TRANSPORTATION RESEARCH RECORD 737

Traffic Control Devices, Geometrics, Visibility, and Route Guidance

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1979

Transportation Research Record 737
Price $6.60
Edited for TRB by Amy E. Shaughnessy and Brenda J. Vumbaco

modes
1 highway transportation
3 rail transportation

subject areas
52 human factors
54 operations and traffic control

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

Notice
The papers in this Record have been reviewed by and accepted for publication by knowledgeable persons 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 these papers are those of the authors and do not necessarily reflect those of the sponsoring committee, the Transportation Research Board, the National Academy of Sciences, or the sponsors of TRB activities.

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Traffic control devices, geometrics, visibility, and route guidance.

(Transportation research record; 737)
1. Electronic traffic controls–Addresses, essays, lectures.
2. Roads–Lighting–Addresses, essays, lectures. 3. Roads–Interchanges and intersections–Design and construction–Addresses, essays, lectures. 4. Traffic safety–Addresses, essays, lectures.
I. Title. II. Series.
TE7.H5 no. 737 [TE228] 380.5s [625.7'94] 80-17543
ISBN 0-309-02992-9

Sponsorship of the Papers in This Transportation Research Record

GROUP 3–OPERATION AND MAINTENANCE OF TRANSPORTATION FACILITIES
Adolf D. May, University of California, Berkeley, chairman

Committee on Traffic Control Devices
Ken F. Kobetsky, West Virginia Department of Highways, chairman
H. Milton Heywood, Federal Highway Administration, secretary
John L. Barker, Robert L. Bleyl, Edward L. Cook, Charles E. Dare, J. R. Doughty, Thomas J. Foody, Paul H. Fowler, Robert M. Garrett, Alan T. Gonseth, Robert L. Gordon, Gerhart F. King, Kay K. Krekorian, Hugh W. McGee, Robert D. McMillen, Joseph A. Mickes, Zoltan A. Nemeth, Harold C. Rhudy, John C. Rice, David G. Snider, Ronald E. Stemmler, James I. Taylor, James A. Thompson, Thomas L. Vanvechten, Clinton A. Venable, Earl C. Williams, Jr., Walter P. Youngblood, Jason C. Yu

Committee on Visibility
Nathaniel H. Pulling, Liberty Mutual Research Center, chairman
Samuel P. Sturgis, Liberty Mutual Research Center, secretary
Albert Burg, Charles W. Craig, Warren H. Edman, Ralph A. Ehrhardt, Eugene Farber, W. S. Farrell, T. W. Forbes, Vincent P. Gallagher, S. A. Heenan, Robert L. Henderson, Charles H. Kaehn, Antanas Ketvirtis, Lee Ellis King, Ken F. Kobetsky, Ralph R. Lau, Richard N. Schwab, Richard E. Stark, Frederick E. Vanosdall, Ned E. Walton, Earl C. Williams, Jr., William L. Williams, Henry L. Woltman, Kenneth Ziedman

Committee on Railroad-Highway Grade Crossings
Otto F. Sonefeld, Atchison, Topeka and Santa Fe Railway Company, chairman
Howard C. Palmer, Safetran Systems Corporation, secretary
Charles L. Amos, William D. Berg, Lucien M. Bolon, Archie C. Burnham, Jr., Louis T. Cerny, Janet A. Coleman, Bruce F. George, Bruce Lanier Gordon, William J. Hedley, John Bradford Hopkins, Robert C. Hunter, Robert A. Lavette, Richard A. Mather, G. Rex Nichelson, Jr., Hoy A. Richards, Eugene R. Russell, C. Shoemaker, Max R. Sproles, Barry M. Sweedler, Richard A. Wiita

Committee on Operational Effects of Geometrics
Stanley R. Byington, Federal Highway Administration, chairman
Robert B. Helland, Federal Highway Administration, secretary
Frank E. Barker, Robert E. Craven, Edwin W. Dayton, J. Glenn Ebersole, Jr., Daniel B. Fambro, Julie Anna Fee, John Feola, John C. Glennon, George F. Hagenauer, Rajendra Jain, Janis H. Lacis, William John Laubach, Jr., William A. McConnell, Woodrow L. Moore, Jr., Robert P. Morris, Thomas E. Mulinazzi, Sheldon Schumacher, James J. Schuster, Robert B. Shaw, C. Robert Shinham, Bob L. Smith, Robert C. Winans

Committee on Traffic Flow Theory and Characteristics
Kenneth W. Crowley, Pennsylvania State University, chairman
Edmund A. Hodgkins, Federal Highway Administration, secretary
Charles R. Berger, Said M. Easa, John W. Erdman, Antranig V. Gafarian, Nathan H. Gartner, John J. Haynes, Richard L. Hollinger, Matthew J. Huber, John B. Kreer, Joseph K. Lam, Tenny N. Lam, Edward B. Lieberman, Roy C. Loutzenheiser, Carroll J. Messer, Paul Ross, Richard Rothery, Joel Schesser, Steven R. Shapiro, A. D. St. John, W. W. Wolman

Committee on User Information Systems
Wallace G. Berger, U.S. Senate Appropriations Committee, chairman
Fred R. Hanscom, BioTechnology, Inc., secretary
Terrence M. Allen, Hebert J. Bauer, Norman J. Cohen, Karen P. Damsgaard, Robert Dewar, J. Glenn Ebersole, Jr., Eugene Farber, John C. Hayward, Robert S. Hostetter, Gerhart F. King, Harold Lunenfeld, Truman Mast, Peter B. Moreland, Theodore J. Post, Martin L. Reiss, Arthur W. Roberts III, Bob L. Smith, Eugene M. Wilson

David K. Witheford and James K. Williams, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units and officers and members are as of December 31, 1978.

Contents

Warrants for Left-Turn Signal Phasing

Kenneth R. Agent and Robert C. Deen, Division of Research, Bureau of Highways, Kentucky Department of Transportation, Lexington

Warrants for the installation of left-turn phasing in Kentucky were developed. A review of the literature was conducted, along with a survey of the policies of other states. Field data on delays and conflicts were taken before and after installation of exclusive left-turn signalization. Left-turn delay studies were conducted at intersections that had varying volume conditions. Analysis of the effect on accidents of adding a left-turn phase was made. The relationship between left-turn accidents and conflicts was investigated. Other types of analyses concerning gap acceptance, capacity, and benefit-cost ratios were also performed. It was found that exclusive left-turn phasing significantly reduced left-turn accidents and conflicts. This reduction was offset in part by an increase in rear-end accidents. Left-turn delay was reduced only during periods of heavy traffic flow. Total delay for an intersection increased after installation of left-turn phasing. Warrants were developed dealing with accident experience, delay, volumes, and traffic conflicts.

A vehicle attempting to turn left across opposing traffic is a common problem. Separate left-turn lanes minimize the problem but may not be the final solution. At signalized intersections, left-turn phasing can be used as an additional aid. However, warrants have not been established for the addition of separate left-turn lanes or signal phasing. In this study, warrants or guides were developed for installing left-turn phasing at signalized intersections that have separate left-turn lanes. Before-and-after data were taken at locations where left-turn phasing had been added. Studies at locations that had varied traffic conditions were made to determine the relationship between various volumes and left-turn delays. The relationship between left-turn accidents and conflicts was investigated. Comparisons of signalized intersections with and without left-turn signals were also made.

SURVEY OF OTHER STATES

Other state highway agencies were requested to describe their procedure used to determine the need for left-turn phasing. Of the 45 states responding, only 6 cited numerical warrants for left-turn phasing. In one state, warrants were proposed. The various numerical warrants used when considering left-turn phasing were as follows (some states had more than one warrant):

1. Product of the left-turn highest-hour volume and the opposing traffic ≥ 50 000;
2. Five or more left-turn accidents within a 12-month period (two states);
3. Cross product of left turns and conflicting through peak-hour volumes > 100 000 (two states, one listing this for traffic-actuated signals only);
4. Delay to left-turning vehicle in excess of two cycles;
5. One left-turning vehicle delayed one cycle or more in 1 h;
6. At a pretimed signal, left-turn volume of more than two vehicles per approach per cycle during a peak hour;
7. Average speed of through traffic exceeds 72 km/h (45 mph) and the left-turn volume is 50 or more on an approach during a peak hour;
8. Left-turning volume exceeds 100 vehicles during the peak hour;
9. More than 90 vehicles/h making a left turn; and
10. For four-lane highways with left-turn refuges, a relationship between left-turn volume, opposing-traffic volume, and posted speed.

Nearly all of the responses listed guidelines that have been used. Following is a list of the general guidelines (areas that should be considered) that were mentioned, some of which were listed by several states: accident experience, capacity analysis, delay, volume counts (peak-hour left-turn and opposing through volumes), turning movements, speed, geometrics, signal progression (consistency with and effect on adjacent signals), queue lengths, right-of-way available, number of opposing lanes to cross, gaps, consequences imposed on other traffic movements, type of facility, sight distance, and percentage of trucks and buses. Several states listed more detailed guidelines involving specific left-turn volumes, etc.

Following is a summary of guidelines used when considering a separate left-turn signal phase: left-turn volume > 500 (two-lane roadway), wherever a left-turn lane is installed on divided highways; 100-150 left-turning vehicles during the peak hour (small cities); 150-200 left-turning vehicles during the peak hour (large cities); at new installations, where left-turn phases already exist at other intersections on the same roadway; average cycle volume exceeds two vehicles turning left from the left-turn bay, and the sum of the number of left-turning vehicles per hour and the opposing-traffic volume per hour exceeds 600 vehicles; high percentage of left-turning vehicles (20 percent or greater); not provided at intersections with left-turn volume < 80 vehicles/h for at least 8 h/day; the number of left-turning vehicles is about 2 per cycle; 120 left-turning vehicles in the design hour; turning volume in excess of 100 vehicles/h, and more than one cycle of the signal needed to clear a vehicle stopped on the red; left-turn volumes of 90-120 in peak hours; and more than 100 turns/h.

RESULTS

Accident Warrant

Before-and-After Accident Studies

Accident data before and after installation of separate left-turn phasing were collected for 24 intersections. The length of the before and after periods was usually one year, but it varied in some cases depending on the available data. There was an 85 percent reduction in left-turn accidents, defined as those occurring when one vehicle turned left into the path of an opposing vehicle. This reduction in left-turn accidents was offset in part by a 33 percent increase in rear-end accidents. There was a reduction of 15 percent in total accidents.

Accident severity was reduced only slightly after installation of the left-turn phasing. Rear-end accidents (which were increased) are less severe than left-turn (angle) accidents (which were decreased). Injury accidents decreased from 13 to 11 percent after left-turn phasing was installed.

Comparison of Accident Rates at Intersections With and Without Left-Turn Phasing

Accident rates at intersections in Lexington, Kentucky, with and without left-turn phasing were compared. Rate were calculated by using 1972 accident data, and the volume data were taken for 1971 through 1973. Volume counts were available for a 12-h period (7:00 a.m. to 7:00 p.m.) at each intersection. The assumption was made that 80 percent of the total daily volume occurred in this 12-h period, so the volumes were multiplied by 1.25 to obtain the 24-h volume. The total rate of the intersection-type accidents was computed in terms of accidents per million vehicles entering the intersection. The left-turn accident rate was calculated, for each approach that had a separate left-turn lane, in terms of left-turn accidents per million vehicles turning left from the approach. Intersections without left-turn phasing (44 intersections) had average annual daily traffic (AADT) of approximately 20 000, compared with slightly more than 32 000 for intersections that had left-turn phasing (16 intersections). The higher AADT affects the accident rate. Calculating rates for only the high-volume intersections (AADT > 25 000) eliminated this variable. There were 13 intersections that had separate phasing and 10 intersections without separate phasing that met this criterion.

The left-turn accident rate was drastically lower for the approaches that had left-turn phasing (0.77 left-turn accidents/million vehicles entering the intersection for all intersections, 0.86 for high-volume intersections) than for approaches without left-turn phasing (2.74 for all intersections and 3.76 for high-volume intersections). The lower rate agreed with the findings of the before-and-after accident studies. The data again showed that left-turn phasing did not reduce the total intersection accident rate. The total accident rate was almost identical at locations with (1.66 for all intersections and 1.63 for high-volume intersections) and without (1.63 for all intersections and 1.69 for high-volume intersections) left-turn phases.

Critical Left-Turn Accident Number

By using the Lexington data base, the average number of left-turn accidents for the approaches with no left-turn phasing was calculated. By using this average number of accidents, the critical number of accidents was also determined. For 1968 through 1972, the average number of left-turn accidents per approach was 0.93 (for 96 approaches with a left-turn lane but no separate phase). For a street that had a left-turn lane in each direction, both approaches were included. The formula for critical accident rate (1) can be converted to calculate the critical number of accidents by substituting accidents divided by volume for the rate. Multiplying both sides of the equation by volume resulted in the following formula for critical number of accidents:

$$N_c = N_a + K\sqrt{N_a} + 0.5 \qquad (1)$$

where

N_c = critical number of accidents,
N_a = average number of accidents, and
K = constant related to level of statistical significance selected (for P = 0.95, K = 1.645; for P = 0.995, K = 2.576).

For P = 0.995, the critical number of left-turn accidents per year per approach was found to be four. Using the high probability increases the likelihood of selecting for improvement only intersections that have a significant left-turn problem. Therefore, four left-turn accidents in one year on an approach would make that approach critical. The number of accidents in a two-year period necessary to make an approach critical was also determined. There was an approximate average of two left-turn accidents on an approach during a two-year period. By using this average of two accidents, the number of left-turn accidents necessary in a two-year period to make an approach critical was found to be six.

The same procedure was used to determine the critical number of accidents for both approaches when a street has left-turn lanes in both directions. For 1968 through 1972, the average number of left-turn accidents for both approaches on a street was 2.1 (for 36 streets with left-turn lanes for both directions at an intersection but no separate phase). This resulted in a critical number of 6 for a one-year period for both approaches. For a two-year period, an average of 4 accidents resulted in a critical number of 10 for both approaches.

Delay Warrant

Before-and-After Delay and Conflict Studies

To determine the change in vehicular delay, studies were conducted before and after installation of left-turn phasing at three intersections that had two-phase, semi-actuated signalization. Left-turn delay was defined as the time from when the vehicle arrived in the queue or at the stop bar until it cleared the intersection. The arrival and departure times of each left-turning vehicle were noted; delay could then be calculated. If the vehicle did not have to stop, a zero delay was noted. The number of left turns was counted. Opposing volumes and left-turn conflicts were also counted during the study period, usually 30 min of each hour.

Because of high volumes involved when determining total intersection delays, the stop-type delay, the time in which the vehicle is actually stopped, was used because it was the easiest and most practical delay to measure (2, 3). The estimating procedure consisted of counting the number of vehicles stopped in each intersection approach at periodic intervals. The interval used was 15 s for two of the intersections and 20 s for the other. The volume on each approach was also counted. The total delay was the product of the total vehicles stopped at periodic intervals and the length of the interval. The delay per vehicle was obtained by dividing the total delay by the volume for that approach. Data were taken for 30 min out of the hour in most cases and were taken during an average of 9 h/day at the three intersections. The delay was calculated for each approach and then combined with left-turn delay to determine total intersection delay. The results of the studies are given in Table 1.

As expected, total delay increased after installation of the exclusive left-turn phasing. Two of the locations were T-intersections at which left-turn phasing was installed on only one approach. The T-intersections had an average increase in delay of less than 1 s, compared with about 5 s at the other intersection. The reason for the difference was clear when the delay for each approach was examined. The T-intersections had one approach on the main street that had a substantial reduction in delay because it was allowed to proceed while

Table 1. Summary of delay and conflict studies.

Item	Dixie Highway and Deering Road, Louisville (T-intersection) Before	After	Change (%)	US-41A and Skyline Drive, Hopkinsville (T-intersection) Before	After	Change (%)	Dixie Highway and Pages Lane, Louisville Before	After	Change (%)
Delay (s/vehicle)									
Total intersection									
All hours	6.8	6.8	0	4.7	5.4	+15	9.4	15.2	+62
Peak hours	11.3	11.2	-1	6.8	6.2	-9	11.7	21.6	+85
Nonpeak hours	6.5	6.8	+5	3.9	5.0	+28	8.9	13.7	+54
Side street	8.4	10.6	+26	20.0	18.2	-9	17.9	24.0	+34
Opposing approach traffic	2.0	4.5	+125	4.7	6.4	+36	6.9	11.9	+72
Unopposed approach traffic	4.0	1.0	-75	2.4	1.7	-29			
Left turn									
All hours	22.1	32.7	+48	15.5	20.5	+32	39.0	38.3	-2
Peak hours	48.8	36.8	-25	30.0	27.8	-7	52.8	44.2	-16
Nonpeak hours	23.9	34.0	+42	11.2	19.2	+71	37.2	36.8	-1
Left-turn conflicts	50	12	-76	42	13	-69	53	3	-94
Total volume (vehicles)	9057	8372	-8	8606	7208	-16	10 531	5036	-52
Left-turn volume (vehicles)	481	492	+2	653	650	0	364	397	+9

the left turns were made, thus increasing its green time. This was the unopposed approach. This reduction in delay compensated for the increase in delay for the approach that was opposing the left turns. Another study had found a 3.5-s increase in delay when left-turn phasing was added on one street (2); increased delay of 8.6-12.5 s/vehicle was observed when additional phasing was installed on all approaches.

Total left-turn delay was not decreased by the addition of left-turn phasing. Delay actually increased at two of the locations and remained the same at the other. Left-turn delay was reduced at all three locations during the peak hour. The data clearly showed that exclusive left-turn phasing will only reduce left-turn delay during periods of heavy traffic flow. The total left-turn delay was reduced at the one location because it had several high-volume hours, while there were only a few hours of heavy volume at the other locations.

Left-turn conflicts were classified into three categories (4). The first type of conflict (basic left-turn conflict) occurred when a left-turning vehicle crossed directly in front of or blocked the lane of an opposing through vehicle. This conflict was counted when the through vehicle braked or weaved. This was the most common type of left-turn conflict. A second type of conflict is a continuation of the first type. If a second through vehicle following the first one also had to brake, this conflict was counted. There were very few of these conflicts. The third conflict consisted of turning left on red. This conflict was counted when the vehicle entered the intersection after the signal turned red. Vehicles that entered the intersection legally and completed their movement after the signal changed were not counted.

Left-turn conflicts were reduced drastically after installation of left-turn phasing. The only conflicts in the after period involved vehicles running the red light. The after-period data were not taken immediately after installation in order to allow drivers to become accustomed to the left-turn phase, but there were still some red-light violations. This large reduction in conflicts corresponded to the accident reduction found at locations where left-turn phasing was added.

There was a slight increase in left-turn volumes after installation of the separate phasing. This could be expected, because drivers would take advantage of the safer movement allowed by the left-turn phase. The total volume happened to be lower during the after studies. The delays during the after period might have been slightly higher if the volumes had been equal to the before-period conditions.

Benefit-Cost Analysis

The benefits and costs of installing left-turn phasing were compared to determine the economic consequences. The benefit considered was the reduction in accident costs. As was discussed above, left-turn accidents were reduced by 85 percent after installation of left-turn phasing, but rear-end accidents increased, partly offsetting the benefits of the reduction. For the 24 intersections where accident data were collected, the average reduction in the number of left-turn accidents was 4.1, compared to a reduction of 3.0 in total accidents. This factor (3.0/4.1) was applied to the 85 percent reduction in left-turn accidents to account for the increase in other accidents. Accident savings resulting from a left-turn phase were then determined by using an average cost of $7112/accident. This cost was calculated by using National Safety Council accident costs and considering the distribution of fatalities, injuries, and property-damage-type accidents in Kentucky. The operating cost considered was that due to the increase in intersection delay.

Benefits and costs were calculated on an annual basis. The cost of installation, when computed as an annual cost, becomes insignificant compared to the delay costs. Therefore, installation costs were not included. Annual delay costs of adding left-turn phasing on one approach (T-intersections) as well as both approaches on a street were tabulated as a function of intersection volume (AADT). An added delay of 1 or 5 s/vehicle was used when phasing was added on one approach or two approaches, respectively. These numbers were obtained from the delay studies. A delay cost of $4.87/vehicle-h was used. This number was derived from a 1970 report that listed values for delay of $3.50/vehicle-h for passenger automobiles and $4.47/vehicle-h for commercial vehicles (5). By using the consumer price index to convert to 1975 costs and assuming 5 percent of the total volume to be commercial vehicles, a delay cost of $4.87/vehicle-h was derived.

The benefit-cost ratio would vary greatly according to AADT and the number of left-turn accidents. As an example, an AADT of 30 000 was used because it was close to the average volume for the Lexington intersections that had left-turn phases. This would result in an annual delay cost of $14 800 and $74 100 for adding phasing to one and two approaches, respectively. The critical number of left-turn accidents in one year was used to determine accident savings. For a T-intersection, the critical number of four yields an annual savings of $17 700. The benefit-cost ratio would be 1.20. For two approaches, the critical number is six, which gives an accident savings of $26 500. Using the delay cost of $74 100 yields a benefit-cost ratio of 0.36.

As a general rule, the savings attributable to accident reduction should offset the increased cost due to delay when street geometry makes left-turn phasing necessary on only one approach that has a critical number of accidents. This situation would be approximated if both approaches must be signalized but left-turn volume on one approach is very low. Since the left-turn phasing would be actuated, this would approximate the T-intersection situation if the left-turn phasing for one approach was used only during a very small percentage of the cycles. However, when a street has relatively high left-turn volumes on both intersection approaches, the cost of increased delay will be much higher than the savings from accident reduction.

Left-Turn Delay

Excessive delay in left turns is one of the major reasons for installing separate left-turn signals. A good delay criterion should include both delay and volume. Multiplying the average delay per vehicle (seconds) by the corresponding left-turn volume yields the number of vehicle-hours of delay. This unit of delay was used in this study. Also, further safeguards were built into the delay warrant. Minimum delay per vehicle and minimum volumes were specified so that neither very low volumes with excessive delays nor very high volumes with minimal delays would meet the warrant. The delay during peak-hour conditions was specified, since these are the conditions that create excessive delays.

Cycle time and the number of vehicles that might turn left during amber periods were considered when determining a minimum left-turn volume. The maximum cycle that normally would be used is 120 s. This would give 30 periods of amber/h for use by left-turning vehicles. Assuming that a minimum average of 1.6 vehicles could turn left during each amber phase, 48 vehicles/h could turn left during amber under peak opposing-flow conditions. Therefore, a minimum left-turn volume of 50 vehicles in the peak hour was specified.

A minimum value necessary for the average left-turn delay was also determined. Since installing a separate left-turn phase would increase total delay at the intersection, the supposition was made that a minimum delay was necessary to left-turning vehicles independent of the left-turn volume. To determine this level of delay, a past survey of engineers was used (6). This survey asked the engineers for their opinion of what constituted maximum tolerable delay for a vehicle controlled by a traffic signal. A mean value of 73 s was found. The criterion used was that 90 percent of all left-turning vehicles be delayed less than this maximum of 73 s.

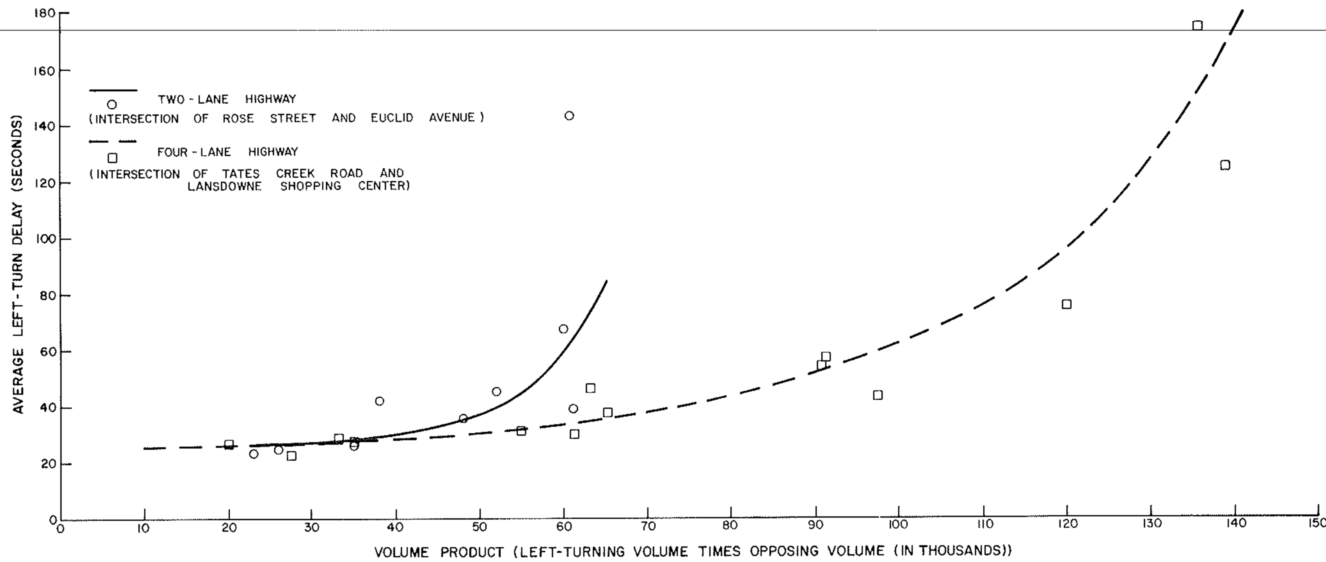
Assuming that the distribution of delays was approximately normal, it was then possible to find the mean of the delay distribution whose 90th percentile value was approximately 73 s/vehicle. From field data, it was found that the ratio of the mean to the standard deviation increased as the mean increased. For average delays approximating 73 s, this ratio was about 1.5. By using this ratio, a value of 35 s for the mean delay was determined. This value of 35 s was used as the minimum average delay necessary, since this constituted the lower bound of excessive delay.

When considering what would constitute excessive delay, the delay to left-turning vehicles turning only on the amber phase was calculated. This would approximate peak-flow conditions when the only gap available to turn left occurs at the end of the amber phase. The maximum delay possible if none of the vehicles had to wait more than one cycle length was determined. The maximum delay possible would occur when the left-turning vehicle arrived at the start of the red phase and departed during the amber phase. This delay would be approximately equal to one cycle. The number of vehicles that could turn left in 1 h during the amber phases was dependent on the cycle length. Since peak-hour conditions were specified, the assumption was made that side-street traffic would be heavy enough to make an actuated signal behave as a fixed-time signal with a constant cycle length. If the cycle length were 60 s, there would be 60 amber phases available to left-turning vehicles. Thirty amber phases would be available during the peak hour at a signal with a 120-s cycle length. If an average of 1.6 vehicles turned left during each phase of amber, 96 vehicles/h could turn left if the cycle length were 60 s. This volume would decrease to 48/h for a cycle length of 120 s. For a maximum delay of one cycle, the total delay for the peak hour was determined to be 1.6 vehicle-h for both cycle lengths. Field experience has shown that during peak conditions the number of vehicles turning left during each phase of amber can become close to 2 if the left-turn volume is heavy. If an average of 2 vehicles turn left during each amber phase, the total left-turn delay becomes 2.0 vehicle-h during the peak hour. Delays in excess of these values could be considered excessive. These delays would apply to the critical approach.

Delay data collected at several intersections were compared with these values to check their validity. As stated earlier, studies were done before installation of left-turn phases at three intersections. During peak-hour conditions before installation, left-turn delays of 2.45, 1.27, and 1.64 vehicle-h were found at those three locations. The location that had a delay of 1.27 vehicle-h also had an average left-turn delay during the peak hour of only 30 s. Six intersections in Lexington that had high left-turn delays were selected for detailed delay studies. Delays were measured on both streets at one of the intersections. Left-turn delays were measured for several hours during the day. The peak-hour delay was ≥2.0 vehicle-h (varying from 1.76 to 5.96) in all but one case. Only two of the critical approaches had peak-hour delays > 2.5 vehicle-h. All of these approaches met the criteria of minimum left-turn delay and volume. The field data show that peak-hour, left-turn delay in excess of 2.0 vehicle-h can occur regularly at locations that have a left-turn problem.

A review of the literature (7) disclosed two peak-hour delay warrants for the installation of traffic signals that had been developed in terms of vehicle hours of delay. One warrant requires that the average side-street vehicle delay in seconds multiplied by side-street volume per hour be equal to or exceed 8000. This is equivalent to 2.2 vehicle-h of delay. Another peak-hour delay warrant for a single, critical left-turn approach was 2.0 vehicle-h of delay. A minimum volume of 100 on the approach during the peak hour was also required. Assuming the delays for side-street vehicles can be applied to left-turning vehicles, a delay of 2.0 vehicle-h

Figure 1. Relationship between volume product and left-turn delay.



during the peak hour could be considered a valid warrant.

Volume Warrant

Relationship Between Left-Turn Delay and Traffic Volumes

Data collected at several intersections have shown that average left-turn delay varied substantially between intersections for any given volume-related product. For example, for a product of left-turn and opposing 1-h volumes of approximately 100 000, the average left-turn delay found at approaches at seven intersections on four-lane streets varied from a low of 15 s to a high of 100 s. Three of the approaches had average left-turn delays of less than 30 s, while three had average delays of 60 s or more. This clearly shows that, even if the calculated product was above the specified warrant value, a left-turn phase should not be added to an existing signal unless a delay study also showed an excessive delay.

Better relationships of delay versus the volume product were found when data from individual intersections were plotted. An important deficiency was found in some currently used volume-product warrants; all but one of these warrants did not define the number of opposing lanes. Data showed that a much higher volume product would be necessary to warrant a left-turn phase on a four-lane street than on a two-lane street. The product was directly proportional to the number of opposing lanes.

Plots of data collected at two intersections are shown in Figure 1. In both cases, the left-turn delay increased sharply after the product of the left-turning and opposing volumes reached a certain level. The increase in delay occurred at a much higher volume product on the four-lane street than on the two-lane street. Plots such as these were prepared for several intersections. The increase in delay did not occur at any specific volume product, and the increase was not as dramatic in some cases. The increase in delay did not occur at all if the volume product remained low. For four-lane streets, plots showing this increase in left-turn delay were drawn for the approaches of seven intersections. The 1-h volume product at which the increase occurred was estimated in each case. It varied from a low of 60 000 to a high of 145 000 and averaged 103 000. For two-lane streets, plots were drawn for approaches of three streets at two intersections. The critical volume product varied from 30 000 to 70 000 and averaged 50 000.

Comparison of Locations With and Without Left-Turn Phases

Plots of peak-hour opposing volume versus peak-hour left-turn volume were made for intersections on both four-lane and two-lane highways with data from Lexington (Figure 2). A point was plotted for each approach at a signalized intersection that had a separate left-turn lane. The only exception was that only the critical approach was plotted for streets that had left-turn phasing if it was obvious that only one approach had a problem. The policy is to install left-turn phasing in both directions although it may only be warranted for one approach.

The objective was to construct a line that separated intersection approaches with and without left-turn phases. An attempt was made to construct a line in which the product of the peak-hour left-turn and opposing volumes was a constant. If such a line could be drawn, this product could be thought of as a warrant based on past practices. Such a line was drawn for both four-lane and two-lane highways. There were only a very few exceptions to the division of the approaches into groups with and without left-turn phasing. The lines represented a product of peak-hour left-turn and opposing volumes of 90 000 for four-lane highways and 60 000 for two-lane highways.

Gap Acceptance

Gap acceptance has been proposed as a criterion for left-turn phasing (8). Although it will not be used as a warrant in this study, it can be used to corroborate other data. Some very rough calculations were made that seemed to agree with field observations.

Data were taken to determine the critical gap for vehicles turning left across opposing traffic. The critical gap was defined as the length of gap at which the number accepted was equal to the number rejected. The gap was measured as the interval in time between

Figure 2. Comparison of volumes at intersections with and without left-turn phasing.

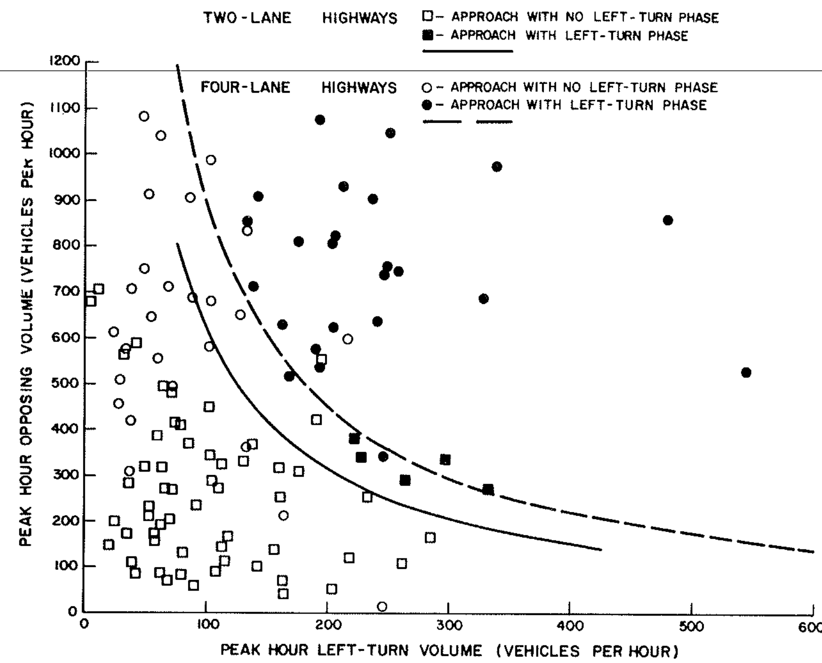
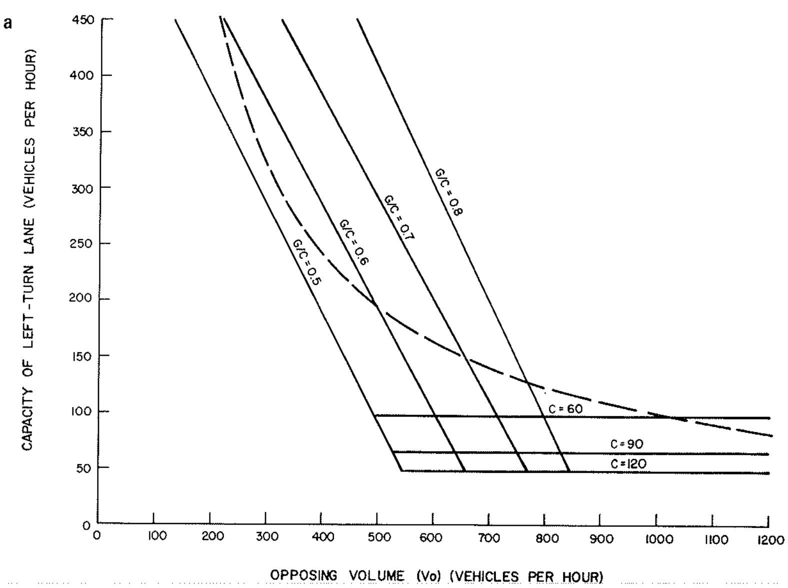


Figure 3. Capacity of a left-turn lane on the basis of a capacity nomograph.



vehicles opposing the left turn. It was measured from the rear of one vehicle to the front of the following vehicle. Observations were made of 500 vehicles attempting to turn left at a signalized intersection. A critical gap of 4.2 s was found.

By using several assumptions, an estimate of the volume of left-turning and opposing traffic necessary to warrant a left-turn phase can be made. The volume at which there are no gaps greater than the critical gap (4.2 s) would be approximately the point at which all left turns must be made during the amber. If the assumption is made that 60 percent of the cycle is green time for the main street, there would be 2160 s of green and amber time per hour on the main street. Making the rough assumption that the vehicles would be equally spaced resulted in volumes of 514 vehicles/h on two-lane highways and 1028 vehicles/h on four-lane highways as the point at which left-turning vehicles could turn only on the amber. It is recognized that vehicles will not be equally spaced under stable flow conditions. This assumption, however, should yield conservative results, since opposing volumes above these volumes will contain gaps greater than the critical gap because of variations in vehicle spacings. However, the results generally agree with field observations that, under average conditions, for opposing volumes of about 500 vehicles/h on two-lane highways and 1000 vehicles/h on four-lane highways, most left turns must be made during the amber period. For a cycle of 60 s, 60 amber periods would be available per hour. Assuming 1.6 vehicles can turn left during each amber period, the capacity of the left-turn lane was 96. Therefore, the critical product of left-turning and opposing volumes was approximately 100 000 for four-lane highways and 50 000 for two-lane highways.

Of course, this critical product would vary as the cycle length or green-time-to-cycle-length ratio for the main street changed. For example, data were taken at one intersection on a four-lane highway that had a cycle of 60 s and a green-time-to-cycle-length ratio of about 0.75 for the main street. For peak-hour opposing volumes of slightly more than 1000/h, most left-turning vehicles did not have to turn during the amber. This was the result of more green time for the main street. By using the same assumptions as before, except that 75 percent of the cycle is assumed to be devoted to the main street, a volume of 1286 vehicles/h was the point at which left-turning vehicles could turn only on the amber. This would yield a critical product of 125 000.

Relationship Between Left-Turn Accidents and Traffic Volumes

By using the same Lexington data base, plots were drawn of the highest number of left-turn accidents in one year for an approach versus the product of peak-hour left-turn volume and opposing volume, as well as just the left-turn volume. The highest accident year was used so that a comparison could be made with the critical accident number. The plots showed that the relationship was very poor in nearly all cases. Plots were drawn for both two- and four-lane highways. With one exception, the maximum coefficient of determination (r^2) was 0.2. The one exception was the plot of accidents versus the product of peak-hour left-turn and opposing volumes for four-lane streets; the r^2 value for this plot was 0.5. Four accidents on an approach in one year had previously been found to be the critical number. This corresponded to a volume product of approximately 80 000. A plot of left-turn accidents versus left-turn volume resulted in an r^2 value of only 0.19. A value of four accidents related to a left-turn volume of 120. The inability to fit a curve to the points makes it hard to draw any valid conclusions from the plots. However, the higher r^2 value for the plot that used the product of left-turning and opposing volumes indicates that this product was a better estimator of left-turn accidents than was left-turn volume.

Capacity Analysis

A capacity analysis is used in several states as a guideline when considering the installation of left-turn phases. The nomograph developed by Leisch was used to develop a warrant curve based on intersection capacity (9). By assuming 5 percent trucks and buses, curves were drawn representing green-time-to-cycle-length ratios of 0.5 to 0.8 and cycles of 60 to 120 s (Figure 3). This figure clearly shows how the left-turn capacity is increased as the green-time-to-cycle-length ratio is increased and the cycle length is decreased. Points above the curves represent intersections where the left-turn volume was above the left-turn capacity that would warrant a left-turn phase. The dashed line in Figure 3 depicts a product of 95 000 for the left-turning and opposing volumes, assuming 5 percent trucks and buses; a green-time-to-cycle-length ratio of 0.6; and a cycle length of 60 s. A deficiency of this procedure is that the number of opposing lanes is not specified.

Selection of Volume-Related Warrants

The preceding sections have dealt with various methods of selecting a critical product of left-turning and opposing vehicle volumes. Although some methods were based on assumptions and collected data and some were based entirely on field data, there was a close agreement of the results. A volume warrant based on all sources of input was developed. The warrant required that the addition of separate left-turn phasing should be considered when the product of left-turning and opposing volumes during peak-hour conditions exceeds 100 000 on a four-lane street or 50 000 on a two-lane street. A limitation is that the left-turn volume must be at least 50. This is based on the same reasoning as that for the minimum volume requirement in the delay warrant. It is important to note that, even if the calculated product exceeds the warrant, a left-turn phase should not be added to an existing signal unless a study shows excessive left-turn delay.

Traffic-Conflicts Warrant

A major reason for installing left-turn phasing is to provide improved safety. An obvious indicator used to warrant a left-turn phase because of a safety problem has been the number of left-turn accidents. A weakness of that indicator is that a substantial number of accidents must occur before any improvement is made. The traffic-conflicts technique has been developed in an attempt to objectively measure the accident potential of a highway location without having to wait for an accident history to evolve.

An attempt was made to find a relationship between left-turn accidents and conflicts. The types of left-turn conflicts counted have been described earlier in this report. The Lexington data base was the source of the accident data. This provided a five-year accident history for the intersection approaches. Comparisons were made for individual approaches that had separate left-turn lanes. The approach also had to be at a signalized intersection. Since conflicts indicate accident potential, the highest numbers of accidents in a one-year and in a two-year period were used in the comparisons. Left-turn accidents were compared with the total number of conflicts (all three types) and with the basic left-turn conflicts (left-turn vehicle crossed directly in front of or blocked the lane of an opposing through vehicle). Conflict counts were taken during peak flow conditions for 1 h. Volume counts were used in selecting times for data collection. Both left-turn and opposing volumes were considered. Peak hours were chosen, because conflicts are highest during these hours; left-turn accidents also reach a maximum during peak-volume hours, and it appeared reasonable that conflict counts should be conducted when accident problems are most acute. It is important to note that conflict data were taken during

Figure 4. Left-turn accidents versus total left-turn conflicts.

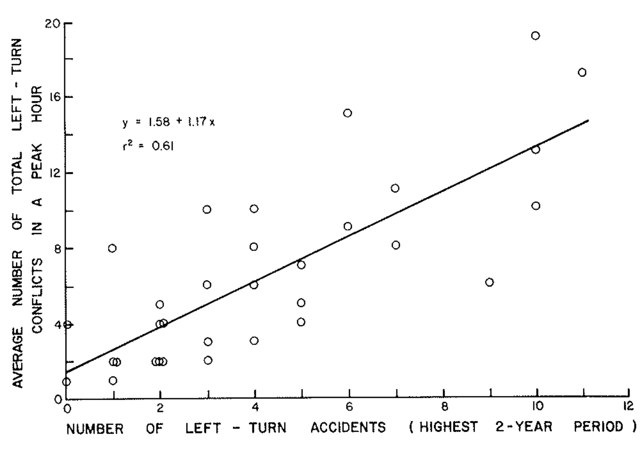
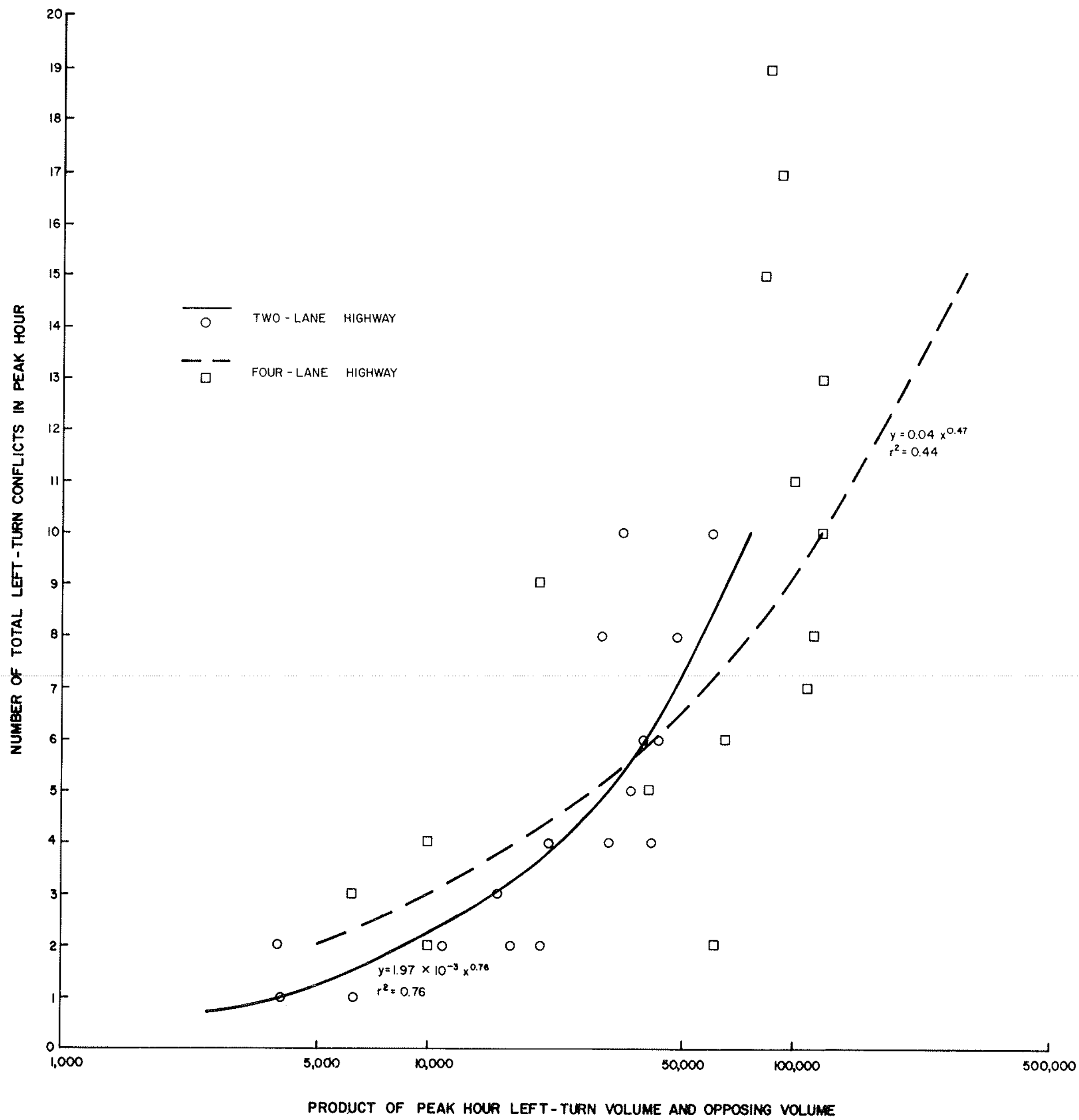


Table 2. Relationship between left-turn accidents and left-turn conflicts.

Variable	Linear Regression Equation[a]	R^2	Critical No. of Conflicts	Range[b]
Number of total conflicts versus				
Highest one-year period of accidents	Y = 1.26 + 1.87X	0.50	8.7	5.4
Highest two-year period of accidents	Y = 1.58 + 1.17X	0.61	8.6	4.8
Number of basic conflicts versus				
Highest one-year period of accidents	Y = 1.42 + 1.13X	0.39	5.9	4.1
Highest two-year period of accidents	Y = 1.70 + 0.69X	0.45	5.8	3.9

[a] Y = number of conflicts; X = number of accidents.
[b] Probability level = 95 percent.

