88

[a]Data were compiled in axles, not vehicles. [b]No lighting.

Table 5. Effect of SRPMs on centerline and edgeline encroachments by two-axle vehicles on two-lane rural curves.

	Before			After				
Site	Total Vehicles	Encroachments	Percent	Total Vehicles	Encroachments	Percent	Change (%)	Level of Significance (%)
Centerline encroachments								
US-206	1044	162	15.5	972	34	3.5	-12.0	>99
NJ-29[a]	675	78	11.6	406	32	7.9	-3.7	95
NJ-29[a,b]	707	14	2.0	733	17	2.3	+0.3	<50
Edgeline encroachments								
NJ-29[a]	1072	107	10.0	609	26	4.3	-5.7	>99
NJ-29[a,b]	450	36	8.0	457	32	7.0	-1.0	<50

[a]No lighting. [b]Control site.

Table 6. Effect of SRPMs on erratic maneuvers by vehicles (three or more axles).

	Before			After				
Site	Vehicles	Encroachments	Percent	Vehicles	Encroachments	Percent	Change (%)	Level of Significance (%)
Exit								
I-295 and NJ-168[a]	457	15	3.3	378	8	2.1	-1.2	68
I-295 and NJ-168, rain[a]	125	6	4.8	137	8	5.8	+1.0	<50
I-295 and NJ-38[a]	444	66	14.9	438	13	3.0	-11.9	99
I-295 and NJ-561[b]	338	14	4.1	547	15	2.7	-1.4	74
I-295 and NJ-168[c]	527	6	1.1	505	10	2.0	+0.9	71
Two-lane rural curves								
US-206[d]	32	8	25.0	43	9	20.9	-4.1	<50
NJ-29[d]	64	11	17.2	36	7	19.4	+2.2	<50
NJ-29[b,d]	50	12	24.0	47	6	12.8	-11.2	84
Bifurcations								
I-287	444	12	2.7	377	4	1.1	-1.6	91

[a]Data were collected during peak periods.
[b]Control site.
[c]Data were collected during off-peak periods.
[d]Includes centerline and edgeline encroachments.

and location 4 was about 1000 ft north of the installation, around a curve. At locations 1, 2, and 3 the markers were visible. Lack of a suitable place for parking the car out of the motorists' view prevented the collection of speeds at a control site south of the installation.

The SRPMs appear to have caused a smoother speed profile through the site in both the northbound and southbound directions (Figure 3, Table 7). This is evidenced by the smaller changes in speed that oc-

Figure 3. Effect of SRPMs on speed at NJ-29.

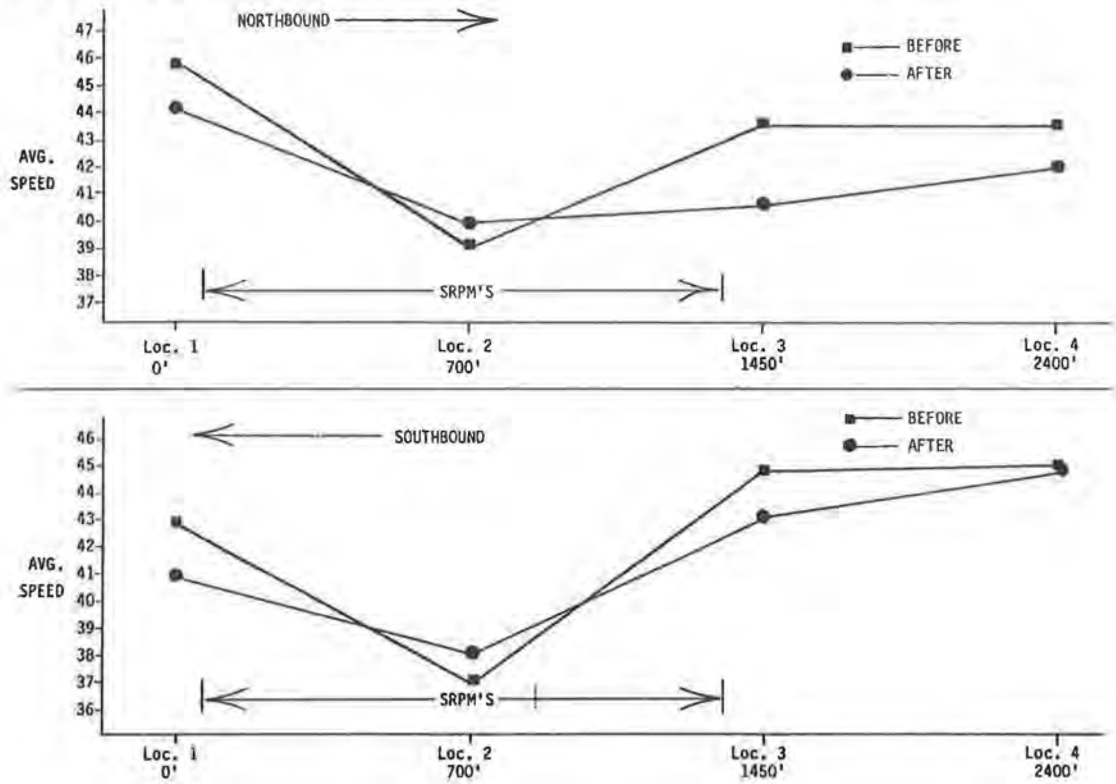
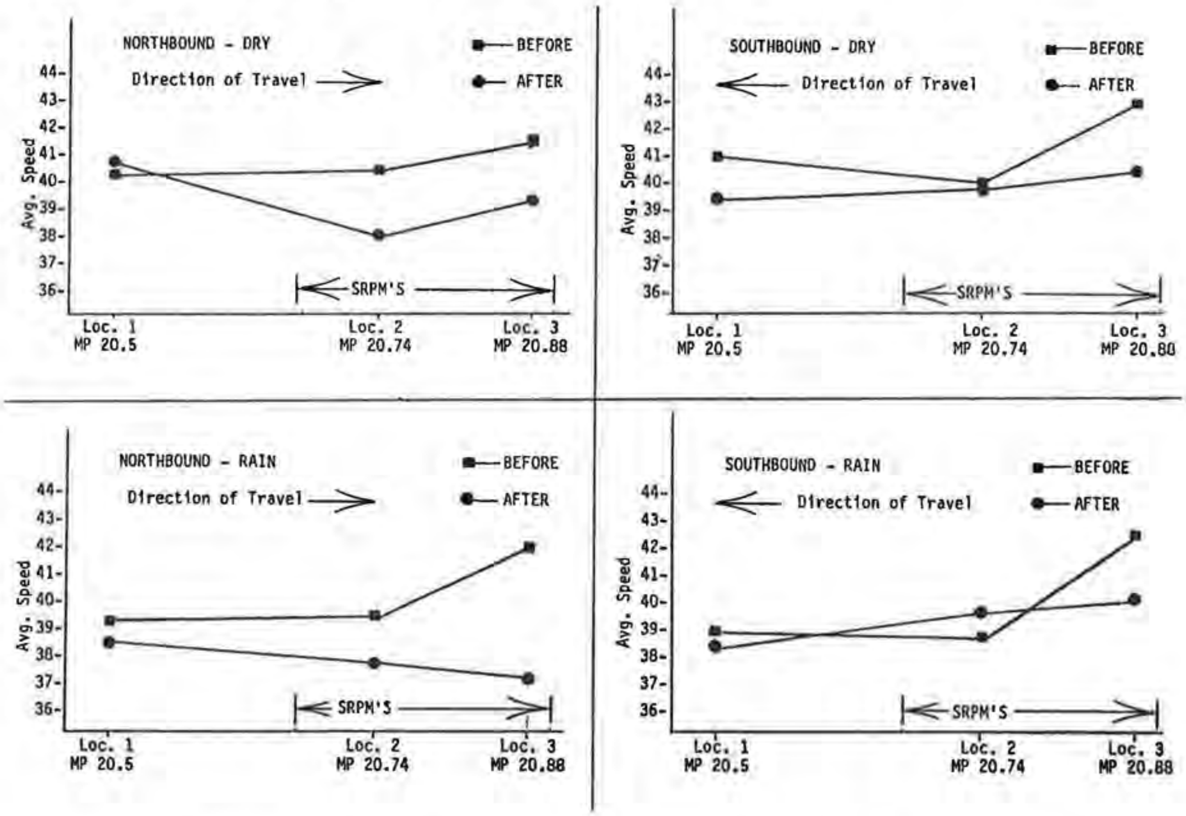


Table 7. Analysis of mean speeds and speed variance at NJ-29.

Location	Direction	Measure	Percentage Before	After	Change	Level of Significance (%)
1	Northbound	$\overline{X}$	45.8	44.1	-1.7	>99
		σ	4.2	4.7		
		σ^2	17.6	22.1	+4.5	>95
		n	187	215		
	Southbound	$\overline{X}$	48.8	40.9	-1.9	>99
		σ	5.2	4.6		
		σ^2	27.0	21.2	-5.8	<95
		n	121	147		
2	Northbound	$\overline{X}$	39.0	39.9	+0.9	97
		σ	3.9	4.1		
		σ^2	15.2	16.8	+1.6	<95
		n	194	177		
	Southbound	$\overline{X}$	36.8	38.0	+1.2	98
		σ	4.5	4.6		
		σ^2	20.3	21.2	+0.9	<95
		n	170	139		
3	Northbound	$\overline{X}$	43.4	40.6	-2.8	>99
		σ	4.4	4.5		
		σ^2	19.4	20.3	+0.9	<95
		n	176	249		
	Southbound	$\overline{X}$	44.7	42.9	-1.8	>99
		σ	4.7	5.0		
		σ^2	22.1	25.0	+2.9	
		n	162	210		<95
4	Northbound	$\overline{X}$	43.4	42.0	-1.4	>99
		σ	4.1	4.5		
		σ^2	16.8	20.3	+3.5	<95
		n	171	192		
	Southbound	$\overline{X}$	44.8	44.8	0	
		σ	4.2	4.1		
		σ^2	17.6	16.8	-0.8	<95
		n	137	185		

Figure 4. Effect of SRPMs on vehicle speeds at NJ-35, Belmar.



curred between the data collection points after tne markers were installed. The lower speeds measured as cars entered the site in the after condition (location 1, northbound and location 3, southbound) indicate that the markers gave the motorists a cue that the curve was near and prompted them to begin deceleration earlier. That speeds increased at the apex of the curve (location 2) after the markers were placed may be due to the increased confidence imparted to the motorists by the improved view of the curve geometry. The combined effect of these phenomena was the smoothing of the speed profile.

At location 4, in the southbound direction, no difference occurred between the speeds collected in the before and after conditions at the 95 percent level of confidence. As previously stated, this was the only true control site where motorists could neither see nor had passed through the installation. The difference in speeds at location 4 for cars traveling north could be a residual effect of the motorists having just traversed the site. Since the SRPMs caused a smoothing of the speed profile, the motorists seem to be continuing this effect by gradually increasing their speed.

NJ-35, Belmar

Speeds were collected at three locations, northbound and southbound, during rain and dry conditions. At location 1 northbound vehicles could neither see nor had passed through the installation, and southbound vehicles had just gone through the site. Location 2 and 3 were in the site roughly at each end of the installation. Lack of suitable parking places prevented speeds from being measured north of the site or at the apex of the curve.

There appears to be a trend toward a smoother speed profile when the markers were present, with the exception of the cars traveling north in the dry condition (Figure 4). As with the previous analysis on NJ-29, this is probably due to the cue the driver receives concerning road geometry that causes an earlier deceleration. In general, speeds were reduced after the SRPMs were installed. Location 2 for southbound cars showed an increase in speed under wet conditions. The control site, location 1 for northbound vehicles, showed insignificant changes in both speed and speed variance when comparing the before and after conditions (see Tables 8 and 9).

DISCUSSION AND SUGGESTED RESEARCH

Raised pavement markers can be successful in reducing erratic maneuvers and traffic conflicts by altering the variable behavior of traffic with regard to lane placement, choice of exit pathway, and vehicle speeds. Although insufficient lengths of road were marked in order to perform an accident analysis, the reduction of erratic maneuvers accomplished infers the safer use of roadways. Alexander and Lunenfeld (4) describe erratic maneuvers as noncatastrophic system failures on a scale that includes accidents as catastrophic failures. They further state that "erratic maneuvers are symptomatic of driver uncertainty at the navigational level and may cause serious problems for the traffic stream." A reasonable assumption is that most (or all) accidents are preceded by some erratic maneuver or that such a maneuver, apparently inconsequential in the absence of other vehicles, may be disasterous when performed with other cars around. Hence, the reduction of erratic maneuvers can be an indicator of a safer and more efficient use of the roadway.

The types of erratic maneuvers reduced by the presence of raised markers were painted gore en-

Table 8. Analysis of mean speeds and speed variance at NJ-35, dry conditions.

Location	Direction	Measure	Percentage Before	After	Change	Level of Significance (%)
1	Northbound	$\bar{X}$	40.3	40.6	+0.3	<50
		σ	4.1	4.7		
		σ^2	16.8	22.1	+4.3	<95
		n	66	123		
	Southbound	$\bar{X}$	41.1	39.5	-1.6	>99
		σ	4.1	4.5		
		σ^2	16.8	20.3	+3.5	<95
		n	94	147		
2	Northbound	$\bar{X}$	40.7	38.1	-2.6	>99
		σ	4.7	3.7		
		σ^2	22.1	13.7	-8.4	>95
		n	97	180		
	Southbound	$\bar{X}$	40.1	39.9	-0.2	<50
		σ	4.3	4.0		
		σ^2	18.5	16.0	-2.5	<95
		n	89	179		
3	Northbound	$\bar{X}$	41.5	39.4	-2.1	>99
		σ	5.3	4.9		
		σ^2	28.1	24.0	-4.1	<95
		n	109	165		
	Southbound	$\bar{X}$	43.1	40.5	-2.6	>99
		σ	4.3	4.1		
		σ^2	18.5	16.8	-1.7	<95
		n	98	148		

Table 9. Analysis of mean speeds and speed variance at NJ-35, rain condition.

Location	Direction	Measure	Percentage Before	After	Change	Level of Significance (%)
1	Northbound	$\bar{X}$	39.2	38.6	-0.6	<50
		σ	5.1	4.7		
		σ^2	26.0	22.1	-3.9	<95
		n	60	33		
	Southbound	$\bar{X}$	38.9	38.6	-0.3	<50
		σ	4.2	5.3		
		σ^2	17.6	28.1	+10.5	>95
		n	53	52		
2	Northbound	$\bar{X}$	39.3	37.7	-1.6	94
		σ	3.5	3.8		
		σ^2	12.3	14.4	+2.1	<95
		n	44	34		
	Southbound	$\bar{X}$	38.7	39.7	+1.0	82
		σ	4.6	3.0		
		σ^2	21.2	9.0	+12.2	>95
		n	67	41		
3	Northbound	$\bar{X}$	42.1	37.2	-4.9	>99
		σ	4.2	5.1		
		σ^2	17.6	26.0	+8.4	<95
		n	37	55		
	Southbound	$\bar{X}$	42.4	40.2	-2.2	>99
		σ	3.8	4.5		
		σ^2	14.4	20.3	+5.9	<95
		n	46	65		

croachments, centerline and edgeline encroachments, and lane changes and encroachments. Fewer gore encroachments should reduce instances of collisions with the physical gore and reduce conflicts between vehicles already in the deceleration lane and those exiting late, through the gore. One site experienced a significant decrease in gore encroachments at traffic volumes of 400 vehicles/h/lane but no change in the erratic maneuver rate when more than twice that many vehicles were on the road. Apparently, the vehicles themselves can block the view of the markers and prevent following cars from reacting to the treatment. The potential for head-on accidents should be reduced when the number of vehicles encroaching on the opposing lane is decreased. On roads with little or no shoulder, reducing the edgeline encroachments may cause a decrease in fixed-object accidents. There is a concern in some circles that edgeline markings may cause motorists to think there is a lane to the right of the edgeline that perhaps coerces motorists to drive off the road. The results of the study point to the opposite view and show a reduction of vehicles traveling over the edgeline.

Wet weather data were collected at two sites before and after the markers were installed. At one exit, NJ-168, the rate of gore encroachments during rain was not significantly affected by markers. However, at the time of data collection, traffic volumes were at the higher rate previously discussed, and the failure of the markers to reduce gore encroachments may be due to the inability of the motorist to view the devices. At the second site, a left-side exit with two exit lanes, the percentage of lane changing and encroachments was significantly higher during rain without the markers but not significantly different from dry conditions when the markers were placed. This is important documented evidence that raised markers provide significant guidance to motorists under adverse weather conditions, when the visibility of painted lines is severely reduced.

That the markers caused reductions in erratic maneuvers at lit and unlit sites was an unexpected occurrence. This result occurred for each type of site--curves, exits, and bifurcations. This suggests that the treatment of areas with overhead lighting such as intersections and interchanges can provide a safety benefit to motorists and should not be excluded from consideration for the sole reason that they are lit.

Due to the expense of installing SRPMs, decisions have to be made about where and when the markers should be used. Whether spot treatments of locations that are considered hazardous or entire roads should be marked could be the subject of future research, perhaps considering the cost/benefit ratio of each situation if it can be shown that accidents are reduced by the placement of SRPMs. Research may also be useful in choosing among the use of the markers on Interstate and primary highways or two-lane rural roads. Although the former would most likely have higher vehicle miles of travel per lane mile of marked roadway, the dark, winding nature of many rural roads, and the presence of fixed obstacles near to the roadway may point to their being considered a higher priority.

ACKNOWLEDGMENT

This project was sponsored by the New Jersey Department of Transportation and the Federal Highway Administration. The contents do not necessarily reflect the official views or policies of the New Jersey Department of Transportation or the Federal Highway Administration. I wish to acknowledge the assistance of the following persons and groups in performing this work: Eugene Reilly, Richard Hollinger, and Arthur Roberts for their administrative and editorial help; Christopher Graf, Adolphus Phillips, Mike Ferrara, Rena Reali, and the Bureau of Maintenance for their help in the installation of the SRPMs. Gail Flynn and Lorraine Stallings for their preparation of this and other reports generated by this project; and the members of the research staff who participated in the data collection effort under, at times, trying conditions.

REFERENCES

1. G.A. Ferguson. Statistical Analysis in Psychology and Education. McGraw Hill, New York, 1966.
2. P.C. Box and J.C. Oppenlander. Manual of Traffic Engineering Studies. Institute of Traffic Engineers, Arlington, VA, 1976.
3. L.J. Pignatoro. Traffic Engineering Theory and Practice. Prentice Hall, Englewood Cliffs, NJ, 1973.
4. G.J. Alexander and H. Lunenfeld. Positive Guidance in Traffic Control. Office of Traffic Operations, Federal Highway Administration, April 1975.

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

STOP Sign Versus YIELD Sign

HARRY S. LUM AND WILLIAM R. STOCKTON

This paper investigates the relative effectiveness of STOP and YIELD signs at low-volume intersections (less than 500 vehicles/day on minor roadway) in rural and urban environments. Traditional rationales for installing STOP signs, such as inadequate sight distance and high volumes on major roadways, are examined. It is shown that the current use of STOP signs is unrelated to sight distance availability and that STOP signs do not categorically reduce accidents at low-volume intersections. Further, no relation is demonstrated between accidents and major roadway volumes up to 6000 vehicles/day. STOP signs are shown to increase road user costs by more than 7 percent over YIELD signs.

The STOP sign is by far the most prevalent traffic control at intersections. Its message is simple and clear, and the expected response of motorists is a complete cessation of motion (1). The distinct color and shape of the STOP sign result in quick recognition by motorists. Despite its clear meaning, Stockton and others (2), in a study sponsored by the Federal Highway Administration (FHWA), reported that less than 20 percent of the motorists voluntarily complied by completely stopping at STOP signs. (Motorists who had to stop at a STOP sign because of traffic conditions were excluded from the computation.) This compliance rate of 20 percent represents an overall average of three states: Florida, Texas, and New York. A total of 140 intersections in urban and rural environments were sampled. At least one roadway had average daily traffic (ADT) of 500 or fewer vehicles; major road volume ranged up to 36 000 vehicles/day and did not meet the Manual of Uniform Traffic Control Devices (MUTCD) (3) volume warrants for traffic signals.

Dyar (4) also investigated driver's observance of STOP signs at rural and urban intersections in South Carolina. He reported a voluntary compliance rate of 11 percent. Stockton, however, noted that the difference in compliance rates among the three states studied was significant. Such low compliance rates indicate that STOP signs are being used indiscriminately; hence, the sign's purpose of providing for orderly and predictable movement of traffic is defeated.

MUTCD REQUIREMENTS

MUTCD states that (3), to be effective, a traffic control device should meet five basic requirements:

1. Fulfill a need;
2. Command attention;
3. Convey a clear, simple meaning;

4. Command respect of road users; and
5. Give adequate time for proper response.

In practice, the second, third, and fifth requirements are generally met without difficulty. The fourth is dependent on the first requirement, which, of course, is the most critical one. In the eyes of the motoring public, the need must be visible and real, not merely perceived by the traffic engineer or unknowledgeable citizen groups or associations. Excessive and indiscriminate use of STOP signs eventually breeds disobedience and contempt for law enforcement.

The MUTCD warrants provide broad guidelines for the use of two-way STOP control. A STOP sign may be warranted at an intersection where one or more of the following conditions exist (3):

1. Intersection of a less important road with a main road where application of the normal right-of-way rule is unduly hazardous;
2. Street entering a through highway or street;
3. Unsignalized intersection in a signalized area; and
4. Other intersections where a combination of high speed, restricted view, and serious accident record indicates a need for control by the STOP sign.

Conditions 1, 2, and 3 deal with the assignment of right-of-way at an intersection. STOP and YIELD signs both have that function, but the YIELD sign is less restrictive in that all traffic does not have to come to a complete stop. Condition 4 is vague and is open to the engineer's interpretation as to when a STOP sign should be used. Unlike signal warrants, guidelines for quantification of the variables [e.g., speed, restricted view, volume (not stated), accident record] are not discussed.

Warrants for YIELD control are somewhat vague (3):

1. On a minor road at the entrance to an intersection where it is necessary to assign right-of-way to the major road, but where a stop is not necessary at all times, and where the safe approach speed on the minor road exceeds 10 mph;
2. On the entrance ramp to an expressway where an acceleration lane is not provided;
3. Within an intersection with a divided highway, where a STOP sign is present at the entrance to the first roadway and further control is necessary at the entrance to the second roadway, and where the median width between the two roadways exceeds 30 ft;
4. Where there is a separate or channelized right-turn lane, without an adequate acceleration lane; and
5. At any intersection where a special problem exists and where an engineering study indicates the problem to be susceptible to correction by use of the YIELD sign.

The first four conditions are fairly straightforward for the application of the YIELD signs; however, it is not clear as to what is meant by "problem to be susceptible to correction by use of the YIELD sign."

Without specific guidelines to follow, the problem of when to use STOP or YIELD signs becomes one of interpretation of the word need by the individual traffic engineer. It would not be surprising then if the views of engineers differ on the need for STOP or YIELD signs. Tables 1 (2) and 2 (2) give criteria for the application of STOP and YIELD signs by six different traffic agencies. They all agree that sight distance is a critical criterion for STOP control; they disagree as to what the critical approach speed (distance) should be.

Sight Distance

A recent study evaluated the effect of sight distance on choice of control. The sight distance standard used was the American Association of State Highway and Transportation Officials (AASHTO) case 2 (5) requirements. AASHTO case 2 sight distance requires that drivers on all approaches have sight distance sufficient for the relative approach speeds to detect a vehicle on a conflicting approach and stop prior to entering the intersection. Since most of the approaches studied had approach speeds in excess of 25 mph (40 km/h), this test was considerably more conservative than the 10 mph (16 km/h) requirement of the manual (3).

Table 3 gives the frequencies of control types used at 179 approaches (140 intersections) for varying degrees of available sight distance. Sight distance availability is defined as the ratio of available sight distance to the required AASHTO case 2 sight distance. An index value of 1.0 indicates adequate sight distance.

The supposition that STOP signs are used at intersections where sight distance is poor is not supported by the data. Table 4 gives an analysis of the data presented in Table 3. Two null hypotheses are tested: (a) STOP and YIELD signs are used independently of sight distance, and (b) whether an intersection is controlled (STOP and YIELD) or uncontrolled is independent of sight distances.

The minimum discrimination information statistics (MDIS) are both less than the tabulated χ^2 value of 7.841 for 3 df at the 5 percent significance level. Hence, the hypotheses are not rejected. STOP control at low-volume intersections is used in spite of adequate sight distance, and uncontrolled intersections are as likely to have poor sight distance, at least in practice.

With respect to driver behavior, we hypothesized that voluntary stop rate would increase as sight distance decreases. Voluntary stop rate was based on the percentage of drivers who stop in the absence of a conflicting major road vehicle. A regression analysis of voluntary stop rate versus sight distance showed a very poor relation ($r = -0.126$).

Voluntary stop rates were very low for all control types studied (stop = 19 percent, yield = 8 percent, no control = 9 percent). Drivers were observed to slow down to whatever speed was required to evaluate the safety of entering the intersection before choosing a course of action. This behavior appeared to be consistent across all levels of sight distance and control type. Observations of more than 3000 individual movements were made at 140 intersections. Of these, only a small portion exceeded a 5-mph (8-km/h) entry speed (stop = 17 percent, yield = 13 percent, no control = 11 percent). Though not tabulated, most of the entries at speeds greater than 5 mph (8 km/h) were made at less than 10 mph (16 km/h). Therefore the imposition of a 10-mph (16-km/h) sight distance criterion ignores the propensity of the vast majority of drivers to slow well below that speed. Further, it unnecessarily restricts the application of yield and no control at locations where there is no evidence that stop control is superior.

Accident Experience

Does the use of STOP signs help to reduce accidents at low-volume intersections? Table 5 was compiled from Stockton's data (2). The entries in the table represent the number of intersections that experience a given number of accidents over a three-year period (1975-1977, inclusive). The table shows that STOP-controlled intersections exhibit a higher

Table 1. Stop control application criteria other than or in addition to MUTCD criteria.

Location	Volume	Accidents	Sight Distance Criteria	Other
Delaware		Two accidents correctable by STOP within 12 months	Safe approach speed < 24 mph	Minor approaches at school crossings; stop control may be applied on major if more than 2500 ft from previous STOP or YIELD or if minor approach serves 15 or more homes
New York State			Critical approach speed < 8 mph	
North Dakota	Major approach ADT >150 or total ADT >250		< AASHTO case 2	
Baltimore, MD	Major volume >100 vehicles/h		Safe approach speed <5 mph	
		2 in 5 years and sight distance criteria	Safe approach speed 5-10 mph	
		2 in 1 year or 3 in 5 years and sight distance criteria	Safe approach speed >10 mph	
Concord, CA	Major volume >1000 vehicles/day (or 100 vehicles/h) and minor volume >500 vehicles/day (or 50 vehicles/h)	4 in 1 year	Critical approach speed of ≤10 mph	
	Major volume >500 vehicles/day (or 50 vehicles/h) and minor volume 250 vehicles/day (or 25 vehicles/h) and/or accident and sight distance criteria	2 or more within 1 year and/or volume and sight distance criteria	Critical approach speed <15 mph	Two or more criteria must be met
Montgomery County, MD			Sight distance along major approach from 35 ft back; on minor approach, <125 ft	

Table 2. Yield control application criteria other than or in addition to MUTCD criteria.

Location	Volume	Accidents	Sight Distance Criteria	Other
Delaware				Minor approach serves 5 or more homes
New York State			Critical approach speed >8 mph	
North Dakota	<150 vehicles/day on major approach and sight distance criteria		> AASHTO case 2 modified for rural and urban separately and other criteria	Rural, gravel roads only; urban, city streets only
Baltimore, MD				At intersections where STOP is not warranted
Concord, CA	Major street, 500 vehicles/day or (50 vehicles/h); peak and minor street, 250 vehicles/day (or 25 vehicles/h) peak	Two or more of correctable type in 12 months if only STOP warrant met	Critical approach speed between 15 and 20 mph	
Montgomery County, MD			Sight distance along major from 35 ft back; on minor, >125 ft	Some control dictated by geometrics, accidents, or volumes

proportion of intersections that have one or more accidents. Had the one-accident intersections been reported for the uncontrolled classification instead of the STOP-controlled classification, it would lend support to the contention that STOP control helps to prevent accidents.

One possible explanation for this deviation from the expected is that STOP signs were erected after an accident had occurred. However, a rechecking of the records did not indicate such was the case. A second possible explanation is that STOP signs were installed at hazardous intersections. Accident records and field visits to these intersections revealed no evidence of potential hazards. Another possible explanation is that the unusual number of accidents at these STOP-controlled intersections occurred at high-volume intersections. The table below (2) gives the distribution of one-accident intersections by volume and control type. The low cell frequencies preclude statistical analyses. It is certainly unconvincing that STOP signs were used at high-volume intersections.

Volume (vehicles/day)	STOP Sign	YIELD Sign
0-1000	7	1
1000-2000	2	0
2000-3000	1	1
3000-4000	3	1
>4000	0	1

Our interpretation is that accidents at low-volume intersections are rare events, but over a period of time, an accident will occur. In Table 5 each control type shows two intersections that had three or more accidents. Our conclusion is that stop control at low-volume intersections does not categorically help to reduce accidents.

Table 3. Distribution of control type by sight distance at 179 approaches.

	Sight Distance Index				
Control Type	0-0.5	0.51-1.0	1.01-1.5	>1.5	Total
STOP	18	26	16	9	69
YIELD	7	27	17	11	62
Uncontrolled	11	14	16	7	48
Total	36	67	49	27	179

Table 4. Analysis of information table for data presented in Table 3.

Component	MDIS[a] ($2\hat{I}$)	df
Independence between STOP and YIELD control	4.885	3
Independence between control and no control	2.338	3
Total independence	7.223	6

[a]In this paper Kullback's information-theoretic approach to the analysis of contingency tables was used instead of the conventional Pearson's chi-square test. The minimum discrimination information statistics (MDIS) whose symbolic representation is $2\hat{I}$ and is asymtotically distributed as χ^2 for large sample and for a wide class of problems (6, p. 393). The formula used to calculate $2\hat{I}$ is $2[\Sigma_i\Sigma_j n_{ij}\ln(n_{ij}) + n\ln(n) - \Sigma_i n_i \ln(n_i) - \Sigma_j n_j \ln(n_j)]$, where $\Sigma_i n_i = \Sigma_j n_j = n$.

Table 5. Distribution of accident frequency by intersection and control type.

	No. of Accidents				
Control Type	0	1	2	≥3	Total
STOP	33	13	0	2	48
YIELD	40	4	2	2	48
Uncontrolled	42	0	0	2	44
Total	115	17	2	6	140

Table 6. Distribution of intersections with accidents by major road volume.

	No. of Intersections		
ADT	Without Accidents	With Accidents	Total
0-1000	68	10	78
1001-2000	12	1	13
2001-3000	9	4	13
3001-4000	7	1	8
4001-5000	2	3	5
5001-6000	4	0	4
>6000	13	6	19
Total	115	25	140

<u>Major Roadway Volume</u>

Both YIELD and STOP signs have the function of assigning the right-of-way, generally to the roadway that has the higher volume. Data collected by Stockton (2) on major roadway traffic varied from 1000 to 36 500 ADT. To determine if there is a relation between volume and accident experience, Table 6 (2) was constructed. Volume was grouped by increments of 1000 ADT except for the last group, which included all intersections with more than 6000 ADT. The total MDIS of 12.0833 is slightly less than the tabulated χ^2 value of 12.6 for 6 df at the 5 percent significance level, which indicates that there is no relation between accident experience and traffic volume.

Our tentative conclusion is that, up to 6000 ADT, the YIELD or the STOP sign may be used to assign the right-of-way. It is not clear, however, that the 6000 ADT is the upper bound of no association between intersection accidents and traffic volume. It will have to be established in future studies how much higher than 6000 ADT the upper bound might be.

PAST SIGNING PRACTICE

Signing for traffic control, unlike signalization, is passive; it cannot accommodate changing traffic conditions. Understandably, the traffic engineer with safety uppermost in his or her mind, would choose the normally conservative but more restrictive STOP sign in preference to the YIELD sign. This may have been acceptable engineering practice years ago, but the proliferation of STOP signs has made drivers skeptical and disbelieving of a need for the STOP sign when they see one. Dyar (4) reported no significant difference in driver's observance of STOP signs with or without special control measures at rural intersections with inadequate sight distance. It is not clear from the report whether these special control measures were used at selected hazardous intersections. The special control measures included

1. STOP sign larger than the standard 30-in sign,
2. STOP signs installed on both the left and right shoulders of the controlled approach,
3. Red flashing lights at STOP-controlled approaches and amber flashing lights for through streets,
4. Larger rectangular overhead sign with the word STOP suspended above the intersection, and
5. Combinations of the above.

It is evident that the STOP sign has lost its meaning. Drivers treat it as a YIELD sign--slow down and then proceed with caution.

YIELD SIGN

The YIELD sign (trapezoidal) was introduced in 1951 in Tulsa, Oklahoma, and incorporated into the 1955 revision of the national MUTCD as an equilateral triangle with one corner pointed downward with black lettering on a yellow background (7), which was later changed to the now familiar red on white. It is less restrictive than the STOP sign and definitely assigns the right-of-way to the major road. The Supreme Court of South Dakota ruled that [State v. Muhs, 137 N.W., 2nd 237, 239 (S.D., 1965)]:

> The only difference between a STOP sign and YIELD sign is the duty always exists to stop and look effectively (at) the STOP sign, and for a YIELD sign the duty is to slow down, effectively look to see if the highway is free from oncoming traffic, and stop if necessary to yield the right-of-way....

