TRANSPORTATION RESEARCH RECORD 597

Traffic Control: Signals and Other Devices

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1976

Transportation Research Record 597
Price $2.60
Edited for TRB by Marianne Cox Wilburn

subject areas
51 highway safety
52 road user characteristics
53 traffic control and operations

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

Notice
The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competence and with regard for appropriate balance.

This report has been reviewed by a group other than the authors 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 this report are those of the authors and do not necessarily reflect the view of the committee, the Transportation Research Board, the National Academy of Sciences, or the sponsors of the project.

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Traffic control, signals and other devices.

(Transportation research record; 597)
1. Electronic traffic controls–Addresses, essays, lectures. 2. Traffic engineering–Addresses, essays, lectures. 3. Traffic safety–Addresses, essays, lectures. I. Title. II. Series.
TE7.H5 no. 597 [TE228] 380.5'08s [625.7'94] 76-30778
ISBN 0-309-02566-4

Sponsorship of the Papers in This Transportation Research Record

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

Control and Communications Section

Committee on Traffic Control Devices
J. R. Doughty, Pennsylvania Department of Transportation, chairman
Zoltan A. Nemeth, Ohio State University, secretary
John L. Barker, Robert L. Bleyl, Frank Carroll, Robert E. Conner, Charles E. Dare, John O. Elstad, Roy D. Fonda, Arthur Freed, Albert L. Godfrey, Sr., Alan T. Gonseth, Robert L. Gordon, John Gray, Rudolph J. Israel, Ken F. Kobetsky, Hugh W. McGee, Robert D. McMillen, Joseph A. Mickes, Juri Raus, James I. Taylor, Francis E. Twiss, Earl C. Williams, Jr., Jason C. Yu

James K. Williams, Transportation Research Board staff

The organizational units and the officers and members are as of December 31, 1975.

Contents

Truck Blockage of Signals

G. F. King, P. Abramson, and C. Duerk, KLD Associates, Inc., Huntington Station, New York

A previously developed and validated model of truck blockage of the line of sight of traffic signals was used to determine the extent of expected blockage as a function of signal location, traffic volume, and composition of approaching traffic. Both parametric analysis and simulation experiments were used. It was found that the relative effectiveness of various possible signal locations in minimizing expected truck blockage varied between one- and two-lane approaches and between the left and right lane of two-lane approaches. The simulation studies covered a wide range of speed, volume, and truck percentage conditions for 10 common two-head signal configurations as well as for 1 single-head configuration. Both horizontal and vertical locations of the individual signal heads were found to have an effect on expected blockage. Increasing volume and increasing truck percentage result in an increase in expected blockage.

A traffic control signal is an information source transmitting on the line of sight. Any interruption of the line of sight between a traffic control signal and a driver adversely affects the timely and accurate reception of the information concerning signal presence and state of the signal. Minimizing the potential of such line of sight interruption is, therefore, a critical element in the design of traffic signal installation. For that reason it received considerable attention as part of National Cooperative Highway Research Program Project 3-23 (1).

A common cause of interruption of the line of sight is the presence of large trucks or buses operating in the traffic stream. The extent of this blockage phenomenon was evaluated by developing an analytical model of the blockage geometry and applying it to highway traffic conditions. The analytical model that has been developed (1) generates a set of blockage curves that are used in the subsequent analysis.

GEOMETRIC PARAMETERS

It is assumed that the intersection approach is a straight and level road throughout the region of interest. In our application, we restrict the length of the approach to 305 m (1000 ft) as measured from the traffic signal position. The lateral position of the traffic signals can be within or beyond the lateral roadway boundaries; signal height is also variable. All trucks and vehicles are assumed to be centered within their respective lanes. Each truck is represented as a rectangular solid in order to simplify the geometrics.

The motorist's eye is assumed to be 1.4 m (4.6 ft) to the right of the left edge of its lane and 1.2 m (4.0 ft) above the pavement. The size of the truck (length, width, and height) and the lane location of the truck can be varied. An average lane width of 3.7 m (12 ft) is assumed; however, this value can be varied. The analysis considers truck blockage for all combinations of truck and vehicle positions. The two basic situations treated are: (a) truck and vehicle in the same lane and (b) truck and vehicle in different lanes. In the first situation, the line of sight blockage is due mainly to the rear profile of the truck; in the other, blockage is caused primarily by the side profile of the truck.

For any given truck-to-signal distance and any vehicle-to-truck separation, a determination can be made of whether the view of the signal from a vehicle is obstructed. A set of curves has been produced that separate the intersection approach into blocked and unblocked regions as a function of the stated specifications. These curves have been designed for use with the Urban Traffic Control System 1 (UTCS-1) traffic simulation model to determine the percentage of time that vehicles are blocked when approaching an intersection. Three general heights were used: 4.9 and 6.1 m (16 and 20 ft) for overhead signals and 2.7 m (9 ft) for post-top mounted signals.

One signal position at a time was analyzed, and multiple signal arrangements were analyzed by combining the results obtained for each signal.

After several trial runs truck height was found to be the most sensitive variable and truck length was found to be the least important. Truck widths do not vary much, and a value of 8 ft (2.4 m) was used throughout.

Because of the insensitivity of truck length, only two truck sizes were selected. These sizes, given in meters, are as follows (1 m = 3.28 ft):

Height	Width	Length
3.0	2.4	10.7
3.8	2.4	10.7

These variable values define the following set of cases that were run:

1. Fourteen lateral signal positions [-8.2 m (-27 ft) and -4.6 m (-15 ft) to +8.2 m (+27 ft) in 0.9-m (0.3-ft) increments];
2. Three signal heights [4.9 and 6.1 m (16 and 20 ft) for all 14 lateral locations and 2.7 m (9 ft) for -8.2, 7.3, and 8.2 m (-27, 24, and 27 ft)]; and
3. Two truck sizes.

PARAMETRIC STUDIES

The effect on truck blockage for the following common signal positions was studied:

1. Far right, post mounted;
2. Far left, post mounted;
3. Center of intersection;
4. Center of each lane, far side of intersection overhead; and
5. Lane line.

Post-mounted signals were assumed to be at a height of 2.7 m (9 ft). Overhead signals were evaluated at two different heights—4.9 and 6.1 m (16 and 20 ft). For this study, the traffic signal is considered to be a point source. The signal positions used are shown in Figure 1.

The results for the single-lane case are shown in Figure 2; the results for the two-lane case are shown in Figures 3 and 4. In each figure, the area to the left of the line corresponding to a given signal position defines the blockage region.

One-Lane Case

Of the five individual signal positions tested, the low-mounted, far left position gave the least blockage; the overhead, center of lane position, at 4.9 m (16 ft), gave the most blockage. The other three signal positions yielded intermediate amounts of blockage with a maximum difference of 8 percent in time of blockage between them.

A traffic stream of 600 vehicles/h/lane at 48 km/h (30 mph) results in a mean space headway of 80.5 m (264 ft). This is equivalent, for the assumed 10.7-m (35-ft) truck, to a vehicle-truck separation of 69.8 m (229 ft). Reference to Figure 2 shows that, at this separation, there is no blockage at any time of the post-

Figure 1. Signal position.

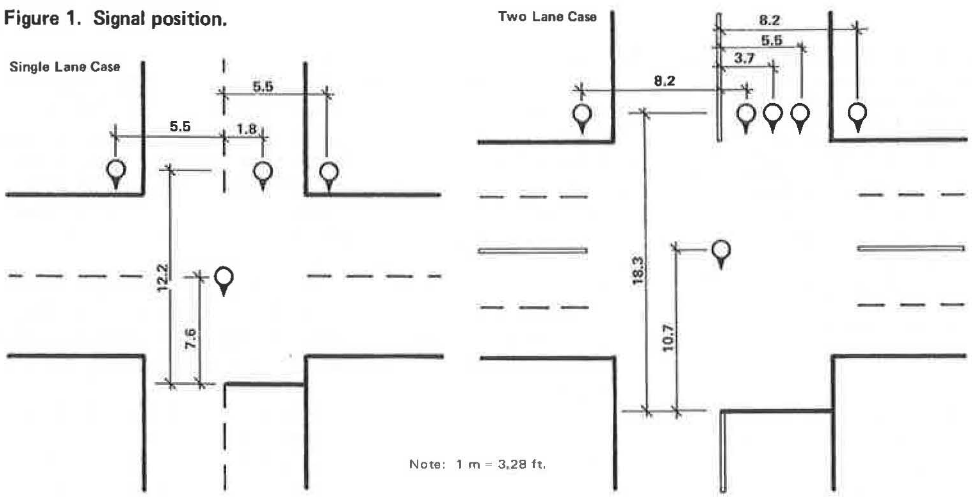
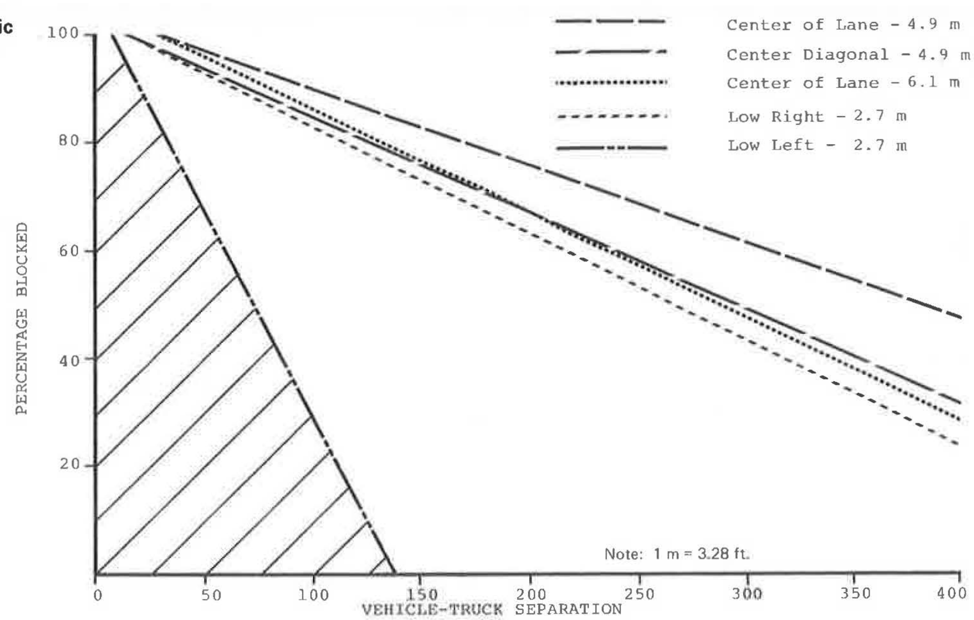


Figure 2. Truck blockage parametric study, single-lane case.



mounted, far left signal head. On the other hand, blockage percentages for the other four positions are as follows (1 m = 3.28 ft):

Position	Blockage
Center of lane at 4.9 m	72
Center diagonal at 4.9 m	62
Center of lane at 6.1 m	62
Post mounted, far right	58

An appreciable amount of blockage (>20 percent) of the far left, post-mounted signal will not occur until the vehicle-truck separation is reduced to 34.1 m (112 ft). This is the mean separation to be expected in a traffic stream of 600 vehicles/h/lane at 27.3 km/h (17 mph) or in a traffic stream of 1080 vehicles/h/lane at 48.2 km/h (30 mph). These values represent D or E level of service according to the Highway Capacity Manual (2). Under these conditions, car-following behavior, rather than signal-observance behavior, is the rule and potential

Figure 3. Truck blockage parametric study, two-lane case (car lane = 2, truck lane = 2).

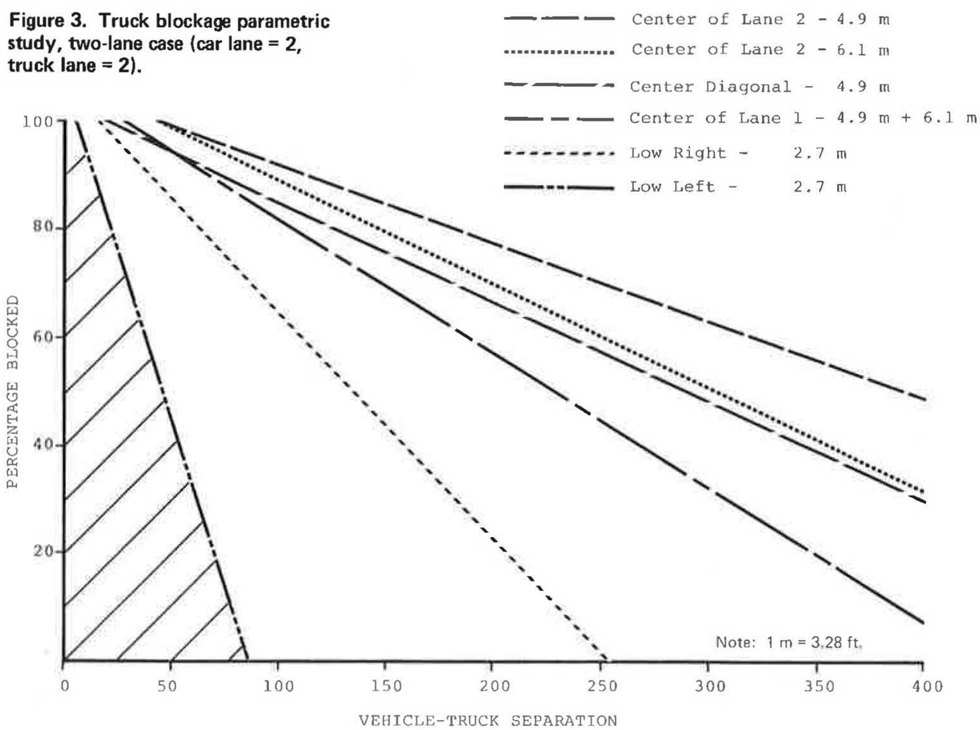
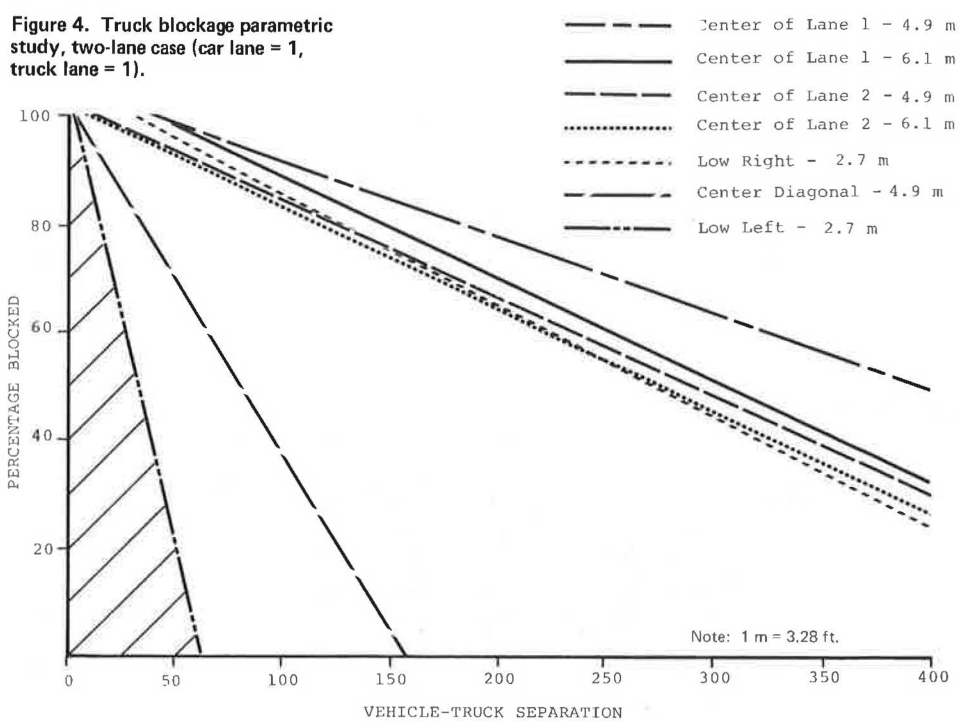


Figure 4. Truck blockage parametric study, two-lane case (car lane = 1, truck lane = 1).



blockage assumes a lesser importance. A similar set of parameters for the other signal positions is given in Table 1.

Two-Lane Case

The two-lane case is presented in somewhat more detail because differences with the single-lane case as well as between the two lanes themselves exist.

A greater number of signal head positions could be considered for the two-lane case. Examination of the graphs shows that, for either lane, same lane position always yields more blockage than opposite lane positions yield. A lane-line signal position is better than a center of lane position for the lane 1 case; however, for lane 2, there is no difference between these positions.

An increase in mounting height in over-the-road signal positions leads to a decrease in blockage percentage,

Table 1. Traffic stream parameters and blockage for single-lane case.

	20% Blockage			50% Blockage			80% Blockage		
Signal Position	Space Headway (m)	Speed for 600 Vehicles/h/lane (km)	Volume at 48 km/h (vehicles/h/lane)	Space Headway (m)	Speed for 600 Vehicles/h/lane (km)	Volume at 48 km/h (vehicles/h/lane)	Space Headway (m)	Speed for 600 Vehicles/h/lane (km)	Volume at 48 km/h (vehicles/h/lane)
Far left, post mounted	44.8	27	1080	32.6	20	1480	20.7	12	2330
Far right, post mounted	138.7	83	348	93.0	56	419	46.3	28	1040
Center of lane, 4.9 m	189.9	114	254	127.7	77	378	62.8	38	769
Center of lane, 6.1 m	146.9	88	329	98.4	59	491	51.2	31	943
Center of intersection	153.3	92	315	100.6	60	480	49.4	31	978

Note: 1 m = 3.28 ft. 1 km = 0.621 mile.

Table 2. Traffic stream parameters at 20 percent blockage for two-lane case.

	Space Headway (m)		Speed for 600 Vehicles/h/lane (km)		Volume at 48 km/h (vehicles/h/lane)	
Signal Position	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
Far left, post mounted	26.5	32.0	16	20	1820	1506
Far right, post mounted	137.8	73.1	83	44	351	661
Center of lane, 4.9 m	186.5	116.1	112	70	259	416
Center of lane, 6.1 m	155.7	116.1	94	70	310	416
Center of opposite lane, 4.9 m	150.0	200.2	90	120	322	241
Center of opposite lane, 6.1 m	142.9	151.5	86	91	338	319
Over lane line, 4.9 m	159.7	200.2	96	120	303	241
Center of intersection, 4.9 m	49.7	149.3	30	90	973	324

Note: 1 m = 3.28 ft. 1 km = 0.621 mile.

Table 3. Traffic stream parameters at 50 percent blockage for two-lane case.

	Space Headway (m)		Speed for 600 Vehicles/h/lane (km)		Volume at 48 km/h (vehicles/h/lane)	
Signal Position	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
Far left, post mounted	20.7	24.4	12	15	2322	1978
Far right, post mounted	93.9	51.2	56	31	514	942
Center of lane, 4.9 m	131.7	79.2	79	47	367	609
Center of lane, 6.1 m	104.2	79.2	63	47	463	609
Center of opposite lane, 4.9 m	100.0	130.1	60	78	485	371
Center of opposite lane, 6.1 m	95.4	103.0	57	62	506	469
Over lane line, 4.9 m	108.2	130.1	65	78	446	371
Center of intersection, 4.9 m	36.0	99.1	22	59	134	488

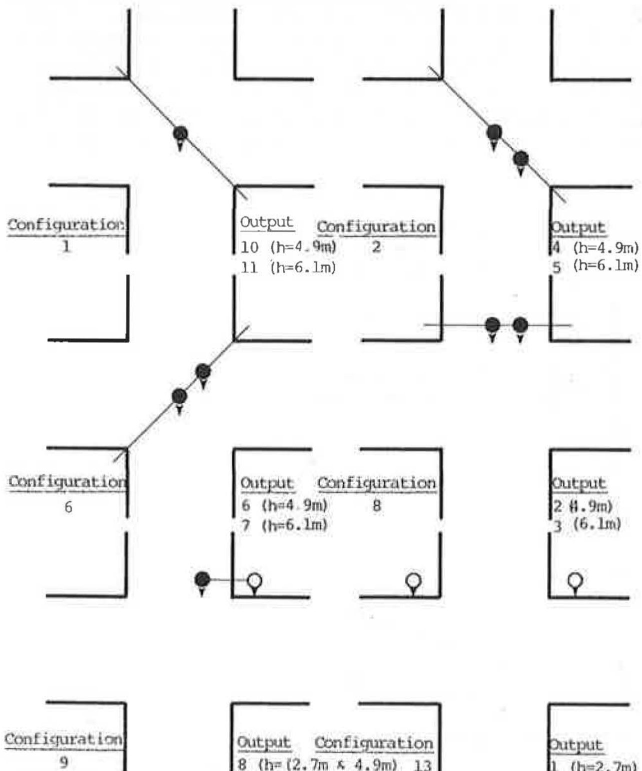
Note: 1 m = 3.28 ft. 1 km = 0.621 mile.

Table 4. Traffic stream parameters at 80 percent blockage for two-lane case.

	Space Headway (m)		Speed for 600 Vehicles/h/lane (km)		Volume at 48 km/h (vehicles/h/lane)	
Signal Position	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
Far left, post mounted	12.2	16.8	7	10	3956	2880
Far right, post mounted	47.2	29.9	28	18	1023	1614
Center of lane, 4.9 m	66.1	43.0	40	26	730	1125
Center of lane, 6.1 m	55.8	43.0	33	26	865	1125
Center of opposite lane, 4.9 m	48.8	65.5	29	39	989	736
Center of opposite lane, 6.1 m	47.2	55.2	28	33	1028	876
Over lane line, 4.9 m	57.9	65.5	35	39	833	736
Center of intersection, 4.9 m	21.6	49.7	15	30	2236	973

Note: 1 m = 3.28 ft. 1 km = 0.621 mile.

Figure 5. Configurations used in blockage analysis.



Note: 1 m = 3.28 ft.

except for the lane 2, opposite lane case where it makes no difference. One major difference between these two lane positions is in the relative efficiency of the far right post mount and overhead center positions. For the lane 1 case, the overhead signal position is to be preferred; for the lane 2 case, the post-mounted signal position dominates.

Tables 2, 3, and 4 give some representative average traffic conditions that will yield the degrees of blockage shown in the graphs.

SIMULATION STUDIES

We determined the expected severity of truck blockage given a defined signal configuration and a specific set of traffic stream parameters (volume, composition, and mean speed) by using simulation.

Simulation Model

The simulation was done by means of the UTCS-1 model (3). The curves generated by the blockage program were incorporated in a subroutine. At each time step, each vehicle is checked for visual blockage of the signal heads.

Five possible conditions are defined.

1. Both signals are visible (condition 0),
2. Right signal is not visible (condition 1),
3. Left signal is not visible (condition 2),
4. Both signals are not visible (condition 3), and
5. Only one signal is visible (condition 1 or 2).

Table 5. Relative blockage percentages for various signal configurations at low approach speed.