Figure 5. Total left-turn conflicts in peak hour versus product of peak-hour left-turn volume and opposing volume.



several peak hours at each of 32 approaches, so that a reliable average number of conflicts per hour could be obtained.

Plots were drawn of left-turn accidents versus left-turn conflicts; see Figure 4 for an example. By using linear regression and the method of least squares, equations of the best-fit lines were determined. The coefficients of determination (r^2) ranged between 0.39 and 0.61. For both conflict categories, the best relationship was found when the two-year accident maximum was considered. Also, better relationships were found between accidents and total conflicts than basic left-turn conflicts; however, data showed the number of basic conflicts to be more consistent from one period of observation to the next. The critical number of left-turn accidents for one approach was previously found to be four for a one-year period and six for a two-year period. By using the linear regression equations, the number of conflicts corresponding to the critical number of accidents was predicted. The equations for one- and two-year accident data gave similar results. The equations predicted that about nine total conflicts or six basic conflicts corresponded to the critical number of accidents. Since the r^2 values were low, the range (confidence interval) within which conflicts could be predicted was determined. A probability level of 95 percent was used. A range of about ±5 was found for total conflicts, and a range of about ±4 was found for basic conflicts. The various findings are summarized in Table 2.

Simply using the predicted number of conflicts related to the critical accident number as a warrant for left-turn signalization would not be very reliable; this is so because of the uncertainty of the prediction equation, as evidenced by the large range in values possible. A warrant that considered the confidence interval would be much more reliable. The upper bound of values in the confidence interval was used as the conflict warrant. Given that number of conflicts, there would be a 95 percent certainty that the potential exists for the critical number of accidents to occur. Therefore, a warrant for left-turn signalization was developed that listed 14 total conflicts or 10 basic conflicts as its criterion.

A recent report included a critical evaluation of the state of the art of the traffic-conflicts technique and listed the results of work done in this area (10). In terms of accidents per conflict, there were 20 left-turn accidents/100 000 left-turn conflicts in one study (11) and 15 left-turn accidents/100 000 left-turn conflicts in the other study (12). If those results are averaged (17.5 accidents/100 000 conflicts) and if 4 left-turn accidents on an approach in a year is considered to be critical, the critical number of left-turn conflicts would be 22 857 in one year. Assuming the conflicts to be equally distributed throughout the year yielded an average of 62.6 conflicts/day. Volume data for Lexington showed that 14 percent of the daily left-turn volume occurred during the peak hour. Applying this factor to conflicts yielded 7.0 conflicts in the peak hour. This agreed with the previous finding: 6 basic left-turn conflicts in a peak hour would give an accident potential of 4 left-turn accidents in one year. Those two studies gave r^2 values of 0.38 and 0.11. The values for r^2 from 0.39 to 0.61 found for the linear regression lines of accidents and conflicts in this study compared favorably.

Conflicts are inherently related to volume. Plots were drawn to determine the relationship between left-turn conflicts and volumes for data collected in this study. Peak-hour conflicts were plotted against the product of left-turn volume and opposing volume. Volumes were counted while the conflict data were collected. Separate plots were drawn for four-lane and two-lane highways. Both total and basic conflicts were used, and it was found that the use of total conflicts gave better results (Figure 5). Several linear regression lines were tried, and the power curve yielded the best-fit line. The r^2 values for these figures indicate that a better relationship exists between left-turn conflicts and volume than between left-turn accidents and volume. A total of nine left-turn conflicts in the peak hour was previously found to correspond to the critical accident number. This number of conflicts related to volume products of 65 000 and 100 000 for two-lane and four-lane highways, respectively. These agree closely with the other findings for critical products.

RECOMMENDATIONS

It is recommended that the following warrants be used as guidelines when considering the addition of separate left-turn phasing. The warrants apply to intersection approaches that have a separate left-turn lane.

1. Accident experience—Install left-turn phasing if the critical number of left-turn accidents has occurred. For one approach, 4 left-turn accidents in one year or 6 in two years are critical. For both approaches, 6 left-turn accidents in one year or 10 in two years are critical.
2. Delay—Install left-turn phasing if a left-turn delay of 2.0 vehicle-h or more occurs in a peak hour on a critical approach. Also, there must be a minimum left-turn volume of 50 during the peak hour, and the average delay per left-turning vehicle must be at least 35 s.
3. Volumes—Consider left-turn phasing when the product of left-turning and opposing volumes during peak hours exceeds 100 000 on a four-lane street or 50 000 on a two-lane street. Also, the left-turn volume must be at least 50 during the peak-hour period. Volumes meeting these levels indicate that further study of the intersection is required.
4. Traffic conflicts—Consider left-turn phasing when a consistent average of 14 or more total left-turn conflicts or 10 or more basic left-turn conflicts occurs in a peak hour.

REFERENCES

1. T. Yamane. Statistics: An Introductory Analysis, 2nd ed. Harper and Row, New York, 1967.
2. G. Gurnett. Intersection Delay and Left-Turn Phasing. Traffic Engineering, June 1969, pp. 50-53.
3. H. Sofokidis, D.C. Tilles, and D.R. Geiger. Evaluation of Intersection-Delay Measurement Techniques. HRB, Highway Research Record 453, 1973, pp. 28-39.
4. S.R. Perkins and S.J. Harris. Traffic Conflict Characteristics: Accident Potential at Intersections. General Motors Corporation, Detroit, MI, Dec. 1967.
5. V.G. Stover, W.G. Adkins, and J.C. Goodknight. Guidelines for Medial and Marginal Access Control on Major Roadways. NCHRP, Rept. 93, 1970.
6. Traffic Signal Warrants. KLD Associates, Huntington, NY, KLD TR No. 17, Nov. 1973.
7. P.C. Box and W.A. Alroth. Warrants for Traffic Control Signals. Traffic Engineering, Nov. 1967, pp. 32-41; Dec. 1967, pp. 22-29;

Jan. 1968, pp. 14-20.
8. J. Behnam. Gap Acceptance as a Criterion for Left-Turn Phasing. Traffic Engineering, June 1972, pp. 40-42.
9. J. Leisch. Capacity Analysis Techniques for Design of Signalized Intersections. Public Roads, Vol. 34, No. 9, Aug. 1967, pp. 171-201; No. 10, Oct. 1967, pp. 211-226.
10. J. C. Glennon and B. A. Torson. Evaluation of the Traffic-Conflicts Technique. Federal Highway Administration, Dec. 1975.
11. W. T. Baker. Evaluating the Traffic-Conflicts Technique. Federal Highway Administration, 1971.
12. P. J. Cooper. Predicting Intersection Accidents. Canadian Ministry of Transport, Ottawa, Sept. 1973.

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

Guidelines for Traffic Control at Isolated Intersections on High-Speed Rural Highways

Ahmed Essam Radwan, Virginia Polytechnic Institute and State University, Blacksburg
Kumares C. Sinha, Department of Transportation Engineering, and
Harold L. Michael, School of Civil Engineering, Purdue University, West Lafayette, Indiana

This paper involves the development of guidelines for traffic control warrants at isolated intersections on high-speed rural highways by using both field studies and traffic simulation. Gap-acceptance and delay studies were performed at stop-sign-controlled rural intersections in Indiana, and the resulting data were used to validate and modify the UTCS-1 program (known now as NETSIM). Two-way stop signs, pretimed signals, semiactuated signals, and fully actuated signals were evaluated over a range of traffic volumes on both major and minor approaches. Annual economic cost was used as a basis to develop criteria for selecting the most appropriate control type. The resulting warrants are expressed in chart form.

The control of vehicular traffic at highway intersections has been one of the most studied areas in traffic engineering. Intersections critically affect the efficiency, capacity, and safety of a highway system. Not enough information is available on traffic control alternatives at isolated intersections on high-speed rural highways, in particular at the intersection of a multilane high-speed major highway and a two-lane minor road located in suburban or rural areas.

The Manual on Uniform Traffic Control Devices (MUTCD) (1) provides general guidelines for stop-sign and signal warrants at intersections; however, these guidelines do not distinguish between pretimed (PR) signals and vehicle-actuated (VA) control. Section 4C-3 of the MUTCD states:

> When the 85 percentile speed of the major street traffic exceeds 64 km/h (40 mph), or when the intersection lies within the built-up area of an isolated community having a population of less than 10 000, the maximum vehicular volume warrant is 70 percent of the requirement above (in recognition of differences in the nature and operational characteristics of traffic in urban and rural environments and smaller municipalities).

According to that statement, the minimum vehicular volume warrant for traffic-signal installation for a four-lane major street intersecting with a two-lane minor street is 420 and 105 vehicles/h (total traffic per approach), for the major and minor streets, respectively. For the warrant for the interruption of continuous traffic, the vehicular volumes are 630 and 53 vehicles/h. Section 4C-11 of the MUTCD reviews principal factors that may lead to selecting traffic-actuated control. However, there is a need for a detailed examination of warrants for traffic controls at multilane high-speed intersections.

DEVELOPMENT OF THE ANALYSIS TOOL

In this study, the UTCS-1S simulation model (the smaller version of the UTCS-1 model) was modified and used for the purpose of evaluating alternative traffic control devices. The single-intersection version of the UTCS had been successfully validated by Cohen (2) by using field data collected from two intersections that differed widely in geometry and location.

Hall (3) modified the UTCS-1S computer program to provide the vehicle fuel-economy and air-pollution measurements; careful study of the velocity patterns created by automobiles traversing the intersection made it possible to estimate fuel consumption and air pollution resulting from the use of various traffic controls.

Two-Way Stop-Sign-Controlled Intersection

For undivided major highways, the gap-acceptance distributions developed by Wagner (4) were used to modify the UTCS-1S model. These distributions represent the gap-acceptance behaviors of drivers stopped at the stop sign of a two-lane street intersecting with a four-lane undivided highway.

For the case of divided highways, the gap-acceptance distributions were developed from field observations made in the present study. Six rural intersections in Indiana were selected for this purpose, and they fulfilled

Table 1. Linear regression equations for average delay per vehicle on divided and undivided stop-sign-controlled intersections.

Regression Equation	R^2	F	Significance of F^a
Y = 2.1368 + 0.001 841 (major volume) + 0.002 113 (minor volume)	0.7177	41.9553	0.000 01
Z = 1.5543 + 0.001 054 (major volume) + 0.010 46 (minor volume)	0.8379	100.8126	0.000 01

Note: Y = Naperian logarithm (average delay per vehicle on undivided major highway); and Z = Naperian logarithm (average delay per vehicle on divided major highway).

$^a\alpha = 0.05$.

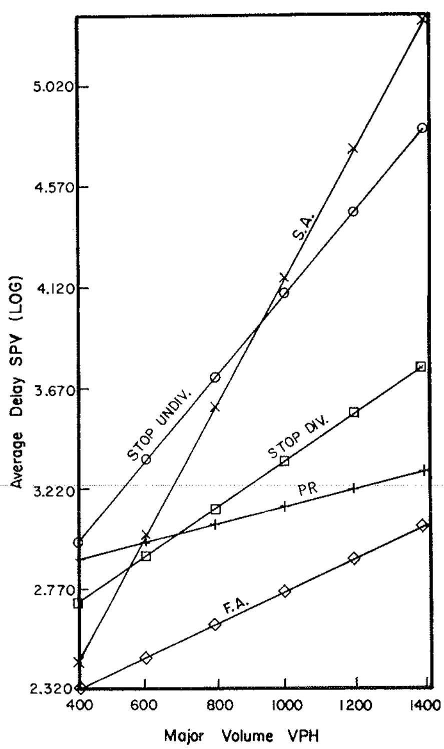
Table 2. Linear regression equations for average delay per vehicle at signalized intersections.

Regression Equation	R^2	F	Significance of F^a
X = 2.6564 + 0.000 386 6 (major volume) + 0.001 861 (minor volume)	0.7453	48.2861	0.000 01
Y = 1.1239 + 0.002 862 (major volume) + 0.003 437 (minor volume)	0.8784	119.2540	0.000 01
Z = 1.8714 + 0.000 724 9 (major volume) + 0.003 175 (minor volume)	0.7365	46.1393	0.000 01

Note: X = Naperian logarithm (average delay for pretimed signals); Y = Naperian logarithm (average delay for semiactuated control); and Z = Naperian logarithm (average delay for fully actuated control).

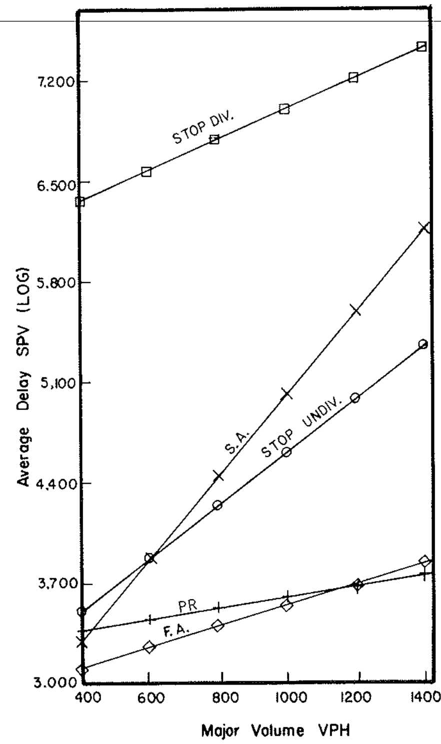
$^a\alpha = 0.05$.

Figure 1. Average delay values for minor-road volume of 50 vehicles/h.



specific requirements. One of these requirements was that the posted speed limit must be 89 km/h (55 mph) for the major road and 64 km/h (40 mph) for the minor road. A slow-motion film technique was adapted for securing the necessary data, and gap-acceptance distributions for different maneuvers were obtained. By using the distribution of driver type embedded in the UTCS program together with the average gap-acceptance values obtained

Figure 2. Average delay values for minor-road volume of 300 vehicles/h.



from the field study, a set of decile distributions of acceptable gaps was developed for various maneuvers. These decile distributions were then embedded in the simulation model.

In order to validate the delay values obtained from the model, stop delays and move-up times were measured from one of the films used in the gap-acceptance field study. Statistical tests showed that there was no significant difference between the simulated and the observed delay values at $\alpha = 0.05$.

After the model was modified and validated, it was used to perform a series of large-scale, two-way stop-sign-controlled intersection simulation runs. Four levels of major traffic volumes (500, 800, 1100, and 1400 vehicles/h), and three levels for minor traffic volumes (100, 200, and 300 vehicles/h), were considered. The purpose of these simulation runs was to develop regression equations for purposes of delay prediction. Three replicate simulation runs were obtained for every major-minor volume combination (400 of simulation time each). Average delay per vehicle measured in seconds for major and minor roads was chosen to be the delay measure for this study. Homogeneity of variances for the replicated data was checked, and it was found that the Naperian logarithm transformation was needed. Test of normality was also checked, and a linear regression equation was fitted to the transformed data for both divided and undivided highways, as shown in Table 1. Velocity profiles developed from the simulation runs were used as an input to an adapted version of the U.S. Environmental Protection Agency's Automobile Exhaust Emission Modal Analysis Model (5). Similar linear regression equations were developed to predict the amount of fuel consumed (per vehicle) within 122 m (400 ft) of either side of the intersection, as a function of major and minor traffic volumes.

Table 3. Costs and estimated lifetimes of control units.

Item	Flasher[a]	Pretimed Signal	Actuated Control
Capital cost ($)	3000-5000	15 000-18 000	(20 000-25 000)[b]
Annual maintenance cost ($)	240	350	550-650
Annual emergency cost ($)	75-100	150	450
Lifetime (years)	15-20	20-25	15-20

[a]Most Indiana rural intersections have both flashers and stop signs.
[b]Two-phase signal.

Table 4. Linear regression equations for annual accidents at signalized and unsignalized intersections.

Regression Equation	R^2	F	Significance of F[a]
X = 2.4613 + 0.000 224 6 (major volume + minor volume)	0.426	5.5738	0.026 0
Y = 1.7235 + 0.000 704 7 (major volume + minor volume)	0.825	85.2934	0.000 01

Note: X = annual number of accidents for stop-controlled intersections; and Y = annual number of accidents for signalized intersections.
[a]α = 0.05.

Signalized Intersection Studies

Signal timing is regarded as a critical variable affecting delay at intersections; for that reason, careful consideration was given to this matter.

Three control alternatives were considered: pretimed (PR), semiactuated (SA), and fully actuated (FA) control systems, each with the same levels of traffic volume. For the PR signal, a cycle length of 80 s was assumed, and the durations of the major- and minor-road phases were timed to minimize delay. For the SA control, a minimum green interval of 36 s was adopted for the major road. Assuming that the detectors are located 55 m (180 ft) back from the stop line on the minor approach, a vehicle extension duration time of 3 s would be sufficient for a vehicle to travel from the detector to the stop line. The initial interval and maximum extension duration times were taken to be 12 s and 24 s for the actuated phase of the minor-road approach. The durations of the initial interval, vehicle extension, and maximum extension were assumed to be 16, 4, and 64 s, respectively, for the major-road phase of the FA control. The corresponding values for the minor-road phase were 12, 3, and 27 s. The time durations of the actuated phases for the SA and FA controls were kept the same under the different levels of traffic demand.

The assumption of a fixed cycle length for all traffic control alternatives might have negated some of the advantages of SA and FA signals, especially at low traffic flow rates. The cycle length assumed for PR signals had a higher value than the optimum cycle length required to minimize traffic delay because the major-road approaches were of the high-speed type and their safety should be incorporated. No information is available regarding the required increase in the optimum cycle length for such intersections; therefore, it was felt that the 80-s cycle duration was a reasonable assumption. Since the initial runs indicated satisfactory results, no changes were made in the signalized logic of the model. Delay regression equations were developed for the three control alternatives, and they are shown in Table 2.

ANALYSIS OF RESULTS

Delay Analysis

After the delay equation for each control alternative was developed, the equations were combined to form composite plots. Figures 1 and 2 show two out of the possible six plots of the average delay values (seconds per vehicle, log scale) versus the major-road volumes (vehicles per hour) for the minor-road volumes of 50 and 300 vehicles/h, respectively. The composite plots indicated that the FA control causes the lowest average delay for the six minor-road traffic volume levels. It can be observed in Figure 1 that the average delay for a divided major highway (one that has a median) is smaller than the average delay for an undivided major highway for a minor-road traffic volume of 50 vehicles/h. As the minor-road volume increases, the delay curve for the divided major highway shifts upward. This can be explained by the fact that the existence of a highway median on the major road allows minor-road vehicles to perform their maneuvers in two movements. This holds for low minor-road volumes; however, an increase in the minor-road volume results in blockage of minor-road vehicles from the median and a consequent spillback occurs. In the case of a spillback, the simulation model randomly assigns a movement decision for a vehicle to determine whether it will join the spillback. This causes disturbance in the major-road flow, resulting in a reduction of speed and a subsequent increase in the average delay for major-road vehicles.

It was noticed that the SA control causes less average delay than the PR control at low minor-road volumes. As the minor-road volume increases, the intersection of the SA control line and the PR control line shifts to the left. Based only on the average delay analysis, the FA control appears to be the best control alternative at any major- or minor-road volume. However, the evaluation of a control alternative must also consider the safety aspects. In addition, the equipment cost should also be considered. It was therefore decided to perform an economic cost analysis that considered the costs of control-unit construction and maintenance, vehicle operation, accidents, and delay.

Economic Cost Analysis

Control-unit construction and emergency and normal maintenance cost data were obtained from the Indiana State Highway Commission. The capital cost, routine maintenance cost, emergency cost, and estimated life for flasher, PR signal, and actuated control signals are shown in Table 3. In the absence of actual information, the SA control costs were assumed to be an average of those for FA and PR controls. In reality, the SA control costs might be higher; however, since the equipment cost is very small compared with the accident and delay costs, this will not affect the results significantly. Assuming that the life of all signals is 20 years and that they have no salvage value, the equivalent uniform annual costs were estimated to be $1587, $5644, $6730, and $13 522 for flasher, PR, SA, and FA controls, respectively.

Vehicle Accident Costs

Accident records for stop-controlled, pretimed signal, and actuated controlled intersections on Indiana state highways were collected for the years 1974, 1975, and 1976. An analysis of variance test was performed, and it was found that the accident rates for stop-controlled intersections are significantly different from those for signalized intersections at α = 0.05. On the other hand, this test showed no significant difference between pretimed signal rates and fully actuated control rates for the same level of significance. Multiple regression equations were developed for stop-controlled and signalized intersections to be used in estimating an annual number of accidents, as shown in Table 4.

One survey of particular relevance to this research was performed by Hejal and Michael (6) to evaluate the direct cost per rural accident in Indiana. By updating the accident cost values to 1978 prices with the aid of the appropriate consumer price indices, the figures were estimated to be $25 954, $5971, and $845 for fatal, personal injury, and property-damage-only accidents, respectively.

In order to determine the average cost of an accident, the study conducted by Abramson (7) was used; in this study, the results of statewide accident information from Illinois, Massachusetts, Utah, and New Mexico were used. By assuming that the results of this study are applicable to the state of Indiana, the fractions of fatal, personal injury, and property-damage-only accidents were estimated to be 0.0041, 0.0826, and 0.9133, respectively.

A recent study by Wuerdemann (8) provided national indirect costs of motor vehicle accidents. An average indirect cost value of $160/accident was adopted from this study. By using the severity fractions, the direct costs, and the indirect-cost value, the weighted average cost per accident was found to be $1595.

Automobile Operating Cost

Knowing the quantity of gasoline consumed in driving a vehicle through an intersection under the four types of control alternatives, as simulated by the UTCS-1S model, permitted gasoline cost calculations. It was assumed that the average cost of gasoline was 17 cents/L (64 cents/gal) in 1978. Federal and state gasoline taxes, which accounted for 3.5 cents/L (13 cents/gal) of this price, are returned to the road user through maintenance benefits. Hence, the actual gasoline operating cost was assumed to be 13.5 cents/L (51 cents/gal).

Winfrey (9) estimated the other automobile operating expenses (tires, oil, maintenance, and depreciation) on the basis of empirical data. By updating these prices to 1978 dollar values, it was found that the other operating costs were 1.980 and 1.806 cents/vehicle for major- and minor-road signalized approaches, respectively. As for

Figure 3. Annual cost for divided major-road intersections controlled by stop sign.

MAJOR VOLUME	MINOR VOLUME	ACCIDENT * COSTS	EQUIPMENT* COSTS	FUEL** MAJOR	FUEL** MINOR	MAJOR ** OTHER	MINOR** OTHER	DELAY COST**	ANNUAL COST
400.	50.	5941.	1587.	57.	14.	56.	16.	580.	271488.
600.	50.	6837.	1587.	85.	14.	84.	16.	846.	389981.
800.	50.	7732.	1587.	114.	14.	112.	16.	1119.	511121.
1000.	50.	8628.	1587.	142.	14.	141.	16.	1402.	635846.
1200.	50.	9523.	1587.	171.	14.	169.	16.	1698.	765425.
1400.	50.	10419.	1587.	199.	14.	197.	16.	2012.	901579.
400.	100.	6165.	1587.	57.	107.	56.	31.	673.	344913.
600.	100.	7060.	1587.	85.	107.	84.	31.	960.	471254.
800.	100.	7956.	1587.	114.	107.	112.	31.	1263.	603596.
1000.	100.	8852.	1587.	142.	107.	141.	31.	1590.	744212.
1200.	100.	9747.	1587.	171.	107.	169.	31.	1947.	896281.
1400.	100.	10643.	1587.	199.	107.	197.	31.	2349.	1064299.
400.	150.	6389.	1587.	57.	278.	56.	47.	808.	462576.
600.	150.	7284.	1587.	85.	278.	84.	47.	1146.	607540.
800.	150.	8180.	1587.	114.	278.	112.	47.	1524.	767147.
1000.	150.	9076.	1587.	142.	278.	141.	47.	1960.	947715.
1200.	150.	9971.	1587.	171.	278.	169.	47.	2479.	1158665.
1400.	150.	10867.	1587.	199.	278.	197.	47.	3120.	1414419.
400.	200.	6613.	1587.	57.	527.	56.	63.	1057.	650728..
600.	200.	7508.	1587.	85.	527.	84.	63.	1530.	845013.
800.	200.	8404.	1587.	114.	527.	112.	63.	2118.	1080912.
1000.	200.	9300.	1587.	142.	527.	141.	63.	2881.	1381100.
1200.	200.	10195.	1587.	171.	527.	169.	63.	3925.	1783608.
1400.	200.	11091.	1587.	199.	527.	197.	63.	5432.	2355146.
400.	250.	6837.	1587.	57.	856.	56.	79.	1672.	1000947.
600.	250.	7732.	1587.	85.	856.	84.	79.	2595.	1359255.
800.	250.	8628.	1587.	114.	856.	112.	79.	3959.	1878972.
1000.	250.	9523.	1587.	142.	856.	141.	79.	6119.	2688839.
1200.	250.	10419.	1587.	171.	856.	169.	79.	9797.	4052819.
1400.	250.	11315.	1587.	199.	856.	197.	79.	16596.	6555946.
400.	300.	7060.	1587.	57.	1262.	56.	94.	3955.	1988608.
600.	300.	7956.	1587.	85.	1262.	84.	94.	7258.	3215792.
800.	300.	8852.	1587.	114.	1262.	112.	94.	13765.	5612461.
1000.	300.	9747.	1587.	142.	1262.	141.	94.	27986.	10824430.
1200.	300.	10643.	1587.	171.	1262.	169.	94.	63092.	23659953.
1400.	300.	11539.	1587.	199.	1262.	197.	94.	163493.	60327835.

*Annual Cost
**Daily Cost

Figure 4. Annual cost for undivided major-road intersections controlled by stop sign.

MAJOR VOLUME	MINOR VOLUME	ACCIDENT * COSTS	EQUIPMENT* COSTS	FUEL ** MAJOR	FUEL** MINOR	MAJOR** OTHER	MINOR** OTHER	DELAY COST**	ANNUAL COST
400.	50.	5941.	1587.	57.	14.	56.	16.	587.	274082.
600.	50.	6837.	1587.	85.	14.	84.	16.	868.	398055.
800.	50.	7732.	1587.	114.	14.	112.	16.	1173.	530967.
1000.	50.	8628.	1587.	142.	14.	141.	16.	1520.	679070.
1200.	50.	9523.	1587.	171.	14.	169.	16.	1939.	853349.
1400.	50.	10419.	1587.	199.	14.	197.	16.	2484.	1073943.
400.	100.	6165.	1587.	57.	107.	56.	31.	656.	339028.
600.	100.	7060.	1587.	85.	107.	84.	31.	943.	465015.
800.	100.	7956.	1587.	114.	107.	112.	31.	1257.	601356.
1000.	100.	8852.	1587.	142.	107.	141.	31.	1620.	755353.
1200.	100.	9747.	1587.	171.	107.	169.	31.	2067.	939970.
1400.	100.	10643.	1587.	199.	107.	197.	31.	2664.	1179341.
400.	150.	6389.	1587.	57.	278.	56.	47.	727.	433057.
600.	150.	7284.	1587.	. 85.	278.	84.	47.	1019.	561378.
800.	150.	8180.	1587.	114.	278.	112.	47.	1345.	701703.
1000.	150.	9076.	1587.	142.	278.	141.	47.	1727.	862580.
1200.	150.	9971.	1587.	171.	278.	169.	47.	2207.	1059371.
1400.	150.	10867.	1587.	199.	278.	197.	47.	2865.	1321154.
400.	200.	6613.	1587.	57.	527.	56.	63.	798.	556346.
600.	200.	7508.	1587.	85.	527.	84.	63.	1098.	687374.
800.	200.	8404.	1587.	114.	527.	112.	63.	1437.	832333.
1000.	200.	9300.	1587.	142.	527.	141.	63.	1840.	1001259.
1200.	200.	10195.	1587.	171.	527.	169.	63.	2360.	1212432.
1400.	200.	11091.	1587.	199.	527.	197.	63.	3092.	1501099.
400.	250.	6837.	1587.	57.	856.	56.	79.	872.	708874.
600.	250.	7732.	1587.	85.	856.	84.	79.	1180.	843044.
800.	250.	8628.	1587.	114.	856.	112.	79.	1533.	993399.
1000.	250.	9523.	1587.	142.	856.	141.	79.	1963.	1171765.
1200.	250.	10419.	1587.	171.	856.	169.	79.	2529.	1399996.
1400.	250.	11315.	1587.	199.	856.	197.	79.	3350.	1721092.
400.	300.	7060.	1587.	57.	1262.	56.	94.	947.	890775.
600.	300.	7956.	1587.	85.	1262.	84.	94.	1266.	1028594.
800.	300.	8852.	1587.	114.	1262.	112.	94.	1636.	1185245.
1000.	300.	9747.	1587.	142.	1262.	141.	94.	2096.	1374711.
1200.	300.	10643.	1587.	171.	1262.	169.	94.	2718.	1623259.
1400.	300.	11539.	1587.	199.	1262.	197.	94.	3646.	1983717.

*Annual Cost
**Daily Cost

Figure 5. Annual cost for divided major-road intersections controlled by PR signal.

MAJOR VOLUME	MINOR VOLUME	ACCIDENT * COSTS	EQUIPMENT* COSTS	FUEL ** MAJOR	FUEL** MINOR	MAJOR** OTHER	MINOR** OTHER	DELAY COST**	ANNUAL COST
400.	50.	9072.	5645.	54.	7.	138.	16.	585.	306826.
600.	50.	11882.	5645.	81.	7.	208.	16.	849.	440898.
800.	50.	14692.	5645.	108.	7.	277.	16.	1114.	575827.
1000.	50.	17502.	5645.	135.	7.	346.	16.	1383.	711722.
1200.	50.	20312.	5645.	162.	7.	415.	16.	1654.	848701.
1400.	50.	23122.	5645.	189.	7.	485.	16.	1929.	986898.
400.	100.	9774.	5645.	54.	14.	138.	31.	653.	340582.
600.	100.	12584.	5645.	81.	14.	208.	31.	919.	475317.
800.	100.	15394.	5645.	108.	14.	277.	31.	1186.	611007.
1000.	100.	18204.	5645.	135.	14.	346.	31.	1457.	747772.
1200.	100.	21014.	5645.	162.	14.	415.	31.	1731.	885746.
1400.	100.	23825.	5645.	189.	14.	485.	31.	2009.	1025077.
400.	150.	10477.	5645.	54.	20.	138.	47.	722.	374684.
600.	150.	13287.	5645.	81.	20.	208.	47.	990.	510161.
800.	150.	16097.	5645.	108.	20.	277.	47.	1260.	646702.
1000.	150.	18907.	5645.	135.	20.	346.	47.	1533.	784442.
1200.	150.	21717.	5645.	162.	20.	415.	47.	1810.	923529.
1400.	150.	24527.	5645.	189.	20.	485.	47.	2091.	1064132.
400.	200.	11179.	5645.	54.	27.	138.	63.	792.	409179.
600.	200.	13989.	5645.	81.	27.	208.	63.	1062.	545488.
800.	200.	16799.	5645.	108.	27.	277.	63.	1335.	682983.
1000.	200.	19609.	5645.	135.	27.	346.	63.	1611.	821815.
1200.	200.	22419.	5645.	162.	27.	415.	63.	1892.	962151.
1400.	200.	25230.	5645.	189.	27.	485.	63.	2177.	1104179.
400.	250.	11882.	5645.	54.	34.	138.	79.	864.	444125.
600.	250.	14692.	5645.	81.	34.	208.	79.	1136.	581365.
800.	250.	17502.	5645.	108.	34.	277.	79.	1411.	719930.
1000.	250.	20312.	5645.	135.	34.	346.	79.	1691.	859986.
1200.	250.	23122.	5645.	162.	34.	415.	79.	1976.	1001723.
1400.	250.	25932.	5645.	189.	34.	485.	79.	2265.	1145350.
400.	300.	12584.	5645.	54.	41.	138.	94.	936.	479584.
600.	300.	15394.	5645.	81.	41.	208.	94.	1211.	617870.
800.	300.	18204.	5645.	108.	41.	277.	94.	1490.	757634.
1000.	300.	21014.	5645.	135.	41.	346.	94.	1774.	899064.
1200.	300.	23825.	5645.	162.	41.	415.	94.	2063.	1042373.
1400.	300.	26635.	5645.	189.	41.	485.	94.	2357.	1187797.

*Annual Cost
**Daily Cost

Figure 6. Annual cost for divided-highway intersections controlled by SA signal.

MAJOR VOLUME	MINOR VOLUME	ACCIDENT * COSTS	EQUIPMENT * COSTS	FUEL ** MAJOR	FUEL ** MINOR	MAJOR** OTHER	MINOR** OTHER	DELAY COST **	ANNUAL COST
400.	50.	9072.	6730.	1.	0.	138.	16.	575.	282189.
600.	50.	11882.	6730.	2.	0.	208.	16.	850.	411039.
800.	50.	14692.	6730.	3.	0.	277.	16.	1158.	552137.
1000.	50.	17502.	6730.	8.	0.	346.	16.	1538.	720762.
1200.	50.	20312.	6730.	23.	1.	415.	16.	2083.	953447.
1400.	50.	23122.	6730.	198.	7.	485.	16.	3036.	1395489.
400.	100.	9774.	6730.	1.	0.	138.	31.	643.	313645.
600.	100.	12584.	6730.	5.	1.	208.	31.	925.	446207.
800.	100.	15394.	6730.	11.	1.	277.	31.	1248.	594623.
1000.	100.	18204.	6730.	31.	3.	346.	31.	1663.	781976.
1200.	100.	21014.	6730.	184.	15.	415.	31.	2290.	1099522.
1400.	100.	23825.	6730.	387.	28.	485.	31.	3465.	1634998.
400.	150.	10477.	6730.	5.	2.	138.	47.	712.	346993.
600.	150.	13287.	6730.	11.	3.	208.	47.	1002.	484001.
800.	150.	16097.	6730.	34.	6.	277.	47.	1345.	646965.
1000.	150.	18907.	6730.	166.	25.	346.	47.	1804.	897288.
1200.	150.	21717.	6730.	347.	43.	415.	47.	2540.	1266526.
1400.	150.	24527.	6730.	577.	62.	485.	47.	4023.	1926721.
400.	200.	11179.	6730.	10.	5.	138.	63.	782.	382723.
600.	200.	13989.	6730.	33.	11.	208.	63.	1084.	531302.
800.	200.	16799.	6730.	142.	36.	277.	63.	1452.	742376.
1000.	200.	19609.	6730.	300.	60.	346.	63.	1967.	1025265.
1200.	200.	22419.	6730.	508.	85.	415.	63.	2848.	1459410.
1400.	200.	25230.	6730.	765.	109.	485.	63.	4769.	2291712.
400.	250.	11882.	6730.	27.	17.	138.	79.	856.	426047.
600.	250.	14692.	6730.	114.	47.	208.	79.	1171.	612283.
800.	250.	17502.	6730.	250.	78.	277.	79.	1570.	846786.
1000.	250.	20312.	6730.	435.	109.	346.	79.	2159.	1168961.
1200.	250.	23122.	6730.	670.	140.	415.	79.	3236.	1686865.
1400.	250.	25932.	6730.	954.	170.	485.	79.	5797.	2764777.
400.	300.	12584.	6730.	81.	60.	138.	94.	932.	495881.
600.	300.	15394.	6730.	195.	97.	208.	94.	1265.	700779.
800.	300.	18204.	6730.	358.	134.	277.	94.	1703.	961903.
1000.	300.	21014.	6730.	571.	171.	346.	94.	2390.	1331567.
1200.	300.	23825.	6730.	832.	208.	415.	94.	3737.	1960163.
1400.	300.	26635.	6730.	1142.	245.	485.	94.	7261.	3401376.

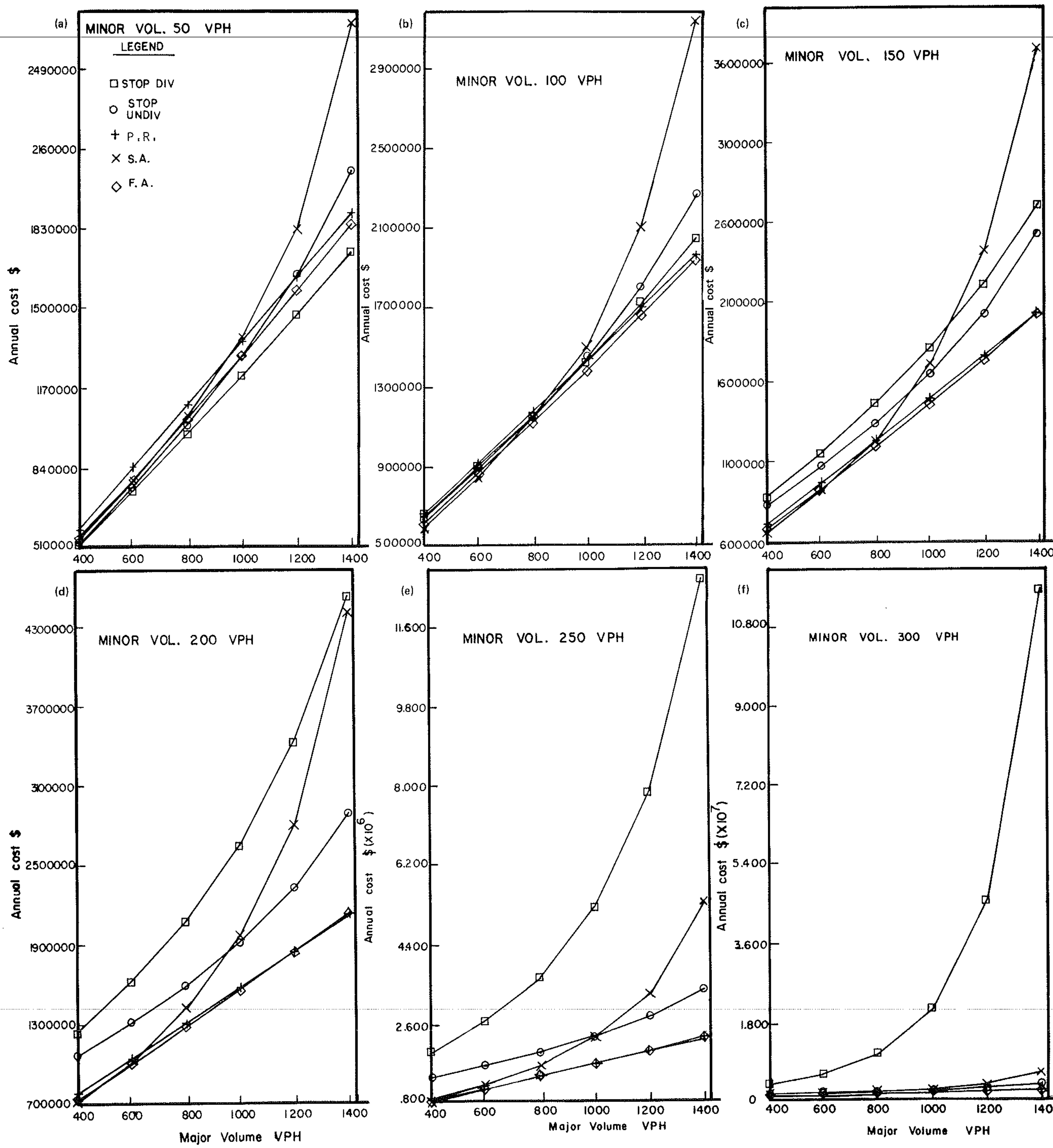
*Annual Cost
**Daily Cost

Figure 7. Annual cost for divided-highway intersections controlled by FA signals.

MAJOR VOLUME	MINOR VOLUME	ACCIDENT* COSTS	EQUIPMENT * COSTS	FUEL ** MAJOR	FUEL** MINOR	MAJOR ** OTHER	MINOR** OTHER	DELAY COST**	ANNUAL COST
400.	50.	9072.	13522.	7.	1.	138.	16.	573.	290991.
600.	50.	11882.	13522.	21.	2.	208.	16.	831.	418679.
800.	50.	14692.	13522.	42.	3.	277.	16.	1092.	549914.
1000.	50.	17502.	13522.	69.	3.	346.	16.	1357.	684938.
1200.	50.	20312.	13522.	104.	4.	415.	16.	1625.	824045.
1400.	50.	23122.	13522.	145.	5.	485.	16.	1900.	967594.
400.	100.	9774.	13522.	14.	3.	138.	31.	640.	324981.
600.	100.	12584.	13522.	31.	5.	208.	31.	900.	454907.
800.	100.	15394.	13522.	55.	7.	277.	31.	1164.	588586.
1000.	100.	18204.	13522.	85.	9.	346.	31.	1431.	726306.
1200.	100.	21014.	13522.	123.	10.	415.	31.	1704.	868421.
1400.	100.	23825.	13522.	168.	12.	485.	31.	1984.	1015363.
400.	150.	10477.	13522.	20.	8.	138.	47.	707.	360101.
600.	150.	13287.	13522.	40.	10.	208.	47.	970.	492296.
800.	150.	16097.	13522.	67.	13.	277.	47.	1237.	628617.
1000.	150.	18907.	13522.	101.	15.	346.	47.	1509.	769284.
1200.	150.	21717.	13522.	142.	18.	415.	47.	1787.	914723.
1400.	150.	24527.	13522.	190.	20.	485.	47.	2073.	1065458.
400.	200.	11179.	13522.	27.	13.	138.	63.	777.	396206.
600.	200.	13989.	13522.	50.	17.	208.	63.	1042.	531162.
800.	200.	16799.	13522.	81.	20.	277.	63.	1313.	670414.
1000.	200.	19609.	13522.	117.	23.	346.	63.	1590.	814101.
1200.	200.	22419.	13522.	161.	27.	415.	63.	1874.	963251.
1400.	200.	25230.	13522.	212.	30.	485.	63.	2167.	1118267.
400.	250.	11882.	13522.	33.	21.	138.	79.	848.	433533.
600.	250.	14692.	13522.	60.	25.	208.	79.	1117.	571308.
800.	250.	17502.	13522.	93.	29.	277.	79.	1392.	713739.
1000.	250.	20312.	13522.	134.	33.	346.	79.	1675.	861332.
1200.	250.	23122.	13522.	181.	38.	415.	79.	1966.	1014713.
1400.	250.	25932.	13522.	235.	42.	485.	79.	2269.	1174273.
400.	300.	12584.	13522.	39.	29.	138.	94.	921.	472232.
600.	300.	15394.	13522.	69.	35.	208.	94.	1195.	613110.
800.	300.	18204.	13522.	106.	40.	277.	94.	1476.	759079.
1000.	300.	21014.	13522.	150.	45.	346.	94.	1765.	910755.
1200.	300.	23825.	13522.	200.	50.	415.	94.	2066.	1068904.
1400.	300.	26635.	13522.	258.	55.	485.	94.	2380.	1234472.

*Annual Cost
**Daily Cost

Figure 8. Annual cost for different levels of minor-road volumes.



stop-controlled intersections, the values were 0.806 and 1.806 cents/vehicle.

Delay Cost

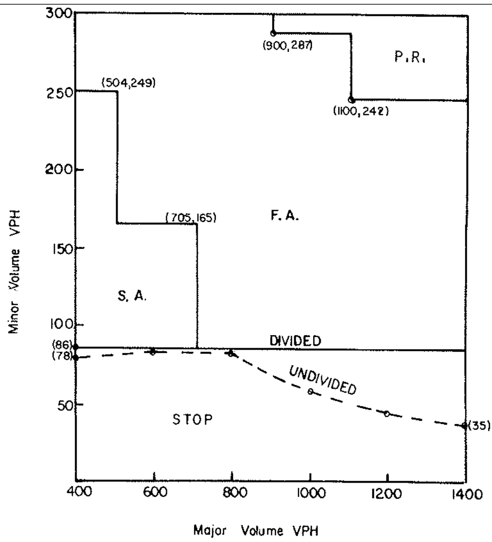
A previous study by Thomas and Thompson (10) provided the magnitude of savings in travel time as a function of motorist income by trip purpose. Knowing the median family income in Indiana, and assuming an average automobile occupancy of 1.5 persons/vehicle and that one-third of the trips on rural highways are for work and the rest for other purposes, the magnitude of delay costs was estimated.

Total Annual Cost

By substituting the individual cost items in the following equation, the total annual cost per combination of traffic volumes on major and minor roads was calculated for each control alternative separately:

$$TAC = AAC + EACMC + \left[\sum_{i=1}^{2} (AFC_i + AOOC_i) \times ADT_i + ADC \times ADT_3 \right] \times 365 \qquad (1)$$

Figure 9. Control alternatives for different major- and minor-road volumes on the basis of minimum total annual cost.



where

TAC = total annual cost ($),
AAC = annual accident cost ($),
EAMAC = equivalent annual construction and maintenance cost ($),
AFC_1 = average fuel cost for major approach ($/vehicle),
AFC_2 = average fuel cost for minor approach ($/vehicle),
$AOOC_1$ = average other operating cost for major approach ($/vehicle),
$AOOC_2$ = average other operating cost for minor approach ($/vehicle),
ADT_1 = average daily traffic for major approach (vehicles/day),
ADT_2 = average daily traffic for minor approach (vehicles/day),
ADC = average delay cost ($/vehicle), and
$ADT_3 = ADT_1 + ADT_2$.

In order to estimate the ADT values, the hourly volumes generated by the simulation model were adjusted by a ratio of peak-hour to average daily traffic. Since peak-hour traffic is not as apparent on rural facilities as it is on urban facilities, a minimum threshold ratio between ADT and peak-hour volumes of 8 percent was used.

The detailed individual cost estimates of the control alternatives are given for each level of major- and minor-road volume combination in Figures 3-7. Furthermore, the total annual cost values are included. These values were then plotted against major-road volumes for different levels of minor-road volumes (50, 100, 150, 200, 250, and 300 vehicles/h), and they are shown in Figure 8. To develop traffic control warrants based solely on traffic demands of major and minor roads, an accepted criterion should be selected. By using the total annual cost as the adopted criterion, a chart (Figure 9) was developed to show boundary lines for the control alternatives under consideration. Similar charts can be developed for other cost criteria.

A closer look at the chart reveals that stop-sign control is warranted for a divided highway when the minor-road traffic volume is less than about 90 vehicles/h for any major-road traffic volume. This value varies from about 80 vehicles/h for a major-road traffic volume of 400 vehicles/h to 35 vehicles/h for a major-road traffic volume of 1400 vehicles/h in the case of an undivided highway. A PR signal is warranted when the minor-road traffic volume is greater than about 240 vehicles/h and the major-road traffic volume is greater than about 1100 vehicles/h.