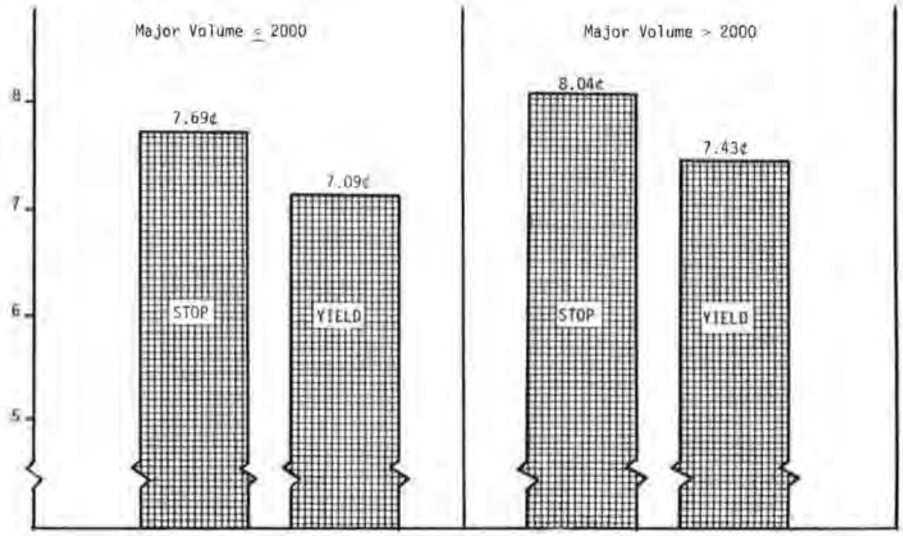
Then, why is the YIELD sign not in greater use? There are several reasons:

1. The application of YIELD signs would require engineering studies, whereas little or none is required for STOP signs if conditions 1, 2, or 3 of the STOP warrant are used;
2. The belief by engineers that a single ultimate policy of stop control prepares them for all eventualities against tort liability; and
3. Political pressure from citizen groups in the mistaken belief that STOP signs offer greater protection than YIELD signs.

RELATIVE EFFICIENCY

Total road user cost per cycle was estimated from

Figure 1. Costs per cycle of stop and yield control.



more than 3000 observations at both stop- and yield-controlled intersections. This cost included both the vehicle operating cost and the delay cost, and was based on entry speed and travel time through the intersection. Figure 1 shows the cost differentials for major roadway volumes above and below 2000 vehicles/day (the point of significant difference in driver behavior). Yield control offers a 7.8 percent reduction in total cost below 2000 vehicles/day, and a 7.6 percent reduction at the higher volume level.

SUMMARY

Traffic control is highly visible and sensitive to public scrutiny. Understandably, the traffic engineer must consider and accommodate all drivers--the novice and the experienced, the familiar and unfamiliar, the defensive and aggressive. The task is not an easy one. The ultimate measure of successful traffic control is a good safety record. This can only happen through public understanding and acceptance of control devices. Many of the STOP signs at low-volume intersections are unjustified (although warranted by MUTCD) and could be replaced by YIELD signs without increasing accident experience. Furthermore, the use of YIELD signs would restore respect and effectiveness of the STOP sign and improve operating efficiency. The path of least resistance of a single policy of STOP control is contraindicated by the low rate of obedience to the STOP signs. Our findings are summarized below:

1. The low rate of driver compliance to the STOP sign is a result of its excessive use at intersections where it is not reasonable and necessary that all motorists stop,
2. There is no relation between major road traffic volume (up to 6000 ADT) and accident experience at low-volume intersections,
3. STOP signs do not reduce accident experience at low-volume intersections,
4. The supposition that STOP signs are being used at locations with poor sight distance as defined by AASHTO is not supported by data, and
5. STOP signs result in a higher road user cost than do YIELD signs.

In keeping with the philosophy that the least-restrictive device consistent with safety and smooth traffic flow should be used, the basic question is asked: When should the STOP sign be used? The answer may not be so easy.

ACKNOWLEDGMENT

Much of the material in this paper has been taken from the study STOP, YIELD and No Control at Intersections sponsored by FHWA. The findings and opinions are ours and do not reflect those of FHWA.

REFERENCES

1. Uniform Vehicle Code. rev. ed., National Committee on Uniform Traffic Laws and Ordinances, Washington, DC, 1968.
2. W.R. Stockton, R.Q. Brackett, and J.M. Mounce. STOP, YIELD and No Control at Intersections. Federal Highway Administration, Rept. FHWA/RD-81/084, June 1981.
3. Manual on Uniform Traffic Control Devices for Streets and Highways. Federal Highway Administration, 1978.
4. R.D. Dyar. An Analysis of Driver Observance of Stop Signs. Clemson Univ., Clemson, SC, unpublished master's thesis, May 1977.
5. S. Kullback. Information Theory and Statistics. Wiley, New York, 1959.
6. A Policy on Geometric Design of Rural Highways. AASHTO, Washington, DC, 1965.
7. G.M. Sessions. Traffic Devices: Historical Aspects Thereof. Institute of Traffic Engineers, Washington, DC, 1971.

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

Energy and Emission Consequences of Improved Traffic Signal Systems

SUK JUNE KAHNG AND ADOLF D. MAY

The primary objective of this study was to evaluate the impacts of selected strategies for improvement of traffic signal systems and to develop policy guidelines for the strategies in light of current realities such as increasing passenger delay on surface streets, high costs and scarcity of fuels, and concern about the environment. The existing simulation and optimization model, TRANSYT6C, was applied to a selected study arterial, San Pablo Avenue in Berkeley, California. Two basic categories of traffic signal timing improvement strategies were evaluated: (a) splits and offsets optimization and (b) optimal cycle length selection. A series of sensitivity analyses was conducted to determine variations in the impact effects of the strategies under different operational environments in terms of changed levels of traffic flow. The effects of different objective functions were also investigated and included. The major findings of this investigation include the following. For a given cycle length, optimization of splits and offsets based on either the minimization of passenger delay or fuel consumption also led to near-minimum value for all other measures of effectiveness. Passenger delay and vehicle emission were further reduced by shorter cycle lengths; however, total stops were further reduced by longer cycle lengths. Fuel consumption was relatively less sensitive to changes in cycle length. As the level of traffic flow increased, a moderate cycle length rather than a short cycle length was preferred in order to minimize fuel consumption. Trade-offs between passenger hours saved per gallon of fuel consumed were identified for different cycle lengths and flow levels.

In recent years emphasis in transportation planning has shifted from long-term, capital-intensive, capacity-increasing construction projects to shorter-term, relatively low-cost projects aimed at using existing transportation facilities more efficiently. The importance of energy conservation and environmental impact analysis is also being stressed. This trend in the transportation engineering field placed heavy emphasis on transportation system management (TSM) as a part of the planning process and as a prerequisite for improvements to increase the capacity of the urban transportation system (1). One of the typical elements of TSM planning is optimization of traffic signals in terms of energy saving, reduction in vehicle emissions, and increase in the productivity of transportation systems.

Control of traffic signals is by far the most common type of control at heavily trafficked intersections in urban areas. Inefficient use of the transportation system results when traffic signals are set without the aim of optimizing them. The byproducts of such situations include greater fuel consumption, increased vehicle emissions, increased travel time, higher accident rate, and less reliable services. According to Federal Highway Administration (FHWA) data (2), fuel consumption could be reduced by 100 000 barrels of crude oil per day if the timing of the 130 000 coordinated, signalized intersections that currently exist along the nation's urban streets were made optimum. Thus, signal retiming optimization is regarded as one of the most obvious TSM strategies to implement and one of the most cost-effective energy, pollution, and cost-conservation measures available in transportation.

Signalized intersections can be classified into two types: (a) an individual intersection and (b) a network that is comprised of two or more intersections and streets that link those intersections. For the analysis of individual intersection capacity and performance, the critical movement method (3) is being developed as a part of a National Cooperative Highway Research Program (NCHRP) project. Other useful analytical methods include the U.S. highway capacity manual (HCM) (4, pp. 111-159); British (5), Australian (6), Swedish models (7); the signal operations analysis package (SOAP) (8); and network simulation (NETSIM) (9) methods. For an arterial network that is comprised of a number of signalized intersections, the coordination of traffic signals along the route is regarded as one of the most efficient ways to improve total system performance by reducing delay, stops, fuel consumption, and vehicle emissions. Cycle length, splits, and offsets of traffic signals in the system need to be evaluated and made optimum to improve total system performance.

One of the most important analytical tools of signal time optimization in arterial network is computer simulation. Traffic simulation models can be used to analyze existing conditions as well as to predict the shorter-term and longer-term impacts of traffic control strategies on selected measures of effectiveness (MOE) like fuel consumption, vehicle emission, travel time, and number of stops. In 1973 modeling efforts for arterial networks in terms of signal optimization were initiated by the Institute of Transportation Studies (ITS) of the University of California at Berkeley. A literature review revealed the existence of the traffic network study tool (TRANSYT) (10) model, developed by the British Transport and Road Research Laboratory (TRRL), that could perform similar tasks. Many versions of the TRANSYT model have been developed by various organizations throughout the world to meet their transportation needs. These modified TRANSYT versions include TRANSYT/6 (11), TRANSYT6C (12), TRANSYT/6N (13), TRANSYT/7 (14), TRANSYT7F (15), and TRANSYT/8 (16). The developers and features of several TRANSYT versions are included in Table 1. Although a traffic performance measure such as delay or travel time is often the only impact considered in most versions, the newly arising considerations of traffic management in terms of fuel consumption, vehicle emission, and priority treatment are addressed directly in the TRANSYT6C model. This model is selected for the purpose of this study because emphasis is placed on various impacts evaluations and more flexible objective functions.

TRAFFIC SIGNAL UPGRADING TOOL

Investigation of traffic signal upgrading strategies in the field can be expensive and time consuming. Unexpected and unnecessary congestion may result and cause negative citizen reaction. There is a need to develop and use computer models to evaluate the impacts of various strategies for upgrading traffic signals in different operating environments.

Overview of TRANSYT6C

TRANSYT6C is a macroscopic, deterministic model used to simulate and optimize arterial network signal timings. The model is based on TRANSYT/6 (11), developed by TRRL, and was extended and tested by Clausen, Jovanis, May, Kruger, and Deikman at ITS to include fuel and emission estimates, spatial and

Table 1. Developers and features of TRANSYT versions.

Model	Developer	Date	Features
TRANSYT/1	D.I. Robertson, Transport and Road Research Laboratory	1967	Base model
TRANSYT/2	Transport and Road Research Laboratory	1968	FORTRAN IV rewrite of TRANSYT/1; provision for more than three-stage signals
TRANSYT/3	Transport and Road Research Laboratory	1970	Reorganized data input facilities and checking
TRANSYT/4	Transport and Road Research Laboratory	1971	Allowed the calculation of initial signal settings to equalize saturation; flow pattern graphs added
TRANSYT/5	Transport and Road Research Laboratory	1972	Provided multiple links at a common stopline and bus progression speed and time-at-stops along a link
TRANSYT/6	Transport and Road Research Laboratory	1975	Included an improved model of stops; program could also be more readily compiled on some smaller computers
TRANSYT/6C	P. Jovanis, A.D. May	1977	Fuel-consumption and vehicle-emission estimation, priority treatment, demand response prediction, and improved performance index; U.S. units
TRANSYT/6N	R. Akcelik, National Capital Development Commission	1978	Link groups, intersection statistics, and search for optimal cycle time
TRANSYT/7	Transport and Road Research Laboratory	1978	Cheaper optimization, bottleneck facility added, and input format modification
TRANSYT/7F	C. Wallace, K. Courage	1981	Time-space diagram, fuel-consumption estimation; U.S. terminology
TRANSYT/8	Transport and Road Research Laboratory	1980	Gap-acceptance model, fuel-consumption estimation, monetary costs of delay and stops, preliminary output for the selection of cycle time, and maximum queue length control

Figure 1. Overview of TRANSYT6C model.

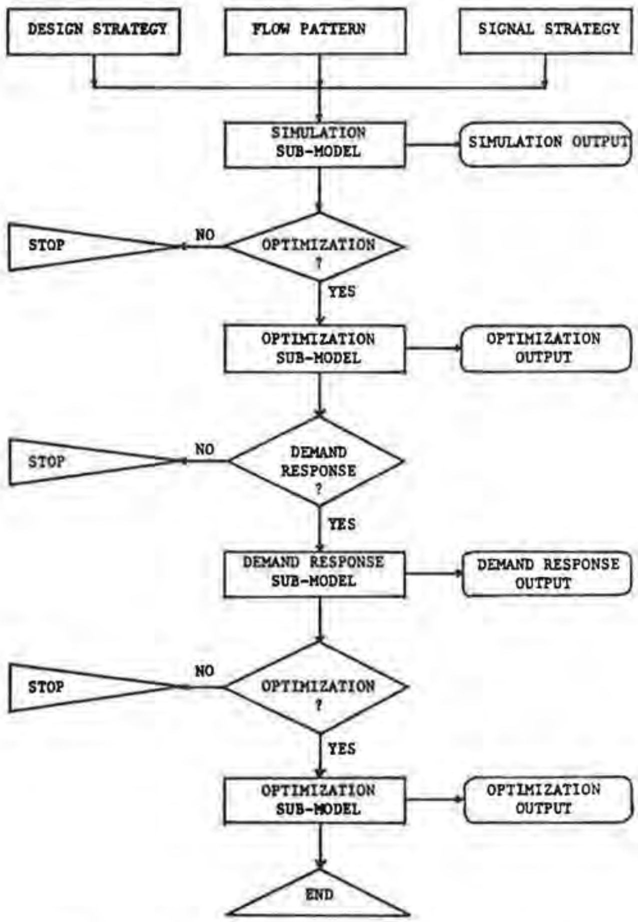
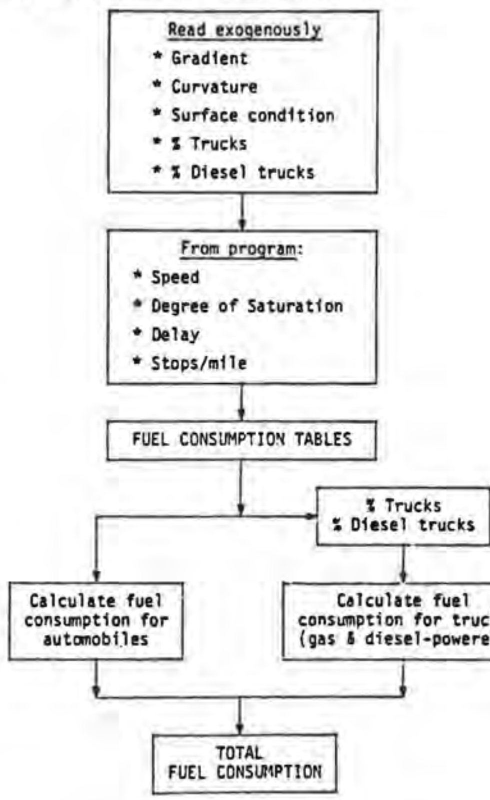


Figure 2. Structure of fuel-consumption submodel.



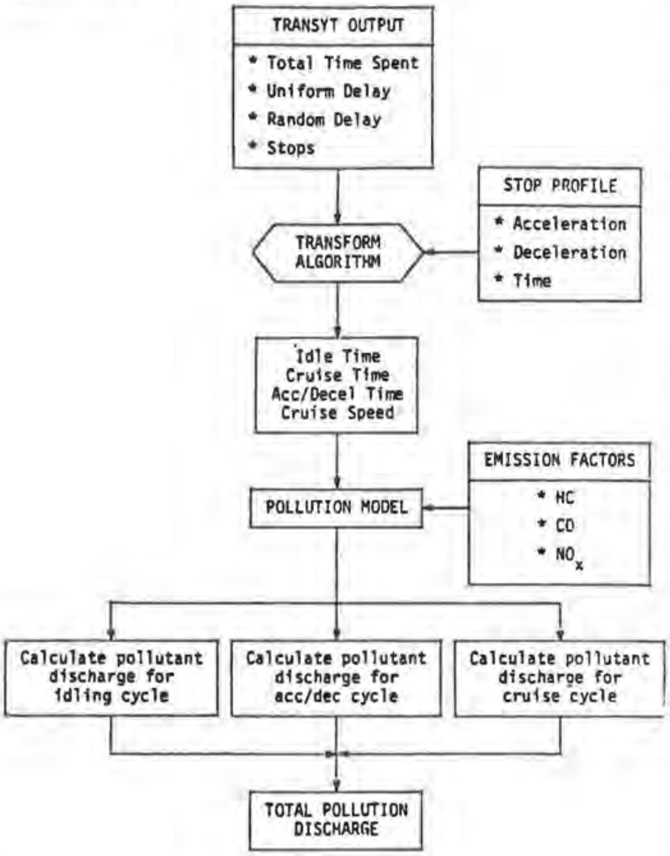
demand responses, and reformulated objective functions. The results of their research have been documented in various papers and research reports (12,17-19).

The program requires as input a description of the roadway design, flow pattern, and signal strategy. The model represents vehicles as platoons that change as vehicles proceed through signals and disperse along a route. The arterial network is represented as a series of nodes (intersections) connected by a series of unidirectional links. It provides as output traffic performance for each link measured by the following variables: estimate of the fuel consumption and vehicle emission impacts, time spent, distance traveled, uniform and random delay, number of stops, maximum uniform queue, and degree of saturation. The individual link values are summed to arrive at measures of system performance. Traffic signal optimization uses hill-climbing techniques that search the response surface for a minimum value of a performance index. After the signals are optimized, the demand response submodel may be engaged. After the demand response occurs, the signals may be optimized again for the new flow conditions. An overview of the TRANSYT6C model is shown in Figure 1.

Table 2. Adjustment factors for change in fuel economy by year.

Year	Factor	Year	Factor
1974	1.000	1981	0.831
1975	1.000	1982	0.791
1976	0.980	1983	0.747
1977	0.955	1984	0.748
1978	0.931	1985	0.670
1979	0.902	1986	0.638
1980	0.871	1987	0.612

Figure 3. Overview of vehicle emission submodel.



Impacts Prediction and Demand Response

Since energy and cost saving is a prime concern, the capability of the TRANSYT model to predict vehicle emission and energy consumption is becoming one of its most important features. With this model, users can generate optimal signal settings that will minimize a weighted formula of fuel, emissions, delay, and stops and also differentiate between priority and nonpriority vehicles. The model can also be used to evaluate alternative design and control plans with reversible lane, priority lane, and one-way street operations.

Fuel Consumption

Energy estimates in the TRANSYT6C model are based on fuel-consumption rates developed by Claffey (20). The tables are entered for each link by using traffic data developed from TRANSYT output and a series of user-specified values that describe geometric conditions. All values between table entries are obtained by linear interpolation. The three driving aspects considered in computation of fuel consumption are cruise, acceleration-deceleration, and stopped time. Additional computations are made to this procedure to calculate fuel consumption for priority vehicle links. Overall structure of the fuel-consumption submodel is shown in Figure 2.

Automobile fuel economy has been improving steadily over the last few years and will continue to improve in the future; therefore, a method of updating the fuel figures in the model may be needed. Based on a California Department of Transportation report (21), annual adjustment factors for fuel economy change by year are shown in Table 2.

According to one of the authors of the report (21), a base year of 1974 is assumed as the year when the average vehicle on the road had fuel economy characteristics similar to those of the Claffey vehicles (20). Use of adjustment factors is simple and no program modifications are required. To update a TRANSYT6C fuel estimate for 1981, for example, multiply it by 0.831.

Vehicle Emissions

An overview of the vehicle emission submodel in TRANSYT6C is shown in Figure 3. Based on a report by Kunselman and others (22) to the U.S. Environmental Protection Agency (EPA), a simplified version of a vehicle emission model was developed and incorporated into the TRANSYT6C model. The model postulates that the amount of a particular pollutant emitted can be computed by multiplying the emission factor for the pollutant by the extent of driving done in each aspect. For each individual link, the three driving aspects considered in calculation of vehicle emission are cruise, idle, and acceleration-deceleration. TRANSYT6C contains separate treatments for automobile and bus links and the additional calculation for bus emissions were included in the model. Individual link emissions are then summed to compute total emissions for the arterial network.

Demand Response

Traffic management strategies may alter an individual choice of route, mode, time of travel, or rate of travel making. The demand responses will result in a change in traffic performance and thus a change in impacts (fuel consumption and vehicle emission). The amount of each type of response depends on the characteristics of the trip, the characteristics of the trip maker, and the characteristics of the transportation system. The TRANSYT6C model applies the general formulation that a response is a function of a stimulus times a sensitivity. The stimulus for demand responses is the change in vehicle travel time computed in TRANSYT6C. The sensitivity reflects the traveler's awareness and opportunity to take advantage of the change in travel time. The sensitivity is user specified and is applied to the submodels of spatial and model demand responses. For the purpose of computing changes in travel time, the study arterial is divided into segments as specified by the user. The segments should correspond as closely as possible to the average trip length (miles) on the arterial. Then, the change in the traveler's trip time for the average trip length is estimated by computing change in travel time for each segment. The demand responses are treated sequentially. A driver is assumed to alter his or her route, if possible, before changing mode.

Performance Index

Today traffic management emphasizes consideration of energy and environmental impacts as well as passenger mobility. In keeping with these new concerns of traffic management, the following performance index (PI) is introduced in the TRANSYT6C model:

$$PI = \sum_{i=1}^{n} [(K_1 d_i)_{NP} + (K_2 S_i)_{NP} + (K_3 f_i)_{NP} + (K_4 d_i)_P + (K_5 S_i)_P + (K_6 f_i)_P + K_7 CO_i + K_8 NO_i + K_9 HC_i] \quad (1)$$

where

$K_1, K_2, \ldots, K_9$ = weighting factors;
i = link i;
n = number of links;
d_i = delay on link i;
S_i = stops on link i;
f_i = fuel consumed on link i (gal);
CO_i = carbon monoxide emitted on link i (kg);
NO_i = nitrous oxide emitted on link i (kg); and
HC_i = hydrocarbons emitted on link i (kg).

NP and P refer to nonpriority and priority vehicles, respectively. By selecting different values for the weighting factors, the user may include or exclude certain variables from the PI. For example, if fuel consumption for priority and nonpriority vehicles is desired in the PI, K_3 and K_6 may be set to one and all other weights set to zero. It is even possible to assign dollar values to each weighting factor and to attempt to set signals to minimize the total cost of the impacts considered. This new performance index permits the direct evaluation of timing signals for energy, emission, and cost savings as well as for traffic performance.

MODEL APPLICATION

In order to test the utility of the TRANSYT model as a tool for upgrading traffic signals and to assist in the development of policy guidelines, the model has been applied to several operational environments. A specific location in the San Francisco Bay Area was used and, through sensitivity analysis, expanded to represent a wide cross section of operational environments.

Site Description and Data Base

In the San Francisco Bay Area, San Pablo Avenue in Berkeley was selected as the study arterial. San Pablo Avenue runs parallel to Interstate 80 in the East Bay and the street is important as an alternate route to travelers on the Eastshore Freeway. It extends from the Oakland central business district (CBD) on the south through Berkeley to Albany, El Cerrito, and Richmond on the north. The street carries two-way operation, three lanes in each direction, on a 74-ft width with no parking on either side at the time of the study. Although the arterial is not heavily congested, it carries a significant number of local buses. The 2.75-mile section on San Pablo used has nine intersections. The representation of San Pablo Avenue used in the TRANSYT6C model application is shown in Figure 4. In the figure the circled numbers represent intersections, and directional arrows represent links.

The study section consists of nine signalized intersections with a common (fixed) 70-s cycle. Because of slightly more critical operational problems, the evening peak hour was selected for the model application. A previous study by ITS (23) developed data to be used in TRANSYT for the study section. Local traffic operations engineers in Berkeley examined the TRANSYT output and agreed that the traffic performance given by the model was a realistic representation of peak-hour conditions (12).

The following characteristics of the study section at the time of the study were used as input to the model: (a) bus flows 13-18 vehicles/h in both directions; (b) average bus occupancy of 30 passengers; (c) average automobile occupancy of 1.2 passengers; (d) vehicle mix of approximately 2 percent trucks and buses, 50 percent of which are diesel; (e) roadway was straight and level; (f) directional split along the study section was approximately 60-40 with predominent flow northbound.

Objective Functions Selection and Optimization

By modifying PI, signal settings may be optimized to satisfy different objective functions. The new PI equation allows a detailed evaluation of impacts and the consequences of different impact objectives. The equations below give traffic management objectives and corresponding PIs employed in the San Pablo study site.

The PI to minimize total passenger delay is

$$PI = \sum_{i=1}^{n} [(d_i)_{NP} + (d_i)_P] \quad (2)$$

Figure 4. Site characteristics.

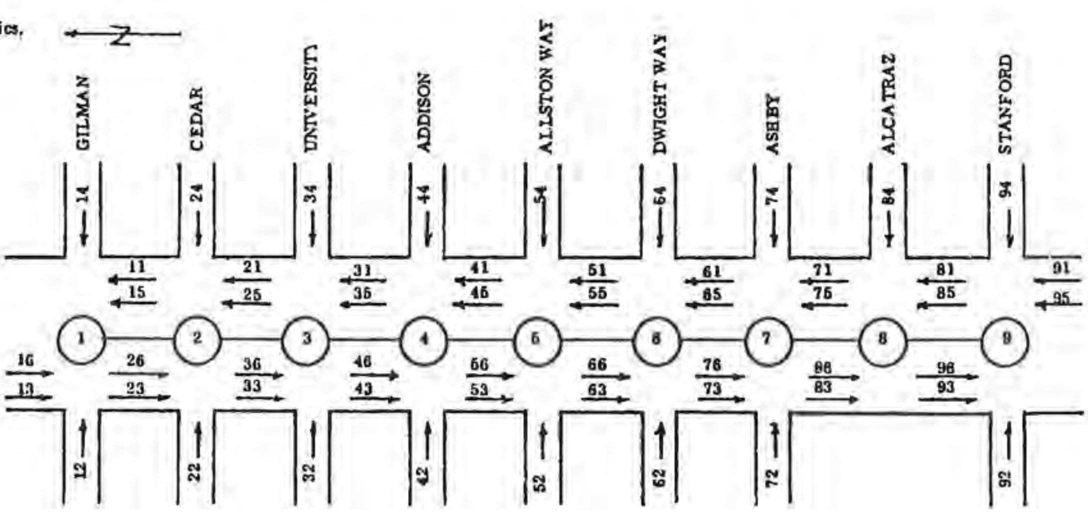


Table 3. Design of experiment with TRANSYT6C.

Flow	Minimization Objective	Run No. for Existing Spacing			Run No. for Half Spacing	
		50-s Cycle	70-s Cycle	90-s Cycle	50-s Cycle	70-s Cycle
Existing	Total delay	1	11	21	31	41
	Priority delay	2	12	—[a]	32	42
	Total stops	3	13	23	33	43
	Total fuel	4	14	24	34	44
	Total emissions	5	15	25	35	45
50 Percent Greater	Total delay	6	16	26	36	46
	Priority delay	7	17	—[a]	37	47
	Total stops	8	18	28	38	48
	Total fuel	9	19	29	39	49
	Total emissions	10	20	30	40	50

Note: Entries in cells are production run numbers.
[a]Run not made.

Figure 5. Effect of optimized splits and offsets on reducing impact under existing flow conditions.

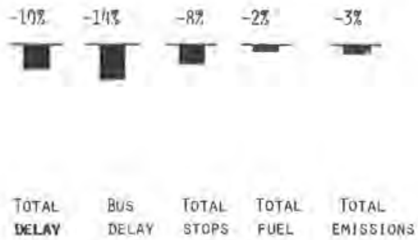
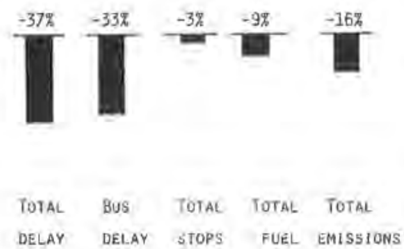


Figure 6. Effect of optimized splits and offsets on reducing impacts under increased flow conditions.



The PI to minimize priority passenger delay is

$$PI = \sum_{i=1}^{n} (d_i)_P \tag{3}$$

The PI to minimize total vehicle stops is

$$PI = \sum_{i=1}^{n} [(S_i)_{NP} + (S_i)_P] \tag{4}$$

The PI to minimize total fuel consumption is

$$PI = \sum_{i=1}^{n} [(f_i)_{NP} + (f_i)_P] \tag{5}$$

The PI to minimize total vehicle emissions is

$$PI = \sum_{i=1}^{n} (CO_i + NO_i + HC_i) \tag{6}$$

where

- n = number of links,
- i = link i,
- d_i = delay on link i (passenger hours),
- S_i = vehicle stops on link i (vehicle stops/h),
- f_i = gasoline consumed on link i (gal/h),
- CO_i = carbon monoxide emitted on link i (kg/h),
- NO_i = nitrous oxide emitted on link i (kg/h), and
- HC_i = hydrocarbons emitted on link i (kg/h).

NP and P refer to nonpriority and priority vehicles, respectively.

In addition to these basic runs with different objective functions, a number of sensitivity tests were performed in terms of different cycle lengths, traffic flows, and signal spacings (although production runs were made for different signal spacings, their results have not been analyzed and are not discussed in this paper) to determine variations in results from the basic runs. Examination of minimum green time due to pedestrian requirements indicated that a 50-s cycle length was the shortest that could be used to the study section. In order to select a cycle length that was equally spaced but greater than the existing cycle length, a 90-s cycle was chosen.

Each of the three cycle lengths was tested under two different flow conditions: existing flows and existing flows increased by 50 percent. Only the cycle length and flow were changed for each specific basic run mentioned above. The impacts for each cycle length after optimization were compared with the impacts for the existing signal timing with the same flow conditions. Table 3 summarizes the design of these sensitivity production runs as well as basic runs.

TRAFFIC SIGNAL UPGRADING POLICY GUIDELINES

Split and Offset Optimization

The results of optimization runs were compared with the existing condition runs in terms of passenger delay, vehicle stops, fuel consumption, and vehicle emissions. Figures 5 and 6 show the effect of optimized splits and offsets on reducing those impacts under existing flow conditions and increased (by 50 percent) flow conditions, respectively. The existing common cycle length of 70 s was employed in both cases.

As can be seen in Figures 5 and 6, all the impacts have been reduced by optimizing splits and offsets of signals along the study network. Under the existing flow condition in Figure 5, the optimization of splits and offsets resulted in 10 percent reduction in total passenger delay (14 percent reduction in bus delay), 8 percent reduction in total vehicle stops, 2 percent reduction in total fuel consumption, and 3 percent reduction in total vehicle emission. The reductions in those impacts by splits and offset optimization under increased flow conditions are greater than those under the existing flow conditions except the reduction in vehicle stops. Figure 6 shows the estimated impacts reduction by the optimization, which ranges from 3 percent reduction in total stops to 37 percent reduction in total passenger delay.

These impact reductions were achieved by adjust-ing existing splits and offsets in the optimization computer runs to minimize total passenger delay. Although splits are usually set to give equal degrees of saturation to two critical traffic flows at an intersection and offsets are set to give good progression to the traffic flow by manual method, the comparison of the optimized signal timings with existing ones reveals that it might be necessary to give a preference to the predominant traffic flow in terms of splits and offsets to optimize the total system performance. Although the flow in Figure 6 was increased by 50 percent from the existing condition in Figure 5, the impact reductions by optimizing splits and offsets are almost three times those obtained in Figure 5. This result implies that even the best signal timing for the present flow level might not be the best one if the flow level were to change in the future. Therefore, it would be desirable to adjust traffic offsets and splits as the traffic flow level changes in the future, even though the present signal timing has been best optimized for the current flow level.

Cycle Length Selection

Figures 7 and 8 show the effect of cycle length with optimized splits and offsets on reducing impacts under existing flow conditions and increased flow conditions, respectively. Under existing flow conditions, as can be seen in Figure 7, the total delay and emissions were further reduced by employing a shorter cycle length (50 s) while further reductions in total steps were achieved by employing a longer cycle length (90 s). Although Figure 8 also shows this general trend, note that measured impacts reduction are less sensitive to the variation of cycle length under increased flow conditions. Compared with other MOEs, in Figures 7 and 8, fuel consumption seems to be relatively less sensitive to the changes in cycle length. As the level of traffic flow increased, a moderate cycle length rather than a short cycle length was preferred in order to minimize fuel consumption.

Although the optimization of splits and offsets could reduce all the impacts for those three different cycle lengths (except total stops in 50- and 70-s cycle length, and total delay in 90-s cycle length), the figures show that the effect of the change in each cycle length on total delay and emission might be opposite to its effect on total stops to some degree (especially under existing flow conditions, in this case).

Depending on the objective of signal timing optimization, the optimal cycle length can vary from a short cycle to a long cycle length. Generally speaking, a reasonably short cycle length is preferred to a long cycle length to minimize total delay and emissions and vice versa. However, several cycle lengths should be tested by using the computer simulation model and the results should be examined carefully before applying the general effect of cycle length described above.

Objective Function Selection

TRANSYT6C can employ various objective functions by simply modifying PI; therefore, it might be necessary to provide general guidelines for the selection of an appropriate objective function for system optimization. Based on the extensive sensitivity analyses of objective functions, the objective function of either minimizing total fuel consumption or minimizing total delay also reduces all other impacts. Thus, either of them is regarded as the best single objective function. Figures 9 and 10 show the effect of these objective functions on reduction in total delay and total fuel consumption. As can be seen, both objective functions reduced not only total delay but total fuel consumption for different cycle lengths employed.

The differences in the results of the objective functions are regarded as a kind of trade-off between fuel-consumption reduction and total delay reduction. Figure 11 is included to discuss this trade-off in terms of passenger hours saved per gallon given up. For the network under study, from 0.5 to 4.0 passenger-h savings have to be given up to save 1 gal of fuel if one attempts to further reduce fuel consumptions below the fuel-consumption-reduction level achieved by the objective function of minimizing total delay. Depending on the perceived relative value of fuel and passenger hours, therefore, either of those objective functions can be employed to meet the specific objective. Alternatively, the relative values expressed in numerical terms can be assigned to weighting factors of fuel and delay items in the PI, respectively.