		Single Lane Approach						Right Lane Approach[a]						Left Lane Approach[a]					
		0		1		2		0		1		2		0		1		2	
Configuration Number	Mounting Height (m)	Green	Red	Green	Red	Green	Red	Green	Red	Green	Red	Green	Red	Green	Red	Green	Red	Green	Red
13	–	15	21	25	30	61	49	12	22	38	46	50	32	8	11	30	46	62	43
8	4.9	44	58	0	0	56	42	23	39	22	23	55	38	31	46	10	14	59	40
	6.1	38	50	0	0	62	50	21	28	14	20	65	52	31	42	5	8	64	50
2	4.9	39	50	4	6	57	44	18	33	30	35	52	32	35	48	2	6	63	46
	6.1	37	46	1	1	63	53	16	27	27	35	57	38	33	44	1	3	66	53
6	4.9	39	49	5	8	57	43	19	36	31	34	50	30	35	48	4	9	61	43
	6.1	37	48	1	1	62	51	17	30	28	35	55	35	34	46	1	3	65	51
9	4.9	39	50	6	9	56	41	35	49	2	3	63	48	27	47	22	20	51	33
	6.1	39	50	1	1	61	49	35	49	1	1	64	50	26	42	22	23	52	35
1	4.9	39	49	61	51	–	–	34	57	66	43	–	–	35	49	65	51	–	–
	6.1	37	48	63	53	–	–	28	48	72	52	–	–	34	47	66	53	–	–

Notes: 1 m = 3.28 ft.
Numbers in column headings refer to number of signal heads visible.
[a]Two approach lanes.

Table 6. Relative blockage percentages for various signal configurations at high approach speed.

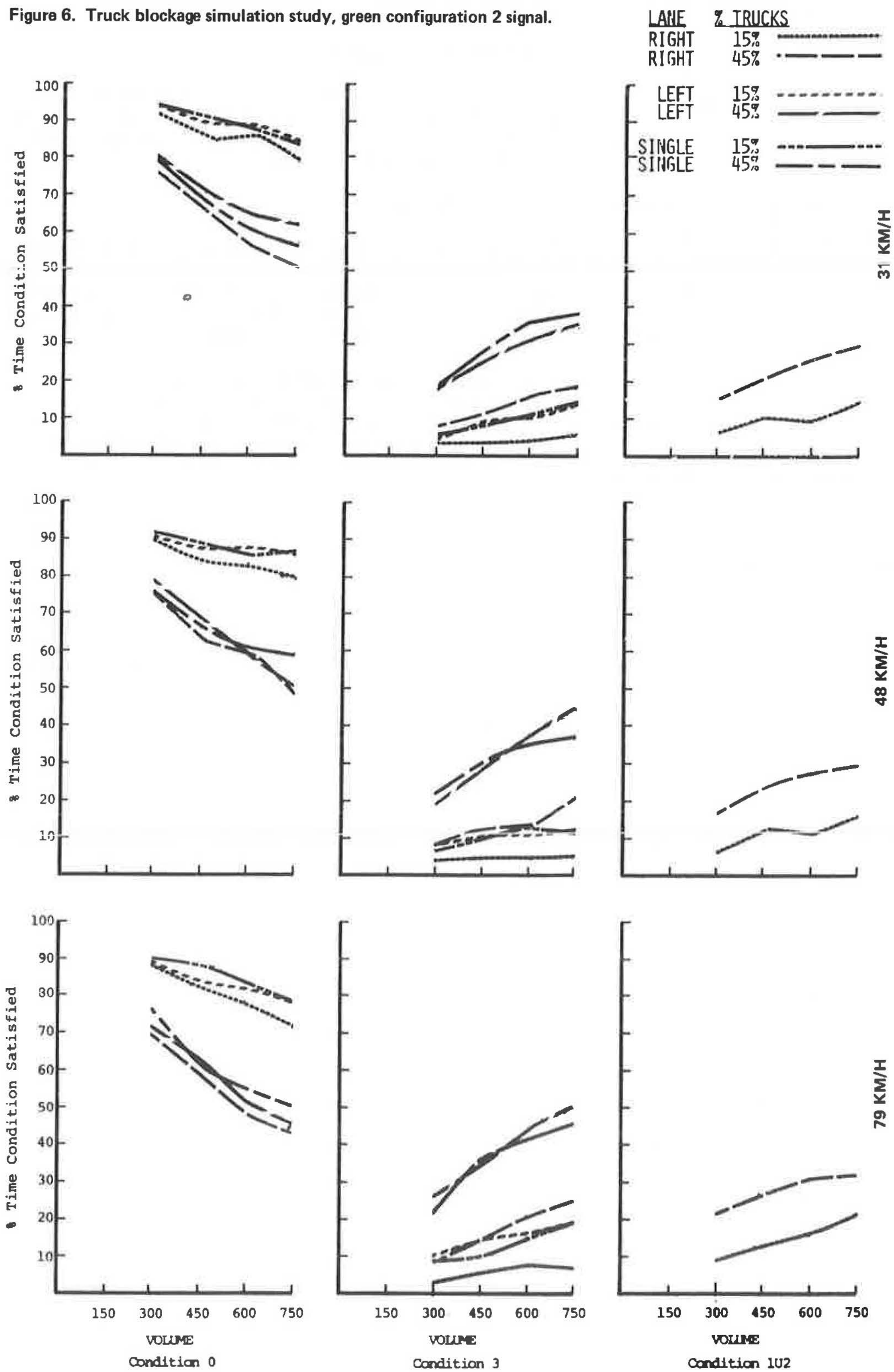
		Single Lane Approach						Right Lane Approach[a]						Left Lane Approach[a]					
		0		1		2		0		1		2		0		1		2	
Configuration Number	Mounting Height (m)	Green	Red	Green	Red	Green	Red	Green	Red	Green	Red	Green	Red	Green	Red	Green	Red	Green	Red
13	–	14	19	31	41	55	40	12	19	38	55	50	26	8	6	30	56	62	38
8	4.9	51	66	0	0	49	34	23	44	22	27	55	29	31	48	10	17	59	35
	6.1	45	59	0	0	55	41	21	34	14	26	65	40	31	45	5	9	64	46
2	4.9	45	59	5	6	50	35	18	36	30	39	52	25	35	54	3	6	62	40
	6.1	43	57	1	1	56	42	16	28	27	39	57	32	33	49	1	3	66	48
6	4.9	45	59	5	7	50	34	19	37	31	39	50	24	35	53	4	8	61	39
	6.1	44	57	1	1	55	42	17	29	28	41	55	30	34	51	1	2	65	47
9	4.9	45	60	6	7	49	33	36	59	2	3	63	38	27	49	22	26	51	25
	6.1	45	59	1	1	54	40	36	59	1	1	63	40	26	44	22	28	52	28
1	4.9	45	59	–	–	55	41	34	53	–	–	66	47	35	54	–	–	65	46
	6.1	44	58	–	–	56	42	28	41	–	–	72	59	34	52	–	–	66	48

Notes: 1 m = 3.28 ft.
Numbers in column headings refer to number of signal heads visible.
[a]Two approach lanes.

If only one signal is present, only conditions 0 and 3 can exist.

At each time step, each vehicle (beginning with the one furthest upstream) is examined to determine the conditions that exist for that vehicle. First, the nearest truck in front of the car in the same lane is determined. If no truck is present, no blockage occurs for that vehicle. If the truck blocks both signals, any truck in the other lane (if it exists) cannot alter the blockage result and no further analysis is needed. If only one signal is blocked, the model finds the first truck in the opposite lane and determines the blockage due to it. Note that trucks themselves are not considered vehicles, and that only the nearest truck in front of the car in the same

Figure 6. Truck blockage simulation study, green configuration 2 signal.

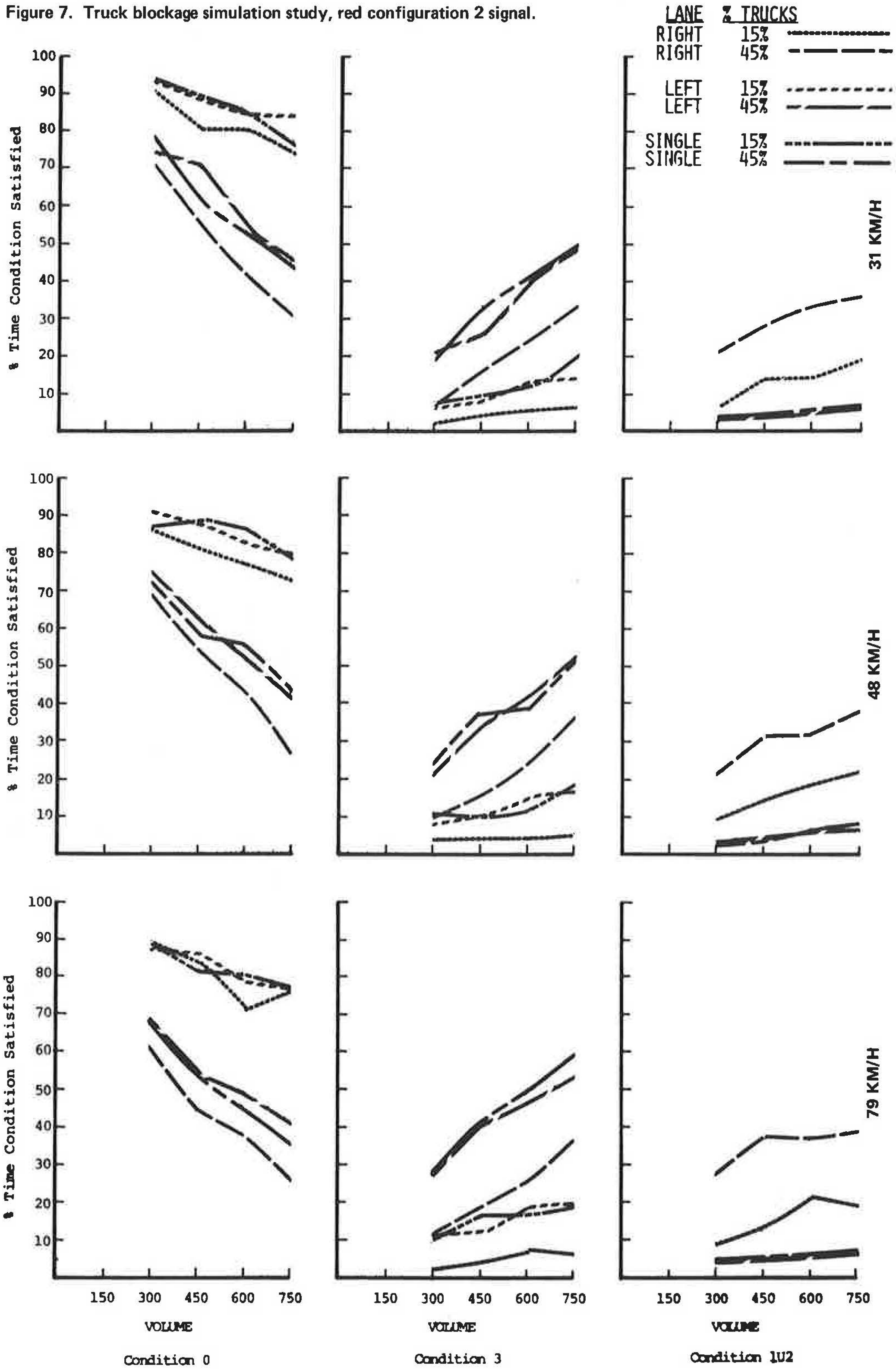


lane is examined, but all trucks in the parallel lane are examined. This analysis is done for all configurations and for each vehicle on the roadway. After the required volume of vehicles is examined, the simulation output yields the accumulated statistics by configuration for the particular case examined.

Selection of Configurations

Eleven signal configurations were examined in detail and are shown in Figure 5. For overhead signals, two mounting heights [4.9 and 6.1 m (16 and 20 ft)] were used.

Figure 7. Truck blockage simulation study, red configuration 2 signal.



Results of Simulation Studies

Figures 6 and 7 show the effect of volume, speed, truck percentage, and signal state on expected amount of truck blockage for configuration 2. Only the results for the lower overhead mounting height are shown. The following general conclusions can be drawn:

1. Truck blockage increases with increasing volumes;
2. Truck blockage increases with increasing truck percentage; and
3. Truck blockage increases with decreasing mean speed.

Truck blockage increases as the mean space headway in the traffic stream decreases, which confirms the same results shown by the parametric studies described earlier.

The simulation study obtained expected blockage percentages for six different common signal configurations. Five of these included overhead-mounted signal heads. These were evaluated at two different heights. Table 5 gives a comparison of these 11 configurations for three lane conditions and a low approach speed. Table 6 gives the same data for high approach speed conditions.

Effect of Configurations

There is relatively little difference among the three configurations that consist of two overhead-mounted signals (configurations 2, 6, and 8). For a single-lane approach, these configurations all showed relatively small percentages of time when only one signal was visible probably because of the small relative lateral displacement of the two signal heads. Consequently, there is very little difference between these configurations and a configuration with a single overhead-mounted signal head (configuration 1).

The two-post-mounted signal (configuration 13) performs best when evaluated on the basis of at least one signal visible. This is due to the excellent performance of the far left signal position as shown in Figures 3 and 4. However, under some conditions, especially in the two-lane cases it performs notably more poorly than some other configurations.

The mixed configuration (configuration 9), in which both signals are in the far right quadrant, shows no appreciable improvement over the all-overhead configurations for the single-lane case. For the two-lane case, there is a considerable difference between the left and right lanes primarily because of the considerable amount of one signal head visibility for the left-lane case.

That the addition of a far left, post-mounted signal head to a two-head overhead or mixed signal configuration would lead to a considerable reduction in truck blockage can be postulated on the basis of this study. A far right high-mounted signal head, although not often used, might even be preferable in preventing cross-traffic and approaching-traffic blockage.

Effect of Mounting Height

Varying the signal mounting height by 1.2 m (4 ft), the maximum variation permitted by the Manual on Uniform Traffic Control Devices (MUTCD) (4) leads to a change in the percentage of signal blockage from 0 to 14 percent. The higher mounting is better whenever a difference exists.

For the three multiple overhead configurations, the difference between mounting heights for any one configuration is greater than the difference between configurations. At their higher mountings, the multiple overhead configurations usually perform marginally better than the two-post configuration for the condition in which two signal heads are visible. However, even at the maximum height, none of the overhead configurations shows as great a percentage as the two-post configuration for the case in which all signals are blocked.

Effect of Approach Speed

Theoretical considerations and the results of the parametric study indicate that the degree of truck blockage to be expected is directly related to the space headway of the truck-vehicle pair. Higher approach speeds, at constant volume, lead to lower densities and, therefore, larger space headways. This basic relationship cannot, however, be equated with the conclusion that blockage is a less severe problem at higher approach speeds.

The amount of blockage to be expected is not a point phenomenon; it must be evaluated over the entire extent of the approach that falls under the influence of signals. This length of approach is not well defined but is definitely speed dependent. For instance, the MUTCD (4) gives a table of minimum sight distances for signals based on 85 percent approach speeds. Translated into travel times, these result in a range of signal viewing times of 3.5 to 8 s. On the other hand, the Traffic Engineering Handbook (5) gives recommended signal head aiming instructions that imply a signal viewing time of approximately 16 s. Therefore, we decided that the simulation study would aggregate blockage over an approach distance equivalent to 10 s at desired mean free speed.

Because of the geometry of the problem (1), the probability that blockage will occur at any point (for a given value of space headway) on the roadway generally increases with the distance of that point from the signal. Because the approach distance increases with specified speed, a far greater prospect for truck blockage for high-speed traffic results because of this factor. Therefore, there are two opposing factors as speed increases: larger space headways and longer approaches corresponding to a constant value of 10-s test period. Examination of the detailed simulation output shows that these two competing factors vary in effect and that the influence of approach speed is not monotonic over all variables. At high volumes, such as 750 vehicles/h/lane, the influence of the configurations incorporating two overhead signal heads generally performed best. If, however, the criterion is changed to require at least one signal head visible, a configuration incorporating a signal head in the far left position has proved to be best.

ACKNOWLEDGMENT

This work was sponsored by the American Association of State Highway and Transportation Officials in cooperation with the Federal Highway Administration and was conducted in the National Cooperative Highway Research Program, which is administered by the Transportation Research Board of the National Research Council.

The opinions and conclusions expressed or implied in the report are those of the authors. They are not necessarily those of the Transportation Research Board, the National Academy of Sciences, the Federal Highway Administration, the American Association of State Highway and Transportation Officials, or of the individual states participating in the National Cooperative Highway Research Program.

REFERENCES

1. Guidelines for Uniformity in Traffic Control Signal

Design Configurations. NCHRP, Project 3-23, interim rept., May 1975.
2. Highway Capacity Manual—1965. HRB, Special Rept. 87, 1965.
3. E. B. Lieberman, R. D. Worrall, and J. M. Bruggeman. Logical Design and Demonstration of UTCS-1 Network Simulation Model. HRB, Highway Research Record 409, 1972, pp. 46-56.
4. Manual on Uniform Traffic Control Devices. Federal Highway Administration, 1971.
5. J. E. Baerwald, ed. Traffic Engineering Handbook. Institute of Traffic Engineers, Arlington, Va., 3rd Ed., 1965.

Stability of Occupancy-Based Library Traffic Control Systems

Daniel J. Mayer, Edwards and Kelcey, Inc., Newark, New Jersey
Leonard G. Shaw, Polytechnic Institute of New York, Brooklyn, New York

The purpose of this paper is to investigate the long-term behavior of a class of closed-loop (feedback) traffic control systems that select an appropriate program from a prestored library on the basis of occupancy information. The subject of traffic control system instability is defined and expounded on. The effects of various factors on traffic control system behavior in general and on stability in particular are explored. Among the factors whose effects are scrutinized are sensor placement, degree of parameter smoothing (system damping), degree of directionality of the control-program library, and the value of threshold and hysteresis used for program selection. Sensor placement in locations sensitive to the formation of queues that require system reaction is advocated in combination with a threshold level composed in its entirety of a hysteresis band. Such a band is shown to provide a good match between traffic conditions and programs handling them and is also shown to reduce or eliminate unstable behavior under certain conditions. The results summarized in this paper were derived from simulating a closed-loop traffic control system operating on a four-intersection corridor. More than 600 simulation runs, each 1 h long (real time), were conducted for various combinations of parameter values. Aggregate delay is used as the measure of effectiveness for comparing these parameter-combination sets.

Recent developments in traffic signal control algorithms have resulted in little demonstrable improvement to measurable parameters of traffic flow quality when compared to older, trustworthy algorithms mainly because the vast majority of research to date consists of the construction and implementation of traffic control programs (set of signal parameters designed to handle a particular static traffic condition) rather than the design of traffic control algorithms (prescribing the control system's behavior for any set of time-varying traffic conditions). More recent work has involved new generations of signal control systems that, by design, use control algorithms within which control programs do not exist as precalculated entities. The lack of success (1, 2) of these algorithms (which should achieve better results than older control systems) is partially attributable to the lack of clear understanding of the behavior of closed-loop traffic controllers. These are controllers that use field sensor data to generate a set of control parameters implemented at the signals being supervised by the system.

This paper is dedicated to the investigation (through simulation) of the long-term (1-h) behavior of a closed-loop traffic control system. This system has a control algorithm that switches among a precalculated set of control programs according to the value of a decision variable calculated from field sensor outputs. The decision variable used here is occupancy (percentage of time a field sensor is covered by vehicles). The simulations are limited to systems that use fixed-time (open-loop) intersection controllers, but similar results are expected for other cases, such as systems with semiactuated intersection controllers. Two different transition algorithms are studied for smooth shifting from one offset program to another when the main algorithm calls for such a change.

Occupancy is extensively used on contemporary traffic control systems. Because this variable is affected both by the control programs implemented and by prevailing traffic conditions, to expect occupancy-based switching to result in control-system instability under a wide range of conditions is reasonable. Simulation studies demonstrated these anticipated oscillations.

Although this paper should yield a firmer basis for the understanding of the behavior of closed-loop traffic control systems and for the design of new generations of control algorithms, parallel studies of volume-plus-occupancy-controlled algorithms are in progress. Qualitatively similar results are expected from those simulations when a significant weight is assigned to occupancy.

TEST CORRIDOR

The corridor selected for demonstrating the long-range closed-loop behavior of the occupancy-based traffic control system consists of the portion of Ashburton Avenue in Yonkers, New York (population 205 000) shown in Figure 1. In Figure 1, arrival rates are in vehicles per hour. Ten percent of the main arrival streams turn away onto the first side street (Nepperhan or Park avenues) in each direction. Side arrival rates refer only to vehicles turning onto Ashburton Avenue (50 percent in each direction). Twenty percent of the main-line flow turn away from Ashburton Avenue at each intersection except at the first intersection. The arrival rates used for most of the simulations depicting unequal flow are 840 vehicles/h in one direction and 540 vehicles/h in the other. The notation used is 540/840. Figure 1 shows equal arrivals of 840/840. Side arrival flows from both directions of each side street were combined for simplicity and are represented in Figure 1 as a T-intersection. One lane of moving traffic in each direction exists on the corridor and on each side street.

CONTROL ALGORITHM

Fixed library control algorithms are examined here. These control algorithms are defined as first generation control software in Urban Traffic Control System (UTCS) terms (3). The selection was made because the vast majority of systems used to date use similar algorithms. Such algorithms consist of a policy that switches among specific, precalculated control programs (the library) in an attempt to fit each implemented program to a set of modified sensor measurements of traffic parameters (such as volume, occupancy, or some combination of the two). Occupancy has been selected as the decision variable in the system investigated here.

Discussion is limited to the corridor previously described with one sensor placed in each direction between St. Joseph and Vineyard avenues. This corridor section, which consists of 4 intersections, was considered small enough so that one sensor in each direction would suffice. The sensors are placed on an internal link of the corridor to reduce the effects of the vehicular arrival process (from outside the study area) on arrivals to the sensed link (modified here by two signals) and because the location is central in the corridor and, unlike an off-central location, has a larger chance of faithfully representing conditions throughout the corridor. The control algorithm used a three-program library made up of:

1. A program favoring one direction of travel,
2. A program favoring both directions equally, and
3. A program favoring the second direction of travel.

Most existing algorithms have the added option of permitting selection of a cycle length; this dimension was not examined in this study. The programs stored in the library are based on offset relationships, the cycle length is fixed at 60 s, and all signal splits are constant.

Three different three-program libraries are considered:

1. Slightly directional with a green-band ratio of 45 to 35 percent of the cycle,
2. Moderately directional with a green-band ratio of 55 to 25 percent of the cycle, and
3. Superdirectional with a green-band ratio of 63 to 17 percent of the cycle.

The program handling average of conditions is common to all three directional-intensity program sets and has a green-band ratio of 40 to 40 percent of the cycle. The directional programs favoring the second direction are symmetric to the ones favoring the first direction. The programs could have been calculated subject to any criterion, but an attempt was made to use the most widely used method: bandwidth maximization. Some attempt is usually made in practice to match the ratio of the incoming volumes (on the corridor) to the ratio of the directional-program bandwidths. This attempt succeeds in matching the ratios for limited time periods only because arrival rates vary with time. One of the main reasons for implementing a closed-loop control system is the unpredictable nature of arrival-rate changes and their times of occurrence.