The warrant for the PR signal rather than the FA control at high traffic volumes on both major and minor roads is probably called for because, as the minor-road volume increases, a queue builds up at the minor-road approach to the intersection, and the fixed cycle length of a PR signal causes less delay than extending the green phase of an FA control to its maximum extension duration.

A control unit equipped with a gap-reduction feature and a variable initial-interval feature (volume-density control) was thought to be a better alternative for reducing average delay per vehicle for this situation until several simulation runs with a volume-density control were tried. The results showed that the average delays per vehicle experienced by using this control were higher than those observed for the PR and FA controls. Since the capital and maintenance costs for the volume-density control are higher than the costs for PR signals, it is doubtful that volume-density control would be a better alternative at high volumes.

CONCLUSIONS

A methodology for selecting the best control alternative at isolated intersections on high-speed rural highways has been presented in this paper. The warrants for a specific traffic control alternative, as indicated in Figure 9, were based on the criterion of minimum total annual cost. The annual cost included (a) annual accident cost, (b) construction and maintenance cost of equipment, (c) fuel cost, (d) nonfuel operating cost, and (e) delay cost.

Similar analysis could be done by considering any of the individual cost items. In addition, the information generated by the model can also be used to evaluate trade-off relationships involving the various cost items with respect to the different control alternatives. Such an evaluation can provide an indication of the relative advantage gained by one control over other alternatives, with respect to, say, safety versus gasoline use.

ACKNOWLEDGMENT

This research was conducted with the financial support of the Joint Highway Research Project of Purdue University and the Indiana State Highway Commission. We are solely responsible for the results and opinions expressed in the article.

REFERENCES

1. Manual on Uniform Traffic Control Devices for Streets and Highways. Federal Highway Administration, 1971.
2. S. L. Cohen. Application of Network Simulation Models to the Analysis of Urban Intersection Performance. Federal Highway Administration, Rept. FHWA-RD-73-95, Sept. 1973.
3. D. L. Hall. Energy, Air Pollution, Delay, and Safety Evaluation of Non-Signalized Control at Low-Volume Intersections. Highway Extension and Research Project for Indiana Counties, Purdue Univ., West Lafayette, IN, Final Rept., Aug. 1977.
4. F. A. Wagner, Jr. An Evaluation of Fundamental Driver Decisions and Reactions at an Intersection.

HRB, Highway Research Record 118, 1966, pp. 68-84.
5. Automobile Exhaust Emission Modal Analysis Model. Office of Air and Water Programs, U.S. Environmental Protection Agency, EPA-460/3-74-024, Oct. 1974.
6. S. S. Hejal and H. L. Michael. The Estimated Direct Costs (1970) of Traffic Accidents on Indiana Rural State Highways. Purdue Univ., West Lafayette, IN, Joint Highway Research Project No. 20, 1970.
7. P. Abramson. An Accident Evaluation Analysis. TRB, Transportation Research Record 486, 1974, pp. 33-41.
8. H. Wuerdemann. National Indirect Costs of Motor Vehicle Accidents. Center for the Environment and Man, July 1973.
9. R. Winfrey. Economic Analysis for Highways. Intext Educational Publications, New York, 1969.
10. T. C. Thomas and G. I. Thompson. Value of Time Saved by Trip Purpose. HRB, Highway Research Record 369, 1971, pp. 104-115.

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

Use of EC-DC Detector for Signalization of High-Speed Intersections

Peter S. Parsonson, Georgia Institute of Technology, Atlanta
Richard A. Day, Wilbur Smith and Associates, Columbia, South Carolina
James A. Gawlas and George W. Black, Jr., Traffic Engineering Division, Gwinnett County, Georgia

This paper describes a new detector-controller configuration intended to minimize the dilemma-zone problem at signalized intersections on high-speed roads. Included are a complete functional and electrical description of the design, the findings of a field test, and a comparison with two existing designs. The new design uses a basic, actuated, digital controller operated in the nonlocking mode. An approach that has a design speed of 89 km/h (55 mph) has an upstream detection loop located 117 m (384 ft) back from the intersection and a middle loop 77 m (254 ft) back. A loop at the stopline is 8 m (25 ft) in length and is connected to a novel extended-call–delayed-call (EC-DC) detector that is able to change from an EC model to a DC unit at the strategic moment during the green interval. In effect, the change disconnects the stopline loop, leaving the other two loops to control the extension and termination of the green. The controller and detectors are off-the-shelf units that require no internal modification. The only special-logic items are two relays mounted on the back panel. The design does not pose a maintenance problem. A test installation in Georgia significantly reduced conflicts associated with the dilemma zone. A comparison with two existing designs shows that the EC-DC configuration costs somewhat more than the EC design but is superior in three operational categories. It is less expensive than the density design and is superior in four operational aspects.

Drivers face a "dilemma zone" or "zone of indecision" at signalized intersections where approach speeds are 56 km/h (35 mph) or higher. If the yellow comes on while the driver is in this zone, an abrupt stop may produce a rear-end collision. The decision to go through, on the red, may produce a right-angle accident. The traffic engineer can install a vehicle-actuated signal controller and appropriate detection in order to attempt to minimize the untimely display of yellow. This paper describes a new detector-controller design to meet this objective.

The new design was made possible by the following three recent breakthroughs in driver-behavior research and hardware technology:

1. The boundaries of the dilemma zones for various approach speeds are now known with reasonable accuracy.
2. Digital controllers permit the unit extension to be tailored to tenths of a second.
3. Loop detectors are now available off the shelf that have both extended-call (EC) and delayed-call (DC) features as standard equipment.

A number of advanced detector-controller configurations for isolated intersections on high-speed roads have been reported in the literature. Most of these were summarized in a design manual published by the Federal Highway Administration in 1977 (1). Concurrently, Zegeer (2) reported excellent results in Kentucky with a particular configuration known as a green-extension system. In 1978, Parsonson (3) discussed the state of the art and concluded that there appears to be an unmet need for a design for high-speed roads that has the following characteristics: (a) loop-occupancy features; (b) basic, actuated controller with nonlocking detection memory; (c) EC detection; (d) short allowable gap, primarily to prevent frequent "max-out" (which may well show a yellow to a vehicle in the dilemma zone); and (e) dilemma-zone protection over a wide range of speeds. Parsonson then proposed a new detector-controller configuration to meet these objectives and analyzed it on the basis of seven criteria set forth for evaluating any proposed design. In response to a discussion of the paper by Clark (4), Parsonson modified the design in his closure (3, p. 42) and stated that his proposal would soon be field tested in Gwinnett County, Georgia.

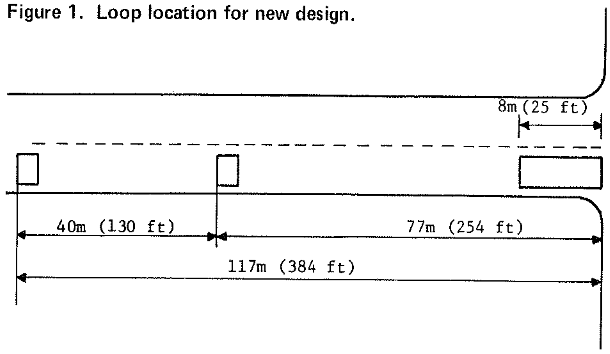
The present paper provides a complete functional and electrical description of the modified design, reports the findings of the Gwinnett field test, and compares the new configuration with the most commonly used existing designs.

FUNCTIONAL OPERATION

The design uses a basic, actuated controller operated in the nonlocking detection-memory mode. Figure 1 shows the detection loops located for approach speeds of up to 89 km/h (55 mph). The upstream detector is located 117 m (384 ft) from the intersection. The table below [from Zeeger (2)] shows that 90 percent of drivers approaching at 89 km/h will decide to stop if the yellow begins just as they reach this point (1 km/h = 0.6 mph; 1 m = 3.3 ft):

Approach Speed (km/h)	Distance from Intersection (m) Probability of Stopping 10 Percent	90 Percent
56	31	77
64	37	86
72	46	99
80	52	107
89	71	117

Figure 1. Loop location for new design.



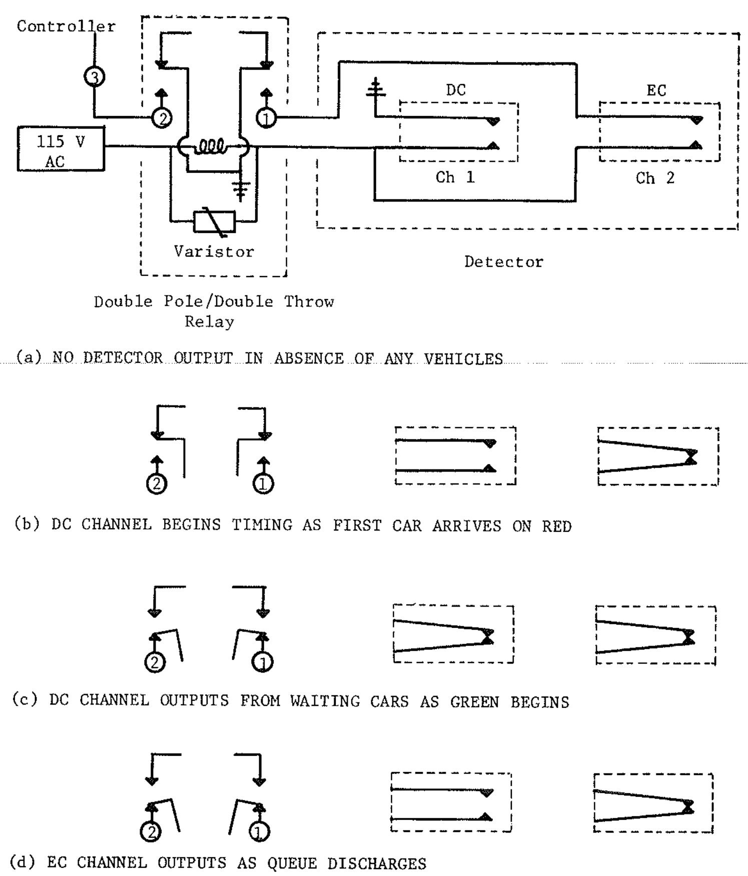
The middle loop is placed 77 m (254 ft) from the intersection. As the table indicates, this is the upstream boundary of the dilemma zone for vehicles approaching at 56 km/h (35 mph).

Both of these loops are small, perhaps 2x2 m (6x6 ft). Their "amplifiers" are of normal design (not EC). They may produce a short pulse when the vehicle enters the loops, or they may be operated in the presence mode.

The loop at the stopline is 8 m (25 ft) in length, which is intended to be long enough to bridge the gap between waiting vehicles, thereby assuring a call from a queue. The detector is a novel EC-DC unit that is able to change from an EC model to a DC unit at the strategic moment during the green interval. Each mode of operation has its own adjustable timer.

A description of the operation of the configuration begins with a start of green. As the waiting vehicles discharge over the stopline loop, its EC-DC detector functions as an EC model. (It has a carry-over output, meaning that it holds or "stretches" the call of a vehicle for a period of seconds that has been set on an adjustable timer.) The controller, meanwhile, is timing a minimum green that Clark (4) has found may need to be as long as 12-18 s in order to meet the expectations of truck drivers. On expiration of the minimum green, probably only five or six vehicles have discharged and there is still no motion over either of the upstream loops. A "stretch" setting of approximately 2 s on the EC-DC detector is intended to produce an unbroken actuation (and an extension of the green) until motion is assured over the middle loop. By this time, discharging traffic is up to speed. A 2-s gap between vehicles appears. The EC detector at the stopline in effect "gaps out" and becomes a DC unit with approxi-

Figure 2. Electrical operation of EC-DC detection at stopline.



mately 5 s of time delay. The full-speed vehicles in transit over the stopline loop do not produce a call. This loop has in effect become disconnected, and the continued extension of the green is controlled by the upstream loops and the unit extension setting of the controller.

If the upstream loops use an amplifier that produces a short pulse when the vehicle enters the loop, then the unit extension setting of the controller is selected to be 2.2 s. It is simple to show that this setting will carry vehicles approaching within the speed range of 64-89 km/h (40-55 mph) through their respective dilemma zones. For example, a vehicle arriving on the green at a speed of 64 km/h will receive a 2.2-s unit extension that will prolong the green until the vehicle reaches the middle detector. The second unit extension holds the green until the vehicle is 37 m (124 ft) from the intersection. The table above shows that this is the downstream boundary of the dilemma zone for that speed, so the design succeeds in delaying the start of the yellow until the vehicle has cleared its dilemma zone.

If the upstream loops use an electronics unit operating in the presence mode, then an actuation will not end until the rear of the vehicle leaves the loop. In this case, a unit extension of 1.9 s is used instead of 2.2 s.

It is useful to apply seven criteria to judge the effectiveness of any proposed configuration for high-speed intersections (3):

1. Does the design detect a vehicle approaching at the design speed before it reaches the dilemma zone? The table above shows that the dilemma zone for 89 km/h begins 117 m (384 ft) from the intersection. Therefore, the design of Figure 1 does meet this criterion.

2. What is the allowable gap imposed by this design? A long allowable gap produces sluggish transfer of the green, increasing both delay and the likelihood of losing dilemma-zone protection because of max-out. With a unit extension of 2.2 s, the allowable gap produced by the two upstream loops is calculated as the travel time from the first loop to the middle one, plus the 2.2 s of extension by the middle loop. A stream of vehicles at the design speed of 89 km/h would thus experience an allowable gap of 1.6 + 2.2 = 3.8 s, while a stream at 64 km/h (40 mph) would hold the green if their time headways did not exceed 2.2 + 2.2 = 4.4 s. These allowable gaps appear to be snappy enough to minimize the extension of green to a maximum interval set to 50-60 s or more.

3. On termination of the green by gap-out, will vehicles approaching at the design speed be clear of the dilemma zone? On gap-out, a vehicle approaching at 89 km/h will be 2.2 s, or 54 m (178 ft), past the middle detector. Figure 1 shows that this is 77 - 54 = 23 m (75 ft) short of the stopline. The table above indicates that at this point the vehicle has left the dilemma zone well behind.

4. On termination of the green by gap-out, will vehicles traveling more slowly than the design speed be clear of the dilemma zone? The initial explanation of this design, above, pointed out that a vehicle approaching at 64 km/h will have just reached the downstream boundary of its dilemma zone on gap-out. All approach speeds from 64 km/h to 89 km/h receive dilemma-zone protection from this design. A slower vehicle, at 56 km/h (35 mph) is also protected because the yellow will appear before the vehicle reaches its dilemma zone. The first detector's unit extension of 2.2 s is not enough to carry this straggler to the middle detector, so gap-out occurs upstream of the dilemma zone.

5. Can a queue waiting at the stopline get into motion without a premature gap-out? The amount of stretch set on the EC timer of the EC-DC stopbar loop is selected to prevent premature gap-out, as explained earlier. The stretch must be long enough to assure motion over the middle loop, yet short enough to assure that the stopbar loop "disconnects" as soon as discharging traffic gets up to speed. The stopbar loop must disconnect (by reverting to the DC mode) before full-speed traffic gaps out over the upstream detectors.

6. Can the design screen out false calls for the green [as, for example, with right turn on red (RTOR)]? The nonlocking controller mode is capable of screening out false calls that appear while the cross-street traffic is holding the green. When the green is at rest on the cross street, the DC mode of the EC-DC stopbar detector will screen RTOR vehicles that enter the approach from a driveway less than 77 m (254 ft) from the intersection. However, with the green at rest on the cross street, a false call at either of the two upstream detectors will bring the green unnecessarily.

7. During the green interval, can a queue of left-turning vehicles hold the green as they wait to filter through gaps in oncoming traffic? This is important on two-lane roads, where an occasional left-turning vehicle can cause a queue to form. When a gap in oncoming traffic appears, a gap-out may occur with some designs before the queue can get into motion over a detector upstream of the stopline. In this design, the EC-DC loop at the stopline permits a queue of left-turning vehicles to hold the green.

ELECTRICAL DESIGN

The test installation (described below) used digital loop detectors. Therefore, this section will discuss the electrical design in terms of "channels" of detection. Three two-channel detectors would be used at a two-phase intersection of a high-speed route that has a low-speed crossroad.

One channel of one detector is connected to the four upstream, small-area loops. Figure 1 shows two of these loops; the other two would be on the opposite approach.

The second channel of that detector is connected to the loops used on the crossroad. Each of these two approaches might have an 18-m (60-ft) loop at the stopline.

Both channels of the second detector are used for one of the EC-DC loops. For example, they could be used with the long loop at the stopline of the approach shown in Figure 1. Both channels of the third detector are associated with the other EC-DC loop, which would be at the opposite stopline of an expanded Figure 1. The operation of these four channels is explained with the aid of Figure 2.

Figure 2a shows a two-channel detector in the dashed-line box on the right. Channel 1 has been switched by hand to the DC mode of operation. Channel 2 has been set to the EC mode. These channels would both be assigned to a stopline loop, as explained above.

Figure 2a also shows a double-pole, double-throw relay that would be installed on the controller cabinet's back panel by the local staff. The relay is connected to the detector, cabinet power, and the controller phase's detector input terminal as shown. A metal-oxide varistor (MOV) is included to eliminate the AC noise coming from the 115-V power source. It is a model GE V150LA20.

In the absence of a vehicle in the loop, Figure 2a shows that neither detector channel is actuated; both sets of output-relay contacts are open. Similarly, the

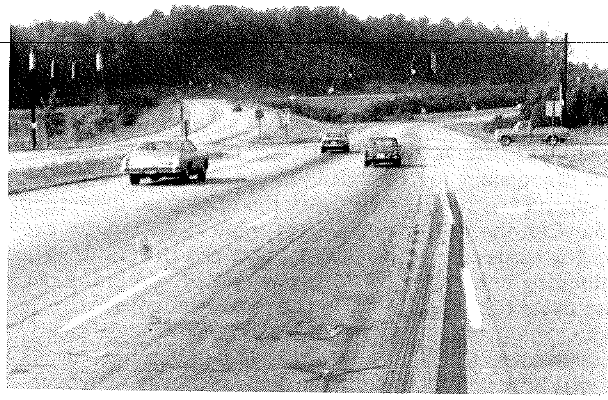


Figure 3. Site of test installation.

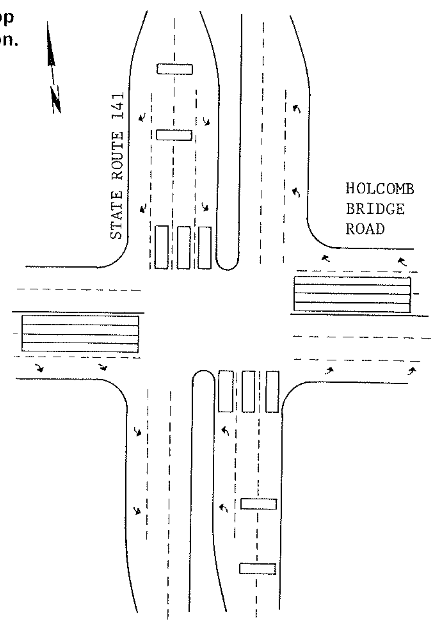


Figure 4. Detection-loop layout of test installation.

Figure 5. Upstream loops do not detect the turn lanes.

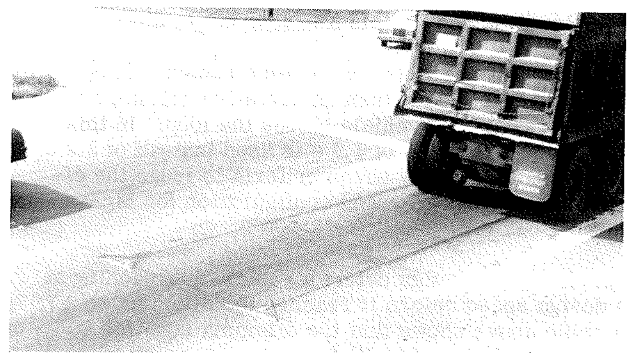


Figure 6. EC-DC loops at the stopline are only 8 m (25 ft) long.

open contacts in the relay to the left indicate that no call is going to the controller.

In Figure 2b, a vehicle has arrived and come to a stop on the loop (because the signal is red). Channel 1 begins to time its delay in issuing a call; meanwhile, the channel 2 contacts close. It is clear from Figure 2a that contact 1 in the relay will not close as a result of this arrival, and no call goes to the controller.

When channel 1 times out its delay, its contacts close, as shown in Figure 2c. A ground is applied to one side of the relay's coil, energizing it and causing contacts 1 and 2 to close. A call in the form of a ground is applied to the controller phase's detector input terminal (contact 3). The controller responds by changing the green signal to this phase at the earliest opportunity.

When the first vehicle leaves the loop on the green interval, the channel 1 contacts open. That ground is removed from the coil. Channel 2 remains energized for a time, however, and the call is extended by means of the ground provided in the relay. As the queue discharges, the loop becomes occupied again and again before the set extension is able to time out. The call to the controller is unbroken; see Figure 2d. Soon the vehicles are crossing the loop at full speed and a gap appears that allows channel 2 to time out an extension. The time-out opens the channel 2 contacts, deenergizing the coil and opening contacts 1 and 2 in the relay. The call to the controller ends, and the system returns to Figure 2a.

TEST INSTALLATION

The design was tested in July 1978, at the intersection of GA-141 and Holcomb Bridge Road in Gwinnett County (in greater Atlanta). Figure 3, looking south on GA-141, shows an open, rural intersection. The skid marks in the dilemma zone were found before the new design went into effect.

In both directions on GA-141, the peak-period speeds at the 15th, 50th, and 85th percentiles are 69, 77, and 86 km/h (43, 48, and 54 mph), respectively. Southbound volumes peak during the morning commuter rush at 550 vehicles/h. The northbound peak, in the afternoon, is close to 700. This four-lane highway has an average daily traffic (ADT) of 17 100 vehicles/day. Holcomb Bridge Road carries an ADT of 10 800 vehicles/day on four lanes.

Before the new detector-controller configuration was installed, GA-141 was signalized by using a small-area detection loop 58 m (190 ft) from the intersection. The table above can be interpolated to show that the median-speed vehicle approaching at 77 km/h (48 mph) has a 22 percent chance of stopping if shown a yellow interval at that location. Therefore, the "before" fully actuated signalization provided an insignificant amount of dilemma-zone protection.

Figure 4 shows the detection-loop layout of the test installation. The loops on the high-speed north-south GA-141 are identical to the design set forth in concept in Figure 1. The upstream loops on each of these approaches (Figure 5) do not detect the low-speed turning

Figure 7. The design requires a digital controller.

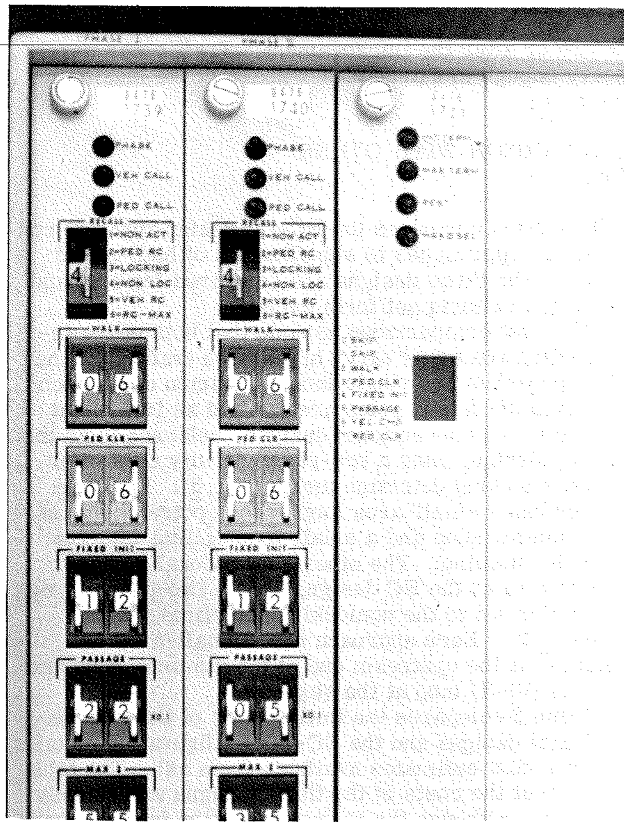
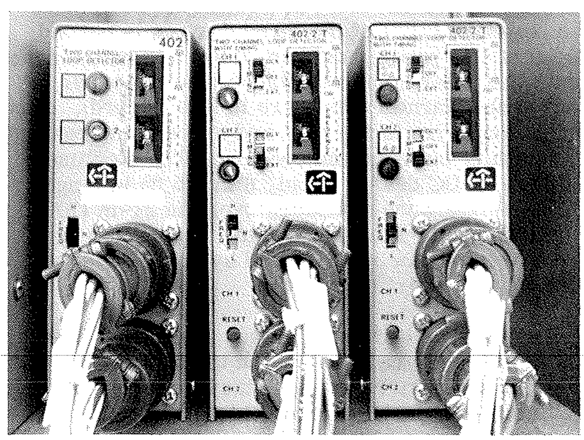


Figure 8. Six channels of detection are required.



traffic. An EC-DC loop at the stopline is shown in Figure 6. The southbound right-turn-only lane in Figure 4 has no detection loop at the stopline because vehicles are permitted to turn right on red.

The east-west cross street (Figure 4) uses conventional loop-occupancy control. Each approach lane has a quadrupole loop 18 m (60 ft) long operated in the presence mode.

The digital controller, shown in Figure 7, has the settings given below.

Item	Phase 1 (GA-141)	Phase 2 (Holcomb Bridge Road)
Detection memory	Nonlock	Nonlock
Fixed initial (s)	12	12

Figure 9. One double-pole, double-throw relay for each high-speed approach.

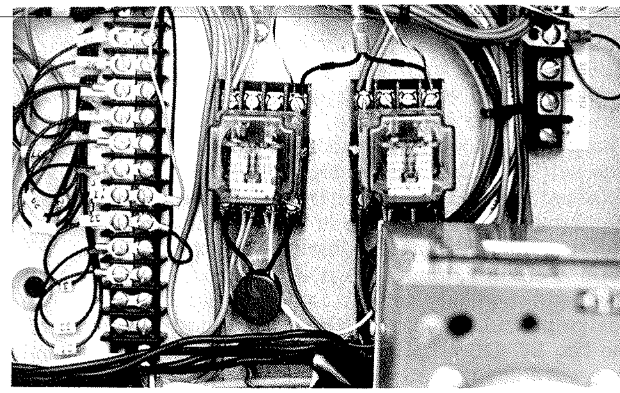
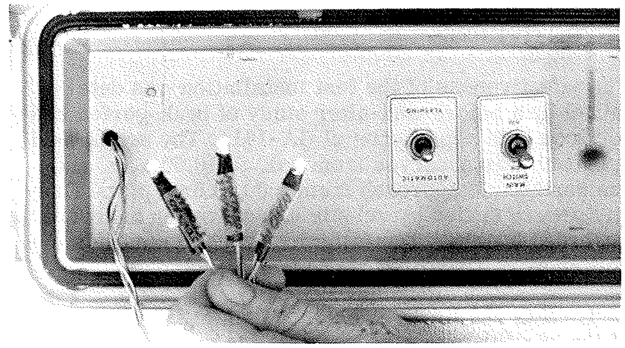


Figure 10. Three pilot lights are stored behind the police panel.



Item	Phase 1 (GA-141)	Phase 2 (Holcomb Bridge Road)
Passage time (s)	2.2	0.5
Maximum interval (s)	55	35
Yellow change (s)	4.3	4.3
Red clearance (s)	1.5	1.7

The yellow-plus-all-red clearance period is unchanged from the "before" settings. The 14 s of fixed initial plus passage time for phase 1 were derived from the Clark recommendations of 12-18 s (4). The 2.2 s of passage time was derived earlier. The maximum setting of 55 s was a compromise of the 50-60 s commonly used and would have been set higher if necessary to avoid max-outs.

The three detector units shown in Figure 8 provide the required six channels of detection. The detector on the left in the figure uses channel 1 to operate the four upstream loops on GA-141. The long stopline loops on Holcomb Bridge Road use channel 2. The center detector and the right one were designed at the factory so that any channel can be switch-selected to either the DC or EC mode of operation. The center detector, serving the stopline loop for the northbound approach on GA-141, is set to DC for channel 1 and EC for channel 2. The right detector is set similarly for the southbound approach. Switches inside these two detectors are set to time 5 s of delay and 2 s of extension at this intersection.

The double-pole, double-throw relays explained earlier are shown in Figure 9. One is associated with the center detector unit and the other with the right unit. The fact that these two relays are the only special logic in the cabinet points up the simplicity of the design. It does not pose a maintenance problem.

Figure 10 shows three pilot lights that were added for convenience in monitoring the operation of the design. The left lamp glows during GA-141 green. It was desired because the north-south signal heads are not visible from the cabinet. The center and right lamps are associated with the stopline loops on the northbound and southbound approaches, respectively. Each glows whenever its loop is outputting a call to the controller. The setting of 2 s of extension on each approach was obtained by trial and error by using these pilot lights. The lamp for an approach is extinguished when its stopline loop disconnects during the green. As explained earlier, the disconnect must occur at the strategic moment, neither too early nor too late. The design provides separate electronics units for each of the two EC-DC stopline loops. If a single two-channel unit were used, instead of the two shown at the center and right of Figure 8, then a timely disconnect of a stopline loop would become difficult. Heavy traffic in both directions would be much more likely to cause extension of the green to the maximum interval.

EFFECTIVENESS

The effectiveness of the test installation was determined by a before-and-after study of peak-period conflicts on both approaches of GA-141. The southbound approach was observed from 7:00 to 10:00 a.m. and the northbound approach from 3:00 to 6:00 p.m. Conflicts were tallied by vehicle classification and were recorded hourly. The following types of conflicts observed were suggested by Zeegeer (2):

1. Run red light: A clearing vehicle is upstream of the stopline when the signal turns red.
2. Abrupt stop: A driver decides at the last instant to stop. The deceleration, particularly within 30 m (100 ft) of the stopline, causes the front end of the vehicle to dip noticeably.
3. Skid: A stop is so abrupt as to produce the sound of skidding.
4. Swerve: An erratic maneuver narrowly averts a collision.
5. Acceleration through yellow: The driver "guns" the engine in order to clear the intersection.
6. Brakes applied before clearing: The driver's indecision causes him or her to apply the brakes and then to clear the intersection.

The observers were stationed 137 m (450 ft) from the stopline and inconspicuously off to the side of the highway. Traffic cones were placed at the boundaries of the dilemma zone for the median approach speed. At each approach, 6 h of "before" and 6 h of "after" data were collected, for a total of 24 h of data. These results are summarized in Table 1, which shows that the test installation caused total conflicts for 12 h to decline from 29 to 9. The two approaches combined therefore experienced a 69 percent reduction in conflicts per hour—from 2.42 to 0.75, as shown below:

Period	No.	Mean	SD
Before	12	2.42	1.62
After	12	0.75	0.75

The t-test results (3.23) indicate that so great a difference in the means could have occurred by chance only one time in 100 ($\alpha = 0.01$). Therefore, the reduction in conflicts was highly significant statistically. The test installation was very effective, despite the fact that the "before" signalization provided a measure of dilemma-zone protection.

Delay was not a factor at this intersection either before or after the improvement. Traffic was never heavy enough to extend either phase to the maximum interval.

COMPARISON WITH OTHER DESIGNS

This section compares the EC-DC design with two conventional approaches to signalization of high-speed roads. The three designs are compared for cost and the seven criteria set forth earlier.

The cost comparisons assume that both of the intersecting roadways carry high-speed traffic and that all approaches are single lane. Common costs such as signal heads and poles are omitted as irrelevant.

One of the conventional designs, referred to as the density design, uses a two-phase density controller that has locking detection memory (1, 3). Each approach has a small-area loop at the upstream end of the dilemma zone and a small-loop calling detector near the stopline. The other conventional design, referred to as the EC design, uses a two-phase basic controller set to the nonlocking detection-memory mode (1, 3). Each approach has a small-area EC detector at the upstream end of the dilemma zone and a 21-m (70-ft) loop at the stopline.

Table 2 compares the initial costs of the two conventional designs and the EC-DC configuration (details of these cost estimates available from us). The table shows that the costs of the three designs are roughly the same, although the EC configuration is 10-15 percent cheaper than the other two designs. Overall, Table 2 suggests that the choice among the three alternatives should depend on operational features and maintainability, rather than on initial cost alone.

The maintainability of the three designs is beyond the scope of this paper; however, there are maintenance-cost data that offer some assistance (see the paper by Parsonson and Tarnoff elsewhere in this Record). The amount of loop wire in the road and exposed to breakage is an important maintenance concern. The stopline loop of the EC-DC design is only 8 m (25 ft) in length, a definite maintenance advantage over the 21-m (70-ft) stopline loop used by the EC design. Operationally, the three designs can be compared by using the seven criteria set forth earlier:

1. Inasmuch as all three designs place their upstream detector in accordance with the probability-of-stopping distances given earlier, they are equal in their ability to detect a design-speed vehicle before it reaches its dilemma zone.
2. The allowable gap imposed by the density design is reduced, usually on the basis of time waiting on the red, to the setting of the minimum gap. It is easy to calculate from the table of stopping distances that, for a typical speed of 80 km/h (50 mph), the shortest setting that would pass a vehicle through its dilemma zone is about 2.5 s. This constitutes a minimum desirable allowable gap; a shorter value would give snappier operation but could leave a vehicle in the dilemma zone. The allowable gap of the EC design is typically 5 s (3). That of the EC-DC design was calculated earlier to be about 4 s. Actually the difference is greater; the ability of the EC-DC design to disconnect its stopline loops gives it a decided superiority to the EC design in real-world operation.
3. On termination of the green by gap-out, all three designs will permit a vehicle approaching at the

Table 1. Traffic conflicts before and after at test installation.

Conflict	Automobiles Before	Automobiles After	Single-Unit Trucks Before	Single-Unit Trucks After	Semitrailers Before	Semitrailers After	Total Before	Total After
Run red light	11	7	1	0	0	0	12	7
Abrupt stop	3	0	1	0	1	0	5	0
Skid	1	0	0	0	0	0	1	0
Swerve	0	0	0	0	0	0	0	0
Accelerate through yellow	3	2	0	0	0	0	3	2
Brake before clearing	7	0	1	0	0	0	8	0
Total	25	9	3	0	1	0	29	9

Table 2. Initial costs of the three designs.

Item	Cost ($) DC	EC	EC-DC
Controller	3345	2200	2200
Detector units	330	650	1000
Labor	432	502	600
Loop wire	48	87	56
Lead-in wire	226	226	226
Messenger cable	160	160	160
Conduit	80	80	100
Sealant	20	45	31
Saw, blade	25	50	50
Relays	0	0	20
Total	4666	4000	4443

Table 3. Comparative of ranking the three designs.

Item	DC	EC	EC-DC
Initial cost	3	1	2
Detection at design speed	1	1	1
Allowable gap	1	2	1
Protection at design speed	1	1	1
Protection at lower speeds	2	2	1
Avoidance of premature gap-out	2	1	1
Screening of false calls	3	2	1
Left-turn queue holding green	2	1	1
Total	15	11	9

design speed to clear the dilemma zone.

4. On termination of the green by gap-out, vehicles traveling more slowly than the design speed will typically be protected by the EC and EC-DC configurations. Slow vehicles will not be protected by the density design if the minimum gap is set low, e.g., 2.5 s. There is a trade-off here between snappy operation and protection to the slower vehicles in the stream. One can be obtained only at the expense of the other. If a density design for 89 km/h (55 mph) is also to protect the vehicle approaching at only 64 km/h (40 mph), the minimum gap must be increased to 4.5 s. This is no shorter than the allowable gap of the EC-DC design. Therefore, the density design offers no allowable-gap advantages. The EC-DC design, with its middle loop, can assure that vehicles approaching at 56 km/h (35 mph) or less will gap out upstream of their dilemma zone.

5. The EC and EC-DC configurations use stopline loops to ensure that a queue can discharge without premature gap-out. The density design relies on its variable initial interval to accomplish this. Dense traffic can defeat the purpose of this feature (3); therefore, the density design is the least attractive of the three on this account.

6. The density design has no ability to screen out false calls for the green, because the controller's detection memory is of the locking type. The other two configurations are better, because of their loop-occupancy features. The EC-DC design is best; the stopline loop operates in the DC mode when the signal is red.

7. It is desirable, especially on two-lane roads, that a queue of vehicles waiting on the green to turn left be able to hold the green. The EC and EC-DC configurations can do this, but the density design cannot.

Table 3 summarizes the comparison of the three designs. For each of the criteria, 1 is assigned to the best design, 2 to the next best, etc. If the designs are equally effective, each receives 1. The table shows that the EC-DC configuration costs somewhat more than the EC design but is superior in three operational categories. It is less expensive than the density design and is superior in four operational aspects. It appears to be a worthy challenger to these conventional designs.

ACKNOWLEDGMENT

We are grateful to the Canoga Controls Corporation for lending the three detectors used in this research.

REFERENCES

1. H. Sackman and others. Vehicle Detector Placement for High-Speed, Isolated, Traffic-Actuated Intersection Control: Volume 2—Manual of Theory and Practice. Federal Highway Administration, Rept. FHWA-RD-77-32, May 1977, 165 pp.
2. C. V. Zegeer. Effectiveness of Green-Extension Systems at High-Speed Intersections. Bureau of Highways, Kentucky Department of Transportation, Lexington, Res. Rept. 472, May 1977.
3. P. S. Parsonson. Signalization of High-Speed, Isolated Intersections. TRB, Transportation Research Record 681, 1978, pp. 34-42.
4. J. D. Clark. Discussion of paper, Signalization of High-Speed, Isolated Intersections, by P. S. Parsonson. TRB, Transportation Research Record 681, 1978, pp. 41-42.

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

Maintenance Costs of Traffic Signals

Peter S. Parsonson, Georgia Institute of Technology, Atlanta
Philip J. Tarnoff, Alan M. Voorhees and Associates, Inc., McLean, Virginia

A current major research project is aimed at developing guidelines for the selection of types of traffic-signal control at individual (noninterconnected, "isolated") intersections. Most traffic engineers prefer to install fully actuated signals at such locations; however, these are somewhat more costly to maintain than semiactuated or pretimed signals. Traffic engineers need to take into account the incremental maintenance costs of the more sophisticated types of control. Almost two years of maintenance-cost data from the California Department of Transportation (Caltrans) are reported for 121 actuated traffic signals of various designs. The costs include those for field maintenance, bench repair, travel, and parts. Also reported are the annual costs of the time required for the field maintenance of almost 1800 pretimed and actuated signals of the New York State Department of Transportation (NYSDOT). The frequency of repair of various types of controllers and detectors is reported for eight other cities and states. The Caltrans and NYSDOT data were merged to reach conclusions for the total annual cost to maintain electromechanical and solid-state controllers, including microprocessors, for various numbers of phases.

A major research project is currently under way to develop guidelines for the selection of the type of traffic-signal control to be used at individual (noninterconnected or "isolated") intersections. The choice of pretimed, semiactuated, basic fully actuated, or density fully actuated signals weighs the low price and economical maintenance of the pretimed controllers against the reduced costs, delays, etc., of the traffic-responsive models. The staff of the research project found that traffic engineers generally prefer to install fully actuated signal control at individual intersections. However, there is some concern over the cost of maintaining fully actuated controls, which is higher than that for semiactuated or pretimed controls. The professional literature sheds little light on the magnitude of these higher costs. Therefore, traffic engineers are at present unable to document the cost-effectiveness of their preference for traffic-responsive control.

An essential element in the selection of type of traffic-signal control is a knowledge of the maintenance burden of the various designs. We believe this paper is the first to provide that information.

The project staff found that traffic-signal maintenance costs are of some concern throughout the country. These costs seem to be a particular worry, however, in some northeastern and upper midwestern states. Loop detection needs to be of especially high quality in the snow belt because of severe winters, deteriorating pavements, and other reasons. There have been low-bid barriers to the purchase of high-quality "amplifiers", and difficulties in hiring and retaining technicians capable of installing and maintaining modern, sophisticated detectors and controllers. There has been a growing feeling that, if a fully actuated controller cannot be kept in fully actuated operation but must be placed on recall to one or both phases, it would then have been more economical to select semiactuated or pretimed control at the outset. It was this climate of concern that prompted the research to determine the maintenance costs of traffic signals.

RESEARCH APPROACH

Telephone contacts were made with many state and local traffic engineering agencies throughout the United States in an effort to obtain maintenance data. Most of the agencies responded that their data are in raw form—handwritten malfunction reports—that could not be summarized at reasonable cost. However, a few were found to have manual tabulations or computerized summaries of raw data in a form susceptible to tabulation at reasonable expense.

1. California Department of Transportation (Caltrans): The Caltrans Maintenance Management System (MMS) included data for 23 recent months on the total cost to maintain 121 actuated traffic signals of various designs. Total cost includes field maintenance, bench repair, travel, and parts.
2. New York State Department of Transportation (NYSDOT): NYSDOT maintains a computerized file of the work hours required for the field maintenance of the approximately 2500 pretimed and actuated traffic signals in its jurisdiction. The costs of bench repair, travel, and parts are not included. Data for two recent years were obtained.
3. Ohio Department of Transportation (DOT): Ohio DOT furnished approximate data for 1976 on the frequency of repair of their 558 actuated traffic signals, divided into three levels of sophistication.
4. Minnesota Department of Highways: The Minneapolis District furnished data covering several years for the frequency of repair of 135 actuated controllers and several hundred detectors of various types.
5. Cincinnati, Ohio: Computerized maintenance summaries for 2½ years were obtained for the frequency of repair of more than 700 controllers and their detectors, which were of various types and ages.
6. Tampa, Florida: A computerized record of frequency of repair was obtained for 1974 for almost 400 controllers of various types and ages.
7. Charlotte, North Carolina: In 1977, Charlotte purchased 72 microprocessor controllers of type 190 design. As of May 1978, 50 had been installed. The project staff obtained 14 months of maintenance data on these signals and 6 months of data on the other 388 signals in Charlotte.
8. Springfield, Illinois: Data on the frequency of repair of Springfield's 144 pretimed and actuated signals were obtained for 1976.
9. Winston-Salem, North Carolina: Four years of detailed maintenance cost data were obtained for one example each of pretimed and fully actuated traffic signals.
10. Washington, D.C.: Maintenance data for one year were obtained on a total of 497 loop detectors.

Caltrans Data

Traffic-signal maintenance-expense records for 121 selected locations are stored in the Caltrans computer. These costs include the dollar expense of all scheduled and nonscheduled field and bench maintenance of all electrical equipment at the location, including lampouts, detector malfunctions, and knockdowns. (Caltrans has a group relamping program; therefore, lampouts should be negligible in these statistics.) Table 1 shows data for the 23 months from July 1, 1976, to May 30,

Table 1. Caltrans annual maintenance costs for selected locations.

	Annual Cost per Signal						
	Two-Phase		Three- to Four-Phase		Five- to Eight Phase		All
Controller Type	N	Cost ($)	N	Cost ($)	N	Cost ($)	Cost ($)
Electromechanical							
Full volume-density	-		1[a]	1162	5[a]	1506	1449
Basic three-phase (nondensity)	-		6	753	4[a]	1209	935
Solid state							
Analog timing, transistors	-		4	657	7	1610	1263
Digital timing, noncomputer							
Brand A	1	429	4	796	9	815	782
Brand B	-		5	949	5	2600	1775
Brand C							
Two-phase	6	597	-		-		597
Two- to four-phase	2	694	11	612	-		625
Five- to eight-phase	1	354	-		10	1623	1508
Digital timing, minicomputer	-		-		33	1004	1004
Digital timing, microprocessor	-		1	421	4	757	690

[a]Apparently using minor movement controllers.

1978, adjusted to 12 months by multiplying the raw cost data by 12/23.

For two-phase digital controllers, Table 1 shows that the annual maintenance cost varies from $354 to $694, depending on the brand of manufacture. The weighted average of these data is $575.

Table 1 data for three- and four-phase, solid-state digital machines are complicated by the fact that the brand B digitally timed controllers had to be modified in design by Caltrans personnel in order to keep them operating acceptably. The model was soon discontinued by the manufacturer. If the maintenance costs for this model are therefore rejected as atypical outliers, then the average of the remaining data is $646. This is very close to the $657 for analog equipment, somewhat less than the $753 required to maintain the electromechanical controllers of basic (nondensity) design, and far less than the $1162 spent to maintain a single electromechanical volume-density controller of early vintage. Of all the three- and four-phase machines, the microprocessor design has the lowest cost—$421. The significance of the microprocessor's advantage here is clouded by the fact that only one location is included.

Controllers of five to eight phases vary widely in maintenance cost; the average is $1067 for solid-state, digital models (again omitting brand B). Once more the basic electromechanical models are somewhat higher than the digital machines, and the volume-density machines are much higher. The solid-state analog controllers were the highest of all—$1610/year—even after discarding a five-phase outlier that consumed $8065 of maintenance funds over the 23-month period. Again, the microprocessor design is significantly less expensive to maintain.

After 16 months of maintenance data had been obtained, Caltrans removed all of the electromechanical volume-density controllers, half of the 10 basic machines, and several solid-state controllers. The 16-month data for these controllers were properly annualized for inclusion in Table 1.

Table 1 suggests these general conclusions:

1. The five microprocessor controllers have significantly lower average maintenance costs than do the other types. These microprocessors are not the new type 170 but are of the special-purpose type that has nonvolatile memory.
2. Electromechanical volume-density controllers are particularly costly to maintain in comparison with their solid-state counterparts.
3. The increase in maintenance cost of three- to four-phase controllers over two-phase models is very small, probably less than $100/year. However, the jump from two-phase to five- to eight-phase controllers could easily double maintenance costs to more than $1000 unless a microprocessor controller is specified.

Caltrans now (1978) has a program under way to replace all 800 of its electromechanical traffic-actuated controllers over a three-year period. During that time, the state plans to install nearly 3000 microprocessor controllers of type 170 design. (These are purchased without factory software and the programs are loaded by the state.) As of early 1978, 50 type 170 controllers had been installed as Caltrans' standard unit for intersection or ramp-metering signal control on all safety or operational improvement projects. Since the first unit was installed in the field in September 1977, there were no maintenance data available as of April 1978. However, Caltrans expects that, since the type 170 has fewer connection points and a lower component parts count than other controllers, it will have an improved mean time between failures (MTBF). The 170's design should also result in a shorter mean time to repair (MTTR), because it is electrically organized in a more logical manner than earlier designs. It contains several self-test features intended to expedite bench repair.