SUMMARY AND FUTURE RESEARCH

The primary objective of this study was to evaluate the impacts of selected strategies to improve traffic signal systems and to develop policy guidelines for the strategies in the light of current realities such as increasing passenger delay on surface streets, high costs and scarcity of fuels, and concern about the environment.

The existing simulation and optimization model, TRANSYT6C, was applied to a selected study arterial, San Pablo Avenue, Berkeley, California. The study network consists of nine signalized intersections that have a 70-s common cycle length. Although evening peak hour flow conditions were used as inputs to the study, the network was not heavily congested at the time of study.

Two basic categories of strategies for improvement of traffic signal timing were evaluated: (a) splits and offsets optimization and (b) optimal cycle length selection. A series of sensitivity analyses was conducted to determine variations in the effects of the strategies under different operational environments in terms of changed traffic flow levels. The effects of different objective functions were also investigated and included. The major findings of this investigation include the following.

1. For a given cycle length, optimization of splits and offsets based on either the minimization of passenger delay or fuel consumptions also led to near-minimum value for all other MOE.
2. Passenger delay and vehicle emissions were further reduced by shorter cycle lengths; however, total stops were further reduced by longer cycle lengths. Fuel consumption was relatively less sensitive to changes in cycle length. As level of traffic flow increases, a moderate cycle length rather than a short cycle length was preferred in order to minimize fuel consumption.
3. Trade-offs between passenger hours saved per gallon of fuel consumed were identified for different cycle lengths and flow levels.

Considerable investigations to reduce various impacts have been conducted through the traffic signal timing improvement by using TRANSYT6C, but constraints of time and budget prohibited the further investigation of potentially fruitful areas of research. Future research may be divided into three basic categories: additional model application and sensitivity analyses, model modification and expan-

Figure 7. Effect of cycle length on reducing impacts under existing flow conditions with optimized splits and offsets.

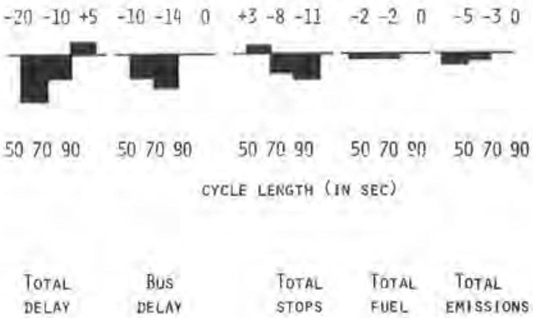
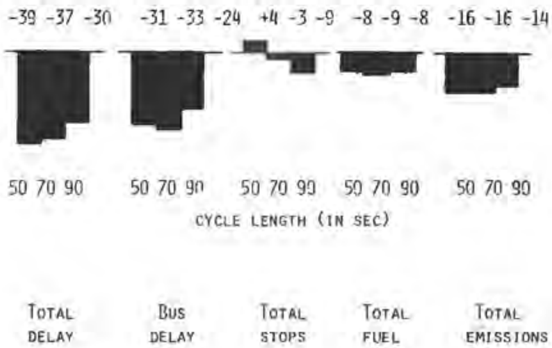


Figure 8. Effect of cycle length on reducing impacts under 50 percent increased flow conditions with optimized splits and offsets.



sion, and development of traffic-responsive signal control system.

Additional model applications include the following:

1. Evaluation of additional traffic management strategies such as bus and carpool lanes or paratransit for various traffic flows that have different composition characteristics;
2. Further testing with the multiple objective function to minimize the total cost of impacts evaluated;
3. Further application of the model to other arterial networks with different characteristics from the study site in terms of geometry, flow pattern, and signal timings; and
4. Sensitivity analysis of improved traffic signal timings to the traffic flow variations such as morning peak flow, evening peak flow, and off-peak flow.

Possible areas for model expansion and modification are as follows:

1. Inclusion of additional impacts such as operating costs, safety, and noise pollution;
2. Inclusion of origin-destination information in the model;
3. Inclusion of additional demand responses such as temporal shift or a change in the rate of trip-making;
4. Modification of the model to reduce the computing time in large networks and handle bottleneck situations; and
5. Field validation studies in terms of fuel-consumption and vehicle-emission estimation.

A possible area for the development of traffic-re-

Figure 9. Effect of objective functions on total delay.

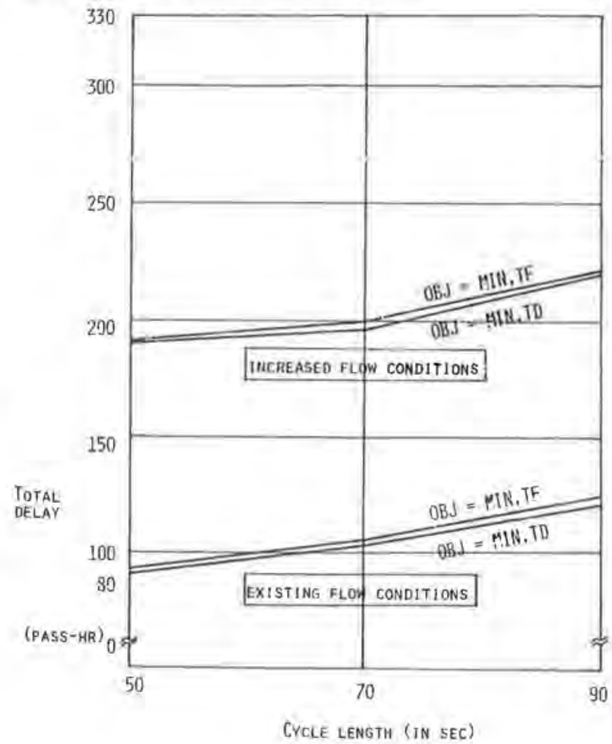
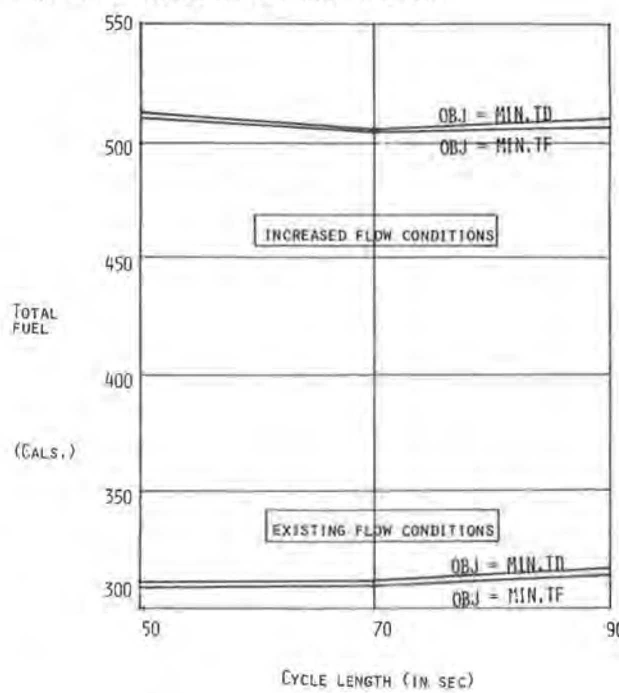
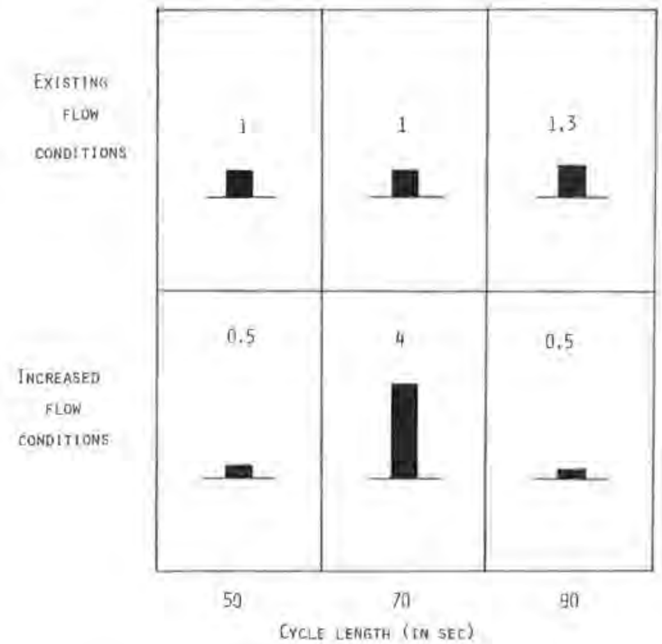


Figure 10. Effect of objective functions on total fuel.



sponsive signal control systems includes dynamic traffic signal control systems that provide adjustments of signal timings for shortand long-term changes in traffic demand, arterial capacities, and operational conditions. The integration of the greatly improved capability of traffic signal controllers and detectors into the dynamic control system for more effective use of arterial systems is

Figure 11. Total delay-total fuel trade-off (passenger hours saved per gallon given up).



another area to be developed.

ACKNOWLEDGMENT

This paper describes and summarizes the results of the study, which used TRANSYT6C, conducted at ITS, University of California, Berkeley. We are indebted to the following former researchers of ITS who have developed TRANSYT6C, applied the model, and made very valuable suggestions and comments on this research: Paul Jovanis of Northwestern University, Thomas Clausen, Abraham Kruger, and Alan Deikman. We also would like to thank Walter Okistu, Victor Siu, and Wai-Ki Yip for their contributions to the computer production runs and the design of experiments with the model. Special appreciation is given to Ken Courage of the University of Florida for pointing out error in the fuel-consumption table (regarding fuel consumed while vehicles are stopped), which have been corrected. Finally, our appreciation is extended to Sylvia Adler for typing this paper.

REFERENCES

1. Transportation System Management, Supplemental Information on the Development and Implementation of TSM Plans. Urban Mass Transportation Administration; Federal Highway Administration, Dec. 1975.
2. Federal Highway Administration. Traffic Signals; Signal Timing Optimization Project; Solicitation of Interest. Federal Register, Vol. 45 No. 58, March 1980.
3. Interim Materials on Highway Capacity. TRB, Circular 212, Jan. 1980, 276 pp.
4. Highway Capacity Manual--1965. HRB, Special Rept. 87, 1965.
5. F.V. Webster and B.M. Cobbe. Traffic Signals. Road Research Laboratory, Crowthorne, Berkshire, England, Tech. Paper 56, 1966.
6. Signalized Intersection Capacity Guide. Australian Road Research Board, Rept. 79, April 1978.
7. Statens Vägerk. Beräknig av Kapacitet, Kölängd, Fördröjning i Vägtrafikanlaggningar. Borlange, Sweden, TV131, Feb. 1977.
8. Signal Operations Analysis Package. Federal Highway Administration, Implementation Package 79-9, July 1979.
9. Traffic Network Analysis with NETSIM--A User's Guide. Federal Highway Administration, FHWA-IP-80-3, Jan. 1980.
10. D.I. Robertson. TRANSYT: Traffic Network Study Tool. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, LR 253, 1969.
11. D.I. Robertson and P. Gower. User Guide to TRANSYT Version 6. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Supplementary Rept. 255, 1977.
12. P.P. Jovanis, A.D. May, and A. Deikman. Further Analysis and Evaluation of Selected Impacts of Traffic Management Strategies on Surface Streets. Institute of Transportation Studies, Univ. of California, Berkeley, 1977.
13. R. Akcelik. TRANSYT Version 6N, National Capital Development Commission Engineering Division, Cleveland, OH, June 1978.
14. P.B. Hunt and J.V. Kennedy. A Guide to TRANSYT/7. Department of the Environment, Department of Transport, Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Rept. LR942, 1978.
15. C.E. Wallace, K. Courage, D.R. Reaves, G.W. Schoene, and G.W. Euler; Univ. of Florida. TRANSYT7F User's Manual. Federal Highway Administration, 1981.
16. R.A. Vincent, A.I. Mitchell, and D.I. Robertson. User Guide to TRANSYT, Version 8. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Rept. 888, 1980.
17. T.J. Clausen and A.D. May. The Analysis and Evaluation of Selected Impacts of Traffic Management Strategies on Surface Streets. U.S. Department of Transportation, Oct. 1976.
18. P. Jovanis and A.D. May. Planning Consequences of Traffic Systems Management, Institute of Transportation Studies, Univ. of California, Berkeley, April 1977.
19. A.D. May, T.J. Clausen, and A.J. Kruger. Development and Application of Traffic Management Models. TRB, Transportation Research Record 630, 1977, pp. 1-6.
20. P.J. Claffey; P.J. Claffey and Associates. Running Costs of Motor Vehicles as Affected by Road Design and Traffic. NCHRP, Rept. 111, 1971, 97 pp.
21. J.A. Apostolos, W.R. Shoemaker, and E.C. Shirley. Energy and Transportation Systems. California Department of Transportation, Sacramento, 1978.
22. P. Kunselman and others. Automobile Exhaust Emission Model Analysis Model. U.S. Environmental Protection Agency, Rept. 460/3-74-005, 1974.
23. R.C. Dowling and A.D. May. Data Collection and Analysis for the Eastshore Freeway (I-80) Corridor. In Freeway Operations Study--Phase IV, Institute of Transportation Studies, Univ. of California, Berkeley, Special Rept. 74-1, June 1974.

Publication of this paper sponsored by Task Force on Traffic Signal Systems.

Possible PASSER II Enhancements

RAMEY O. ROGNESS

The PASSER II computer program for optimization of arterial signal timing has been accepted by usage. It has been used extensively during the last few years. The program's ability to select multiphase sequences for a maximum bandwidth progression solution has led to its increasing use and application. The PASSER II maximum bandwidth solution has been well accepted and implemented throughout this country. The theory, model structure, methodology, and logic in the PASSER II computer program has been evaluated and documented. An evaluation was undertaken to determine if several enhancements to the PASSER II program as related to a revised green split procedure, a minimum delay cycle length, and number of alternate optimal solutions could improve the utility of the solution and would be useful measures. The comparison was to the existing PASSER II computer program and comparison TRANSYT program runs. This evaluation showed, for the three scenarios considered, that a revision for the green split routine provided equal saturation splits. An advisory minimum delay cycle length calculation would provide useful guidance in the selection of the cycle length range to consider. Other measures, like a minimum delay performance measure, alternate optimal solutions, and improved delay measure, could provide useful results.

The PASSER II computer program for optimization of signal timing on arterials has been accepted by usage and is being used extensively. The program's ability to select multiphase sequences for a maximum bandwidth progression solution has led to its increasing use in the last few years. The PASSER II maximum bandwidth solution has been well accepted and implemented throughout this country. The theory, model structure, methodology, and logic in the PASSER II computer program have been evaluated and documented.

The PASSER II computer model was developed by Messer and others (1) and modified to an off-line computer program by Messer (2). It was developed primarily for high-type arterial streets (i.e., those that have intersections with protected left-turn lanes and phases) (3). It is applicable for the timing for modern eight-phase controllers.

The PASSER II computer program can be classified as a macroscopic deterministic optimization model. It uses a platoon level representation for fixed (uniform) traffic volumes and speeds. The optimization procedure is an implicit enumeration of the minimum interference values and uses a variant of the half-integer synchronization approach for relative offsets. The unique advantage the PASSER II program has over other optimization programs for signalization is that it can be used to consider and select multiple phase sequences (4).

The optimal bandwidth solution is selected as the lowest minimum interference sum. Two measures are used to determine the worthiness of the solution--efficiency and attainability.

CYCLE LENGTH

An investigation was conducted to determine whether the cycle length selected by the PASSER II program corresponded to the minimum cycle length (over the range studied) for the traffic network study tool (TRANSYT) program (5). This investigation was part of a larger study to develop a heuristic programming approach to arterial signal timing (6).

A four-signal arterial street was selected for the evaluation. It was considered large enough to permit signal and link characteristics not to especially affect the traffic behavior and results of the study.

It was decided to evaluate three cycle lengths to evaluate the minimum delay and progression solution interaction. The three cycle lengths selected were 80, 90, and 100 s, which appeared to be representative values and still provided a nominal range and three solution points.

To permit some range of spacing, three intersection spacings were considered to study the effect of the interrelation between cycle length and intersection spacings--full scale, half scale, and quarter scale. The morning peak-period volume condition was used for the evaluation.

The arterial street selected, Skillman Avenue, was not considered ideal for either progression or minimum delay objectives. Figure 1 shows the four intersections used. In general, all intersections are high-type and all signalization is multiple phase with protected turning. Figure 2 shows the full-scale intersection spacing.

Table 1 lists the three intersection spacings considered.

COMPARISON OF CYCLE LENGTH SOLUTIONS

For three spacing scenarios (full scale, half scale, and quarter scale), runs were made for the three cycle lengths (80, 90, and 100 s). The efficiency of the PASSER II optimal solutions for the three intersection spacings and cycle lengths are provided in the table below

	Efficiency by Cycle Length		
Scenario	80s	90s	100s
Full scale	0.341	0.299	0.398
Half scale	0.382	0.393	0.346
Quarter scale	0.349	0.328	0.310

For the full-scale scenario, the optimal cycle length is 100 s. For the half-scale scenario, the optimal cycle length is 90 s. An 80-s cycle length is the optimal solution for the quarter-scale scenario. The efficiency of the solution obtained varies as the cycle length changes for each of the three scenarios. The shape of the efficiency curve for the cycle lengths for each of the scenarios is shown in Figure 3. For the full-scale scenario, the efficiency curve is nonmonotonic and illustrates the effect of cycle length on the progression efficiency of the arterial street.

Number of Alternate Optimal Solutions

For the phasing combination for the optimal solution, the PASSER II program outputs the last phasing sequence that was considered. Although alternate optimal phasing arrangements may exist to the phasing sequence selected, the program has no means to identify these. Different phasing arrangements may satisfy the progression criteria. From a progression standpoint, there is no advantage of these alternative solutions over the one selected.

Although which of the alternate optimal phasing arrangement that PASSER II selected does not affect the progression solutions, the phasing arrangement may have an important effect on the minimum delay solution. Also one of the alternate optimal phasing arrangements may allow the traffic engineer to pick the type of phasing he or she would like to use for other than progression considerations (i.e., safety and consistency). A physical change in the phasing arrangement can be a major undertaking in the field; therefore, these alternate optimal phasing arrangements can be important.

For each of the cycle lengths, the alternate optimal phasing arrangements were determined by explicit enumeration. The number of optimal phasing sequences for each cycle length and spacing scenario is listed in Table 2.

The phasing alternatives were numerically described for each intersection as

1. Left-turns first,
2. Through movements first,
3. Leading green, and
4. Lagging green.

As an example of the phasing description of the four-signal arterial overall, a phasing combination of 1342 would be intersection one left-turns first (1), intersection two leading green (3), intersection three lagging green (4), and intersection four through movements first (2). Each phasing arrangement will be indicated by using a dash as the delimiter mark. The example would be shown as 1-3-4-2.

For the full-scale scenario, there was only 1 optimal phasing sequence for the 80-s cycle length. There were 17 alternative optimal sequences for 90 s, and 21 alternative optimal phasing sequences for 100-s cycle length.

For the half-scale scenario, there were 9 alternative optimal phasing arrangements for the 80-s cycle length. There were 4 alternative optimal sequences for 90 s, and 12 alternative optimal phasing arrangements for the 100-s cycle length.

The quarter-scale scenario had only one optimal phasing sequence for the 80-s cycle length. There were three optimal phasing arrangements for 90 s. Four phasing arrangements were optimal for the 100-s cycle length.

Overall, for each of the scenarios, there were few alternative optimal phasing arrangements for the 80-s cycle length. There were several alternative optimal phasing arrangements for the 90-s cycle length. For the 100-s cycle length there were more

Figure 1. Location of Skillman Avenue, Dallas.

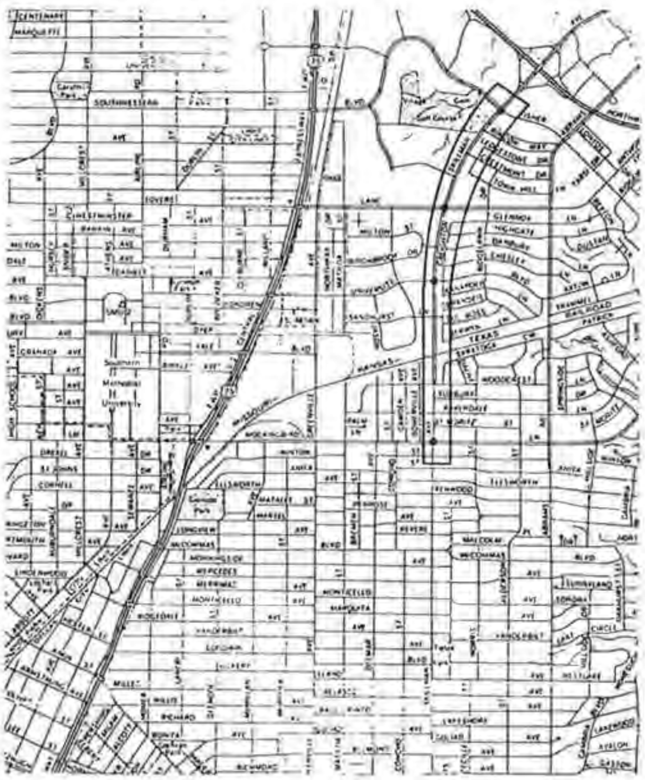


Figure 2. Skillman Avenue line drawing with spacing.

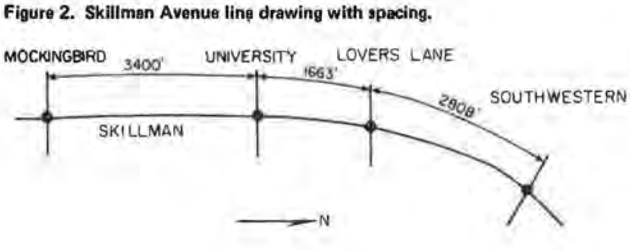


Table 1. Intersection spacing scenarios.

Link	Link Spacing (ft) Full Scale	Half Scale	Quarter Scale
South of Mockingbird	3000	3000	3000
Mockingbird to University	3400	1700	850
University to Lovers Lane	1663	832	416
Lovers Lane to Southwestern	2808	1404	702
North of Southwestern	3000	3000	3000
Cross street approaches	2000	2000	2000
Entry links	1000	1000	1000

Figure 3. PASSER II efficiency versus cycle length for scenarios.

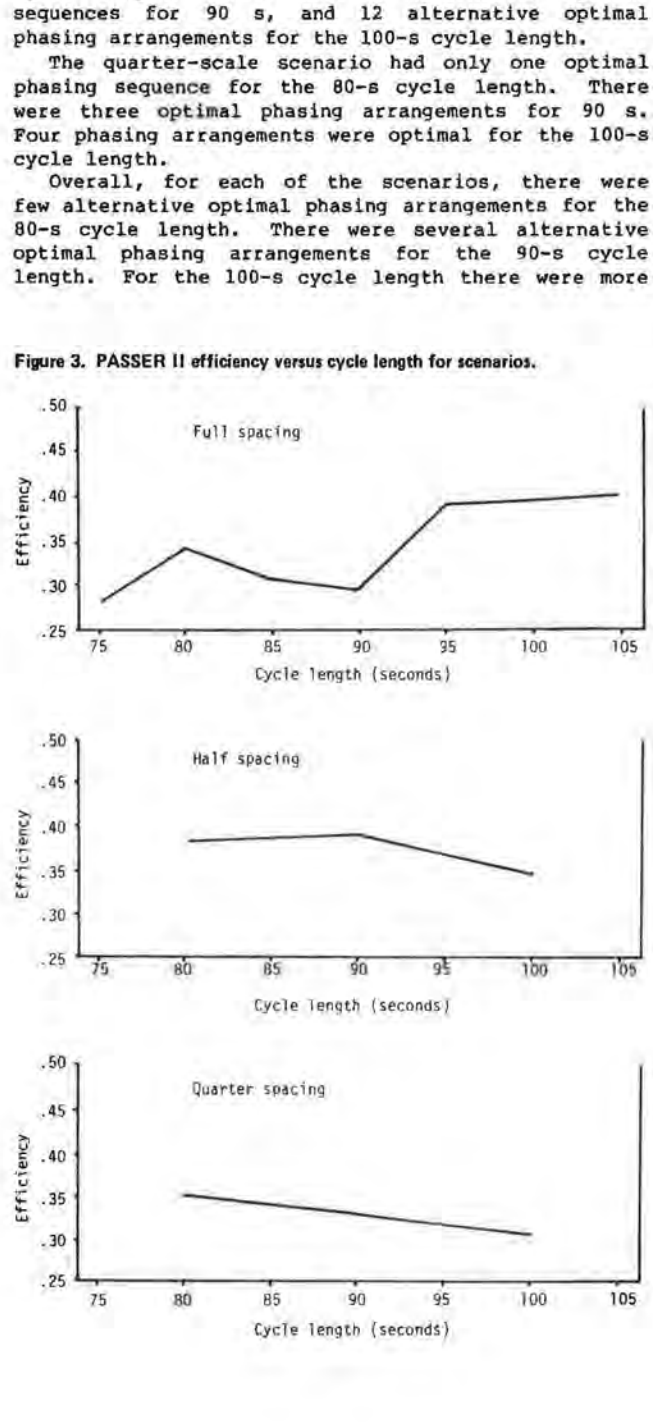


Table 2. Number of PASSER II alternate optimal solutions for three intersection spacing and cycle lengths.

	PASSER II Alternate Optimal Solutions by Cycle Length		
Scenario	80 s	90 s	100 s
Full scale	3-4-4-3[a]	3-1-1-4	2-1-1-4
		3-1-2-4	2-1-2-4
		3-1-3-4	2-1-3-4
		3-1-4-4	2-2-1-4
		3-2-1-4	2-2-2-4
		3-2-2-4	2-2-3-4
		3-2-3-4	2-4-1-4
		3-2-4-4	2-4-2-4
		3-3-1-4	2-4-3-4
		3-3-2-4	3-1-3-1
		3-3-3-4	3-1-3-2
		3-3-4-4	3-1-3-4
		3-4-1-4	3-2-3-1
		3-4-2-4[a]	3-2-3-2
		3-4-3-4	3-2-3-4
		3-4-4-4	3-3-3-1
			3-3-3-2
			3-3-3-4
			3-4-3-1
			3-4-3-2
			3-4-3-4[a]
Half scale	2-1-4-1	4-1-4-3	4-1-1-3
	2-1-4-2	4-2-4-3	4-1-2-3
	2-2-4-1	4-3-4-3	4-1-4-3
	2-2-4-2	4-4-4-3[a]	4-2-1-3
	2-3-4-1		4-2-2-3
	2-3-4-2[a]		4-2-4-3
	3-3-1-3		4-3-1-3
	3-3-2-3		4-3-2-3
	3-3-4-3		4-3-4-3[a]
			4-4-1-3
			4-4-2-3
			4-4-4-3
Quarter scale	4-3-3-4[a]	4-3-3-1	4-3-3-1
		4-3-3-2	4-3-3-2
		4-3-3-4[a]	4-3-3-3
			4-3-3-4[a]

[a]PASSER II selected optimal solution.

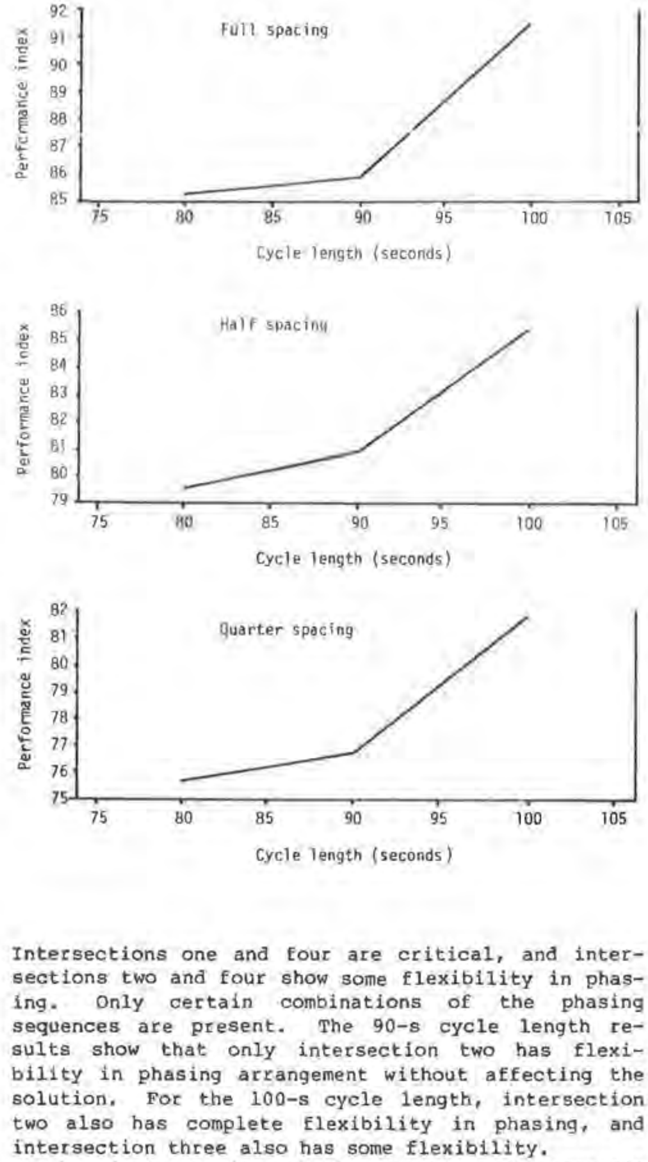
alternative optimal phasing arrangements.

The number of alternate optimal phasing sequences would appear to increase with larger cycle lengths. The effect of intersection spacing would be an additional factor. This arises from the greater flexibility permitted by the larger green times and space periodicity. These increases allow different phasing arrangements at the noncritical intersections without affecting the efficiency of the optimal solution.

For the full-scale scenario, there is no flexibility at the 80-s cycle length and no alternate optimal solutions. The 90-s cycle length has some flexibility. At this cycle length, the first and last intersections are critical and the phasing arrangement is for intersection one leading green and intersection four lagging green. For the two inside intersections (two and three) that are noncritical, any phasing combination is possible because of the flexibility. Flexibility is even greater with the 100-s cycle length. There appears to be two sets of critical intersections for this cycle length. Either intersections one and four with phasing of through movements first (2) and lagging green (4), respectively, or intersections one and three with phasing of leading green (3) are critical. There is not complete flexibility of the phasing of the noncritical intersections, however, since not all four arrangements were present.

The intermediate spacing of the half-scale scenario shows slightly different results. There are alternate optimal phasing arrangements for the 80-s cycle length, as can be observed from Table 2.

Figure 4. Plot of performance index versus cycle length for scenarios.



Intersections one and four are critical, and intersections two and four show some flexibility in phasing. Only certain combinations of the phasing sequences are present. The 90-s cycle length results show that only intersection two has flexibility in phasing arrangement without affecting the solution. For the 100-s cycle length, intersection two also has complete flexibility in phasing, and intersection three also has some flexibility.

The short spacing of the quarter-scale scenario again shows limited ability of alternate optimal phasing. The only flexibility in phasing is for intersection four for the larger cycle length.

The combination of longer green time, spacing periodicity, and flexibility in phasing arrangement alternatives for a given optimal solution is apparent. For each cycle length and intersection spacing combination, the intersections that are noncritical cause the alternative optimal phasing arrangements for the progression optimal solution.

TRANSYT Runs

Corresponding runs were made for the scenarios by using the TRANSYT6B program. For the three spacing scenarios, the lowest TRANSYT performance index was for the 80-s cycle lengths. The relation between TRANSYT performance index and cycle length is illustrated in Figure 4. These data would indicate that the average delay would increase monotonically as the cycle length is increased from 80 s (for the cycle length range considered).

Comparison of PASSER II and TRANSYT

The comparison of cycle lengths for TRANSYT and PASSER II results would indicate, for this study, that the PASSER II optimal cycle length is not usually the minimum delay solution cycle length from TRANSYT. This comparison shows that, potentially, a minimum delay cycle length would need to be incorporated as part of the PASSER II program to alleviate this problem.

The determination of a minimum delay cycle length used the previous runs to provide a calculation to find the minimum delay cycle length. The evaluation was whether the cycle length calculation permitted the selection of the minimum delay cycle length for each scenario.

Green Split Routine

The procedure used in the PASSER II program for traffic signal timing for splits is to distribute the available effective green time in proportion to the critical movement volume to saturation capacity flow rate ratios. The methodology is derived from Webster's concept and uses the critical lane analysis approach (7).

The routine has a limitation in common with many others--the initial green split between the arterial street and the minor street at each intersection is made on the ratio to the critical lane volume sum. The street time is then allocated (split) between the opposing movements in the ratio of the relative critical lane volumes. A check is then made whether each movement time exceeds its minimum green time. When the movement green time is less than the minimum green time, the remainder of the time needed is taken off the paired opposing movement time and the times are adjusted.

In the case where the times are adjusted, the minimum green times are satisfied, but the saturation ratio and opposing movement green time has been changed. For the critical intersection or a critical movement, the adjustment can result in poor operation from increased saturation on the opposing link. The method results in the split between the arterial street and the cross street remaining the same.

In certain situations, with the model arterial being one, this taking from the opposing movement can cause poor performance measures. The STAR1 routine in TRANSYT follows a slightly different approach by using equal saturation. From the TRANSYT runs comparison, the existing PASSER II green split routine outperforms the STAR1 routine for one scenario. A revised PASSER II green split routine was developed by using a modified equal saturation basis.