The control algorithm selected for this investigation causes program switching to occur according to the size of the difference between the smoothed values of the occupancy decision variable in the two directions of travel; the algorithm will switch from a balanced program to a directional program when the size of that difference exceeds a predetermined threshold. The decision variable S used to determine the need for program switching is devised in two steps. First, at the termination of each control interval (one 60-s cycle in this case), the cumulative sensor output during the past control interval is used to update a running average of the sensed parameter. The updating is done as follows:

$$\bar{X}_{n+1} = (AVP \times \bar{X}_n) + [(1 - AVP) \times X_{n+1}] \quad (1)$$

where

$\bar{X}_n$ = smoothed sensor output after control interval n,
X_{n+1} = cumulative sensor output in control interval n + 1, and
AVP = averaging period value.

The AVP value is a fraction between 0 and 1 that defines the number of intervals at which the smoothed sensor output $\bar{X}_n$ has a certain percentage level effect into the future. The following tabulation gives the number of future control intervals in which a sensor reading accumulated during one cycle contributes at least 10 percent of the total smoothed value:

AVP Value (%)	Duration (s)	AVP Value (%)	Duration (s)
0	1	0.6	5
0.1	2	0.7	7
0.2	2	0.8	11
0.3	2	0.9	22
0.4	3	1	00
0.5	4		

Second, the running average of the sensor outputs in each direction is scaled to represent the percentage of maximal value that this parameter can be expected to reach at oversaturated conditions. For example, if the number of seconds that the northbound sensor is occupied in the nth 60-s control interval is X_n^N, the scaled, smoothed occupancy value will be $y_n^N = 100\ (\bar{X}_n^N/30)$, and program switching will depend on the difference $(y_n^S - y_n^N) = S_n$. This normalization uses 30 s/min as the full-occupancy value, which allows the rare possibility of greater than 100 percent normalized occupancy.

When S exceeds a preset threshold level, a program favoring the direction with the larger smoothed decision variable is instituted. If S does not exceed this threshold level, the average-conditions offset program is used. A hysteresis band is commonly used in control systems with this type of algorithm to prevent program-selection oscillation when the value of S hovers near the threshold level for a period of time. For example, a directional program may be invoked when the threshold level is exceeded by S, but the balanced program is returned to only if S drops below the threshold level minus the hysteresis band (Figure 2). (No program switching is allowed during program transition following a previous threshold passage.)

DEFINITION OF SYSTEM STABILITY

Under certain conditions, a fixed library, closed-loop traffic control system may switch repeatedly among control programs. If such repeated switching takes place under constant vehicular arrival rates at the termini of the controlled street network, the control system might be said to be unstable. More specifically, a traffic signal control system will be defined as unstable if a set of feasible, constant vehicle-arrival rates (one value for each network arrival terminus) into the controlled street network exist that cause the control system to vary its implementation of control programs without settling at the steady-state program most appropriate for that set of constant vehicle-arrival rates. The term "most appropriate" is defined here in terms of the criterion on which the control algorithm is based. This definition of system stability holds for any type of traffic signal control system except ones in which the most appropriate control policy for certain sets of arrival rates consists of oscillation between two or more control programs (4).

EXISTENCE OF SYSTEM INSTABILITY

Control systems that use volume as the decision variable might experience instability when a temporary blockage occurs that affects flow over a sensor location. This instability manifests itself by the selection of a control program that is not the most appropriate for the field conditions. This is the major reason that most manufacturers of signal control equipment avoid the use of volume as the sole decision variable; some manufacturers use a linear combination of volume and occupancy. Such a combination may result in instability similar to that experienced by the occupancy-based system studied here but with a different range of parameters under which such instability occurs.

When occupancy is used as the decision variable and the flow of vehicles into the corridor is such that the difference S between the two directions is below the preset threshold level, the system can be expected to reach a steady state at the average traffic conditions program subject to the influences of other stability-affecting parameters that will be described. When vehicular inputs or other conditions are such that S exceeds the threshold level indicating preferential offset, the system might oscillate in its selection of the control programs without reaching a steady state. This potential instability occurs as follows (Figure 3):

1. The system selects a preferential offset as a result of a threshold passage;
2. The preferential offset helps to reduce S because it reduces occupancy in the preferred direction at the expense of increasing occupancy in the opposite direction; and
3. When S is less than threshold minus hysteresis, the average traffic conditions program is implemented until S exceeds the threshold level once again, and the process is repeated.

BASIC MECHANISMS AFFECTING STABILITY

The basic mechanisms affecting control system stability may be divided into three groups:

1. Traffic-flow conditions,
2. Physical attributes, and
3. Control-algorithm parameters.

The control system designer has little or no control over traffic-flow conditions. One way in which ω may rise above the switching threshold is simply by unequal vehicular-input flows. A second way involves the occurrence of a blockage downstream from a sensor location. Such a blockage, when sufficiently close to affect the flow of traffic over the sensor, might start an instability process when the smoothed occupancy in the blocked direction increases sufficiently to cause S to cross the threshold level. A third way involves the occurrence of a blockage upstream from a sensor location; such a blockage can reduce the smoothed-occupancy parameter in the blocked direction sufficiently to push S over the threshold level, favoring flow in the opposite direction.

Most physical attributes are beyond the control of the control system designer. For example, the physical layout of the corridor, the number of lanes in each direction, and the length of the links between intersections are generally fixed. These physical features play a major role in the preparation of the control program library. One major physical attribute controllable by the system designer is the location of the sensors in which data regarding the decision variable originate.

Control-algorithm parameters that are under the control of the system designer and might affect system stability are

1. Averaging period,
2. Threshold level and hysteresis band,
3. Directional intensity of control programs, and
4. Program transition.

SIMULATION PROGRAM

A microscopic program was designed to have the following characteristics:

1. Vehicular-clearance signal phases assumed to be negligible in their effect on system behavior;
2. Initial delay of 4 s before the first vehicle moves at the onset of green with 2-s headways thereafter [3.7 and 2.1 s respectively, based on the numbers given by Greenshields, Schapiro, and Ericksen (5)];
3. Constant headway of 2 s in saturation flow regardless of velocity;
4. Free flow urban velocity of 48 km/h (30 mph); and
5. Shock-wave negative velocity of 24 km/h (15 mph).

The simulation is detailed enough to supply data on vehicular passage over sensor locations so that the proper decision variable values may become available to the control algorithm. This requirement necessitated a microsimulation model in which shock waves are propagated by iteratively tracing individual vehicles through the corridor. The simulation is time based, and individual vehicles are advanced starting at the upstream end of the system during each 1-s simulation interval. The need for long-term observation of closed-loop system behavior dictated an efficient simulation program operating on a street network small enough to yield a high real-to-simulated time ratio. Each simulation run of 3 min corresponds to 1 simulated hour over a two-directional corridor with four signalized intersections (on an IBM 370/135).

The simulation program is a derivative of one described by Longley (6). This simulation differs from that of Longley in the method of tracing vehicles through the network, in the intersection model, and in the sequence of vehicular propagation. Details of the simulation are available elsewhere (7).

Simulation runs conducted to investigate the behavior of the control algorithm acting on unbalanced directional traffic flows were initialized by 20 min of balanced flow before each 60-min run having unbalanced flow. This step type of perturbation to which the system was subjected at the beginning of each run had a twofold purpose:

1. To investigate the effect of such a step type of input on system instability and its duration and
2. To provide information on the quickness of system response to flow changes.

Corridors simulated with balanced arrival flows were initialized in one of the following two methods:

1. The same balanced flow as was provided during the actual simulation period (usually 840/840) with a 4-min preferential control-program perturbation immediately following this initialization period or
2. Unbalanced flow (usually 540/840) for the 20-min duration of initialization followed by the actual 60-min simulation run at the balanced flow (usually 840/840).

Simulations were carried out with both deterministic flows having the indicated rates and stochastic arrivals whose mean rates matched the indicated rates.

SIMULATION RESULTS

Figures 4 and 5 show typical simulation results for the oscillation of S across threshold levels and the corresponding switching among control programs. The following sections present interpretations of these results.

Smoothing

Insufficient smoothing of the decision variable is a major cause of control system instability. Figure 4 shows the behavior of an insufficiently smoothed (AVP = 0.4) system under constant, equal, regular (nonstochastic) saturated arrival rates (1000 vehicles/h in each direction). Note that this simulated system, with splits of 63 percent at each intersection (excluding start-up delay), carries its maximal load (at corridor termini) at about 1000 vehicles/h. Figure 5 shows a sufficiently smoothed (damped) system behavior under the same saturated conditions at a larger AVP value of 0.6. It would seem that, for saturated conditions at the given system parameters, an AVP value of at least 0.6 is called for to prevent unwanted (in this case) control-program oscillations. Results in this case indicate that aggregate delay is reduced when the algorithm parameters are adjusted to eliminate oscillations.

The simulation of systems with stochastic arrivals (in this case with an average of 800 vehicles/h in each direction) yielded the same general results. The system stabilized with an appropriate AVP value usually slightly higher than the one sufficient to stabilize an equivalent deterministically generated arrivals system with the same arrival rates (balanced).

At unbalanced vehicular arrival rates, instability may occur subject to threshold, smoothing parameter, directional flow difference, and the like. Where a system is shown to be unstable for a particular set of unbalanced arrival rates, the smoothing parameter AVP has an effect on the period of control system oscillations and, sometimes, on the existence of such oscillations. At arrival rates of 600/1000 (one direction saturated) and an AVP value of 0.5, the average control-program oscillation period is 5 min. As the AVP lengthens to 0.7, the average oscillation period increases to 5.4 min. At the long AVP of 0.9, the average control-program oscillation grows to 7.6 min. Figure 6 summarizes the relationship between average oscillation period and AVP values with stochastic arrivals of 400/840. The oscillation period generally increases with the AVP value and with approach to an infinite period, which indicates that oscillation disappears at high AVP values.

Threshold and Hysteresis

The effect of the threshold level on system behavior and stability varies with the values assigned to the smoothing parameter, sensor location, control-program directional intensity, and size of the hysteresis band modifying that threshold level. In addition, the threshold level strongly influences the measure of effectiveness (aggregate delay) value derived from the simulations. The threshold should be set in a manner that causes control-program transition when the directional program invoked by this action is better equipped to handle the then-current field conditions. Hysteresis may play a minor or a major part in the threshold mechanism and have a strong effect on system oscillations.

Hysteresis as a Minor Threshold Component

It is clear that the smaller the value of threshold used is the larger is the AVP value required to avoid unnecessary oscillations in control-program implementation. For a system simulated with stochastic inputs, statistical variations in S are sufficiently large to require larger stabilizing AVP values than those needed for systems simulated with deterministic inputs. These larger AVP values damp the system to a point where transient response in reaction to a step input change in arrival rates is almost nonexistent. The critical AVP value (selected from the range of 0.1, 0.2, . . . , 1) changes the duration of transient behavior from greater than 60 min to insignificant values for the case of 800/800 stochastic arrival rates. The critical values of AVP for various threshold levels are shown in Figure 7.

Hysteresis as Major Threshold Component

A major role may be assigned to the hysteresis band in order to

1. Cause the system to match the proper program to traffic-flow conditions and keep this program in operation until the need for it disappears and
2. Reduce oscillations to enable operations at a relatively short averaging period for faster system reaction to flow changes.

Both of these goals can be achieved by putting the entire switching-definition burden on the hysteresis band that is set at the full program-switching threshold level. This hysteresis band setting causes program switching away from the balanced program, as usual, at threshold and program switching away from the directional program only when S returns to cross the 0 percent mark. In this manner, the system is allowed to operate at a directional program if that program tends to equalize occupancy in both directions, which is a desirable trait. The previous arrangement of a threshold level with a small hysteresis band ensured system instability if the directional program was sufficiently matched to the traffic flow to equalize occupancy in both directions (Figure 8). The algorithm operating with a large hysteresis band does not abandon the directional program if it is successful in equalizing occupancy in both corridor directions and is effective in terms of lowering aggregate delay as well (Figure 9). Note that this new arrangement enables the system to operate at low damping (short AVP), insuring fast system response without excessive oscillations.

Figure 1. Ashburton Avenue with simulated arrival rates shown.

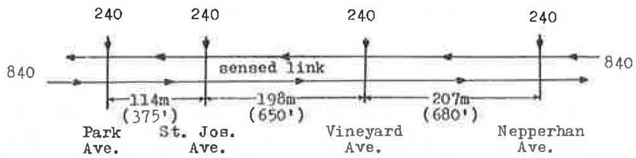


Figure 2. Switching rule.

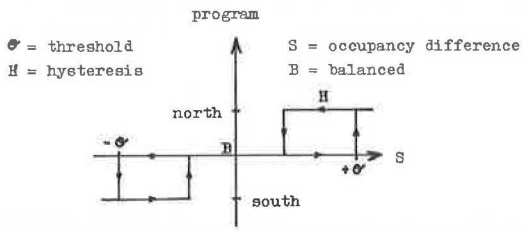


Figure 3. Oscillation at unbalanced flows.

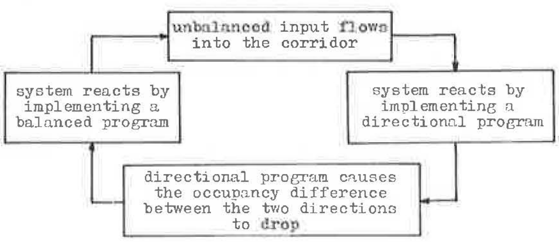


Figure 4. Instability: insufficient damping at balanced, saturated flows for deterministic arrivals.

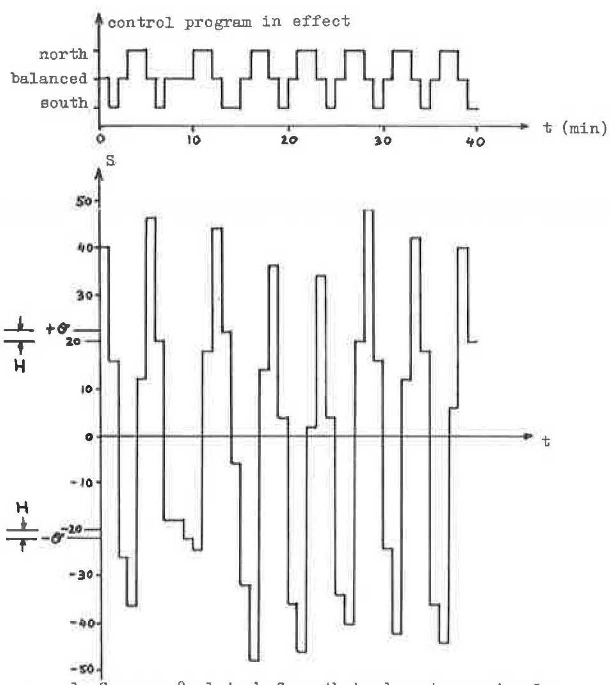


Figure 5. Stability: sufficient damping at balanced, saturated flows for deterministic arrivals.

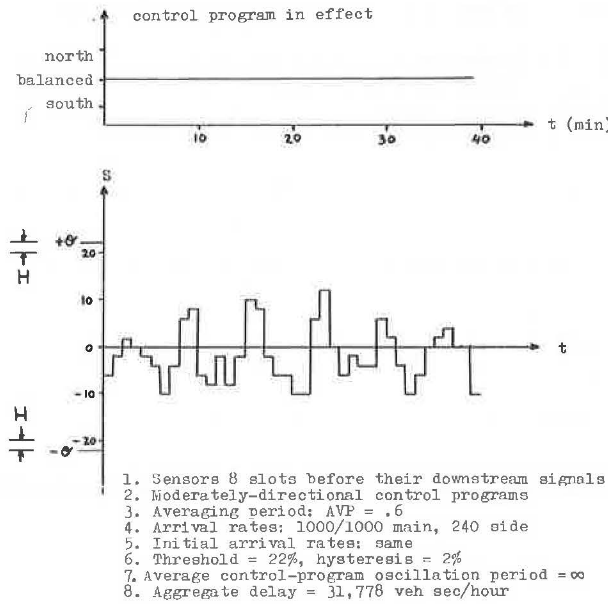


Figure 6. Average oscillation period versus AVP.

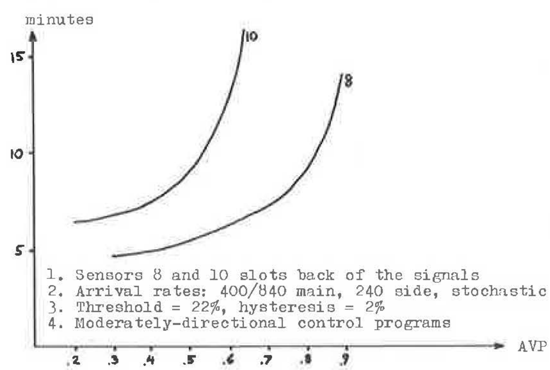
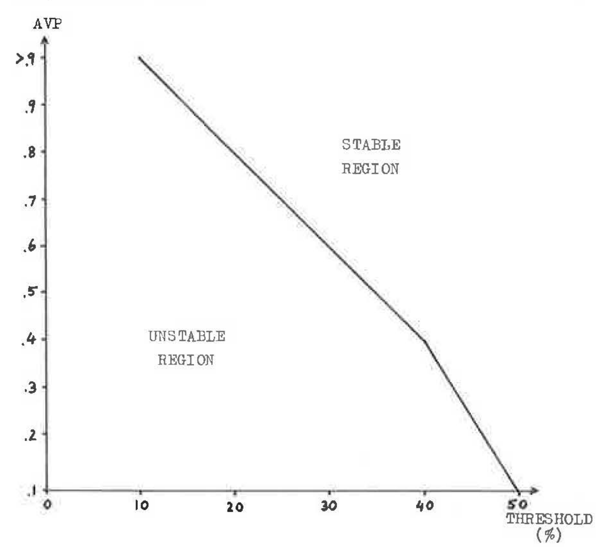


Figure 7. Critical AVP values versus threshold level for stochastic 800/800 arrival rates.



Ideally, each program (balanced or directional) should have a threshold and a band that enable switching away from it in the same manner usually provided for in the balanced program only. Hence switching away from a directional program should not occur at the 0 percent threshold (as used here) but within a "forgiveness" range of a percentage away from 0 percent as in switching away from the balanced program. This feature should be incorporated into future systems; however, this study is addressed to the improvements that can be made in using existing equipment; therefore, the limit to threshold and hysteresis operation was set at a large hysteresis band equal to but not larger than the threshold level.

Control-Program Directional Efficiency

Three control-program sets were compared:

1. The slightly directional program with a green bandwidth ratio of 45 percent/35 percent,
2. The moderately directional program with a green bandwidth ratio of 55 percent/25 percent, and
3. The superdirectional program with a green bandwidth ratio of 63 percent/17 percent.

For each program set, a correlation was perceived for aggregate delay at an AVP value, the number of program transitions, and the suitability of the directional intensity

Figure 8. Usual hysteresis band (2 percent) causes small oscillation period at AVP = 0.4 for stochastic arrivals.

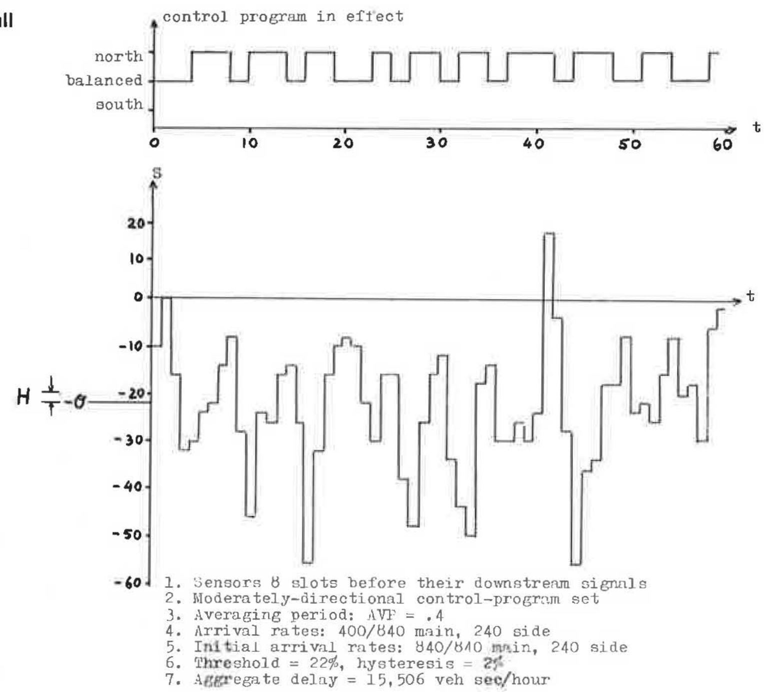
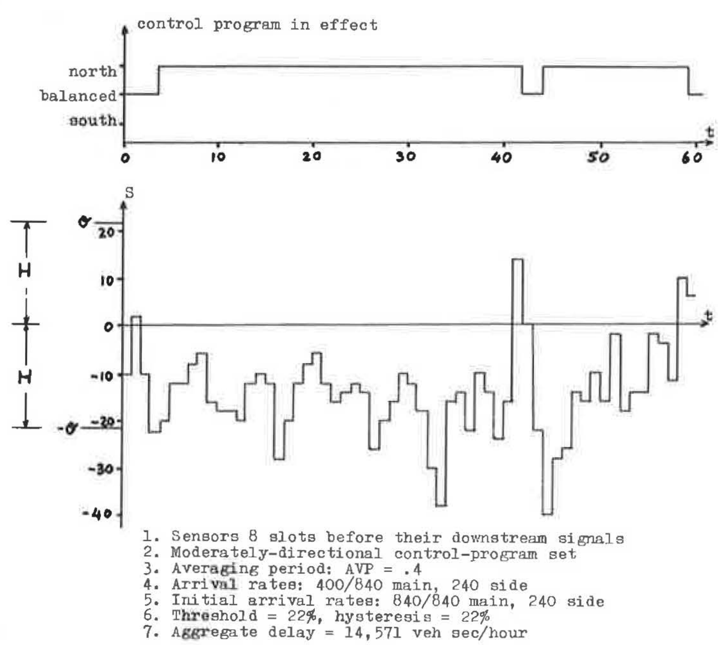


Figure 9. Full hysteresis band (22 percent) enables large oscillation period at AVP = 0.4 for stochastic arrivals.



of the program set to handle the actual flow conditions on the corridor. For example, in general, the fewer program transitions there were, the less the aggregate delay would be. If two different directional intensity program sets exhibit the same rate of program transitions, then the suitability of the directional program influences the delays experienced.

Figure 10 shows the typical behavior of the system through one control-algorithm excursion beyond the threshold in the smoothed-occupancy plane. The axes represent the smoothed value of occupancy (in seconds per minute) in each direction. Successive breakpoints have coordinates (y_n^N, y_n^S). It is interesting to note that the more directional a program set is the faster its points move in this plane and the more nearly orthogonal are its trajectory and the switching lines.

Figure 10. Typical paths in the smoothed-occupancy plane through one excursion into directional-program use.