Although the available California data do not include the data on the pretimed controller needed for this project, they do furnish total-cost benchmarks for other types. These benchmarks are incorporated in the comprehensive conclusions below.

NYSDOT Data

New York maintains a computerized inventory of its more than 2500 stop-and-go signals, flashers, and beacons. The work hours for the field portion of the maintenance of all of these signals are similarly catalogued. The top part of Table 2 provides a summary furnished by NYSDOT for a recent 12-month period for all of the regions except one. It includes controller maintenance only, not detector maintenance as well. The data show that, although the work hours per call are relatively independent of the type of controller, the work hours per signal increase with greater sophistication of the controller. At the request of the project staff, NYSDOT furnished a detailed computer printout of the controller and detector maintenance experience for the next 12-month period, October 1, 1976, to September 30, 1977. Selected data from the controller printout were tabulated by the project staff

Table 2. Portion of NYSDOT controller field-maintenance data for October 1, 1975, to September 30, 1976, and for October 1, 1976, to September 30, 1977.

Controller Type	No. of Signals	No. of Calls	Calls per Signal	Regular Work Hours	Overtime Work Hours	Total Work Hours	Work Hours per Call	Work Hours per Signal
October 1, 1975, to September 30, 1976								
Pretimed	168	479	2.85	704	605	1 309	2.73	7.79
Semiactuated	1243	5 500	4.42	8 201	5 442	13 643	2.48	10.98
Fully actuated	557	3 960	7.11	6 577	6 277	12 854	3.25	23.08
Flashing	506	537	1.06	950	744	1 694	3.15	3.35
Beacon	43	36	0.84	63	36	99	2.74	2.30
Total	2517	10 512	4.18	16 496	13 103	29 599	2.82	11.76
October 1, 1976, to September 30, 1977								
Electromechanical								
Pretimed	84	154	1.83	412	234	646	4.19	7.69
Semiactuated	583	1 520	2.61	3 918	1 345	5 263	3.46	9.03
Fully actuated	251	994	3.96	2 692	1 142	3 834	3.86	15.27
Volume-density	24	115	4.79	246	139	385	3.35	16.04
Mixed electromechanical and solid state								
Semiactuated	473	1 037	2.19	3 666	1 501	5 167	4.98	10.92
Fully actuated	194	626	3.23	1 772	826	2 598	4.15	13.39
Solid state								
Analog timing								
Semiactuated	28	27	0.96	178	36	214	7.93	7.64
Fully actuated								
Two- to four-phase	72	260	3.61	713	376	1 089	4.19	15.12
Five- to eight-phase	22	162	7.36	305	274	579	3.57	26.32
Digital timing, fully actuated								
Two- to four phase	37	54	1.46	155	154	309	5.72	8.35
Five- to eight-phase	14	58	4.14	145	155	300	5.17	21.43
Total	1782	5 007	2.81	14 202	6 182	20 384	4.07	11.44

Table 3. Cost of work hours for field maintenance of selected NYSDOT controllers for October 1, 1976, to September 30, 1977.

Controller Type	Cost per Signal ($)			
	Two-Phase	Three- to Four-Phase	Five- to Eight-Phase	All
Electromechanical				
Pretimed	-	-	-	82
Semiactuated	92	-	-	92
Fully actuated	134	197	-	158
Volume-density	208	109	-	170
Mixed electromechanical and solid state				
Semiactuated	113	-	-	113
Fully actuated	113	155	-	140
Solid state				
Analog timing				
Semiactuated	75	-	-	75
Fully actuated	166	154	293	191
Digital timing, fully actuated	70	144	243	135

and are shown in the bottom part of Table 2 (in order to expedite the manual tabulation of the computer output, only those models installed at four or more locations in the state were included). The data in the bottom part are much more detailed in their breakdown by controller type than are the data above for the previous year.

Table 3 reduces the data from the bottom part of Table 2 to a dollar cost of the work hours for field maintenance for each type of controller by using NYSDOT-supplied labor costs of $9.00/h for regular time and $13.50/h for overtime (including an 80 percent overhead factor).

Tables 2 and 3 are incomplete because of their omission of detector maintenance data. Such data are tabulated by the NYSDOT by manufacturer rather than by type of traffic signal. The tabulation for 1976-1977 showed that 6190 detector-related service calls required 17 713 regular h and 7483 overtime h. The project staff distributed these calls and work hours among the various types of actuated controllers as judiciously as it could, according to the number of actuated phases of each controller type. These data were then merged with the controller-only data in the lower part of Table 2 and Table 3. The results are shown in Tables 4 and 5 as estimated data. Like Table 3, Table 5 uses the NYSDOT-supplied wage rates.

Table 5 is the most important because it presents controller-plus-detector costs, as does Table 1 for Caltrans. However, there are important differences between the two tables: Table 5 includes only the cost of field work hours, while Table 1 also includes the cost of repair vehicles, parts, and bench labor. By using the pretimed controls as a baseline at a field-maintenance cost of $82/year, Table 5 suggests these general conclusions:

1. A step up to two-phase semiactuated control will add approximately $110/year to the cost of maintenance, regardless of whether the controller is of electromechanical or solid-state design.
2. A further step up, from two-phase semiactuated control to any two-phase fully actuated controller that is not digital, will cost $143/year. A digital machine will reduce that increase in cost to only $76/year.
3. A two-phase electromechanical volume-density controller costs about $65 more per year to maintain than any basic controller of nondigital design and about $130 more than a digital model.
4. Basic actuated controllers of three and four phases cost an average of $462/year to maintain (if the "mixed electromechanical and solid-state" data are discarded as outliers). This is $380 more than pretimed control, $127 more than two-phase fully actuated nondigital control, and $194 greater than digital control.

Table 4. Estimated maintenance data for selected NYSDOT controllers and detectors for October 1, 1976, to September 30, 1977.

Controller Type	No. of Signals	No. of Calls	Calls per Signal	Regular Work Hours	Overtime Work Hours	Total Work Hours	Work Hours per Call	Work Hours per Signal
Electromechanical								
Pretimed	84	154	1.83	412	234	646	4.19	7.69
Semiactuated	583	2 874	4.93	7 797	2 984	10 781	3.75	18.49
Fully actuated	251	2 385	9.51	6 678	2 826	9 504	3.98	37.86
Volume-density	24	245	10.21	618	296	914	3.73	38.08
Mixed electromechanical and solid state								
Semiactuated	473	2 132	4.51	6 801	2 825	9 626	4.52	20.35
Fully actuated	194	1 820	9.39	5 191	2 270	7 461	4.10	38.46
Solid state								
Analog timing								
Semiactuated	28	95	3.39	373	118	491	5.17	17.54
Fully actuated								
Two- to four-phase	72	687	9.54	1 936	893	2 829	4.12	39.29
Five- to eight-phase	22	366	16.64	890	521	1 411	3.86	64.14
Digital timing, fully actuated								
Two- to four-phase	37	259	7.00	740	401	1 141	4.41	30.84
Five- to eight-phase	14	188	13.43	517	312	829	4.41	59.21
Total	1782	11 190	6.28	31 953	13 680	45 633	4.08	25.61

Table 5. Cost of work hours for estimated maintenance of selected NYSDOT controllers and detectors for October 1, 1976, to September 30, 1977.

Controller Type	Cost per Signal ($) Two-Phase	Three- to Four-Phase	Five- to Eight-Phase	All
Electromechanical				
Pretimed	-	-	-	82
Semiactuated	189	-	-	189
Fully actuated	330	489	-	391
Volume-density	398	398	-	398
Mixed electromechanical and solid state				
Semiactuated	210	-	-	210
Fully actuated	309	619	-	399
Solid state				
Analog timing				
Semiactuated	177	-	-	177
Fully actuated	365	449	684	474
Digital timing, fully actuated	268	448	633	411

Table 6. Frequency of controller repair by the Minneapolis District of the Minnesota Department of Highways.

Controller Type	Age (years)	Years of Data	No. of Signals	Annual Calls per Signal
Electromechanical, fully actuated				
Two-phase	0-5	5.0	8	2.40
Three- to five-phase	0-5	3.0	4	4.70
Total	0-5	-	12	2.92
Solid state				
Analog timing				
Semiactuated	0-5	5.0	2	1.10
Fully actuated				
Three-phase	0-5	4.9	11	1.84
Five-phase	0-5	5.0	27	3.19
	6-10	2.5	23	2.40
Digital timing, fully actuated				
Three-phase	0-5	3.45	13	1.34
Five- to eight-phase	0-5	3.0	24	2.05
Total	0-5	-	100	2.37
	6-10	2.5	23	2.4

5. Solid-state controllers of five to eight phases cost an average of $659/year to maintain. This is $210/year more than a three- or four-phase analog or digital machine.

Ohio DOT

Ohio furnished maintenance data for its 558 traffic signals as shown below.

Controller Type	No. of Signals	No. of Calls	Annual Calls per Signal
Electromechanical			
Actuated (basic)	296	412	1.39
Volume-density	84	169	2.01
Solid state			
Analog timing	178	237	1.33
Digital timing	Few	Unknown	—

The table shows primarily that electromechanical volume-density controllers require significantly greater maintenance than do their basic counterparts, and much more than modern solid-state controllers.

Minnesota Department of Highways

Table 6 shows the frequency of repair of 135 controllers in the Minneapolis District of the Minnesota Department of Highways. The table indicates a distinct advantage of solid-state over electromechanical design. As expected, the greater the number of phases, the more frequent the repair. The table indicates that the frequency of repair of solid-state controllers does not increase with the age of the unit.

The available data included the frequency of repair of the 811 loop detectors and 12 magnetic detectors used with the 112 controllers in Table 6. It was found that the loop detectors averaged 0.24 failures per detector per year, and the magnetic models averaged 0.26/year.

Cincinnati, Ohio

Cincinnati has used a computerized traffic control equipment maintenance summary for five years. These summaries have been used to reduce the number of chronically malfunctioning intersection controls from 17 in

1973 to only 2 today. The city has more than 700 traffic signals.

Table 7 provides a summary of 2½ years of computerized record keeping—March 1975 to August 1977. The project staff removed all "normal cycle" reports (indicating no malfunction found by the repair crew). The staff also removed all failure reports associated with system features, such as coordination units, since the emphasis in this project is on individual intersections.

Table 7. Frequency of controller repair in Cincinnati from March 1975 to August 1977.

Controller Type	Age (years)	No. of Signals	Annual Calls per Signal
Electromechanical			
Pretimed	0-5	31	1.81
	6-10	35	2.17
	11-15	-	-
	16-20	139	1.99
	>20	72	1.74
	All	277	1.92
Semiactuated	0-5	2	1.50
	6-10	48	2.58[a]
	11-15	80	1.70
	16-20	54	1.65
	>20	38	1.68
	All	222	1.87
Fully actuated, operated semiactuated	0-5	16	1.13
	6-10	45	2.76[b]
	All	61	2.33
Solid state			
Semiactuated	0-5	2	1.50
Fully actuated, operated semiactuated	0-5	12	1.00
	6-10	7	1.14
	All	19	1.05

[a] High because of a single model.
[b] High because 20 units of an early-design phase-modular controller experienced 3.85 calls/signal each year.

Table 7 does not indicate any significant increase in maintenance load with an increase in sophistication from pretimed to semiactuated to fully actuated controls. Rather, the evidence is that the solid-state actuated equipment is more reliable than the pretimed.

Table 7 shows that the frequency of repair of electromechanical equipment increases with age to approximately the 10th year and then decreases with greater age. This same phenomenon is evident also in the data presented below for Tampa, Florida.

Detector maintenance over two years in Cincinnati is shown below:

Detector Type	No. of Detectors	Annual Failures per Detector
Pressure	23	0.17
Magnetic	81	0.26
Loop	151	0.29
Sonic	37	0.32

The data indicate that the pressure detector is significantly more reliable than the other types listed. The data for magnetic and loop detectors are strikingly similar to the Minneapolis data reported above.

Tampa, Florida

A computerized record of frequency of repair was obtained for almost 400 controllers for 1974. The record is summarized in Table 8; detector maintenance data are not included. Except for the pretimed controllers and the most recently purchased solid-state controllers operated as semiactuated, the maintenance load is extremely heavy compared with that reported above for Ohio, Minnesota, and Cincinnati. The higher rate for Tampa may be due to the severe lightning storms experienced frequently in Florida. Many of the Tampa rates are of the same order of magnitude as those obtained for New York State (Table 4).

Table 8. Frequency of controller repair in Tampa for 1974.

Controller Type	Age (years)	No. of Signals	Annual Calls per Signal
Electromechanical			
Pretimed	0-5	32	1.69
	6-10	93	2.33
	11-15	5	0.40
	16-20	1	1.00
	>20	45	2.76
	All	176	2.26
Semiactuated	0-5	1	0
	6-10	11	6.82
	11-15	38	8.71
	16-20	11	4.45
	>20	12	9.08
	All	73	7.73
Fully actuated	0-5	1	1.00
	6-10	3	4.33
	11-15	5	11.80
	16-20	4	7.50
	>20	1	5.0
	All	14	7.71
Semiactuated, operated fixed	0-5	1	2.00
	11-15	1	11.00
	All	2	6.5
Fully actuated, operated semiactuated	11-15	1	3.0
	16-20	2	5.0
	All	3	4.33
All			4.09
Solid state			
Semiactuated	6-10	29	5.48
	11-15	4	6.00
	All	33	5.55
Fully actuated	0-5	35	6.17
	6-10	33	9.09
	11-15	2	30.0
	All	70	8.23
Fully actuated, operated semiactuated	0-5	8	1.75
	6-10	8	5.00
	11-15	2	23.0
	All	18	5.56
All			7.04

Charlotte, North Carolina

Charlotte has a variety of actuated equipment of both electromechanical and solid-state design and has for many years provided adequate funding for traffic engineering operations. Therefore, maintenance data were readily available and, in addition, there was experience with a significant number of microprocessor controllers.

Their total of 438 controllers includes 72 type 190 microprocessors received in 1977. Unlike the five microprocessors reported on by Caltrans (Table 1), the Charlotte models are of the type that include volatile memory with battery backup in the event of power failure. The program for the type 190, unlike that for the type 170, is provided by the factory.

The installation of the Charlotte microprocessors began in March 1977. At the time of the visit by project personnel in October 1977, 24 microprocessors had been installed for an average of only about three months. The latest data available to us were as of December 1978; 69 microprocessors had been installed.

Table 9 summarizes six months of 1977 data for 438 Charlotte signals, except that the microprocessor controller data were updated to December 1978. The 69 microprocessor controllers were installed gradually between March 1977 and December 1978. Data on two installed less than a month before the December update

Table 9. Frequency of traffic-signal repair in Charlotte for April to September 1977.

Controller Type	Age (years)	No. of Signals	Annual Calls per Signal
Electromechanical			
Semiactuated	>20	75	1.95
Semiactuated (PR)	16-20	37	2.49
Fully actuated	0-5	3	0.67
	11-15	63	1.43
	>20	13	0.91
	All	79	1.29
All			1.83
Solid state			
Pretimed[a]	6-10	167	0.54
Semiactuated	0-5	1	2.00
	6-10	3	0.00
	All	4	0.50
Semiactuated (TPR)	6-10	11	1.82
	11-15	6	0.00
Fully actuated			
Digital, noncomputer	0-5	4	0.00
Microprocessor[b]	0-5	67	0.99
Analog, noncomputer	11-15	7	0.57

[a]These have four-phase frames but are operated in two phases, with only two load switches and without detectors or actuation module; a central digital computer operates them as pretimed controllers.
[b]March 1977 to December 1978.

were discarded, leaving data on 67 that had service records of 1-21 months. The project staff calculated the frequency of calls per year for each microprocessor individually, by using the number of months that each had been in place. This procedure was more precise than one that assumed that all 67 controllers had been in service for an average of 10.5 months. The table shows that Charlotte's electromechanical signals require service once or twice a year and that the new microprocessor controllers require close to one service call annually.

Springfield, Illinois

Springfield furnished maintenance data for a 12-month period in 1976-1977. These data for 144 signals are summarized in the table below, which shows an unusually high failure rate for semiactuated controllers.

Controller Type	No. of Signals	Annual Calls per Signal
Pretimed	117	2.37
Semiactuated	21	3.95
Fully actuated	6	1.67

The city traffic engineer explained that the city had had excellent operational results with two-phase semiactuated controllers for many years. Their maintenance problems began in 1975, when multiphase fully actuated controllers were purchased and operated semiactuated in an arterial system.

Winston-Salem, North Carolina

Four years of detailed maintenance cost data were obtained for one pretimed and one fully actuated controller at locations selected by the city as fairly typical; six of the eight calls for the pretimed controller were for preventive maintenance. The data, summarized below, show a very low cost to maintain the controllers (city costs of $5/h for labor, truck, and supplies were increased by 80 percent to account for fringe benefits and overhead). However, the record of the loop-detector maintenance shows 33 trips in four years to retune, replace, and cut new loops.

Controller Type	Calls per Year	Maintenance Cost per Year ($)
Pretimed	2.16	17
Fully actuated, three-phase, solid-state, digital	2.05	18
Loop detection for above controller	8.46	340

Washington, D.C.

A total of 497 loop detectors were installed as a part of the Urban Traffic Control System Research Project sponsored by the Federal Highway Administration. The installations were made only after a thorough study by the contractor of the available (crystal) electronics units and the procedures and materials for installing the loop wire and lead in. In the first year, there were 33 failures of the electronics units, for an annual rate of 0.07 failures/detector. During that period, 26 loops failed because of utility excavations; if these failures are added, the total annual rate becomes 0.13 failures/detector (1).

CONCLUSIONS

The foregoing findings provide the basis for conclusions about the total cost to maintain various types of controllers. Table 10 reflects these findings.

It was found that the Caltrans Maintenance Management System offers the only available data base of total maintenance costs, including both field and bench work, parts, and travel. It seemed appropriate, therefore, to plot the first points from those data. The table indicates the values determined directly from the Caltrans data in Table 1. The NYSDOT data (Table 5) were given second preference, because work hours were available only from that source.

For electromechanical equipment, the coordination point selected between the Caltrans and NYSDOT data sets was for fully actuated three- or four-phase controllers. The ratio of California total cost to NYSDOT field cost for that cell is 753 ÷ 489 = 1.54. The values in the other cells of Table 5 were multiplied by 1.54 to obtain the values shown in Table 10. For solid-state equipment with analog timing, the coordination point between the Caltrans and NYSDOT data was again taken for the fully actuated three- or four-phase controllers. The ratio of the two cells is 657 ÷ 450 = 1.46, which is reassuringly close to the 1.54 calculated for electromechanical equipment. This factor was used to obtain the remaining values for solid-state analog controllers.

The factors of 1.46 and 1.54 indicate essentially that, for every dollar spent on work hours for field maintenance of actuated equipment, an additional 50 cents is required for the other items that constitute the total cost as defined by Caltrans. These items include the truck and its fuel, the parts used in the repair work, and the cost of the bench labor. If these items cost about the same for pretimed equipment as they do for actuated models, then it would be appropriate to derive the total cost of pretimed controller maintenance as 1.50 × $82 (from Table 5) or $123. However, bench-repair labor is certainly less for pretimed equipment than for actuated designs. Therefore, the project staff arbitrarily assigned a reduced cost of $115 for that entry in Table 10.

Table 10. Derived conclusions on the total annual cost to maintain various types of traffic signals.

Controller Type	Cost per Signal ($) Two-Phase	Three- to Four-Phase	Five- to Eight-Phase	All
Electromechanical				
Pretimed	-	-	-	115
Semiactuated	291[a]	-	-	-
Fully actuated	508[a]	753[a]	1209[a]	-
Volume-density	613[a]	1162[a]	1506[a]	-
Solid-state				
Analog timing				
Semiactuated	258	-	-	-
Fully actuated	532	657[a]	1610[a]	-
Digital timing (except microprocessor), fully actuated	575[a]	661[a]	1090[a]	-
Microprocessor, fully actuated[b]	-	421[a]	757[a]	-

[a]Taken directly from the Caltrans Maintenance Management System data in Table 1.
[b]These data are for a few controllers from a single manufacturer. Other microprocessor controllers may have different maintenance requirements (see Charlotte data in Table 9).

ADEQUACY OF DATA

These data were gathered to assist in the future selection of type of control—pretimed, semiactuated, basic fully actuated, and density fully actuated. Data on microprocessor controllers must be included. In this context there are two fundamental inadequacies in the available data.

One is that none of the data sets provides the total maintenance costs for each of the four types of control. The Caltrans data quote total cost—field and bench labor, travel, and materials—but do not include pretimed control or the new type 170 microprocessor. The NYSDOT data include pretimed equipment, but only the cost of the field work hours can be obtained; bench labor, travel, and parts are not covered. Most of the other sources quote only frequency of repair, not dollar cost.

Another difficulty with these data is that future consideration of actuated control—at least for the future as we can see it now—will focus on the microprocessor controller and the digital loop detector. Almost all of the available maintenance data predate these recent innovations.

Respondents in California, New York, and Charlotte, for example, make it clear that microprocessor designs of type 170 (user programmed) and type 190 (factory programmed) are showing a longer MTBF and a shorter MTTR than have the other controllers reported herein. (However, hard data on this superiority are skimpy so far.) Presumably other designs of microprocessors will show similar benefits when their records are tabulated.

Moreover, the digital loop detector is proving to be significantly more effective than its analog counterpart. New York State, for example, found in 1978 that some digital loops are successfully operating even though they are in such poor condition that the locations had been scheduled for reinstallation of new loops. A number of other respondents indicated that they are extremely impressed with the digital unit's sensitivity and ability to operate under adverse conditions of loop condition, temperature, etc.

It is notable that this research project was conceived at a time when some states—particularly those in the Northeast and upper Midwest—were experiencing great difficulty in maintaining actuated controllers and loop detectors of conventional design. In 1977 and 1978, the microprocessor and the digital loop detector began to change this situation completely for some agencies. New York State, for example, now is able to consider selecting fully actuated control at individual intersections. It seems clear that this research project has been overtaken by technological breakthroughs that greatly diminish the potential attraction of pretimed or semiactuated control at individual intersections.

ACKNOWLEDGMENT

This paper reports some of the preliminary findings of National Cooperative Highway Research Project 3-27, Guidelines for Selecting Traffic-Signal Control at Individual Intersections.

We acknowledge with thanks the cooperation and contributions of Gerald R. Bloodgood and Kenneth McDaniel, Division of Operations, California Department of Transportation; David A. Green, Safety Standards and Systems Bureau, and David J. Russo, Traffic Control and Engineering Bureau, New York State Department of Transportation; Geno P. D'Ippolito, Bureau of Traffic, Ohio Department of Transportation; Dennis R. Eyler, Minnesota Department of Highways; Benjamin W. McKay, Division of Traffic Engineering, Cincinnati, William Holsonback, Division of Engineering, Tampa; John Clark, Maintenance Services Section, and Bruce E. Friedman, Traffic Signal Section, Charlotte Traffic Engineering; Ronald L. Habegger, city traffic engineer, Springfield, Illinois; and Donald E. Holloman, Traffic Systems Division, Traffic and Transportation Department, Winston-Salem.

REFERENCE

1. The Urban Traffic Control System in Washington, D.C. Office of Research and Development, Federal Highway Administration, Sept. 1974, p. 27.

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

Reflectorization of Railroad Rolling Stock

Richard G. McGinnis, Department of Civil Engineering, Bucknell University, Lewisburg, Pennsylvania

This paper examines the effectiveness of retroreflectors on the sides of railroad rolling stock as a means of reducing highway-railroad grade-crossing accidents, and it estimates the benefits and costs of reflectorizing the U.S. fleet of railroad cars. Factors that affect the amount of reflected light received by a driver (including reflector characteristics, vehicle-reflector positioning, reflector cleanliness, headlight cleanliness and beam usage, windshield transmittance, and atmospheric conditions) were analyzed, and expected reflector illuminance levels were predicted. Under conditions expected in railroad operation, the analysis indicates that 15-cm (6-in) square delineators of high-intensity-grade reflective sheeting will permit detection distances sufficient for safe stopping in most highway situations, even under low-beam headlight illumination. Benefits were estimated from the 1975 Federal Railroad Administration accident data. Accidents were categorized into four groups based on the speeds of the train and motor vehicle and the collision point on the train. Reflector effectiveness for each of these groups was estimated by considering the type of crossing warning device, daylight accident rates, weather conditions, presence of obstructions, human factors associated with nighttime driving, and the train and motor vehicle speeds. The costs of a reflectorization program were estimated and a cost-effectiveness analysis was performed to assess the impact of visibility at grade crossings on annual benefits, since no reliable information is available on this important factor.

Conflicts between trains and automobiles at highway-railroad grade crossings have long been recognized as a major safety problem. Since the 1920s, the railroads and various local, state, and federal government agencies have worked to reduce hazards at the 220 000 public grade crossings in the United States.

Statistics indicate that efforts to improve the safety at grade crossings have been effective: in 1928 there were 2568 fatalities resulting from grade-crossing accidents (1); in 1977, this figure was 63 percent lower, even though vehicle kilometers of travel increased more than 800 percent during the same period (2).

Unfortunately, the problem of grade-crossing accidents has still not been completely solved. There were more than 12 000 grade-crossing accidents reported to the Federal Railroad Administration (FRA) in 1977. Consequently, FRA and the Federal Highway Administration are continuing their programs to reduce the hazards of railroad-highway grade crossings.

Most grade-crossing safety programs have been aimed at improving the warning devices at the grade crossing, but another approach is to improve the conspicuity of the train, so that motorists can actually detect it near a crossing. At some crossings, for example, street lights have been installed to improve nighttime visibility. On-train devices have also been proposed. Recently, interest has been high in the use of strobe lights on locomotives to improve both day and nighttime train visibility. Also, the use of reflectors on the sides of railroad cars has long been advanced by some as an effective way of increasing nighttime train visibility.

The purpose of this study was to examine the effectiveness of reflectors on the sides of railroad cars as a means of reducing grade-crossing accidents. The use of reflectors on railroad cars has been discussed in many documented studies (1, 3-8), but the conclusions reached in these investigations are not consistent and indicate that the effectiveness of reflectors in reducing grade-crossing accidents may be either very considerable or absolutely minimal. This paper provides both an in-depth analysis of reflector effectiveness and an examination of the benefits and costs of reflectorizing the sides of the U.S. railroad car fleet.

REFLECTORIZATION

Reflectorization has its greatest safety potential for accidents that occur at night and involve a motor vehicle striking the side of a train. In many of these accidents, the motorist is apparently unable to see the train in time to stop the vehicle safely. Reflectors on the side of a railcar will reflect light from a motor vehicle's headlights back toward the vehicle and, to the driver, such reflectors will appear as light sources or "bright spots" against a dark background.

The approach taken to analyze the effectiveness of reflectorization was to examine first the factors that affect the amount of reflected light that can be expected at various distances from a grade crossing and then to compare these light levels with visual detection standards to see whether detection (and perception) of a train's presence is likely.

The type of reflector that would be used on railroad cars is called a retroreflector or reflexreflector. Retroreflectors reflect incident light back toward the light source in a narrow beam. Retroreflective materials are used extensively for highway signs, pavement markings, and motor vehicle markings.

The amount of light received by an observer from a retroreflector is affected by six factors: the reflective intensity of the reflector, its size, the intensity of the original light source, atmospheric transmissivity, windshield transmittance, and its distance from the observer. The relationship between these factors and illuminance received by the observer is given by Equation 1:

$$E_e = (I_s\ A\ R\ t^{2d}\ W)/d^4 \qquad (1)$$

where

E_e = illuminance received by the observer (lx),
I_s = intensity of the light beam toward the reflector (cd),
A = area of the reflector (m^2),
R = reflective intensity of the reflector [$(cd/lx)/m^2$],
t = transmissivity of the atmosphere per meter,
W = windshield transmittance, and
d = distance between the observer and the reflector (m).

A FORTRAN computer program was written to compute reflector illuminance received by a driver for various reflector (train) locations. The program used headlamp luminous-intensity distributions and retroreflector properties to determine expected reflector brightness. Values were computed for reflector locations from 30 m (100 ft) to 244 m (800 ft) in front of the motor vehicle and from 122 m (400 ft) to the left to 122 m to the right of the projected path of the motor vehicle. In addition, the program allowed for variation in the size, efficiency, and placement height of reflectors; the conditions of headlights, windshields, and atmosphere; and the intersection angle between the train and the motor vehicle.

Reflective Intensity

The reflective intensity of a reflector depends on the grade of reflective material and on the incidence and divergence angles. The incidence angle is the angle from the light source to a line normal to the reflective surface, and the divergence angle is the angle between the line of sight of the observer and the path of light from the source (Figure 1).

The divergence angle is a function of the distance between the driver's eyes and the light source and the distance between the reflector and the light source. Because the distance between the light source and the driver's eyes is a constant, the divergence angle decreases as the distance between the vehicle and the reflector increases (Figure 1). In the analysis, dimensions for a typical U.S. passenger vehicle were used (9) and produced divergence angles of 2° to 0.14°.

The overall efficiency of a retroreflector is maximized when the divergence and incidence angles are both zero. Furthermore, since both the divergence and incidence angles vary inversely with reflector-vehicle separation, reflector efficiency will increase with separation between train and motor vehicle.

Retroreflective sheeting material is currently available in two grades: engineering grade and high-intensity grade. Analyses in this study are based on the reflective qualities of high-intensity-grade reflective sheeting, since the threefold to fourfold increase in reflectivity that high-intensity grade provides over engineering grade is needed to produce illumination levels that are sufficiently bright at long distances for grade-crossing safety. The low range of divergence angles expected also contributed to the selection of high-intensity grade.

Reflector efficiency is defined as the proportion of the original reflectivity that a reflector maintains under given operating conditions. Reflector efficiency decreases with time because of deterioration of the reflective material and accumulation of dirt and grime. The average efficiency of the reflectors on a fleet of cars would depend on the frequency of reflector replacement, the level of reflector maintenance, the operating environments of the railcars, and the durability and dirt-resistant qualities of the reflective material.

Limited data are available on the decreased reflector efficiency that can be expected from continuous use of retroreflectors on railroad rolling stock. However, a leading manufacturer of reflective materials advertises that silver high-intensity reflective sheeting used on vertical surfaces for highway signs will have a reflective intensity of 200 (80 percent of original specified reflectivity) after 10 years of service and proper cleaning of the material, while the effective performance life is decreased to seven years in areas of abundant sunshine. In general, experience with high-intensity sheeting in highway use indicates an effective performance life of 12-14 years (10). Indications are, however, that the railroad environment is more severe than that experienced by highway signing and that a shorter life may consequently be expected for reflective sheeting used on railroad rolling stock. Nevertheless, the reflector efficiency question cannot be definitively answered before field tests of reflectors on railcars have been performed.

In the absence of reliable data, a reflector efficiency of 0.50 has been used in this study. Since the reflective intensities have been computed conservatively, the actual reflectivities used in the analysis represent approximately 30-40 percent of the reflective intensities of new silver high-intensity sheeting.

Reflector Size

In the analysis, a reflector size of 0.023 m^2 (0.25 ft^2) was used since it is the largest size that can still be viewed as a "point source" under most conditions expected at grade crossings.

Motor Vehicle Headlight Systems

The amount of light beamed on a reflector (and ultimately back to the driver) is a function of the location of the reflector in relation to the headlights, the type of headlight system, the mode of headlight operation (high beam or low beam), and the maintenance level of the headlights (alignment and cleanliness).

In most operational situations, the retroreflectors on the railcars will be located above the horizontal axis of the motor vehicle's headlight system. Under high-beam operation, a substantial amount of light is beamed upward; however, very little light is directed upward in the low-beam operational mode. But the amount of light incident on the reflector surface is enhanced at long distances due to the decreasing vertical angle between the reflectors and the headlight axis. For example, for a vertical separation of 0.3 m (1.0 ft) (see Figure 2) low-beam headlight intensity is 1500 cd at 30 m (100 ft) and 4500 cd at 244 m (800 ft).

The use of high-beam lights was studied by the Southwest Research Institute (11), which found that less than 25 percent of the 23 176 vehicles observed in an open road situation (high beams appropriate) actually used

Figure 1. Divergence and incidence angles of retroreflectors.

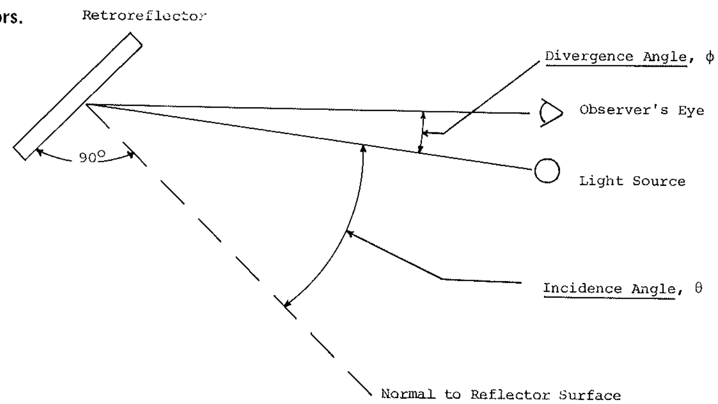
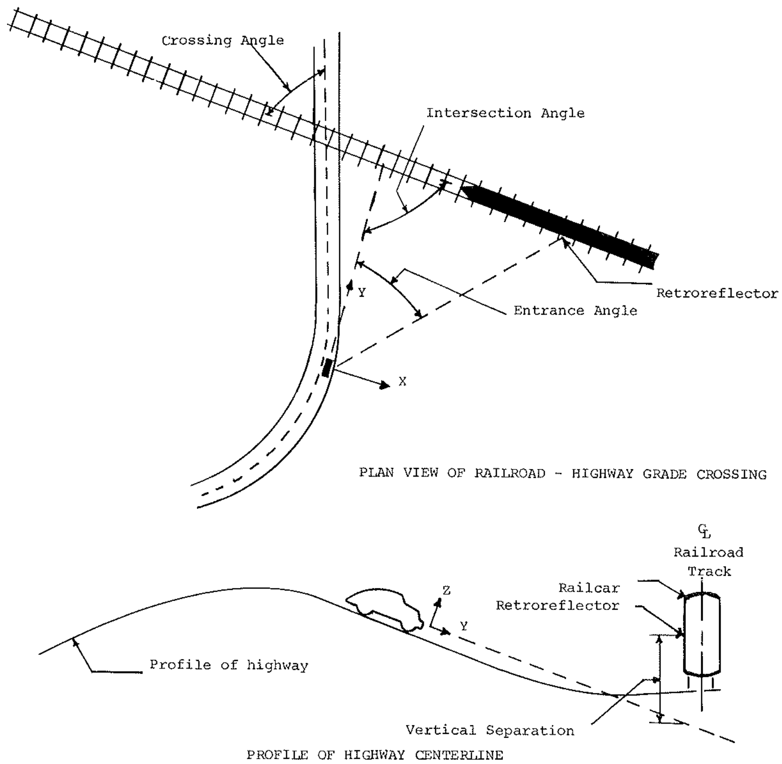


Figure 2. Visibility of retroreflectors at a typical railroad-highway grade crossing.



their high beams. Therefore, if reflectors are to be highly effective, they must be visible under low-beam illumination.

A headlight efficiency of 0.85 was used in all analyses. This figure is consistent with research findings (12, 13) for operation during dry-roadway conditions. During wet-road conditions, light reductions of 50 percent are not uncommon. However, recent research on the visibility of reflectorized overhead highway signs (14) indicates that sign illumination increases by a factor of more than two under wet-road conditions because of the increased amount of light reflected up from the wet pavement. Thus, the 0.85 headlight efficiency used in the analyses should be applicable to most driving conditions. Effects of improper aim of headlights were not included in the analyses because of inadequate data.

Atmospheric Conditions

Atmospheric conditions affect the efficiency of any reflector. Fog and haze, for example, reduce all visibility, including that of light bounced off a retroreflector. In the analyses, a "light haze" condition [8-km (5-mile) daytime visibility] was used.

Windshield Conditions

A windshield transmits only a portion of the total light incident on it. For untinted windshields, the transmittance is about 87.5 percent, but only about 72.5 percent of the light is transmitted through tinted windshields (15). Tinted windshields are known to decrease visibility distances at night; however, these decreases are usually less than 10 percent (16, 17). In the analysis, a windshield transmittance of 70 percent was used.

Detection Level

Detection of reflected light depends primarily on its brightness and the contrast with its surroundings (3). The threshold illumination level for a point source viewed against a background luminance of 0.0034 cd/m^2 (0.001 foot lambert) (overcast, moon) is 24.7×10^{-9} lx (2.3×10^{-9} footcandles) (18). This value represents the illumination level required for 98 percent probability of detection when the observer knows precisely where to look for the light, and it must be increased 5 to 10 times if the light is to be easily found. The Federal Aviation Administration's (FAA) detection level for pilots is 7.8 times this minimum threshold. If the light signal is to attract the attention of an observer who is not actively looking for it, then increases of 100 to 1000 times the threshold level are needed (19).

For the study, a three-region criterion was used to assess the detectability of various reflector illumination levels. It was assumed that the FAA detection level for pilots is the practical minimum illumination that can be expected to be detected by highway users in the vicinity of railroad-highway grade crossings. A driver familiar with the sight of railcar reflectors, approaching a grade crossing that he or she knows has high train volumes, should be able to detect a reflected light source at this level. Most drivers, however, would require an illumination level significantly higher than the FAA threshold for detection.

An illumination level of 1000 times the minimum threshold [24.7×10^{-6} lx (2.3×10^{-6} footcandles)] should be sufficient to make the reflector detectable to all but the few drivers who are completely oblivious to their driving environment. In the region between 100 and 1000 times the minimum threshold (24.7×10^{-7} lx to 24.7×10^{-6} lx), the reflector "probably" would be detected. Between the FAA threshold and the 100-times level, the reflector could "possibly" be detected.

RESULTS

Figure 3 shows the three ranges of reflector visibility

Figure 3. Visibility regions for an intersection angle of 90° for low beams and for high beams.

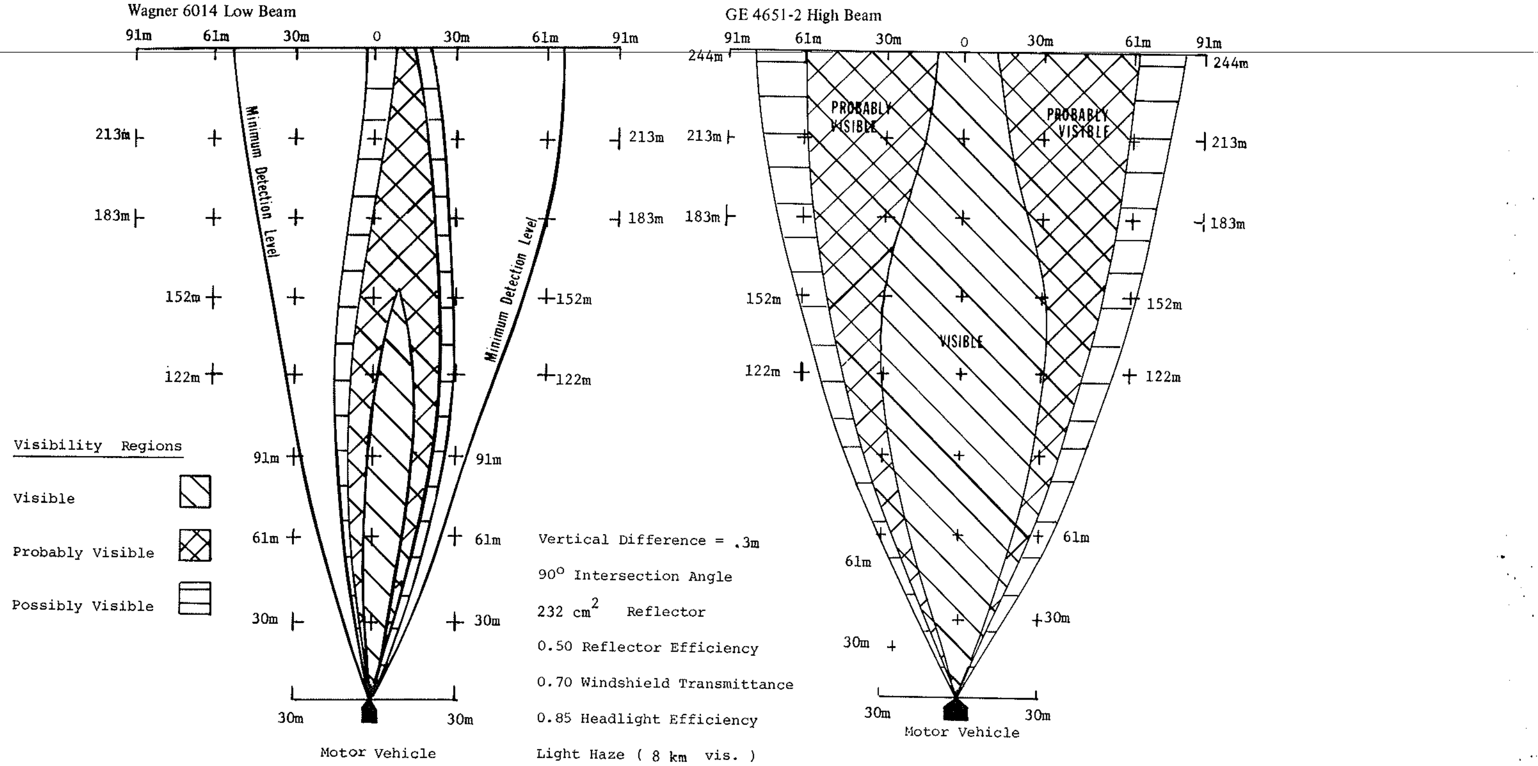
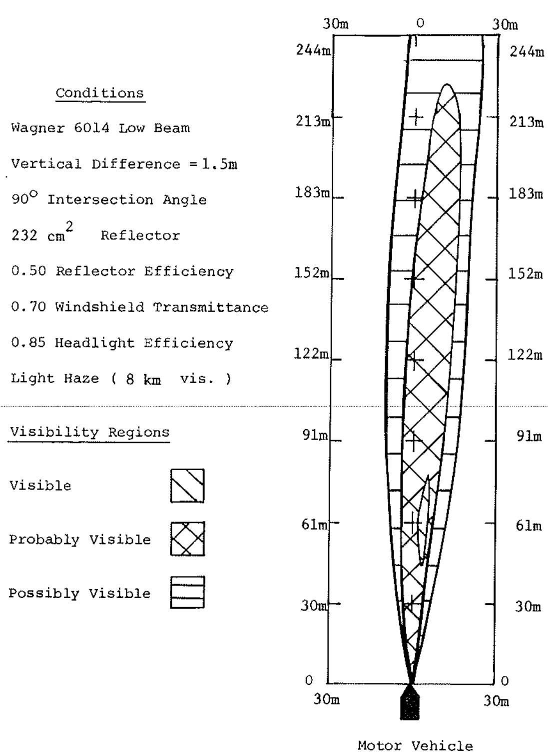


Figure 4. Visibility regions for a 1.5-m vertical difference with low beams.



expected from a two-lamp low-beam system and a four-lamp high-beam system. As expected, the "visible" region is much larger for high-beam illumination than it is for low-beam illumination. Even so, the "visible" region for the low beams extends 152 m (500 ft) from the vehicle and the "probably visible" region beyond 244 m (800 ft). For the crossings represented by these figures, there is little question that reflectors would be visible with high-beam illumination and would most likely also be visible with low-beam illumination. Separate analyses performed indicate that changing the intersection angle to 45° does little to affect the visibility of the reflector.

The effectiveness of the reflectors is greatly influenced by the position of the reflector on the railroad car. Figure 4 shows the visibility regions for a reflector located 1.5 m (5 ft) above the plane of headlights, and a comparison of Figure 4 with Figure 3 for low beams shows that the impact of raising the reflector 1.2 m (4.0 ft) is a reduction of the "visible" region to practically zero, although the "probably visible" region still extends beyond 229 m (750 ft). The impact of a high reflector on visibility is much less when illumination is by high beams.

Reflector Effectiveness

The analytical evaluation of reflector effectiveness indicates that retroreflectors on the sides of railroad cars should be detectable at distances between 152 m (500 ft) and 305 m (1000 ft) if illuminated by low-beam lights and between 274 m (900 ft) and 610 m (2000 ft) if illuminated by high-beam lights. Before any conclusions may be drawn about the effectiveness of the reflectors in eliminating grade-crossing accidents, two questions must be answered: How much sight distance is needed for safe stopping, and how inadequate is the visibility of unreflectorized cars?

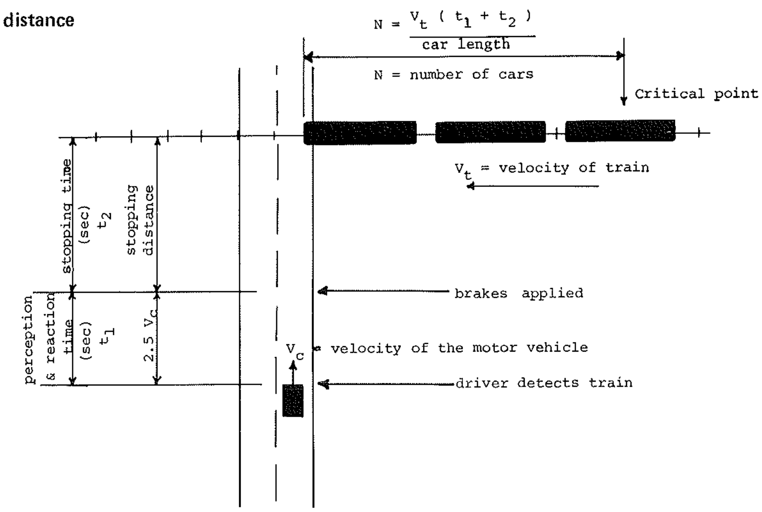
Stopping distance for speeds of 16 km/h (10 mph) to 113 km/h (70 mph) for dry, wet, and icy pavements were computed by using a 2.5-s perception and reaction time. A stopping distance of 152 m (500 ft) should be

Table 1. Nighttime visibility distances for reflectorized and unreflectorized railroad cars.