The revised routine developed retains the existing green split routine through determining which movements do not have their minimum green times satisfied. This deficit time for each movement green is used to calculate an equivalent vehicular volume. This volume is added to the original volume. The green times are recalculated by using the critical lane analysis. These revised green times are checked against the respective green times for each movement. Although the possibility exists for one to four deficit times per intersection, in actuality only two of the movement deficits determine the split allocation.

If after the first recalculation the minimum green time still remains unsatisfied, the resultant deficits are redetermined, the equivalent volumes computed, and the critical lane analysis and green splits are recalculated. After this second recalculation, the resulting green times were considered to be close to the minimum green times (for the critical movements). The incremental improvement for additional recalculations would be small.

At this point, the proposed procedure would revert to the existing procedure. The test for satisfying the minimum green is done. The original deficit movements are adjusted to their minimum green times. This is done for a threefold purpose. First, there may still exist original deficit (critical) movements that have not been increased to the minimum green times (that must be satisfied). The second reason is that the noncritical deficit movements could have been overcompensated and exceed their minimum green times (and should be adjusted back to their minimum times). The third reason is that the revised green splits may have caused a satisfied initial movement time (i.e., no initial deficit) to end up slightly deficit. At this step, the original deficit movements and any currently deficit movements are set to their minimum green times and a corresponding adjustment is made to their paired opposing movement time.

The effect of this revised green split routine is to cause (if two) the two critical deficit movements for an intersection to have equal saturation ratios. The other movements saturation ratio (and green time) are affected from this adjustment to the original split. However, the effect may be added or reduced green time, depending on their relation to the deficit movements.

A comparison between the original green split times and the revised green split times for each intersection showed differences (Table 3). This arose because of the heavy through movement and light left-turn volumes (opposing). The result was that the arterial had several deficit movements. This result is hidden in the present program output unless a visual comparison is made to the minimum green time and the paired opposing movement. The revised procedure green split times for the 90-s and the 100-s cycle lengths are compared with the original green splits in Tables 4 and 5.

The mixed performance for the PASSER II splits versus the STAR1 splits is not surprising. In the original routine the deficits are compensated from the opposing movements, which usually are at a critical level of saturation. As the cycle length is increased (from 80 s), the number of deficit movements and movement time is reduced. This could explain why the original PASSER II green split routine gave mixed results for the different cycle lengths. The revised procedure does not modify the original green splits unless there is a deficit movement. It would appear that the poorer performance of the PASSER II green split results (at the lower cycle lengths) would improve with the revised procedure. At the higher cycle lengths, where the PASSER II results were sometimes better than those of STAR1, it would appear that the revised procedure would only slightly alter the original green splits.

To evaluate the revised green split routine, the calculations were manually performed and input into the existing PASSER II program as minimum green times to force the desired splits. The phasing combinations that are optimal for the three scenarios were determined. The original alternate optimal PASSER II phasing combinations and the best TRANSYT phasing sequence for each scenario were rerun.

In most cases the revised results provided a slightly lower efficiency and bandwidth, since more green time was provided to an intersectional cross movement. The recalculation of the green splits yielded more green time for the cross street and less for the main street. The results are provided in Table 6 and Figure 5.

The effect of the revised green split routine on a progression solution depends on which movements are deficit. For example, if both a cross street movement and a through movement are (equally) deficit, the green split between the cross street and the main street would not change. The allocation of green time for the cross street movements and the main street movements individually will be changed.

Minimum Delay Cycle Length

Although progression has been widely accepted for arterial signals, Webster (8) originally recognized that the consideration of delay minimization was necessary in the selection of the cycle length to be used for each intersection. Webster's cycle calculation, however, is only applicable to two-phase signal operation.

A direct estimate of the minimum delay cycle length that is appropriate for the highly saturated conditions would be desirable. A possible approach is to use the flow ratios directly with the lost time to estimate the minimum delay cycle length. By using a modified critical movement analysis approach, the y_is for each intersection would be determined. The opposing movement y_is would be

Table 3. Comparison of PASSER II green splits from original and revised procedure for 80-s cycle.

Street	Movement	Original (SPLIT1)	Revised (SPLIT2)	Street	Movement	Original (SPLIT1)	Revised (SPLIT2)
Mockingbird	1[a]	10.0	10.0	University	1	10.0	10.0
	2	32.2	30.9		2	51.4	51.8
	3	13.1	12.7		3	10.0	10.0
	4	29.1	28.2		4	51.4	51.8
	5	20.9	21.6		6	18.6	18.2
	6	16.9	17.5		8	18.6	18.2
	7[a]	10.0	10.0				
	8	27.8	29.1				
Lovers Lane	1[a]	10.0	10.0	Southwestern	1[a]	10.0	10.0
	2	39.0	39.0		2	36.3	37.0
	3[a]	10.0	10.0		3[a]	10.0	10.0
	4	39.0	39.0		4	36.3	37.0
	5	10.0	10.0		5	12.7	12.0
	6[a]	21.0	21.0		6[a]	21.0	21.0
	7	10.0	10.0		7[a]	10.0	10.0
	8	21.0	21.0		8	23.7	23.0

[a]Deficit minimum green movements.

Table 4. Comparison of PASSER II green splits from original and revised procedure for 90-s cycle.

Street	Movement	Original (SPLIT1)	Revised (SPLIT2)	Street	Movement	Original (SPLIT1)	Revised (SPLIT2)
Mockingbird	1[a]	10.0	10.0	University	1	10.0	10.0
	2	37.6	36.0		2	59.2	59.5
	3	14.5	14.0		3	10.0	10.0
	4	33.1	32.0		4	59.2	59.5
	5	23.5	24.4		6	20.8	20.5
	6	18.9	19.6		8	20.8	20.5
	7[a]	10.0	10.0				
	8	32.4	34.0				
Lovers Lane	1[a]	10.0	10.0	Southwestern	1[a]	10.0	10.0
	2	46.7	47.2		2	42.3	43.7
	3	10.5	10.9		3[a]	10.0	10.0
	4	46.2	46.3		4	42.3	43.7
	5	12.3	11.8		5	14.8	14.3
	6[a]	21.0	21.0		6	22.9	22.0
	7	10.5	10.5		7[a]	10.0	10.0
	8	22.8	22.3		8	27.7	26.3

[a]Deficit minimum green movements.

Table 5. Comparison of PASSER II green splits from original and revised procedure for 100-s cycle.

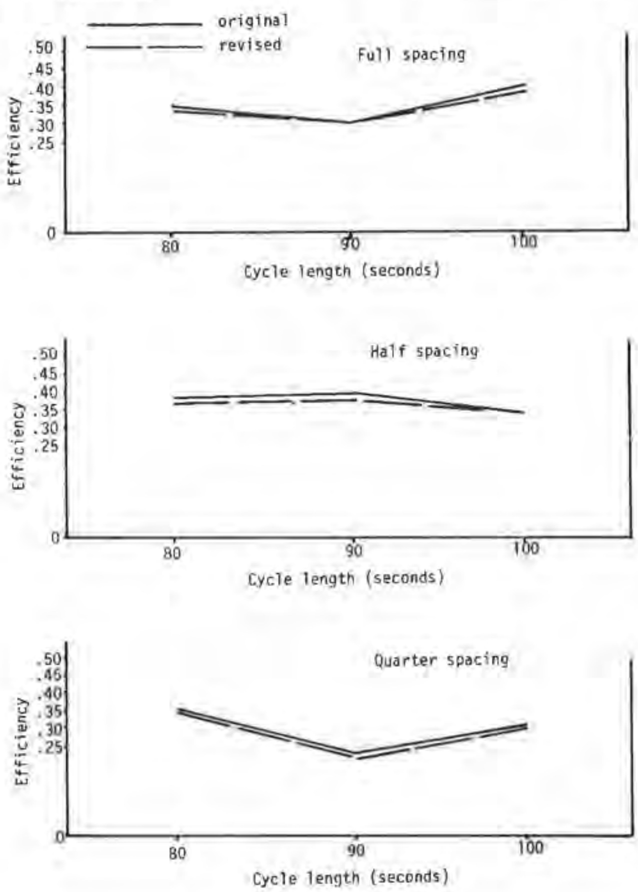
Street	Movement	Original (SPLIT1)	Revised (SPLIT2)	Street	Movement	Original (SPLIT1)	Revised (SPLIT2)
Mockingbird	1[a]	10.1	10.0	University	1	10.0	10.0
	2	42.8	41.1		2	67.0	67.2
	3	15.9	15.4		3	10.0	10.0
	4	37.0	35.7		4	67.0	67.2
	5	26.1	27.1		6	23.0	22.8
	6	21.0	21.7		8	23.0	22.8
	7[a]	10.0	10.0				
	8	37.1	39.0				
Lovers Lane	1[a]	10.0	10.0	Southwestern	1[a]	10.0	10.0
	2	53.3	53.5		2	48.3	49.5
	3	11.4	11.5		3[a]	10.0	10.0
	4	51.9	51.8		4	48.3	50.0
	5	15.7	15.5		5	16.3	15.7
	6[a]	21.0	21.0		6	25.4	24.5
	7	11.3	11.4		7[a]	10.0	10.0
	8	25.4	25.1		8	31.7	30.2

[a]Deficit minimum green movements.

Table 6. Comparison of PASSER II selected solutions for original and revised split procedures.

Scenario	Cycle Length (s)	Original		Revised	
		Phasing Sequence	Efficiency	Phasing Sequence	Efficiency
Full scale	80	3443	0.341	3443	0.335
	90	3424	0.299	3424	0.298
	100	3434	0.398	1434	0.384
Half scale	80	2342	0.382	1341	0.369
	90	4443	0.393	4343	0.377
	100	4343	0.346	4343	0.345
Quarter scale	80	4334	0.349	4334	0.343
	90	4334	0.328	4334	0.321
	100	4334	0.310	4333	0.302

Figure 5. Plot of efficiency versus cycle length for scenarios from original and revised procedures.



summed. A determination would be made of which pair was the greater for each street. These maximum paired y_is would be summed for the intersection and the following equation used to estimate the appropriate cycle length, i.e.,

$$\Sigma(Y_i/X_i)\,C + nl \leq C \tag{1}$$

i.e., critical, or

$$C \leq nl/[1 - \Sigma(Y_i/X_i)] \tag{2}$$

where

C = cycle length,
nl = lost time,
Y_i = flow ratio (g/s), and
X_i = desired x-ratio.

This procedure does not evaluate for deficit green times, therefore, an adjustment for low y_is was proposed to compensate for this shortcoming. The adjustment for the critical movement y_is is if one of the opposing pair y_is is less than 0.10 and the other movement is greater than 0.25, then the low y_i (less than 0.10) is multiplied by 1.1. This provides for a compensation for those movement y_is that are critical and likely would be deficit. The steps in the procedure are as follows:

Step 1: Calculate y_is,
Step 2: Adjust the y_is if necessary, and
Step 3: Calculate C from the equation.

By using the model arterial intersection data and a x-ratio of 0.85, the following minimum delay cycle lengths were determined: Mockingbird, 91 s; University, 39 s; Lovers Lane, 73 s; and Southwestern, 87 s.

If a cycle length was selected within the range of 0.75 and 1.25 of the minimum delay cycle length, the delay might only be increased slightly; i.e.,

$$0.75\,C_{min} \leq C \leq 1.25\,C_{min} \tag{3}$$

By using this recommended range the proposed procedure calculates the minimum cycle lengths for each intersection shown in the table below. In actuality the low cycle length determined may be less than the sum of minimum greens. Since the minimum greens are the absolute lowest cycle length possible, a check must be made for the range of cycles determined.

Intersection	Cycle Length (s)	
	Low	High
Mockingbird	68	113
University	29	49
Lovers Lane	54	90
Southwestern	65	108

For the four intersections, the sum of minimum greens is 57, 41, 62, and 60 s, respectively. The low cycles for University and Lovers Lane both violate this limit and would need to be changed to the limit. Because of the three-phase signal at University Drive, its cycle length range is much lower than that of the other (four-phase) signals. Its cycle length range also falls below the remaining signals sum of minimum green. Because of this, the cycle length range that would be selected must fall outside of University Drive's minimum delay cycle length. From the cycle length range of the remaining three signals, it would appear that a range of 68-90 s could be selected.

SUMMARY

The proposed enhancements appear to have the capability to alleviate certain of the differences between the PASSER II solutions and the TRANSYT solutions. A minimum delay cycle length range procedure is needed to consider the effect of delay in a progression solution. The need for a revised green split routine in the PASSER II program was apparent in the differences between the STARl and TRANSYT solutions and the PASSER II split solutions.

The procedures developed provide only estimates for the minimum delay cycle length and equal saturation green splits. More rigorous and complete procedures could be developed to provide a better estimate.

ACKNOWLEDGMENT

This paper presents a portion of my Ph.D. dissertation under the direction of Carroll J. Messer at Texas A&M University. The Civil Engineering Department provided the computer resources needed to accomplish this research. The contents of this paper reflect my views, and I am responsible for the facts and accuracy of the data presented herein.

REFERENCES

1. J.D. Carvell and W.R. McCasland. Traffic Management in a Dallas, Texas Freeway Corridor. 47th Annual Meeting, Compendium of Technical Papers, ITE, Mexico City, Oct. 1977, pp. 93-116.
2. C.J. Messer, R.N. Whitson, C.L. Dudek, and E.J. Romano. A Variable-Sequence Multiphase Progression Optimization Program. HRB, Highway Research Record 445, 1973, pp. 24-33.
3. C.J. Messer, H.E. Haenel, and E.A. Koeppe. A Report on the User's Manual for Progressional Analysis and Signal System Evaluation Routine--PASSER II. Texas Transportation Institute, College Station, Res. Rept. 165-14, Aug. 1974.
4. R.H. Whitson and J.D. Carvell. A Variable Phase Sequence Progression Optimization System. In Traffic Control and Transportation Systems, American Elsevier, New York, 1974.
5. D.I. Robertson and P. Gower. User Guide to TRANSYT, Version 6. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Supplementary Rept. 255, 1977.
6. R.O. Rogness. A Heuristic Programming Approach to Arterial Signal Timing. Texas A&M Univ., College Station, Ph.D. dissertation, Aug. 1981.
7. C.J. Messer and D.B. Fambro. Critical Lane Analysis for Intersection Design. TRB, Transportation Research Record 644, 1977, pp. 26-35.
8. F.V. Webster; Road Research Laboratory. Traffic Signal Settings. Her Majesty's Stationary Office, London, Tech. Paper No. 39, Oct. 1957.

Publication of this paper sponsored by Task Force on Traffic Signal Systems.

Evaluation of Signal Timing Variables by Using A Signal Timing Optimization Program

ANDREW C.M. MAO, CARROLL J. MESSER, AND RAMEY O. ROGNESS

This paper presents the results of a limited study to evaluate the effects of signal timing variables on the selection of the signal timing plan and the resulting measures of effectiveness from a signal timing optimization program. The TRANSYT computer program was used for the evaluation. Several series of sensitivity tests were performed to study the interrelations among number of signalized intersections, signal spacing, cycle length, and traffic flow conditions. The evaluation showed varying effects of the signal timing variables on the results. There appeared to be consistency in results for different signal system configurations (number of signals). With fixed signal spacing and number of signals, the measure of effectiveness (performance index) increased with volume level and cycle length. The effect of signal spacing illustrated differences in the behavior of the performance index. These results show the trade-offs between signal spacing and cycle length for a fixed number of signals and traffic volume level. As the cycle length was increased, the performance index also increased (although sometimes only slightly). This may suggest the use of the shortest practical cycle length for a progressive operation.

With ever-increasing loads being placed on urban traffic facilities from growing traffic demands, the retention of urban mobility depends to a very large extent on the effective use of urban street signal systems. The signalized intersections of urban arterials are a critical element of the urban street system. Traffic congestion and other operational deficiencies are common along arterial streets. Excessive or unnecessary delays, stops, and fuel consumption are experienced due to the inefficient operation of the signalization system. The safe and efficient movement of arterial traffic is almost totally a function of the signal timing variables. By virtue of their operation, traffic signals cause delay to motorists (1). The intersection characteristics usually determine the efficiency and capacity of the entire street system (2). The need exists to develop improved traffic control technology for facilitating the optimal use of available capacity (3).

Improvement of the effectiveness of the traffic control parameters would contribute to reducing the congestion and to relieving those conditions that impede the flow of traffic. The selection of a signal timing plan is complicated by the large number of alternatives available and the interrelations among the signal timing parameters (4). A considerable amount of research has been done on coordination of traffic signals on urban arterial streets (5). Efforts have been directed toward computerized signal timing optimization programs that would provide for signal timing plans superior to those in use.

The maximum bandwidth progression solution has been the approach preferred by traffic engineers (6-8). This arises in part from the lack of computational complexity in use and the ability to visualize the goodness of the results. Although progression has been widely accepted and used, concerns have arisen as to whether it provides a good arterial solution at the expense of the cross street traffic. Other methods for setting arterial traffic signals are the minimum delay solution and the combination of minimum delay with fuel consumption. Even with the theoretical development and computational efficiency of progression and minimum delay techniques, the final criteria is that both techniques have been accepted as providing a good solution (9).

Settings for fixed-time coordinated traffic signals are based on safety of traffic, capacity of the intersection, and delay minimization (10). Signal timing plans must take into account not only the needs of the individual intersections but also the requirements that arise from the time relations between adjacent intersections and their signals (11).

SIGNAL TIMING VARIABLES

The signal timing variables that determine a signal timing plan are cycle length, green splits, phase sequence, and offsets. The relative efficiency of a coordination timing plan is dependent on traffic and movement volumes, signal spacing, speeds, intersection capacity, and the number of signals. Although all of these variables determine which timing plan is the best for fixed-time signal timing optimization, these variables are considered to be fixed and deterministic for any specific solution.

To evaluate the effect of these variables on the signal timing plan selected as best from a signal timing optimization program and the resultant measures of effectiveness, a set of cases were developed to study several of the variables and their interrelations. Several series of sensitivity tests were performed to study the interrelations by using the results from the TRANSYT6B program and its measure of effectiveness, performance index (PI) as the basis for comparison. The variables that were considered were signal system configuration, intersection spacing, cycle lengths, and traffic flow conditions.

BASE CONDITIONS

Several assumptions were made to simplify the hypo-

Figure 1. PI versus number of signals for varied spacing and fixed volume and cycle length.

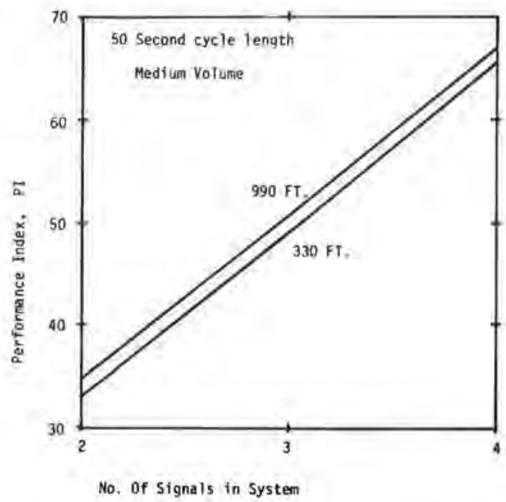


Figure 2. PI versus cycle length for varied volumes and fixed number of signals and spacings.

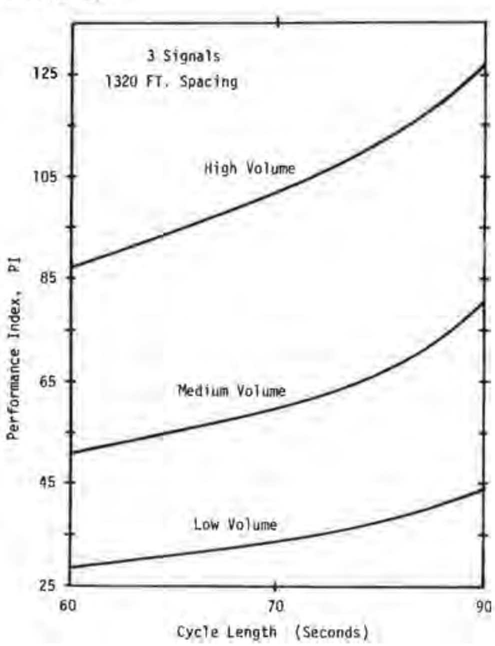


Figure 3. PI versus cycle length with varied volumes and spacings for two-signal system.

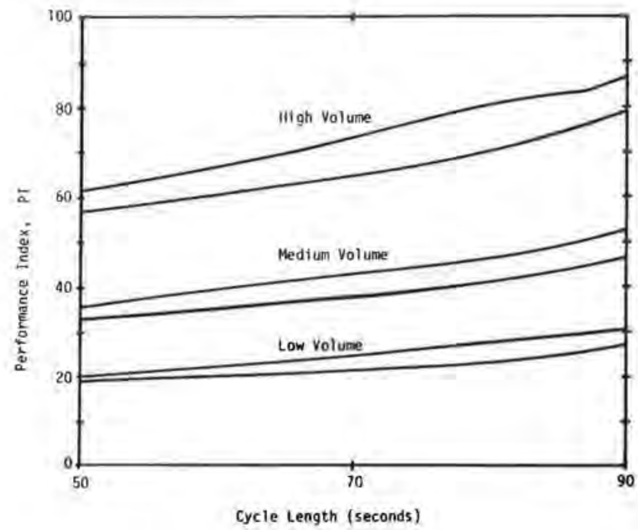
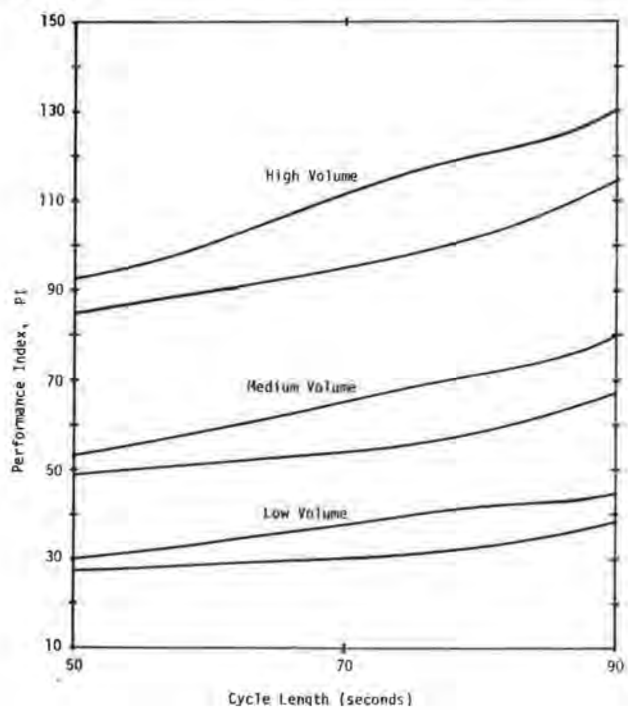


Figure 4. PI versus cycle length with varied volumes and spacings for three-signal system.



thetical street scenario. It was assumed that the basic arterial street and signal control consisted of the following:

1. Uniform arterial grid spacing;
2. Traditional two-phase signal operation;
3. Three lanes at the intersection for each approach consisting of a separate left-turn lane, a through lane, and a combined through plus right-turn lane;
4. Twelve-ft traffic lanes;
5. Saturation flow rates per hour of 1750 pcus for through plus right turns and 1200 pcus for left turns ;
6. Turning movements for an approach of 10 percent for left turns and right turns and 80 percent for through movements; and
7. Average operating speed throughout the system of 34 mph (55 km/h) in both directions.

Signal System Configuration

The signal system configuration concerned the number of signalized intersections. The signal system configurations considered were

1. A signal system comprised of two signals,

Figure 5. PI versus cycle length with varied volumes and spacings for four-signal system.

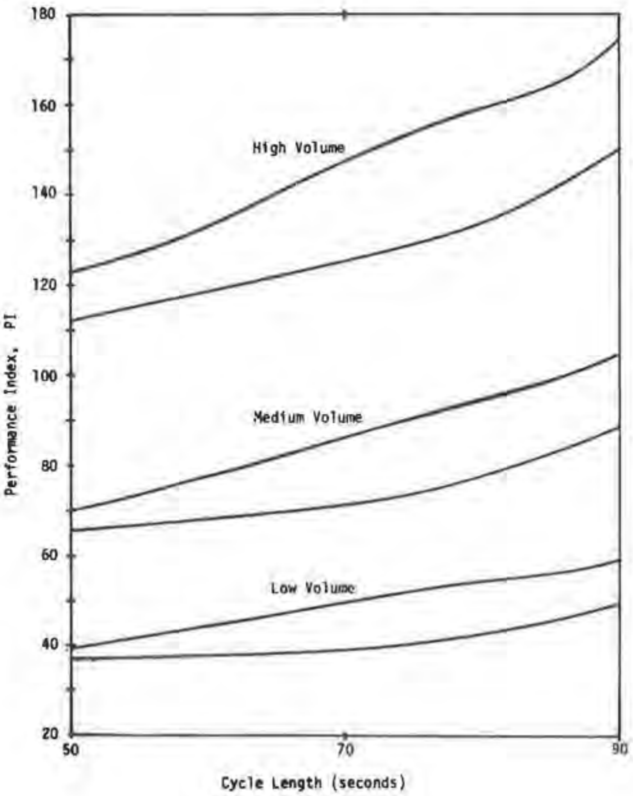


Figure 6. PI versus cycle length with varied spacings and fixed volume for two-signal system.

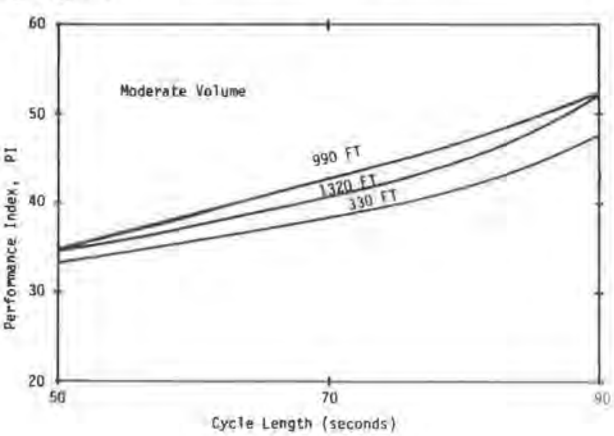


Figure 7. PI versus cycle length with varied spacings and fixed volume for three-signal system.

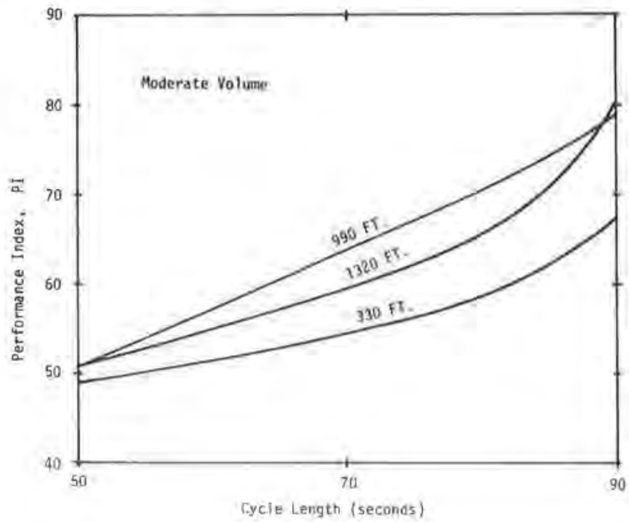


Figure 8. PI versus cycle length with varied spacing and fixed volume for four-signal system.

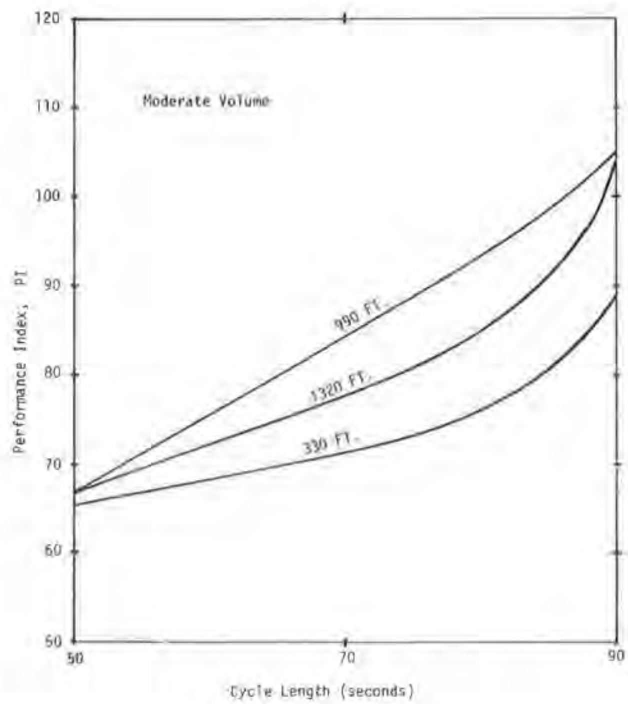


Figure 9. PI versus signal spacing with varied cycle length and fixed volume for two-signal system.

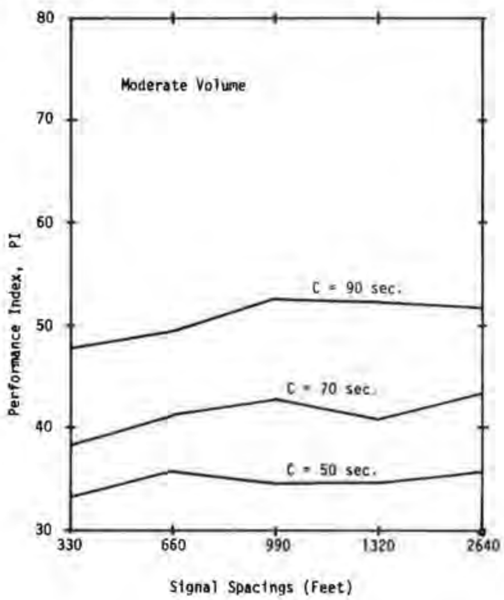
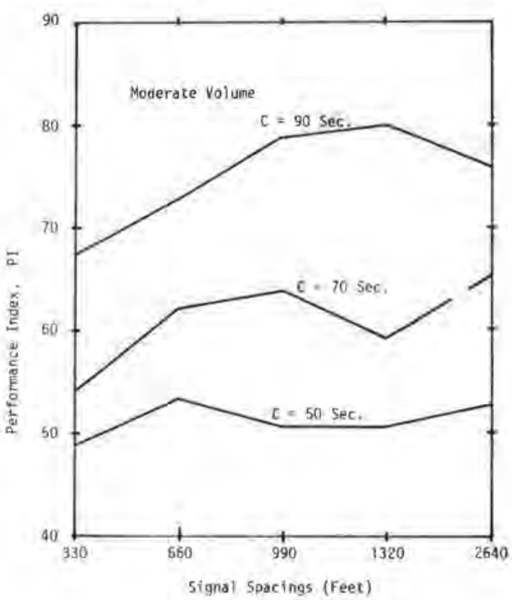


Figure 10. PI versus signal spacing with varied cycle length and fixed volume for three-signal system.



2. A signal system comprised of three signals, and
3. A signal system comprised of four signals.

Traffic Flow Condition

Traffic volume was another variable that was evaluated at three levels. The first level was 80 percent of the saturation flow capacity to represent high-volume traffic conditions on the arterial street. The second level was 60 percent of the saturation flow capacity to represent medium-volume traffic flow conditions. The third level was 40 percent of the saturation flow capacity to represent low-volume traffic conditions.

Figure 11. PI versus signal spacing with varied cycle length and fixed volume for four-signal system.

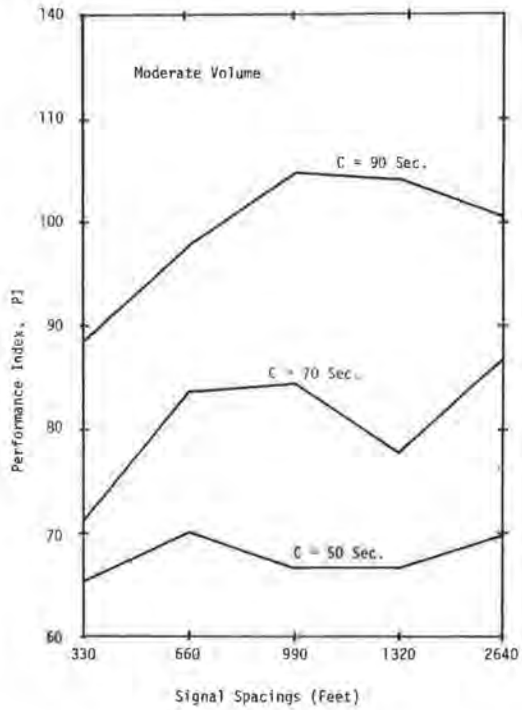
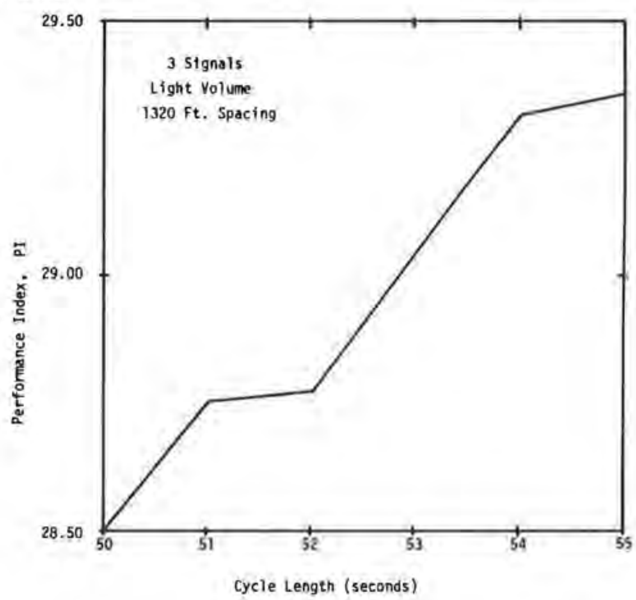


Figure 12. PI versus cycle length for fixed volume, spacing, and number of signals.