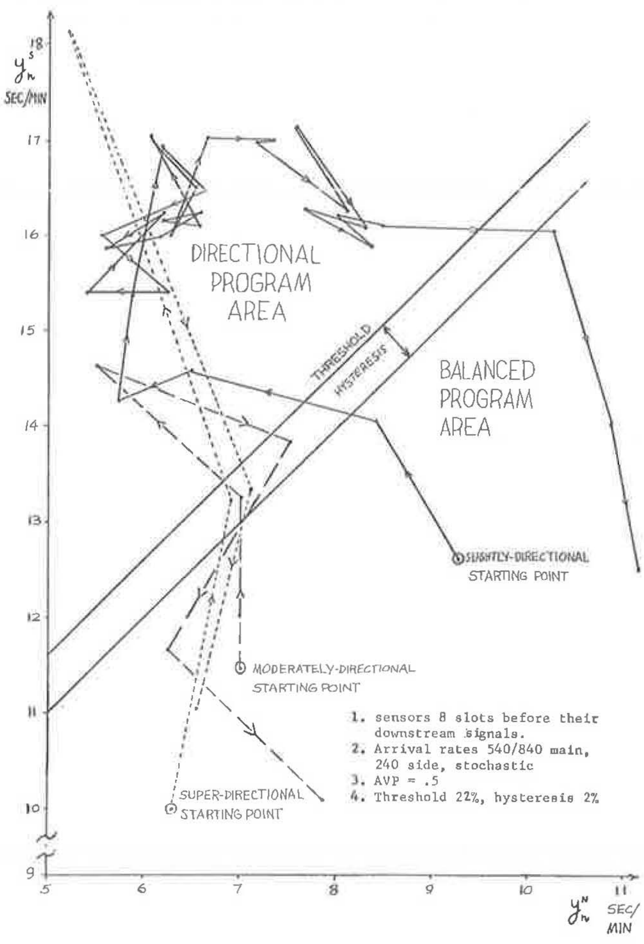
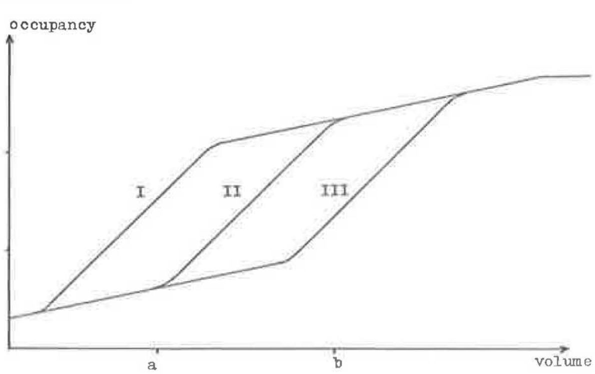


Figure 11. Sensitivity of occupancy to volume variations for three sensor locations.



a = volume below which the queue never reaches sensor II
b = volume above which the queue always reaches sensor II
c–d = good range for threshold level fixing

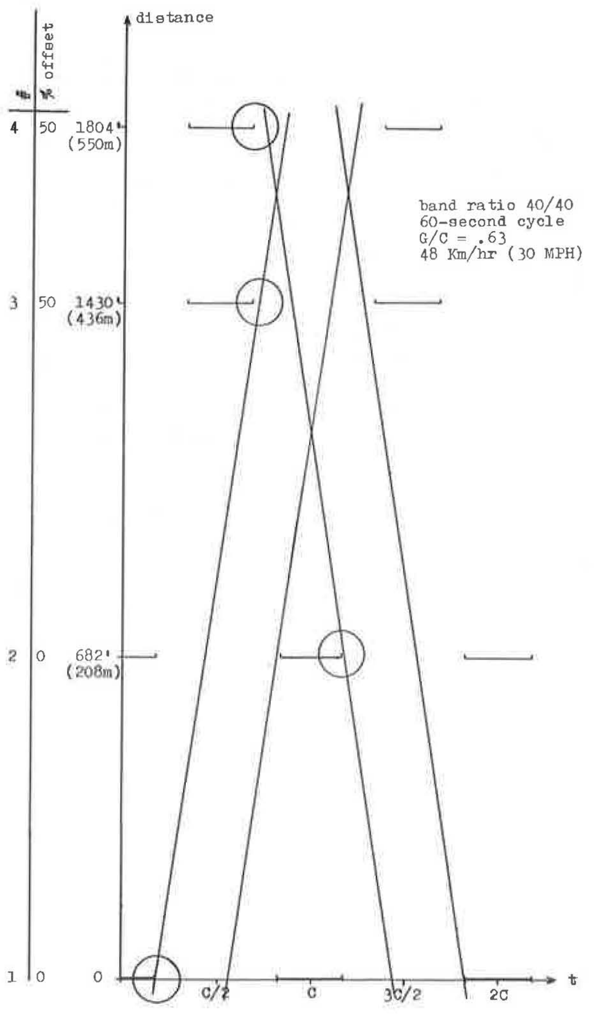
Sensor Location

Ideally, the sensors in occupancy-based traffic control systems should be placed to enable the control algorithm to detect and reduce exessive queues accumulating at the intersections immediately downstream of each sensor, thereby attempting a reduction in delay and number of stops.

Placing each sensor far from its downstream signal tends to yield useful results in the sense of vehicular volume detection. However, such placement does not yield data on accumulated queues within a useful range. The time needed for such queues to reach the length necessary to significantly affect the smoothed occupancy parameter slows down the response of the system. Conversely, placing each sensor too close to its downstream intersection is meaningless because a minimal queue is to be expected most of the time, and sensitivity to queue variations would be low. Two sensors placed on one link can help in accurately determining the length of a queue. However, in the context of the traffic control algorithm discussed here, the selection of one sensor location (on the sensed link) at an appropriate distance from its downstream intersection should suffice for the provision of proper control data. After numerous trial simulations, the sensor locations were selected to be eight vehicle slots behind their respective signals on the central link of the four-intersection corridor [1 vehicle slot = 6.7 m (22 ft)]. Signal-system manufacturers generally recommend locating the sensors as far away from probable queues as possible yet at locations most likely to reflect a change in flow necessitating a change in program. This ambivalence is a result of a combination of the need for fast response and apprehension about queues reaching the sensors. However, it is precisely the placement of a sensor in an area that may be affected by queuing vehicles at certain arrival rates that allows differences in flow conditions between the two corridor directions to be easily detected. The location of a sensor in a spot sensitive to certain queues ensures program switching when the current control program does not handle a queuing problem adequately. Thus a better chance for matching the control program to the flow conditions exists. A directional program need not be invoked when flow conditions are not equal in the two corridor directions as long as the control program in effect is adequate for handling these conditions so that no excessive queues occur. Sensor location should be selected so that the queue affects the occupancy parameter to provide a program transition when volume in the dominant direction is so large that the balanced program cannot handle it well. Figure 11 shows qualitative sensitivity curves of the occupancy parameter in relation to directional volume for three sensor locations. The first sensor is closest to its downstream intersection, and the third is the farthest away.

Symmetric placement of sensors in both corridor directions in relation to their downstream intersections is advocated. Potentially bad results of gross asymmetric sensor placement occur at balanced flows because a sensor in one direction is occupied by a signal-caused queue more often than the sensor located in the opposite corridor direction is, thus forcing the system into a directional program that attempts to alleviate the occupancy discrepancy and sends the system back to a

Figure 12. Average traffic program causes queue bias.



balanced program, which starts the process anew.

Queue Bias

Care should be exercised in calculating the control programs used. The use of band-maximizing techniques would generally result in the creation of a queue bias on the sensed link even though the sensors are placed symmetrically with respect to their downstream intersections. The bias problem may be serious enough to affect the behavior of the control systems, because the threshold level is modified by queue bias. Figure 12 shows the occurrence of queue bias due to the locations of the beginning and the end of each green phase in relation to the through band. The resulting simulated queue bias is shown in Figure 13 in terms of an occupancy-difference bias.

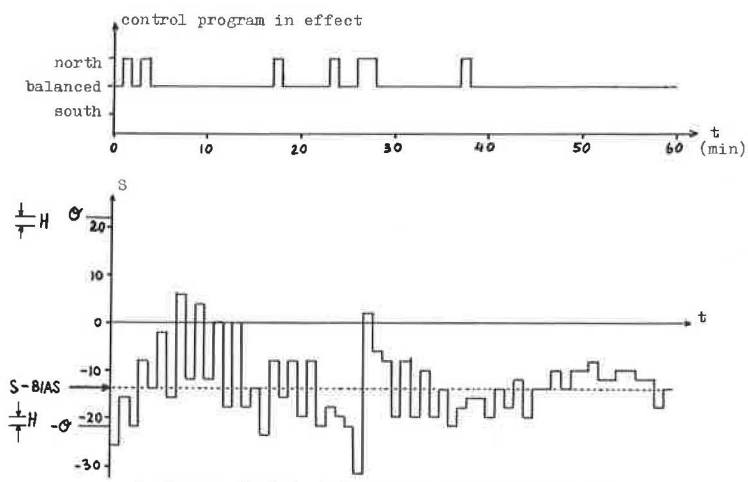
The circles drawn in Figure 12 point out the critical points causing unequal queues on the sensed link between intersections 2 and 3. The average queue at intersection 2 (going from 3 to 2) is larger than the average queue at intersection 3 going in the opposite direction. Hence, a larger occupancy value is derived, on the average, in direction 2 (intersection 3 to intersection 2). The green band is limited at the beginning of the platoon in direction 2 at intersection 2, although in direction 1 the leading edge of the platoon is limited at intersection 1 and intersection 3 allows for some lead time to clear at least a portion of the queue consisting of side-turning traffic and main-traffic residue, before the main platoon arrives from intersection 1. The arrival rates of 840/840 are sufficiently large to cause significantly larger queues or shock waves or both over sensor 2 than would be caused over sensor 1.

Queue bias was corrected in some of the simulations by incorporating a corresponding shift in the decision thresholds. An alternative approach would be to change the method of calculation of the control programs in order to get equal queues when arrival rates are equal. This could be achieved by shifting the offsets derived from band-maximizing methods to equate queues in both directions on the sensed link.

Control-Program Transition

Two program transitions were tested for their effect on

Figure 13. Negative bias in the S parameter and its effect on system behavior by modifying the threshold level for deterministic arrivals.



1. Sensors 8 slots before their downstream signals
2. Mildly-directional control-program set
3. Averaging period: AVP = .5
4. Arrival rates: 840/840 main, 300 side
5. Initial arrival rates: 540/840 main, 300 side
6. Threshold = 22%, hysteresis = 2%
7. Aggregate delay = 24,420 veh sec/hour

systems stability. The dwell type of transition (displaying main street green at each intersection until transition there is complete) was compared to the immediate type of transition (transferring immediate control to the new program subject to minimal phase durations). No significant difference in system behavior was detected, but the aggregate delay in the case of immediate transition is generally slightly lower, even though the program oscillation frequency is generally higher, and more transitions occur when this type of transition is used because no new transitions are initiated while a transition is in progress.

SUMMARY

Instability, as evidenced by unnecessary switching among offset programs in a library, does indeed occur for various combinations of control parameters. These instabilities occur mainly because of the development of unequal queues and shock waves on the sensed links. When these differences result in an occupancy difference that surpasses the threshold level, oscillation in control-program implementation occurs if the directional program invoked is sufficient to return the decision variable to a level lower than the hysteresis-modified threshold. These results bear out the general advice given by system manufacturers who recommend placing the sensors as far away from downstream intersections or from expected queues as possible. The manufacturers' recommendations for the use of occupancy as the decision variable historically evolved from the occurrence of saturated conditions or road blockages, which yielded false volume reading. By switching to occupancy as the decision variable, the equipment manufacturers were trying to improve the reliability of sensor information in relation to actual field conditions. However, the case in which occupancy should yield better results than volume is precisely the one in which instability will potentially occur. Cases in point are those of saturation or road blockages that occur in the vicinity of sensor locations. Similar results are anticipated from analogous studies of volume plus occupancy control now in progress.

The use of a threshold level with a large hysteresis band is strongly recommended whenever occupancy sensors are used to their full potential in detecting queues. Such a setting ensures that a directional program will remain in effect through most of its useful range rather than having the balanced control program take over at an inappropriate time, starting an instability cycle. The technique used in this study may be used to simulate other control algorithms, the feasibility of which in actual use is being investigated.

That the vast majority of previous simulations were used to test control programs on specific, static traffic-flow conditions rather than complete control algorithms is surprising. This situation will undoubtedly change. However, efficient simulation programs are necessary to meet this end because such an undertaking requires the observation of control-algorithm operation over long periods of time with a range of parameter settings.

ACKNOWLEDGMENTS

This paper is based on a portion of Daniel J. Mayer's dissertation, which was submitted to the faculty of the Polytechnic Institute of New York in partial fulfillment of the requirements for the degree of doctor of philosophy in systems engineering (7). Leonard G. Shaw, the coauthor, was the major advisor.

We wish to thank Kenneth W. Crowley and William R. McShane of the Department of Transportation Planning and Engineering for their numerous constructive suggestions.

REFERENCES

1. M. Koshi. On-Line Feedback Control of Offsets for Area Control of Traffic. In Traffic Flow and Transportation, Newell (ed.), Elsevier, New York, 1972.
2. J. Holroyd and D. I. Robertson. Strategies for Area Traffic Control Systems: Present and Future. V.K. Transport and Road Research Laboratory, England, Rept. LR 569, 1973.
3. Urban Traffic Control and Bus Priority System. Sperry Rand Corp., Washington, D.C., Rept. FHWA-RD-73-9, 1972.
4. D. C. Gazis. Optimum Control of a System of Oversaturated Intersection. Operations Research, No. 12, 1964.
5. B. D. Greenshields, D. Schapiro, and E. L. Ericksen. Traffic Performance at Urban Street Intersections. Bureau of Highway Traffic, Yale Univ., Technical Rept. 1, 1947.
6. D. Longley. A Simulation Study of a Traffic Network Control Scheme. Transportation Research, No. 5, 1971.
7. D. J. Mayer. Some Aspects of Closed-Loop Traffic-Signal Control Systems. Polytechnic Institute of New York, Brooklyn, June 1975.

Some Results on Guidelines for Treatment of Traffic Congestion on Street Networks

William R. McShane and Louis J. Pignataro, Department of Transportation Planning and Engineering, Polytechnic Institute of New York, Brooklyn

The guidelines reported in this paper are intended for use by the practicing traffic engineer. The concept of the guidelines, the recommended approach, and a framework for addressing a particular problem are outlined. The framework includes both preliminary identification of the cause of the problem and categories of the treatments available. A sequence of treatment classes is recommended. Proper signalization and provision of added space (bays, lanes, and the like) are of prime importance. Some of the important results or recommended actions are presented in individual sections within the paper—shorter cycle length to avoid spillback, equity offsets to compensate when oversaturation occurs, and a number of nonsignal considerations. Some of the flow charts to aid in decision making are shown.

The problem of traffic congestion, traffic saturation, and traffic oversaturation presents traffic engineers with one of their most difficult tasks. NCHRP Project 3-18(2) (1) addressed this topic. As part of that project, a document on guidelines was prepared. This paper reports on the structure of those guidelines and on some relevant specific results contained therein.

The guidelines are intended to aid traffic engineers in executing their duty by reminding them of available options, by uncovering some subtleties that can be overlooked, and by presenting quantitative insight into the relative benefits of various options. In this way, an appreciation can be obtained of when various options are effective or necessitated or both, of how much impact can be expected, and what combinations are most effective.

SUMMARY OF THE GUIDELINES

Unequivocal statements of when particular techniques or combinations of techniques are better than others could not be developed. However, certain categorical statements can be made, and a logical analysis framework can be specified.

There is a logical set of steps to take to treat the problem of congestion and saturation.

1. Address the root causes of congestion first, foremost, and continually.
2. Update and, if necessary, improve the signalization.
3. Provide more space by use of turn bays and parking restrictions.
4. Consider both prohibitions and enforcement realistically to determine whether an effort would be futile or whether it might merely transfer the problem.
5. Take other available steps, such as allowing right turn on red (RTOR) while recognizing that the benefits will generally not be so significant as either signalization or more space.
6. Develop site-specific evaluations where there are conflicting goals, such as providing local parking versus moving traffic.

The following sections provide an exposition of these key elements in the recommended method of approaching the problem. The framework by which a problem should be considered is then presented. The framework has two components:

1. A focus on the identification of the problem in terms of probable cause and
2. A focus on the categorization of the possible solutions so that they may be readily found within these guidelines.

Root Causes

The problem should be attacked at the root causes, which are

1. Land use policies (concentrations of movement implied in some land use distribution, use of on-street facilities for loading and unloading of goods, and multiplicity of access and egress points and standing queues on the rights-of-way);
2. Demand pattern (concentration of work trips in a short period and unrestricted hours of goods activity);
3. Size of demand (number of vehicles, particularly low-occupancy private automobiles);
4. Use of street space [inefficient curb space management (parking, moving lanes, and the like)]; and
5. Pedestrian conflicts (lack of grade separation in areas of extreme intensity).

Engineers should continually educate other specialists and the public about this need to attack root causes. However, in the time frame of local, site-specific problems that they must address, the guidelines must often suffice.

Signalization

It is difficult to overstate how often poor signalization is the basic problem. After the signalization is improved through reasonably short cycle lengths, proper offsets (including queue clearance), and proper splits, many problems disappear. Sometimes, of course, there is just too much traffic. At such times, equity offsets to aid cross flows and different splits to manage the spread of congestion are appropriate if other options cannot be called on.

Discussions with and surveys of traffic engineers have revealed that a systematic consideration of signalization for congestion and saturation is rarely done. The procedures contained in the guidelines are recommended for use. Study of representative traffic patterns lends strong credibility to the conclusion that minimal-response (preplanned) signals policies generally suffice.

Space

If a problem cannot be remedied by signalization, then more space may be needed. Left-turn bays and, where appropriate, right-turn bays can aid individual movements as well as remove impediments to the through flows. Without question, additional lanes are a benefit. However, this tends to be an arterial-long solution that engineers often don't like.

Two-way left-turn lanes offer special advantages particularly along strip development sites. One-way systems, arterials with unbalanced lanes, and reversible lanes offer advantages, but also represent either major implementation problems or site-specific treatments. One-way systems require studies quite beyond the scope of congestion, although that may be the prime motivator for such a study. Unbalanced lanes require certain volume patterns to be useful.

Prohibition and Enforcement

Before instituting any prohibition or enforcement program, the engineer must decide whether it can be enforced strictly enough to realize most or all of the projected benefit (curb parking prohibition to provide a moving lane) and whether it will simply transfer or even accentuate the overall problem (circulation of vehicles that would otherwise be double-parking). Only then can the engineer consider that there is a potential benefit.

Other Attempts

Some solutions that are available can have either a net benefit or a net disbenefit depending on the site and the situation. RTOR is such a case. If it allows vehicles to "escape" from a congested arterial, it is quite suitable. If, however, it allows vehicles to "steal" available space on such an arterial, then it is inappropriate.

The question of prohibitions such as those affecting turning arises. These can be used only if alternate routes exist. Often, this is not the case.

More Detailed Evaluation

Very often, application of these guidelines will clarify the issue and identify a solution. In some cases, the final decision will rest on conflicting desires that might be usefully viewed in economic terms. Is removal of five parking spots worth the delay savings to the traffic stream? Are off-street goods facilities justified economically? Are pedestrian phases justified in terms of total person-minutes saved? What is a proper allocation of curb space?

If necessary, engineers can develop such an analysis for their individual cases. More general treatment of such situations is recommended for future research. Some of this type of work on curb space management for goods facilities has been done (2).

RANGE OF SOLUTIONS AVAILABLE

The fact that there is substantial traffic congestion virtually ensures that one is dealing with signalized intersections. However, it does not follow that one has only signal remedies at hand. Indeed, the possible treatments may be broadly classified as signal (timing and coordination) and nonsignal treatments.

Within the signal classification, there are two major subclassifications: minimal response (preplanned) and responsive signal control. It is not at all well established that highly responsive control is better than preplanned signal plans particularly for the heavier flow range. This is an indication that is being reinforced by trends in major computer-based study projects. Within the nonsignal classification, there are also two major subclassifications: regulatory and operations. Regulatory action consists of enforcement and of prohibitions. Operations, as used herein, consists of all other traffic measures.

The role of enforcement cannot be minimized. Many problems can be traced to the lack of enforcement of existing traffic regulations. In other cases, certain treatments are precluded initially because it is anticipated that there will not be adequate enforcement to have the measure work. Within this section, the following topics are addressed:

1. Improvement to be sought and
2. Available solutions.

The material on these topics combined with the method of approaching the problem previously presented represents the essence of the recommended framework for attacking the problem of congestion and saturation.

Improvements Desired

Before enumerating possible treatments, it is appropriate to dwell on the ends to be achieved. In other words, what improvements are being sought? The following is a set of the most common improvements that traffic engineers may wish to make when they face a traffic congestion problem:

1. Reduced geographic spread of congestion,
2. Reduced rate of spread of congestion,
3. Increased throughput,
4. Reduced delay,
5. Reduced stops, and
6. Improved regularity of service.

These are stated in the broad terms usually encountered as goals or objectives. Some of the items in this list are really secondary to other items for given flow levels. Figure 1 shows the primary objectives that should be sought by the engineer. These are dependent on the traffic level. First and foremost, the engineer must realize that at the more extreme flow levels the objective becomes the avoidance of spillback. All else follows from this. This is the explicit objective. The mathematical niceties of minimum stops or minimum delay or both collapse in the face of intersections blocked by vehicles.

Some comments on Figure 1 are in order. First, the primary objective to which engineers should address themselves does change depending on the flow level. Second, the major index of performance [measure of effectiveness (MOE)] also changes. However, both sets are well correlated to queue-extent measures; therefore, queue or occupancy patterns or both—particularly during red and at the onset of green—are good indicators across the entire range of conditions.

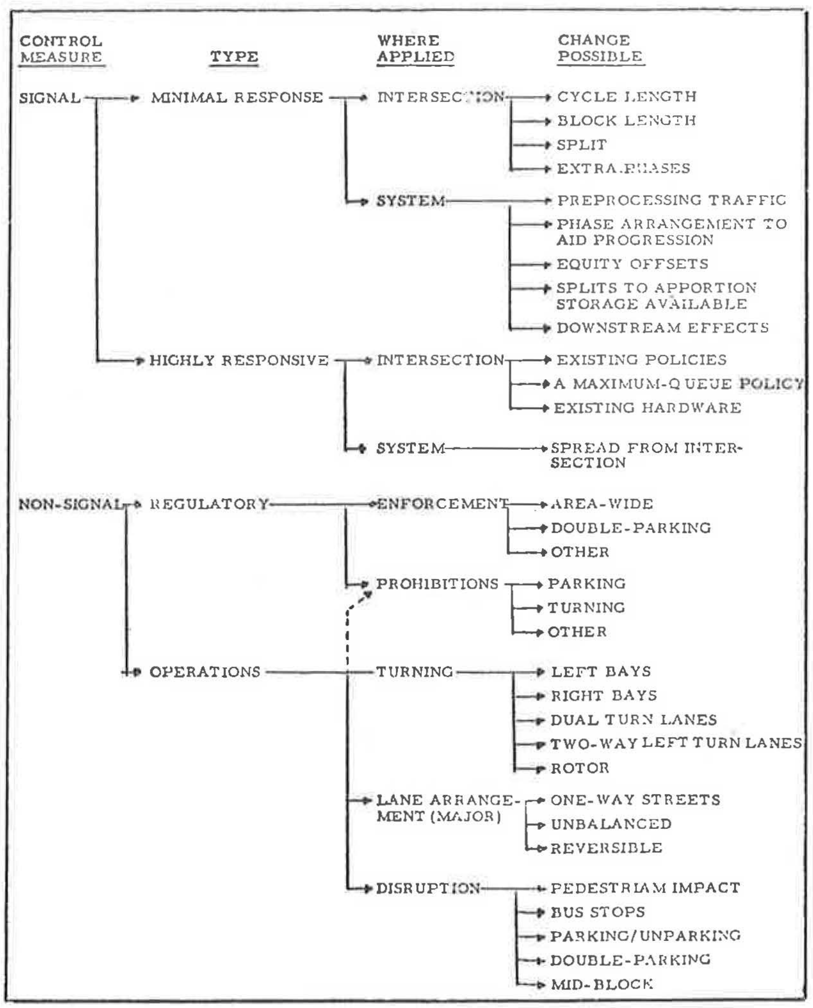
A special word of attention is appropriate on the desire to improve the regularity of service. During simple congestion, the variance as well as the mean of the delay per cycle increases as demand approaches capacity. Thus the individual driver will be exposed to greater variability in his or her individual experience from day to day. Improvements that minimize the mean delay will also enhance the regularity of the delay suffered.