Type of Rail Car	Approximate Detection Distances[a] (m)					
	Two-Lamp Low-Beam System			Four-Lamp High-Beam System		
	Possibly Visible	Probably Visible	Visible	Possibly Visible	Probably Visible	Visible
Empty flat car						
Unreflectorized						
Black	<30	<30	<30	91	46	<30
Red	30	<30	<30	137	91	<30
White	91	46	<30	213	152	46
Reflectorized	366	305	152	610	457	274
15-m box car						
Unreflectorized						
Black	30	<30	<30	183	137	46
Red	46	30	<30	244	213	76
White	137	76	30	488	335	137
Reflectorized	366	305	152	610	457	274

Note: 1 m = 3.3 ft.
[a]Based on data reported in the 1947 edition of the IES Lighting Handbook.

Figure 5. Relationship between vehicle stopping distance and critical point on train.



adequate for most highway driving speeds and, since visibility distances of reflectors even when illuminated by low beams exceed 152 m, retroreflectors on the sides of railcars should provide adequate visibility to allow for safe stopping under most conditions experienced at railroad-highway grade crossings.

In order to assess the visibility of existing unreflectorized railcars, visibility ranges comparable to the three used for the reflectors were estimated for a standard 15-m (50-ft) boxcar and for an empty 15-m (50-ft) flatcar (Table 1).

Visibility ranges of unreflectorized railroad cars are significantly shorter when illumination is provided by low-beam headlights rather than by high beam. With the exception of dark-colored empty flatcars, visibility distances for unreflectorized cars illuminated by high-beam headlights seem to be adequate for safe operation at normal highway speeds. On the other hand, illumination by low-beam headlights does not even allow for safe stopping distance at 32 km/h (20 mph).

Given the low visibility of unreflectorized railcars illuminated by low-beam headlights and the fact that most drivers fail to use high-beam lights when they should, it follows that the increased visibility distances provided by reflectorization of railcars should be effective in eliminating certain grade-crossing accidents. The extent of the benefits anticipated from reflectorization is discussed next.

Benefits of Reflectorization

Accidents were classified into four groups. Category 1 consisted of accidents in which the motor vehicle strikes the train at a point that is far enough back along the train to indicate that the driver could have stopped safely if he or she had detected the train's presence just as it started crossing the highway. To determine which accidents met this criterion, a "critical point" (see Figure 5) on the train was computed by using the motor vehicle speed, the train speed, and the condition of the pavement (dry, wet, or icy). If the motor vehicle hit at or behind the critical point, the accident was included in category 1 if the location hit was not the first car or unit and in category 2 if the location hit was the first car or unit. Accidents involving a motor vehicle hitting a train forward of the critical point were included in category 3. Category 4 comprised all accidents in which the train hit the motor vehicle.

Category 1 includes the accidents most likely to be eliminated by reflectorization. Assuming that reflectors are effective, then the only nighttime accidents in this category that would not be eliminated are those that occur at grade crossings where the view of the tracks is obscured, those that occur because of motor vehicle equipment failures, or those that occur because of human factors such as poor eyesight, intoxication, attempted suicide, sleeping at the wheel, or bad judgment.

A smaller proportion of the accidents in category 2 is expected to be rectified by reflectorization. In order for an accident to be included in category 2, it must have had a critical point of less than 15 m (50 ft) (one car length). In some cases the short critical distances were caused by a blank in the data field for either the motor vehicle speed or the train speed.

Categories 3 and 4 contain those accidents least likely to be eliminated by reflectorization. In order for reflectors to be effective in preventing accidents from these categories, the train would have to be visible before it reached the grade crossing. Since the analytical studies of reflector effectiveness (Figure 3) do indicate that trains would be visible at up to 61 m (200 ft) before they reach the grade crossing, it is likely that some of the category 3 and 4 accidents could be prevented by reflectorization.

Calculation of Benefits

A three-step process was used to estimate the number of accidents that would be eliminated by reflectorization. First, the number of accidents that were potentially caused by nighttime visibility problems was estimated from the 1975 FRA computer-file accident data. Next, accidents occurring under circumstances in which reflectors would not be effective (e.g., bad weather, visual obstructions, intoxicated drivers) were eliminated. Finally, the accidents were reduced to reflect the proportion of grade crossings in which highway-railroad geometry does not allow for effective use of reflectors.

A comparison was made of the accident rates at night (and dawn or dusk) with those that occur during daylight. Relative accident rates for each of the four categories of accidents are given in the table below.

	Category			
Item	1	2	3	4
Passive warning				
Dawn or dusk	3.7	3.1	1.4	1.7
Night	9.2	4.0	0.9	1.4
Active warning				
Dawn or dusk	3.2	2.6	1.5	2.1
Night	7.5	3.6	1.8	2.0
All crossings				
Dawn or dusk	3.4	2.9	1.4	1.9
Night	8.6	3.9	1.2	1.6

The accident rates are expressed as ratios and indicate the relative occurrence rate of each accident category in relation to the daylight rate. For example, the value of 9.2 for category 1 accidents occurring at night at crossings that have passive controls indicates that this particular type of accident is 9.2 times more likely to occur at night than it is during daylight. Variations in train traffic volumes by time of day have not been considered in determining these relative accident rates. It is assumed that train volumes at night are equal to or less than daylight volumes and thus do not add to the decreased exposure rate that occurs at night.

Some accident reduction is expected at actively protected crossings. Previous studies of reflectorization have limited the benefits to passively protected crossings on the assumption that actively protected crossings already inform the motorist of the impending presence of a train and that reflectors would add nothing to warn the driver. A study of driver behavior at signalized railroad crossings (20) found a surprisingly high rate of "critical incidents" (vehicles not stopping for the signal or zigzagging around fully descended gates) during signal alarm periods. Furthermore, the fact that the nighttime category 1 accident rate at actively protected crossings is more than seven times the daytime rate indicates that visibility is most likely a contributing factor in these accidents. Since no program of reflectorization could hope to provide visibility levels better than those experienced in daylight conditions, the daylight accident rates were used as the upper limits on effectiveness of reflectorization.

The relative proportions of travel occurring during the day, dawn or dusk, and night periods were used to compute the number of accidents that corresponded to the daylight accident rate. For example, there is 32 percent as much travel at night as there is during daylight; thus, one would expect 32 percent as many accidents to occur at night as occur during the day if visibility and other nighttime-related phenomena are not a problem. The numbers of accidents potentially caused by nighttime visibility problems were obtained by subtracting 32 percent of the daylight accidents from the night accidents and 6.3 percent of the daylight accidents from the dawn or dusk accidents. These values are shown in Table 2.

It is assumed that the daylight accident rates include those accidents caused by motor vehicle equipment failure and human factors. It seems reasonable to assume that accidents resulting from motor vehicle equipment failures, attempted suicide, heart attacks, or bad judgment should be equally likely to occur at night as they are during the day. On the other hand, accidents resulting from human factors such as poor eyesight, intoxication, or sleeping at the wheel are more likely to occur at night than during the day.

Category 1 accidents represent 3.8 percent of all grade-crossing accidents during the day. Since this category of accidents is caused primarily by visibility problems that should not exist during the day, its occurrence rate should represent the nonpreventable accidents discussed above. Accident data from Pennsylvania (22) were available in a form that allowed comparison between grade-crossing accidents and general highway accidents. A total of 3.6 percent of the 274 grade-crossing accidents that occurred in Pennsylvania in 1976 were caused by motor vehicle equipment failure and human factors.

Reflectors are effective only when the driver is able to see them and perceive that the reflectors are on a train. Visibility of the reflectors can be affected by physical obstructions, weather conditions, and human factors such as poor eyesight and intoxication. The ability to perceive and react to the situation is affected by the drivers' attentiveness or degree of intoxication. It was estimated that 65 percent of all drivers were alert.

Table 2 shows the percentages of accidents that occurred under the various conditions that would permit reflectors to be effective. Data on the presence of obstructions and adverse weather (fog, snow, or ice conditions) were obtained from the FRA accident file.

Alcohol and Other Human Factors

Certain causal factors in accidents are more prevalent during the night than during the day. Accidents caused by excessive use of alcohol, drowsiness, and poor eyesight fall into this category.

Limited data are available on the roles of alcohol and other human factors in railroad-highway grade-crossing accidents. Data from Pennsylvania (22); Alameda and Sacramento Counties, California (23); and Dade County, Florida (24) were used in assessing the impact of these factors on reflector effectiveness. On the basis of the results of these studies, it is esti-

Table 2. Factors that affect the effectiveness of reflectors.

Accident Type	Total Accidents	Caused by Daylight Factors	Caused by Nighttime Factors	Percentage With No Obstructions	Percentage Without Adverse Weather	Percentage With Acceptable Speeds	Reflector Effectiveness (%)
Passive warning							
Category 1							
Night	396	43	353	93.3	90.6	100.0	49.0
Dawn or dusk	31	8	23	96.4	75.0	100.0	35.5
Category 2							
Night	375	93	261[a]	90.1	94.9	100.0	38.7
Dawn or dusk	56	18	38	83.0	94.3	100.0	33.9
Category 3							
Night	129	150	0[a]	90.1	83.5	57.8	0.0
Dawn or dusk	41	30	11	79.5	87.1	57.8	7.1
Category 4							
Night	1349	984	365	90.0[b]	93.0[c]	25.0	3.7
Dawn or dusk	336	194	142	80.0[b]	93.0[c]	25.0	5.1
Total	2713	1520	1193	89.0	92.2	50.3	16.1
Active warning							
Category 1							
Night	255	34	221	95.6	94.7	100.0	51.0
Dawn or dusk	21	7	14	94.7	89.5	100.0	38.1
Category 2							
Night	237	65	172	95.2	97.8	100.0	43.9
Dawn or dusk	33	13	20	87.5	96.9	100.0	33.3
Category 3							
Night	145	82	63	90.6	88.5	57.8	20.3
Dawn or dusk	24	16	8	95.7	87.1	57.8	9.7
Category 4							
Night	1157	571	586	90.0[b]	93.0[c]	25.0	6.7
Dawn or dusk	238	112	126	90.0[b]	93.0[c]	25.0	7.2
Total	2110	900	1210	91.4	93.4	46.5	17.9
All crossings	4823	2420	2403	90.0	92.7	48.6	16.9

[a] For category 3 accidents at passive warnings, the nighttime accident rate was less than the daylight rate. The anomaly is probably due to the misclassification of accidents into category 2 because of missing data for train or motor vehicle speed; 21 accidents were subtracted from the category 2 accidents to make up for the deficit in category 3 accidents.
[b] Estimated.
[c] From 1975 FRA Rail-Highway Grade-Crossing Accidents/Incidents Bulletin (21).

mated that 35 percent of the accidents involve drivers who are sufficiently impaired that they would not be expected to detect and perceive the presence of a train from illuminated reflectors.

Effects of Highway-Railroad Geometry

Very little information about the geometry (vertical and horizontal) of railroad-highway grade crossings is available. The Association of American Railroads (AAR)-FRA Grade-Crossing Inventory contains information about the crossing angle of the highway and railroad, but it contains nothing about the vertical or horizontal alignments of the two routes. The grade-crossing geometry, along with natural and manmade obstructions, determines the visiblity at a crossing.

The visibility requirements necessary to eliminate category 1 and category 2 accidents are different from those needed for category 3 and category 4 accidents. In category 1 and 2 accidents, it is only necessary to see the highway-railroad intersection. In order for category 3 and 4 accidents to be eliminated, it is necessary to see the train at some point before it reaches the crossing. The actual distance up the track that the train is required to be visible depends on the train speed and the motor vehicle speed.

The proportion of accidents in which the train speed and motor vehicle speed are both such that the train would be within the range of the motor vehicle's headlights soon enough for the driver to stop is shown in Table 2. For category 1 and 2 accidents, this value is 100 percent, since the train does not have to be seen until it is across the intersection.

The overall effectiveness of reflectorization (assuming adequate crossing geometry for proper visibility) was found by multiplying the percentages for "no obstructions," "without adverse weather," "alert drivers" (65 percent), and "acceptable speeds" by the proportion of total accidents that were caused by nighttime factors. These effectiveness values are shown in Table 2.

Table 3 summarizes the maximum benefits anticipated from reflectorization. These are the benefits that would accrue if all crossings had the proper geometry to allow for adequate nighttime visibility. The numbers of fatalities and injuries and the amounts of property damage were obtained from the FRA 1975 computer-file accident data. Property-damage figures include damage to the motor vehicle, the train equipment, and the track and signal structures.

Costs of Reflectorization

The costs of a reflectorization program were divided into five categories of costs: initial material costs, initial installation costs, annual replacement costs for reflectors destroyed by vandals or train operations, annual maintenance costs for cleaning reflectors, and program implementation costs. Costs are based on the following assumptions:

1. High-intensity, high-tack reflective sheeting is used at a cost of $\$23.14/m^2$ ($\$2.15/ft^2$).
2. Each railroad car is equipped with four (two per side) 15×15-cm (6×6-in) squares of reflective sheeting.
3. Each locomotive is equipped with six (three per side) 15×15-cm squares of reflective sheeting.
4. Five percent wastage of material occurs.
5. The installation rate is 30-60 reflectors/work hour.
6. Labor costs are \$20/h.
7. No special handling of cars is required for installation or maintenance (work will be done during required inspections).

Table 3. Maximum annual benefits of reflectorization.

Accident Type	Total Accidents	Reflector Effectiveness (%)	Reduction in Accidents	Reduction in Fatalities	Reduction in Injuries	Reduction in Property Damage ($)
Passive warning						
Category 1						
Night	396	49.0	194	17	92	257 740
Dawn or dusk	31	35.5	11	0	3	5 030
Category 2						
Night	375	38.7	145	5	47	98 510
Dawn or dusk	56	33.9	19	0	7	11 100
Category 3						
Night	129	0.0	0	0	0	0
Dawn or dusk	41	7.1	3	0	1	2 980
Category 4						
Night	1349	3.7	50	3	13	48 124
Dawn or dusk	336	5.1	17	2	4	20 441
Total	2713	16.1	439	27	167	443 925
Active warning						
Category 1						
Night	255	51.0	130	7	64	136 850
Dawn or dusk	21	38.1	8	0	1	10 900
Category 2						
Night	237	43.9	104	4	50	70 370
Dawn or dusk	3	33.3	11	0	6	6 490
Category 3						
Night	145	20.3	29	4	16	39 040
Dawn or dusk	24	9.7	2	0	0	2 160
Category 4						
Night	1157	6.7	77	5	21	77 770
Dawn or dusk	238	7.2	17	1	4	19 820
Total	2110	17.9	378	21	162	363 400
All crossings	4823	16.9	817	48	329	807 325

Table 4. Estimated costs of reflectorizing U.S. railroad rolling stock.

Cost Category	Unit Cost ($)	Estimated Cost Ranges (1977 $) First Cost ($000 000s)	Estimated Cost Ranges (1977 $) Equivalent Annual Cost[a] ($000 000s)
Material	23.14/m²	4.0	0.8
Installation	0.33-0.67/reflector	2.3-4.6	0.5-0.9
Annual replacement (5 percent/year)	0.90-1.25/reflector	-	0.3-0.4
Maintenance	0.25-0.50/reflector		
Once a year		-	1.7-3.5
Once in two years		-	0.9-1.7
Program implementation			
Research and development	100 000		
Program development	100 000		
Public education	100 000		
Administration (per year)	125 000		0.2
Total		6.3-8.6	2.7-5.8

Note: 1 m² = 10.7 ft².
[a] Discount rate = 10 percent.

8. The reflective material has a seven-year economic life.
9. The discount rate is 10 percent.

Table 4 contains a summary of the cost estimates for the reflectorization program. Ranges of costs are given for items that cannot be estimated exactly, due to insufficient documented data. Annual costs are expected to be between $2.7 and $5.8 million. The cost of maintenance is the area that has the highest degree of uncertainty. It is also a major component of the total project cost. Research is needed to answer the questions about the frequency of maintenance required and its associated cost.

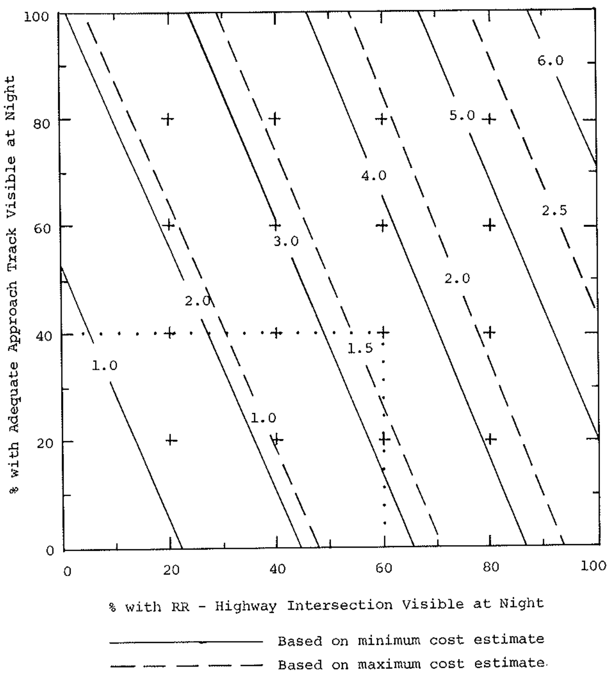
Another unknown that affects the cost of the project is the optimum pattern to be used in placing the reflectors on the railcars. Cost estimates in Table 4 assume that two reflectors are placed on each side of each car. It may be desirable to use additional reflectors on high freight cars to provide a delineating effect that will reduce driver perception time. Again, field research is needed to determine the best pattern to be used. Additional annual costs for extra delineators on high-side cars could run as high as $1.5-3.2 million.

Cost-Effectiveness Analysis

The difficult task of assigning dollar values to the benefits that result from savings in human life and injury was accomplished by using values determined by the National Highway Traffic Safety Administration (NHTSA) (25). NHTSA has made a considerable effort to establish the societal costs of motor vehicle fatalities and injuries. If a reflectorization program is to receive funding, then its cost-effectiveness should be compared with the cost-effectiveness of other proposed safety programs to see whether it merits the spending of scarce dollars. Thus, the absolute values of the benefits assigned to injuries and fatalities is less important than the consistency of values used when comparing the cost-effectiveness of several competing projects.

A value of $318 000 has been used as the average societal cost of a fatality; this is the NHTSA 1975 value updated to 1977 dollars by using a 6 percent annual inflation rate. A value of $5000 has been used as the average societal cost of an injury. This value falls be-

Figure 6. Ranges of benefit/cost ratios for the reflectorization program.



tween the costs established by NHTSA for a moderate injury and that for a severe, but not life-threatening, injury. Property-damage values were obtained from the FRA Grade-Crossing Incident and Rail Equipment Accident files and were updated to 1977 dollars.

The anticipated annual benefits shown in Table 3 were converted into dollars by using the values given above. The actual level of benefits depends on the proportion of grade crossings that have suitable nighttime visibility.

Figure 6 shows the expected benefit/cost ratio for the reflectorization program (based on four reflectors per railcar) as a function of the proportion of grade crossings that have suitable geometry to allow for proper visibility. The solid lines represent the benefit/cost ratio that would result if the project costs were equal to the minimum cost estimate. The broken lines are based on the maximum cost estimate. For example, the dotted lines on Figure 6 show that, if 60 percent of the U.S. grade crossings have geometry such that the railroad-highway intersection is adequately visible and 40 percent of the grade crossings have geometry that allow for adequate visibility of a train as it approaches the intersection, then the expected benefit/cost ratio for a reflectorization program would be between 1.6 (maximum cost estimate) and 3.5 (minimum cost estimate).

Throughout the analysis an attempt has been made to estimate quantities conservatively. The effectiveness analyses were done by assuming low-beam headlight illumination. Much greater visibility is obtained by high-beam lights, and at least 25 percent of the drivers can be expected to use them.

It is important to note that the majority of the benefits are to society and not to the railroads. Other benefits to the railroads may result if liability costs are reduced by the decreased number of accidents. It is possible, however, that liability costs could actually increase if a federal regulation requiring reflectors were passed. With a regulation in force, a dirty or missing reflector could provide the avenue for a negligence suit against the railroad.

FUTURE RESEARCH NEEDS

Research needs to be done to determine the size, pattern, and location of retroreflectors on the sides of railroad rolling stock that will optimize motorist detection and perception of a train's presence.

Research should be conducted to examine the decrease in reflectivity of retroreflectors that is caused by continuous use in railroad environments. This research is needed to determine whether maintenance is required.

Further research investigating driver behavior in the vicinity of railroad-highway grade crossings with both active and passive warning devices should be conducted.

ACKNOWLEDGMENT

The research described in this paper was performed while I was employed by FRA as part of its Summer Professor's Program. Other individuals who contributed to this report are Christa Cudebec, Nancy Ulmer, and Helen Mathias, all of Bucknell University. I wish to express my appreciation to Bruce F. George, Edward S. Lord, and Richard E. Snow of FRA for their advice and assistance in this work. The contents of this paper reflect my views, and I am responsible for the facts and the accuracy of the data reported herein. The contents of this paper do not necessarily reflect the official views or policy of the U.S. Department of Transportation.

REFERENCES

1. Prevention of Rail-Highway Grade-Crossing Accidents Involving Railway Trains and Motor Vehicles: Brief of Railroad Respondents. Interstate Commerce Commission, Docket No. 33440, Feb. 10, 1964.
2. Rail-Highway Grade-Crossing Accidents/Incidents Bulletin for 1977. Office of Safety, Federal Railroad Administration, 1977.
3. H. I. Stalder and A. R. Lauer. Effective Use of Reflectorized Materials on Railroad Boxcars. HRB, Bull. 89, 1954, pp. 70-75.
4. A. R. Lauer and V. W. Suhr. An Experimental Study of Four Methods of Reflectorizing Railway Boxcars. HRB, Bull. 146, 1956, pp. 45-50.
5. Prevention of Rail-Highway Grade-Crossing Accidents Involving Railway Trains and Motor Vehicles: Brief of Railroad Respondents. Interstate Commerce Commission, Docket No. 33440, Nov. 15, 1962.
6. Report to Congress: Railroad-Highway Safety, Part I—A Comprehensive Statement of the Problem. Federal Railroad Administration and Federal Highway Administration, Nov. 1971.
7. J. R. Lepkowski and W. F. Mullis. Considerations for Improving the Conspicuity of Railroad Rolling Stock, Phase V Report. National Bureau of Standards and Federal Railroad Administration, FRA-AR 20033, Sept. 1973.
8. J. B. Hopkins and A. T. Newfell. Guidelines for Enhancement of Visual Conspicuity of Trains at Grade Crossings. Transportation Systems Center and Federal Railroad Administration, FRA-OR&D-75-71, May 1975.
9. K. G. Cook; Century Research Corporation. Reflector Analysis. National Highway Safety Bureau, U.S. Department of Transportation, Aug. 1969.
10. R. N. Robertson. Use of High-Intensity Reflective Materials in Highway Signing: A Literature Re-

view. Virginia Research Council, Charlottesville, Aug. 1978.
11. C. T. Hare and R. H. Hemion. Head-Lamp-Beam Usage on U.S. Highways. Southwest Research Institute, San Antonio, TX, and Federal Highway Administration, Rept. No. AR-666, Dec. 1968.
12. N. T. Cox. The Effect of Dirt on Vehicle Head-Lamp Performance. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR 240, 1968.
13. K. Rumar. Dirty Headlights: Frequency and Visibility Effects. University of Uppsala, Uppsala, Sweden, Rept. 136, 1973.
14. H. L. Woltman and W. P. Youngblood. Indirect Factors Affecting Reflective Sign Brightness. TRB, Transportation Research Record 611, 1977, pp. 35-37.
15. P. L. Connolly. Vision, Man, Vehicle, and Highway. In The Prevention of Highway Injury, Highway Safety Research Institute, Ann Arbor, MI, 1967, pp. 122-149.
16. W. Heath and D. M. Finch. Effect of Tinted Windshields and Vehicle Headlighting on Night Visibility. HRB, Bull. 68, 1953, pp. 1-15.
17. E. Wolf, R. A. McFarland, and M. Zigler. Influence of Tinted Windshield Glass on Five Visual Functions. HRB, Bull. 255, 1960, pp. 30-46.
18. IES Lighting Handbook, 5th ed. Illuminating Engineering Society, New York, 1972.
19. F. C. Breckenridge and C. A. Douglas. Development of Approach and Contact-Light Systems. Illuminating Engineering, Vol. 40, 1945, p. 785.
20. G. J. S. Wilde and others. An Observational Study of Driver Behavior at Signalized Railroad Crossings. Canadian Institute of Guided Ground Transport, Kingston, Ontario, Feb. 1976.
21. Rail-Highway Grade-Crossing Accidents/Incidents Bulletin for 1975. Office of Safety, Federal Railroad Administration, 1975.
22. Statistical Summary for the Period January through December 1976, Encompassing All Railroad Crossing Accidents. Pennsylvania Department of Transportation, Harrisburg, 1977.
23. J. A. Waller. The Role of Alcohol in Fatal Collisions with Trains. Northwest Medicine, Sept. 1968, pp. 852-856.
24. Alcohol, Violence, and the Automobile. Medical Examiner's Office, Dade County, FL, 1968.
25. 1975 Societal Costs of Motor Vehicle Accidents. National Highway Traffic Safety Administration, U.S. Department of Transportation, Dec. 1976.

Discussion

Louis T. Cerny, Erie Western Railway Company, Huntington, Indiana

The reflectorization of railroad rolling stock appears to me to be a good idea. Certain factors would make it appear even more favorable than the study shows. In most cases the reflectors would be moving, causing increased probability of detection, and more than one reflector is likely to be in view at all times. I do not believe these additional favorable effects were taken into consideration.

Another favorable aspect is that the cleaning and application of the reflectors is a low-skill job and, because of the national nature of the car fleet, the work could be concentrated at railroad terminals in areas of high unemployment. This project would appear to involve direct federal funding because of the mobile nature of the railroad vehicles, which are not restricted to one state.

It appears, however, that a certain effect that weakens the case for reflectorization was not considered. In a substantial percentage of cases, headlights from highway traffic in the opposing direction can be seen through the spaces between the moving train cars, or under the bodies of the cars between the wheels, thus creating a very eye-catching effect that is more visible than the reflectors, and this is an additional circumstance under which the reflectors would not be effective.

In category 3 and 4 accidents, the locomotive's headlight would normally be visible long before the reflectors, because its visibility does not depend on reflectivity and because it is much higher off the ground than the reflectors. The illumination of objects around the crossing by the locomotive headlights as the train approaches may also attract more attention to the train than the reflectors would, especially since this effect precedes the arrival of the train.

I am skeptical that reflectors on the cars could create a significant increase in attracting a motorist's attention when the crossing is protected by gates (which normally have flashing lights on the gate in addition to those on the mast); I believe, therefore, that accidents occurring at gated crossings are very unlikely to be prevented by reflectors. In some cases, the gate may actually block the view of the reflector. Is there any reason that the distinction between gates and flashers and flashers alone was not made? Perhaps the accidents that occur at gated crossings should be taken as the limit of the effectiveness of reflectors, rather than daylight conditions.

Perhaps a further analysis of category 2 accidents should be made. The paper indicates that in "some cases" accidents fell into this category because of a "blank in the data field." Notes in Table 2 indicate that some adjustment was made, but no justification is given. It would appear that a similar adjustment would be needed in the "active warning" category.

More than four reflectors per car (two per side) would probably be needed on cars more than 18 m (60 ft) long. Common types of cars, such as piggyback, automobile racks, and automobile parts cars are about 26-27 m (85-90 ft) long. It would seem that the maximum distance between reflectors should be about 9 m (30 ft). I feel that answers to these questions, which reflect both positively and negatively on the project, are worth evaluating.

John B. Hopkins, Transportation Systems Center, U.S. Department of Transportation, Cambridge, Massachusetts

McGinnis has carried out a comprehensive and penetrating analysis that appears to achieve as definitive an answer as can reasonably be expected concerning the costs and benefits of railcar reflectorization. His paper brings out many of the uncertainties that are inherent to our basic lack of knowledge concerning accident causation, driver behavior, reflector degradation in the railroad environment, etc. In most cases I find his assumptions and estimates to be quite reasonable. It is my purpose in these brief comments to address only a few aspects in which I feel the ambiguities are so important as to warrant special attention. My aim is not to criticize, for I have no substantive complaints with the study. Rather, I wish to emphasize sources of uncertainty that I consider to be relevant to interpretation of the results, particularly with reference to formulation of policy in this area.

McGinnis has effectively placed an upper bound (the daylight accident rate) on the safety benefits that might be associated with reflectorization. The question then becomes one of estimating appropriate reductions from this value due to various limitations. One could quibble over matters of headlight aim, the assumption of low-beam operation, stopping distances, etc. However, these are minor points, and they tend to balance one another. More complex is the need to assess whether those accidents identified as relevant are truly caused by visibility problems of a type that could be mitigated by reflectors. My subjective view is that reflector effectiveness as shown in Table 2 is somewhat optimistic, or at least represents only a reasonable upper bound, particularly at crossings that have active-warning systems. For example, I find it quite unlikely that 51 percent of the drivers who fail to respond to conventional railroad-crossing flashing lights (some with gates) for night-related reasons will be deterred any more effectively by railcar reflectors. This is a relatively important question, since Table 3 shows that 46 percent of the expected accident reduction is to occur at such crossings.

A factor that affects both cost and effectiveness is reflector maintenance. One can envision many possible maintenance scenarios, each with its own benefit-cost implications. To my mind, the most realistic assumption is that of no maintenance at all. This substantially reduces estimated costs (by 33 percent for the "minimum cost" case and by 60 percent for "maximum cost"), while having a negative but indeterminate effect on safety. (It is appropriate to mention here that other types of reflectors could be used. For example, plastic devices used as highway delineators have somewhat less desirable optical characteristics in this application, but they appear to perform well in a rather dirty environment for many years without cleaning or replacement.)

In the context of policy formulation, another set of factors takes on real significance. These involve the effects of other activities that are expected to improve grade-crossing safety. For example, there are now under way major efforts to improve both passive- and active-warning systems and to achieve more widespread installation of train-activated devices. Reflectorized crossbucks, improved flashing lights, and increased use of gates are of obvious significance to the subject. Serious government and industry consideration is currently being given to widespread installation of locomotive-mounted strobe lights, which should do all that can be done through visibility enhancement to prevent the accidents McGinnis places in categories 3 and 4 (collisions occurring close to the front of the train). These represent 31 percent of the total estimated fatality reduction, which would thus be eliminated as a potential reflector benefit. There could also be a very significant impact on categories 1 and 2. (It is not claimed that strobe lights will necessarily prevent these accidents. However, for potential collisions near the locomotive, if strobes do not help, reflectors are unlikely to succeed either.)

The basic conclusions of the paper, as presented in Figure 6, assume maximum and minimum cost estimates. I suggest that for a more realistic estimate one should use a single no-maintenance cost assumption that still has two curves, based on minimum and maximum estimates of benefits. For the no-maintenance scenario, with full consideration of the limitations on potential safety effectiveness described above, reasonable minimum and maximum benefits might be approximately 25 and 75 percent of the values projected in the paper. The net effect of these modifications, which reduce both costs and benefits, is relatively small; I infer a subjective "most likely" benefit-cost ratio probably in excess of 1.0 but less than 2.0. It should be noted at this point that the benefits accrue primarily to society in general and only to a limited degree to the railroads. Installation at railroad expense would thus almost certainly have a benefit-cost ratio for them well below 1.0. From either the societal or railroad viewpoint, there may well be other investments in crossing safety that can be expected to yield greater benefits. To keep this matter in perspective, note that the above estimates imply a maximum saving of 12 to 36 lives/year, prior to correction for geometric factors that could easily diminish the benefits by another factor of 2 to 4. The net effect on crossing safety would thus be an improvement of approximately 1-2 percent. Thus, while reflectorization may ultimately prove to be a worthwhile step, with significant benefits, it does not appear to be of major importance to crossing safety in general.

I am in full agreement with the research needs McGinnis has identified, and I would only add reflector type and cost to the reflector optimization study. At the same time, the relatively limited promise of reflectorization, and the difficulty of obtaining definitive answers to these questions, seems to warrant only a modest priority for such research.

Otto F. Sonefeld, Atchison, Topeka, and Santa Fe Railway Company, Chicago, Illinois

As McGinnis indicated, the reflectorization of railroad rolling stock has been the subject of debate many times over the past few decades. Arguments favoring reflectorization have generally failed to show significant evidence of the effectiveness of this approach, particularly when compared with substantial argument in favor of other grade-crossing safety activities.

The McGinnis study is perhaps the most comprehensive look at this subject to date, although it leaves many questions unanswered. The problem, in my opinion, is in the attempt to draw fairly firm conclusions from data that do not lend themselves to such detailed analysis. McGinnis has done a commendable job under these circumstances, but it has required making certain assumptions that I feel should be more critically explored.

The part of the study that deals with reflector visibility distances appears to be well documented and reasonable. One of the shortcomings of previous proposals for reflectorization has been the inability of engineering-grade reflective material to function effectively in the severe railroad environment without diligent maintenance. The introduction of high-intensity reflective material would seem to diminish this problem, although it is not clear that even the use of that material would produce the 0.50 rate of efficiency used in the study. The recommended location for the reflective material is the most severe environment on a railcar.

This is not to say that diligent maintenance could not overcome this problem; however, experience in the automatic car identification (ACI) program does not indicate the capability or will of the rail industry to properly maintain the reflective material unless there is a return to the industry far greater than that provided by the ACI program. Such benefits are not apparent. In any event, the long-term effectiveness of the material seems open to question. Certainly, the cost-benefit ratios would be affected by increased maintenance requirements.

In the same vein, although the report acknowledges that a motorist may have a problem perceiving the recommended light source as a railroad train on a crossing, the perception time used in the report appears to assume that a motorist immediately recognizes the light source as a crossing hazard and comes to a prompt halt. I would suggest that a train crossing is unusual enough in the total traffic scheme that a longer perception time would be required to recognize it for what it is.

In developing the "critical point" used as the basis for analysis of the FRA accident reports, apparently vehicle speeds as stated on the reports are used. If so, it appears that vehicle speeds would tend to be consistently understated, inasmuch as the accident reports require speed at the time of impact, not the approach speed at which the decision to stop must be made. If I understand the rationale behind development of the critical point, this would then have the effect of placing that point further back in the train and thus of reducing both the number of accidents shown in categories 1 and 2 and the number of vehicles that would realize any benefits from reflectorization.

Highly important are the assumptions in this study that result in a finding that category 1 accidents at active-warning crossings are 7.5 times more likely to occur at night than during daylight. This further translates to a finding that reflectorization would prevent 51 percent, or 130, of these accidents. Raw data for 1975, however, show a total of 704 nighttime accidents at active-warning crossings of the ran-into-train variety, and 650 during daylight. The exact methodology for derivation of the figures in the study is not known, but my reaction is that the study figures are excessively high, compared with actual figures. This, of course, has a significant effect on the cost-benefit analysis in the report.

Also important to this analysis is the number of accidents used as the base figure, that is, potentially preventable. If 6:00 p.m. to 6:00 a.m. is a reasonable period in which to categorize nighttime accidents, the FRA report for 1975 shows only 1658 ran-into-train accidents in that period, many of which would involve striking the locomotive. The study, on the other hand, appears to be using a base figure of 4823 potentially preventable accidents. I do not understand these differences.

Another problem that is acknowledged but not used in the cost-benefit study is the number of crossings at which vertical and horizontal alignment is such as to eliminate these crossings as candidates for improvement by reflectorizing cars. I would suggest that the number is sizable.

Another category of accidents that is not discussed in the report, but that could possibly be elimin•ted from consideration for treatment by car reflectorization, is those ran-into-train accidents that occurred at illuminated crossings. This involves a minimum of 576 accidents in 1975 (677 in 1977), although the FRA report does not break these into nighttime and daylight accidents.

These comments are not meant to belittle the basic concept of reflectorizing rolling stock. Undoubtedly there are many crossing situations that lend themselves to this treatment. Whether they are of the magnitude suggested in the study is, in my opinion, a matter that requires more rigorous examination.

McGinnis correctly suggests further research into various aspects of this matter. I agree with these suggestions and, as indicated by my comments in this discussion, I would also suggest further refinements or clarifications of some of the critical factors involved in the development of costs and benefits associated with this subject.

Author's Closure

It does not seem that blinking lights from opposing headlights shining through the spaces between moving railcars will affect the results of this study; if what Cerny says is true, and I think it is, then drivers aided by these blinking lights are already seeing the train and are safely stopping. Thus, they are not becoming FRA accident statistics and would not be touched by the potential benefits of reflectorization.

Cerny indicated a concern about the impact of locomotive headlights on potential reductions of category 3 and 4 accidents from reflectorization. There are problems in the use of locomotive headlights as a means of informing motorists about the impending danger of an approaching train. Locomotive headlights are placed close together, giving the impression of a single light, and are aimed in a very narrow beam. First, the lack of space between the two lamps does not allow a motorist to judge distance in the way he or she can with widely spaced automobile lamps. Second, the narrow beam of the locomotive headlight makes detection of these lights difficult for approaching vehicles. In a study conducted on the visual conspicuity of trains at grade crossings (8), Hopkins and Newfell concluded that a beam width of up to 150° would be required if visibility to a great majority of vehicles is to be achieved. Very little light is visible from a locomotive headlight at angles of greater than 15° to 20°; thus, locomotive headlights cannot be assumed to be effective in providing visibility to approaching vehicles.

The missing data responsible for the misclassification of certain category 3 accidents into category 2 do not seem to be too important in regard to the final study results. A sensitivity analysis was conducted to determine the impact of changing perception and reaction time on accident classification. This analysis indicated that the results are very insensitive to reaction and perception time, which also indicates that the results would be fairly insensitive to variations in vehicle and train speeds.

All three discussants expressed concern about the high effectiveness of reflectors shown at actively controlled crossings. I would point out that the analysis indicated that 51 percent of category 1 accidents would be eliminated at active crossings if all crossings had adequate visibility. I suspect that more of the accidents at active-warning crossings than at passive-warning crossings are caused by restricted visibility at grade crossings and would not be eliminated by reflectorization. However, this question cannot be answered until more is known about actual visibility at grade crossings.

Sonefeld questioned the source of several figures and the exact methodology used in determining them. A more detailed description of the methodology is available in an FRA publication (26). The 130 accidents referred to by Sonefeld represent 51 percent of the 255 category 1 accidents that occurred at night at crossings that have active-warning systems (see Table 2). The base figure of 4823 potentially preventable accidents includes accidents in which the motor vehicle was struck by the train, i.e., category 4 accidents.

Without specific regulations to require the cleaning of reflectors, Hopkins' no-maintenance scenario is probably the most realistic. However, it certainly would be nice to have some research on the question of the impact of lack of reflector maintenance on reflector brightness.

Hopkins' suggestion for using a single cost estimate with estimates of minimum and maximum benefits to give a more realistic idea of the program's benefit/cost ratio is impossible until better cost data are available on installation costs and, more importantly, until information on grade-crossing visibility is obtained, so that ranges of benefits can be established. At this point, it is impossible to estimate minimum benefits.

REFERENCE

26. R. G. McGinnis. The Benefits and Costs of a Program to Reflectorize the U.S. Fleet of Railroad Rolling Stock. Federal Railroad Administration, 1979.

Publication of this paper sponsored by Committee on Railroad-Highway Grade Crossings.

Accident and Operational Guidelines for Continuous Two-Way Left-Turn Median Lanes

C. Michael Walton and Randy B. Machemehl, Department of Civil Engineering, University of Texas at Austin
Thomas Horne, city of Beaumont, Texas
William Fung, Federal Highway Administration

An investigation was begun to provide highway designers and traffic engineers with more definitive information on the installation of left-turn median lanes. Primary emphasis was on documentation of experiences with continuous two-way left-turn median lanes; however, for purposes of comparison, channelized one-way left-turn median lanes (raised and flush markings) were included. This paper presents a summary of the detailed investigation of the literature on left-turn lanes, the results of a survey of current practices and standards in Texas, results of field studies, and guidelines for use. A literature survey and analysis of questionnaires returned by representatives from Texas cities and the Texas State Department of Highways and Public Transportation suggested areas in which definitive guidelines were required. Based on the analysis of these two phases of the study, field studies were conducted that concentrated on operational characteristics, accident experience, and currently accepted practices. The analysis of the data collected on left-turn-lane sites revealed many characteristics, patterns, and relationships of accidents and operational experiences. A brief summary of the conclusions and findings is included, and recommendations are provided to improve current practices. In the operational characteristics phase of the study, emphasis was placed on the lateral placement of vehicles in the left-turn lane and the entering and maneuvering distances of vehicles within the lane. These suggest the characteristics of driver behavior that can be used by traffic engineers and highway designers in determining the optimum design elements for two-way left-turn lanes.

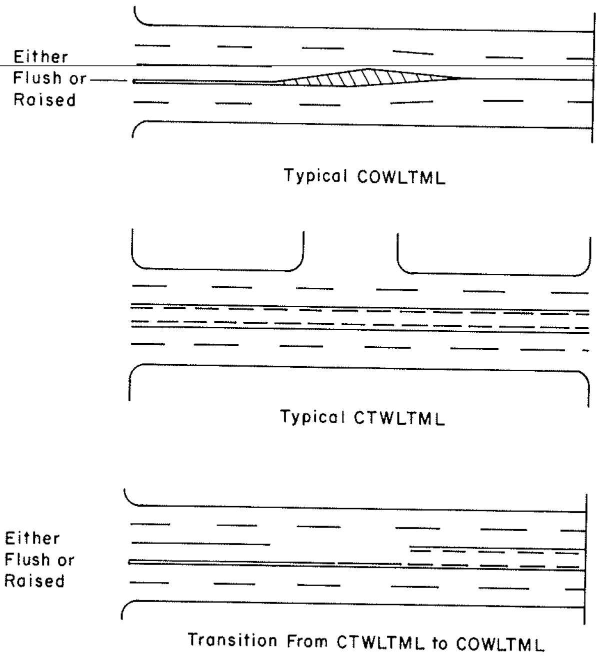
In recent years, there has been increased emphasis on improving the capacity and safety of existing traffic facilities through low-cost improvements or modifications. One concern among highway designers and traffic engineers is the treatment of medians on non-controlled-access highways in urban areas and the development of design and operational standards for median improvements. Although many guidelines have been developed to aid traffic engineers in considering left-turning vehicles, there are still many unanswered questions about how and when special median facilities should be provided.

Basically, three types of left-turn facilities are considered in this study: raised channelized one-way left-turn median lane (raised COWLTML), flush COWLTML, and continuous two-way left-turn median lane (CTWLTML).

A COWLTML (Figure 1) is a median left-turn lane that provides space for speed changes and storage for left-turning vehicles traveling in only one traffic direction to turn at a designated location along a two-direction roadway. A CTWLTML is a left-turn median lane that provides common space for speed changes and storage for left-turning vehicles traveling in either direction and that allows turning movements at any location along a two-way roadway. Raised channelization is generally defined as the use of a curb or other "nontransversible" delineator, while flush channelization generally refers to the use of paint, buttons, tile, or other easily transversible markings.

Although such median lanes have been in operation for some time, very little information has been compiled about their operational differences and about trade-offs between each type of left-turn facility. Therefore, the primary objective of this paper is to present the results of a study that was designed to (a) review previous studies related to traffic operations of left-turn lanes, (b) collect and analyze data for evaluating the operational characteristics of left-turn facilities, (c) identify relationships and characteristics of accidents associated with left-turn-lane facilities, and (d) develop guidelines for design and operational decisions for median treatments. The results presented should enable traffic engineers to better understand the impacts and trade-offs among various types of left-turn facilities in their decision-making process and will facilitate the design of left-turn lanes for individual sites.

Figure 1. Typical types of left-turn lanes.



BASIC DESIGN CONSIDERATION

From a review of pertinent publications, a list of design considerations and implications that focused on access, accidents, and congestion issues was prepared. These three issues are vital considerations in determining the need for a left-turn lane and in determining the type and design details of the facility. The list presented below contains some of the major access considerations that may affect safety, traffic flow, cost, feasibility, and public acceptance of left-turn-lane designs (1-3):

1. What is the abutting retailer's preference in type of access?
2. What is the driver's preference in type of access?
3. How is parking affected?
4. What changes are expected in movement volumes, lane use, traffic composition, etc.?
5. What pedestrian needs exist or are expected?
6. What changes in traffic control are anticipated?
7. What other access does the abutting property have?
8. What controls are there over driveway location, frequency, etc.?
9. What other possible uses of the median area now exist or are anticipated?

Halsey (4) developed a summary of causes of traffic difficulties that lead to traffic accidents and congestion. Items of major importance in left-turn design include angles of movement, velocity differences, acceptable speeds, convergence, divergence (changing number of lanes), and capacity. These basic causes of traffic difficulties manifest themselves in four types of friction: intersectional, marginal, medial, and internal-stream friction. All four are frequently present in left-turn-lane operations.

Several studies (2, 5-7) present principles that are intended as guides to aid the traffic engineer in alleviating friction and minimizing the effects of basic causes of accidents and traffic congestion.

Guidelines for Use of Left-Turn Lanes

A list of warrants and guidelines derived from review of the literature has been developed for use in design of left-turn lanes (8). Included is a tabulation of the documented conditions under which left-turn lanes have been installed or programmed for installation. The following items provide a summary of these guidelines:

1. In general, warrants and guidelines for use with CTWLTMLs indicate average daily traffic (ADT) of 10 000-20 000 on facilities that have four through lanes and an ADT of 5000-12 000 on facilities that have two through lanes.
2. Warrants and guidelines for use with COWLTMLs usually indicate only that the ADT volume should exceed 10 000. Volumes at COWLTML sites in the literature ranged from 14 400 to 31 200 vehicles/day on facilities that have four through lanes.
3. Through-lane speeds of 48-80 km/h (30-50 mph) are common on CTWLTML sites.
4. COWLTMLs are commonly used on streets that have through-lane speeds greater than or equal to 48 km/h.
5. CTWLTML widths range from 3 to 4.6 m (10-15 ft).
6. Lane widths of 3.7 m (12 ft) are consistently recommended for COWLTMLs.
7. Land uses along CTWLTML sites are most commonly classified as commercial. Some sites are found in industrial areas that have commercial activity.
8. Land use was not found to be as important a consideration at COWLTML sites as it was at CTWLTML sites.

Sawhill and Neuzil (8) also provide a discussion of an opinion survey of city and state engineers in Texas. Questionnaires were mailed in October 1975 and January 1976 to the 25 district engineers of the State Department of Highways and Public Transportation (SDHPT) and to city engineers in 48 Texas cities ranging in population from approximately 18 000 to 1 233 000 (1970 census figures).