Signal Spacing and Cycle Time

Five different spacings between the intersection stop lines were established as the variable levels for signal spacing. This permitted a more detailed evaluation of the spacing effect on choosing cycle lengths. The spacings considered were 330 ft (101 m), 660 ft (201 m), 990 ft (302 m), 1320 ft (402 m), and 2640 ft (805 m).

Three common cycle times of 50, 70, and 90 s were selected for the cycle lengths considered.

EVALUATION AND ANALYSIS

A total of 140 cases were analyzed to study the effects of traffic signal variables on the performance of a signal system. For each combination, the optimal result obtained from the TRANSYT6B program was used for the basis for the evaluation. TRANSYT's performance index (a weighted measure of stops and delay) was used as the comparative measure of effectiveness.

For the two signal spacings used in Figure 1, the shape of the PI curves is nearly straight lines. The slope of these curves is almost identical. This would indicate that consistency exists for the different signal system configurations.

For the conditions of a fixed signal spacing and number of signals, an increase in the traffic volume levels increased the PI. An increase in the cycle length also increased the PI. This is shown on representative Figure 2. Whenever the PI increases, the quality of the traffic conditions becomes worse and the level of service goes down.

The range of PI was evaluated in terms of the five signal spacings considered. The PI for signal system configuration is shown in Figures 3-5. For a given cycle length and volume level, the range of values of the PI is due to the differences in the quality of progression for the signal spacing considered. The range of PI is greater at the higher volume condition. The range of PI is also greater as the number of signals is increased. The interaction of cycle length and a fixed signal spacing is also indicated by the varying slope of the PI curve.

Three of the spacings (330, 990, and 1320 ft) are illustrated in Figures 6-8 to show the inconsistent characteristics of signal spacing versus cycle lengths for the signal system configurations. The effect of the number of signals on PI for the signal spacing can also be seen. The quality of progression is the cause for these inversions. The shape of the PI curves are similar at the short and intermediate cycle lengths for the number of signals. The shape of the PI for the long signal spacing for the larger cycle length has different characteristics.

The relation between PI and signal spacing was further studied at the three cycle lengths for the different signal configuration and traffic volume conditions. The PI varied with the signal spacing. The minimum and maximum values of the PI did not coincide for the three cycle lengths. These differences in PI and signal spacing for the cycle lengths are illustrated in Figures 9-11 for the three signal system configurations. The figures show the trade-offs between signal spacing and cycle length to change the PI. For the same cycle length the differences in the value of the PI are due to the differences in the quality of progression. The shape of the PI curves appears similar for the three signal system configurations. The effect of increasing the number of signals appears to increase the slope and range of the PI curves. The minimum and maximum performance values for the signal spacing, however, appear to change for the three signals and the 90-s cycle length.

To further study the effect of cycle length on the value of PI as the cycle length is varied, a range of the cycle length near the optimal cycle length for a 1320-ft signal spacing and a progression speed of 34 mph was evaluated. Based on the space periodicity concept of progression for the 34-mph speed and 1320-ft spacing, the optimal cycle length falls within the range of 50-55 s. To study the effect of the cycle length on the value of the PI near this optimal progression cycle length, the cycle length was varied from 50 to 55 s in 1-s increments. The effect on PI is displayed in Figure 12. Comparison of the PI value to the cycle length for the conditions modeled as the cycle length is increased shows that PI always increases, although sometimes it may be only slightly.

SUMMARY OF FINDINGS

The findings of this limited hypothetical study are that, in all cases studied, an increase in the traffic volume increased the performance index. An increase in the cycle length also increased the PI in all cases studied. The effects of signal spacings depend on the resulting quality of progression. For a given set of traffic volume, cycle length, and signal spacing, the signal system performance appears to be optimized by operating at the lowest practical cycle length with the best progression possible for that cycle length.

ACKNOWLEDGMENT

This paper presents the results of a special topic research study completed by Mao during his graduate study under Rowan at Texas A&M University. Special acknowledgment is extended to Neilon J. Rowan and Donald L. Woods for their guidance and assistance in accomplishing the study. The contents of this paper reflect our views and we are responsible for the facts and accuracy of the data presented herein.

REFERENCES

1. P.J. Wilkes, A.J.R. Pawley, and A.D.J. Cook. Urban Traffic Control. Highway Engineer, Journal of the Institution of Highway Engineers, Vol. 26, No. 6, June 1979, pp. 26-29.
2. L. Casciato and S. Cass. Pilot Study of the Automatic Control of Traffic Signals by a General Purpose Electronic Computer. HRB, Bull. 338, 1962, pp. 28-39.
3. H.L. Michael. Opportunities in Transportation Engineering Funding and Intersection Management. ITE Journal, Vol. 50, No. 3, March 1980, p. 14.
4. J.H. Kell. Intersection Delay Obtained by Simulating Traffic on a Computer. HRB, Highway Research Record 15, 1963, pp. 73-97.
5. R.L. Vecellio. Traffic Behavior on One-Way Signalized Arterials. ASCE, Transportation Engineering Journal, Vol. 103, No. TE5, Sept. 1977, pp. 575-589.
6. R. Mayer. Cologne's Automatic Traffic Control System. Traffic Engineering and Control, Vol. 4, No. 5, Sept. 1962, pp. 256-259.
7. G. Breenkotter. Hamburg's Traffic Signal Control. Traffic Engineering and Control, Vol. 4, No. 8, Dec. 1962, pp. 438-443.
8. G. Charlesworth. Improving London's Traffic Flow. Traffic Engineering and Control, Vol. 3, No. 10, Feb. 1962, pp. 603-609.
9. J.C. Lai, M.G. Hazzard, P.D. Christopherson, and R.M. Riddle. Safety Aspects of Traffic Signal Control on Arterial System. Federal Highway Administration, FHWA-RD-77, July 1977.
10. J.A. Sosin. Delays at Intersections Controlled by Fixed-Cycle Traffic Signals. Traffic Engineering and Control, Vol. 21, No. 8/9, Aug.-Sept. 1980, pp. 407-413.
11. G. Pavel. Centralized Computer Control of Traffic Signals. Traffic Engineering and Control, Vol. 9, No. 5, Sept. 1967, pp. 232-236.

Publication of this paper sponsored by Task Force on Traffic Signal Systems.

Arterial Progression—New Design Approach

CHARLES E. WALLACE AND KENNETH G. COURAGE

This paper proposes a new approach for the design of traffic signal timings to coordinate the progression of traffic on arterial highways. The two most popular signal optimization policies in use today are the maximal bandwidth approach and the minimum delay and stops approach. The new approach is proposed as a measure of the quality of progression perceived by the driver. It deals with progression opportunities (PROS), and the policy is to maximize the number of PROS available on an arterial signal system. It differs from the maximal bandwidth approach by considering progression opportunities that occur outside the traditional through progression band. Arterial progression design based on this approach will usually show decreased stops and delay compared with the maximal bandwidth design without suffering the loss of perceived progression associated with direct minimization of stops and delay. The number of progression opportunities presented to the driver at any point in time is, by definition, the number of successive green signals that will be encountered at the design speed without stopping.

The coordination of traffic signals on arterial highways is an extremely effective way of reducing excessive fuel consumption and annoying stops that cause delay as well as wear and tear on vehicles. As the sophistication of signal controllers has improved over the years, coordinated signal systems have been able to use a variety of phase sequences and other control parameters to improve traffic flow on these facilities, which remain the backbone of the urban transportation system. Likewise, the methods of optimizing signal settings have been enhanced by the increasing power and decreasing cost of off-line computational capabilities of digital computers.

At one time signal settings were determined from the time-space relation of signal timing and traffic flow by using manual methods. As researchers began to use computers to reduce the computational effort and to increase analysis flexibility, the objective of the design program was still based on the time-space relation and on maximizing the through bands to accomodate platoons of traffic. One of the first popular computer programs of this type was signalized arterial (SIGART), which produced offsets that maximized bandwidths based on cycle length, free speeds, and intersection spacing (1). SIGART could also favor one direction over the other to account for directional imbalances in demand by time of day. Other similar models have been proposed as well.

More recent models, progression analysis and signal system evaluation routine (PASSER) II (2) and maximal bandwidth (MAXBAND) (3), are based on the same underlying objective (maximizing through bandwidth) but, unlike earlier models, these models also take into account traffic demands to determine cycle length and splits. They are also more powerful in the functional aspect because, in addition to offsets, a range of cycle lengths, alternative phase sequences, and phase lengths can be optimized.

Maximal bandwidth is an appropriate design approach for arterials but does not adapt well to two-dimensional networks. Thus, the development of signal optimization strategies for networks has generally been based on minimizing a disutility, which has generally been a function of delay, stops, and, in some models, queue length. The traffic signal optimization program (SIGOP) (4) and traffic network study tool (TRANSYT) (5,6) are the more prominent models in this area.

Although the disutility approach is well accepted for network signal optimization, it has not been readily accepted for applications on arterials because through progression bands based on minimizing disutility may not be as clean as those produced by the maximal bandwidth method. A school of thought, nonetheless, contends that the disutility approach is indeed applicable to arterial design since the overall objective [i.e., minimizing delay and stops (and, optionally, other disutility values)] is actually more valid than simply maximization through bandwidth, which does not explicitly recognize the presence of traffic demand as a function of time. Thus, two somewhat conflicting design strategies might yield substantially different signal timings.

NEW APPROACH: PROGRESSION OPPORTUNITIES

The maximal bandwidth approach will clearly produce offsets and other signal timing parameters that result in good through green bands, albeit this approach does not recognize partial progression opportunities (i.e., over short sections of the arterial) or the actual presence of demand with respect to the timings produced. On this latter point, it is assumed that traffic will conform to the signal timing and that relatively intact platoons will propagate through the entire length of the arterial. Particularly on long arterials, the bandwidth approach may produce signal timings that produce large system stops and delay.

On the other hand, the more realistic disutility models necessarily consider the actual traffic demand, because it is requisite to this approach that the traffic flow be simulated accurately. Designs based on this method automatically consider all traffic demands, thus the short trip, partial progression, and demand-dependent considerations are taken into account. However, progression bands produced by disutility models are often neither continuous nor wide.

A logical question is, "Can these methods be combined?" Indeed they can. The progression opportunities (PROS) model was initially developed (7,8) to improve only the maximal bandwidth policy. (The original concept was referred to as forward link opportunities but the acronym FLOS led to obvious confusion.)

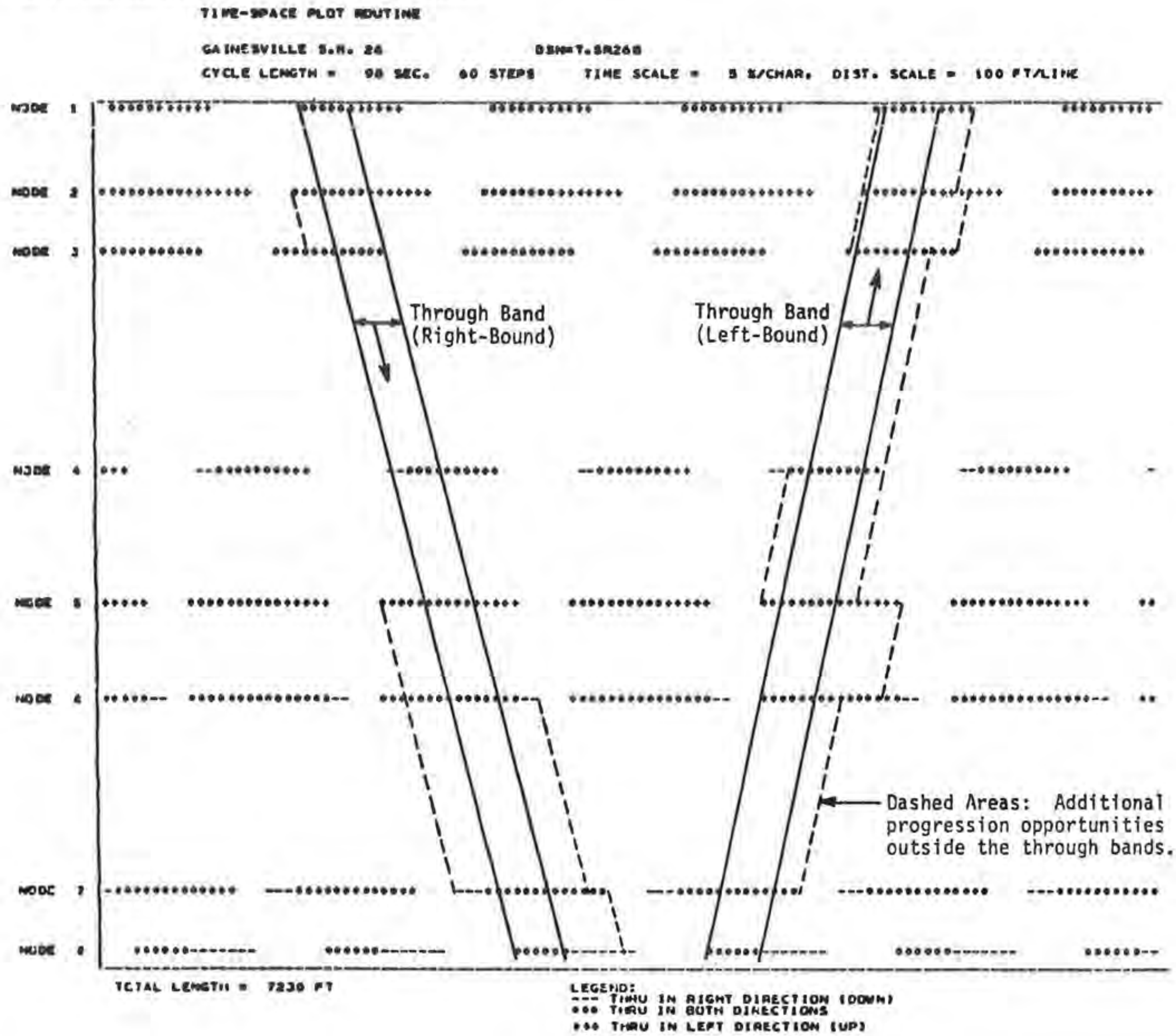
A progression opportunity is defined simply as the opportunity, presented at a given traffic signal and at a given point in time, to travel through a downstream signal without stopping. The number of progression opportunities presented to the driver at any time is determined by the number of successive green signals that will be encountered at the design speed without stopping. PROS can be determined for short increments of time, then accumulated to evaluate the total progression potential for any given set of signal timings.

PROS are based on a binary status function as follows (for one direction):

$$S_{jt} = \begin{cases} 1, \text{ if signal } j \text{ is green at time } t \text{ and signal } j+1 \text{ is green at} \\ \quad \text{time } t + T_{j,j+1} \\ 0, \text{ otherwise} \end{cases} \tag{1}$$

where S_{jt} is the status of signal (j) at time (t), where t ranges from 1 to the cycle length; and T is

Figure 1. Time-space diagram of maximal bandwidth optimization.



the travel time at progression speed between the specified intersections.

A single forward progression opportunity exists at intersection j whenever S_{jt} and $S_{(j+1)t}$ are 1 at any time increment (t).

PROS are then calculated for one direction by

$$PROS_{it} = -S_{it} + \sum_{j=i}^{n} \prod_{k=i}^{j} S_{kt} \qquad (2)$$

where $PROS_{it}$ is the forward progression opportunities from intersection (i) (of which there are n) at time interval (t).

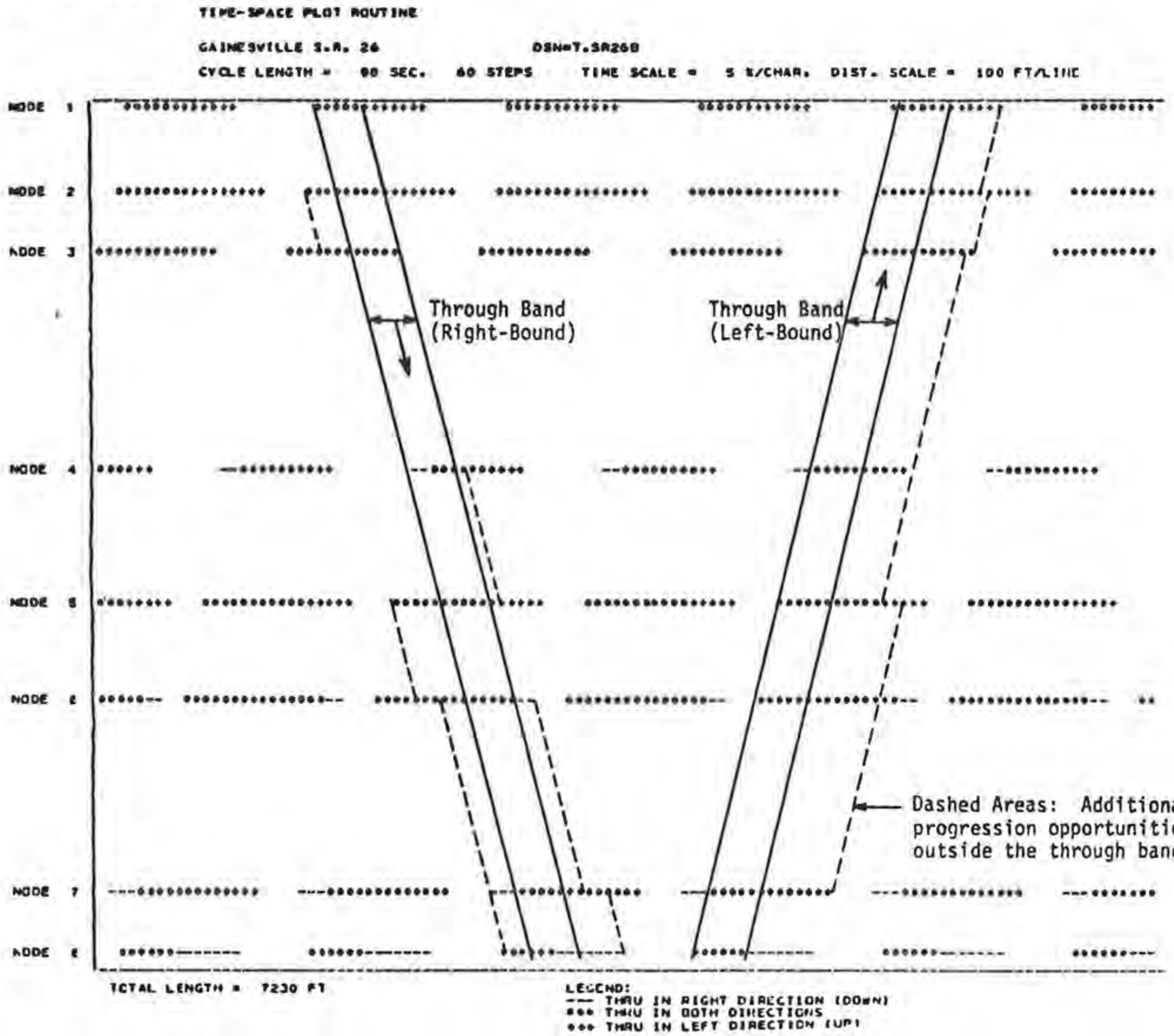
The product term in Equation 2 is necessary to count only those successive intersections for which all status variables are unity (for time increment t). It is necessary to decrement the sum by one (if S_{it} is equal to unity) to indicate that the value of PROS represents the number of downstream forward progression opportunities from intersection i.

The PROS concept can best be visualized by use of a diagram. Consider a standard time-space diagram for the through links on an arterial (Figure 1). This time-space diagram represents a maximal bandwidth solution that uses PASSER II (2). The through bands are indicated by the solid lines and other partial progression opportunities are indicated by the dashed lines. Notice that the placement of the green at intersection 5 is arbitrary because this signal is not critical to the through bands.

If the offsets are adjusted to maximize PROS, thus considering the partial progression opportunities, the time-space diagram in Figure 2 results.

Perhaps a more useful illustration can result from an alteration to the traditional time-space diagram. If the signal offsets are adjusted for travel time, the time-space diagram can be adjusted such that the progression speed has zero slope. When this is done, the distance between intersections is no longer relevant (at progression speed) and the distance scale can be collapsed into a dimensionless scale where only relative location (or order) of intersections is pertinent. The PROS can then be shown for each intersection for each time increment (t) as illustrated in Figure 3. The circled two indicates that there are two forward

Figure 2. Time-space diagram of maximal PROS optimization.



progression opportunities from intersection five at time increment 20. The rows that have PROS indicated for all signals at each time increment (thus PROS at the first signal rightbound equals seven) represent the through progression bands.

Figure 3 is the PROS analysis of the maximal bandwidth optimization shown in Figure 1. When Figure 2 is adjusted to a time-location diagram, Figure 4 results, which clearly indicates superior overall progression. Note also that the value for the circled example has increased to three.

The general model for total aggregate PROS is as follows:

$$PROS = C/T \sum_{k=1}^{2} \sum_{i=1}^{N} \sum_{t=1}^{T} PROS_{jt}, j = \begin{cases} 1 & k=1, \text{rightbound or} \\ N+1-i & k=2, \text{leftbound} \end{cases} \tag{3}$$

where C is the cycle length in seconds and all other variables have been previously defined.

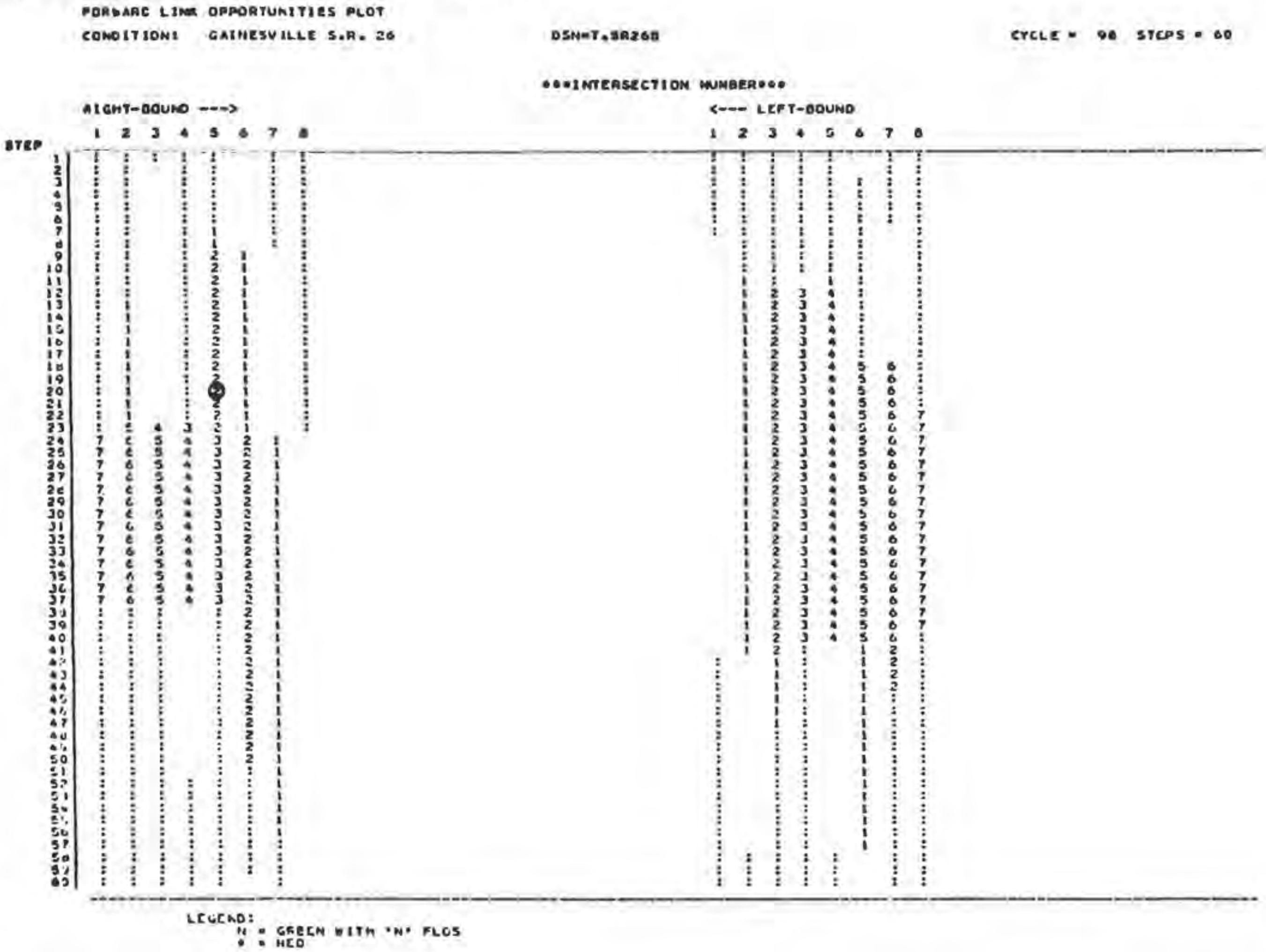
In the examples of Figures 3 and 4, the total PROS increased from 1978 in the maximal bandwidth solution to 2131 (or 7.7 percent) when offsets were changed to maximize PROS (i.e., max PROS in Equation 3).

To evaluate the PROS concept, the TRANSYT-6C model (9) was modified to perform the PROS calculations and to optimize offsets (and optionally splits) based on Equation 3. The use of TRANSYT permitted simulation of the alternative design strategies to evaluate their effectiveness.

A summary of comparative results of five arterials of differing configurations for the maximal bandwidth and PROS optimization policies is given in Table 1. As noted, most measures of effectiveness (MOE) were improved, albeit by very small magnitudes, by using the PROS optimization concept. In all these analyses, splits were based on balanced demand per capacity as determined by PASSER II, and these were held constant. Thus, only offsets were allowed to change.

The simple PROS optimization suffers the same disadvantage as the maximal bandwidth approach in that the actual traffic demand is not considered explicity. Variations of the objective functions were tested in which PROS were weighted by a number of other characteristics--namely, total demand, link length, link travel time, and stopline arrival pattern (i.e., a time-dependent demand weighting). None of these strategies demonstrated a significant

Figure 3. Time-location diagram of maximal bandwidth optimization.



improvement in the PROS optimization.

EXPANDING PROS MODEL

Although initial studies indicated only slight improvement in progression and other measures, it was evident that the PROS optimization strategy is at least as effective as maximal bandwidth. Additional studies were undertaken to improve the manner in which the PROS concept was implemented to determine whether a single model could improve the basic time-space approach.

For example, splits can be considered in the optimization. When splits were optimized by using a simple PROS maximization, side street times were seriously affected. Since side streets are not considered in the PROS optimization, TRANSYT forces them to their minimums. The execution of multiple runs with different minimums was one way of overcoming this problem, but a more direct approach was desired.

Another problem with the PROS offsets was that, despite increased total progression opportunities, platoons were often propagated into the backs of queues, thus causing delay to through traffic. This is evident in Figure 4 for the leftbound direction since the leading edge of the band is essentially flat at time increment 22. This effect was responsible for the only limited improvements, and in some cases actual disimprovements, of the PROS optimizations given in Table 1.

Finally, the logical question of comparing these two policies with the minimal disutility must also be addressed. TRANSYT has a disutility function called the performance index (PI), which is computed as follows:

$$PI = \sum_{i}^{n} (w_{di}\, d_i + k w_{si}\, s_i) \qquad (4)$$

where

d_i = delay on link i, of which there are n links (vehicle·h/h);
s_i = stops on link i (vehicles/s);
k = stop penalty, which equates stops to delay; and
w_{di}, w_{si} = individual weights for link i.

For the purposes of this research, the stop penalty (k) was set to eight and the individual link weights were all set to unity. Note that the PI considers all links, including minor movements.

The objective of a normal TRANSYT optimization is to minimize PI, which is the equivalent of maximizing its inverse. A logical extension of the PROS concept was to redefine the TRANSYT objective function as follows:

$$\max\,(PROS/PI) \qquad (5)$$

subject to minimum phase length constraints, as usual.

By using this formulation, splits can be optimized in addition to offsets because the minor movements will be accounted for in the PI. This optimization function attempts to maximize main street progression, subject to maintaining sufficient green time on the minor approaches. Equation 5 will also

Figure 4. Time-location diagram of maximal PROS optimization.

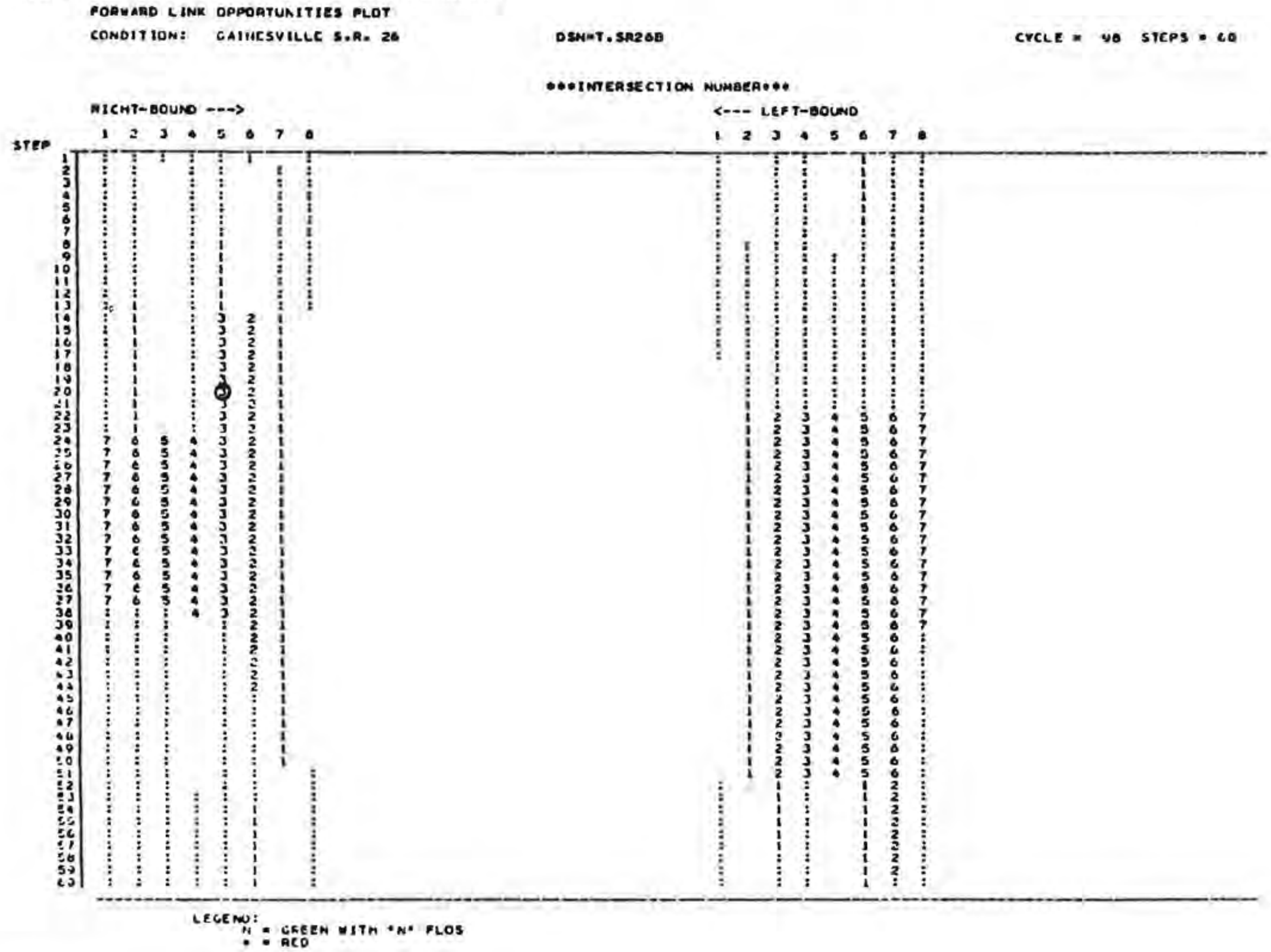


Table 1. Comparison of maximal bandwidth and PROS optimizations.

		Buffalo, Tampa		FL-26, Gainesville		FL-7A, Fort Lauderdale		Beech Daly, Detroit		FL-7B, Fort Lauderdale	
Characteristic	Optimization	No.	Change (%)	No.	Change (%)	No.	Change (%)	No.	Change (%)	No.	Change (%)
No. of signals		5		8		12		16		20	
Length (ft)		3450		7230		29 900		32 250		34 450	
Avg spacing (ft)		690		1033		2718		2140		1813	
Cycle length[a]		60		98		102		87		106	
Total bandwidth	BW	42.0		52.3		34.0		36.2		51.2	
	PROS	43.0	2.4	52.3	0	35.7	5.0	36.2	0	51.2	0
PROS	BW	675.0		1 977.9		2 686.0		5 573.7		11 150.4	
	PROS	684.0	1.3	2 131.5	6.7	2 779.5	3.5	5 663.6	1.6	11 719.2	5.1
Total delay (vehicle-h/h)	BW	46.17		67.77		208.49		209.03		384.73	
	PROS	46.16	~0	66.81	-1.4	205.88	-1.3	208.32	-0.3	374.87	-2.6
Delay on artery (vehicle-h/h)	BW	21.07		33.46		90.42		94.74		179.54	
	PROS	21.06	~0	33.75	0.9	88.98	-1.6	93.54	-1.3	168.91	-5.9
Total stops (%)	BW	62.1		45.9		70.7		61.6		58.6	
	PROS	61.4	-6.7	44.9	-1.0	69.8	-0.9	60.7	-0.9	57.2	-1.4
Stops on artery (%)	BW	52.5		35.8		62.3		51.1		47.8	
	PROS	51.5	-1.0	34.5	-1.3	61.4	-0.9	49.9	-1.2	46.0	-1.8
Fuel consumption (gal/h)	BW	88.58		199.62		537.66		721.50		1 138.59	
	PROS	88.36	-0.2	197.22	-1.2	534.91	-0.5	719.48	-0.3	1 131.37	-0.6

Notes: All MOE as estimated by TRANSYT-6C, plus PROS MOE.
BW = Bandwidth by using PASSER II; PROS = forward progression opportunities, offsets only.
[a]Based on PASSER II solution; phase sequences are based on PASSER II.