There are times when a basic solution has been

Figure 1. Dependence of desired objectives on traffic condition.

TRAFFIC CONDITION	Congestion →	Saturation/Oversaturation
OBJECTIVE TO BE DESIRED	{Minimize Delay, Minimize Stops} →	{Avoid Spillback, Provide Equitable Service}
IMPACT OF USING OR TRYING TO ACHIEVE ABOVE OBJECTIVE (i. e., RESULTS)	• Intersection-specific congestion, not area-wide • Throughput Adequate • Service Regularity Enhanced • Enhancements to Special Groups (such as right-turners) Possible	• Reduce Spatial Extent and Rate of Spread of Congestion • Potential for occurrence of area-wide Congestion reduced, and service regularity thus enhanced • Throughput not Impeded • Special Treatment of User Groups Possible
MAJOR INDEX OF PERFORMANCE	Delay, Stops	Queue Extent and Occupancy

Figure 2. Range of solutions available.



achieved and some possible benefits can be realized by additional improvements addressed to specific subgroups. Right-turn bays are examples, for they frequently have little impact on a measure such as average delay of all vehicles although they have truly substantial benefits for a smaller segment of the traffic stream—those turning right.

Solutions Available

Figure 2 shows the range of solutions available. There is no simple statement of an ordered list of recommended solutions in decreasing order of preference. There are, however, indications of how much of one solution must be implemented to have an equivalent impact of so much of another solution. The engineer must then use this knowledge in conjunction with local conditions and practices to reach a decision. There are also indications of how best to use two solutions in conjunction with each other.

It can be stated that there is a simple set of initial steps that can be followed as an elimination checklist (Figure 3).

The engineer must reach a preliminary judgment of the underlying cause of the problem. At the same time, he or she must be assured that the solution is not trivial. Extensive queues may drive an engineer into the depths of these guidelines too quickly. Such problems can arise because of poor offsets, outdated splits, and excessive cycle lengths. As a first step, therefore, the engineer should prepare a preliminary opinion on the underlying cause. Given a preliminary opinion, there are a number of possible solutions that one is tempted to consider. Much of the guidelines are addressed to the candidate solutions, the considerations involved, and the relative merits of each.

CYCLE LENGTH AND BLOCK LENGTH

Two questions must be addressed. Do long cycle lengths have any virtue in their own right? Does block length enter into the cycle length determination?

Cycle Length and Capacity

One of the most prevalent erroneous beliefs in the traffic engineering community is that the capacity of an intersection increases substantially as cycle length is increased. This has been questioned in the past (3), and studies (4, 5) have provided data to support such questioning. Lack of substantial capacity increases with increasing cycle length is rooted in at least three factors: (a) Loss time per cycle is not that severe because of both usage of the amber and lower start-up delays than are often assumed; (b) the use of longer greens is inefficient because of increasing headways; and (c) a demand to fill rather long green times cannot be provided. The last item is just another manifestation of the storage problem.

Block Length and Storage

Cycle length may not be as powerful a capacity improver as one might think. However, no evidence was offered that there is a positive good to short cycle lengths in some cases.

To avoid a high potential for spillback, a minimum condition is that the moving platoon not exceed the available link storage. Thus

$$\pounds \leqslant f_1(\xi C/3600) \qquad (1)$$

where

ξ = vehicle storage length,
$\pounds$ = link storage distance,
C = cycle length in seconds, and
f_1 = flow in passenger cars per hour per lane.

$\pounds$ need not be the physical length of the link. If a policy decision is made that the stored vehicles should come no closer than within 61 m (200 ft) of the upstream intersection, then $\pounds$ is 61 m (200 ft) less than the physical length. Such a policy decision is in accord with the avoidance of the perception of congestion.

In order to avoid the situation shown in Figure 4, excessively long platoons must be avoided. Equation 1 may be rewritten as a constraint on cycle length:

$$C \leqslant (3600/f_1)(\pounds/\xi) \qquad (2)$$

Clearly, two contrary forces are at work. As the total critical land flow (all approaches) increases, cycle length increase brings some benefit; at the same time, flow increases on any one approach decrease the maximum cycle length permitted. Figure 5 shows block length that is adequate for cycle length set.

EQUITY OFFSETS

Unfortunately, avoiding spillback is not always possible, for there may be too many vehicles attempting to enter a particular link. With the extreme traffic congestion, it is not uncommon to see vehicles storing themselves in the intersection, to the detriment of the cross traffic. One common solution to this spillback problem is to place a traffic control officer at this site to prevent such events. Another approach is an intensive ticketing program for such offenders. The former approach is not only historically more effective, but it is also the one demanded by a public afflicted with spillback.

A possible alternative solution exists in changing the basic concept of what the offset is supposed to accomplish. However, this solution should not be implemented until one is certain that a better offset cannot alleviate the problem. The treatment to be presented now is only for that period after the best possible offset has failed because of the size of the volume demanding access to the link.

The treatment, shown in Figure 6, is as follows:

1. Allow the oversaturated direction to have green until the vehicles blocking the intersection just begin to move;
2. Switch green to the cross traffic; and
3. Allow the cross stream to move until it has had an equitable input into the oversaturated link or at least to the intersection.

This offset, the equity offset, is not determined in the usual fashion. The upstream red should begin L/V_{ACC} s after downstream green initiation, where V_{ACC} is the acceleration wave speed in meters per second. Assume g_1 as the green time at the upstream intersection (percentage of cycle) and g_{c1} as the green time at the critical intersection. Thus

$$t_{off} = g_1 C - (L/V_{ACC}) \qquad (3)$$

where C = cycle length in seconds and L = physical block length in meters. Typically, $V_{ACC} \simeq 5$ m/s (16 ft/s).

The original link must have unavoidable saturation. Neither any signal nor any available nonsignal remedy could have helped it. Only then is this link given up on and the best possible done for other traffic.

Figure 7 shows an arterial on which the volume ensures oversaturation, at least of links 2 and 3. There are no turns. The equity offset for link 2 is computed as -1.5 s so that a simultaneous offset will happen to provide an equity offset function.

Figure 8 shows the queue per lane on link 1, the cross-stream link. The only offset that is varied is that in link 2. Figure 8 clearly indicates that the equity offset (offset = 0) is quite important to the cross-street traffic (link 1).

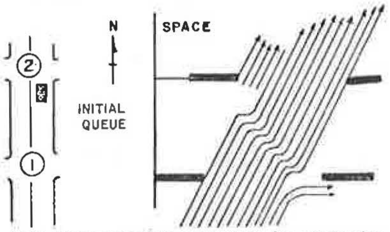
DOWNSTREAM BLOCK LENGTH

If a decision is made that the cycle length at the critical intersection (CI) is to be larger than the downstream

Figure 3. Initial classification and elimination checklist.

Apparent Problem	Initial Steps
Area-Wide Congestion	• Identify the Critical Intersection (CI). • It is unlikely that there are two CI's. If there are, it is likely that each can be considered with its own area of influence. • Do not erroneously identify the intersection downstream of the CI as the CI. • Classify the type of oversaturation
Spillback in a Link	• Determine whether a simple split adjustment is sufficient. • Determine whether the cycle length is too long for the block length and flow. • Determine whether the offset is poor for the primary and secondary flow mix. • Identify any special blockages in the link (double parking, queues for garages, car washes, etc.).
Single Intersection	• Isolate primary symptom
→ Single Approach	• Check split and offset as above. • Check burden caused by turns.
→ Two Approaches, same Right-of-Way	• Check same as one approach, but with emphasis on interference with each other.
→ Two or More Approaches More than One Right-of-Way	• Consider methods for increasing net capacity. • Consider methods for minimizing spatial extent of possible oversaturation and area-wide congestion.

Figure 4. Cycle length too large for block length and flow.



Note: Stoppage prevents some vehicles from discharging intersection 1. It also causes conditions that could lead to a blockage of intersection 1. For example, if there were any delays in moving through link 1-2, the cross stream traffic at intersection 1 would be quickly affected.

Figure 5. Block length adequate for cycle length set.

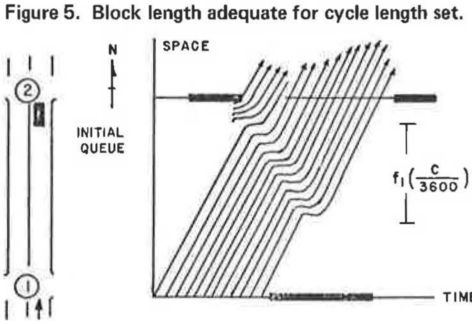


Figure 6. Equity offsets.

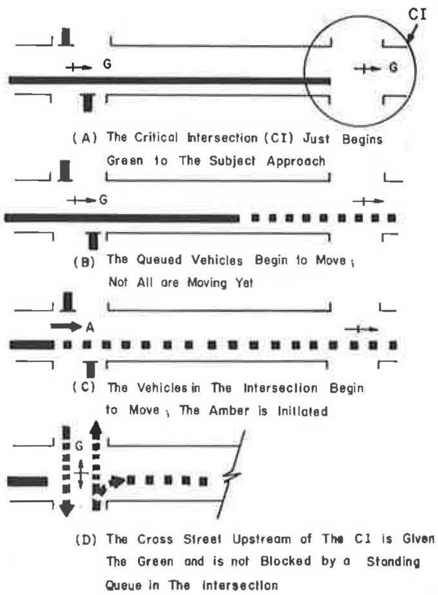


Figure 7. Impact of offsets.

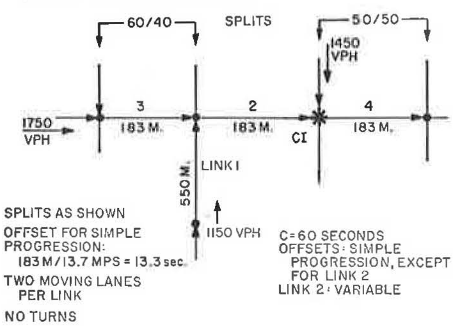
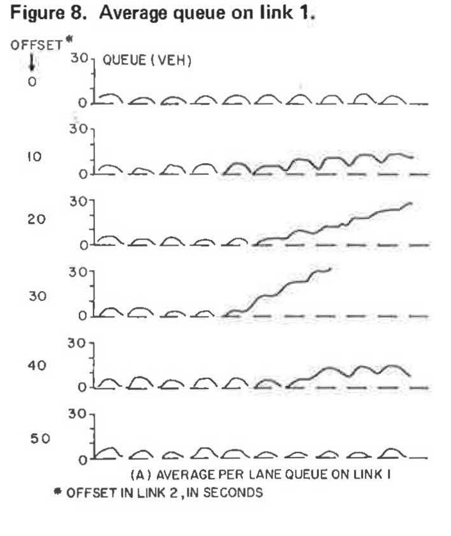


Figure 8. Average queue on link 1.



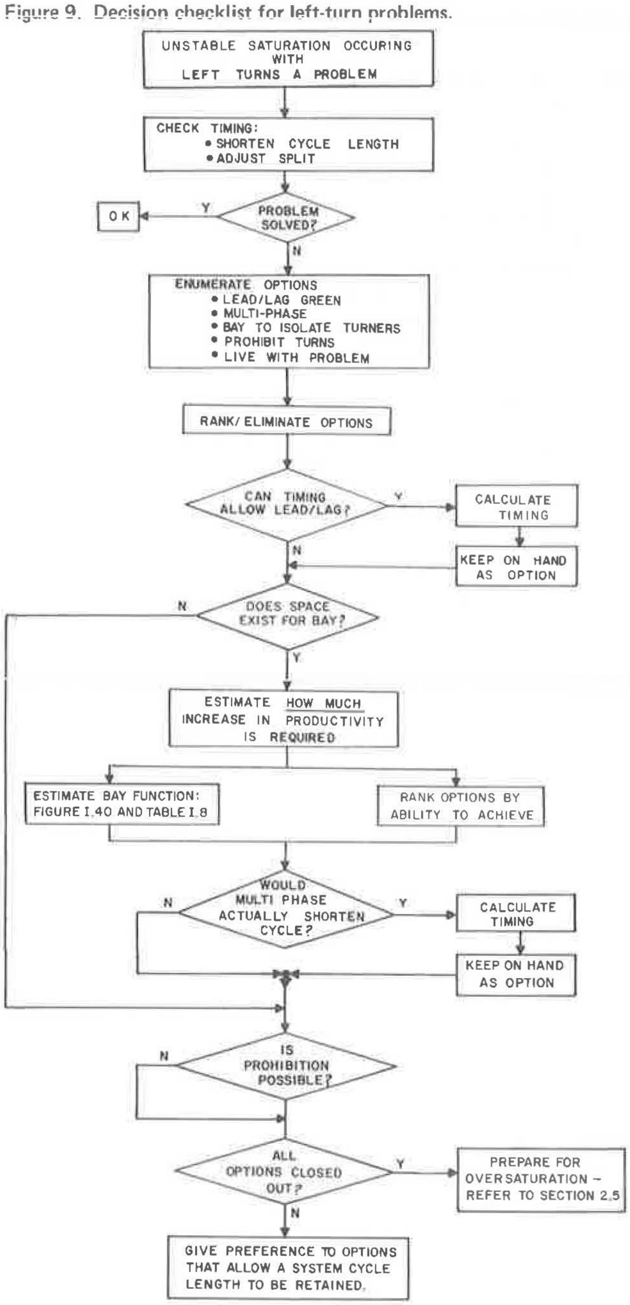
(A) AVERAGE PER LANE QUEUE ON LINK 1
* OFFSET IN LINK 2, IN SECONDS

cycle lengths (perhaps because the CI is to have multiple phases), the queue extent to be stored may be shown to reach a maximum of more than twice the single-cycle discharge of the CI into the link.

SPLIT

For congested flow, the standard rule of proportioning available effective green in the ratio of the critical lane flows will not suffice. It should be appreciated that, as the demand approaches capacity, greater queues will be experienced, as will greater delays and greater fluctuations (variance) in delay per cycle. For unstable saturation and for oversaturation, a different concept should prevail. Clearly, there are situations in which the CI simply cannot handle the total demand put on it. Must the same sense of equitable treatment still hold? We recommend that the split be apportioned so that the rate of growth of congestion in both (or all) directions be equalized; both directions should exceed their respective links or defined system boundaries at the same time. This is addressed in the guidelines.

Figure 9. Decision checklist for left-turn problems.



EXTRA PHASES

As a rule, multiple phases should be avoided particularly because they generally require an increase in the overall cycle length. Other options should be considered: turn bays, shorter cycle lengths, parking restrictions, leading or lagging greens or both, and turn prohibitions. Still, there are cases when multiple phasing is clearly necessary. Even when the left-turn volumes are less than 120 vehicles/h, there are conditions under which a left-turn phase can be added without increasing the cycle length required.

ENFORCEMENT AND IMPACT

Two of the most chronic violations that aggravate the congestion and oversaturation problems are intersection blockage and parking regulation violations. Equity offset represents an attempt to circumvent the first and avoid or delay the need for on-scene traffic control officers.

The UTCS-1 simulation was used to study the impact of double-parkers in a 183-m (600-ft), 3-lane link. The resultant curve can also be used to estimate the impact of adverse uses of a curb lane from which parking was removed to increase capacity.

RIGHT-TURN BAYS

The creation of a right-turn bay allows

1. An increase in productivity or, if it is desired, a decrease in the effective green allocated to the phase and
2. A decrease in the local delay with the right turners realizing most of the delay savings.

The increase in productivity, expressed as a percentage, can be comparable to the right-turn percentage.

The length of the turn bay should be approximately the same length (slightly longer) as the queue that typically forms. In this way, maximum presorting can occur. Thus, in practical terms, short cycle lengths and this objective complement each other, for released platoons are smaller, and the necessary length is easier to achieve.

An example in the guidelines illustrates, however, that much of the benefit is achieved by the existence of a bay of even moderate size. Still, short cycle lengths aid presorting and should be used as a companion measure.

PROBLEMS DUE TO LEFT-TURN MOVEMENT

Figure 9 shows a decision checklist to be used in considering a problem that arises because of a left-turn movement. The final decision must be evaluated with due consideration of the three problems.

1. Is an alternate route available for the left turners? How much does it adversely impact them? Can the alternate route afford to be impacted by the additional flow?
2. How many parking spaces must be removed to aid the flow means of a turn bay or even an additional lane? What is their economic value?
3. What delay is being suffered now?

Frequently, the decisions can be reached by systematically thinking of the options as shown in Figure 9 and keeping these issues in mind. Sometimes a benefit-cost or cost-utility decision would be required for a "close decision" or highly sensitive issue.

TWO-WAY TURN LANES VERSUS UNBALANCED FLOW

The engineer may judge that heavy congestion or even stable saturation at intersections is sometimes inevitable. At mid block, however, it is the opinion of some that the engineer, and the public, will generally find a lower (but significant) amount of congestion to be equally unattractive. The engineer therefore may be solving a congestion problem at mid block and a saturation problem at the intersection.

Two-way left-turn lanes can remedy such congestion impact if space permits. There is some evidence that they can substantially improve the accident situation (6). The option of a two-way turn lane may solve the mid-block congestion problem. Given that the additional lane will now be added, however, opens the possibility that it can be used at the intersection by the through flow—if unbalanced flow is implemented. This will allow a reduced green for this approach, perhaps to the benefit of other phases and thus the system. The guidelines contain a checklist similar in concept to the information shown in Figure 9 for mid-block congestion or for unbalanced flow and reversible lanes. Of course, this use of unbalanced flow requires planning. Can the opposing direction accommodate its own turners without unduly impeding its continuing vehicles?

Note that any decision involving two-way turn lanes versus unbalanced flow in this context considers only the congestion and saturation issue. Data on accident advantages are not sufficient to say whether there is an accident benefit of two-way turn lanes (for example, removal from the traffic stream as opposed to decreasing the density) that should override.

DUAL TURN LANES

When turning volumes are extremely heavy, both capacity and queue extent may dictate use of two lanes for a turning movement. Data in the literature and discussions with engineers responsible for such sites indicate that there is no downward correction factor needed for either of the lanes.

OTHER NONSIGNAL REMEDIES

The guidelines incorporate consideration of all remedies (possible treatments) shown in Figure 2 including left-turn bays and RTOR. Other results are incorporated into the guidelines, including

1. Rules of thumb on when to use simultaneous and other progressions,
2. Rules of thumb for productivity increases due to left-turn bays,
3. Circumstances under which multiple phases may actually not increase the cycle length, and
4. Illustrations of the relative impact of alternative treatments.

More important than any of these specifics, however, is the tutorial approach of the guidelines and the development of a systematic methodology. Once acquainted with these guidelines, engineers can better sort out the issues in their own applications and consider more implications.

CONCLUSIONS

The problem of congestion and saturation is widespread, and is not approached in any consistent manner. There are definite measures that can be taken, but preventive action addressing the root causes must be given a high priority. Among the measures that can be taken, those relating to signalization generally can have the greatest impact. There are distinct signal plans for avoiding spillback and for living with spillback. The nonsignal remedies are in no way to be minimized, particularly those that provide space either for direct productivity increases or for removing impedances to the principal flow. The guidelines produced in this work provide both a tutorial and an illustrated reference in what techniques to consider and how to consider them systematically. The interested reader is referred to Appendix 1 of NCHRP Project 3-18(2) (1) for a more detailed exposition.

ACKNOWLEDGMENT

This study was conducted under National Cooperative Highway Research Program Project 3-18(2) (1). The opinions and findings expressed or implied in this paper are ours. They are not necessarily those of the Transportation Research Board, the National Academy of Sciences, the Federal Highway Administration, the American Association of State Highway and Transportation Officials, or the individual states participating in the National Cooperative Highway Research Program.

REFERENCES

1. Traffic Control in Oversaturated Street Networks. NCHRP, Project 3-18(12), final rept.
2. P. A. Habib. Accommodating Goods-Movement Vehicles in the City Center. Polytechnic Institute of New York, Brooklyn, PhD dissertation, June 1975.
3. K. Moskowitz. Long Cycles, Short Cycles, and "Lost Time" at Signalized Intersections. Paper presented at the Workshop on Intersection Capacity, TRB, Jan. 1974.
4. G. M. Webb and K. Moskowitz. Intersection Capacity. Traffic Engineering, Jan. 1956.
5. R. L. Carstens. Some Traffic Parameters at Signalized Intersections. Traffic Engineering, Aug. 1971.
6. M. R. Hoffman. Two-Way Left-Turn Lanes Work. Traffic Engineering, Aug. 1974.

Guidelines for Application of Selected Signs and Markings on Low-Volume Rural Roads

William R. Stockton, John M. Mounce, and Ned E. Walton, Texas A&M University, College Station

Existing standards and guidelines for the application of signs and markings are unsuited and inefficient for use on low-volume rural roads (roads with less than an average of 400 vehicles/day). To alleviate this inadequacy, several potentially hazardous situations were evaluated to ascertain actual needs for signs and markings as they relate to economy and safety. These evaluations were based on recent research and on probability of conflict analyses with regard to the needs for signing and marking of intersections, horizontal curves, and sections of inadequate passing sight distance. The research revealed that more efficient intersection control can be attained from the careful application of stop signs and crossroad warning signs based on approach speed, sight distance, and combined intersecting volumes. Treatment of horizontal curves can be made more efficient through the application of more stringent guidelines without adversely affecting safety. Striping of no-passing zones was found to be very inefficient in most instances because the probability of conflict in these situations is virtually nil; guidelines for alternative treatments are presented. Overall, the authors felt that application of guidelines suited to the rural context would result in savings in time, money, and frustration on the part of responsible agencies.

Low-volume rural roads [roads with less than 400 vehicles/day of average daily traffic (ADT)] make up the bulk of the public roadways operated in this country. Their existence is essential to the various aspects of rural life. "Farm-to-market" and country roads provide accessibility for communities as well as perform as the major avenue of agricultural commerce. Forest roads and park roads are necessary for the operation, maintenance, and accessibility of national forests and parks.