The engineers were asked to weight site characteristics in order of importance in determining the type and need for a left-turn lane and to rank CTWLTMLs, raised COWLTMLs, and flush COWLTMLs according to how well each satisfied certain site characteristics. Demand for midblock left turns was ranked as the most important site characteristic and was followed by (in order of average weight) peak through-traffic speed, number of through lanes, block spacing, pedestrian movements, public (driver's) preference, and abutting retailer's preference.

Although the respondents as a whole showed no distinct preference for left-turn lane type for many street and traffic characteristics, CTWLTMLs were preferred over COWLTMLs in areas of demand for midblock left turns, peak through-traffic volume, strip commercial land use, through-traffic speed of more than 48 km/h, four-through-lane facilities, long block spacings, driver's preference, and abutting retailer's preference. COWLTMLs were preferred over CTWLTMLs by the survey respondents in the areas of restricted sight distance and pedestrian movements. Flush COWLTMLs were usually ranked between CTWLTMLs and raised COWLTMLs. Other results of this survey are summarized below.

1. City engineers in Texas indicated that they desired maximum speed limits in CTWLTMLs to be less than the usual posted speed limits for arterial-street through

lanes, yet speed limits for CTWLTMLs are rarely posted separately.

2. Guidelines suggested for CTWLTML widths range from 3 to 4.6 m (10-15 ft). The survey also indicated that city engineers in Texas desire the CTWLTML width to increase as the through-lane speed increases.

3. Major effects that the survey respondents believed to be due to left-turn-lane installations include substantial (yet sometimes varied) effects on the number of accidents (especially those involving left-turning vehicles), capacity, delay, and travel time at the sites.

4. All engineers in Texas who responded to the survey had an average of about five years' experience with CTWLTMLs. City engineers had about three years' experience with COWLTMLs. District engineers had about six or more years' experience with COWLTMLs.

5. Engineers in Texas have a wide range of opinions on left-turn-lane design practices and conditions for use, but they generally feel that CTWLTMLs are more frequently misused than are COWLTMLs.

6. Approximately half of the district engineers responding to the survey and three-quarters of the city engineers responding use different signs and markings at major intersections than at midblock locations on CTWLTMLs. The most common difference was the transition of the CTWLTML to a COWLTML with inclusion of a gap in the marking for entering the lane.

Related Studies

Studies that are related to left-turn lanes range from studies of individual installations to projects that cover a wide range of improvements. These studies have provided a great deal of valuable information to aid in understanding the effects of left-turn installations; however, application of the findings of these studies to warrants is difficult because the relationships between accidents and site characteristics have not been fully determined. Previous studies related to left-turn lanes may be generally classified as before-and-after (or parallel) accident studies, operational studies (which may also be before-and-after studies), general access studies, and studies that use regression techniques. The summary of findings presented below draws primarily from the more extensive studies.

Operational Studies on CTWLTMLs

Studies on CTWLTMLs have been done by a variety of state and local agencies, but most were focused on accidents and only a few were related to traffic operational aspects. With respect to operational aspects of CTWLTMLs, two major studies were found. One was conducted by Sawhill and Neuzil of the University of Washington (8) and another was conducted by Nemeth of Ohio State University (7).

Sawhill and Neuzil (8) made their operational study in terms of (a) travel distance within a CTWLTML prior to a left-turn maneuver during rush and nonrush hours, (b) general observations and commentary on users' behavior related to CTWLTMLs, and (c) the use of vehicle turn-signal indicators prior to a left-turn maneuver. Their findings include the following observations:

1. Drivers decelerate or stop in the through lane before entering the CTWLTML.

2. Seventeen percent of the out-of-town drivers make their left turns from the through lane without making use of the CTWLTML.

3. Most drivers complete the left-turn entry maneuver into the left-turn lane within 12-15 m (40-50 ft) of beginning the intersection entry.

4. The average travel distance within a CTWLTML for the local driver is 61 m (200 ft) and for the out-of-town driver is 43 m (140 ft). This distance is longer during the rush hour than during the nonrush hour for the local driver, but it is relatively consistent for the out-of-town driver.

5. Automobiles entering the roadway from driveways make little use of the CTWLTML as an acceleration lane; however, truckers do make use of it for their left-turn movement.

6. Few drivers use the CTWLTML as a passing lane.

7. Approximately 80 percent of the drivers use their turn-signal indicators prior to a left turn into a driveway, but only 40 percent signal when entering the roadway from a driveway.

Sawhill and Neuzil also stated that additional research in signing is needed to familiarize the out-of-town drivers with the proper use of the CTWLTML. It was recommended that the width of the median lane be 3-4 m (10-13 ft).

Nemeth (7) initiated four before-and-after operational studies on CTWLTMLs in Ohio. Major study parameters were traffic conflicts, travel time, left- and right-turning volumes, and traffic volume on each lane. Traffic conflict, as defined by Nemeth, is "any instance in which a main-flow vehicle must either swerve or brake to avoid an accident." He further classified the conflicts into cross conflict, opposing conflict, rear-end conflict, and weaving.

Of the two sites studied by Nemeth in a before-and-after context, one site involved the conversion of a four-lane arterial into a three-lane roadway, and the other involved restriping a four-lane highway section into a five-lane section. The conclusion of the analysis of the first site was that the conversion resulted in increased travel times, increased weaving, and some reduction in total conflicts. In the second case, an increase in volumes was noted, with an insignificant change in travel speeds. Conflicts attributable to braking were noted to have decreased after some initial increase due to driver confusion about the pavement markings. Recommendations are presented in the form of relevant discussion on such topical areas as adjacent lane use, access conditions and requirements, traffic volume, speed limit, spacing of existing intersections, economic conditions, and safety.

Operational Studies on COWLTMLs

Rowan and Williams (9) performed a study on channelization by measuring the tension of drivers through a highway study section. The study was performed during the three stages of a channelization installation. The first stage had no channelization, and the final stage had a divisional island with a special approach-end treatment. The results were inconclusive, due to the small number of responses and the variability in drivers. Rowan also performed a speed study before and after the installation of divisional island channelization. Those results were also inconclusive.

Shaw and Michael (10) conducted a study to aid in the establishment of warrants for the implementation of left-turn lanes in Indiana. They collected delay and accident-rate data at 11 intersections and used multiple regression techniques to develop equations to predict suburban delay time, rural delay time, suburban accident rates, and rural accident rates in terms of several operational variables. Their final presentation was a cost-benefit analysis in which the cost was the construction cost and the benefits were the reductions in accidents and delay.

Another element considered to be an important left-turn operational characteristic is gap acceptance. Ring and Carstens (11) classified the gap characteristics into types that they termed gap, lag, critical gap, and critical lag. They coupled on-site investigations with arterial modeling in an effort to explain gap-acceptance phenomena surrounding left-turn maneuvers. They concluded that gap acceptance is dependent on following and opposing queue length and that left-turning vehicles adjust speed to minimize the need for complete stops. These behavioral aspects, although difficult to predict, were put in a multiple regression model to estimate the number of vehicles that were forced to stop and the magnitude of delays of the stopped vehicles. The final presentations of Ring and Carstens were two equations for estimating the cost-benefit ratio in which the cost was the construction cost and the benefit was the accident reduction and delay savings.

Another left-turn gap-acceptance study was conducted by Dart (12) at both channelized- and unchannelized-approach signalized intersections. He found that drivers rarely accepted a gap of less than 2 s or rejected a gap longer than 8 s and that there was no appreciable difference between channelized and unchannelized approaches.

Volume Warrants

Volume warrants for left-turn lanes are typically presented in graphical form and relate the percentage of left-turning traffic to other volumes. Ring and Carstens (11) developed a series of graphs for determining whether a left-turn lane is warranted at a rural intersection that also considers the posted speed, the annual accident-cost reduction, and the percentage of trucks. Glennon and others (1) presented a volume warrant chart for sections or intersections that requires the percentage of left turns, advancing volume, and opposing volume.

Accidents at Channelized Intersections

Accident studies related to left-turn lanes at intersections (or high-volume driveways) have found significant decreases in accident rates when one-way left-turn lanes were added. Wilson (13) presented a summary of before-and-after studies that compared channelized left-turn lanes at unsignalized intersections using raised bars, curbs, and paint for channelization. The data showed statistically significant reductions in accident rates for projects that used each type of channelization.

Foody and Richardson (14), in a comparison of intersections with and without left-turn lanes (LTLs), found a great deal of variability in accident rates. The table below shows the comparison of sites Foody and Richardson developed on a basis of signalization and the existence of a left turn lane, in terms of accidents per million vehicles per leg per year. Although significant differences ($p = 0.05$) were shown in comparing total accident rates and those for "all others" (both signalized and nonsignalized), the variability of left-turn accident rates caused the subset averages for the left-turn accident rates to show no statistical difference.

	Nonsignalized		Signalized	
Type of Accident	With LTL (N = 33)	Without LTL (N = 134)	With LTL (N = 61)	Without LTL (N = 135)
Left turn	0.12	1.20	0.37	0.65
All others	0.92	3.15	1.17	1.82
Total	1.04	4.35	1.54	2.47

Shaw and Michael (10) used multiple regression to evaluate delays and accidents at intersections. Equations were developed for estimation of delays and accidents at suburban intersections with left-turn-lane channelization that explained 69 percent of the variation in delay and 61 percent of the variation of accident rates by means of eight and seven variables, respectively. The most important variables in predicting the accident rates were related to ADT, the number of approach lanes, and the average speeds of nondelayed through vehicles.

Accident Experiences on Designated Sections

Glennon and others (1) evaluated numerous access techniques by using information available in the literature and estimating average values of accidents, running times, cost-benefit ratios, and other measures of effectiveness. Table 1 shows the general accident warrants for access control techniques developed for left-turn and total accident rates on routes or at points (1). Estimates of accident reduction were prepared for COWLTMLs and CTWLTMLs. For raised COWLTMLs, it was assumed that accidents would generally be reduced by 50 percent at intersections and major driveways and that at minor driveways all left-turn accidents would be eliminated and there would be a slight increase in right-turn accidents. For flush COWLTMLs, it was assumed that accidents would be reduced by 28 percent, and for CTWLTMLs by 35 percent.

Other references have already shown that there is a great deal of variability in reduction of accidents by channelized lanes. Table 2 shows that there is also a great variability in accident reductions as a result of CTWLTML installations. The variabilities in accident reductions, and their unaccountability, make applications of reductions to a specific proposed installation very difficult.

In summary, no quantitative information related to both COWLTMLs and CTWLTMLs was found in any single reference. Only subjective comments in regard to both types of left-turn lanes were found. Accident analysis for a particular type of left-turn lane was the common approach of the few studies on left-turn lanes. Operational characteristics were mentioned in only a few of those studies; the common study elements were delays and gap acceptance on COWLTMLs and conflicts and entrance distances on CTWLTMLs. Although the previous studies provided valuable information, a more definitive basis for relating accident numbers and rates to site conditions was needed.

METHODOLOGY

The technique selected for an accident or operational study depends primarily on the nature of the available data and the study objectives. In most research applications that deal with design features of roadways, the purpose of accident and operational analysis is to investigate relationships between these parameters and various site or roadway characteristics for a number of chosen cases in order that the effects of certain conditions can be estimated. Four common analysis techniques used in such studies are regression analysis, before-and-after studies, comparison and individual case studies, and performance-standard studies.

In developing guidelines for the use of left-turn lanes, many different basic sets of conditions must be examined. It is also desirable to investigate many different variables within these basic subsets. The before-and-after study approach was impractical in this study

Table 1. Warrants for access control techniques on routes or at points, based on annual number of driveway-related accidents.

Item	Left-Turn Accidents: Low ADT (<5000 vehicles/day)	Left-Turn Accidents: Medium ADT (5000-15 000 vehicles/day)	Left-Turn Accidents: High ADT (>15 000 vehicles/day)	Total Accidents: Low ADT (<5000 vehicles/day)	Total Accidents: Medium ADT (5000-15 000 vehicles/day)	Total Accidents: High ADT (>15 000 vehicles/day)
Level of development (driveways/km)						
Low (<48)	2.66	5.18	7.70	3.8	7.4	11.0
Medium (48-96)	7.91	15.47	23.03	11.3	22.1	32.9
High (>96)	10.50	20.58	30.66	15.0	29.4	43.8
Driveway ADT (vehicles/day)						
Low (<500)	0.18	0.31	0.43	0.26	0.44	0.62
Medium (500-1500)	0.44	0.77	1.05	0.63	1.10	1.50
High (>1500)	0.68	1.19	1.61	0.97	1.70	2.30

Note: 1 km = 0.6 mile.

Table 2. Results of before-and-after studies on CTWLTMLs.

Source	Sections	Total Length (km)	Through Lanes	Date Installed	Before Period (years)	After Period (years)	Change in Number of Accidents (%): Total	Left Turn	Rear End	Right Angle	Side-swipe	Other
Sawhill and	1	1.66	4	1958	4	4	-26	+140	-28			-30
Neuzil (8)	1	2.4	4	1961	3	1	-6	-29	-19			+16
Conradson and Al-Ashari (15)	4	10.6	4	1964-1969	1	1	-33	-45	-62	+14	-7	+6
Busbee (16)	1	2.7	4	1974	1	1	-38	-90				

Note: 1 km = 0.6 mile.

due to the limited availability of time. Before-and-after and comparative parallel studies have already been conducted in many areas and can help provide information on possible accident reductions. The performance-standard study approach is undesirable due to difficulties in establishing standards for comparison and to the large number of variables. Since we wished to study operational as well as accident relationships, two study approaches were taken: regression analysis for accidents and comparison and individual case study for operations.

The identification of important variables was undertaken in an extensive review of related literature, and consideration was given to how the data would be used. The literature expressed the data in many different forms and, in some cases, provided statistical parameters, such as means, standard deviations, significance, and levels, that aided in predicting the variability and relative importance of each variable. Transformations used in the studies also provided hints of possible transformations of data for the regression analysis.

Selection of data to be collected was based on the relative importance of the data and the degree of difficulty anticipated in collecting the data. Collection of data that would not generally be available or easily obtained by the traffic engineer was not considered practical. It was considered desirable to be able to separate accidents by location, type, severity, cause, etc., in order that accident characteristics might be more easily compared for different lane types and accident groupings. Site data were tabulated by block or subblock, in order that the sites could be examined at different levels of detail. Vehicle kilometers for the block were calculated and summed over the total length of each section when several blocks were combined. Several dummy variables were used in the analysis as simple tests of whether the existence of signals on the ends of the midblock sites, the existence of parking, or the existence of three-leg intersections could account for differences between sites. The total number of variables was 63.

Accident Study Data Analysis

Accident data for left-turn-lane sites were analyzed by using standard regression techniques. Purposes of the analysis were (a) to provide insight into the characteristics of the sites and accidents that were being used in the analysis and (b) to describe existing field applications of various left-turn lane types.

Equations Developed

Sections were formed by combining midblock and intersection data in a manner that provided as much homogeneity as possible for lane-type markings, parking, lane widths, etc., at each site. Features such as railroad tracks and highly skewed intersections were avoided. The sections averaged approximately 0.72 km (0.45 mile) in length; extremely long sections that remained homogeneous rarely occurred, and extremely short sections were avoided.

The sections were analyzed with and without the inclusion of intersection accidents. This enabled an examination of the effects of intersection accidents on the total number of accidents, thereby providing another means of comparing lane types with the evaluation of the variability of other factors with and without intersection accidents included. The inclusion of intersection accidents generally improved the predictability of equations concerning accidents and accident severity and lessened the predictive ability of the equations related to the critical accident rate and the average damage scale. [Full details of the 46 equations are available on request from the authors.]

Ten equations were developed by using individual midblock sites (short sections between two adjacent intersections), excluding all intersection accidents. Due to the poor predictability of accidents at midblock sites and the large numbers of variables entering the equations (up to 11), individual midblock sites were quickly dropped from the analysis. Separation of the midblock sites by lane type did little to improve the equations.

The sites were examined with combinations of lane types and with separation of the CTWLTML sections. COWLTML sections were too few in number for an adequate regression analysis. The predictive abilities of the equations generally improved slightly when the CTWLTML sections were considered by themselves, indicating that some differences probably exist between characteristics of the CTWLTML sites and those of the COWLTML sites.

Checks of Regression Assumptions

Plots of residuals versus dependent and independent variables were examined to identify inadequacies of the models and to provide clues for possible variable transformations that might improve the equations. The plots of residuals versus dependent variables for the single midblock sites exhibited linear residual patterns, with positive residuals on one end of the dependent variable range and negative residuals on the other. These patterns, which resulted from the large number of site variables that had zero values on the short sections and from a mixture of lane types, rendered the midblock-site equations inadequate for predictive purposes. Similar patterns were observed for the equations developed by using mixed lane types. Although the patterns were not as strong as in the case of the midblock sites, the equations would still be judged inadequate. These patterns illustrate further that there are differences between the CTWLTML sites and the COWLTML sites.

Residual patterns similar to those related to the midblock-site equations were also observed for equations predicting the severity index, critical rate, and average damage scale, for reasons similar to those previously discussed. For the section equations developed from the CTWLTML sites, the residual patterns were extremely slight or exhibited the normal absence of pattern. The equation that was chosen for predictive purposes on CTWLTML sections presented no residual problems.

Regression Analysis Results

Examination of the regression equations, residual plots, extreme cases, etc., revealed many important relationships between accident and site characteristics. The following is a summary of the most important findings of the regression analysis, with a concentration on CTWLTML equations.

Important Variables

In order to identify the variables that are of greatest importance in relation to accidents at the study sites, a maximum level of five independent variables per equation was set.

Dependent Variables

The best dependent variables for prediction of all types of accidents on CTWLTML sections appeared to be (in order of value) the number of accidents per mile, the number of accidents, and the number of accidents per million vehicle miles. [Customary units are retained in the names of the variables since customary units were used in developing the equations.] The left-turn accident variables followed the same pattern. The amounts of variability explained by the equations were generally higher for the CTWLTML sections when the intersection accidents were included.

The severity index and average damage scale were very unpredictable, as was expected. The equations for prediction of the severity index and average damage scale also were found inadequate, due to previously mentioned residual plot patterns. The critical accident rate was used as a dependent variable to aid in spotting unusual conditions. (The R^2 values, however, are somewhat misleading, since the critical rate was developed by using vehicle miles, the primary independent variable for predicting the critical rate.)

Independent Variables

The most consistently important independent variables were weekday ADT, number of signals (or number of signals per mile), number of driveways (or number of driveways per mile), and city size. Other important variables were vehicle miles of travel (per weekday), percentage of commercial land use, and the existence of curbside parking. The relationships indicated that independent variables expressed as rates are most appropriately associated with dependent variables that are also expressed as rates.

ADT has frequently been related to accident rates, since it is a measure of both exposure and congestion. Vehicle miles of travel is a measure of interaction between the ADT and the section length. The number of signals and number of driveways are logical entries since both are indirect measures of level of development and conflicting movements. It is also important to note that the number of signals on the site is important even when intersection accidents are not included. The inclusion of a signal variable illustrates the importance of signal effects on accidents that do not actually occur at intersections. The city-size variable may be a measure of the differences in traffic characteristics of the cities in which the sections were located.

As might be expected, percentage of land use classified as commercial appeared to influence accident numbers and rates. Commercial-land-use influences appeared to be more prevalent on the CTWLTML sections in the prediction of left-turn accidents, illustrating the importance of commercial land use in generating midblock left turns and the greater need for left-turn provisions in commercial areas. The high colinearity between percentage of commercial land use and number of driveways per mile (0.671) generally deterred both variables from entering the same equation.

It is also important to note the absence of other variables that were considered to be important in the literature. Lane widths were not shown to be of major importance in the analysis, which may be primarily due to the fact that the average—3.6 m (11.7 ft)—is adequate. Similarly, there is no evidence from the analysis that present speed limits are unsafe or that posted speed limits for CTWLTMLs significantly reduce accident numbers or rates.

Prediction of Accident Rates

CTWLTMLs

The best dependent variable for predicting accident rates on CTWLTML sections is the number of accidents per mile. This equation also provides logical independent variables that consistently demonstrate relationship to accidents. These independent variables are weekday ADT, number of signals per mile, number of driveways per mile, and city size. The equation developed is

$$\begin{aligned}\text{Number of accidents per mile} = {} & -43.5 + 0.002\,03(\text{ADT}) + 0.000\,175\,(\text{city population}) \\ & + 0.491\,(\text{number of driveways per mile}) \\ & + 9.20\,(\text{number of signals per mile}) \qquad (1)\end{aligned}$$

The standard error for the residuals is approximately 33 accidents/mile, the F_{reg} is approximately 34, and the value of R^2 is approximately 0.75.

Although the equation shows that the number of accidents per mile increases with each of the independent variables, Table 3 better illustrates the magnitude of the expected accident rates. The average observed accident-per-mile rate for the CTWLTML sites with intersection accidents included is 77.9 accidents/mile. The average rate for the values in Table 3 is 79.5 accidents/mile.

Non-CTWLTMLs

Although there were too few non-CTWLTML sites for a regression analysis, comparison of these sites with the CTWLTML sites can provide some insight into differences in the lane types. Table 4 presents a tabulation of COWLTML and reversible-lane-site accident rates in comparison with estimated accident rates for CTWLTML sites with the same characteristics. This expedient comparison shows a consistent overestimation of accident rates on raised COWLTML sites by the accident-rate equation developed for CTWLTML sites. The comparison also illustrates part of the reason why equations developed for all lane types in combination were not satisfactory.

Operational Study Data Collection

Five operational situations were selected to represent typical left-turn installations. These situations were (a) short blocks, (b) offset intersections, (c) offset driveways, (d) one-side left-turns only, and (e) other commonly used situations. Selection of sites for operational study involved reviewing locations in several cities and making an inventory of those sites that fitted selection criteria. These criteria were based on land use, type of left-turn facility, average daily traffic volume, posted speed limit, and type of delineation. Twenty sites were selected in Austin and Fort Worth, Texas. Nine of the 14 sites in Austin are CTWLTMLs; 4 of these are CTWLTMLs that have transitions from CTWLTMLs to either raised or flush COWLTMLs at the intersection. The five other Austin sites are either raised or flush COWLTMLs. The remaining six sites, in Fort Worth, have either an extreme width or a different delineation. A brief summary of the characteristics of the sites is shown in Table 5.

Various operational characteristics mentioned in the literature were considered in the data selection process. The data requirements adopted for this study were entrance distance, maneuvering distance, lateral placement, traffic volume, and conflicts.

Entrance distance is the distance from an intersection to where a vehicle enters the turn lane before making a left-turn maneuver. These data apply to CTWLTML facilities, since the COWLTML has specific openings provided for left-turn entry. The entrance distance for each car that entered each CTWLTML facility was recorded by two observers, who noted the distance from the stopping line of the intersection at which the left front wheel touched the CTWLTML line. Maneuvering distance is the distance required for the left-turning vehicle to fully enter the left-turn lane. The spot where the left front wheel touched the CTWLTML and the spot where the right rear wheel touched the CTWLTML were estimated by the same two observers. The distance between these spots is the maneuvering distance.

Lateral placement is the lateral position of the vehicle within the lane. Data were collected through the use of a movie camera set on the roadside as far as possible from the roadway (in order to minimize influence on the driver). One still photograph was also taken whenever a vehicle entered the median left-turn facility. Three reference markers were used; two outside markers located the outer edges and the third marker located the center of the left-turn lane.

A clipboard counter was used to record the combined total for the through-lane volume, left-turn volume, and opposing volume. These volume counts were made simultaneously with the distance data collection and used as relative descriptors of the site. Conflict data include any friction caused by vehicles turning left over the study section. Only the peak period was observed, since the higher volume would normally generate more conflicts.

Theoretically, five types of conflicts were identified as pertinent to the operation of CTWLTMLs: (a) head-on conflict, (b) conflict between a vehicle in the CTWLTML and a left-turning vehicle from a minor street as it enters the CTWLTML, (c) conflict between a vehicle in the CTWLTML and a vehicle that starts to enter the CTWLTML, (d) conflict between a left-turning vehicle from the through lane (not using the CTWLTML) and a straight-through vehicle, and (e) conflict between a vehicle in the CTWLTML and a left-turning vehicle from the through lane.

In a flush COWLTML, fewer types of conflicts are possible, since fewer choices are available to the drivers. These consist of the following: (a) conflict between a left-turning vehicle and a straight-through vehicle in the through lane, (b) conflict between a left-turning vehicle in the left-turn lane and a left-turning vehicle from the opposite direction, and (c) conflict between a left-turning vehicle and a straight-through vehicle in the opposite direction.

On a raised COWLTML, even fewer conflict types are possible, since conflicts with the opposite stream of traffic are eliminated. The only possible type of conflict is one between a left-turning vehicle and a through vehicle in the through lane.

Operational Study Data Analysis

Data were analyzed by means of variance techniques to ascertain the effects of different lane widths, different delineation systems, and different types of left-turn facilities. Results of the analyses provided some basic information on the proper width of the left-turn lane, the proper delineation system, and other related operational characteristics that can be used to develop criteria for the left-turn-lane design. Lateral placement of the vehicle in the left-turn median lane, as well as entering and maneuvering distances, was analyzed in three interrelated efforts. In the lateral placement study, the effects of lane widths, pavement markings, types of median turn lane, and location of the raised island were investigated. For the entrance distance, a study was made on (a) entrance distance during peak and off-peak periods, (b) entrance distance at midblock and intersection locations, (c) entrance-distance behavior for different types of pavement markings, and (d) entrance-distance behavior for different types of through lanes. The maneuvering-distance portion of the study was concerned with the same general locations and configurations as the entering-distance study.

Accident Analyses

1. Comparisons of general accident statistics for raised COWLTML sites and CTWLTML sites reveal similar patterns by hour of day, number of vehicles involved, and severity.
2. Raised COWLTML sites have a greater proportion

Table 3. Estimated accidents per lane on four-lane urban streets [average section length = 0.71 km (0.44 mile)].

Signals per Mile	Driveways per Mile[a]	ADT <15 000 (avg = 10 540) Population 50 000	250 000	400 000	ADT = 15 000-20 000 (avg = 17 500) Population 50 000	250 000	400 000	ADT >20 000 (avg = 24 500) Population 50 000	250 000	400 000
>3 (avg = 4.63)	>60	72.3	107.3	133.5	86.4	121.4	147.6	100.6	135.6	161.8
	40-60	53.9	88.9	115.1	68.0	103.0	129.2	82.2	117.2	143.4
	<40	40.4	75.4	101.6	54.5	89.5	115.7	68.7	103.7	129.9
1-3 (avg = 2.0)	>60	48.1	83.1	109.3	62.2	97.2	123.4	76.4	111.4	137.6
	40-60	29.7	64.7	90.9	43.8	78.8	105.0	58.0	93.0	119.2
	<40	16.2	51.2	77.4	30.8	65.3	91.5	44.5	79.5	105.7
0	>60	29.7	64.7	90.9	43.8	78.8	105.0	58.0	93.0	119.2
	40-60	11.3	46.3	72.5	25.4	60.4	86.6	39.6	74.6	100.8
	<40	0.0	32.8	59.0	11.9	46.9	73.1	26.1	61.1	87.3

[a]Average values used in developing the table are for >60: 87.7; for 40-60: 50; for <40: 22.7.

Table 4. Comparison of accident rates by lane type.

Lane Type	Number of Through Lanes	ADT	Population	Signals per Mile	Driveways per Mile	Actual Accidents per Mile	Estimated CTWLTML Accidents per Mile	Error[a] (Actual - CTWLTML)
Raised COWLTML	6	29 562	407 000	4.17	39.6	166.7	145.5	+21.2
	6	31 134		4.65	39.5	127.9	153.2	-25.3
	6	32 706		3.13	84.4	253.1	317.5	-64.4
	4	15 483		0.0	16.1	41.9	67.1	-25.2
	4	13 921		0.0	31.3	12.5	71.4	-58.9
	4	13 591		0.0	0.0	9.4	55.4	-46.0
	4	14 477		0.0	81.8	65.9	97.3	-31.4
	4	14 477		0.0	100.0	76.3	106.3	-30.0
	4	14 477		2.1	62.5	64.9	107.0	-42.4
	4	8 323	283 700	0.0	17.0	36.2	31.4	+4.8
	6	13 660		3.2	35.5	29.0	81.0	-52.0
Flush COWLTML	4	17 197	407 000	0.0	23.3	46.4	74.1	-27.6
Reversible	2	13 223	283 700	2.0	56.0	66.0	78.9	-12.9
	2	11 367		2.9	5.9	35.3	59.2	-23.9

[a]Average error = 29.7 (SD = 24.3); average error (raised) = -31.8 (SD = 26.1); average error (four-lane, raised) = -33.4 (SD = 20.3).

Table 5. Summary of selected sites for operational study.

Location	Type of Left-Turn Lane	ADT[a]	Speed Limit (km/h)	Delineation
Austin				
5th and Lamar	CTWLTML	31 110	56	Single line of white buttons; yellow square buttons at intersection approach.
6th and Lamar	CTWLTML	31 110	56	Single line of white buttons; yellow square buttons at intersection approach.
45th and Lamar	CTWLTML and raised COWLTML	25 780	64	Standard CTWLTML marking[b] at midblock; opening; raised island at approach.
45th and Guadalupe	CTWLTML and flush COWLTML	23 210	56	Standard CTWLTML with buttons; opening; yellow square buttons at approach.
Anderson and Burnet	CTWLTML	22 570	64	Standard CTWLTML with buttons; large round buttons at approach.
Denson and Airport	CTWLTML	19 060	72	Standard CTWLTML with buttons.
Barton Spring and Lamar	CTWLTML and raised COWLTML	29 940	64	Single line of white buttons; raised island at approach.
Riverside and Congress	CTWLTML and flush COWLTML	21 340	56	Standard CTWLTML at midblock; opening; yellow square buttons at approach; six lanes with parking on one side.
32rd and Red River	CTWLTML	12 240	48	Standard reversible lane[b] marking; two lanes; reversible lane during peak period.
45th and Lamar	Raised COWLTML	21 680	64	Standard COWLTML with raised island.
19th and Lamar	Raised COWLTML	25 790	56	Standard COWLTML with raised island on the right side.
45th and Guadalupe	Flush COWLTML	20 730	56	Standard COWLTML with buttons.
Congress and 19th	Flush COWLTML	25 040	48	Standard COWLTML.
26th and Guadalupe	COWLTML	26 980	56	Continuous one-way with buttons.
Fort Worth				
Cockell and Berry	CTWLTML	19 500	56	Single line with buttons; double line with buttons at intersection; six lanes.
Wichita and Mansfield	Raised COWLTML	14 500	64	Raised island; metallic buttons 30 cm in diameter on the other side.
Bigham and Camp Bowie	Raised COWLTML	28 700	56	Raised island; ceramic buttons 20 cm in diameter on the other side.
Guliford and Camp Bowie	Raised COWLTML	32 200	56	Raised island; ceramic buttons 20 cm in diameter on the other side.
University and West Settlement	Flush COWLTML	16 700	48	Ceramic buttons 20 cm in diameter on both sides.
East Vickery and South Main	Flush COWLTML	8 000	48	Metallic buttons 30 cm in diameter on both sides.

Note: 1 km = 0.6 mile; 1 cm = 0.4 in.
[a]Obtained from 1975 volume count furnished by the Texas Department of Highways and Public Transportation.
[b]See Manual on Uniform Traffic Control Devices (17).

of intersection and intersection-related accidents than CTWLTML sites—75 percent and 55 percent for raised COWLTML sites and CTWLTML sites, respectively. CTWLTML sites have a higher proportion of driveway and nonintersection accidents.

3. The most frequently noted factors contributing to accidents on CTWLTML and raised COWLTML sites are unsafe speed and failing to yield right-of-way. Together these factors accounted for 56 percent and 24 percent of the two-vehicle cases for CTWLTML sites and raised COWLTML sites, respectively. Following too closely is a contributing factor in 42 percent of the two-vehicle accidents for raised COWLTML sites, compared with 14 percent for CTWLTML sites. The analysis of factors contributing to accidents illustrates the effects of the greater freedom of movement possible with CTWLTMLs, which allow continuous access to abutting property.

4. Analysis of factors related to accidents on the study sites indicated that the percentage of cases involving driveway maneuvers on CTWLTML sites was twice that on raised COWLTML sites. CTWLTML sites had only small percentages of midblock accidents involving vehicles slowing or stopping to make left turns.

5. The best dependent variable for estimation purposes was found to be the number of accidents per mile.

6. Little success was found in predicting accident severities or damage measures.

7. The most consistently important independent variables for prediction of accidents and rates were weekday ADT, number of signals (or signals per mile), number of driveways (or driveways per mile), and city size. Secondary variables were vehicle miles, percentage of commercial land use, and the dummy variable for existence of parking.

8. Independent variables notably absent from the equations were those related to lane widths plus speed limits.

9. A "best" predictive equation was selected and a table was developed that illustrated the effects of the independent variables on the number of accidents per mile on CTWLTML sites.

Operational Analyses

In regard to the operational analyses, the following findings [more completely documented in Walton and others (18)] were developed.

Lateral Placement

1. In reference to CTWLTMLs, lane widths of 3.4 m (11 ft) and 3.7 m (12 ft) have no significant adverse effect on traffic operations, but lane widths of approximately 4.6 m (15 ft) or more created some confusion among drivers.

2. In reference to flush COWLTMLs, lane widths of 3.2 m (10 ft 6 in) to 3.8 m (12 ft 6 in) showed no significant operational variation.

3. Lane widths of 2.6 m (8 ft 6 in) to 3.2 m for COWLTMLs produced significant variations.

4. Standard CTWLTML markings and white single-line button markings were interpreted differently by drivers, and the use of paint or buttons for delineation showed some operational variation in terms of driver response and vehicle positioning.

5. Raised COWLTMLs with paint markings and flush COWLTMLs with 30-cm (12-in) diameter metallic buttons on both sides of the lane were comparable in terms of vehicle queueing in the lane.

6. There were significant differences between CTWLTMLs and flush COWLTMLs with 30-cm diameter metallic buttons on both sides of the lane.

7. In a raised COWLTML, drivers tend to position the vehicle away from the raised barrier.

Entrance Distance

1. Traffic volume, especially the left-turning and the adjacent through-lane traffic volume, has a significant effect on entrance distance.

2. Entrance distances to left turns at midblock and at intersection approaches are different.

3. The type of lane delineation has significant effects on entrance distance.

4. Entrance distance varies with the number of through lanes.

5. There is a wide range of entrance distances on CTWLTMLs. The majority of drivers observed entered the CTWLTML 45-75 m (150-250 ft) from the intersection, while very few drivers entered the lane less than 30 m (100 ft) from the intersection.

Maneuvering Distance

1. Although there is a range of maneuvering distances, a large number of observed drivers completed the left-turn entry in 15 m (50 ft).

2. Traffic volume and the number of through lanes were found to influence maneuvering distance.

3. Maneuvering distances are shorter at midblock than at intersection approaches.

CONCLUSIONS AND RECOMMENDATIONS

The study findings suggest a wide range of guidelines for consideration by highway designers and traffic engineers. The guidelines refer to urban arterials and are recommended for use in addition to standard traffic engineering practice. These guidelines should, however, provide a higher level of user confidence and a basis for comparing information gained from other sources.

CTWLTMLs are an effective and efficient means of providing an enhanced level of service on many urban arterials. They are especially effective in locations of strip commercial development and frequent driveway openings that experience moderate left-turn demand. Raised and flush COWLTMLs are effective at major intersections that experience high left-turn demand.

CTWLTML lane widths and posted speed limits of the urban arterial were found to be adequately accounted for in standard practice by highway designers and traffic engineers. In other words, a minimum of a 3.4-m (11-ft) lane, with a 3.7-m (12-ft) requirement desirable for CTWLTML facilities, is recommended. Any lane width over 4.6 m (15 ft) was found to create some driver confusion regardless of the speed of the through traffic or the legal speed limit. Therefore, the following provides a summary of recommended guidelines found in this study for left-turn median lanes.

1. Existing site conditions should be carefully inventoried and assessed when considering left-turn-lane improvements or installations. The findings of this or any other study should be considered only as guides, not warrants, for left-turn-lane improvements or installations.

2. The text table on page 00, along with Wilson (13), may be used for estimating improvements in accident rates due to left-turn channelization at individual intersections.

3. Table 1 should be used as a general guide for consideration of access control techniques.

4. Existing accident locations, contributing factors, and related factors should be used as guides in deter-

mining the potential effectiveness of left-turn lane types.

5. Table 3 and Equation 1 should be used as guides for determining the potential effectiveness of a CTWLTML.

In general, CTWLTMLs provide for increased flexibility, e.g., the inherent characteristic of additional storage space for short blocks. The fear of conflicts and a resultant increase in accidents after implementation is unfounded. In fact, most "anticipated" conflicts rarely occur; if they occur, they are handled with typical driver judgement. It was observed that the signing and pavement-marking procedures in the Manual on Uniform Traffic Control Devices (MUTCD) (17), sections 3B-12 and 2B-17 (as amended in Volumes 1-8), are effective in informing drivers of CTWLTML operations. We believe that signing contributes marginally to driver awareness and that pavement markings (lane delineation and symbol messages) are mandatory. Speed limits imposed on many CTWLTML locations serve little purpose because of the characteristic use of the facility.

In regard to raised or flush COWLTMLs, no significant driver-conflict problems were observed. Adequate storage space for the left-turning queue was the primary design element that created any concern.

In reference to raised lane markers (e.g., ceramic or metallic buttons), other minor observations of interest are that 1.3-cm (0.5-in) high square buttons and 7.6-cm (3-in) high, 20-cm (8-in) buttons installed at the intersection approach to separate opposing traffic were not observed to be very effective in prohibiting left turns from the opposing traffic and that 30-cm (12-in) metallic buttons are effective in separating through-lane traffic and left-turn-lane traffic. However, there are several disadvantages: (a) the buttons are difficult to maintain and clean, (b) they can create hazards to motorcyclists, and (c) they may force motorists who entered the left-turn lane by mistake to turn at the intersection. Few vehicles were observed returning to the through lane at the intersection and few vehicles entered the left-turn lane by crossing through the space between buttons.

The CTWLTML, as is appreciated by most practitioners, is an excellent option and is recommended for use where these guidelines suggest it as an effective alternative.

ACKNOWLEDGMENT

The contents of this report reflect our views, and we are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

REFERENCES

1. J. C. Glennon, J. J. Valenta, B. A. Thorson, and J. A. Azeh. Guidelines for the Control of Direct Access to Arterial Highways: Draft Volume 2—User's Manual. Midwest Research Institute, Kansas City, MO, 1975.
2. Technical Council Committee, Institute of Transportation Engineers. Continuous Left-Turn Lanes. Traffic Engineering, May 1964.
3. J. C. McMonagle. Relation of Traffic Signals to Intersection Accidents. HRB, Bull. 74, 1953, pp. 46-53.
4. M. Halsey. Traffic Accidents and Congestion. Wiley, New York, 1941.
5. An Operational Review. City of Fort Lauderdale, Rept. 71-04, July 1971.
6. N. C. Jouzy and H. L. Michael. Use and Design of Acceleration and Deceleration Lanes in Indiana. HRB, Highway Research Record 9, 1963, pp. 25-51.
7. Z. A. Nemeth. Development of Guidelines for the Application of Continuous Two-Way Left-Turn Median Lanes. Ohio State Univ., Columbus, Final Rept. EES 470, July 1976.
8. R. B. Sawhill and D. R. Neuzil. Accidents and Operational Characteristics on Arterial Streets with Two-Way Median Left-Turn Lanes. HRB, Highway Research Record 31, 1963, pp. 20-56.
9. N. J. Rowan and T. G. Williams. Channelization. Texas Transportation Institute, Texas A&M Univ., College Station, 1966.
10. R. B. Shaw and H. L. Michael. Evaluation of Delays and Accidents at Intersections to Warrant Construction of a Median Lane. HRB, Highway Research Record 257, 1968, pp. 17-33.
11. S. L. Ring and R. L. Carstens. Guidelines for the Inclusion of Left-Turn Lanes at Rural Highway Intersections. Engineering Research Institute, Iowa State Univ., Ames, 1970.
12. O. K. Dart, Jr. Left-Turn Characteristics at Signalized Intersections on Four-Lane Arterial Streets. HRB, Highway Research Record 230, 1968, pp. 45-59.
13. J. E. Wilson. Simple Types of Intersection Improvements. In Improved Street Utilization Through Traffic Engineering, HRB, Special Rept. 93, 1967, pp. 144-159.
14. T. J. Foody and W. C. Richardson. Evaluation of Left-Turn Lanes as a Traffic Control Device. Ohio Department of Transportation, Columbus, Nov. 1973.
15. B. A. Conradson and N. Al-Ashari. Evaluation of Four Safety Projects. Traffic and Safety Division, Michigan Department of State Highways, Lansing, 1972.
16. C. B. Busbee, Jr. Cost-Effectivensss of a Two-Way Left-Turn Lane. Institute of Transportation Engineers, Dec. 1975.
17. Manual on Uniform Traffic Control Devices for Streets and Highways. Federal Highway Administration, 1971.
18. C. M. Walton, T. W. Horne, and W. K. Fung. Design Criteria for Median Turn Lanes. Center for Highway Research, Univ. of Texas, Austin, March 1978.

Discussion

Stanley L. Ring, Department of Civil Engineering, Iowa State University, Ames

The authors have made a significant contribution toward a better understanding of the use of CTWLTMLs. Although this technique for improving traffic operations is more frequently used now than it was in the past, a surprisingly small amount of significant research has been conducted on the subject. The authors' review of the literature contains a number of studies done in the 1960s.

The development of a "best" regression equation to predict accident rates was a basic objective of the study. The authors document the final equation selected with suitable supportive information for the readers. It would appear, however, that a word of caution regarding the application of Equation 1 should be given. The uniqueness of the study sites and the operating characteristics of the traffic flow are related to the specific model presented. In view of this applicable environmental constraint, the reader should be cautioned in making applications to other environments.

Also, it should be noted that a relatively large negative intercept constraint exists in Equation 1. Again, a word of caution regarding the extension of the predicting range beyond the data would be in order.

The operational study phase provided some useful results. I was, however, somewhat concerned with the procedure for measuring vehicle entrance and maneuvering distance and lateral placement. Observations from the side of the road "as far as possible from the roadway (in order to minimize influence on the driver)" would appear to introduce considerable judgment decisions and estimating because of visual shortcomings and parallax. A number of photographic or video studies from elevated positions have been made in similar situations and have yielded more reliability.

It has been reported that, in certain locations, vehicles entering the roadway from an entrance tend to use the CTWLTML for storage or for merging with adjacent through traffic at a more convenient angle. In some cases, trucks especially benefit from these facilities. It would have been helpful if this aspect were observed and reported for the reader's total knowledge of benefits.

Other questions arose in regard to increased potential for U-turns, signal progression as an independent variable, and changes in total vehicle delay. Answers may be available in the detailed project report, or they may have been beyond the scope of this study. These questions may be related to the limitations of the manual observation technique.

The authors are to be commended overall for an excellent study.

John C. Glennon, Overland Park, Kansas

I would like to commend the authors on their paper. It offers a significant contribution to our knowledge about left-turn median treatments. I have no other direct comments about the paper, but I would like to recommend further operational studies. My observation is that several jurisdictions around the country are still reluctant to try the two-way left-turn median. Although some jurisdictions may be afraid of increased accidents, more are probably reluctant because of the lack of convincing evidence on significant benefits.

It seems clear that two-way left-turn medians would offer substantial delay reductions where traffic volumes are moderate to high in strip commercial areas. Yet no studies have clearly shown the capacity improvement value of two-way left-turn medians under various traffic-flow and commercial-development conditions. The studies I recommend, therefore, would be aimed at delineating the traffic operational benefits of two-way left-turn medians under various conditions.

Zoltan A. Nemeth, Department of Civil Engineering, Ohio State University, Columbus

Let me begin by congratulating the authors of this valuable contribution to the very limited literature on CTWLTMLs. With the exception of a few states and cities, this ingenious traffic engineering device is not being used to its full potential. In fact, it is not being used at all in many cities, although the date of the first installation goes back to at least 1950, when it was introduced in Michigan. There are various misconceptions underlying the resistance, and every bit of evidence regarding the safety and effectiveness of CTWLTMLs should be shared with the traffic engineering community.

Every traffic engineer recognizes that urban arterials usually perform two conflicting functions, namely the provision of access to abutting land uses and the provision of flow for through traffic. By permitting parking in the curb lane, for example, access is being favored, while the removal of curb parking increases the level of service for the traffic flow. CTWLTMLs can do both: They improve access to driveways and reduce delay to through traffic.

The main purpose of my comments is to support some of the findings of this paper. The basis for my comments is mostly the information gathered in two different surveys of traffic engineers around the country: an earlier one in connection with sponsored research (7) and a recent one in connection with the work of a committee I chair for the Institute of Transportation Engineers (this survey resulted in 106 responses from 29 states).

My comments will be directed at three areas of the subject paper.

1. Among the findings of the accident analysis is the information that "unsafe speed," "failing to yield the right-of-way," and "following too closely" are the major contributing factors in 70 percent of the accidents at CTWLTML sites. These factors are among the contributing factors commonly listed on standard accident reporting forms and, in general, are not very useful for the purposes of cause-and-effect accident studies. The more important conclusion is that "CTWLTML sites had only small percentages of midblock accidents involving vehicles slowing or stopping to make left turns."

I would like to add that, among the traffic engineers who responded to the above-mentioned surveys, only 11 percent indicated that accident problems were experienced at CTWLTMLs. Furthermore, all but one of the respondents who reported accident problems also indicated that some problems existed with improper use of the median lane. In contrast, only 50 percent of the total survey population reported problems with improper use. In other words, few CTWLTMLs have accident problems and those that do also have problems with improper use. As one of the respondents to the survey explained, "motorists sometimes stop in the median lane at an angle, with the rear of the car protruding into the through lane. This causes some rear-end accidents, which most often do not involve the left-turning vehicles themselves."

Also, sideswipe accidents occur when some drivers enter the CTWLTMLs too early and travel down the lane only to be struck by another driver entering the CTWLTML nearer to the left-turn point.

Sometimes right-angle collisions occur between vehicles entering the CTWLTML from the through lane and vehicles exiting from a driveway and making a left turn into or across the CTWLTML.

Generally, the responses indicate that the incidence of other types of accidents (especially head-on collisions) was very rare.