Table 2. Comparison of maximal bandwidth, TRANSYT, and PROS optimizations.

Characteristic	Optimization	Buffalo, Tampa		FL-26, Gainesville		FL-7A, Fort Lauderdale		Beech Daly, Detroit		FL-7B, Fort Lauderdale	
		No.	Change (%)	No.	Change (%)	No.	Change (%)	No.	Change (%)	No.	Change (%)
Total bandwidth	BW	42.0		52.3		34.0		36.2		51.2	
	PI	37.0	-11.9	47.4	-9.4	40.8	20	21.7	-40	26.5	-48.2
	PROS/PI	45.0	7.1	58.8	12.4	45.9	35	46.4	28.2	53.0	3.5
PROS	BW	675.0		1 977.9		2 686.0		5 573.7		11 150.4	
	PI	649.0	-3.8	1 977.5	~0	3 207.9	19.4	5 871.0	5.3	10 475.6	-6.0
	PROS/PI	740.0	9.6	2 232.7	12.9	3 501.9	30.4	6 901.9	23.8	12 727.8	14.2
Total delay (vehicle-h/h)	BW	46.17		67.77		208.49		209.03		384.73	
	PI	45.20	-2.1	62.70	-7.5	199.10	-4.5	196.49	-6.0	360.84	-6.2
	PROS/PI	46.29	0.3	64.38	-5.0	202.44	-2.9	201.79	-3.5	366.75	-4.7
Delay on artery (vehicle-h/h)	BW	21.07		33.46		90.42		94.74		179.54	
	PI	20.06	-4.8	28.50	-14.8	76.44	-15.5	79.30	-16.3	152.56	-15.0
	PROS/PI	18.28	-13.2	26.30	-21.4	73.48	-18.7	75.03	-20.8	146.87	-18.2
Total stops (%)	BW	62.1		45.9		70.7		61.6		58.6	
	PI	50.3	-2.6	39.9	-6.0	67.4	-3.3	56.4	-5.2	55.2	-3.4
	PROS/PI	59.5	-2.6	40.3	-5.6	68.1	-2.6	57.1	-4.5	55.1	-3.5
Stops on artery (%)	BW	52.5		35.8		62.3		51.1		47.8	
	PI	50.3	-2.2	28.6	-7.2	55.8	-6.5	43.7	-7.4	43.0	-4.8
	PROS/PI	46.1	-6.4	28.3	-7.5	56.1	-6.2	43.7	-7.4	42.7	-5.1
Fuel consumption (gal/h)	BW	88.58		199.62		537.66		721.50		1 138.59	
	PI	87.61	-1.1	193.79	-2.9	526.04	-2.2	707.94	-1.9	1 118.86	-1.7
	PROS/PI	87.42	-1.3	194.35	-2.6	526.94	-2.0	709.77	-1.6	1 119.62	-1.7

Notes: All characteristics, cycle lengths, and phase sequences as per Table 1.
BW = PASSER II bandwidth, PI = TRANSYT splits and offsets, PROS/PI = PROS/PI with offsets and splits.

result in offsets that tend to clear the existing queues before the progressed platoons arrive.

Table 2 contains the comparative results of the same five arterial highways by using Equation 5 as the objective function. Again, all values are based on TRANSYT-6C estimates of MOE. The standard TRANSYT PI optimization is also included for comparison.

Several significant observations can be drawn from the results in Table 2, which are summarized below (with all comparisons referenced to the maximal bandwidth optimization as the base condition):

1. Optimization based on PROS/PI always increased both bandwidth and total PROS; PI optimization alone had mixed effects, but was consistently less effective than PROS/PI.
2. As expected, total system delay was consistently lowest by using the TRANSYT minimization of the PI, and the PROS/PI optimization generally reduced total delay as well.
3. Significant reductions in main street delay occurred with both PI and PROS/PI optimizations, with the latter consistently superior.
4. The percentages of total and main street stops were also consistently lower with the PI and PROS/PI methods, again with the latter being generally better.
5. Reductions in fuel consumption were mixed between these two techniques, but both were better than the results by using the maximal bandwidth technique.

One might reasonably ask why the PROS/PI strategy would increase bandwidth more than would a maximal bandwidth optimization technique. The answer lies in that not only are offsets better aligned for the progression of actual platoons, but also the split optimization based on system disutility yields better splits than the balanced demand per capacity techniques common to maximal bandwidth algorithms.

On the basis of these analyses, the PROS approach in general, and the PROS/PI optimization strategy in particular, offer significant potential as design approaches that recognize both design objectives of maximizing bandwidth and reducing system disutility.

The PROS would appear to be a reasonable indicator of perceived progression.

FUTURE RESEARCH

Although the PROS model appears to have merit, and the model has been automated by incorporation into TRANSYT-6C, several areas of additional research and development are needed.

First, the concept needs to be field tested. TRANSYT is sufficiently realistic that its estimates of several pertinent MOE suggest that traffic will operate more efficiently and with good progression by using the PROS (particularly the PROS/PI) optimization function, but the true test is field validation.

Second, additional sensitivity studies are needed to refine the model parameters further. Several weighting factors have been tested in a preliminary fashion. Among these, weighting of PROS by both a platoon dispersion factor (an inverse function of travel time) and by the stopline arrival pattern have shown promise in further improving the PROS optimization.

Third, the PROS model needs to be incorporated into a more recent version of TRANSYT (or some other model such as SIGOP) to improve the computational efficiency.

Finally, if this concept proves worthwhile, it should be incorporated into standard packages such as the Arterial Analysis Package, currently in preparation for the Federal Highway Administration.

REFERENCES

1. SIGART: Program Documentation Manual. Metropolitan Toronto Traffic Control System, Toronto, April 1965.
2. C.J. Messer, H.E. Haenel, and E.A. Loeppe. A Report on the User's Manual for Progression Analysis and Signal System Evaluation Routine--PASSER II. Texas Transportation Institute, College Station, Rept. 165-14, Aug. 1974.
3. J.C.D. Little and M.D. Kelson. Optimal Signal Timing for Arterial Signal Systems, Volume 1: Summary Report. Massachusetts Institute of Technology, Cambridge, April 1980.
4. E.B. Lieberman and J.L. Woo. SIGOP II: Program

to Calculate Optimal Cycle-Based Traffic Timing Patterns. Federal Highway Administration, 1975.
5. D.I. Robertson. TRANSYT: A Traffic Network Study Tool. Ministry of Transport, Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Rept. LR 253, 1969.
6. D.I. Robertson and P. Gower. User Guide to TRANSYT, Version 6. Ministry of Transport, Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Supplementary Rept. 255, 1977.
7. C.E. Wallace. Development of a Forward Link Opportunities Model for Optimization of Traffic Signal Progression on Arterial Highways. Univ. of Florida, Gainesville, Ph. D. dissertation, 1979.
8. C.E. Wallace. Forward Link Opportunities: A New Concept for Optimizing Traffic Signal Progression. 11th Annual Pittsburgh Conference on Modeling and Simulation, Pittsburgh, May 1980.
9. P.P. Jovanis, A.D. May, and A. Diekman. Further Analysis and Evaluation of Selected Impacts of Traffic Management Strategies on Surface Streets. Institute of Transportation Studies, Univ. of California, Berkeley, Rept. UCB-ITS-RR-77-9, 1977.

Publication of this paper sponsored by Task Force on Traffic Signal Systems.

Macroscopic Traffic Delay Model of Bus Signal Preemption

A. ESSAM RADWAN AND JAMIE W. HURLEY, JR.

Productivity enhancement of public transportation is an essential goal, and bus signal preemption at intersections is one of the transportation system management strategies that strives for this goal. Improvements in bus speed and reductions in delay are the anticipated benefits accrued from such strategy. A macroscopic traffic delay model, which applies stochastic procedure, is presented to evaluate different bus preemption signal strategies at an isolated intersection. The model permits the user to evaluate a certain operational strategy provided for bus traffic on both main and cross streets. The signal controller modeled in this paper has a green extension and red truncation capabilities. A comparison between preemption on both main and cross street and preemption on main street only is provided to validate the model's logic. Sensitivity analyses were implemented and it was found that the delay savings due to signal preemption are sensitive to saturation flow rate and to bus passenger load. Potential applications and further enhancement are suggested.

Transportation and traffic engineers realize the importance of system productivity and its major role in minimizing passenger delays and maximizing passengers throughput. Several transportation systems management (TSM) strategies have been identified to achieve such a goal, and one of those is the provision of bus priority treatment at urban intersections by means of signal preemption strategies. The federal government currently can fund the capital costs of TSM projects and it is necessary to investigate the worthiness of bus preemption.

Bus preemption demonstration experiments were conducted in Los Angeles, Miami, and Melbourne, Australia (1-3). All studies concluded that bus signal preemption could reduce total passenger delay. Two bus signal preemption studies (4,5), one in Sacramento, California, and the other in Concord, California, reported similar results and showed that, with low bus frequencies, the added delays to automobile occupants are negligible.

Two computer simulation models were developed and tested for bus signal preemption (6,7). These models are of a microscopic nature in which the status of the vehicle with regard to its location, speed, and delay is updated every small time interval. The Urban Traffic Control System-Bus Priority System (known as UTCS-BPS) and the network simulation-bus priority system (NETSIM-BPS) computer programs are, perhaps, the only packages available that provide bus preemption at urban intersections (8,9).

The complexity and high cost involved in developing and validating such software packages lead to the consideration of other macroscopic approaches. An analytical model of bus preemption, by using a deterministic vehicle arrival process, was developed for the purpose of evaluating pretimed signal priority treatments at isolated intersections (10). Delay values derived from this model are believed to be underestimated due to the deterministic nature. Another analytical model, which uses a stochastic approach, was developed to evaluate and assess priority treatment of buses at signalized intersections (11). The model provides green extension and red truncation signal strategies, however, it is limited to one direction signal capability (preemption on main street only).

Evaluation of signal strategy effects on traffic flow requires, in general, a detailed analysis of a vehicle's speed, location, acceleration and deceleration capabilities, and the status of the signal. Use of microscopic computer simulation packages for system evaluation and justification can be an accurate way. However, the time and cost involved in running the program constrains and sometimes prohibits the completion of an extensive analysis. A solution for this problem is to use macroscopic analytical models that can reasonably do the job with possibly 10 percent the price of the microscopic model. The analytical models cited in the literature lack the ability to evaluate the bus signal preemption option on both main and cross street approaches.

DELAY MODEL AT AN INTERSECTION

Several models have been developed for estimating queues and delays at signalized intersections. Winsten and coworkers were the first to use the binomial distribution in an analysis of delays at pretimed signals (12). The Poisson distribution that describes the arrival of vehicles at intersections has been used by Adams, Webster, and Wardrop (13-15). Newell used a model in which the arrival headways were assumed to have a shifted exponential distribution (16). Most models assume departures at equal time intervals, providing a queue exists and the first departure is at the start of the effective green time.

One of the better known models for delay is the one developed by Webster (14) by using data resulting from computer simulation of intersection operation. Because Webster's delay model has been tested at several locations in England and the United States and has been proven to be reliable, it was adapted for this paper. The average delay per vehicle as given by Webster is determined from the following formula:

$$d = [C(1-\lambda)^2/2(1-\lambda X)] + [X^2/2q(1-X)] - 0.65(C/q^2)^{1/3} X^{(2+5\lambda)} \qquad (1)$$

where

- d = average delay per vehicle on the particular intersection approach;
- C = cycle time;
- λ = proportion of the cycle that is effectively green for the phase under consideration (g/c);
- q = flow;
- S = saturation flow; and
- X = degree of saturation; this is the ratio of the actual flow to the maximum flow that can be passed through the intersection from this approach and is given by $X = q/\lambda S$ (if d and c are in vehicles per second).

The third term of Equation 1 was found to range from 5 to 15 percent of the total mean delay, and Allsop suggested (17) that the average delay may be taken as

$$d = 9/10 \left\{[C(1-\lambda)^2/2(1-\lambda X)] + [X^2/2q(1-X)]\right\} \qquad (2)$$

Equation 2 was used to develop an analytical model described in this paper. The basic concept of the model was to investigate all possible bus detection events at an intersection and list the corresponding signal cycle lengths and splits. Cycle lengths, proportions of the cycle that were effectively green, degrees of saturation, and flow rates were substituted in Equation 2 to determine the total delay per approaching vehicle. Appropriate adjustments and assumption were made to calculate passenger car delays and bus delays. For each bus detection event, the probability of signal preemption was estimated by assuming a Poisson distribution of vehicle arrivals. The expected delay figures were then calculated and compared with the initial delay figures for no preemption.

MODEL ASSUMPTIONS

The following are the assumptions made to formulate the analytical model:

1. Pretimed signal controller with a two-phase plan and a cycle length determined from Webster's optimum cycle formula (14):

$$C_o = (1.5L + 5)/(1 - Y) \qquad (3)$$

where

- C_o = optimum cycle time (s),
- L = total lost time per cycle (5 s/phase), and
- Y = sum of the maximum ratios of flow to saturation flow;

2. Minimum red phase durations for main and cross streets are determined from Webster's minimum cycle formula:

$$C_m = L/(1 - Y) \qquad (4)$$

3. Absolute minimum cycle length of 40 s and absolute maximum cycle length of 120 s;
4. Minimum green phase duration of 12 s;
5. Detectors set up around 250-ft upstream of the intersection with average time required for a bus to clear the intersection of 7 s; (The location of the detector does not allow for nearside bus stops.)
6. Green extension and red truncation strategies are provided;
7. Saturation flow rate of 1800 passenger car equivalent per hour per lane;
8. Bus weight of 2.25 passenger car equivalent; and
9. Average bus load of 35 passengers and a passenger car load of 1.4 passengers.

MODEL DEVELOPMENT

Probability Expressions

The general case of bus preemption (preemption on main and cross streets) was developed first. Some assumptions concerning preemption priorities were made:

1. Main street green extension,
2. Cross street green extension,
3. Main street red truncation, and
4. Cross street red truncation.

The minimum red phase constraints, estimated from Webster's minimum cycle length, combined with the green detection time period within which green extension is requested (last 6 s of green) created four possible operational scenarios, as shown in Figure 1. The four possible scenarios are as follows:

$R_{1min} > (G_2 - 6)$ and $R_{2min} > (G_1 - 6)$,
$R_{1min} > (G_2 - 6)$ and $R_{2min} < (G_1 - 6)$,
$R_{1min} < (G_2 - 6)$ and $R_{2min} > (G_1 - 6)$,
and
$R_{1min} < (G_2 - 6)$ and $R_{2min} < (G_1 - 6)$.

In general, a total of 10 operational cases exist for any signal cycle:

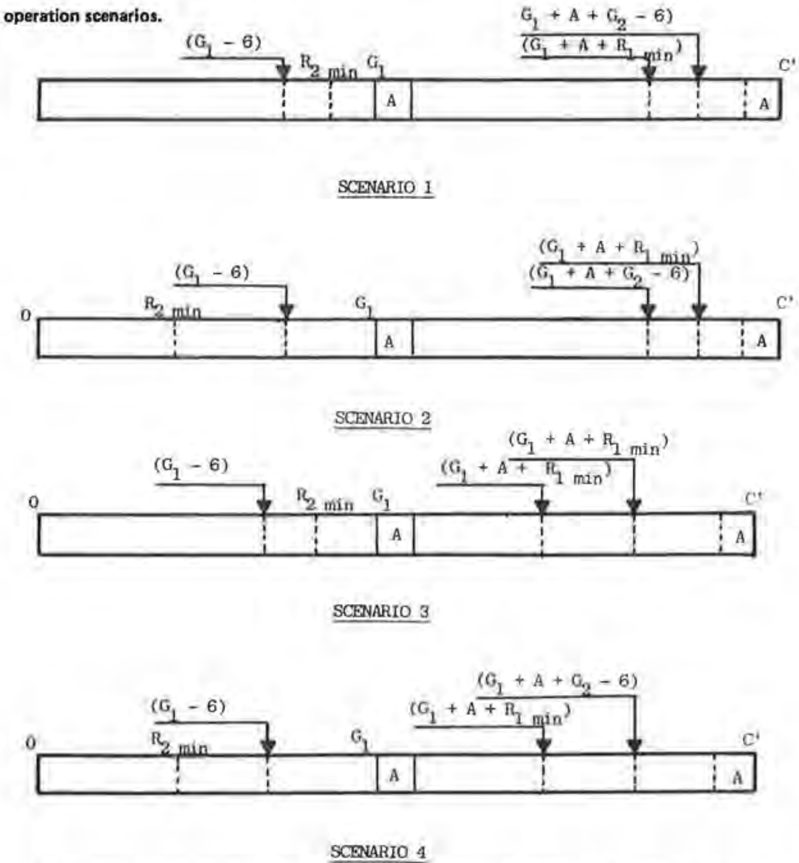
Case 1: No buses in a cycle,
Case 2: Buses arrive but there is no preemption,
Case 3: Main street green extension,
Case 4: Cross street green extension,
Case 5: Main street red truncation with red phase = R_{1min},
Case 6: Main street red truncation between ($G_1 + R_{1min}$) and ($G_1 + A + G_2 - 6$),
Case 7: Main street red truncation after ($G_1 + A + G_2 - 6$),
Case 8: Cross street red truncation with red phase = R_{2min},
Case 9: Cross street red truncation between R_{2min} and ($G_1 - 6$), and
Case 10: Cross street red truncation between ($G_1 - 6$) and G_1.

Each of these cases has unique characteristics that may include cycle length, splits, saturation flow rates, and special bus delay terms. The 10 cases listed are the possible operational cases that can occur for scenario 4. Examination of Figure 1 would reveal that the number of cases for Scenarios 1, 2, and 3 is 8, 9, and 9 cases, respectively.

To develop the probability of each of these cases it was necessary to divide the cycle into six appropriate intervals:

Interval 1: from ($C' - A$) to C', and from 0 to (R_{2min});
Interval 2: from (R_{2min}) to ($G_1 - 6$);
Interval 3: from ($G_1 - 6$) to G_1;
Interval 4: from G_1 to ($G_1 + R_{1min}$);

Figure 1. Four possible signal operation scenarios.



G_1: Green phase on main street

G_2: Green phase on cross street

A: Amber phase

$R_{1\,min}$: Minimum red phase on main street

$R_{2\,min}$: Minimum red phase on cross street

C': Webster's optimum

Interval 5: from $(G_1 + R_{1min})$ to $(G_1 + A + G_2 - 6)$; and
Interval 6: from $(G_1 + A + G_2 - 6)$ to $(C' - A)$.

The probability of no arrivals during each of the six intervals was developed by assuming a Poisson arrival distribution. The probability expressions are presented in a matrix format (Figure 2) with the element S_{ij} as the probability of interval i from approach j. Main street approaches and cross street approaches are subscripted 1 and 2, respectively.

The probability expression for each case that corresponds to each scenario was then derived as a function of the X-matrix components. A summary of the probability terms is shown in Figure 3. The sum of all the probabilities of each scenario is equal to unity.

Cycle Lengths

The cycle lengths that correspond to the 10 operational cases are given in the list below.

Case 1,

$$C = C' \tag{5}$$

Case 2,

$$C = C' \tag{6}$$

Case 3,

$$C = \sum_{n=1}^{3} (p_1 q_1^{n-1}) Y + C' \tag{7}$$

Case 4,

$$C = \sum_{n=1}^{3} (p_2 q_2^{n-1}) Z + C' \tag{8}$$

Case 5,

$$C = G_1 + A + R_{1\,min} \tag{9}$$

Case 6,

$$C = G_1 + A + R_{1\,min} + (G_2 - R_{1\,min} - 6)/2 \tag{10}$$

Case 7,

$$C = G_1 + A + G_2 - 3 \tag{11}$$

Case 8,

$$C = G_2 + A + R_{2\,min} \tag{12}$$

Case 9,

$$C = G_2 + A + R_{2\,min} + (G_1 - R_{2\,min} - 6)/2 \tag{13}$$

Figure 2. Probability of no bus arrival matrix.

$$X_{11} = \exp^{*}\left[-\frac{B_1 R_{2\,min}}{3600}\right] \qquad X_{12} = \exp\left[\frac{-B_2 R_{2\,min}}{3600}\right]$$

$$X_{21} = \exp\left[\frac{-B_1(G_1 - R_{2\,min} + A - 6)}{3600}\right] \qquad X_{22} = \exp\left[\frac{-B_2(G_1 - R_{2\,min} + A - 6)}{3600}\right]$$

$$X_{31} = \exp\left[\frac{6\,B_1}{3600}\right] \qquad X_{32} = \exp\left[\frac{-6\,B_2}{3600}\right]$$

$$X_{41} = \exp\left[\frac{-B_1\,(R_{1\,min})}{3600}\right] \qquad X_{42} = \exp\left[\frac{-B_2\,(R_{1\,min})}{3600}\right]$$

$$X_{51} = \exp\left[\frac{-B_1(C' - G_1 - A - R_{1\,min} - 6)}{3600}\right] \qquad X_{52} = \exp\left[\frac{-B_2(C' - G_1 - A - R_{1\,min} - 6)}{3600}\right]$$

$$X_{61} = \exp\left[\frac{-6\,B_1}{3600}\right] \qquad X_{62} = \exp\left[\frac{-6\,B_2}{3600}\right]$$

where: B_1 and B_2 are bus flow rates for main street and cross street, respectively.

* Napierian exponent

Figure 3. General probability expressions.

<table>
<tr><th rowspan="2">Case Number</th><th colspan="4">Scenario Number</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td>P_1 =</td><td colspan="4">$\exp\left[\frac{-(B_1 + B_2)\,C'}{3600}\right]$</td></tr>
<tr><td>P_2 =</td><td colspan="4">$\exp\left[\frac{-6\,[(B_1 + B_2) + B_2(G_1 + A) + B_1(C' - G_1 + A)]}{3600}\right] - \exp\left[\frac{-(B_1 + B_2)\,C'}{3600}\right]$</td></tr>
<tr><td>P_3 =</td><td>$(1 - X_{31})$</td><td>$(1 - X_{31})X_{12}X_{22}$</td><td>$(1 - X_{31})$</td><td>$(1 - X_{31})X_{12}X_{22}$</td></tr>
<tr><td>P_4 =</td><td>$(1 - X_{62})X_{31}$</td><td>$(1 - X_{62})X_{31}X_{12}X_{22}X_{32}$</td><td>$(1 - X_{62})X_{31}X_{41}X_{51}$</td><td>$(1 - X_{62})X_{31}X_{41}X_{51}X_{12}X_{22}X_{32}$</td></tr>
<tr><td>P_5 =</td><td>$(1 - X_{41})X_{31}X_{62}$</td><td>$(1 - X_{41})X_{31}X_{62}X_{12}X_{22}X_{32}$</td><td>$(1 - X_{41})X_{31}$</td><td>$(1 - X_{41})X_{31}X_{12}X_{22}X_{32}$</td></tr>
<tr><td>P_6 =</td><td>$(1 - X_{51}X_{61})X_{31}X_{41}X_{62}$</td><td>$(1 - X_{51}X_{61})X_{31}X_{41}X_{62}X_{12}X_{22}X_{32}$</td><td>$(1 - X_{51})X_{31}X_{41}$</td><td>$(1 - X_{51})X_{31}X_{41}X_{12}X_{22}X_{32}$</td></tr>
<tr><td>P_7 =</td><td>-</td><td>-</td><td>$(1 - X_{61})X_{31}X_{41}X_{51}X_{62}$</td><td>$(1 - X_{61})X_{31}X_{41}X_{51}X_{62}X_{12}X_{22}X_{32}$</td></tr>
<tr><td>P_8 =</td><td>$(1 - X_{12})X_{62}X_{31}X_{41}X_{51}X_{61}$</td><td>$(1 - X_{12})$</td><td>$(1 - X_{12})X_{62}X_{31}X_{41}X_{51}X_{61}$</td><td>$(1 - X_{12})$</td></tr>
<tr><td>P_9 =</td><td>$(1 - X_{22}X_{32})X_{62}X_{12}X_{31}X_{41}X_{51}X_{61}$</td><td>$(1 - X_{22})X_{12}$</td><td>$(1 - X_{22}X_{32})X_{62}X_{12}X_{31}X_{41}X_{51}X_{61}$</td><td>$(1 - X_{22})X_{12}$</td></tr>
<tr><td>P_{10} =</td><td>-</td><td>$(1 - X_{32})X_{12}X_{22}X_{31}$</td><td>-</td><td>$(1 - X_{32})X_{12}X_{22}X_{31}$</td></tr>
</table>

Case 10,

$$C = G_2 + A + G_1 - 3 \tag{14}$$

where

C' = Webster's optimum,
$p_1 = \exp(-7\ B_1/3600)$,
$q_1 = 1 - p_1$,
$Y = (n-1)\,[(3600/B_1) - 7\exp(-7\ B_1/3600) - (3600/B_1)\exp(-7\ B_1/3600)] + 7$,
$p_2 = \exp(-7\ B_2/3600)$,
$q_2 = 1 - p_2$, and
$Z = (n-1)\,[(3600/B_2) - 7\exp(-7\ B_2/3600) - (3600/B_2)\exp(-7\ B_2/3600)] + 7$.

The cycle length expression of cases 3 and 4 (green extension logic) was attained by adding the expected green extension to Webster's optimum cycle length. We assumed that no more than three buses can request green extension in any cycle. The expected green extension period was calculated by summing the product of the probability of exactly n arrivals to produce one headway greater than 7 s times the expected extension of n arrivals. The probability term is explained by a geometric distribution:

$$p_1(n) = p_1 q_1^{(n-1)} \tag{15}$$

where q_1 equals $1 - p$ and p_1 equals $\exp[-7\ B_1/3600]$ = probability of headway > 7.

As for the expected extension of n arrivals, the following derivation describes it:

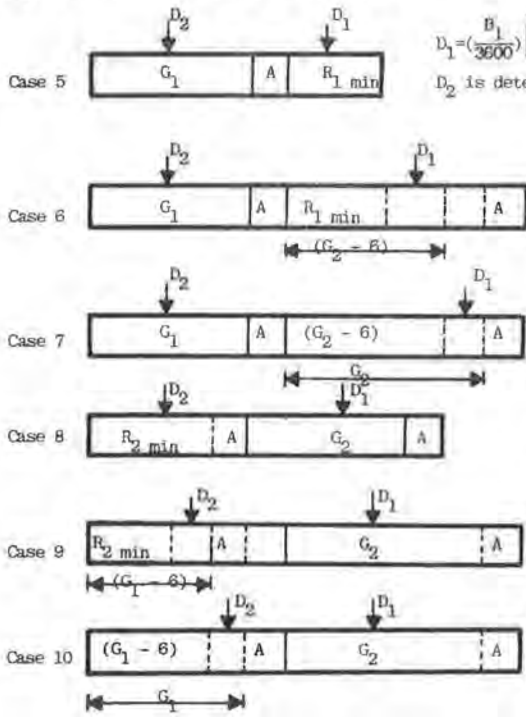
$$\text{Expected headway/(all headway} < 7) = \int_0^7 Y\lambda\exp(-\lambda Y)dY$$
$$= 1/\lambda - 7\exp(-7\lambda) - 1/\lambda\exp(-7\lambda) \tag{16}$$

Table 1. Bus delay terms used on model.

Case Number	Main Street	Cross Street
1. No buses in a cycle	Webster; S is adjusted for no bus arrivals	Webster; S is adjusted for no bus arrivals
2. Buses arrive, no preemption	Webster	Webster
3. Main street green extension	Webster	Webster
4. Cross street green extension	Webster	Webster
5. Main street red truncation ($R = R_{1\,min}$)	Compound delay model 1[a]	Webster
6. Main street red truncation ($G_1 + R_{1\,min} < R < G_1 + A + G_2 - 6$)	Compound delay model 2[a]	Webster
7. Main street red truncation ($G_1 + A + G_2 - 6 < R < C'$)	Compound delay model 3[a]	Webster
8. Cross street red truncation ($R = R_{2\,min}$)	Webster	Compound delay model 4[a]
9. Cross street red truncation ($R_{2\,min} < R < G_1 - 6$)	Webster	Compound delay model 5[a]
10. Cross street red truncation ($G_1 - 6 < R < G_1$)	Webster	Compound delay model 6[a]

[a]Shown in detail in Figure 4.

Figure 4. Red truncation compound delay models.



Case 5: $D_1=(\frac{B_1}{3600})\left[\frac{(R_{1\,min})^2}{2}\right]N_c+p(Stop)(13.78\ B_1)$

D_2 is determined from Webster with ($C=G_1+A$)

Case 6: $D_1=(\frac{B_1}{3600})\left[\frac{(G_2-6-R_{1\,min})^2}{2}\right]N_c+p(Stop)(13.78\ B_1)$

D_2 is determined from Webster with ($C=G_1+A+R_{1\,min}$)

Case 7: $D_1=(\frac{B_1}{3600})\left[\frac{(6)^2}{2}\right]N_c+p(Stop)(13.78\ B_1)$

D_2 is determined from Webster with ($C=G_1+A+G_2-6$)

Case 8: D_1 is determined from Webster with ($C=G_2+A$)

$D_2=(\frac{B_2}{3600})\left[\frac{(R_{2\,min})^2}{2}\right]N_c+p(Stop)(13.78\ B_2)$

Case 9: D_1 is determined from Webster with ($C=G_2+A+R_{2\,min}$)

$D_2=(\frac{B_2}{3600})\left[\frac{(G_1-6-R_{2\,min})^2}{2}\right]N_c+p(Stop)(13.78\ B_2)$

Case 10: D_1 is determined from Webster with ($C=G_2+A+G_1-6$)

$D_2=(\frac{B_2}{3600})\left[\frac{(6)^2}{2}\right]N_c+p(STop)(13.78\ B_2)$

Where: D_1 and D_2 are delay figures of buses arriving during the time periods indicated by arrows.

Total bus delay = D_1+D_2

N_c = Number of Cycles per hour.

$$\lambda = B_1/3600 \tag{17}$$

from Equations 16 and 17 the expected extension of n arrivals is defined as:

$$Y = (n-1)[(3600/B_1) - 7\exp(-7B_1/3600) - (3600/B_1)\exp(-7B_1/3600)] + 7 \tag{18}$$

Delay Estimation

Webster's delay model, shown in Equation 2, provides the average delay per approaching vehicle at a pretimed signalized intersection. The model does not differentiate between passenger cars and buses, therefore, some assumptions and adjustments had to be made to count for the difference. The average delay per vehicle was divided into stop time delay and delay due to speed change cycles. The derivation of both delay components is as follows:

Total delay = Total passenger car stopped time delay + passenger car time lost due to acceleration and deceleration + total bus stopped delay + bus time lost due to acceleration and deceleration.

The time lost in the queue was neglected in this derivation, and the speed profile adopted for estimating time lost due to acceleration and deceleration was as follows:

Vehicle Type	Speed: Initial (ft/s)	Final (ft/s)	Rate of Change (ft/s^2)
Passenger car	0	20	+8
	20	V	+4
	V	0.90V	-1
	0.90V	0	-7
Bus	0	V	+2
	V	0	-4

Table 2. Saturation headway and bus passenger load sensitivity results.

Item	Total Passenger Delay Savings (passenger-s)	
	Main Street	Cross Street
Saturation headway		
1800 vehicles/h	19 769	−2442[a]
1980 vehicles/h	18 747	1651
2160 vehicles/h	17 353	2083
Passenger bus load		
20	11 572	−2755[a]
35	19 769	−2442[a]
50	27 965	−2127[a]

[a]Delay was increased.

where V is the target speed in feet per second. This unimpeded speed profile was borrowed from the UTCS-BPS computer program (9).

It was assumed that the average target speed of heavily traveled urban streets is 25 mph. The time lost per passenger car speed change maneuver based on the same distance was found to be 5.56 s, and the corresponding value was 13.78 s/bus.

The probability of stopping more than once, as defined by Webster, was applied for the delay estimation. The breakdown of the delay components is

$$\text{Total delay per approach} = X + p(\text{stop})[5.56V_1 + 13.78B_1] \tag{19}$$

where

X = total stopped delay,
V_1 = hourly passenger car flow,
B_1 = hourly bus flow, and
$p(\text{Stop})$ = probability of stopping = $(1 - \lambda)/(1 - Y)$ and λ and Y were defined earlier in Equation 7.

$$\text{Total delay per passenger car per approach} = [X(V_1)/(V_1 + B_1)] + p(\text{stop})(5.56V_1) \tag{20}$$

The delay term defined in Equation 20 applies only for passenger cars and buses that operate under normal cycle length and phase splits with no preemption. The delay terms of buses for all possible signal cases are listed in Table 1 and detailed in Figure 4. As for passenger cars, Webster's model was assumed to apply to their delay estimation, and their benefit from signal preemption is reflected in the (G/c) term. The probability expressions and the delay equations were then coded into a computer program. The program calculates internally the total delay of passenger cars and buses under both preemption and nonpreemption strategies and provides the total delay saving (or losses) due to preemption.

Model Testing and Sensitivity Analyses

The model was applied to the following hypothetical setting:

Main street passenger car volume = 500 cars/h,
Cross street passenger car volume = 500 cars/h,
Main street buses = 40 buses/h,
Cross street buses = 10 buses/h, and
Saturation flow rate = 1800 vehicles/h.

The hypothetical setting resulted in operational strategies of scenario 1. The total delay per hour of passenger cars and buses for no preemption and preemption and the total passenger delay gains are given in the table below.