Heretofore, application of traffic control devices on rural roads has been restricted to those guidelines and regulations contained in the Manual on Uniform Traffic Control Devices (MUTCD) (1). However, those guidelines, which were developed primarily for major highways and city streets, are easily recognized as impractical for application on low-volume rural roads. Adherence to existing MUTCD guidelines not only is unnecessarily expensive but also produces considerable visual clutter in the rural environment. Therefore, a reduction in the levels of signing and marking on low-volume rural roads has been given careful consideration. This paper contains the guidelines developed for the application of warning and regulatory signs on low-volume rural roads and the analyses that led to their development.

Of primary importance in the reduction of the level of signing and marking is the corresponding effect on safety. To assess this effect, three major potential hazard situations were analyzed—intersections, horizontal curves, and sections of insufficient passing sight distance, or no-passing zones. Two of the situations, intersections and no-passing zones, were analyzed by using a probability of conflict technique. Safety on horizontal curves was based on research by Ritchie and others (2) and field observations made during the course of this research.

One of the overriding concerns throughout the conduct of the research was development of guidelines that not only were easily understood and readily implementable but also were truly suited to the rural situation. Guidelines contained in the MUTCD may result in too little intersection control and too much horizontal curve and no-passing zone warning if applied in rural areas. Therefore, a combination of economic analysis, engineering judgment, and field observation was applied to produce the guidelines contained herein. The analyses presented are abridgments of the actual research. Detailed descriptions of the research may be obtained from the Texas Transportation Institute.

INTERSECTION CONTROL

The analyses and guidelines developed for treatment of low-volume rural intersections stemmed from the question, What is the probability of accident occurrence at a low-volume rural intersection?

Analysis

The initial step in determining the probability of an accident was the determination of the probability of conflict. From this determination, the expected number of accidents per year can be estimated.

For the purpose of analysis, eight assumptions are made.

1. Conflict is defined as that maneuver of vehicle B that makes the driver of vehicle A change speed or direction to maintain a comfortable clearance interval.

2. Average speed is 64 km/h (40 mph) or approximately 18 m/s (60 ft/s), and no intersection control or signing is provided.

3. Any two vehicles approaching the intersection from conflicting directions in such a way that the second vehicle would enter the intersection within 3 s after the first vehicle enters the intersection are said to be in conflict; that is, one or both vehicles must make a speed change maneuver to provide comfortable clearance.

4. Effects of sight distance are not considered in the analysis portion.

5. All vehicles arrive during a 12-h period from 7 a.m. to 7 p.m. (All vehicles probably do not arrive between 7 a.m. and 7 p.m., but, because this assumption covers the worst condition, it is used here.)

6. All arrivals are random; that is, they follow a Poisson distribution.

7. Only one arrival per approach is possible during one 3-s interval; that is, all approaches are single-lane approaches, and all headways are greater than three seconds.

8. The possibility of vehicles arriving on three approaches within a 3-s interval is negated because the probability of such an occurrence is a maximum of 2.01×10^{-5} for the volumes under consideration.

The probability that two vehicles will be in conflict is the product of the probability that either vehicle is in the conflict region during the interval Δt(3 s). Or

$$P(\text{conflict}) = P(\text{vehicle A in conflict region during } \Delta t) \times P(\text{vehicle B in conflict region during } \Delta t) \quad (1)$$

This probability of conflict analysis revealed that, on the average, 0.68 conflict/day could be expected on two intersecting roadways of 100 vehicles/day ADT each. ADTs were incremented by 25 vehicles/day on each facility to provide an expected number of conflicts E(C) for all ADT combinations up to 400 by 400 (800 vehicles/day ADT combined intersecting volumes). Expected number of conflicts ranged from 0.04/day for a combined ADT of 50 vehicles/day (25 by 25) to 10.67/day for a combined ADT of 800 vehicles/day. Selected values for E(C) given in Table 1 reveal that the highest expected number of conflicts for a given combined ADT occurs when the intersecting volumes are approximately equal. This indicates that the "worst case" condition may not be the intersection of a minor road with a major road but actually may be the intersection of two very similar roads.

Given, then, the expected number of conflicts, what is the probability of an accident? Data from a study by Perkins and Harris (3) indicated that about 33 accidents occur in every 100 000 conflicts for the situation in question, or

$$P(A,C) = 0.000\,33 \quad (2)$$

where P(A, C) = probability of an accident given a conflict. Other data indicated that P(A, C) ranges from 0.000 25 to 0.000 35. Therefore, to examine worst case conditions, a value of P(A, C) = 0.000 35 was chosen.

Then the probability of an accident P(A) is given by

$$P(A) = P(A,C) \times P(C) \quad (3)$$

Multiplying the probability of an accident occurrence in a given 3-s interval by the number of such intervals in a day yields the expected number of accidents per day. Thus multiplying by 365 yields the expected number of accidents per year E(A).

For the two intersecting facilities of 100 ADT each, E(A) = 0.087. From the selected values of E(A) given in Table 2, it can be seen that one or more accidents per year can be expected above a combined ADT of approximately 700 vehicles/day. However, the absolute number of expected annual accidents is not of sole importance. Of equal or greater importance is the estimated annual cost of accidents in the no-control alternative as it relates to the estimated annual cost of the two-way-stop-control alternative.

Estimated annual cost of accidents at a particular intersection is the product of estimated cost per accident and estimated number of accidents per year. The primary determinant in accident cost is severity. Results of a study by Burke (4) showed little variation in severity over the ADT range 0 to 400. However, as would be expected, severity (5) as well as the proportion of fatalities (6) was found to increase with speed. Combining the results of these two studies, we developed a weighted accident cost equation:

$$\text{Cost} = F_P(A) + F_I(B) + F_F(C) \quad (4)$$

where

F_P = proportion of property-damage-only accidents,
A = average cost of property-damage-only accidents [\$318 (4)],
F_I = proportion of injury accidents,
B = average cost of injury accidents [\$1955 (4)],
F_F = proportion of fatal accidents, and
C = average cost of fatal accidents [\$13 781 (4)].

Combining the proportional factor for each type of accident with the average cost of that type of accident in the preceding equation resulted in a weighted average cost per accident for each speed group. For example, the weighted average cost of 32-km/h (20-mph) accidents would be found as follows:

$$\text{Cost/accident} = 0.750(\$318) + 0.248(\$1955) + 0.002(\$13\,781) = \$750 \quad (5)$$

These costs and the proportional factors from which they were derived are given in Table 3. Average yearly accident cost per intersection by speed for each ADT combination is given by the product of expected number of yearly accidents, E(A) (Table 2) and weighted average cost per accident (Table 3). These costs were compared with costs associated with the use of two-way stop control. Two-way-stop control costs included expected accident cost (approximately 20 percent of that of no control) and additional annual motor vehicle operating costs due to the stop control. Additional operating cost is the difference between the cost of continuing through the intersection at the approach speed and the cost of slowing to a stop from the approach speed and returning to the previous speed. As would be expected, the costs of stopping and regaining running speed increase with higher running speeds. Table 4 (7) gives additional operating costs for each speed group and the compilation of expected cost of two-way-stop control on a facility with an ADT of 100 vehicles/day.

Selected values of costs associated with no control and two-way-stop control are compared in Tables 5, 6, and 7.

Careful examination of the estimated cost tables reveals that, up to combined volumes of 200 vehicles/day, the expected annual accident costs associated with no control are less than the accident and operating costs associated with two-way-stop control. At higher ADT values, these expected costs become equal; as the ADT values become higher still, the no-control alternative becomes more expensive. As a result of increased running speeds, this breakpoint between the economic justification of the two control alternatives increases as the speed on the intersecting roadways increases. These analyses showed that the no-control alternative was more economical up to the following combined ADT (1 km/h = 0.61 mph):

Speed (km/h)	Combined ADT	Speed (km/h)	Combined ADT
32	300	80	700
48	520	96	720
64	650		

The calculation of these breakpoints is derived by equating the costs of the no-control alternative and the cost of the two-way-stop-control alternative as represented in the following equation:

$$E(A) \times C_A = (ADT \times 365 \times C_S) - 0.2[E(A) \times C_A] \quad (6)$$

which can be simplified to

$$0.8[E(A) \times C_A] = T_Y \times C_S \quad (7)$$

where

$E(A)$ = expected number of yearly accidents with no control [for equally split traffic volumes (Table 2)],
C_A = weighted average cost per accident (Table 3),
T_Y = yearly traffic volume = ADT × 365, and
C_S = additional motor vehicle operating cost with two-way-stop control (Table 4).

Thus, for each approach speed there is a point below which stop control is not economically justified. However, as mentioned previously, economy is not the only necessary consideration. Although two-way-stop control may not be economically justified, adequate visibility of a crossing roadway is vital in the absence of signing. Because it is highly likely that a situation will arise in which stop control is not justified and crossroad visibility is inadequate, a standard crossroad warning sign (W2-1 in MUTCD) is necessary. Criteria for the use of a crossroad sign were based on sight distance requirements specified by American Association of State Highway and Transportation Officials (7). The inclusion

Table 1. Expected number of conflicts per day.

Facility B ADT	Facility A ADT 100	200	300	400
100	0.68	1.36	2.03	2.70
200	1.36	2.70	4.04	5.37
300	2.03	4.04	6.04	8.03
400	2.70	5.37	8.03	10.67

Note: Values are in vehicles per day.

Table 2. Expected number of accidents per year.

Facility B ADT	Facility A ADT 100	200	300	400
100	0.087	0.174	0.259	0.345
200	0.174	0.345	0.516	0.686
300	0.259	0.516	0.772	1.026
400	0.345	0.686	1.026	1.363

Note: Values are in vehicles per day.

Table 3. Weighted average cost per accident by speed.

Speed (Km/h)	Proportional Factors F_P	F_I	F_F	Weighted Average Cost/Accident ($)
32	0.750	0.248	0.002	750
48	0.720	0.277	0.003	812
64	0.660	0.322	0.008	969
80	0.580	0.400	0.020	1242
96	0.410	0.783	0.077	1733

Note: 1 km/h = 0.621 mph.

Table 4. Expected annual costs associated with two-way-stop control.

Approach Speed (km/h)	Operating Cost/Stop ($)	Stops/Year	Annual Operating Cost ($)	Average Cost/Accident ($)	Expected Number of Accidents	Expected Annual Accident Cost ($)	Expected Annual Cost of 2-Way Stop[a] ($)
32	0.0022	36 500	81	750	0.0174	13	94
48	0.0040	36 500	145	812	0.0174	14	159
64	0.0059	36 500	216	969	0.0174	17	233
80	0.0083	36 500	302	1242	0.0174	22	324
96	0.0116	36 500	422	1733	0.174	30	452

Notes: 1 km/h = 0.621 mph.
Operating cost per stop is based on Cleveland (5).
[a]Annual operating cost plus expected annual accident cost.

Table 5. Accident costs per year for no control and two-way-stop control at 32-km/h approach speeds.

Facility B ADT	Type of Control	Facility A ADT 100	200	300	400
100	None	65	130	194	259
	2-way stop	94	107	120	133
200	None	130	259	387	514
	2-way stop	107	213	238	264
300	None	194	387	579	770
	2-way stop	120	238	357	395
400	None	259	514	770	1022
	2-way stop	133	264	395	526

Notes: 1 km/h = 0.621 mph.
Values are in dollars.

Table 6. Accident costs per year for no control and two-way-stop control at 64-km/h approach speeds.

Facility B ADT	Type of Control	Facility A ADT 100	200	300	400
100	None	84	169	251	334
	2-way stop	233	250	266	283
200	None	169	334	500	665
	2-way stop	250	499	532	565
300	None	251	500	748	994
	2-way stop	266	532	798	847
400	None	334	665	994	1320
	2-way stop	283	565	847	1129

Notes: 1 km/h = 0.621 mph.
Values are in dollars.

of the crossroad warning sign as part of low-volume rural intersection control was, in our opinion, a necessary safety measure in the absence of stop control and adequate sight distance. Although the erection of four crossroad signs is more expensive than two stop signs, the savings in motor vehicle operating costs over the life of the signs more than offset the additional capital cost of the crossroad signs.

Guidelines

The analyses coupled with engineering judgment and many hours of field observation in rural areas resulted in certain recommended guidelines for safe and economic low-volume rural intersection control. Stop signs should be on low-volume rural roads (paved or unpaved) that intersect paved highways provided that the low-volume road

1. Serves 10 or more residences,
2. Has an ADT of 50 vehicles or more, or
3. Is 8 km (5 miles) long or longer.

Two guidelines should be followed unless two things can be shown.

1. The combined ADT for the two intersecting roadways is less than the following for the corresponding lower approach speed of the two facilities (1 km/h = 0.621 mph):

Approach Speed (km/h)	Combined ADT	Approach Speed (km/h)	Combined ADT
32	300	80	700
48	500	96	720
64	640		

2. The sight distance on each approach is at least the same as the following for the corresponding approach speed (1 km/h = 0.621 mph; 1 m = 3.28 ft):

Approach Speed (km/h)	Sight Distance (m)	Approach Speed (km/h)	Sight Distance (m)
32	27	80	66
48	39	96	78
64	54		

Sight distance is defined here as a triangle of clear visibility with legs of a length equal to the distance shown for the corresponding speed. This triangle shall apply from all directions of approach. For example, approach speeds on two intersecting facilities are 80 km/h (50 mph) and 64 km/h (40 mph) respectively. A driver approaching the intersection on the 80-km/h (50-mph) facility must, 66 m (220 ft) from the intersection, have clear visibility throughout a cone of vision extending 54 m (180 ft) in each direction along the crossing roadway (Figure 1).

For intersections that meet the ADT requirements for no control but do not meet the sight distance requirements, a standard crossroad sign, W2-1, may be used in advance of the intersection instead of two-way-stop control.

The requirements for intersection control just given can be determined from Figure 2. The procedure includes three steps.

1. Enter combined ADT in part A and project horizontally to intersect with lowest approach speed. If the intersection of these two lines is above the curve (shaded area), stop here and install stop signs on the minor approach or approaches.
2. Enter combined ADT in part A and project horizontally to intersect with lowest approach speed. If the intersection is below the curve, project intersection point downward into part B.
3. Enter shortest sight distance on lower speed approach and project horizontally to intersect line drawn in step 2. If this intersection point lies below the line, no control is needed. If the intersection point lies above the line (shaded area), a standard crossroad sign is needed on all approaches.

HORIZONTAL CURVES

Aside from the elements of geometric design, use of warning signs is one of the primary methods of improving safety on horizontal curves. In an effort to provide guidelines for the application of curve warning signs on low-volume rural roadways, existing practices, recent research, and subjective data obtained in this study were assimilated. Recommendations based on these elements were developed.

Analysis

The MUTCD provides minimal guidelines for the application of curve signs and advisory speed plates. Several states have developed specific warrants for curve signs within the requirements of the MUTCD. These warrants require the availability of ball bank indicators or detailed curve data. The objective of this endeavor was to establish guidelines for curve signing in lay terms to permit ready application. The primary assumption made was that supplemental driver information (signs, markings, and the like) is more critical in night driving than in day driving. Using the equation

$$S = 0.277V_1T + \left\{[0.277^2(V_1^2 - V_1^2)]/2a\right\} \quad (8)$$

where

S = required deceleration distance in meters,
T = perception-reaction time,
V_1 = approach speed in kilometers per hour,
V_2 = safe curve speed in kilometers per hour, and
V_3 = deceleration rate in meters per second^2

required distances for deceleration to safe curve speed that were calculated assuming an average deceleration rate of -2.1 m/s^2 (-7 ft/s^2). The addition of a perception-reaction time of 2 s yielded the minimum distance at which a driver must be aware of an impending situation. These distances are shown for various combinations of approach and curve speeds in Figure 3.

For certain combinations of approach and curve speed, the roadway itself generally provides adequate information for proper vehicular maneuvers. High beam visibility distance [about 90 m (300 ft)] was assumed to be the upper limit at which the roadway provides adequate information. A line was drawn on Figure 3 through the 90-m (300-ft) contour. Distances to the upper left of the contour line require advance supplemental information; distances to the lower right do not. Calculated data points were compared with field observations. A close correlation was found between calculated critical speed differentials and those curves observed to be hazardous.

In general, at approach speeds greater than 48 km/h (30 mph), a differential of 16 km/h (10 mph) between approach speed and safe curve speed required perception-reaction-deceleration distances necessitating advance warning. This advance warning can be provided through the use of standard (W2-1) curve signs. Speed differentials of 24 km/h (15 mph) are characteristic of more severe curvature and should be identified with a curve sign (W2-1) and an advisory speed plate (W13-1).

The relative degree of risk associated with this reduced level of signing on curves can be evaluated based on driver characteristics in a curve maneuver. The important question to be answered is whether the reduced level of signing (fewer or no signs) contributes to potentially hazardous operations. To determine the effect of signing level, Ritchie and others (2) conducted a study in 1968 that involved the relationship between forward velocity and lateral acceleration in curve driving. In a subsequent study, the previous research was expanded to determine the driver's choice of curve speed as a function of curve and advisory speed signs (2).

The study was based on the actions of 50 subjects negotiating sections of roadways containing 162 curves that required deceleration from normal operating speed. Four levels of signing were evaluated: (a) no signs, (b) curve signs, (c) curve signs with advisory speed plaques, and (d) curve signs without advisory speed plaques. In addition, all curves were lumped together to obtain an overall condition. The significant results of the study were as given in Table 8 (2).

1. As forward velocity increased, lateral acceleration decreased, indicating that, at higher speeds, drivers tend to provide themselves with a greater margin of safety on curves.
2. Drivers were more cautious on curves without signs than on curves with signs. Mean lateral accelerations on curves with signs ranged from 0.280 to 0.159 *g*; on curves without signs, they ranged from 0.259 *g* to 0.124 *g*.
3. Except at very low speeds, greater lateral acceleration (0.268 to 0.161 *g*) was produced on signed curves with advisory speed plaques than on signed curves without advisory speed plaques.
4. Below 64 km/h (40 mph), posted advisory speeds were exceeded more often than above 64 km/h (40 mph).

The conclusion of Ritchie and others was that the experimental data do not support the hypothesis that the roadway signs are responsible for the inverse relationship between speed and lateral acceleration. Roadway signs serve to reduce uncertainty and increase the confidence with which the driver proceeds. Therefore, the reduced level of signing on curves on low-volume rural roads can be effected without appreciable decrease in level of safety.

Guidelines

Based on the foregoing analyses and associated assessment of relative degree of risk and on engineering judgment founded on field observations, guidelines were developed.

1. Curve signs (W1-2) should be placed in advance of all curves with intersecting angles of 45 deg or more on paved roadways and 60 deg or more on unpaved roadways unless it can be shown that the posted speed limit is 55 km/h (35 mph) or less or that the combination of normal approach speed and safe curve speed requires a perception-reaction-deceleration distance of less than 90 m (300 ft) [the combination of the speeds produces a point to the lower right of the 90-m (300-ft) contour line in Figure 3].
2. Advisory speed plates (W13-1) should be used in conjunction with curve warning signs when the safe curve speed is 8 km/h (5 mph) below that speed warranting a curve sign (the combination of the speeds produces a point to the upper left of the appropriate line in Figure 3).

NO-PASSING ZONES

Because most low-volume rural roads follow the existing horizontal and vertical curvature of the terrain, there can be a considerable amount of inadequate passing sight distance. Treatment of this condition, with respect to the MUTCD, requires the use of standard no-passing-zone stripes on all such sections. Because this practice may be unnecessarily expensive, an evaluation of the need for such a practice is necessary. The probability of conflict technique was again employed for this determination.

Analysis

For analysis purposes, all passing maneuvers were assumed to be undertaken without regard for oncoming vehicles (as soon as a driver overtakes a slower vehicle, he or she pulls out to pass). This assumption produces unrealistic results that will be adjusted later.

In the basic situation for development of probability of conflict, a driver in vehicle A traveling at 80 km/h (50 mph) overtakes vehicle B traveling at 64 km/h (40 mph). Without regard for safe passing sight distance, the driver in vehicle A pulls into the opposing traffic lane to pass vehicle B. Before vehicle A can return to the right lane, vehicle C, traveling in the opposite direction, comes into conflict with vehicle A. The necessary determination in this evaluation is the probability of occurrence of this situation. To begin with, the probability of vehicles A and B being in this passing situation is the probability of simultaneous arrival (within a Δt of 2 s) of two or more vehicles, which is given by

$$P(x) = 1 - [P(0) + P(1)] \tag{9}$$

Based on the maximum low-volume rural road ADT of 400 vehicles (200 vehicles in each direction), the probability of such an occurrence in any 2-s interval is 4×10^{-5}. Over an entire day, the expected number of potential passing situations is 0.864.

Assuming that the following vehicle passes at a constant speed of 80 km/h (50 mph), the length of time that vehicle A is encroaching on the opposing lane is determined as follows:

$$t = d/0.277v \tag{10}$$

where

t = time left lane occupied,
d = distance traveled in left lane in meters, and
v = average speed in kilometers per hour.

For an assumed speed of 80 km/h (50 mph), the duration of encroachment on the opposing lanes is approximately 11 s. Therefore, if an opposing vehicle arrives during that 11-s interval, there will be a conflict. The probability of such an arrival P(A) in the opposing lane is 0.049 65. The probability that the passing maneuver will occur during the 11-s critical interval is

$$[(P)\,(11/2)] \times 0.000\,04 = 0.000\,22 \tag{11}$$

The probability that both events will occur and thus cause a conflict is the product of the respective probabilities.

$$P(C) = P(P) \times P(A)$$

$$= 0.000\,22 \times 0.049\,65$$

$$= 1.09 \times 10^{-5} \qquad (12)$$

Over the course of a year, the expected number of conflicts would be 15.6, or about one conflict every 3 weeks. However, this figure is based on total disregard for passing sight distance.

Table 7. Accident costs per year for no control and two-way-stop control at 96-km/h approach speeds.

Facility B ADT	Type of Control	Facility A ADT 100	200	300	400
100	None	151	302	449	598
	2-way stop	452	482	512	542
200	None	302	598	894	1189
	2-way stop	482	965	1024	1083
300	None	449	894	1338	1778
	2-way stop	512	1024	1536	1624
400	None	598	1189	1778	2362
	2-way stop	542	1083	1624	2162

Notes: 1 km/h = 0.621 mph.
Values are in dollars.

Figure 1. Required sight distance triangle for no intersection control.

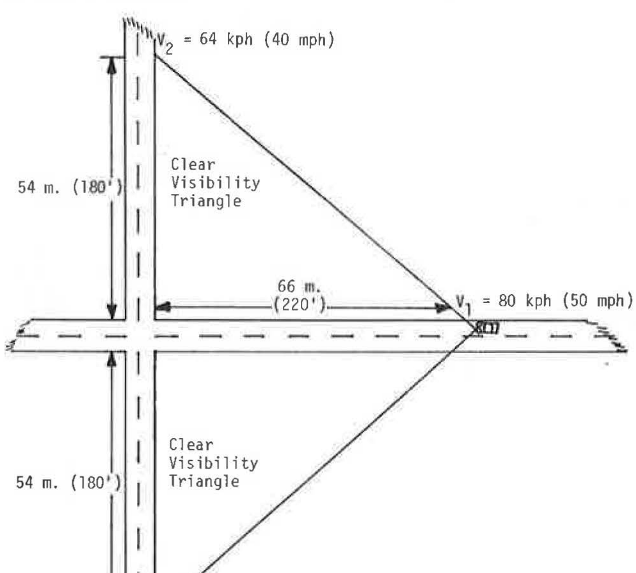


Figure 2. Intersection signing needs diagram.