Some other improper uses reported in the surveys include turning improperly from the through lanes, passing slower vehicles for many blocks in the CTWLTML before turning left, truckers stopping for loading or unloading, bicyclists using it as a bike way. Improper uses that are due to unfamiliarity will, of course, diminish in time. At the beginning, however, education of the public is important, and the cooperation of the enforcement agencies is needed. (In an extreme case of noncooperation, one respondent reported that, in an early use, the police considered CTWLTMLs to be median divider islands and ticketed left-turning vehicles.)

I would like to add that CTWLTMLs can, in case of emergency, provide a path for emergency vehicles, a detour lane during blocking of through lanes by construction or vehicle breakdown, or even a place for storing snow removed from the through lanes.

2. The accident prediction equation (Equation 1) includes only four independent variables, and they are readily available. This should make it easy for others to test the equation and compare results. It was surprising at first to find that operating speeds were not included among the independent variables. A closer look at the study-site characteristics revealed, however, that speed limits ranged mostly from 48 to 64 km/h (30-40 mph) [only one site had a 72-km/h (45-mph) speed limit]. In this range, apparently, speed has no significant effect on accident statistics. It would have been interesting to see the effect of higher operating speed. Fifty-seven percent of the respondents to our survey suggested that the speed limit should be less than 88 km/h (55 mph) on arterials where CTWLTMLs are to be introduced. On the other hand, there are several examples of CTWLTMLs working properly even at 105 km/h (65 mph) (prior to imposition of the 88-km/h speed limit). One respondent stated that there is "no magic involved in the speed-limit sign. We have TWLTLs for miles on open rural unposted state highways...." It is probable, however, that, if the frequency of midblock left turns justifies CTWLTMLs, then the lower speed limit is also justified by the intensity of roadside development.

3. The paper states that the authors believe that signing contributes marginally to driver awareness but that pavement markings, including arrows, are essential. The MUTCD (17) requires both signs and pavement markings.

Our surveys found that 96 percent of the respondents complied with the MUTCD for pavement markings and 76 percent did for signing. At least one respondent even expressed concern that the expenses involved in the required signing may keep some agencies from installing CTWLTMLs. Some agencies reported that they use signs to comply with MUTCD, but they do not really think that they are necessary. Some 50 percent of the reporting agencies use overhead signs as well as ground-mounted signs. Some interesting comments were received from them regarding their policy on overhead signs: They use overhead signs where obliteration of pavement markings by snow can be expected, where curb parking or roadside development can detract from ground-mounted signs, where frequent improper use has been observed during use of pavement markings and ground-mounted signs, and on major multilane streets that have frequent signalized intersections. They reported that overhead signs are spaced at quarter to half points between major intersections, no less than 305 m (1000 ft) apart at other locations, and no less than 46 m (150 ft) away from major intersections to ensure adequate visibility for turning vehicles. These comments indicate that many agencies have a rational approach to the decision on overhead versus side-mounted signs.

MUTCD itself has gone through several changes regarding the subject of signing CTWLTMLs. I am not at all sure that I am aware of all the relevant changes, but let me attempt to summarize briefly my understanding of the evolution of the relevant sections in MUTCD: (a) Section 2 B-17 stated that signs "shall" be used, while Section 3 B-12 stated that signs "should" be used with pavement markings (1971); (b) Change M-24 (9/27/74) eliminated the contradiction by changing "should" to "shall" in Section 3 B-12; (c) Change Sn-156 (6/29/76) stated that "The R3-9 or R3-9a sign shall be mounted overhead and over the two-way left-turn lane when there are more than three lanes"; and (d) Change: Reconsideration of Ruling Sn-156 (9/19/77): "The post-mounted R3-9b sign may be used as an alternate to or a supplement to the overhead-mounted R3-9a sign."

In conclusion, let me state again that two-way left-turn lanes provide a good solution to the problem created by midblock left turns. Starting-up problems can be expected when they are first introduced in an area. However, they provide such an obviously needed service that drivers will soon get used to them and the level of improper use will drop to a minimum, as with other forms of traffic control.

Authors' Closure

We wish to express our appreciation to those who submitted discussions. These discussions emphasize that, in general, the vast majority of current experience with CTWLTMLs indicates that accident problems are not a primary deterrent to the use of these facilities. Excessive accident rates seem to be related to improper use of CTWLTMLs; the combination of appropriate control devices and driver experience should have a positive effect on this situation.

MUTCD (17) does require that signs, as well as pavement markings, be used along CTWLTMLs. The reconsideration (9/19/77) of Change Sn-156 allows a choice between post- and overhead-mounted signs. Opinions expressed by the vast majority of those contacted during this reserach indicate that driver response to pavement markings is clearly more positive than that for post-mounted or overhead-mounted signs. As noted in the discussions, areas where markings are often obliterated by snow have definite need for effective signage.

In summary, the CTWLTML is an effective and efficient treatment for midblock turn problems. Its controlled use is recommended for a substantial variety of conditions.

Publication of this paper sponsored by Committee on Operational Effects of Geometrics.

Abridgment

Benefit-Cost Analysis of Advance Treatment for No-Passing Zones

Donald L. Woods and Graeme D. Weaver, Texas Transportation Institute, Texas A&M University, College Station

It was the objective of a recent Federal Highway Administration (FHWA) research project (1) to develop improved criteria and guidelines for establishing passing and no-passing zones, regulatory traffic control devices, and traffic regulations, enforcement methods, and legal requirements that can be uniformly applied throughout the nation for the safety and benefit of all drivers. Alternative methods of establishing and designating passing and no-passing zones were developed to enhance current practice to supply the safety needs. The suggested system was selected because it exhibited the greatest feasibility on the basis of driver understanding, continuity of accepted type of demarcation (longitudinal line adjacent to roadway centerline), satisfaction of the driver's visual information needs, and practicality of installation and maintenance with existing mechanized striping equipment. It consisted of a dotted yellow line adjacent to the roadway centerline throughout the downstream end of the passing zone in conjunction with the standard NO PASSING ZONE pennant sign at the beginning of the no-passing zone's solid yellow barrier stripe.

The details of criteria development, driver understanding studies, and research methodology are included in the final report (1) and are not discussed here, but the salient research results are summarized below:

1. To avoid ambiguity in definitive terminology relating to the passing maneuver, the distance elements involved in the maneuver were defined as illustrated in Figure 1.
2. Minimum passing sight distance ($2/3d_2 + d_3 + d_4$) is defined as the sight distance at which a passing driver at the critical position (passing and passed vehicles abreast) must be able to perceive an opposing vehicle to permit safe completion of the maneuver.
3. A concept of advance marking treatment preceding a no-passing zone demarcation system with a NO PASSING ZONE pennant at the beginning of the solid yellow no-passing-zone line was suggested as a system of traffic control devices that would provide the needed driver information to designate that the passing zone will soon end and that a no-passing zone will be reached.
4. Establishment of a no-passing-zone system should be predicated on minimum passing sight distance for a specified operating speed. The advance marking treatment begins at the point at which sight distance first becomes less than the specified minimum passing sight distance and continues for a distance equal to the pass-completion distance for the specified operating speed; the no-passing zone begins at the downstream end of the advance marking treatment and continues until the minimum passing sight distance again becomes available.

PURPOSE OF THE ECONOMIC ANALYSIS

Traditionally, one of the primary considerations in evaluating a proposed change in concept or practice involves the question, What benefits can be expected from the change? Administrators in the highway profession are constantly forced to apportion already scarce funds to those areas in which they believe the greatest return will be achieved.

The concept of advance notification is a philosophy generally regarded among traffic engineers as being highly desirable in eliciting proper driver response and, therefore, as contributing noticeably to highway safety. The need for advance warning of a no-passing zone is evidenced by the rapidly spreading adoption of the NO PASSING ZONE pennant sign, the primary approved traffic control device for this purpose. In many cases, the administrative decision to adopt the concept of advance warning for the no-passing zone appears not to have been made on the basis of a detailed benefit-cost analysis but rather as the result of the conviction that driver conditioning to potential hazard is in itself beneficial and the cost of implementing such a system will be offset by improved operations and safety and reduced potential for litigation, although these elements may not have been quantified precisely.

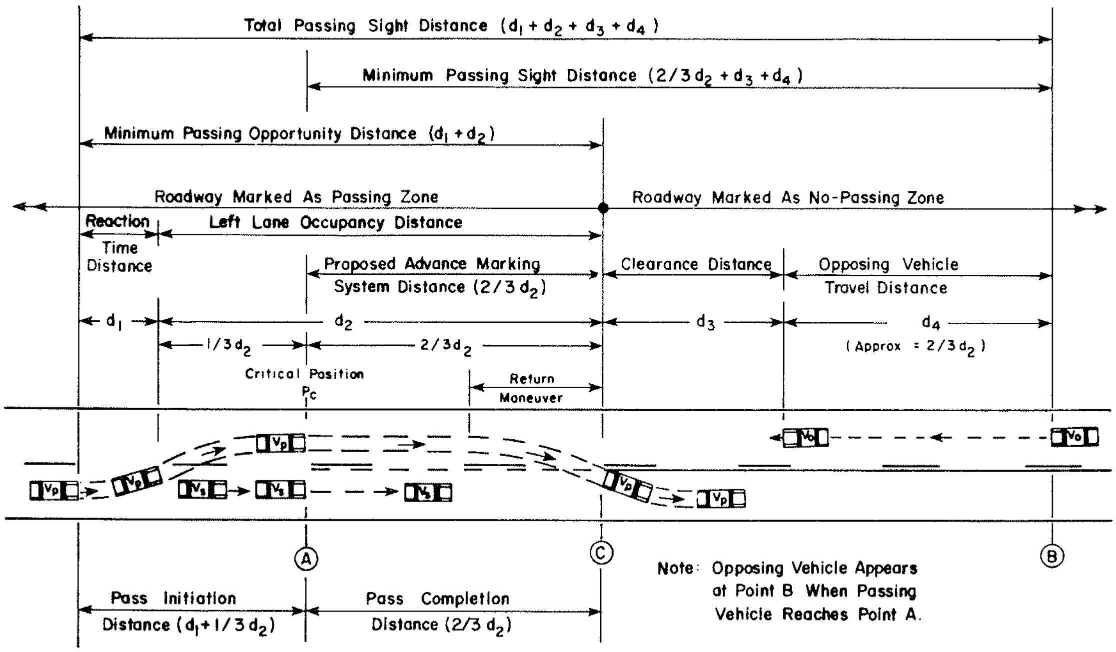
The concepts developed in the research regarding criteria, application, and design of the traffic control devices by which no-passing zones should be established and designated (1) were critically reviewed by 36 traffic engineers in various parts of the country and by a group of members of the National Advisory Committee on Uniform Traffic Control Devices and FHWA in a two-day workshop. In general, the concepts remained unchallenged; the primary concern expressed was that there was a definite need to evaluate the benefits that could reasonably be expected from implementation of the proposed treatments for no-passing zones. This would provide a framework for basing a decision to evaluate, in actual operation, a concept that theoretically appeared sound and reasonable. In response to this rather unanimous expression of opinion, an economic analysis was conducted to predict the expected benefit-cost ratio of application of the advance-warning no-passing-zone treatment proposed in the research effort. This abridged paper presents the results of the economic analysis.

General Approach

The intent during the economic analysis was to be conservative, so that the resulting benefit-cost ratio would represent a very conservative estimate of the relative value of the system. In all probability, the benefits would be substantially greater. To ensure conservatism, several assumptions were made: (a) relatively short sign life was assumed, (b) only those drivers who "clip" the no-passing zone (complete the pass beyond the start of the solid yellow line) were assumed to benefit from the advance treatment, and (c) an interest rate on the high side was assumed.

The approach adopted to compare expected benefits with expected costs on a nationwide basis included estimating the (a) costs of proposed no-passing-zone advance treatment nationwide, (b) number of no-passing zones nationwide, (c) number of passing maneuvers executed annually on two-lane highways nationwide, (d) number of passing maneuvers that involve "clipping", (e)

Figure 1. Distance elements and terminology to define passing design and operations.



number of accidents that involve sight-restricted passing maneuvers, (f) accident reduction due to application of advance treatment, (g) number of lives saved, (h) reduction in injury and property-damage-only (PDO) accidents, and (i) dollar savings of advance treatment, as well as (j) determining the expected benefit-cost ratio.

These tasks were accomplished by using a combination of state-supplied information, previous research regarding passing operations (2), accident statistics from several state studies of advance treatment and from national statistics, and field measurements on the passing maneuver by state agencies. National Highway Traffic Safety Administration (NHTSA) cost values were used in computing benefit-cost ratios.

Estimation of Cost Elements

Expansion Factor for Cost Data

A survey of the states regarding the cost of signing and marking operations was coordinated in 1974 by FHWA, and the results were made available for this research study. Since costs have risen appreciably, the 1974 costs were converted to equivalent 1978 costs by using three independent scaling approaches: the construction-cost-index ratio (3), the consumer-price-index ratio (4), and the bid-price trend on federally aided highway contracts (5). The three approaches indicated 25-35 percent increases in costs from 1974 to 1978. A midpoint value of 30 percent was selected, and the 1978 costs were estimated by applying a factor of 1.30 to the 1974 average costs.

Estimation of Proposed No-Passing-Zone Treatment Costs

The cost of the proposed no-passing-zone treatment is based on (a) sign life of seven years, (b) marking life of eight months, (c) interest rate of 10 percent, (d) marking cost of 8.2 cents/m (2.5 cents/ft), (e) advance treatment of 168 m (550 ft) [88.5 km/h (55 mph) operation], and (f) average pennant sign cost of $51 each, installed.

The marking cost based on 168 m of treatment of 4.6 m (15 linear ft)/12-m (40-ft) pattern, 8.2 cents/m, and an eight-month repainting schedule (1.5 times/year) is

Annual cost = 168 m × (4.6 m ÷ 12-m pattern) × $0.082/m × 12/8 = $7.69/zone

The cost of the advance dotted line actually would be less than this value, because it would be placed at the same time the centerline is placed.

The expected sign costs, by using the $51 pennant cost and a capital recovery factor of 0.205 (10 percent for seven years) is

Annual sign cost = $51.00 × 0.205 = $10.46/zone

The expected total annual cost per no-passing zone is $18.15, the sum of the marking and signing costs identified above.

Estimation of the Number of No-Passing Zones

The precise number of no-passing zones nationwide is not documented. In fact, this information was received from relatively few states in the FHWA survey. Therefore, it was necessary to estimate the number by using several sources of available data. Byington (6) reported that there are approximately 1.2 no-passing zones/km (2/mile) on two-lane highways in Virginia. This would translate to approximately 1 116 408 no-passing zones nationwide. The topography in Virginia suggests that this estimate is probably high for a nationwide estimate. By extrapolating from observation along 160 km (100 miles) of two-lane roadway in 28 states, an estimate of 835 000 no-passing zones was obtained.

The FHWA survey obtained data from four states. Based on an average of 9125 no-passing zones/state, a total national estimate would be 456 250. This figure represents the smallest probable value that could reasonably be expected.

The three basic esimating procedures produced a wide range of estimates. The average is about 800 000, which compares favorably with the estimate obtained by sampling 160 km in several states. For this reason, the estimate of 835 000 no-passing zones was selected as the expected value and used throughout the economic analysis.

Estimation of the Benefits of the Advance No-Passing-Zone Treatment

Estimation of the Annual Number of Passing Maneuvers on Two-Lane Highways

The 1974 Highway Statistics (7) presents a summary of the length of two-lane roadway in each of several average daily traffic (ADT) classifications. The Highway Capacity Manual (8) presents a relationship between the peak loadings and the frequency with which the loadings occur on the two-lane highway system.

By multiplying the average percentage of the ADT by the midpoint value of the ADT ranges, expected hourly volumes for each ADT range can be obtained. By multiplying the number of hours represented by the peak-period percentage by the length of roadway in each ADT classification, hourly kilometers at each ADT level can be generated. By using relationships between total hourly volume and passes per kilometer per hour (9, 10), the average number of passing maneuvers for each ADT group can be determined. The product of the hourly kilometers at each ADT level and the average number of passing maneuvers for that group summed over all ADT groups produces an estimated 7451.2×10^6 passing maneuvers/year.

Estimation of the Number of Passing Maneuvers Performed with No On-Coming Vehicle in Sight

Normann (11) determined that 59.7 percent of all passing maneuvers were performed in the absence of an opposing vehicle. The product of this percentage and the total number of passing maneuvers performed (7451.2×10^6) provides an estimate of the number of passing maneuvers executed when available sight distance was the limiting factor (P_{SR}).

$$P_{SR} = 0.597 \times 7451.2 \times 10^6 = 4448 \times 10^6$$

Estimation of the Number of Passing Maneuvers That Involve Clipping

A study by the Michigan State Highway Department (12) revealed that clipping (completing the pass beyond the start of the solid yellow line) occurred in 14-17 percent of the total number of passing maneuvers on a two-lane highway. Assuming that 15 percent of passing drivers would clip in the normal passing situation and in the absence of opposing traffic, the total number of passing maneuvers for which the advance treatment could be beneficial is estimated as

$$P_{CLIP} = 0.15 \times 4448 \times 10^6 = 667 \times 10^6$$

This number represents the annual number of passing maneuvers on two-lane highways that end beyond the start of the no-passing zone.

Estimation of Passing Accidents That Involve an Illegal Sight-Restricted Passing Maneuver

Data from the Federal Aid Fatal and Injury Accident Rate Study in 1974 (13) reveal that the fatality and injury accident rates for rural federal aid highways were 1.99 and 37/100 million vehicle-km (3.2 and 59.5/100 million vehicle miles), respectively. The number of fatalities and injury accidents eliminated can be estimated by multiplying these rates and the annual vehicle kilometers. The total vehicle kilometer of travel per day was estimated as 849.1×10^6 (527.7×10^6 vehicle miles), which represents 309 800 million vehicle-km (192 600 million vehicle miles) annually; therefore,

$$\text{Fatal accidents} = 1.99 \times 309\ 800 \times 10^{-2} = 6160$$
$$\text{Injury accidents} = 37 \times 309\ 800 \times 10^{-2} = 114\ 600$$

The 1974 edition of Accident Facts (14) indicates that 19 300 fatal accidents, 230 000 injury accidents, and 2 240 000 PDO accidents occurred on rural state roads in 1973. This resulted in 420 000 injuries and 23 300 fatalities. Expanding the injury accident figures by the ratio of PDO accidents to injury accidents, an estimate of the number of PDO accidents on federal aid highways can be obtained:

$$\text{PDO accidents} = (2\ 240\ 000 \div 230\ 000) \times 114\ 600 = 1\ 116\ 000$$

The distribution of two-lane-highway accident severity therefore becomes 6160 fatal accidents, 114 600 injury accidents, and 1 116 000 PDO accidents (total = 1 236 760 accidents).

A study conducted by the Franklin Institute Research Laboratories (9) concluded that approximately 10 percent of the accidents on two-lane roadways were related to passing. Therefore, an estimate of the total number of passing-related accidents is 10 percent of the values above:

$$A_{PASSING} = 0.10 \times 1\ 236\ 760 = 123\ 676$$

The passing-related accidents in which sight distance is the visual restriction can be estimated by the product of the total passing-related accidents ($A_{PASSING}$) and the ratio of the number of passing manuevers that involve clipping (P_{CLIP}) to the total number of passing maneuvers:

$$A_{CLIP} = 123\ 676 \times [(667 \times 10^6) \div (7451 \times 10^6)] = 11\ 071$$

Thus, an estimated 11 071 passing-related accidents occur annually in which the maneuver was initiated in the passing zone and involved an illegal sight-restricted completion.

Estimation of the Number of Accidents That Would Be Eliminated

The human-factors studies in this research indicated that approximately 74 percent of the subject drivers occasionally clip during the passing maneuver. Further, about 69 percent understood the meaning of the advance treatment without prior education. Thus, approximately 51 percent of the drivers (product of the two percentages) could be expected to respond correctly to the advance treatment system. The number of clipping accidents that could be expected to be eliminated through extensive use of the proposed advance no-passing-zone treatment can be estimated as

$A_{ELIM_{low}} = 0.51 \times 11\ 071 = 5646$ accidents/year

With minimal education, 94 percent correct understanding of the advance treatment was demonstrated (1). Therefore, the long-term effects could be estimated as

$A_{ELIM_{high}} = 0.94 \times 0.74 \times 11\ 071 = 7700$ accidents/year

Estimation of the Number of Lives Saved

The accident investigation (1) indicated that approximately 3.5 percent of all passing accidents involve a fatality. Applying this factor to the low and high estimates of accident reduction above produces an estimate of the number of fatalities that can be expected to be eliminated annually by application of the advance treatment:

Fatalities eliminated$_{low}$ = $0.035 \times 5646 = 198$
Fatalities eliminated$_{high}$ = $0.035 \times 7700 = 270$

Estimation of the Number of PDO and Injury Accidents Eliminated

PDO and injury accidents represent the difference between total accidents eliminated and the number of fatal accidents. Kemper and others (15) stated that 42 percent of the nonfatal passing accidents result in injury. The number of PDO and injury accidents eliminated can be estimated by applying this factor to the low and high estimates above:

Injury accidents eliminated$_{low}$ = $0.42 \times (5646\text{-}198) = 2288$
Injury accidents eliminated$_{high}$ = $0.42 \times (7700\text{-}270) = 3121$
PDO accidents eliminated$_{low}$ = $0.58 \times (5646\text{-}198) = 3160$

Estimation of the Potential Dollar Savings

NHTSA (16) estimated the total societal costs of automobile accidents at \$287 175 for fatalities, \$8085 for injuries, and \$520 for PDO accidents. By using these values, the expected annual savings for the United States as a result of using the advance no-passing-zone treatment can be estimated as

Annual savings$_{low}$ = $(198 \times 287\ 175) + (2288 \times 8085) + (3160 \times 520)$ = \$77 million
Annual savings$_{high}$ = $(270 \times 287\ 175) + (3121 \times 8085) + (4309 \times 520)$ = \$105 million

The annual dollar saving per no-passing zone in the nation is this saving divided by the total number of zones. As previously indicated, the low estimate of the number of no-passing zones in the United States is about 500 000; the high estimate is about 1 100 000; the most probable estimate is about 835 000. The estimated annual dollar savings per zone are shown below.

Low estimate = \$77 000 000 ÷ 1 100 000 = \$70/zone
High estimate = \$105 000 000 ÷ 500 000 = \$210/zone
Most probable estimate = \$91 850 000 ÷ 835 000 = \$110/zone

BENEFIT-COST COMPARISON

The annual cost per zone of the advance treatment is \$10.71 (low), \$22.11 (high), or \$18.15 (most probable). The expected benefit-cost ratio for the system is determined as the ratio of the savings to the annual cost of treatment. The following table presents these values and a summary of the economic analysis.

Item	Low Estimate	High Estimate	Most Probable Estimate
Annual cost per zone of advance treatment (\$)	10.71	22.11	18.15
Number of no-passing zones nationwide	456 250	1 116 400	835 000
Number of accidents eliminated by advance treatment	5646	7700	6673
Fatalities eliminated	198	270	234
Injury accidents eliminated	2288	3121	2704
PDO accidents eliminated	3160	4309	3734
Annual savings resulting from advance treatment (\$000 000s)	77	105	92
Annual savings per zone (\$)	70	210	110
Benefit-cost ratio of advance treatment	3.2	19.6	6.1

REFERENCES

1. G.D. Weaver and D.L. Woods. Passing and No-Passing Zones: Signs, Markings, and Warrants—Final Report. Texas Transportation Institute, Texas A&M Univ., College Station, TTI Rept. 3453-1, Sept. 1978 (rev. Jan. 1979).
2. G.D. Weaver and J.C. Glennon. Passing Performance Measurements Related to Sight-Distance Design. Texas Transportation Institute, Texas A&M Univ., College Station, TTI Rept. 134-6, July 1971.
3. Engineering Cost Index. Engineering News Record, March 28, 1978, pp. 30-31.
4. Consumer Cost Index. Wall Street Journal, April 14, 1978, p. 28.
5. W.B. Ledbetter and others. Techniques for Rehabilitating Pavements Without Overlays: A System Analysis. Federal Highway Administration, FHWA-RD-77-132, 1977, appendixes to vol. 2.
6. S.R. Byington. Economic and Location Criteria for a Passing-Aid System. Federal Highway Administration, May 1967.
7. 1974 Highway Statistics. Federal Highway Administration, U.S. Department of Transportation, 1974, p. 223, Table SM-110.
8. Highway Capacity Manual—1965. HRB, Special Rept. 87, 1965.
9. Franklin Institute Research Laboratories. Identification and Evaluation of Remedial Aid Systems for Passing Maneuvers on Two-Lane Rural Roads, Volume 2. Federal Highway Administration, Tech. Rept. 1-201, Feb. 1970.
10. Highway Capacity Manual—1950. U.S. Government Printing Office, 1950.
11. O.K. Normann. Progress in Study of Motor Vehicle Passing Practices. HRB, Proc., Vol. 19, 1939, pp. 205-217.
12. Study of No-Passing-Zone Signing. Michigan State Highway Department, May 1965.
13. Fatal and Injury Accident Rates on Federal-Aid and Other Highway Systems. Federal Highway Administration, U.S. Department of Transportation, 1974.
14. Accident Facts—1974. Statistics Division, National Safety Council, Chicago, 1974.
15. W. Kemper, P. Huntington, and S. Byington. Overtaking and Passing Vehicle Accidents. Public Roads Magazine, Vol. 37, No. 3, Dec. 1972, pp. 81-88.
16. B.M. Faigin. 1975 Societal Costs of Motor Vehicle Accidents. National Highway Traffic Safety Administration, Dec. 1976.

Publication of this paper sponsored by Committee on Operational Effects of Geometrics.

Abridgment

Shoulder Improvements on Two-Lane Roads

David I. Davis, Federal Highway Administration, Homewood, Illinois

An objective approach toward the development of geometric design standards has long been sought. Such an approach should link design variables to user benefits and impacts and should include economic considerations. While the relationship between shoulder design and highway safety has been extensively studied, an objective design approach has yet to be developed.

The approach outlined in this paper uses traffic volume as a measure of user exposure, traffic speed as a measure of user risk, and 1975 dollars as a common economic factor to measure costs, benefits, and eventually the net worth of shoulder improvements. The approach, as illustrated, is based on relationships between accident rate and shoulder width that are drawn from the literature (1). These relationships appear to remain consistent among various independent studies and are generally supported by the conclusions of numerous before-and-after evaluations of shoulder improvement projects conducted in various states during the last few years (2-6). The relationships shown in Figure 1 were developed by R. W. Sanderson in 1977 (7).

These relationships can be converted directly into a format that illustrates for given traffic volumes how many accidents can be expected to be prevented as shoulder widths increase (Figure 2). Plotted in this fashion, sample data show that the first 1 m (3 ft) of width over 0.6 m (2 ft) prevents the greatest number of accidents; additional width prevents diminishing numbers of accidents.

Each accident represents a societal cost; each prevented accident therefore represents a user benefit. Widely used societal cost figures (8) can be effectively linked to the relationships between travel speed and accident severity that were established by Solomon (9) and are shown in Figure 3. User benefits per prevented accident are estimated to be \$3257 at 48 km/h (30 mph) and \$5735 at 96 km/h (60 mph). Equivalent uniform annual shoulder-surfacing costs were calculated to be \$1042/km/m (\$512/mile/ft) of width, assuming 1975 cost of approximately \$5/m^2 (\$0.46/ft^2), 6 percent interest, and a 20-year service life.

For each annual average daily traffic (AADT) range shown in Figure 2, a relationship such as that depicted in Figure 4 can be constructed. This relationship depicts user benefits—for travel speeds of 48 km/h and 96 km/h—and improvement costs for shoulder widths of 0.6-3.0 m (2-10 ft). The algebraic difference between the benefits at a given travel speed and the corresponding cost for any chosen shoulder width can be plotted as the net worth of the shoulder improvement, as shown in Figure 5. The peak of the net-worth curve is reached at the shoulder width that provides the highest positive difference between benefits and costs—the optimum shoulder width.

If this procedure is repeated for each AADT range shown in Figure 2, the optimum shoulder widths for particular travel speeds may be plotted as they are shown in Figure 6. This figure indicates that for 48-km/h traffic the optimum shoulder width is 0.6 m for 0-1000 AADT, 1.9 m (6 ft) for 1000-5000 AADT, 2.4 m (8 ft) for 5000-6000 AADT, and 3.0 m for AADT over

Figure 1. Accident frequency for various traffic volume ranges versus shoulder width.

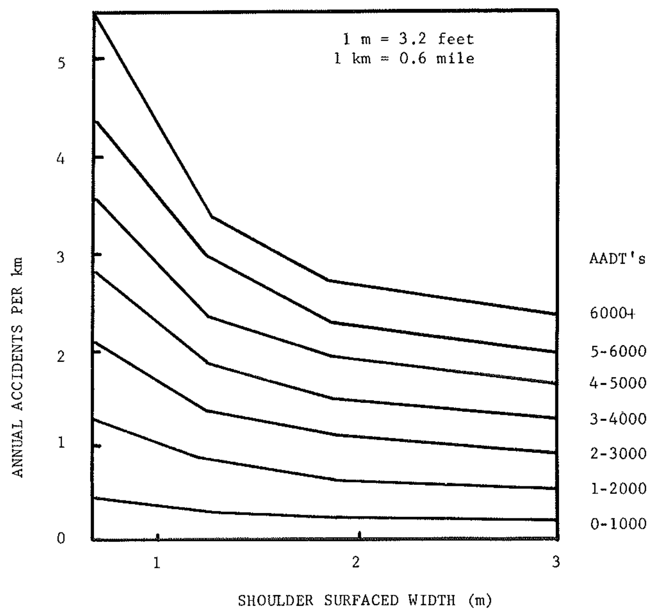


Figure 2. Expected reduction in accident frequency for various traffic volume ranges versus shoulder width.

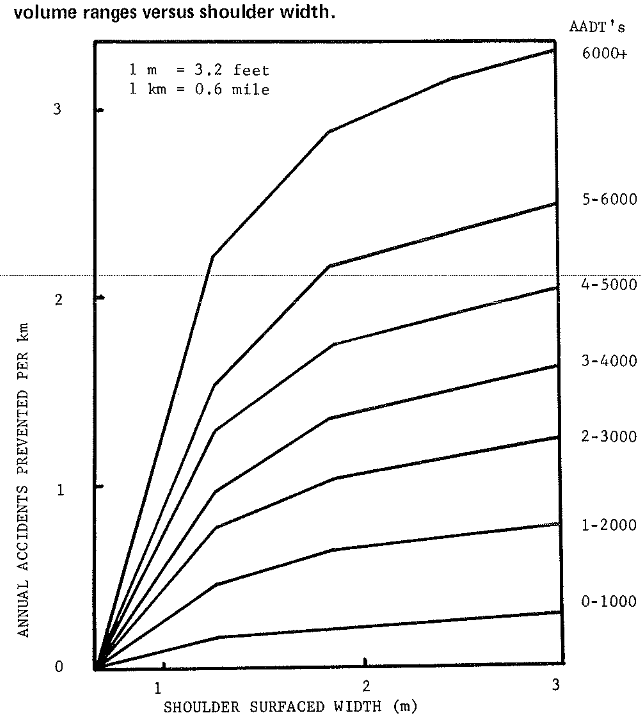


Figure 3. Accident severity versus travel speed.

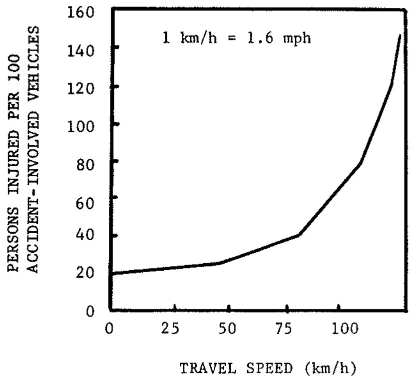


Figure 4. Annual safety benefits and equivalent uniform annual improvement costs for two travel speeds versus shoulder width (AADT 1000-2000).

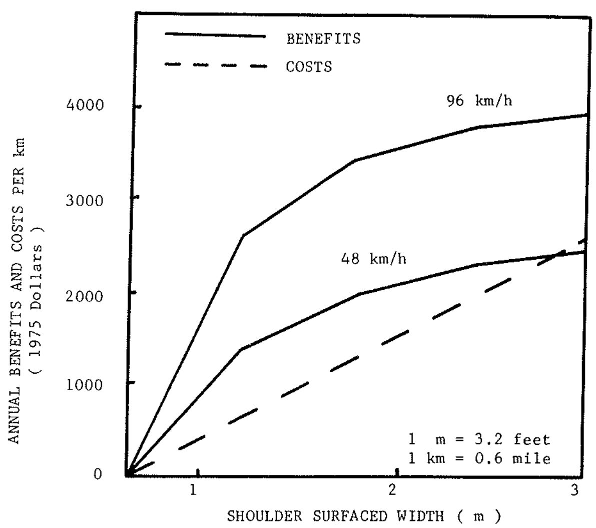


Figure 5. Net worth of shoulder widening per kilometer improved versus shoulder width (AADT 1000-2000).

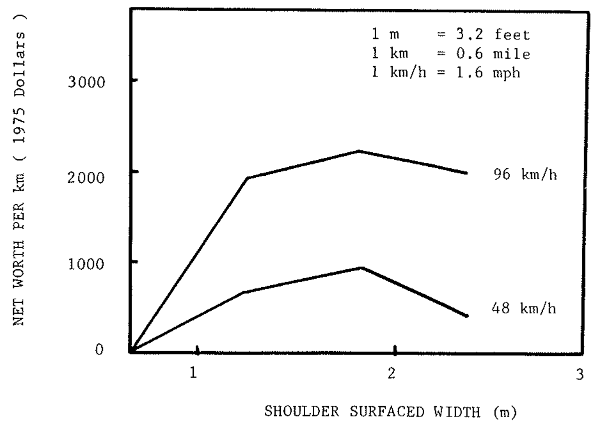
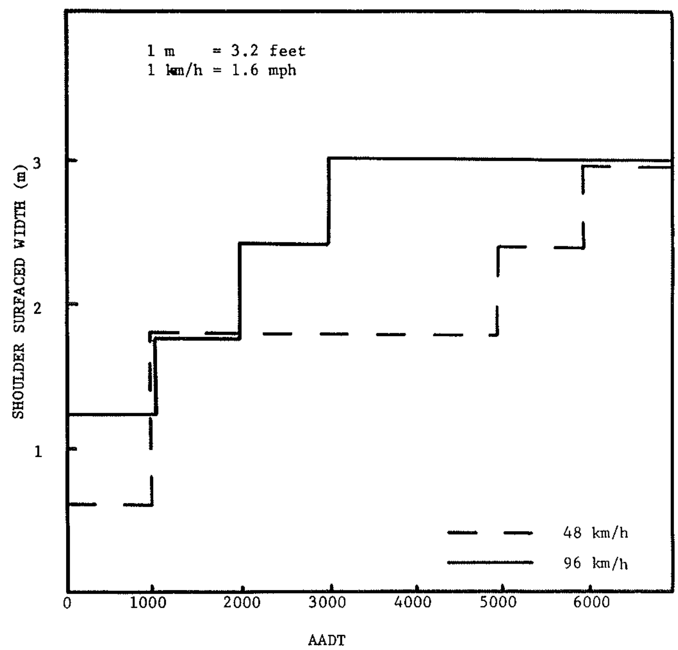


Figure 6. Optimum shoulder widths for various traffic volumes and travel speeds.



6000. For 96-km/h traffic, the optimum width is 3.0 m for highways that have AADT over 3000.

While the greatest need for improved data is generally to eliminate certain apparent assumptions in this approach and is particularly acute for roads that have low volumes (especially AADT under 400), it is thought that such objective techniques can now be used to guide the judgment of engineers who will continue to work toward the improvement of geometric standards.

REFERENCES

1. V. V. Silyanov. Comparison of the Pattern of Accident Rates on Roads of Different Countries. Traffic Engineering and Control, Jan. 1973.
2. C. L. Heimbach. Investigation of the Relative Cost-Effectiveness of Paved Shoulders on Various Types of Primary Highways in North Carolina for the Purpose of Establishing Priority Warrants. North Carolina State Department of Transportation and Highway Safety, Raleigh, 1972.
3. T. J. Foody and M. D. Long. The Identification of Relationships Between Safety and Roadway Obstructions. Ohio Department of Transportation, Columbus, 1974.
4. E. A. Rinde. Accident Rates versus Shoulder Widths. California Department of Transportation, Sacramento, 1977.
5. Evaluation of Shoulder Improvements. Arizona Department of Transportation, Phoenix, 1978.
6. Before-and-After Accident Study, Route 360, Chesterfield County. Virginia Department of Highways, Richmond, 1955.
7. R. W. Sanderson. Analysis of Accident Experience versus Shoulder Width. Transport Canada, Technical Memorandum No. TMRS 7701, April 1977.
8. B. Faigin. 1975 Societal Costs of Motor Vehicle Accidents. U.S. Department of Transportation, 1976.
9. D. Solomon. Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle. Bureau of Public Roads, U.S. Department of Commerce, 1964.

Publication of this paper sponsored by Committee on Operational Effects of Geometrics.

Modern Rotaries: A Transportation Systems Management Alternative

Kenneth Todd, Washington, D.C.

Although many rotaries, large and small, were constructed in the United States before World War II, they have received little attention in American literature in recent years. The present paper describes British developments in this field. In Britain, where rotaries (or "roundabouts") are a common form of intersection control, the capacity of large, conventional rotaries has been increased by 20-50 percent by reducing the size of the center island and enlarging the space available at the point of entry into the circulatory roadway. Rotaries need not be large to achieve high capacity. Small designs have been developed that generally have operating and safety characteristics superior to those of conventional alternatives, particularly when their geometry is devised to reduce approach speed and facilitate gap acceptance. At previously signalized intersections, capacity increases of 5-30 percent were obtained. Three-leg (T and Y) rotaries gave the best results for safety and capacity. Pedestrian and other fatal and injury accidents decreased substantially at sites converted from other forms of control, though an increase in property-damage-only incidents was observed at several locations. However, except at three-leg sites, fatal and injury accidents almost doubled after the center islands of conventional rotaries were reduced in size. Considerable savings in right-of-way acquisition and construction costs can be obtained on arterials, where the need for additional lanes to increase road capacity is greatly reduced by the use of rotaries. Rotaries can also reduce cost and land take when used as alternatives to, or in conjunction with, grade separation. The rotary concept can be employed in a grid of one-way streets, where blocks form rectangular center islands.

The cure for many traffic congestion problems is seen in the construction of wider roads. Yet it would often be more economical to increase the capacity of critical intersections rather than to widen a road in its entire length. The present constraints on government expenditure, the growing public resistance to large-scale highway construction, the energy crisis, and the problem of air pollution have given added impetus to the search for simple, inexpensive, and readily available techniques that relieve traffic congestion, increase street capacity, and make best use of existing facilities.

It was recognized many years ago that a rotary could be designed to have a capacity equal to that of the intersecting roads. In the United States, little work has been done in this field. The purpose of the present article is to acquaint the reader with developments in Britain, where small rotaries have been built with operational and safety characteristics that are superior to conventional alternatives.

A rotary has been described as a series of T-intersections at which a number of streets join a one-way circulatory roadway (1). For its successful operation, it is essential that entering drivers are advised by a right-of-way rule or a yield sign to give precedence to those within the rotary.

In the United States, rotaries have been unpopular because of the large area required and the low capacity of their weaving sections. The American Association of State Highway Officials (AASHO) used to recommend rotaries for complex intersections of five legs or more, as well as for intersections that would otherwise require multiphase signal control, but considered their capacity limited to about 3000 vehicles/h (2), though traffic circles in the United States were reported to handle 5000 and 7600 vehicles/h (3, 4). Fairly high design speeds and the concept that the length of a weaving section contributed materially to its capacity led to large layouts, where time gained by the higher speed is lost on longer travel distance. In practice, observed mean speed is about 24-28 km/h (15-17.5 mph) on urban rotaries (5, 6), so that the slow-down, as well as the extra distance traveled, contribute to what is called the "geometric delay".

However, small rotaries in the form of "dummy cops", "mushrooms", and small traffic circles were used extensively 50 and 60 years ago (7-9). The dummy cop was a trestle or stone column, often illuminated, that stood in the center of an intersection. The mushroom, also called a bumper or button, consisted of an iron disc about 45 cm (18 in) in diameter and 15 cm (6 in) high; some contained an electric light and others were fitted with reflectors. Small traffic circles were pioneered by early traffic engineers, notably William Phelps Eno (7, 8), George Guy Kelcey (9, 10), and Theodore M. Matson (11). These circles entered the literature under the heading of "channelized intersections", but in the United States they have not been researched and developed as intensively as they have been more recently in Britain.

MODERN ROTARIES IN BRITAIN

In Britain, rotaries are called "roundabouts" and are widely used for intersection control, particularly in rural areas, where signal control is rare. In an effort to overcome the disadvantages inherent in conventional, large rotaries—size, cost, low capacity, and extra travel distance—the Road Research Laboratory [now called the Transport and Road Research Laboratory (TRRL)], on the initiative of Frank C. Blackmore, made an intensive study of the design and application of smaller layouts specifically built for high capacity at low speed. Experiments conducted first on test tracks and later on public roads showed that a reduction in the size of the center island and an increase in the entry width could raise the capacity of large rotaries by 20 percent (12) and in some cases by 50 percent (13, 14). In terms of road-user costs, some conversions paid for themselves within 5-10 weeks (14). When very small center islands were installed, such layouts provided greater capacity than signals at sites previously considered too restricted for rotaries (15). In the first public road experiment, at Peterborough, Huntingdonshire, the replacement of signals by a "mini-roundabout" with a center island of 3 m (10 ft) in an inscribed circle of 29 m (95 ft) raised the saturation capacity of a T-intersection from 3700 to 4700 vehicles/h (16). A 20 percent capacity increase to 3300 vehicles/h was obtained at Longbenton (Figure 1) after signals were replaced by a center island of 2.5-m (8-ft) diameter in an inscribed circle that had been enlarged from 22 m (72 ft) to 24 m (79 ft) (15).

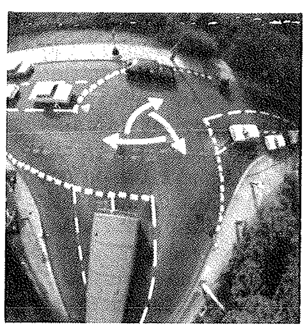
In its smallest form, the center island is made crossable by vehicles and, finally, reduced to zero, with the rotary principle indicated by circular markings in paint or thermoplastic. In this form, a rotary might be visualized as a three- or four-way yield to the left (in countries where traffic drives on the right, or to the right in those where traffic drives on the left), a technique commonly referred to as "offside priority".

A film by TRRL shows the operation of a rotary that has a zero center island in an inscribed diameter of 17 m (58 ft) at a previously signalized T-intersection in South Benfleet, Essex (Figure 2). Saturation capacity rose from 1800 to 2400 vehicles/h. Peak-hour queues were reduced from 2 km (1¼ miles) to 50 m (165 ft) (15).

Over the last 10 years, the use of mini-roundabouts with center islands of 4 m (13 ft) or less and other offside-priority rotaries that have center islands of up to about one-third of the inscribed diameter has spread widely in Great Britain. They are employed to increase the capacity of existing large, conventional rotaries or as control where the intersection of a major and minor road had become dangerous or caused excessive delays, as turn and speed control in residential areas, and for new intersections. Where they replaced signals, they increased capacity by 5-30 percent and significantly reduced peak and off-peak delays, as well as the number of stops. A variety of at-grade rotary layouts is shown in Figure 3a-f.

Figure 1. Mini-roundabout at Longbenton, Newcastle-upon-Tyne.

Figure 2. Fish-eye view of a zero roundabout, Benfleet, Essex.



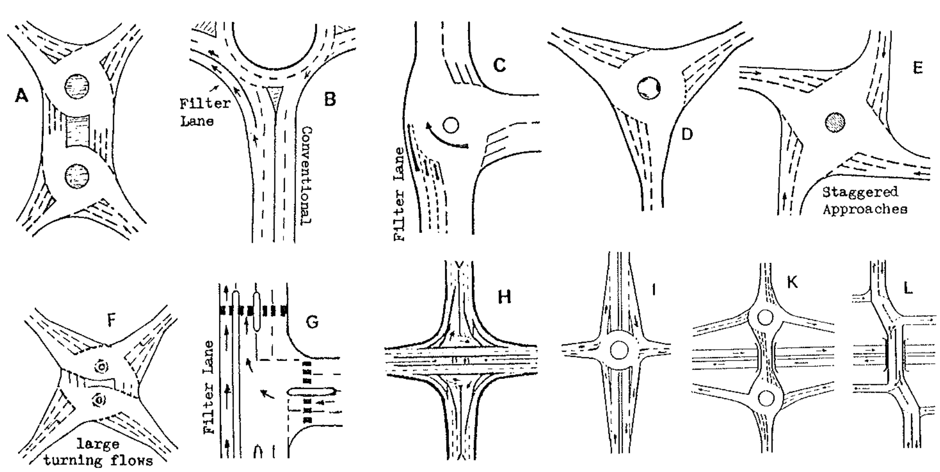
The operation of three-leg rotaries has been found to be safer, easier, and more reliable than that of others. The Y-intersection is the most satisfactory layout; it usually gives good visibility and a natural slow-down effect, which recommends it for new and redesigned sites (17). At some four-leg sites it has been found an advantage in capacity, safety, or convenience to change the intersection into a pair of adjacent three-leg rotaries (18) (Figure 3a). This feature can obviate the reconstruction of two offset "dog-leg" T-intersections into a single straight-through layout, a solution often proposed for signalized intersection improvements.

Rotaries eliminate the need for left-turn prohibitions (where traffic drives on the right), and they require no median to accommodate left-turn lanes. Unlike signals, they are immune to equipment failure and need little maintenance and no operating expenditures. They have a further advantage that is important to the resident, the taxpayer, and the environmentalist: they greatly reduce the need for highway widening. To raise capacity and accommodate the movement of high-volume platoons, signalized arterials are often widened to six or eight lanes. By contrast, the capacity of an offside-priority rotary is not affected by the width of the approach roads (19) but depends on the number of vehicles able to accept each suitable gap. In principle, if space permits, a rotary's capacity can be increased until it reaches or exceeds that of the approach roads, a feature recognized in the United States in the 1930s (20). It is usually less expensive and less controversial to enlarge a congested intersection than to build additional lanes where right-of-way acquisition would affect numerous adjoining properties (21).

From this it is apparent that the conversion of an existing six- or eight-lane signalized arterial to rotary control would, without detriment to total traffic flow, free the curb lanes for the exclusive use of buses, carpools, bicycles, or parking, provided all traffic may use the entire intersection area and its immediate approaches and exits (22).