Results	Main Street Passenger Cars	Main Street Buses	Cross Street Passenger Cars	Cross Street Buses
Vehicle delay, no preemption (s)	8887	991	9 583	263
Vehicle delay, preemption (s)	8425	445	11 855	242

Savings (losses) attributed to signal preemption were 19 769 passenger-s for main street traffic but a delay of 2442 passenger-s was found for cross-street traffic. Total intersection savings were therefore 17 327 passenger-s. The results proved to be consistent and as expected in the sense that bus signal preemption helped both main street and cross street buses, with more benefits to main street. The main street bus delay saving amounted to 122 percent due to the preemption and the corresponding cross street saving was 9 percent. The bus delay saving on cross streets did not offset the passenger car delay loss, hence a total passenger delay loss was observed (2442 passenger-s).

A set of sensitivity analysis were implemented on the saturation headway and bus passenger load and they are given in Table 2. The increase in saturation headway caused less savings for main street and higher savings to cross street. As the results show, the model proved to be sensitive to cross street passenger delay savings between saturation headways of 1800 and 1980 vehicles/h.

An increase in the passenger bus load of 15 passengers caused a delay savings increase of 8197 passenger-s of main street passenger delay. This shows the significance of improving the bus passenger load. On the contrary, the reduction in delay losses on cross streets were insensitive to the increase in the passenger bus load.

The last step in the analysis was to test the model for bus preemption for main street only. This was attained by using a zero bus flow on the cross street. The total passenger saving for the basic hypothetical setting was found to be 21 549 passenger-s (4222 passenger-s higher than the preemption logic on both streets). This result is as expected because the zero bus flow on the cross street resulted in higher saturation flow rate, which provided higher total intersection delay savings, as proved earlier in Table 2.

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This paper presented an analytical model for evaluating and testing a possible bus preemption strategy at an isolated intersection. Four possible signal operational scenarios were identified and their corresponding probability terms were fully documented. Webster's delay model was adopted to estimate the average delay per vehicle per approach. The model was applied to a set of hypothetical demand rates to further validate the logic.

Sensitivity analyses were implemented, and it was concluded that the model was sensitive to an increase in the saturation flow rate from 1800 to 1980 vehicles/h and that the delay saving was insensitive beyond that point. In addition, the model results were found to be sensitive to bus passenger load.

The analytical model presented in this paper can be incorporated in the traffic network study tool (TRANSYT) computer package to develop signal splits and offsets in urban networks with provision for bus signal preemption on a main transit artery, an option that TRANSYT can not handle. The model can also be extended and enhanced to evaluate bus preemption strategies for a multiphase signal operation.

REFERENCES

1. H.K. Evans and G.W. Skiles. Improving Public Transit Through Bus Preemption of Traffic Signals. Traffic Quarterly, Vol. 24, No. 4, Oct. 1970.
2. J.A. Wattleworth, K.G. Courage, and C.E. Wallace. Evaluation of Bus-Priority Strategies on Northwest Seventh Avenue in Miami (Abridgment). TRB, Transportation Research Record 626, 1977, pp. 32-35.
3. A.J. Richardson and K.W. Ogden. Evaluation of Active Bus-Priority Signals. TRB, Transportation Research Record 718, 1979, pp. 5-12.
4. W.J. Elias. The Greenback Experiment--Signal Preemption for Express Buses: A Demonstration Project. California Department of Transportation, Sacramento, Rept. DMT-014, 1976.
5. C.D. Kinzel and others. City of Concord, Evaluation of Bus Priority Signal System. TJKM Transport Consultants, Walnut Creek, CA, 1978.
6. K. Wood. Bus-Actuated Signal Control at Isolated Intersections--A Simulation Model. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Supplementary Rept. 373, 1978.
7. R.A. Vincent, B.R. Cooper, and K. Wood. Bus-Actuated Signal Control at Isolated Intersections--Simulation Studies of Bus Priorities. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Laboratory Rept. 813, 1978.
8. G. Radelat. Bus Priority Systems Supplement to the Subroutine Documentation of the MITRE's Version of the Original UTCA-1 Model. Traffic Systems Division, Office of Research, Federal Highway Administration, 1973.
9. J.F. Torres and B. Mikhalkin; JFT Associates. Integration of BPS Strategies with NETSIM. FHWA, Oct. 1980.
10. I.B. VanBilderbeek. Priority Treatment of Buses at Traffic Signals. Proc., Technology Assessment Review, Organization for Economics and Development, Directorate for Scientific Affairs, Paris, France, 1979.
11. J. Jacobson and Y. Sheffi. Analytical Model of Traffic Delays Under Signal Preemption: Theory and Application. Paper presented at 59th Annual Meeting, TRB, 1980.
12. M. Beckmann, C.B. McGuire, and C.B. Winsten. Studies in the Economics of Transportation, Yale Univ. Press, New Haven, CT, 1956.
13. W.F. Adams. Road Traffic Considered as Random Series. Journal of the Institute of Civil Engineers, Vol. 4, 1936, pp. 121-130.
14. F.V. Webster. Traffic Signal Settings. Road Research Tech. Paper No. 39, Road Research Laboratory, London, 1958.
15. J.G. Wardrop. Some Theoretical Aspects of Road Traffic Research. Proc., Institute of Civil Engineers, Vol. 1, 1952, pp. 325-362.
16. G.F. Newell. Statistical Analysis of the Flow of Highway Traffic Through a Signalized Intersection. Quarterly Applied Mathematics, No. 13, 1956, pp. 353-369.
17. R.E. Allsop. Delay at a Fixed Time Traffic Signal. I: Theoretical Analysis, Transportation Science, Vol. 6, No. 3, pp. 260-285, 1972.

Publication of this paper sponsored by Task Force on Traffic Signal Systems.

Estimation of Average Phase Durations for Full-Actuated Signals

FENG-BOR LIN

A model for estimating the average green durations that result from full-actuated signal control is described. The model is developed primarily on the basis of probabilistic interactions between traffic flows and the control. It is structured according to three flow and control conditions: (a) the control employs motion detectors, (b) right-turn-on-red is either prohibited or does not affect the signal timing, and (c) left turns are made only from exclusive left-turn lanes. The discussions are focused on the formulation of the model. Applications of the model are also illustrated. The model can be used either manually or with the aid of a simple computer program.

Traffic-related phenomena at a signalized intersection, such as lane capacity, delays, queue length, and passenger car equivalent of left-turn vehicles, are influenced by the cycle splits and cycle length of the signal control. Under a full-actuated control, the cycle splits and the corresponding cycle length vary from one cycle to another. Consequently, it becomes desirable to estimate the average cycle splits of a full-actuated control to facilitate rational planning, design, and operation of signalized intersections. However, reliable and convenient methods for estimating full-actuated cycle splits are currently not available. This weakness may become increasingly critical when more full-actuated controls are used for intersection control.

To alleviate this problem, this paper presents a model that can be used either manually or with the aid of a simple computer program to obtain estimates of average full-actuated cycle splits. The model is structured on the basis of the following conditions:

1. The control relies on motion detectors to obtain information on traffic flow,
2. Right-turn-on-red is either prohibited or does not affect the signal timing, and
3. Left turns are made only from exclusive left-turn lanes.

In the following, a model for estimating the average cycle splits of a two-phase control is illustrated. The model is then expanded for application to cases that involve multiphase controls.

CONTROL LOGIC AND FLOW CHARACTERISTICS

A typical full-actuated signal control that employs motion detectors has the following control parameters for each signal phase i:

I_i = initial portion of green duration,
U_i = unit extension,
$(G_{max})_i$ = maximum allowable green duration,
Y_i = yellow duration, and
S_i = detector setback.

The logic of the control is simple. If phase i receives the green light, the first stage of its green duration will include the initial portion. After the time of the initial portion is completed, the green will be extended for an interval equal to the unit extension. If a vehicle actuates a detector of this phase during this unit extension, the green will be extended from the moment of actuation for another unit extension. The green can be extended repeatedly in the same fashion until the maximum allowable green is reached. If no vehicles actuate the detectors during an extended interval of one unit extension, the green will be terminated at the end of that unit extension.

In a given time interval vehicles in a traffic lane can be assumed to arrive at the upstream side of an intersection at random. With this approximation, the arrival pattern in each lane can be represented by a Poisson distribution (1):

$$P_j(n/t) = (\lambda_j t)^n \exp(-\lambda_j t)/n! \tag{1}$$

where $P_j(n/t)$ is the probability of having n arrivals in the jth lane in time interval t, and λ_j is the average flow rate in the jth lane during the same time interval t.

The headway distribution associated with $P_j(n/t)$ can be approximated by a shifted negative exponential function (1):

$$F_j(h \geq t) = \exp[-(t - \tau)/(H_j - \tau)] \tag{2}$$

where

$F_j(h \geq t)$ = probability that a vehicle headway h in the jth lane is greater than or equal to t,
$H_j = 1/\lambda_j$ = average headway of vehicles in in the jth lane, and
τ = minimum headway of vehicles in a traffic lane, equal approximately to 1 s.

When more than one traffic lane is present, several vehicles can cross a given reference point simultaneously. Therefore, the minimum headway of the vehicles in the combined flow approaches zero. Under this condition, the headway distribution of the combined flow can be approximated by

$$\begin{aligned} Z(h \geq t) &= \exp(-\lambda_1 t) \exp(-\lambda_2 t) \ldots \exp(-\lambda_m t) \\ &= \exp[-(\lambda_1 + \lambda_2 + \ldots + \lambda_m)t] \end{aligned} \tag{3}$$

where $Z(h \geq t)$ is the probability that a headway h of a combined multilane flow is greater than or equal to t and m is the number of traffic lanes involved.

Equations 2 and 3 can be combined into a single probability distribution:

$$f(h \geq t) = \exp[-\lambda(t - \tau)] \tag{4}$$

where

$$\lambda = 1/(H_1 - \tau) + 1/(H_2 - \tau) + \ldots + 1/(H_m - \tau) \tag{5}$$

and

$$\tau = \begin{cases} 1 & \text{if } m = 1 \\ 0 & \text{if } m > 1 \end{cases} \tag{6}$$

The probability density function associated with $f(h \geq t)$ is

$$f(h = t) = \lambda \exp[-\lambda(t - \tau)] \tag{7}$$

BASIC MODEL FOR TWO-PHASE CONTROL

Under full-actuated control, vehicles in each phase can extend the green according to the control logic and the settings of the control parameters. Figure 1 shows a representative two-phase timing sequence that results from such a control. The time interval $[\beta_i(i = 1,2)]$ as shown in the figure represents the length of the green beyond the initial portion for phase i. This time interval can be divided into two components:

1. D_{ni}--This component is the additional green extended by n vehicles that form moving queues upstream of the detectors after the initial portion of time has elapsed.
2. E_{ni}--This is the additional green extended in the absence of the moving queues by vehicles that cross the detectors at headways of no more than one unit extension after the initial portion of time has elapsed or after the moving queues disappear.

The portion of the green represented by D_{ni} precedes that represented by E_{ni}. Since the green in any cycle cannot exceed $(G_{max})_i$, D_{ni} should not exceed $(G_{max})_i - I_i$ and E_{ni} can at most equal $(G_{max})_i - I_i - D_{ni}$. Collectively, these individual constraints can be replaced by $I_i + D_{ni} + E_{ni} \leq (G_{max})_i$ without affecting the estimated values of the average green durations.

The values of D_{ni} and E_{ni} vary from one cycle to another. Thus, estimates of their expected values are necessary in order to determine the average green duration for each phase. Let the average values of D_{ni} and E_{ni} be denoted, respectively, as D_i and E_i. Furthermore, assume for the time being that no opposed left turns are involved. Then, D_i and E_i can be estimated according to the procedures described below.

Value of E_i

After the initial portion of phase i has elapsed two things may happen. One is that there are no moving queues upstream of the detectors in the lanes associated with this phase. The other is that moving queues may exist in some or all of the lanes. In the former case, the green can be extended if vehicles of phase i actuate at least one detector within U_i seconds of each other after the initial portion of time has elapsed. In the latter case, vehicles can still extend the green in the same manner after the moving queues disappear.

The probability that the green will be extended exactly k times is $[f(h \leq U_i)]^k f(h > U_i)$. Let the average length of each extension be denoted as J. Then, the additional green that results from the k extensions is $kJ + U_i$. This $kJ + U_i$ represents the value of E_{ni} in a given cycle for phase i. As indicated previously, this extended portion of the green cannot exceed $(G_{max})_i - I_i - D_{ni}$. For estimating the average green of phase i, however, this individual constraint may be neglected as long as the collective constraint $I_i + D_{ni} + E_{ni} \leq (G_{max})_i$ is satisfied. If we neglect the individual constraint, the expected value of E_{ni}, (i.e., E_i) can be conveniently estimated from

$$E_i = \sum_{k=0}^{\infty} (kJ + U_i)[f(h \le U_i)]^k f(h > U_i)$$
$$= U_i + J\{\exp[\lambda(U_i - \tau)] - 1\} \tag{8}$$

where

$$J = \left[\int_{\tau}^{U_i} t\, f(h = t)dt\right] / f(h < U_i)$$
$$= \{-[U_i + (1/\lambda)] + [\tau + (1/\lambda)] \exp[\lambda(U_i - \tau)]\} / \{\exp[\lambda(U_i - \tau)] - 1\} \tag{9}$$

Therefore,

$$E_i = -(1/\lambda) + [\tau + (1/\lambda)] \exp[\lambda(U_i - \tau)] \tag{10}$$

The sum of I_i and E_i cannot be allowed to exceed $(G_{max})_i$. Therefore, the value of E_i obtained from Equation 10 should be limited to a maximum of $(G_{max})_i - I_i$. If E_i exceeds $(G_{max})_i - I_i$, the average green will reach $(G_{max})_i$ and the estimation of D_i becomes unnecessary.

Value of D_i

After a green duration is terminated for phase i, arriving vehicles will begin to accumulate and form a stationary queue in each of the lanes associated with this phase. Once the light is turned to green again, the vehicles in the queues will start moving downstream while additional arriving vehicles may join the queues. If the number of the vehicles in a stationary queue is large and the flow rate of the same lane is high, a moving queue may exist upstream of the detector in that lane by the time the initial portion of time has elapsed. When this happens and a reasonably long unit extension is implemented, vehicles in the moving queue may cross the detection area at headways of no more than one unit extension. Under this condition, the green will be extended continuously until the moving queue disappears from every lane or until the maximum allowable green is reached.

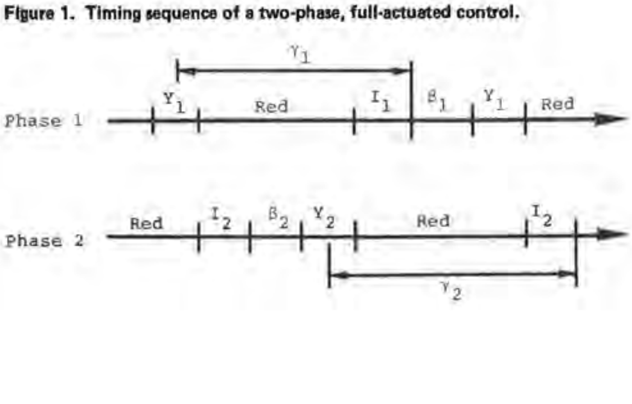
The growth and decay of a moving queue in a given cycle depends on the number of arrivals in time period γ_i (refer to Figure 1) and also on the flow rate of the traffic lane. γ_i represents a time interval in a cycle approximately from the beginning of the yellow duration of phase i until the end of the initial portion. For a two-phase control, the values of γ_1 and γ_2 are as follows:

$$\gamma_1 = 0.5Y_1 + I_2 + \beta_2 + Y_2 + I_1 \tag{11}$$

$$\gamma_2 = 0.5Y_1 + I_1 + \beta_1 + Y_1 + I_2 \tag{12}$$

In these two equations, it is assumed that 50 percent of the yellow duration in each phase will be used as the green duration by the vehicles in that phase. Therefore, vehicles that arrive in the first half of the yellow duration will enter the intersection instead of waiting for the next green.

Figure 1. Timing sequence of a two-phase, full-actuated control.



To allow a moving queue to extend the green, the first queueing vehicle immediately upstream of the detector should be able to move into or across the detection area before the first unit extension of time has elapsed. This feature of the control can generally be ensured if the following condition is satisfied:

$$w(N + 1) < I_i$$

where w is the average time required for each queueing vehicle to start moving after the light has turned from red to green (approximately 1.5 s), and N is the maximum number of queueing vehicles that may be stored between the stop line and the detector.

The above condition implies that the first vehicle immediately upstream of the detector should have started moving toward the detector before the initial portion of time has elapsed. This will allow the vehicle a time interval equal at least to one unit extension to move into the detection area at a sufficiently high speed (2-3 mph) to actuate the detector. In this study, this condition is considered satisfied.

To extend the green continuously, the vehicles in the moving queue should cross the detection area at headways of no more than one unit extension. Studies of queue discharge headways at intersections have revealed that a moving queue formed by signal interference usually reaches a steady average headway of about 2.2 s (2). Therefore, if the unit extension is set at a value of 3.5 s, as recommended by the Southern Section of the Institute of Traffic Engineers (3), every vehicle in the moving queue should have little difficulty extending the green.

Assume that the unit extension is greater than the headways at which vehicles in a moving queue cross the detection area. Furthermore, let n represent the number of vehicles in a lane that have arrived during γ_i. If the nth vehicle is in the queue, the time (B_n) required for this vehicle to reach the detector after the initial portion of time has elapsed can be approximated by

$$B_n = nw + [2(nL - S_i)/A]^{0.5} - I_i \tag{13}$$

where L is the average longitudinal space occupied by a vehicle (approximately 25 ft) and A is the average acceleration rate of a vehicle from a standing position (approximately 6 ft/s²).

If $B_n \le 0$, no moving queue exists upstream of the detector in a lane by the time the initial portion of time has elapsed. If B_n is greater than zero, the n arrivals will create a moving queue to extend the green after the initial portion of time has elapsed. By using representative values of w = 1.5 s, L = 25 ft, A = 6 ft/s², S_i = 120 ft, and I_i = 12.5 s, the smallest n needed to form a moving queue is 7 vehicles. In the following discussion the smallest n needed to form a moving queue will be denoted as n_{min}.

During time interval B_n additional vehicles may arrive and join the queue and, thus, continue to extend the green. Let μ be the rate at which the queueing vehicles move across the detector and λ_j, as defined previously, be the flow rate of the jth lane associated with phase i. If $\mu > \lambda_j$, the average time required for the moving queue to disappear from the upstream side of the detector in the jth lane is

$$\Delta_{nj} = \mu B_n / (\mu - \lambda_j) \tag{14}$$

The value of Δ_{nj} as given in this equation equals the time interval from the moment the initial portion of time has elapsed to the moment the moving queue disappears from the upstream side of the detector. To ensure that the maximum allowable green is not exceeded, a limiting value $(B_n)_{max}$ should be imposed on B_n. Given the value of E_i as obtained from Equation 10, this constraint can be stated as $I_i + \Delta_{nj} + E_i \leq (G_{max})_i$ or

$$B_n \leq (B_n)_{max} = [(\mu - \lambda_j)/\mu]\,[(G_{max})_i - I_i - E_i] \tag{15}$$

If phase i has m lanes, the longest Δ_{nj} in a cycle for this phase will determine the green extended by the moving queues in that cycle. In other words, D_{ni} for a given cycle is the maximum of Δ_{n1}, Δ_{n2}, ..., Δ_{nm}. Since the longest Δ_{nj} may not always be produced by the same lane flow in every cycle, the contribution of the moving queue in each lane to the extension of the green has to be estimated.

The probability that there will be no moving queue in the jth lane in a cycle is $P_j(n < n_{min})$. This probability is a function of the flow rate λ_j of the jth lane and the time interval γ_i. It can be determined from Equation 1 as the sum of $P_j(n = 0/\gamma_i)$, $P_n(n = 1/\gamma_i)$, ..., $P_j(n = n_{min} - 1/\gamma_i)$. The values of $P_j(n < n_{min})$ for $n_{min} = 7$ and various combinations of λ_j and γ_i are given in Figure 2.

When m lanes are associated with phase i, the probability that there will be no moving queue in any of the lanes in a cycle is the product of $P_1(n < n_{min})$, $P_2(n < n_{min})$, ..., $P_m(n < n_{min})$. Accordingly, the probability that there will be at least one moving queue in a cycle for phase i is

$$p_i = 1 - P_1(n < n_{min}) \cdot P_2(n < n_{min}) \ldots P_m(n < n_{min}) \tag{16}$$

Given that there is at least one moving queue in a cycle, the probability that the flow in the jth lane will result in the longest Δ_{nj} can be approximated by

$$g_j = \exp(0.0075\lambda_j) / \sum_{k=1}^{m} \exp(0.0075\lambda_k) \tag{17}$$

where λ_j is measured in vehicles per hour.

This equation was developed through a simple probabilistic simulation of the number of arrivals in each lane during time period γ_i and the resulting Δ_{nj} for various combinations of lane flows. Analytical formulation of g_j is possible but the computation requirements associated with it are prohibitively tedious.

Given that the jth lane has a moving queue and that this moving queue results in the longest Δ_{nj} in a cycle for a phase, the expected length of the green extended by the moving queue in this lane can be approximated by

$$\left[\sum_{n=n_{min}}^{\infty} P_j(n/\gamma_i) \cdot \Delta_{nj}\right] / [1 - P_j(n < n_{min})]$$
$$= [\mu/(\mu - \lambda_j)] \sum_{n=n_{min}}^{\infty} [B_n P_j(n/\gamma_i)] / [1 - P_j(n < n_{min})]$$
$$= [\mu/(\mu - \lambda_j)] \cdot \alpha_j \tag{18}$$

The values of α_j for B_n that satisfies Equation 15 are given in Figure 3 for various combinations of λ_j, γ_i, and $(B_n)_{max}$.

Taking into account the contribution of each lane flow to the extension of the green, the value of D_i can be determined as

$$D_i = \sum_{j=1}^{m} [\mu/(\mu - \lambda_j)]\, \alpha_j p_i g_j \tag{19a}$$

This equation is applicable if $\mu > \lambda_j$. If $\mu \leq \lambda_j$, then

$$D_i = (G_{max})_i - E_i - I_i \tag{19b}$$

With the aid of Figures 2 and 3, the estimation of D_i can be simplified. Table 1 gives an example computation of D_i for a phase that involves three lanes.

The value of D_i can also be estimated from a simple Monte Carlo simulation (4). Table 2 compares the values of D_i estimated from Equation 19a and those generated from such a simulation. The differences between the values obtained from the two methods are negligibly small.

Figure 2. Value of $P_j(n<n_{min})$ for n_{min} = 7.

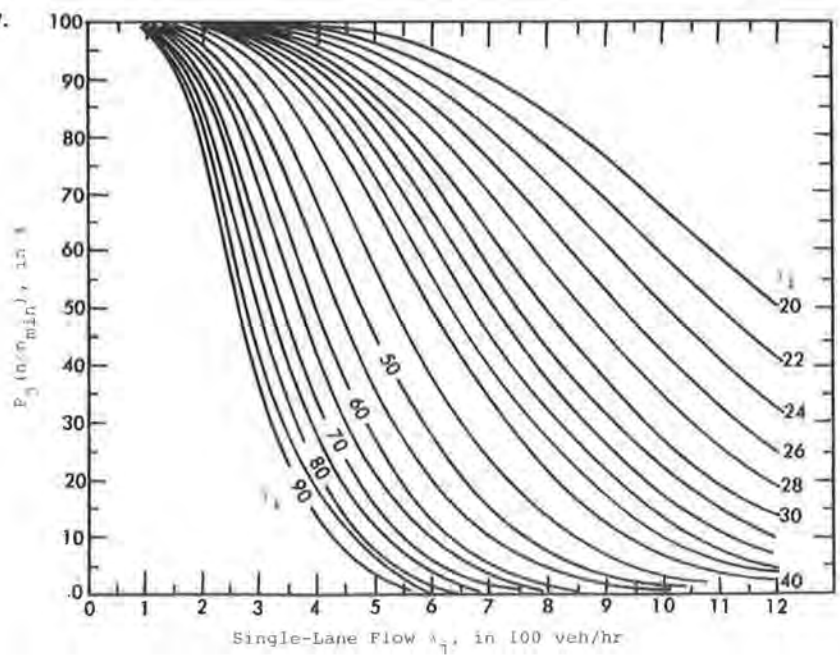


Figure 3. Value of α_j as a function of $\lambda_j\gamma_i$ and $(B_n)_{max}$.

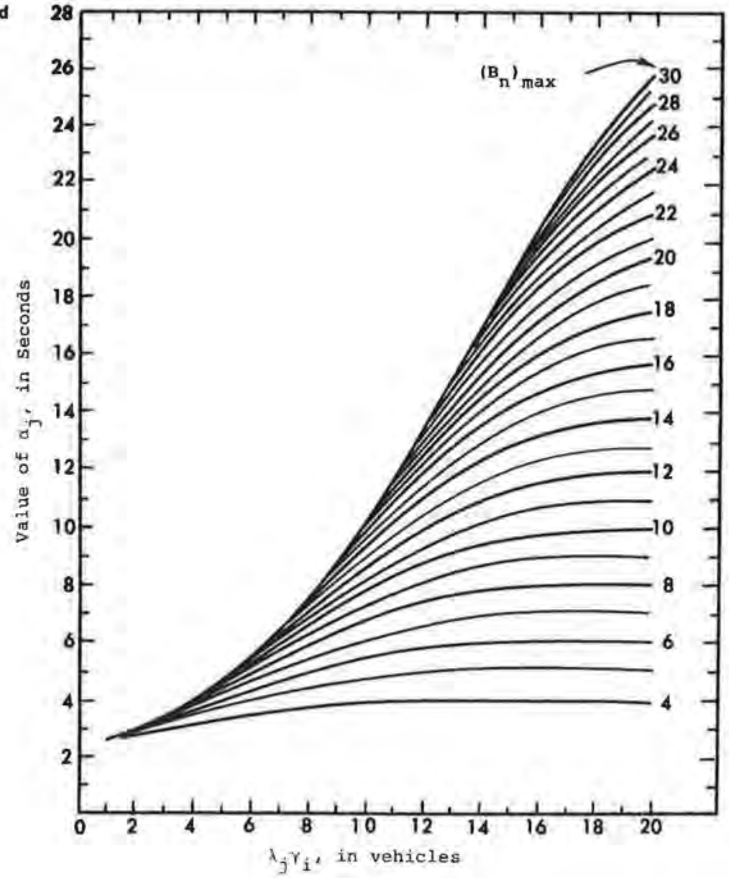


Table 1. Example computation of D_i.

Parameter	Lane, j 1	2	3	Source
$P_j(n < n_{min})$	0.076	0.975	0.675	Figure 2
ρ_i	0.950	0.950	0.950	Equation 16
$\exp(0.0075\lambda_j)$	403	4	20	
ε_j	0.944	0.009	0.047	Equation 17
$(\mu - \lambda_j)/\mu$	0.50	0.88	0.75	
$(B_n)_{max}$	15	26.4	23.5	Equation 15
$\lambda_j\gamma_i$	11.1	2.8	5.6	
α_j	10.8	3.4	5.3	Figure 3
D_i	20.1	20.1	20.1	Equation 19a

Note: This table is based on the following set of data:
$(G_{max})_i - I_i - E_i = 30$ s,
$n_{max} = 7$,
$\gamma_i = 50$ s,
$\lambda_1 = 800$ vehicles/h,
$\lambda_2 = 200$ vehicles/h,
$\lambda_3 = 400$ vehicles/h, and
$\mu = 1600$ vehicles/h.

Table 2. Comparison of simulated and estimated D_i (μ = 1600 vehicle/h).

Case	Flow Rate Lane, Lane j (vehicles/h) 1	2	3	4	Simulated	Estimated
1	800	200	400	0	21.1	20.1
2	600	200	400	0	10.2	9.4
3	600	800	300	100	21.6	19.2
4	500	500	300	300	8.1	7.4
5	500	300	200	100	5.2	5.8
6	1000	800	400	200	29.1	26.6
7	1000	800	400	200	9.2	9.2
8	600	800	300	100	3.2	3.8
9	1000	400	0	200	7.8	7.5
10	600	300	200	200	0.7	1.1
11	800	500	200	100	3.2	3.5

Note: For cases 1 through 6, $\gamma_i = 50$ s and $(G_{max})_i - I_i - E_i = 30$ s; for cases 7 through 11, $\gamma_i = 25$ s and $(G_{max})_i - I_i - E_i = 20$ s.

A Model

Given I_i, E_i, and D_i, the average green (G_i) for phase i can be determined as

$$G_i = I_i + D_i + E_i \tag{20}$$

In this equation D_i is a function of γ_i, which, in turn, depends on β_i (refer to Figure 1 and Equations 11 and 12). The length of β_i varies from one cycle to another. As an approximation, β_i can be used to represent the average length of the extended portion of the green for phase i. With this approximation, the sum of I_i and β_i equals the average green of phase i and Equations 11 and 12 can be rewritten as

$$\gamma_1 = 0.5Y_1 + Y_2 + I_1 + G_2 \tag{21}$$

$$\gamma_2 = 0.5Y_2 + Y_1 + I_2 + G_1 \tag{22}$$

Thus, γ_1 depends on G_2 and γ_2 varies with G_1. Since G_1 and G_2 are unknown, an iteration process has to be used to determine their values. The iteration process may include the following steps:

Step 1: Let $G_2 = I_2 + U_2$;
Step 2: Determine γ_1 from Equation 21 and use Equations 10, 19, and 20 to obtain an estimate of G_1; denote the estimate as G_1;

Step 3: Use $\tilde{G}_1$ in Equation 22 to determine γ_2 and use Equations 10, 19, and 20 to obtain an estimate of G_2; denote the estimate as $\tilde{G}_2$; and

Step 4: Compare the assumed G_2 and the estimated $\tilde{G}_2$; if the difference is sufficiently small (e.g., $\leq$1 s), stop the iteration. Otherwise, use the estimated $\tilde{G}_2$ as a new trial value and go back to the second step.

This process generally requires less than two iterations to obtain estimates of G_1 and G_2 that are within 1 s of the final solutions.

Comparisons with Simulation Results

Prior to the current study, a microscopic simulation model has been developed for analyzing the performance of pretimed and traffic-actuated signal controls. The model contains a traffic flow processor and a signal processor. The flow processor generates random arrivals in each lane at an arbitrarily specified distance from the intersection. The arriving vehicles in each lane are then processed downstream according to a set of car-following, gap acceptance, and car-signal interaction behaviors. The location, speed, and acceleration of each vehicle are updated once every second. The processing of vehicles is dependent only on signal indications. The signal indications for each 1-s interval are determined by the signal processor according to the logic of a control.

Table 3 gives the average green durations (G_1 and G_2) obtained from the simulation model and from Equations 10, 19, and 20 for 12 combinations of flows. It shows that Equations 10, 19, and 20 result in estimates that are practically the same as those generated from the complex simulation model. Similar comparisons based on other combinations of $(G_{max})_1$, $(G_{max})_2$, and lane flows reveal that the differences between the simulated and the theoretical values of G_1 and G_2 are generally less than 2 s. In these comparisons, the simulation model employs (a) a driver reaction time of 1.5 s in a stationary queue that corresponds to $w = 1.5$ s in Equation 13 and (b) a minimum spacing of 25 ft between two successive vehicles that corresponds to $L = 25$ ft in the same equation. Furthermore, the value of μ in Equation 19a is assumed to be 1600 vehicles/h (i.e., about 1 vehicle in every 2.2 s) and that of A in Equation 13 equals 6 ft/s². The settings of the control parameters are $S_i = 120$ ft, $I_i = 12.5$ s, $U_i = 3.5$ s, and $Y_i = 3.5$ s. It can be shown that a 15-20 percent variation in the value of any w, L, A, and μ from the assumed values has negligible effects on the estimated values of G_1 and G_2.

TRANSFORMATION OF LEFT-TURN FLOWS

When left-turn flows are not opposed they can be treated essentially the same as straight-through flows. When they are opposed, however, left-turn vehicles may not freely actuate the detectors to extend the green duration. Consequently, a left-turn vehicle is less effective than a straight-through vehicle in extending a green duration.

The complexity of the interactions between opposed left turns and the signal control makes it difficult to examine analytically the influence of the turning movements on the average green durations. For this reason, the microscopic simulation model described previously was used as a tool to identify a mechanism for transforming an opposed left-turn flow into an equivalent straight-through flow. The simulation model has proved to be capable of providing a reasonable representation of opposed left-turn movements (5).

Based on an examination of more than 50 combinations of flow patterns that involve opposed left turns of up to 900 vehicles/h, the simulated average greens reveal that an opposed left-turn flow can be transformed into a straight-through equivalent flow according to the following constraint:

$$(Q_s)_{max} = 900 - Q_o \qquad (23)$$

where $(Q_s)_{max}$ is the maximum straight-through flow equivalent of an opposed left-turn flow in vehicles per hour, and Q_o is the opposing flow rate in vehicles per hour.

Equation 23 implies that, if an opposed left-turn flow exceeds its $(Q_s)_{max}$, then it can be transformed into a straight-through flow with a flow rate equal to $(Q_s)_{max}$. Otherwise, it can be treated directly as a straight-through flow. For example, a left-turn flow of 400 vehicles/h opposed by a flow of 600 vehicles/h has a $(Q_s)_{max}$ of 300 vehicles/h. Therefore, the left-turn flow of 400 vehicles/h is equivalent to a straight-through flow of 300 vehicles/h. If the left-turn flow is 200 vehicles/h instead, then this left-turn flow can be treated directly as a straight-through flow.

Figure 4 shows a comparison of simulated average green durations for a number of traffic patterns that involve opposed left turns and the corresponding durations estimated from Equations 10, 19, and 20 based on transformed left-turn flows. The differences are generally less than 4 s. The settings of the control parameters used in the comparison are $U_i = 3.5$ s, $I_i = 12.5$ s, $S_i = 120$ ft, $Y_i = 3.5$ s, and $(G_{max})_i$ ranges from 35 to 50 s.