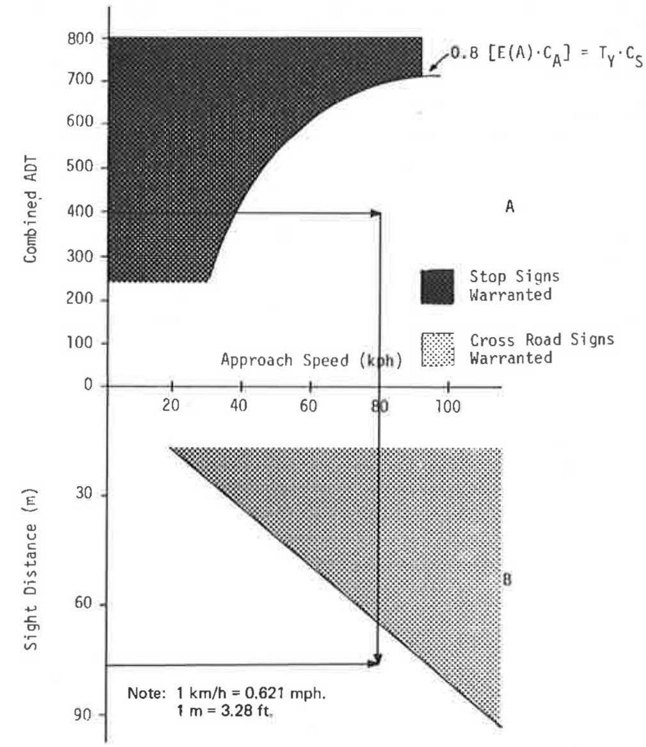


Figure 3. Required deceleration distances on horizontal curves.

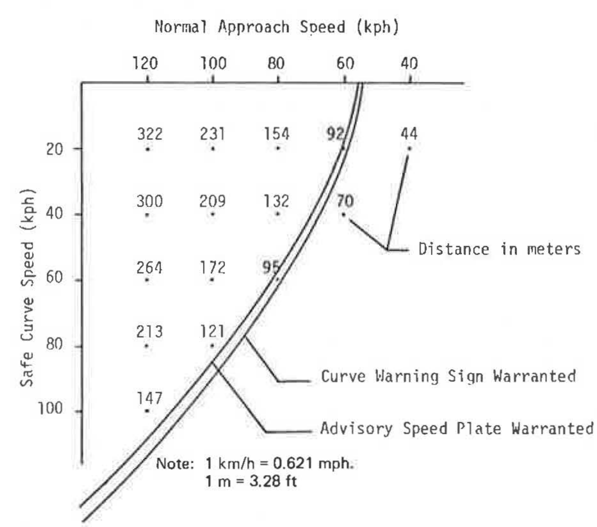


Table 8. Lateral acceleration in gravitational units as a function of forward velocity and type of roadway sign.

Forward Velocity (km/h)	All Curves N	Mean	σ	With Signs N	Mean	σ	Without Signs N	Mean	σ	With Advisory Speed N	Mean	σ	Without Advisory Speed N	Mean	σ
<32	9	0.264	0.055	2	0.280	0.024	7	0.259	0.062	1	0.263	0	8	0.264	0.059
32 to 40	6	0.257	0.070	5	0.270	0.071	1	0.193	0	4	0.268	0.081	2	0.234	0.059
40 to 48	11	0.228	0.061	6	0.257	0.061	5	0.193	0.043	6	0.257	0.061	5	0.193	0.043
48 to 56	16	0.201	0.051	10	0.222	0.053	6	0.165	0.021	10	0.222	0.053	6	0.165	0.021
56 to 64	20	0.212	0.042	13	0.223	0.035	7	0.192	0.048	12	0.224	0.037	8	0.195	0.045
64 to 72	28	0.172	0.051	21	0.183	0.051	7	0.140	0.139	19	0.185	0.053	9	0.146	0.035
72 to 80	35	0.142	0.043	18	0.159	0.037	17	0.124	0.042	13	0.161	0.043	22	0.130	0.039
80 to 88	37	0.129	0.041	4	0.174	0.028	33	0.124	0.040	3	0.169	0.032	34	0.126	0.041

Note: 1 km/h = 0.621 mph.

Assuming that about 30 percent passing sight distance was on our example roadway and that the ordinary prudent driver would take advantage of this visibility, the expected number of conflicts per year is reduced by 30 percent to about eleven. Although this number may seem a bit high to be tolerable, it applies to the worst case—400 vehicles/day and total disregard for safety on sections on inadequate passing sight distance by all drivers. Because a majority of dirvers probably would not attempt a passing maneuver without at least marginal sight distance, the actual number of conflicts is more likely 2 or 3/year. Yet this figure is applicable only for 400 vehicle/day facilities. The average facility examined (about 150 vehicle/day) would produce over the long term only about one conflict every 3 or 4 years.

This analysis indicates that there may be inefficient striping of no-passing zones on low-volume rural roads according to MUTCD requirements. MUTCD-recommended striping might prevent a conflict every few years, but there is no reason to believe that every conflict will result in an accident. Conceivably, a paint stripe would not prevent any accidents throughout the entire life of the paint.

Guidelines

Although the probability of conflict in a passing maneuver has been shown to be minute, the elimination of all signs and markings relative to passing does entail some risk. Yet the degree of risk involved does not appear to justify the expense of standard MUTCD striping. The following alternatives are offered as a substitute for MUTCD striping.

A PASSING HAZARDOUS warning sign should be used to indicate extended sections of inadequate passing sight distance on all unmarked paved roadways and all unpaved roadways. Such signs should have attached to them supplementary plates indicating the length of the section. Subsequent PASSING HAZARDOUS signs and supplementary plates should be erected beyond the intersections with paved roadways. The distances on these subsequent supplementary plates should indicate the number of kilometers remaining in the section from that point.

If centerline definition is desired on paved roadways with insufficient passing sight distance, a double narrow line may be used instead of the PASSING HAZARDOUS signs. The double narrow line consists of two 3.8-cm (1.5-in) yellow lines separated by a 2.5-cm (1-in) space. This line should be used only for extended sections of insufficient passing sight distance; intermittent sections of restricted sight distance within which striping is deemed necessary should be striped according to current MUTCD guidelines. Because vehicle wheel paths on roadways less than 6.1 m (20 ft) wide tend to overlap the centerline and obliterate painted pavement markings, such roadways should not be striped.

SUMMARY

The results of this research indicate that considerable benefit can be derived from a reevaluation of the needs for signs and markings on low-volume rural roads. These benefits include not only obvious monetary savings from reduced levels of signing and marking but also considerable savings in time and frustration on the part of the engineer responsible for the operation of these roadways. Guidelines presented in this paper were developed solely for the rural context and are thus more readily applicable to that environment than are the guidelines offered in the MUTCD. Although the recommendations presented by no means cover all control devices or all situations, they do provide guidance in three most crucial areas—intersections, horizontal curves, and no-passing zones.

1. Low-volume rural intersection control can be efficiently achieved through guidelines based on an economic analysis. Primary variables governing the application of regulatory-warning devices are approach speed, ADT, and sight distance. Below 200 vehicles/day combined entering volume, stop control is inefficient and should not be used except in rare cases. Crossroad signs are advocated for use instead of stop signs at certain locations described in the guidelines.
2. Existing signing practices produce more curve warning signs than are necessary. The guidelines presented describe a more efficient and pragmatic technique for signing of horizontal curves. This reduced level of signing was shown not to adversely affect safety because drivers tended to be more cautious on unsigned curves.
3. Guidelines were developed that are more efficient than existing standards for traffic control in sections of inadequate passing sight distance. Analyses showed that the potential for accidents in no-passing zones is virtually nil on these roadways. Recommendations contained in this paper would virtually eliminate standard striping of no-passing zones and replace that practice with a PASSING HAZARDOUS sign or a more economical double narrow line.

We found, in general, that standard practices for signing and marking of highways are inefficient and unsuited to the rural environment. The recommended guidelines should provide for a much more orderly, pragmatic, and efficient application of control devices on low-volume rural roads.

ACKNOWLEDGMENT

We wish to express appreciation to Carroll J. Messer of the Texas Transportation Institute for his assistance throughout the course of the research. This project was sponsored by the Federal Highway Administration. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration.

REFERENCES

1. Manual on Uniform Traffic Control Devices. Federal Highway Administration, 1971.
2. M. L. Ritchie and others. Choice of Speed in Driving Through Curves as a Function of Advisory Speed and Curve Signs. Human Factors, Dec. 1972, pp. 533-538.
3. S. R. Perkins and J. I. Harris. Traffic Conflict Characteristics—Accident Potential at Intersections. General Motors Research Laboratories, Warren, Mich.
4. D. Burke. Highway Accident Costs and Rates in Texas. Texas Transportation Institute, Texas A&M Univ., College Station, Research Rept. 144-1F, Dec. 1970.
5. D. E. Cleveland. Traffic Control and Roadway Elements—Their Relationship to Highway Safety. Highway Safety Research Institute, Univ. of Michigan, Ann Arbor, 1970.
6. A Policy of Geometric Design of Rural Highways. AASHO, Washington, D.C., 1965.
7. R. W. Winfrey. Economic Analysis for Highways. International Textbook Co., Scranton (now IEP, New York), 1969.

Optimal Control of Isolated Traffic Signals

K.-L. Bång, Swedish Transport Research Commission, Royal Academy of Engineering Sciences

The purpose of the study was to examine the properties of conventional fixed-time (FT) and vehicle-actuated (VA) control of isolated, signalized intersections and to develop and test a self-optimizing control strategy giving special consideration to buses and pedestrians. A discussion of the possibilities to apply the developed methods for coordinated signal control was also to be included.

The project was carried out in five stages involving the following main studies:

1. Criteria for signal control,
2. Literature inventory of control strategies,
3. Development of strategies,
4. Simulation, and
5. Field tests.

An unabridged version of this paper is available elsewhere (1).

CRITERIA FOR OPTIMAL SIGNAL CONTROL

Most existing control methods lack explicit criteria for control. To develop and test new control techniques such criteria must be defined and the effectiveness of the control with respect to the criteria must be estimated.

The overall criterion can be to minimize the total community cost at a given traffic demand. From this, lower order criteria such as minimizing the sum of travel time, vehicle operating costs, and environmental costs caused by the traffic can be derived. The cost function has to include for each category of traffic the number of stops as well as delay time multiplied by appropriate unit cost figures. Operationally, the lower order criteria can be applied through a series of short-term predictions at regular intervals of resulting costs if the green light is extended or changed to another phase. Development of a control strategy along these lines is described in the following sections.

DEVELOPMENT OF A SELF-OPTIMIZING CONTROL STRATEGY

A. J. Miller (2) suggested a simple self-optimizing strategy based on the criterion of minimizing total vehicle delay. In Miller's strategy, the decision to extend a phase is made at regular intervals by the examination of a control function. This function represents the difference in vehicle-seconds of delay between the gain made by the extra vehicles that can pass the intersection during an extension and the loss of the queuing vehicles in the cross street resulting from that extension. The same basic idea has also been used to develop a control strategy within the framework of this project. The method has been named traffic optimization logic (TOL). In the TOL method, the extension of the green light is based on calculations at regular intervals h of a control function Φ. This function represents the gain or loss in community cost resulting from extension of the prevailing green light with h s. Figure 1 shows this method in the form of a flow chart.

ESTIMATES OF THE CONTROL VARIABLES FOR PRACTICAL APPLICATION

The TOL method requires that all approaches of the intersection be continually surveyed. Two methods to obtain the necessary traffic information have been tested in the field studies:

1. Derivation from passage detectors situated 30 and 120 m from the stop lines and
2. Direct observation of the analog output from long loop detectors encircling each lane in the approaches. The long loops give an output signal that is roughly proportional to the number of cars within the loop. The loops are divided into segments to provide information on vehicle positions (Figure 2).

Buses are identified by using selective detector equipment with a small passive unit in the bus and loops embedded in the pavement. Pedestrians cannot yet be quantitatively detected although simulations assuming that this can be done have been performed.

SIMULATION

The simulation model used for the studies is a further

Figure 1. Flow chart showing basic principle of TOL.

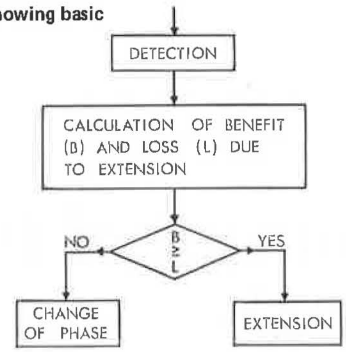


Figure 2. Test site detector installations.

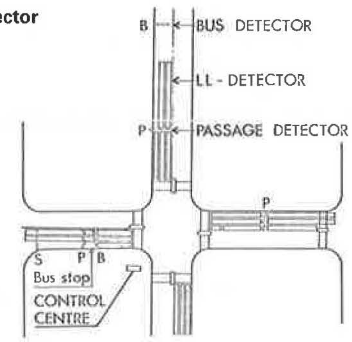
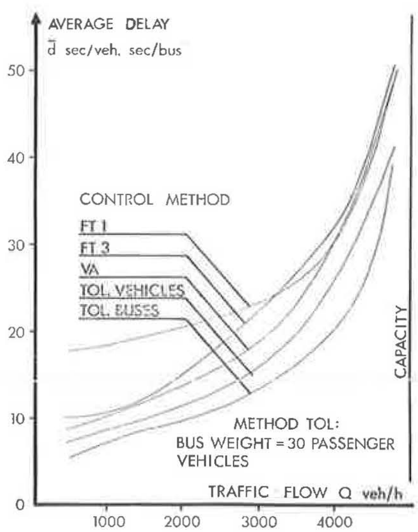


Figure 3. Simulation results for two-phase, four-way intersection.



development of a model presented by Klijnhout (3). The model is event scanning and is written in PL/1. Separate programs are developed for the different control strategies.

The simulation results indicated that TOL was considerably more effective than conventional VA and FT control (Figure 3). The vehicles should preferably be detected far in advance of the intersection (100 to 200 m), and the h between the calculations of Φ should be kept as short as possible.

Simulations were also performed that assumed that the pedestrians could be quantitatively detected and considered in the control function. The pedestrian delay was reduced by 10 to 15 percent without significant increases of the vehicle delay.

FIELD TESTS

Field tests comparing TOL with FT and VA were performed in two intersections in Stockholm. A minicomputer (PDP 11/05) mounted in a mobile van served as

Figure 4. Field test results, average delay versus traffic flow.

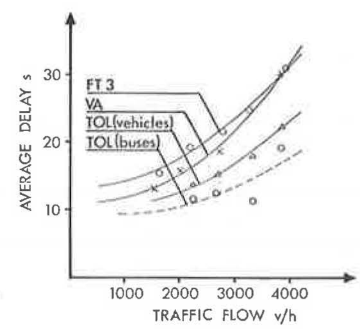


Figure 5. Field test results, proportion of stopped vehicles versus traffic flow.

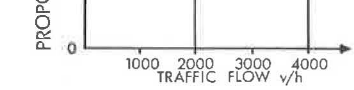


Table 1. Field test results.

Variable	Traffic	Control Mode		
		FT	VA	TOL
Traffic flow, h	Automobiles	2600	2740	2620
	Buses	18	12	20
Average delay, s	Automobiles	21.5 ± 0.9	20.2 ± 3.0	15.5 ± 0.4
	Buses	20.4 ± 2.0	20.0 ± 3.0	12.2 ± 1.2
Proportion stopped, %	Automobiles	68	67	59
	Buses	63	62	55

Note: Limits refer to the 95 percent confidence interval.

Table 2. Benefit/cost ratios for replacement of VA by TOL control.

Traffic	Factor Considered	Benefit/Cost Ratio
Buses	Reduced operational costs	0.13
	Reduced time costs	2.42
	Total	2.55
Automobiles	Reduced operational costs	3.15
	Reduced time costs	7.10
	Total	10.25

Note: Total reduced operational costs for both buses and automobiles = 3.28. Total of all reduced costs for both buses and automobiles = 12.80.

both signal controller and data recording unit during the field experiments. TOL is likely to require a minicomputer or microcomputer even for regular operational use because of the complex control function.

The data collection was performed automatically by using information of signal status, vehicle detector passage times, and the analog detector signals as input. Based on these data, all relevant variables, such as delay, stops, green times, and the like, were derived and recorded on tape. The data on the tapes were later further reduced by a special computer program for statistical analysis.

Examples of results from the field tests are shown in Figures 4 and 5. Figure 4 shows the relationship between average delay and total intersection traffic flow. It shows that TOL gave considerably lower delay values than the other methods did for all traffic volumes tested. Furthermore, it gave an extra reduction of the bus delay. Compared to VA, TOL gave a 20 to 25 percent improvement for the ordinary vehicles and a 20 to 40 percent improvement for the buses. The differences between the curves concerning the vehicles are significant at the 0.05 level.

Table 1 gives the field test results for the whole test period. The differences in mean delay for vehicles are significant at the 0.01 level. The difference for bus delay between TOL and the other methods is also significant. The improvement obtained by replacing VA with TOL has been shown to be cost effective even if only the reduction of vehicle operating costs (largely energy consumption) is considered (Table 2).

COMMENTS

The results of the studies indicate strongly that the TOL strategy when compared to conventional FT and VA control gives substantial reductions of average delay and proportion of stopped vehicles. Further improvements can be given to the buses if they are weighted higher than the other vehicles. The TOL strategy can also be applied for coordinated signal control of nearby intersections. In this case, information on queues and discharge rates is transmitted from each controller to the nearby intersections that are still independently controlled. This method should result in a very flexible type of coordination and offer good possibilities for individual bus priority. A research program to test this method is under way in Sweden.

REFERENCES

1. K.-L. Bång. Optimal Control of Isolated Traffic Signals. Traffic Engineering and Control, Vol. 17, No. 7, July 1976.
2. A. J. Miller. A Computer Control System for Traffic Networks. Proc., 2nd International Symposium on Theory of Traffic Flow, London, 1963.
3. J. J. Klijnhout. A Digital Simulation Model for Single Intersections With Traffic Lights. Traffic Engineering and Control, Vol. 13, No. 4, Aug. 1971.

Inductive Loop Vehicle Detector: Installation Acceptance Criteria and Maintenance Techniques

James W. Ingram, Division of Construction and Research, California Department of Transportation

The field of vehicle detection is filled with design pitfalls and maintenance frustrations. This paper represents only one step toward relieving the pressures on the technician and the traffic engineer. Emphasis is on enabling the traffic engineer to predict during the design phase whether the system will work with available loop detector amplifiers. Emphasis is also on outlining an exact method of evaluating the sensitivity of a loop and lead-in system before the loop detector amplifier is attached to the system.

INSTALLATION DESIGN CRITERIA

When critical, sensitivity may be calculated for loop and lead-in systems during the design phase to verify that the loop and lead-in system will function with available amplifiers.

Procedure

The percentage of change in inductance for the worst case vehicle is calculated for the system in question. This is compared to the sensitivity threshold of the amplifier selected or specified.

Example

According to Table 1 and Figure 1, the percentage of change in inductance $\%\Delta L$ due to the passage of a Honda 100 motorcycle over the center of a 4-series-parallel connected-loop system with 183 m (600 ft) of lead-in is 0.06 percent and the loop inductance is 80 μH. The lead-in inductance is about 130 μH [0.72 μH/m $\times$ 183 m (0.22 μH/ft $\times$ 600 ft)]. If we use the information on lead-ins from Table 1,

$$\%\Delta L = 0.06 \times [80/(80 + 130)] = 0.023 \tag{1}$$

Because the loop has been installed within 2.5 cm (1 in) of the road surface, no reduction in sensitivity need be accounted for. Because the loop amplifiers purchased in California are tested for a 0.03 percent ΔL high sensitivity threshold, the 0.023 percent ΔL may well not result in an output from brand X or Y amplifier. Brand Z may advertise a high sensitivity threshold of 0.02 percent or less, and it may work, but a better approach might be to split the four loops into two sets of two and provide a second lead-in and amplifier. If we use the same tables and procedures, the $\%\Delta L$ for each of the two systems (two 1.83-m-square (6-ft-square) series connected loops with 183 m (600 ft) of lead-in) is found to be 0.046. This is a comfortable margin of sensitivity.

INSTALLATION ACCEPTANCE CRITERIA

Because it was not practical for every jurisdiction to use the same vehicle as a standard for the acceptance of new loop detector installations, choosing an easily fabricated and conveniently carried device that might closely model a vehicle seemed desirable. Such a device might well model the worst case vehicle (that vehicle most difficult for the system to detect, such as a small motorcycle). One such device is a shorted loop of wire mounted around the edges of a 0.61-m-square (2-ft-square) piece of plyboard. Although this device may well be used to test the operation of a loop detector system (system equals loop + lead-in + amplifier), testing the sensitivity of the loop and lead-in portion of the system separately from the amplifier may be more desirable. A second device is required, one that will measure the loop and lead-in sensitivity to the vehicle model. This device, when attached to the loop and lead-in, would measure $\%\Delta L$ due to the presence of the vehicle model. Installation acceptance testing, which includes the measurement of loop resistance, insulation resistance, and inductance, should also include a test for the sensitivity of the loop and lead-in system to a "standard" vehicle or vehicle model.

Procedure

With the loop detector amplifier disconnected, the test device (if equipped with a standard amplifier connector) may be plugged directly into the harness. If the inductance change measuring device is a loop oscillator-frequency counter, the frequency of oscillation (f_1) is recorded. With the vehicle model placed in the center of the loop [±15 cm (6 in)], the new frequency (f_2) is recorded. The percentage of change in frequency %Δf is then calculated according to the formula $100\ [(f_2 - f_1)/f_1]$. It can be shown that %ΔL is approximately equal to 2 times %Δf. This is a good approximation for ΔL up to 10 percent and for loops where Q, the quality factor, is greater than 5. %ΔL is then calculated by multiplying the results by 2. The result is then compared with the data in Table 1.

Table 1. Expected values of change in inductance due to a Honda 100 or the vehicle model.

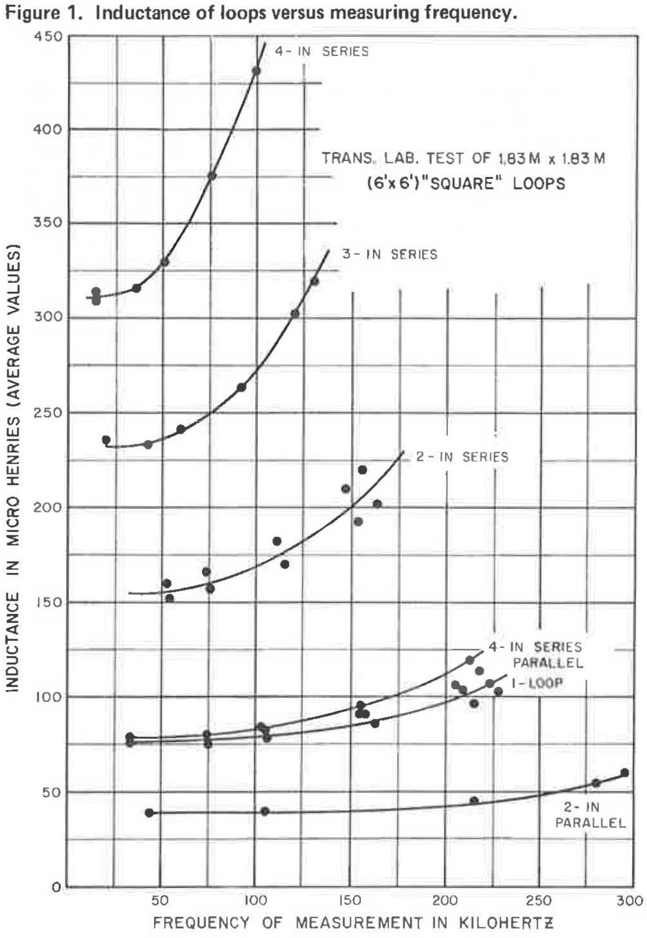
Expected Inductance (μH)	Valid Range of Measuring Frequency (kHz)	Expected Change in Frequency (%)	Change in Inductance (%)	Loop Configuration
79	34 to 100	0.12	0.12	1 loop
158	24 to 75	0.06	0.12	2 in series
40	47 to 150	0.06	0.12	2 in parallel
235	19 to 50	0.04	0.08	3 in series
80	34 to 100	0.03	0.06	4 in series parallel connected
312	10 to 30	0.03	0.06	4 in series

Notes: All values may vary by +10 percent.

All loops are 1.83 by 1.83-m square (6 by 6-ft) 3-turn loops made with 20 metric gauge (12 AWG) wire and less than 4.6 m (15 ft) of lead-in.

Lead-in inductance is generally 0.72 μH/m (0.22 μH/ft). Adding lead-in will reduce the percentage of change by the ratio of loop inductance to loop plus lead-in inductance. Thus the table values for percentage of change in inductance or frequency must be multiplied by this factor when lead-in lengths exceed 4.6 m (15 ft). A further reduction of about 2.4%/cm (6%/in) results in a burial depth greater than 2.5 cm (1 in). Thus a loop buried at a depth of 7.5 cm (3 in) below the surface will have an 18% reduction in both frequency and inductance change values.

Figure 1. Inductance of loops versus measuring frequency.



Example

A new detector installation, consisting of four 1.83-m-square (6-ft-square) series-parallel connected three-turn loops and 76 m (250 ft) of lead-in, is being tested for acceptance. The following measurements are taken:

1. Frequency of loop oscillator with no vehicle on the loop (f_1 = 57 994 H_z) and
2. Frequency of loop oscillator with vehicle model in center of loop (f_2 = 58 005 H_z).

%ΔL is calculated

$$100[(f_2 - f_1)/f_1] = 11/57\,994 = 0.019 \text{ percent } \Delta f \quad (2)$$

Then

$$\%\Delta L = 2 \times \%\Delta f = 2 \times 0.019 \text{ percent} = 0.036 \text{ percent } \Delta L \quad (3)$$

Table 1 predicts the following: for the four series-parallel connected loops, %ΔL = 0.06. The reduction due to 76 m (250 ft) of lead-in is calculated according to the notes in Table 1:

$$\%\Delta L = 0.06[80/(80 + 55)] = 0.36 \quad (4)$$

The measured value is 5 percent higher than the table predicted; the sensitivity of this loop is acceptable.

MAINTENANCE TECHNIQUES

The loop-oscillator, frequency-counter test device will enhance maintenance capability.

More detailed information on maintenance techniques is available elsewhere (1) and describes a loop-oscillator, frequency-counter test device and gives procedures and data so that the device can be used to isolate problems with loop detector systems. These procedures would enable the signal technician to

1. Evaluate the condition of a loop detector system at any point in its life cycle (preventative maintenance);
2. Predict the failure of a system before it fails;
3. Isolate failures to the loop, lead-in, amplifier, or splices;
4. Eliminate crosstalk problems; and
5. Determine the cause of intermittent behavior.

REFERENCE

1. J. W. Ingram. The Inductive Loop Vehicle Detector Installation Acceptance Criteria and Maintenance Techniques. Division of Construction and Research, California Department of Transportation, Sacramento, Jan. 1976.

Traffic-Responsive Ramp Control Through the Use of a Microcomputer

B. C. Fong and R. L. Donner, California Department of Transportation

Traffic-responsive ramp control has been proved effective in the reduction of traffic congestion on major metropolitan freeways. Equipment used at most of the existing traffic-responsive ramp-control installations involved the use of large-scale process computers and microcomputers, which are not the most cost-effective alternatives. A low-cost ramp controller is needed that can serve traffic-responsive ramp controls either for single-ramp local operations or for operations involving a series of ramp locations that link to a central computer for multilevel controls.

Since 1974, the California Department of Transportation has been conducting researches and evaluations on the application of microcomputer for traffic-responsive ramp controls. The department found that recent advancements in the large-scale integration metal oxide semiconductor technology have made possible many applications for microcomputers. The second-generation microcomputers have sufficient capabilities to replace large-scale computers and minicomputers in most traffic control functions. The attractive features of microcomputers include low cost, small physical size, and operability within the ambient temperature range of -18°C to 60°C (-27.78°F to 15.56°F) without air conditioning. The physical size of a typical microcomputer chip is approximately 51 by 16 by 4 cm (2 by 0.6 by 0.15 in), and a read-only-memory chip with 1024 by 8 bits of storage capacity is approximately 28 by 10 by 2 cm (1.1 by 0.4 by 0.1 in). A simple set of microcomputer chips can now be bought for less than $100.

The California Department of Transportation awarded the first contract to Honeywell, Inc. for the manufacture of 200 Type 140 Controllers based on a design developed by the department. The Type 140 Controllers employ a Motorola M6800 microprocessor chip, 2048 by 8 bits of programmable read-only-memory (PROM) and 1024 by 8 bits of complimentary metal oxide semiconductor (CMOS) random access memory (RAM). The M6800 microprocessor can address up to 65 000 by 8 bits of memory and input-output locations. Instructions set includes functions for data transfers between working registers and with memory locations; logical and arithmetic computations on contents of working registers; increment, document, and rotate working registers; conditional and unconditional branchings; and input and output of data from and to peripheral devices. The minimum instruction execution time is approximately 2 μs. The PROM chips are manufactured by Intel by means of the silicon gate process. Each of the Intel 2708 PROM chips can store 1024 by 8 bits of information and is field programmable and erasable. The CMOS RAM is used for temporary storage of traffic data and calculation results. Backup battery power is provided in the controller to prevent loosing of the RAM content in the event of commercial power failure.

Ramp-control strategies and programs are presently being developed by the State. The Type 140 controllers are currently being used in the Los Angeles, San Diego, and San Francisco areas for local and multilevel ramp-control operations. Use of the microcomputer probably will become standard on California highways for both ramp and intersection controls, especially as a result of the growing emphasis on upgrading the operation of existing facilities rather than the building of new facilities.

Speed Reduction in School Zones

Charles V. Zegeer, James H. Havens, and Robert C. Deen, Bureau of Highways, Kentucky Department of Transportation

The use of flashing beacons together with signing has become somewhat standard throughout the country to alert drivers to the presence of school children and to regulate vehicle speed in school zones. Yellow beacons, usually two flashing alternately, may be used with both warning signs and regulatory signs. The only regulatory signs related to school zones are speed limit signs. Both hazard identification beacons and speed limit sign beacons are intended to operate only during hours when the warning and speed regulations are in effect. The effectiveness of signs and flashing lights in reducing speed in school zones has been questioned.

The purpose of this study was to determine the effectiveness of flasher beacons in reducing vehicle speeds in Kentucky. Speed measurements were made during flashing and nonflashing periods at 48 locations. The physical characteristics of each site were identified and compared to speed reductions. A large sample of flashers (120 of 424 school flashers currently maintained by the Bureau of Highways in 33 counties in central, northern, and northeastern Kentucky) was inspected to ascertain condition and operation. This information was helpful in determining the reliability of the beacons in everyday operation.

In Kentucky, pedestrians between the ages of 5 and 9 represent less than 10 percent of the total population but account for more than 16 percent of all pedestrian fatalities. This percentage exceeded all other age groups (1). Of the 167 pedestrian deaths in 1973, 27 were child fatalities (5 to 9 years old). Approximately 600 children pedestrians (5 to 14 years old) were injured in Kentucky by motor vehicles.

Fourteen findings and conclusions were based on analysis of physical and geometric features of the sample locations.

1. Speed reductions attributable to flashers were statistically significant at the 95 percent level at 84 percent of the locations; the average speed reduction was 5.8 km/h (3.6 mph). Seventy-one percent of the locations showed speed reductions of less than 6.4 km/h (4 mph). Only two locations yielded speed reductions of more than 16.1 km/h (10 mph).
2. The 85th percentile speeds decreased by about 8.0 km/h (5 mph) for all locations. The higher speed locations had lower reductions [3.2 km/h (2 mph)] than the low-speed locations [6.4 km/h (4 mph)].
3. The 85th percentile speeds at all locations during flashing periods exceeded the 40.2-km/h (25-mph) limit by about 30.6 km/h (19 mph).
4. Uniformity of driving speeds was the same at low-speed [40.2 to 56.3 km/h (25 to 35 mph)] and medium-speed [57.9 to 72.4 km/h (36 to 45 mph)] locations whether the flashers were on or not. However, at high-speed locations [74 to 88.5 km/h (46 to 55 mph)], a 15 percent drop of vehicles in the 16.1-km/h (10-mph) pace was noted, which indicates that the intervehicle accident potential is increased when the flashers are on.
5. Crossing guards contributed to a drop in vehicle speeds of about 14.5 km/h (9 mph), and the average speeds were under 40.2 km/h (25 mph) at four of the five locations. Without the crossing guards at these same locations, the speed reduction averaged only 4.34 km/h (2.7 mph). Crossing guards were stationed at about 10 percent of all locations.
6. Regular speed enforcement in school zones by police agencies caused average speed reductions of 13.5 km/h (8.4 mph) at seven locations.
7. Speed reductions at high-speed locations were slightly higher than at other locations. However, the average speeds exceeded the 40.2-km/h (25-mph) limit by about 29.0 km/h (18 mph) at high-speed locations compared to 15.6 km/h (9.7 mph) and 6.8 km/h (4.2 mph) at medium- and low-speed locations respectively. Only 8 percent of the vehicles traveled below the speed limit when flashers were not operating.
8. Pedestrian volumes (increasing from 50 to 400/day) in the school zones contributed to a slight decrease in vehicle speeds [3.2 km/h (2 mph)]. Also school bus volumes (increasing from 0 to 32 buses/day) contributed to a slight decrease in vehicle speeds [about 3.2 km/h (2 mph)].
9. Highway width did not appear to affect speed reductions. Short sight distances between motorists and school flashers contributed to the ineffectiveness of flashers at five locations.
10. Average decreases in speed of less than 1.6 km/h (1 mph) during flashing periods were attributed to traffic

volume increases at only two locations.

11. Signalized or stop-sign intersections adjacent to or between school flashers resulted in virtually no speed reductions in 4 of 5 such locations. Excessively long flashing periods at 10 locations resulted in speed reductions of less than 4.2 km/h (2.6 mph). School flashers at 3 locations, with a recent history of inappropriate flashing, yielded an average speed reduction of only 2.7 km/h (1.7 mph).

12. Several flasher installations were not warranted because of low pedestrian volumes and low vehicle speeds and volumes. A few continually flashing lights were also found.

13. Nearly all school flasher locations have favorable as well as unfavorable features that contribute to driver compliance or noncompliance with the 40.2-km/h (25-mph) speed limit. A single, significant defect can render the flasher ineffective.

14. About 14 percent of the school flashers were defective or malfunctioned. Major malfunctions included inoperative clocks and defective bulbs or fuses. Other deficiencies included flashers mounted among commercial signing, obstructed view, deteriorating signs, worn pavement markings, nonuniform signs, and erratic flashing periods.

REFERENCE

1. C. V. Zegeer. Pedestrian Accidents in Kentucky: 1972-1973. Bureau of Highways, Kentucky Department of Transportation, March 1975.

Abridgment

Evaluation of Raised Pavement Markers for Reducing Incidences of Wrong-Way Driving

Frank D. Shepard, Virginia Highway and Transportation Research Council

In an attempt to stop drivers who enter an interchange ramp going the wrong way, a means of alerting them to their error is being sought. In view of the fact that wrong-way drivers must fail to see or properly interpret the directional signs, warning signs, and pavement markings placed in the intersection for their guidance, something beyond conventional devices is obviously needed. A concept that is believed to have merit involves the placement of raised pavement markers on off-ramps in such a configuration that the driver will be alerted as a result of viewing an unexpected phenomenon. Although such markers have been used for this purpose, they have been placed in the shape of an arrow, transverse line, or other configurations similar to markings normally seen by the motorist. The preliminary research reported in this paper was undertaken to determine the efficacy of random configurations of such markers.

The investigation was limited to one Interstate interchange, and only an off-ramp under night conditions was considered. Also only one type of raised pavement marker was considered for testing—a corner cube monodirectional red marker. The markers, which possess good reflective qualities, were placed to reflect only the light of a vehicle traveling in the wrong direction. Forty-five markers were randomly placed in a section 36.58 m (120 ft) long starting approximately 32.00 m (105 ft) from the end of the ramp and extending 3.05 m (10 ft) past pre-existing wrong-way signs. With this placement, the motorist could turn completely into the off-ramp before crossing the marked section. Also, it was felt that termination of the marked section in the vicinity of the wrong-way signs would help call attention to them.

The evaluation was a subjective one concerned primarily with the visibility characteristics, or attention-getting qualities, of the configuration. Each of 16 test subjects was shown the experimental installation from a vehicle (lights on high beam and low beam) that entered and proceeded onto the off-ramp only a short distance before backing out. Before viewing the markers, subjects were told only that their opinion of some experimental materials was desired and that they should not be alarmed if certain unexpected maneuvers were made. Note was made of the initial opinions and reactions of each subject relative to the effectiveness of the marking system. Questions were then asked concerning their thoughts about the effectiveness of the system in preventing wrong-way entries together with any thoughts on the number of markers, the shape of the configuration, and the location of the configuration with respect to signs.

The results of the investigation showed that the raised pavement marking system was effective in alerting drivers because they viewed an unexpected phenomenon. Also, it was thought that the marking system did help to call attention to the wrong-way signs and that it would be effective in causing a wrong-way driver to realize his or her mistake and act accordingly. Based on the finding of this initial work, the system has been implemented at two sites for further study.

A degree of bias can be expected when testing subjects in the manner described; however, because this research was intended as a first step in determining the feasibility of using raised pavement markers to alert wrong-way drivers, such a subjective evaluation was deemed to be appropriate.

The effect of the marking system on intoxicated or drowsy drivers cannot be surmised from the results of this evaluation, nor can the reactions of passengers to the markings be inferred. However, if only a small number of subjects, all of whom thought the system was effective, would have been prevented from going the wrong way, implementation of the system should be seriously considered because of its simplicity and low cost.

Abridgment

Platoon Dispersion Characteristics on One-Way Signalized Arterials

Robert L. Vecellio, Department of Civil Engineering, Auburn University

The overall objective of this study was to investigate platoon movements on urban arterials and to relate variation in platoon behavior to variation in signal control and changes in traffic volume.

RESEARCH PROGRAM

The research efforts were structured around four principal phases: (a) review of the literature, (b) collection and reduction of basic data on platoon movement, (c) identification of platoon dispersion characteristics, and (d) model development.

The first phase of research established that the research approach adopted for this investigation is unique; it differs in two important aspects from all previous studies of platoon behavior. These were the collection of comprehensive data on platoon movement by using a helicopter-mounted aerial camera and use of extensive urban signalized arterials as study sites.

The second phase of research consisted of the collection and reduction of continuous velocity and spacing data as platoons of traffic traveled through progressive signal systems during peak hours. Two study sites, each consisting of nine fixed-time signalized intersections, were selected in the Columbus, Ohio, area. Spacing between signals ranged from approximately 107 to 747 m (350 to 2450 ft). For the platoons photographed, vehicle trajectories were constructed to provide visual representations of traffic movement. A total of approximately 28 000 time-space positions were determined and served as the sample data for this study.

The third phase of research consisted of identifying platoon characteristics for traffic traveling on the signalized arterials. It was established that improper signal offsets, the presence of initial queues at interior signalized intersections, and high frequency of lane changes at a specific location can cause inefficiency in the operation of a progressive signal system.

The principal variables affecting platoon movement through linear signal systems were identified as signal spacing, signal offset, and platoon size. It was established that lane of travel exhibits no significant effect on the behavior of platoons traveling from signal to signal. Finally, it was determined as a result of viewing time and space patterns of selected traffic variables that platoon movement can best be described by patterns of mean velocity or mean spacing, traffic density, and the coefficient of variation of velocity.

The final phase of research involved the development of a mathematical model to simulate the behavior of a group of vehicles progressing through a series of signalized intersections. Written in the IBM simulation language, GPSS/360, the model assumptions include passenger car movement, no turning traffic, no entering traffic from adjacent lanes, and no consideration of signal visibility as a factor affecting platoon behavior. These assumptions were made because the study sites for which two models were specifically developed reflected these conditions.

To apply the model to a specific one-way signalized arterial, some field data are necessary to estimate certain variables. The following input is required before the model can be implemented: signal timing at each intersection, signal offset between intersections, distances between signalized intersections, saturation flow at each signalized intersection, lost time at each signalized intersection, storage capacity between signalized intersections, arrival distribution of traffic at the initial signal, and travel-time parameters between signalized intersections. Regression equations were developed to relate the mean and standard deviation of travel-time distributions to signal spacings and signal offsets for a given level of traffic volume.

The model can be used to generate queue, delay, and travel-time characteristics for traffic traveling through the simulated street system. Statistical agreement between observed and simulated queue length distribution for the two study sites revealed the model to be an adequate representation of traffic operations through a signalized arterial.

CONCLUSIONS AND RECOMMENDATIONS

As a result of this research, a method of timing a linear system of signals along a one-way street allowing for the dispersion of traffic is available. Also the effect of a change in linear signal system timing on expected queue lengths and mean delays per vehicle can be predicted.

As a result of this completed investigation of platoon dispersion characteristics, the following recommendations for future research are made: (a) application of the model to the study of additional one-way signalized arterials having different signal spacings and offsets from those of the sites analyzed in this project, (b) generation of a family of regression equations relating travel-time parameters to signal spacings and signal offsets for a variety of traffic volume levels, and (c) a comprehensive sensitivity analysis of all variables incorporated in the model to precisely identify the effect of a change of one model variable on model output.

ACKNOWLEDGMENTS

The research reported in this paper was sponsored by the Ohio Department of Transportation in cooperation with the Federal Highway Administration while I was associated with the Transportation Research Center of Ohio State University.

Progressive Signal System in a Network of Arterial Streets

Tapan K. Datta, Wayne State University
William C. Taylor, Michigan State University
David M. Litvin, Goodell-Grivas, Inc.

Computer-aided systems to coordinate traffic signals on urban roadways have been developed to achieve a smooth flow of traffic uninterrupted by the red phase of the signals. Most of these systems, however, have neglected one or more of the important traffic variables and their interrelationships and therefore may produce timing patterns, which, when implemented under certain traffic conditions, may not reduce vehicle stoppages and delays.

A new model, the traffic signal optimization model (TRASOM), has been developed that considers the three basic traffic variables: speed V, volume Q, and density K. The model allows different road segments to have different Q-K-V relationships because traffic flow depends not only on the geometrics of the road but also on the environment and traffic conditions. Because the Q-K-V relationship defines roadway capacity as a function of speed, this relationship is an essential factor in the proper design of a progressive signal system. The effect of this relationship on the design of the optimum timing plan becomes increasingly critical as the volume increases. By using a characteristic V-Q relationship as an added constraint in determining a progressive speed, one can produce a solution that is consistent with observed flow characteristics and that avoids those solutions that are infeasible.

TRASOM has two major advantages: (a) a guarantee of a feasible product and (b) a method for measuring and expressing incremental improvements in system performance. The variables and constraints used in this model include the following:

1. Independent variables (traffic arrival rate and distribution, lane usage, roadway geometrics and conditions, and service rates and function);
2. Dependent variables (traffic throughput, progression speed, concentration, total and average delay, queue characteristics, and system efficiency);
3. Control variables (cycle lengths, cycle splits, and offsets); and
4. Imposed constraints (pedestrian crossing time, maximum and minimum red time, V-Q red time, and maximum and minimum cycle lengths).

The objective of TRASOM is to search for the combination of control variables that results in the optimum set of dependent variables within the given constraints. The measures of performance in this decision-making process may include percentage of throughput, progression speed, travel time, average and total delay, queue characteristics for left-turning traffic, or any combination of such variables. The minimization of average network delay is an objective function used in the model.

TRASOM first determines optimal linear solutions for all the roadways constituting the network and then fits in the intersecting nodal offsets according to a sequential strategy. The linear optimal solution thus establishes the optimal progressive traffic flow on each roadway rather than obtaining flow through individual links at various progression speeds.

For system optimization, each linear system in the street network is rank ordered on a priority system. The assigned priorities do not change the cycle splits; however, they do establish the sequence used in determining feasible network solutions. A linear system with no or only one higher priority system crossing it retains the optimal timing plan it would have had as a single street. Any linear street crossed by two or more streets with higher priorities may not retain its optimal linear solution because new solutions are obtained for these intersections that treat the offsets as fixed. If, following the network analysis, there is no feasible solution for one or more linear streets for a particular cycle length and set of constraints, there will be no solution produced for the entire network for that cycle length unless specifically requested by the analyst.

A feasible solution for each cycle length tested is printed for the convenience of the analyst and is followed by summary statistics or expected traffic performance characteristics produced by performing macrosimulation and other analysis based on the designed cycle length, offsets, and splits. The output format identifies the signal numbers, names of cross streets, distances between intersections, approach volumes, and rank-ordered priority of all the cross streets for each linear system. Summary statistics include expected left-turning queue length for all approaches, average delay at all approaches, average number of stopped vehicles at each intersection and average percentage of throughput on the main street. The optimum network solution for a given demand is then determined by comparing the system attributes for the optimal solution at each cycle length.

TRASOM requires the following input data: distance between intersections, intersection counts, lane usage, turning movements, roadway capacity, speed and volume data at selected sections of roadway, and roadway geometrics. Any special features, such as multiphase signals, special left turn phase, and special turning prohibitions, can also be incorporated into the model.

The TRASOM program is written in FORTRAN IV and is capable of describing timing patterns for various network sizes. Recent projects have ranged from a 20-signal linear street to a 93.24-km^2 (36-$mile^2$) area containing 200 signals. Compared to other computer-aided systems, TRASOM has simple data requirements, an output that is easy to interpret, and small implementation costs.