IMPACT OF FLOW VARIATION ON CYCLE SPLITS

The model described previously can be applied to any time period within which the flow rate in each lane does not vary significantly. If the flow rate varies significantly with respect to time, an adjustment of the average green as estimated from Equations 10, 19, and 20 may be necessary.

One way to account for the impact of the flow variation on the cycle splits is to divide the time period into several intervals, each of which has a more or less constant flow rate. With this approximation, the model can be applied to each time interval and the resulting average green for each interval can then be used to obtain an estimate of the average green for the entire period. Figure 5 shows three approximated hourly flow patterns with varying degrees of flow variation. Pattern A has a constant average 5-min flow rate. Pattern B has a moderate variation in its average 5-min flow rate. The flow variation represented by pattern C is substantially higher than that of pattern B.

The extent of the variation in the flow rate can be conveniently defined in terms of the peak-hour factor described in the Highway Capacity Manual (6). The peak-hour factors associated with patterns A, B, and C are, respectively, 1.0, 0.85, and 0.70. Based on pattern B and pattern C, Equations 10, 19, and 20 were applied to each 5-min interval to obtain estimates of average green durations for various combinations of hourly lane volumes. The results are given in Table 4 along with estimates generated directly from the microscopic simulation model. This table shows that repeated application of Equations 10, 19, and 20 can adequately account for the impact of the flow variation on the average green of a signal phase.

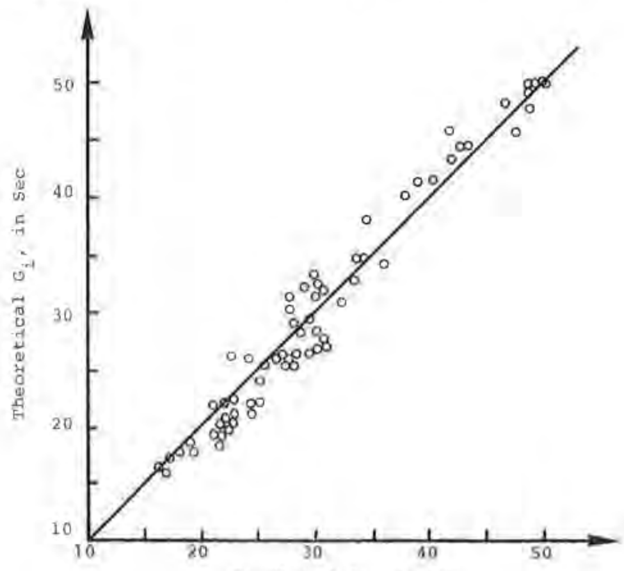
The repeated application of Equations 10, 19, and 20, however, is tedious unless it is aided with a

Table 3. Simulated values of G_1 and G_2 versus values obtained from $(G_{max})_1 = (G_{max})_2 = 35$ s.

Flow Rate (vehicles/h)				Avg Green (s)			
Phase 1		Phase 2		Simulation Model		Equations[a]	
Lane 1	Lane 2	Lane 1	Lane 2	G_1	G_2	G_1	G_2
600	200	300	100	22.1	17.6	20.7	17.0
600	200	600	200	23.0	23.0	22.3	22.3
600	200	750	250	25.1	29.2	24.3	28.0
600	200	900	300	27.0	34.1	25.8	32.4
600	200	1200	400	27.4	35.0	26.7	34.8
600	200	1500	500	27.4	35.0	26.7	35.0
450	150	300	100	19.0	17.4	18.1	16.9
450	150	600	200	18.9	21.2	18.6	21.1
450	150	750	250	19.8	27.3	19.3	25.8
450	150	900	300	20.4	31.1	20.1	30.5
450	150	1200	400	20.8	35.0	21.0	34.5
450	150	1500	500	20.8	35.0	20.1	35.0

[a] Values were obtained from Equations 10, 19, and 20.

Figure 4. Simulated and theoretical average greens that involve opposed left turns.

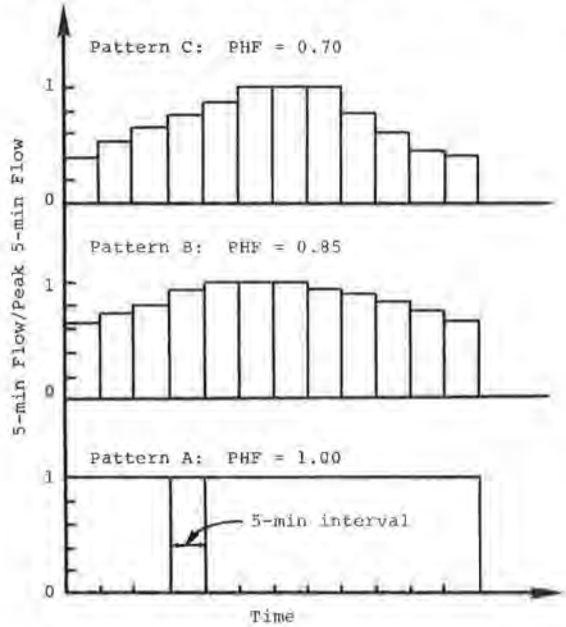


computer. Therefore, it may also be desirable to examine the affect the variation in the flow rate may have on the average cycle splits.

Figure 6 depicts a typical relation between the average green of a signal phase and the peak-hour factor (PHF). Based on the simulation output such as the one shown in this figure, the reduction in average green due to a lower peak-hour factor can be plotted against the average green obtained on the basis of a peak hour factor of 1.0. Figure 7 presents such reductions at four levels of maximum allowable green: 35, 40, 45, and 50 s. This figure indicates that, for all practical purposes, the impact of the flow variation can be neglected if the maximum allowable green for a phase is less than 35 s. For a higher setting of the maximum allowable green, an adjustment may be readily obtained from the figure.

For example, if the lane flows of a given signal phase have a peak-hour factor of 0.7 and the average green duration as estimated from Equations 10, 19, and 20 is 42 s, then Figure 7 indicates that the

Figure 5. Flow patterns with different degrees of flow variation.



estimated green should be reduced by 7 s if the maximum allowable green is 50 s and by 6 s if the maximum allowable green is 45 s.

Note that Figure 7 gives at best an approximate picture of the impact of the flow variation on the average green duration. A simple chart that can be used with an equal degree of accuracy for estimating the impact of the flow variation under varying traffic and signal control conditions is difficult to develop. If a greater degree of accuracy is desired, the repeated application of Equations 10, 19, and 20 may become necessary. I have a simple computer program that implements these equations for general applications.

MODEL FOR MULTIPHASE CONTROL

If the signal phases of a multiphase control are arranged in sequence without overlapping, Equations 10, 19, and 20 are still applicable. The only parameter that requires modifications is γ_i. Under a multiphase control, the value of γ_i can be generalized into the following form:

$$\gamma_i = 0.5Y_i + I_i + \sum_k (Y_k + G_k) \tag{24}$$

where Y_k, G_k are the yellow duration and the green duration of a competing phase k ($k \neq i$), respectively.

The estimation of G_i ($i = 1,2,\ldots$) still has to rely on an iteration procedure that may include the following steps:

Step 1: Transform opposed left-turn flows into equivalent straight-through flows according to the condition set forth in Equation 23;

Step 2: Let $G_i = I_i + U_i$ for $i = 2,3,4,\ldots$;

Step 3: Use Equation 24 to determine γ_1 (i.e., $\gamma_1 = 0.5Y_1 + I_1 + Y_2 = G_2 + Y_3 + \ldots$);

Step 4: Use Equations 10, 19, and 20 to obtain an estimate of G_1; denote this estimate as $\bar{G}_1$;

Step 5: Use $\bar{G}_1$ in Equation 22 to determine γ_2 (i.e., $\gamma_2 = 0.5Y_2 + I_2 + Y_1 + \bar{G}_1 + Y_3 + \bar{G}_3 + \ldots$) and then obtain an estimate of G_2 from Equations 10, 19, and 20; denote the estimate as $\bar{G}_2$; use $\bar{G}_1$ and $\bar{G}_2$ to estimate G_3; continue this task until an estimate

Table 4. Average green durations that result from nonuniform average flow rates for $(G_{max})_1 = (G_{max})_2 = 45$ s.

Hourly Flow Rate (vehicles/h)				Avg Green Duration(s), Peak-Hour Factor = 0.70				Avg Green Duration(s), Peak-Hour Factor = 0.85			
Phase 1		Phase 2		Equations[a]		Simulation Model		Equations[a]		Simulation Model	
Lane 1	Lane 2	Lane 1	Lane 2	G_1	G_2	G_1	G_2	G_1	G_2	G_1	G_2
200	200	200	200	17.1	17.1	17.6	17.5	16.8	16.8	17.4	17.5
200	200	600	600	17.2	25.8	18.0	25.8	16.9	24.5	17.8	25.7
200	200	800	800	17.5	32.1	18.5	32.0	17.1	32.8	17.9	35.1
200	200	1000	1000	17.6	36.7	17.7	37.9	17.3	39.6	17.7	41.2
600	600	200	200	25.8	17.2	25.7	17.7	24.5	16.9	25.8	17.4
600	600	600	600	30.5	30.5	28.2	28.4	30.2	30.2	31.7	32.4
600	600	800	800	31.9	35.2	31.4	38.9	33.0	38.4	34.2	40.2
600	600	1000	1000	32.7	38.4	32.3	39.6	34.4	42.6	35.4	42.8

[a]Values were obtained from Equations 10, 19, and 20.

Figure 6. Variation of average green duration related to peak-hour factor.

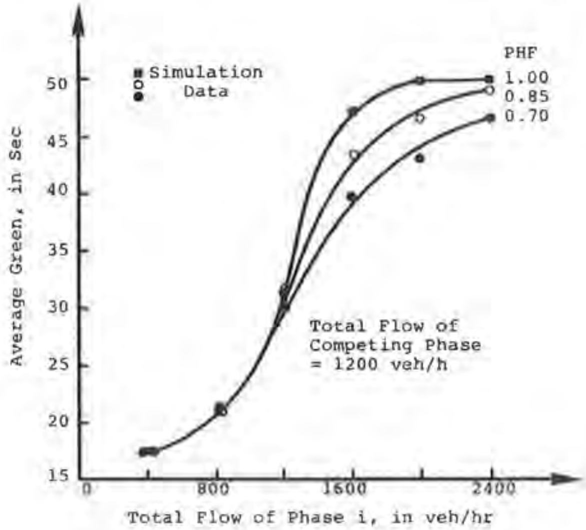
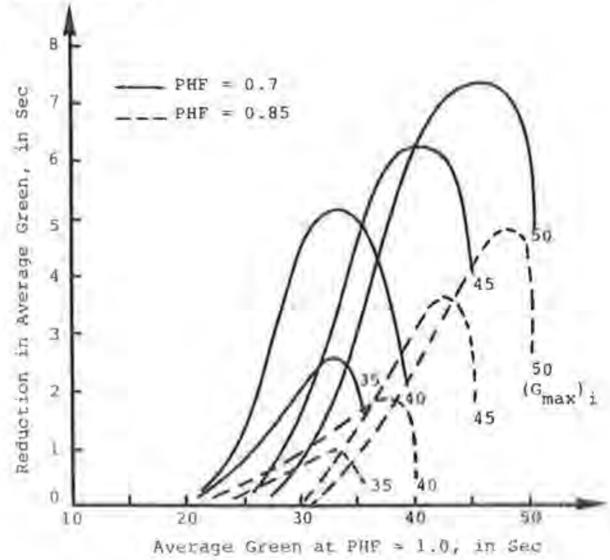


Figure 7. Reductions in average green durations at peak-hour factors smaller than 1.0.



of the average green is obtained for the last phase; and

Step 6: Compare the assumed $G_2, G_3, \ldots$ with the estimated $\bar{G}_2, \bar{G}_3, \ldots$; if any pair of G_i and $\bar{G}_i$ (i = 2,3,...) has a significant difference (e.g., >1 s), use $\bar{G}_2, \bar{G}_3, \ldots$ as the new trial values for $G_2, G_3, \ldots$ and go back to step 3.

This procedure generally requires only two iterations to obtain the needed estimates.

As in the case of the two-phase control, the impact of the flow variation on the average green can be taken into account in two ways. One is to use Figure 7 to estimate the needed adjustments in the values of the average greens. The other requires repeated application of the iteration procedure to successive time intervals. The computer program mentioned previously can also be used for this type of application.

CONCLUSIONS

The model as represented by Equation 10, 19, and 20 provides traffic engineers with a convenient tool to estimate full-actuated cycle splits. With the aid of charts to simplify computations, the model can be readily and manually used. It can also be implemented in the form of a simple computer program that requires limited computing facilities. The model can be incorporated into current methodologies that are being used for the planning, design, and operation of signalized intersections.

REFERENCES

1. D.L. Gerlough and M.J. Huber. Traffic Flow Theory. TRB, Special Rept. 165, 1975, 222 pp.
2. G.F. King and K.M. Wilkinson. Relationship of Signal Design to Discharge Headway, Approach Capacity, and Delay. TRB, Transportation Research Record 615, 1976, pp. 37-44.
3. Small-Area Detection at Intersection Approaches. Tech. Committee 18, Southern Section of the Institute of Traffic Engineers, Traffic Engineering, Vol. 44, No. 5, 1974, pp. 8-17.
4. D.R. Drew. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
5. F.B. Lin and T. Nadratowski. Effects of Signal Control on the Performance of Left-Turn Lanes. Paper presented at 60th Annual Meeting, TRB, 1981.
6. Highway Capacity Manual--1965. TRB, Special Rept. 87, 1965.

Publication of this paper sponsored by Task Force on Traffic Signal Systems.

Abridgment

Prescription for Demand-Responsive Urban Traffic Control

NATHAN H. GARTNER

State-of-the-art traffic control strategies in urban networks are calculated off-line, stored in a computer's memory, and selected for implementation on the street by various real-time criteria such as time-of-day, level of congestion, or special events. Many research efforts have been directed toward the development of new strategies that would relieve the traffic engineer of the constant burden of data collection and strategy revision and, at the same time, provide an improved level of service. This paper reviews results from past studies and analyzes their implications with respect to the development of improved generations of urban traffic control strategies. It then proposes a prescription for demand-responsive control that has the potential for overcoming many of the deficiencies of past efforts and leading to a significant improvement in urban traffic performance.

The large variety of control hardware and strategy software now available to the traffic engineer and system designer is changing continuously (1). The last decade has seen the introduction of computer-based traffic control systems in ever-increasing numbers. Several hundred such systems have already been installed and many more are under development throughout the world.

Strategies are commonly calculated off-line by arterial or network optimization techniques and are then stored in the computer's memory for implementation by various on-line criteria. Attempts have also been made to develop strategies that are calculated on-line in response to prevailing traffic conditions. The goal has been to improve traffic performance through adaptive control as well as to relieve the traffic engineer from the constant burden of data collection and strategy revision. These attempts have met with mixed success.

The emergence of new microprocessor technologies has given new impetus and new opportunities for the development of such strategies. The purpose of this paper is to assess the current status of strategy development and to offer a prescription for future developments.

STRATEGY DEVELOPMENT AND TESTING

Foremost among the computer-control strategies conducted during the past 15 years is the Urban Traffic Control System (UTCS) research project, which was conducted by the U.S. Department of Transportation (2).

The project was directed toward the development and testing of a variety of network control concepts and strategies, divided into three generations of control, as shown in Table 1. The different generations can briefly be characterized as follows.

First-Generation Control

First-generation control (1-GC) uses prestored signal timing plans that are calculated off-line based on historical traffic data. The plan that controls the traffic system can be selected on the basis of time-of-day (TOD), by direct operator selection or by matching from the existing library a plan best suited to recently measured traffic conditions (TRSP). The matching criterion is based on a network threshold value composed of volumes and occupancies. Frequency of update is 15 min. One-GC software also includes logic to enable a smooth transition between different signal-timing plans, a critical intersection control (CIC) feature that enables vehicle-actuated adjustment of green splits at selected signals, and a bus priority system (BPS) at specially instrumented intersections. Plans can be calculated by an off-line signal optimization method; traffic network study tool (TRANSYT)-generated plans were selected for testing in UTCS.

Second-Generation Control

Second-generation control (2-GC) is an on-line strategy that computes and implements in real-time signal timing plans based on surveillance data and predicted volumes. The optimization process [an on-line version of the traffic signal optimization program (SIGOP)] is repeated at 5-min intervals; however, to avoid transition disturbances, new timing plans cannot be implemented more often than every 10 min.

Third-Generation Control

The third-generation control (3-GC) strategy was conceived to implement and evaluate a fully responsive, on-line traffic control system. Similar to 2-GC, it computes control plans to minimize a network-wide objective by using predicted traffic conditions for input. The differences compared with 2-GC are that the period after which timing plans are revised is shorter (3-5 min), and cycle length is required (a priori) to vary among the signals as well as at the same signal during the control period (CP).

Analysis of the dynamics of control plan generation and implementation for the three UTCS strategies is important. In 1-GC the traffic pattern (volume and occupancy) during interval n-1 is used to make a decision whether a new plan should be called from the library for interval n. No prediction is used. Two-GC and 3-GC are similar in concept. Detector measurements are accumulated up to, and including, interval n-1. These data are used during interval n to predict volumes and speeds and to generate the timing plans that are then implemented in interval n+1. In both strategies the traffic data used in the timing plan that is being implemented are displaced by at least one interval from the corresponding measured flows. The different UTCS control strategies were designed to provide an increasing degree of traffic responsiveness through a reduction of the update interval, with a view to improving urban street network performance. However, results of extensive field testing showed that the expectations were not entirely fulfilled (2,3).

One-GC, in its various modes of operation, performed overall best and demonstrated that it can provide some measurable reductions in total travel time over that which could be attained with a well-timed three-dial system. Two-GC had a mixed bag, but was overall inferior compared with 1-GC. Three-GC, in the form tested in the UTCS system, seriously degraded traffic flow under almost all the conditions for which it was evaluated. A summary of the results is given in Table 2 (3). Similar results were experienced in the Glasgow (4) and Toronto (5-7) experiments.

DISCUSSION OF RESULTS

From the studies cited above one may erroneously

Table 1. Characteristics of UTCS strategies.

Feature	First Generation	Second Generation	Third Generation
Update interval, control period	15 min	5-10 min	3-5 min, variable
Control plan generation	Off-line optimization, selection from library by time-of-day, traffic responsive, or manual mode	On-line optimization	On-line optimization
Traffic prediction	None	Historically based	Smoothed values
Critical intersection control	Fine tuning, splits	Fine tuning, splits and offsets	NA
Cycle length	Fixed within each section	Fixed within groups of intersections	Variable in time and space, predetermined for control period

Table 2. Comparison of results of UTCS strategies.

Generation	Traffic Responsive Strategy	Change in Aggregate Vehicle Minutes of Travel with Respect to Base (%)			
		Morning Peak	Off Peak	Evening Peak	All Day Avg
First	Arterial	-2.6	-4.0	-12.2	-5.7
	Network	-3.2	+1.9	-1.6	-1.3
Second	Arterial	-1.3	-3.8	+0.5	-2.1
	Network	+4.4	+1.9	+10.7	+5.2
Third	Arterial	+9.2	+24.0	+21.2	+16.9
	Network	+14.1	-0.5	+7.0	+8.2

conclude that a library of timing plans generated off-line, based on historical data (from another month, perhaps another year but for the same time period of the day), is more effective than timing plans generated on-line, based on very recent data (past 15, 5, or 3 min). However, a closer examination of those studies indicates that the expectations of the researchers were not fulfilled--not because their rationale was wrong (that traffic-responsive control should provide benefits over fixed-time control) but because of a failure of the models and procedures that they implemented to deliver the desired results. A major cause for this failure appears to be in the measurement-prediction cycle used by the procedures.

Most available traffic control methods claim to be traffic responsive in some sense. Even 1-GC strategies are traffic responsive to a certain extent--plans may be replaced at 15-min intervals in response to predicted traffic volume changes. But, these methods are not truly responsive: They do not respond to actual traffic conditions but to hypothetical conditions--the hypothesis is only as good as the model and the predictions used in the optimization. This is the most critical aspect in all the responsive strategies listed above. The traffic-flow process and the optimization procedure form an inseparable closed-loop control system. The control values can only be effective if an accurate model is used in the optimization. However, all the different generation-control strategies do not have an accurate model; they use an abstract model that is calibrated by predicted (thus inherently inaccurate) smoothed volume data. Such a model cannot take account of short-term fluctuations; in essence, by aggregating and smoothing the data, the information content that is most important for on-line demand-responsive control is destroyed.

Large discrepancies were observed (sometimes in excess of 50 percent) when comparing the performance of 2-GC and 3-GC predictors with actual volumes over successive 5-min intervals (8). When aggregated over shorter than 5-min periods, the discrepancies can be even larger. Moreover, suppose one could predict the volume in each cycle with complete accuracy (i.e., with a zero mean error value). Even then the resulting real-time control strategy might be ineffective. For example, the following numbers represent vehicle arrivals for two cycles, grouped into 5-s intervals, on a signal-controlled approach with a 60-s cycle time:

1 1 2 1 1 2 0 2 0 0 1 1

0 1 0 0 1 1 2 1 2 1 1 2

During both cycles the flow is the same (12 vehicles), yet the optimal control strategy for each should be entirely different because of the different distribution of the arrivals within the cycle.

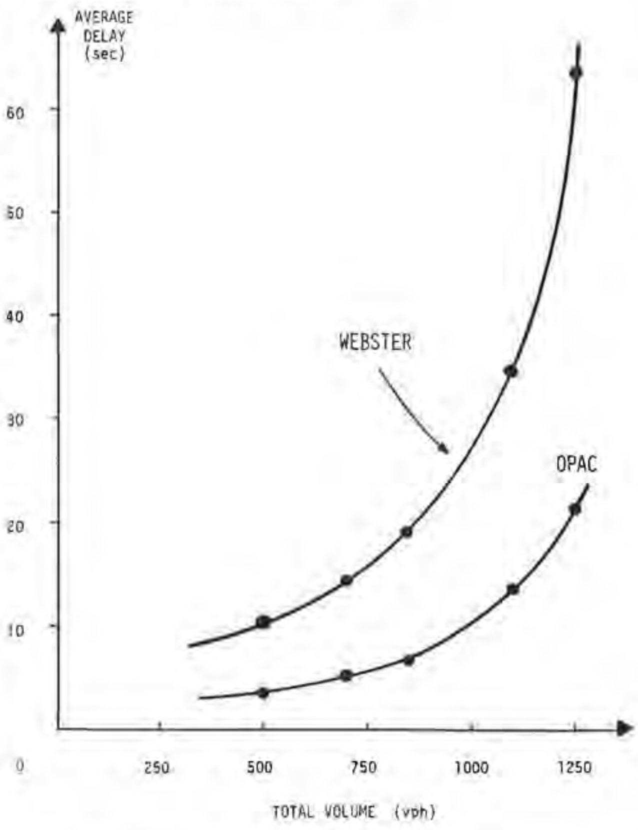
Clearly, an effective demand-responsive traffic control system requires the development of new concepts and not merely the extension of existing concepts toward shorter time frames (i.e., going down from hourly intervals to 15-, 5-, or 3-min intervals, or a cycle time) and using predicted values that are less and less reliable. Data from detectors provide information about past traffic behavior, but a traffic-responsive system must make decisions that result in good control in the future. Ways must be devised to predict future traffic behavior from past detector measurements.

PRESCRIPTION

On-line traffic control strategies should be capable of providing results that are better than those produced by the off-line methods. Simulation studies have indicated that, if under ideal conditions complete information on vehicle arrivals was available, responsive control strategies could reduce as much as 50 percent of the delay incurred by using existing nonresponsive strategies (9,10). To achieve this goal, the following requirements for the development of an effective demand-responsive traffic control system are proposed.

1. The system shall provide better performance than off-line methods. This is the primary criterion, everything else is secondary. Although it may seem self-evident, it was not always explicitly recognized in the development of responsive strategies. In some cases it was superseded by less relevant criteria such as mainstreet platoon progression or variable cycle time.
2. Development of new concepts is needed and not merely the extension of existing concepts. As demonstrated by the experiments reviewed in this paper, effective responsiveness is not achieved by implementing off-line methods at an increased frequency. New methods have to be developed.
3. The system must be truly demand-responsive; i.e., adapt to actual traffic conditions and not to historical or predicted values that may be far off from the actual.
4. It should not be arbitrarily restricted to control periods of any length but should be capable of updating plans at any time, at any location.

Figure 1. Comparison of average delay per vehicle.



5. The system should not be encumbered by a network model structure that requires extensive centralized computer capability. The model should be decentralized in its decisionmaking and use only those data that are directly pertinent to the decisions it has to reach. Decentralization increases the overall computation power, simplifies the data requirements and processing, and enhances the effectiveness of the control strategies that are generated.

6. The system should obviate the conventional notions of offset, split, and cycle time, which are inherent in all existing signal-optimization methods. The pattern of any individual signal should consist of a continuously varying, demand-responsive, sequence of on (effective green) and off (effective red) times that are only subjected to appropriate lower and upper bounds.

Can such a system be realized? The likelihood of its development is greatly enhanced by the continuous improvements in microprocessor technologies. By combining the potential capability of the microprocessor with demand-responsive strategies such as those proposed by Miller (11) or those used in SCOOT (12), SCAT (13), or the optimization policies for adaptive control (OPAC) programs (14), it is to be expected that substantial advances in the state of the art can be achieved.

An example of the potential benefits that can be expected from truly demand-responsive strategies is shown in Figure 1 (14). It compares the average delay at a two-phase signal-controlled intersection when timings are determined by Webster's method and by an OPAC strategy. OPAC is a demand-responsive strategy that dynamically optimizes signal timings. It uses a rolling horizon concept based on a combination of measured and calculated arrival patterns and can be implemented by a microprocessor. This strategy can provide, under ideal conditions, up to 60 percent reduction in delay with respect to a fixed-time strategy. Although such performance may be hard to expect in real life, this result is an indication of the tremendous opportunities that microprocessor-based demand-responsive strategies can offer. Undoubtedly, much more research and experimentation would be needed to take advantage of these opportunities.

ACKNOWLEDGMENT

This paper is based, in part, on a research project conducted under sponsorship of the Office of University Research, U.S. Department of Transportation. The opinions expressed in the paper, however, are mine and not necessarily those of the sponsoring agency.

Discussion

K. Todd

The UTCS evaluation study (3) reported a statistically significant reduction in delay through TRSP versus three-dial, averaging 3.9 percent throughout the test area during the evening peak period. On the strength of this result, the author, as others have done before him, considers TRSP more effective than a well-timed three-dial system.

The listed improvement was primarily brought about through delay reductions for outbound traffic on sections 1 and 2. For example, on outbound Wisconsin Avenue (route 11, section 2), delay was reduced by 15.3 percent (15). Without such improvements, TRSP would have given little or no advantage over three-dial, possibly a degradation.

Note that the evaluation study omitted to measure whether this reduction in delay might not have been accompanied by longer delays at critical downstream intersections beyond the test area. Was this 15 percent cut in delay merely a faster way of getting to the nearest bottleneck, where cars had to wait that much longer? The situation is similar to that often encountered after construction of an overpass: Congestion is transferred to another location.

To be complete, an evaluation study must include losses caused by transitions between timing plans, transitions between test sections operating on different timing plans, and transitions between the test area and outlying areas. In the absence of information on whether delay reductions within the UTCS test area might not have been accompanied by longer delays elsewhere, claims that UTCS and other computer-based control systems can bring an improvement over a well-timed three-dial system should be treated with reserve, just as evaluation studies should be treated with reserve if they apply results obtained from only a certain percentage of the intersections within the system to the entire system.

Can real-time control dispense with historical data? The computer not only has to receive correct information on vehicle arrivals, it must also be able to determine the subsequent action that will bring the desired result; e.g., the least delay throughout the entire system.

Assume the most simple example: A single side street vehicle arrives at a red signal on an arterial. Without additional information as to future side street arrivals, the least delay would be produced by giving this side street vehicle green immediately or as soon as a convenient gap is detected on the arterial. If historical data showed a side street arrival rate of, say, 60 vehicles/h, a different decision would have to be made in order to hold the side street red until the interruption of main street traffic produces no greater delay than the cumulative side street delay. (The example takes no account of stops or stop penalties.) If additional information on side street arrival distribution and main street volume fluctuations could be known, a different strategy would have to be devised; however, this information cannot exist, nor can the effect of each control decision on the remainder of the street system be foreseen. The example can be expanded to the far more complex situations found in systems that comprise more than one side street approach.

Real-time control can measure a variety of parameters, but it cannot, without historical data, predict the number of turning movements; nor can a computer know the delay that turning traffic will produce, because the delay will depend on the turning vehicle's position in the platoon, the gap distribution in opposing traffic, and the presence of pedestrians walking with the green.

From these and more complex examples it may be concluded that real-time control has to be enhanced by historical data, that historical data can produce only a coarse prediction, and that factors needed for truly effective real-time control are not available and cannot be predicted correctly, nor can the effect of the computer's control decisions on subsequent traffic movement be assessed. An attempt to predict future traffic behavior from past detector measurements means trying to predict the unpredictable. Many mathematicians are confident it can be done.

Author's Closure

Todd addresses two issues--one concerns the validity of the UTCS evaluation study and the other is on the use of historical data in real-time control.

Regarding the UTCS evaluation study, I believe that it merits much wider analysis and discussion than has appeared so far in the literature or than I have done in my brief comments. But the issues raised by Todd are not germane to the subject matter of my paper. They should be directed to those who have conducted the evaluation study. Whether Todd's hypotheses are true or false would have no effect on the paper's analysis. In any case, I cannot support them in lieu of scientific evidence.

Concerning the second issue, the use of historical data, I would like to point out that my paper is in the nature of a review and analysis. It does not present any methodological details. Those are described elsewhere (12) and will be included in forthcoming papers. Therefore, any discussion on this topic is merely an expression of opinion.

Historical data have an important role in real-time control, but not in the way they were used in the 2-GC or 3-GC strategies that were implemented in UTCS. A great deal of useful information about future traffic behavior can be derived from detector measurements and effectively used in controlling traffic.

REFERENCES

1. JHK and Associates. An Approach for Selecting Traffic Control Systems. NCHRP, Manual, March 1977.
2. J. MacGowan and I.J. Fullerton. Development and Testing of Advanced Control Strategies in the Urban Traffic Control System. Public Roads, Vol. 43, Nos. 2, 3, and 4, 1979-1980.
3. R.D. Henry, R.A. Ferlis, and J.L. Kay. Evaluation of UTCS Control Strategies--Executive Summary. FHWA, Rept. FHWA-RD-76-149, Aug. 1976.
4. J. Holroyd and D.I. Robertson. Strategies for Area Traffic Control Systems Present and Future. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR569, 1973.
5. Improved Operation of Urban Transportation Systems. Corporation of Metropolitan Toronto, Toronto, Vol. 1, March 1974.
6. Improved Operation of Urban Transportation Systems. Corporation of Metropolitan Toronto, Toronto, Vol. 2, Nov. 1975.
7. Improved Operation of Urban Transportation Systems. Corporation of Metropolitan Toronto, Toronto, Vol. 3, Nov. 1976.
8. W.R. McShane, E.B. Lieberman, and R. Goldblatt. Developing a Predictor for Highly Responsive System-Based Control. TRB, Transportation Research Record 596, 1976, pp. 1-2.
9. D.I. Robertson. Traffic Models and Optimum Strategies: A Review. Proc., International Symposium on Traffic Control Systems, Berkeley, CA, Aug. 1979, pp. 262-288.
10. P.J. Tarnoff. Concepts and Strategies--Urban Street Systems. Proc., International Symposium on Traffic Control Systems, Berkeley, CA, Aug. 1979, pp. 1-12.
11. A.J. Miller. A Computer Control System for Traffic Networks. Proc., 2nd International Symposium on the Theory of Traffic Flow, London, 1963.
12. SCOOT: A Traffic Responsive Method of Coordinating Signals. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Leaflet LF679 Dec. 1977 and Leaflet LF930 March 1981.
13. A.G. Sims. The Sydney Coordinated Adaptive Traffic System. Proc., Engineering Foundation Conference on Research Directions in Computer Control of Urban Traffic Systems, ASCE, New York, 1979, pp. 12-27.
14. N.H. Gartner. Demand-Responsive Decentralized Urban Traffic Control. Office of University Research, U.S. Department of Transportation, Dec. 1980.
15. J.L. Kay, J.C. Allen, and J.M. Bruggeman. Evaluation of First Generation UTCS/BPS Control Strategy. FHWA, Executive Summary, Rept. FHWA-RD-75-26, 1975.

Publication of this paper sponsored by Task Force on Traffic Signal Systems.

The **Transportation Research Board** is an agency of the National Research Council, which serves the National Academy of Sciences and the National Academy of Engineering. The Board's purpose is to stimulate research concerning the nature and performance of transportation systems, to disseminate information that the research produces, and to encourage the application of appropriate research findings. The Board's program is carried out by more than 270 committees, task forces, and panels composed of more than 3300 administrators, engineers, social scientists, attorneys, educators, and others concerned with transportation; they serve without compensation. The program is supported by state transportation and highway departments, the modal administrations of the U.S. Department of Transportation, the Association of American Railroads, the National Highway Traffic Safety Administration, and other organizations and individuals interested in the development of transportation.

The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purpose of furthering knowledge and of advising the federal government. The Council operates in accordance with general policies determined by the Academy under the authority of its Congressional charter, which establishes the Academy as a private, nonprofit, self-governing membership corporation. The Council has been the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine.

The National Academy of Sciences was established in 1863 by Act of Congress as a private, nonprofit, self-governing membership corporation for the furtherance of science and technology, required to advise the federal government upon request within its fields of competence. Under its corporate charter, the Academy established the National Research Council in 1916, the National Academy of Engineering in 1964, and the Institute of Medicine in 1970.