In the United States, the widespread use of larger cars is likely to result in capacity at all types of intersections about 20 percent lower than that in Britain, where rotaries have been built to handle 5000-6000 or more vehicles an hour. Marble Arch, in London,

Figure 3. Various intersection and interchange layouts (traffic drives on left; not to scale).



though not designed as a high-capacity rotary but as a one-way system with a 16-m (52-ft) roadway, carries 8600 vehicles/h in the peak hour. Hyde Park Corner, a one-way system with an 18-m (56-ft) roadway, handles more than 10 000 vehicles/h in the peak hour, as well as 2000 through an underpass.

Scott Circle, in Washington, D.C., which had a circulatory roadway of 22 m (72 ft) prior to its reconstruction, carried 7600 vehicles/h (4) before it was fitted with signals in 1926.

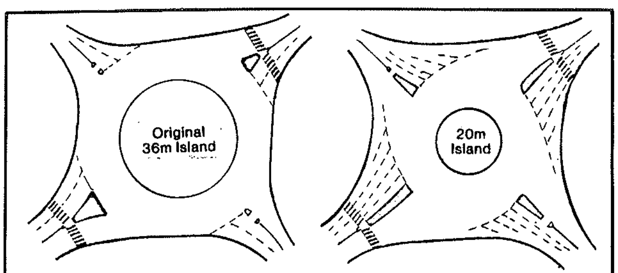
In Paris, the Place de l'Etoile, where 12 streets converge on a roadway 38 m (125 ft) wide, was reported in 1956 to handle close to 20 000 vehicles/h (23), in spite of a rule that gives right-of-way to entering drivers and thereby prevents exploitation of a rotary's full potential.

IMPROVING ROTARY PERFORMANCE

Principles of modern offside-priority rotary design are shown in Figure 4 [adapted from Blackmore and Marlow (24)]. Since a rotary operates on gap acacceptance, measures that facilitate the gap-acceptance task are beneficial in reducing delay and increasing capacity.

A higher rate of vehicle discharge is obtained when an approach is flared into a wider entry, with the yield line brought forward into the inscribed circle. By adding more entry lanes, output from one approach can be raised independently of the output from others and

Figure 4. Design principles for offside-priority rotaries (traffic drives on left).



Design Principles for Offside Priority Roundabouts:
1. Smaller center island to allow greater circulation width and entry width.
2. Yield line moved forward to increase entry width.
3. Greater number of narrower entry lanes in a given width.
4. Yield line left open near curb to encourage filtering.
5. Curb adjustment and narrower deflection islands to allow larger entry flare.

the effect of an unbalanced flow can be countered. Conversion from a conventional one to this type of rotary brings a fundamental change in operational characteristics. As can be seen in Figure 4, no proper weaving sections remain, and traffic no longer weaves. Depending on a driver's initial position at the yield line, the path selected and the size of the rotary, the vehicle's trajectory varies from an intersecting movement at right angles to merging at an acute angle.

Where drivers are reluctant to spread out into the wider entrances provided, they should be encouraged to do so by means of education (24), by a more gradual flare (25), and through initial police supervision of a newly installed rotary.

On entering a rotary, a driver should select a path, proceed along it at an even pace, and not cut in front of another vehicle but drop behind it. Correct use of turn signals, switched on shortly before the exit, helps others to make efficient use of gaps. Bicycle riders are served best if they keep to the outer curb lane and treat the rotary as a succession of T-intersections, at each outlet signaling their intention to turn off or continue ahead.

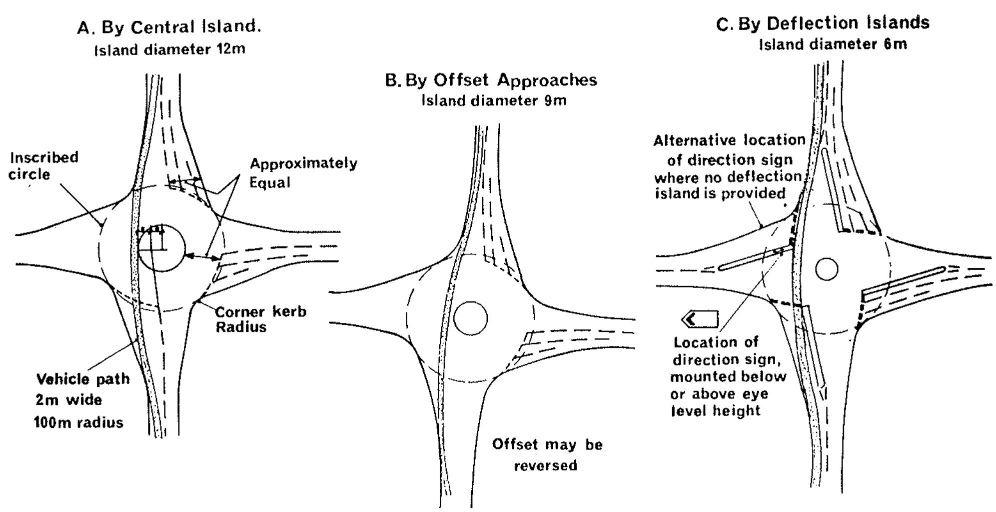
The slow-down necessary at the approaches to a rotary can be assisted by a deflection of the entry lanes (Figure 5) (17) or by a staggered road alignment (Figure 3e). These methods help to eliminate the visual impression that a fast, continuous road lies ahead. They give a gyratory effect with an acute angle of entry and thereby allow drivers to adjust their speed to that of the circulating stream. Entry gaps as short as 3 s have been found acceptable, and even 2 s in ideal conditions (1, 26). Deflection islands should be narrow, so that they do not restrict entry width, yet wide enough [about 1.2 m (4 ft)] to provide a refuge for pedestrians.

Advance warning signs, illuminated yield signs, flashing beacons, chevron boards, and other devices should be employed where necessary to counter excessive approach speed.

"Filter" lanes, provided to give unimpeded movement to drivers who are taking the first exit, have added considerably to the output from an individual approach road (24, 27) (Figure 3b). A yield line kept open ended near the curb has given a similar effect (24, 25) (Figures 3c and 4). At conventional T-intersections, through vehicles can likewise be separated from turning traffic if they are channeled into a filter lane that is indicated by a curb, barrier, or pavement markings (22, 28) (Figure 3g).

The maneuver of slowing down for a rotary and

Figure 5. Alternative methods of providing vehicle-path deflection (diameter of inscribed circle = 32 m; traffic drives on left).



accelerating back to normal travel speed imposes a certain delay that, depending on the speed differential, amounts to about 2-10 s. Since it occurs all day, this delay should be taken into account in cost-effectiveness studies, as well as the extra time a driver takes to negotiate all but the smallest rotaries.

PEDESTRIANS

Rotary control appears to bring considerable benefits to both pedestrian and vehicle safety. This can be attributed to a smooth flow and to comparatively simple decisions made at low speed. Crosswalks, placed two to three car lengths away from the yield line across the throat of the entry flare and divided by a center refuge (Figure 4), allow the driver to deal with pedestrian and vehicular conflicts in separate stages. Likewise, the refuge allows pedestrians to deal separately with vehicles from opposite directions. Where necessary, additional "stepping-stone" refuges can help pedestrians to cross one traffic lane at a time. At high-volume locations, guardrails have been provided. Pedestrians at rotaries generally have to walk a longer distance, but their average delay has been found to be less than that at signals (19). Antiskid surfaces are often used at crosswalk approaches, as are zigzag pavement markings, to indicate a parking and passing prohibition (29).

SAFETY

U.S. literature has mentioned the comparatively good safety record of rotaries; accidents are usually of a minor nature and involve property damage only (PDO) (2, 5). Probably the first before-and-after accident study at a rotary was conducted in Los Angeles in 1922. The installation of a traffic circle with a 37-m (120-ft) inscribed diameter and a center island of 12 m (40 ft) at Western Avenue and Wilshire Boulevard, reportedly capable of handling 5000 vehicles/h, reduced accidents by half and the cost of property damage by 75 percent (3). Published studies in the United States are rare, but those in Britain indicate that rotaries have a better safety record than major-minor road intersections or signals and that safety increases with the size of the center island.

Manning considered that rotaries tend to have at least 25 percent fewer accidents than signals (30). Garwood and Tanner found a reduction of 83 percent in fatal and serious injury accidents at three sites where rotaries had replaced signals and an overall reduction in injury accidents of 56 percent (31). A 10-year study by H. C. Smith compared 12 signalized intersections and 12 rotaries that had similar traffic volumes and found the rate of injury accidents at signals twice as high and the rate of pedestrian injuries three times as high (32).

A study by TRRL recorded personal injury accidents before and after installation of small rotaries at 150 locations. Since the number of accidents at any one site was small and the before and after periods were not always of the same duration for any site, the analysis was performed according to a method devised by Tanner (33) that uses the British national average for a control. Site conversions varied from simple, low-cost markings to expensive changes in design, so that the result could only be a measure of the composite effect averaged over the various sites. British national accident statistics include fatal and injury accidents only; they do not list PDO collisions, which do not have to be reported to the police.

In 48- and 64-km/h (30- and 40-mph) speed-limit zones, 88 sites previously operating as major-minor road intersections showed a decrease of 34 percent in slight injury accidents and of 46 percent in serious injury and fatal accidents; three-leg sites with raised center islands performed best. At 13 sites previously signalized, serious injury and fatal accidents were reduced by 62 percent and all injuries by 31 percent. However, when the center islands of 31 conventional rotaries were reduced in size, the reverse trend occurred except at three-leg sites, and the accident total was almost doubled, thus testifying to the inherent safety of conventional layouts. In higher-speed-limit zones, the study included no signalized intersections but otherwise endorsed the above results (34).

In Newcastle-upon-Tyne, a city that pioneered mini-roundabouts, pedestrian accidents were halved, according to a study that covered 22 years of mini-roundabout operation (35), even though traffic at all sites increased.

A before-and-after study at 38 locations where rotaries had replaced major-minor road intersections was conducted in the greater London area. The total number of injury accidents fell by 39 percent, fatal plus serious injury accidents by 64 percent, wet-road accidents by 51 percent, overall pedestrian accidents by 46 percent, fatal plus serious pedestrian injury accidents by 70 percent, and accidents involving two-wheeled vehicles by 18 percent. There was no change in the total number of rear-end collisions. Again, the larger rotaries performed best—overall accident reduction of 52 percent. At sites with curbless center islands, total accident reduction was only 23 percent; these sites showed an increase of 60 percent in rear-end collisions and of 7 percent in accidents involving two-wheeled vehicles (36).

Despite the general reduction in injury accidents and their severity, more sideswipes and other PDO incidents have been observed at several sites (14, 19, 37); this is attributed to a higher level of activity at high-capacity rotaries. To allow a true assessment of the effect of a change in traffic control devices, it would be useful if studies included every incident, the traffic volumes, and accidents within the neighboring roadnet. For example, an increase in accidents at an intersection that handles more traffic after the changeover may be accompanied by less congestion and fewer collisions on the approaches and on residential streets through which commuters had previously made a detour. Unfortunately, today's reporting and record-compiling methods rarely allow such full analysis.

THE ROTARY AND THE RESIDENT

Citizens often clamor for stop signs to discourage the use of residential streets by commuters and to control speeding. The Manual on Uniform Traffic Control Devices (38) advises that stop signs should not be used for speed control. The signs have been found ineffective and even counterproductive for that purpose (39).

Small rotaries have been installed to control speeding in residential areas in Seattle (40) and San Francisco (41). They are also used for that purpose in some of Britain's more than 30 New Towns, which rely on limited access and grade separation for their traffic control within the primary road system and, at the more important at-grade intersections within the secondary system, almost exclusively on rotaries. Rotaries are said to have a more attractive appearance and to create a more pleasant environment than signals, with their profusion of lights and the disconcerting effect of stop-start movement and the noisy revving of engines (42). Because delays are shorter than at signals, motorists have less incentive to filter through residential side

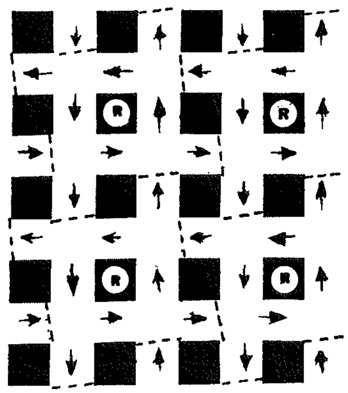
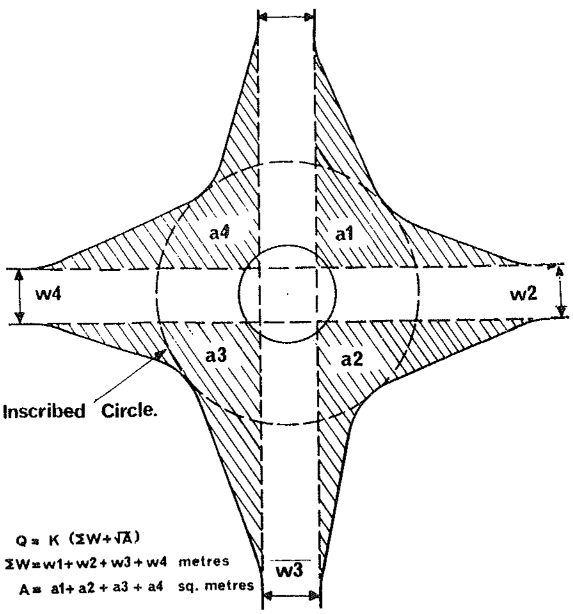


Figure 6. Rotaries in a one-way grid (traffic drives on right).

Figure 7. Calculation of capacity.



streets (32). Flowers, trees, and greenery can be planted on the center island. As evidenced in Washington, D.C., by Scott Circle, Grant Circle, and many others, they can embellish the civic scene by serving as convenient emplacements for the graven images of warriors and politicians.

ONE-WAY GRIDS

The engineer in search of the least restrictive form of traffic control for a system of one-way streets may wish to follow in the footsteps of Pratt (43) and Malcher (44), who recognized in the late 1920s that a grid of one-way streets forms a series of rotaries. The AASHO Blue Book (2) likewise states that parts of an existing street system may be designated as the one-way road of a rotary. In a one-way grid, every second block forms the center island of a rectangular rotary, with the surrounding streets representing the circulatory roadway (Figure 6). Yield signs can be installed in such a manner that, when traveling along the entire length of a street, a driver yields to traffic on the left at every second intersection; at the intervening intersection those on his or her right yield. Great flexibility can be obtained in such a system when the right-of-way on a street is reversed to permit entry into a heavy traffic stream from a street carrying very low volumes, or when the direction of a one-way street is reversed to meet different capacity requirements. Since a block also represents a median, pedestrians can move within a one-way grid in the same way as at other rotaries, with additional intermediate refuges acting as an aid to crossing high-volume vehicle streams.

GRADE SEPARATION AND THE ROTARY

The rotary has been called a compromise between a signalized intersection and grade separation and recommended where the limitations imposed by cost and space prevented construction of the latter (45, pp. 9-12). Today's inflation and the growing resistance of the public to large-scale highway construction projects may make the high-capacity rotary at comparatively low cost a more widely accepted alternative to grade separation. Cost as well as a shortage of land may prohibit the construction of a conventional cloverleaf interchange where a two- or three-level rotary design would reduce cost and space requirements to a minimum (Figure 3h, i). At diamond interchanges, two small or zero-center rotaries linked by a single bridge (Figure 3k, l) have been recommended for economy in place of the conventional two-bridge, single rotary (17). This construction can be employed to relieve or eliminate the problem that often exists at signalized diamond interchanges where traffic backs up along the exit ramp onto the freeway.

CAPACITY CALCULATION

From road experiments, TRRL developed the following formula for calculating the capacity of offside-priority rotaries with small center islands (Figure 7) (17):

$$Q = K(\Sigma W + \sqrt{A}) \tag{1}$$

where

- Q = practical capacity (vehicles/h);
- K = efficiency constant of 70 for three-leg intersections, 50 for four-leg intersections, and 45 for five-leg intersections (for rotaries whose center islands are less than 4 m (13 ft) in diameter, a reduction of 10 percent should be made);
- ΣW = sum of basic widths (not half-widths) of all intersecting roads (m); and
- A = area added to the intersection by the flared approaches and exits (m^2).

TRRL has now adopted a new procedure that assesses capacity on an entry-by-entry basis in terms of entry geometry (25, 46).

Accurate results have been obtained by the rule-of-thumb formula (19)

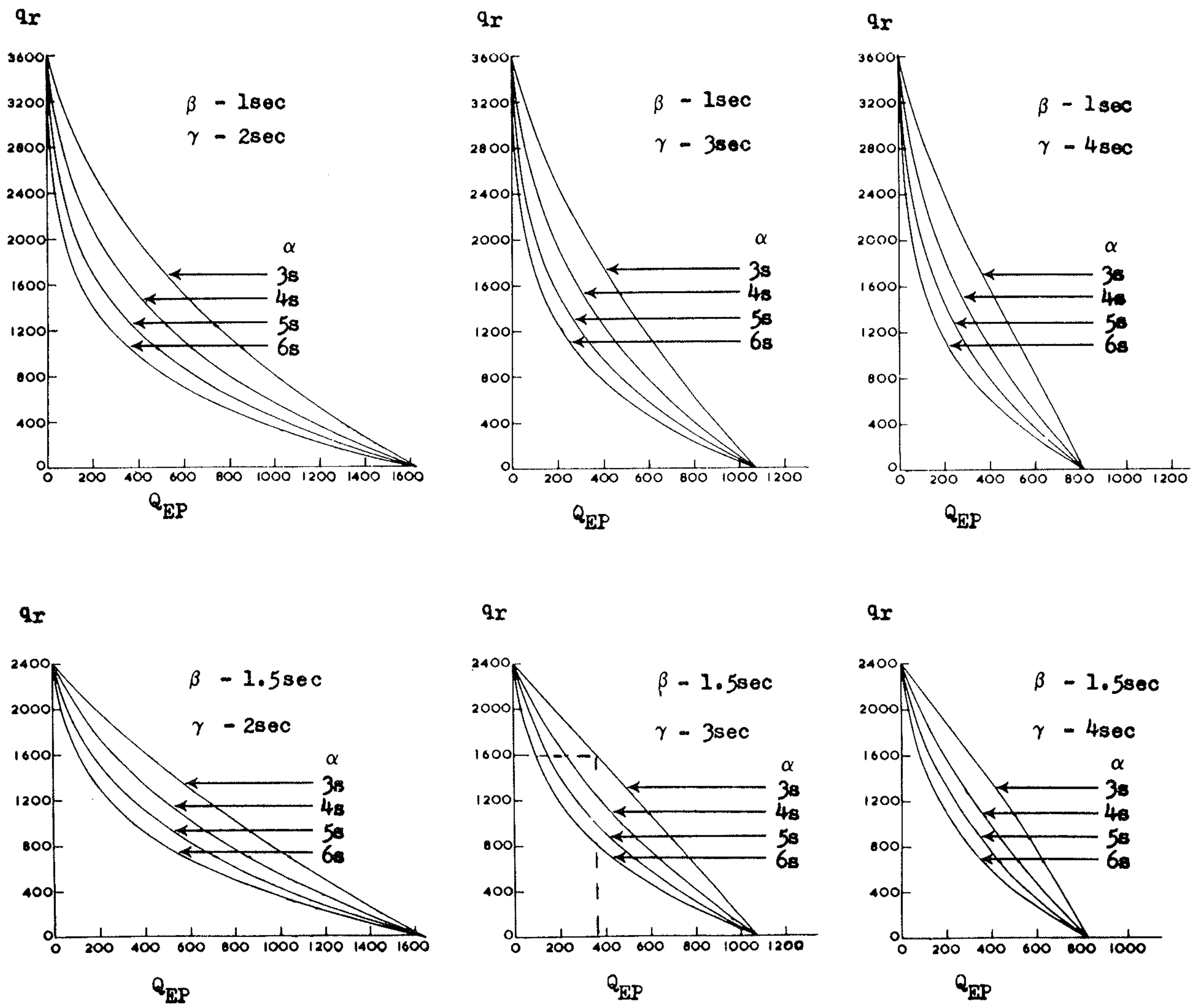
$$Q = KD \tag{2}$$

where

- Q = saturation capacity (passenger vehicles/h),
- K = 150 for three-leg intersections and 140 for four-leg intersections, and
- D = diameter of the inscribed circle (m); for an oval intersection, D is the mean of the major and minor axes.

For design purposes, the flows should not be more

Figure 8. Bennett's design graphs.



Bennett's design graphs, based on Tanner's formula $Q_{EP} = \dfrac{q_r(1-\beta q_r)}{e^{q_r(\alpha-\beta)}(1-e^{q_r\gamma})}$

than 85 percent of the volumes calculated from these formulas.

R. F. Bennett (1) considered each approach as an entry into a T-intersection. By modifying Tanner's formula (47), he developed the graphs shown in Figure 8, where $\overline{Q}_{EP}$ is the practical entry flow per lane from the approach leg (passenger vehicles/h), q_r is the circulatory flow per 7.5 m (24 ft) of roadway passing the entry, β is the minimum headway on the rotary, α is the minimum gap acceptable to entering drivers, and γ is the move-up time of vehicles in the entering queue. Suggested values for α, β, and γ are as follows:

1. $\alpha = 3$ s for level urban approaches, 4 s for level rural or uphill urban approaches, and 5 s for uphill rural approaches.
2. $\beta = 1$-1.5 s (for peak flows, 1 s gives more realistic results).
3. $\gamma = 2$ s for a downhill approach, 3 s for a level approach, and 4 s for an uphill approach.

For example, if there is a level approach to an urban rotary that has a circulatory flow (q_r) of 1600 vehicles/h and values of $\alpha = 3$, $\beta = 1.5$, and $\gamma = 3$, then the entry flow per lane of 350 vehicles/h can be read from the fifth graph in Figure 8. Therefore, if the entry flow is 850 vehicles/h, three entry lanes will be required.

SUMMARY

In Britain, the capacity of large, conventional rotaries has been raised by reducing the size of the center island and increasing the entry width to give a higher rate of vehicle discharge. Small rotaries have been developed to overcome the drawbacks of large layouts. The simpler designs in the form of T- and Y-intersections gave the best performance for safety and capacity. The rotary principle can also be applied to a grid of one-way streets, where a block forms the center island of a rectangular rotary. The following benefits (in comparison with conventional alternatives) can be obtained:

1. Stops and delay are drastically reduced.
2. Capacity within the existing intersection area is raised by 5-30 percent.
3. Space permitting, an intersection and its immediate approaches and exits can be enlarged to handle more traffic, without a need to widen the streets along

their entire length; the need for six- and eight-lane arterials is greatly reduced.

4. On existing six- and eight-lane highways, exclusive lanes for buses, carpools, or bicycles can be operated without detriment to overall traffic performance.

5. There is no left-turn prohibition (in countries where traffic drives on the right, or right-turn prohibition where traffic drives on the left), and no median is needed to accommodate left-turners (or right-turners, respectively).

6. Compensation is made for unbalanced flows.

7. There are low maintenance costs, no operating costs, and no equipment failures.

8. Rotaries are aesthetically pleasing, give smoother flow, and have lower noise levels.

9. There is less incentive for commuters to filter through residential areas.

10. Safety is increased, and road-user decisions are simpler.

11. Signal-free operation of one-way streets is possible.

12. Less land is taken and there are lower construction costs at interchanges.

ACKNOWLEDGMENT

Permission from the County Surveyors' Society, TRRL, and the Institution of Highway Engineers to reproduce illustrations from their literature is gratefully acknowledged.

REFERENCES

1. R. F. Bennett. The Design of Roundabouts Since the Priority Rule. Journal of the Institution of Highway Engineers, Sept. 1971, pp. 13-20.
2. A Policy on Geometric Design of Rural Highways. AASHO, Washington, DC, 1965.
3. Traffic Circle Relieves Crowded Street Intersection. Engineering News-Record, Vol. 89, No. 13, 1922, p. 520; Vol. 90, No. 4, 1923, p. 164.
4. R. B. Brooks. Traffic Circles to Relieve Intersectional Congestion. Nation's Traffic, March 1928, pp. 26-27.
5. L. J. Tuttle and E. H. Holmes. The Design of Street and Highway Intersections. Public Roads, July 1932, pp. 73-88.
6. K. J. Carter. Driver Behaviour at Roundabouts. Univ. of Newcastle-upon-Tyne, thesis, May 1966.
7. W. P. Eno. The Science of Highway Traffic Regulation. Brentano, New York, 1920.
8. W. P. Eno. Simplification of Highway Traffic. Eno Foundation, Saugatuck, CT, 1929.
9. G. G. Kelcey. Traffic Engineering. American Gas Company, Elizabeth, NJ, 1922.
10. G. Kelcey. Channelization of Motor Traffic. Proc., First Annual Texas Traffic Engineering Conference, Austin, TX, 1940, pp. 250-271.
11. T. M. Matson and L. Williams. Elements of Intersection Redesign. Bureau of Street Traffic Research, Yale Univ., New Haven, CT, 1940.
12. Improving the Capacity of Roundabouts. Surveyor, March 29, 1969, pp. 34-37.
13. F. V. Webster and F. C. Blackmore. Improving Road Capacity. Science Journal, Aug. 1968, pp. 69-74.
14. A. Attwood. A Study of Roundabout Modifications. Traffic Engineering and Control, May 1973, pp. 22-27.
15. Mini-Roundabouts in England and Wales. County Surveyors' Society, Chelmsford, England, 1972.
16. F. C. Blackmore. Capacity of Single-Level Intersections. Road Research Laboratory, Crowthorne, England, Rept. LR 356, 1970.
17. Roundabout Design. Highways Directorate, U.K. Department of the Environment, London, Tech. Memorandum H2/75, 1975.
18. Capacity of At-Grade Junctions. Organization for Economic Cooperation and Development, Paris, 1974, pp. 61-66.
19. D. Poole. Design and Performance of Mini-Roundabouts. Municipal Engineering, June 8, 1973, pp. 1347-1351; June 15, 1973, pp. 1389-1392.
20. Manual on Uniform Traffic Control Devices. AASHO, Washington, DC, 1935.
21. R. S. Millard. Roundabouts and Signals. Traffic Engineering and Control, May 1971, pp. 13-15.
22. Traffic Congestion Relief: Alternatives to the Construction of Additional Lanes on Signalised Arterial Roads. CIPAG Bulletin, Sept. 1976.
23. A. Herzog. Traffic Signals at the Place de l'Etoile. Proc., 8th International Study Week on Traffic Engineering and Safety, London, 1956.
24. F. C. Blackmore and M. Marlow. Improving the Capacity of Large Roundabouts. Transport and Road Research Laboratory, Crowthorne, England, Rept. LR 671, 1975.
25. R. M. Kimber and M. C. Semmens. A Track Experiment on the Entry Capacities of Offside-Priority Roundabouts. Transport and Road Research Laboratory, Crowthorne, England, Supplementary Rept. 334, 1977.
26. R. I. Strickland. A Study of Merging Movements. In Studies of Weaving and Merging Traffic, Bureau of Highway Traffic, Yale University, New Haven, CT, 1940, pp. 59-130.
27. M. J. Philbrick. A Filter Lane for Left-Turning Traffic at Slough-Central Roundabout (M4/355). Transport and Road Research Laboratory, Crowthorne, England, Rept. LR 634, 1974.
28. H. J. Klar. Designing Safe and Efficient High-Volume Driveways and Intersections. Traffic Engineering, Jan. 1973, pp. 32-37; April 1975, p. 70.
29. Pedestrian Safety. U. K. Department of the Environment, London, 1973.
30. R. L. Manning. Accidents at Controlled Junctions. Road Research Laboratory, Crowthorne, England, Note RN/2133, 1954.
31. F. Garwood and J. C. Tanner. Accident Studies Before and After Road Changes. In Public Works Municipal Services Congress, 1956, pp. 325-354, 378-380.
32. Roundabouts Preferred to Signal-Controlled Junctions. Surveyor and Municipal Engineer, Aug. 19, 1961, p. 1021.
33. J. C. Tanner. A Problem in the Combination of Accident Frequencies. Biometrika, Vol. 45, No. 3-4, 1958, pp. 657-658.
34. H. Green. Accidents at Offside-Priority Roundabouts with Mini or Small Islands. Transport and Road Research Laboratory, Crowthorne, England, Rept. LR 774, 1977.
35. D. Poole. Mini-Roundabouts. Paper presented to the Institution of Municipal Engineers, Oct. 1971.
36. N. Lalani. The Impact on Accidents of the Introduction of Mini, Small, and Large Roundabouts at Major-Minor Priority Junctions. Traffic Engineering and Control, Dec. 1975, pp. 560-561.
37. F. Cooke. The Mini-Roundabout: Its History,

Design, and Use. Civil Engineering Technician, Nov. 1973, pp. 7-23.
38. Manual on Uniform Traffic Control Devices for Streets and Highways. Federal Highway Administration, 1978.
39. R. F. Beaubien. Stop Signs for Speed Control? Traffic Engineering, Nov. 1976, pp. 26-28.
40. L. C. Orlob. Traffic Diversion for Better Neighborhoods. Traffic Engineering, July 1975, pp. 22-25.
41. W. Marconi. Speed Control Measures in Residential Areas. Traffic Engineering, March 1977, pp. 28-30.
42. C. B. Horman and H. H. Turnbull. Design and Analysis of Roundabouts. Proc., Australian Road Research Board, Vol. 7, No. 4, 1974, pp. 58-82.
43. A Proposed Traffic Plan. Nation's Traffic, June 1929, pp. 31, 37.
44. F. Malcher. The Steadyflow System. Harvard University Press, Cambridge, MA, 1935.
45. A Policy on Grade Separation for Intersecting Highways. AASHO, Washington, DC, 1944.
46. M. G. M. Glen, S. L. Sumner, and R. M. Kimber. The Capacity of Offside-Priority Roundabout Entries. Transport and Road Research Laboratory, Crowthorne, England, Special Rept. SR 436, 1979.
47. J. C. Tanner. A Theoretical Analysis of Delay at an Uncontrolled Intersection. Biometrika, Vol. 49, 1962, pp. 163-170.

Discussion

G. F. Hagenauer, De Leuw Cather and Company, Chicago

I think we all, to some extent, share the author's concern about the phasing out of existing rotary intersections and the lack of interest in constructing new rotaries in this country. I think we can agree with his assessment of their many advantages. As a recent brief article on traffic circles in Transportation Research News (48) noted,

> In view of their many apparent advantages, it would seem logical that traffic circles would continue to be universally popular. The need for expensive and complicated conventional traffic control devices may be reduced or even eliminated. The graceful lines of circles would seem aesthetically preferable to the angularity of typical at-grade intersections and their control hardware clutter. The continuous flow of vehicles through a circle provides a quieter and cleaner environment than the stopping and starting actions associated with conventional intersections.

The most recent U.S. publication in Todd's references is Policy on Geometric Design of Rural Highways, published in 1965 (2), which lists advantages similar to those mentioned above.

The way in which they are described, however, is not strongly persuasive. For example, the statement "there is little speed reduction and no delay from stopping" applies only to low volume levels. It is further stated that accidents "are of minor nature" but, in the United States, accidents have been a concern at rotaries, not because of their severity but because of their frequency.

The reasons for the demise of rotaries in the United States are probably best summarized in the listing of disadvantages in the AASHO Manual:

1. "A rotary can accommodate no more traffic than a properly designed channelized layout. In some cases, rotaries have been eliminated and replaced with a channelized intersection, resulting in better operation." This statement by AASHO is contrary to the assertion in Todd's paper that a rotary can be "called a compromise between a signalized intersection and a grade separation."
2. "A rotary does not operate satisfactorily when, on roads of four or more lanes, the traffic volumes on two or more intersecting legs approach capacity at the same time." This is an extension of the capacity limitation found in U.S. experience. When capacities of approaches to rotaries are exceeded, there is no way to adjust the capacity of each approach to reflect traffic demand volumes. As a result, disproportionate backups are generated on the leg or legs that have the highest demands.
3. "Rotaries require more right-of-way and roadway, and generally cost more, than other at-grade intersections. Access must be controlled for a rotary to function properly. This control may be difficult to obtain when approach highways or streets do not also have access control." These characteristics practically eliminate the consideration of replacing existing high-volume intersections with rotaries.
4. "Rotaries are not directly suitable for conditions that require the crossings of large movements of pedestrians. Orderly flow of vehicles is interrupted where this requirement must be met and it is frequantly necessary for rotaries to be operated with traffic signal controls. This, however, violates the rotary concept of continuous movement." This statement seems very logical—that pedestrians (or bicycles) would find it difficult to cross continuously moving vehicles in several traffic lanes. Yet, Todd states that "Rotary control appears to bring considerable benefits to both pedestrian and vehicle safety" and "Pedestrians at rotaries generally have to walk a longer distance, but their average delay has been found to be less than that at signals."
5. "Rotaries are not readily adapted to high-speed roads. They must be extremely large to provide the proper weaving lengths between intersection legs. Large rotaries add additional travel mileage, particularly for left-turning vehicles. This characteristic should be weighed against delays likely to occur at alternate channelized layouts. For safety and proper operation, particularly of those not familiar with the layout, numerous signs clearly visible both day and night are essential. Signing which avoids confusion on the part of drivers, however, is difficult of attainment." One of the many problems for users of rotaries, particularly strangers, is disorientation. As a result, care must be taken so that sufficient signing and marking is supplied to properly guide motorists; yet oversigning, which would produce confusion, must be avoided. In this regard, it is interesting to note that signing and marking recommendations for rotaries are not commonly found in current federal or state manuals of uniform traffic control devices.
6. "Lighting is desirable and landscaping of the extensive unpaved areas is required. The cost of these items should be weighed against installations required for alternate channelized designs. . . . Rotaries are not readily adaptable to stage development. Attempts at stage development generally result in some over-design when measured by immediate traffic needs."

In light of these many disadvantages, the generating of new enthusiasm for rotary retention or construction in the United States appears difficult, if not impossible. The most discouraging advice, however, appears in the report of the Committee on Urban Street Design of the Institute of Transportation Engineers (ITE). The report, Guidelines for Urban Major Street Design, is described as "a proposed recommended practice"; it is in the final stages of review and approval by ITE's Technical Council and Board of Direction. Rotary intersections are mentioned on only one page of this very large report: "rotary intersections are not recommended....Rotaries are undesirable because of the high number of weaving movements that must be accommodated in a short distance."

It would appear that the handwriting has been on the wall for quite a few years. Todd's paper is based on extensive references from both the United States and overseas. The latter, however, are mostly from Great Britain, where rotaries have been uniquely successful. Those from this country, on the other hand, are principally from the years prior to 1940, and their applicability to present conditions can be questioned.

Since at least the 1960s, however, U.S. engineers on the state and local levels and consultants have been working to produce good operation in existing rotaries and, based on their experience, have recommended that new rotaries should not be constructed. In the face of these facts, a resurgence of interest in rotaries in this country would seem extremely unlikely.

REFERENCE

48. Traffic Circles: Their Past, Present, and Future. Transportation Research News, Sept.-Oct. 1978, pp. 4-6.

Robert J. Nolan, Bureau of Traffic Engineering, New Jersey Department of Transportation, Trenton

Although the New Jersey Department of Transportation is not constructing any new circles on its highway system, we do have approximately 60 circles on our system, most of them built in the 1920s and 1930s. No one is advocating building any more. Therefore, insofar as our highway system is concerned, the basic problem is not how to design new traffic circles but what is the best approach to improve the operation of those we have.

Todd states that, in some instances, traffic circles have been used to replace traffic signals. This is just opposite to our practice. We have reconstructed a number of circles by cutting through them and signalizing the resulting intersections. We have found that such a technique results in a smoother flow, with less congestion and few accidents. At circles that have no more than four approaches and large enough radii to handle the storage of turning vehicles, this approach has worked very well.

We have had no experience with reducing the size of the circle, as Todd mentioned. We believe that the basic problem lies not in the size of the circle or even in the shape of the circle (for in New Jersey we have many "circles" that vary in shape from ellipsoid to ovoid) but in the conflicts generated between overly aggressive and overly timid drivers. Our observations lead us to believe that, if this conflict can be resolved or ameliorated, the flow through the circle will improve.

Therefore, our latest approach has been to try to improve flow into the circle by installing traffic signals on one or more of the approaches 100 m (300 ft) or more in advance of the circle in order to meter the flow into the circle. Since this approach used no additional right-of-way, it does not disturb the land use adjacent to the circle, nor does it involve lengthy and complicated construction.

Since traffic circles in our state tend to be the preferred locations for gasoline stations, shopping centers, and diners, the last item is not a minor one. We have at present one circle that has five approach roadways, all signalized, with detectors in both the approach roadways and the circle proper, wired into a computer installed in a trailer inside the circle. Studies now under way at that location, Ellisburg Circle, show that we can achieve total hourly volumes in excess of 4000 vehicles/h but that above 4400 vehicles/h the flow tends to become unstable and "lockups" occur beyond that point.

Jack E. Leisch, Jack E. Leisch and Associates, Evanston, Illinois

Todd's paper is an enlightening account of the design and operation of rotary intersections in Great Britain. Normally the subject would draw little interest in the United States.

The rotary as a real traffic carrier has been written off and, for the most part, considered a relic in modern transportation engineering practice in the United States. Of course, the rotary may have its place for local or distributor-type traffic at low volumes for aesthetic or architectural purposes and occasionally for maintaining low speed levels in residential communities. Such minor use, however, is incidental, since this discussion focuses on situations where the ability and necessity to handle traffic is of primary concern.

Although this paper at first glance may not seem relevant, it does command interest and perhaps some analysis because of the unusually optimistic and highly favorable account presented in terms of increased capacity, operational efficiency, safety, and a multitude of other advantages attributed to the use of rotaries in Great Britain. The troublesome aspect of this review is that nearly every performance level or characteristic cited as an attribute and an advantage for the British situation is considered a failure and a disadvantage for the American condition.

According to our experience in the United States, it is hard to conceive that rotaries could improve safety, lessen delays, and reduce pedestrian-vehicular hazards. What is really difficult to comprehend is that "mini-roundabouts" with center islands less than 4 m (13 ft) in diameter have become popular in Great Britain during the past 10 years. Their advantage is that the capacity has been increased 50 percent by means of reducing the size of the center island (of previous designs) and increasing the entry width.

Although it is assumed that such design and attendant performance could not be achieved and certainly would not be attempted in the United States, we must recognize that Todd's presentation draws on actual experience and research, and it consequently deserves proper attention. There must be an explanation for such diversity in design application and operational experience.

The decline in the use of rotary intersections in the United States, which began some 30 or 35 years ago, is supported by specific and valid reasons. Also the adherence to the rotary, or roundabout, in Great Britain to the present time—and its increased use in recent years—is also fully understandable. The claims of functional advantages or disadvantages, or operational efficiency or inefficiency, as the case may be, can be fully comprehended and explained if they are properly analyzed.

Let us first consider the course that the United States pursued in its highway development and what goals were to be achieved in traffic operations: increase in capacity; increase in speed, followed by improvement in safety; and a desire to maintain higher levels of service. The willingness to make changes, coupled with available administrative, engineering, and budgetary resources, during a period of intensive improvement programs, provided the means to convert rotaries to other forms of intersections. Although it was only part of the total improvement programs, the replacement of the rotary was a step toward the goals designed to improve operations.

Contrary to highway development and programming of improvement in the United States, the approach in Great Britain during the past several decades was apparently to maintain the operation of its existing roundabouts and to continue to provide new facilities as they were considered appropriate. Such perseverance in the use of the rotary-type intersection has encouraged the engineer to experiment and improve its operation and, in the process, to justify its existence. At the same time, the driving public has continued to acclimate and grow with the use of rotary configurations. The incentive to maintain their existence and to improve their operation has apparently produced the results Todd reported.

The obvious question that cannot be avoided is, Should we in the United States return to building rotaries, or should rotaries in Great Britain be replaced with the more conventional intersection configurations and controls? The obvious answer is that each country will, and perhaps should, continue its present practice. If each considers it has a degree of success with its approach, and there is a reasonable explanation on the part of each, the discussion can thus be appropriately concluded. However, it is difficult to brush this issue off by implying that practice and drivers are sufficiently different in each country that the same design may work well in one and not in the other.

It may be true that the education, orientation, attitude, and tolerance of drivers may differ from area to area, but an issue so major cannot be so simply explained away. It is therefore appropriate to look at the problem from a much broader point of view in an attempt to obtain a better explanation.

I believe the main issues in the question before us have to do with philosophy, human factors, and the broad principles of system configuration. It is here that the real differences may be found. Although the following is stated in the light of U.S. perspective and reasoning, the points made would appear to have universal application, although emphasis and values may differ.

1. Philosophy: What degree of serviceability, quality of operation, or level of service for the driver are considered appropriate? What allowances for comfort, convenience, freedom to maneuver, and speed of operation—as policy—are provided and adhered to in design and operations? This refers not just to an intersection point but to a linear facility or a network of facilities. Another philosophical approach to design and operation is to make it "simple," "direct," "uniform," "easy to follow and maneuver," and with "least possible points of conflict."

2. Human factors: As an extension of the service-level considerations, there are further operational aspects relating to the driver. This concerns the more in-depth characteristics of driver behavior and the reasons behind it. In addition to his or her physiological characteristics, the driver's psychological and emotional makeup play a significant role in how to design the facility and its operation. For this reason, the number and complexity of tasks the driver must perform should be kept to a minimum.

3. Principle of system configuration: An important feature of highway facilities, essential for optimum operation, is the principle of nonoverlapping routes. A major effort to eliminate overlapping routes has been evident in recent years. The principle applies to a single intersection or to a length of highway and its interconnections. Accordingly, improvements and corrections to the system are continually eliminating staggered intersections, rotaries, cloverleaf interchanges, other forms of intersections and interchanges that have weaving sections, and direct overlappings of routes in which two roadways join on one roadway and (after some distance) separate as independent roadways.

These three points embody the reasons for not using rotaries in the United States. The mini-roundabout described by Todd would probably receive an even less enthusiastic response. It would seem that introducing the mini-roundabout would be somewhat akin to reverting back to intersections without control. There is evidence that, at heavily traveled intersections, approximately as much traffic can be handled without signal control as with it if properly regimented traffic proceeds intermittently. Although this has been demonstrated, its application would not be accepted here, nor would the larger rotary, which presents the driver with a multitude of overlapping tasks, worries, and harassments during periods of heavy flow.

It is problematical to what extent these principles, philosophy, or policies would apply or have any bearing on design and operation of rotaries in other countries. Certainly the differences are not in the number of vehicles measured, accidents reported, or left turns (or right turns) accommodated, but in human factors aspects and in quality and uniformity of operation. The standard by which operational quality and driver satisfaction is measured, coupled with other local characteristics, determines the success or failure of a certain design.

Author's Closure

I would like to thank the discussants for their stimulating comments. I agree that many rotaries in the United States do not perform as well as they should. Their decline, in spite of so many advantages, is perhaps due to a lack of effort to eliminate their many disadvantages. It was said long ago that "the failure to function well should not be attributed to the fact that they are traffic circles, but to the fact that many of the most fundamental requirements for successful operation are lacking" (49). The reasons for this failure may be worth looking into more closely.

Built long before the automobile arrived, the earliest rotaries in the United States were an architectural feature, not a traffic control device. The outer curbline of many runs concentric with the center island rather than in the opposite direction. The resulting tight entry-curb radii slow the entering vehicles excessively and force them to swing out into the circulatory roadway, leaving dead pavement areas. Nevertheless, these rotaries do handle traffic moving at low speeds. Their proper operation was most severely impeded in the days when streetcars were present.

As the use of rotaries spread into rural areas in the 1920s, speeds below 40 km/h (25 mph) were found unsatisfactory in conjunction with highway speeds of 64 km/h (40 mph) or more. Design speeds of not less than 40 km/h were recommended, and the minimum radius required for the design speed determined the size of a rotary. To avoid excessive dimensions, the maximum recommended design speed was 64 km/h.

The designs provided that entering vehicles should merge and then weave with circulating traffic; the larger the weaving volume, the longer the weaving section had to be (50, p. 13). It led to the belief that rotaries had to be very large to handle heavy traffic and that the capacity of existing layouts could only be raised by enlarging the inscribed diameter or by cutting a road through the center.

When urbanization spread and brought more traffic, rotaries that had primarily been built to speed requirements were found to lack capacity. Many had been built with two-lane approaches merging into a two-lane circulatory roadway. Two heavy streams of traffic, each on two lanes, cannot be expected to merge successfully into two lanes on any type of road.

A rotary does not function properly when drivers within the circle tend to yield to fast traffic approaching on what seems to them a major road; some center islands had even been shaped to keep speed fairly high along the busier road. A rotary functions worse still where the basic yield-to-the-right rule gives the right-of-way to entering drivers. This prevents vehicles within the rotary from leaving and causes traffic to lock. I am told it happens frequently in New Jersey, but the American literature does not seem to be aware of the problem. We should remember that by the early 1930s, the heyday of the rotary, all states had adopted the yield-to-the-right rule. (Stop or yield signs placed at the entries will cure the locking by reversing the right-of-way, but the stop sign violates the principle of continuous movement and the yield sign did not appear on the scene until 1951, long after the rotary had fallen out of favor.) A similar problem existed in Britain, where no legally defined right-of-way was in force until 1966, when legislation was introduced that required entering drivers to yield. This cured the locking, but many signals that had been put up to overcome it were never removed. In the United States, some jurisdictions (Delaware, Massachusetts, Maine, New York, North Carolina, Rhode Island, Virginia, and the District of Columbia) have adopted similar laws, beginning in 1952.

Hagenauer's comments rely largely on quotations from the 1965 AASHO Blue Book (2), whose section on rotaries is almost a verbatim copy of that in the 1954 edition. The 1954 edition, in turn, gave an abridged version of A Policy on Rotary Intersections (50), published in 1942. Hence, the comments are based on experience with large layouts of pre-1942 design that have design speeds of 40-64 km/h (25-40 mph), many of which are beset with the problems just outlined. Smaller rotaries that have design speeds below 40 km/h were specifically excluded from consideration in the 1942 publication, which said (p. 1),

> According to definition, even a center post at a street intersection produces rotary traffic flow, but a rotary is more commonly considered to have a central area or island of some size. In this discussion, the name rotary is applied only to those layouts in which the radius of any portion of the central island is at least 75 feet [23 m]. Rotary intersection layouts with small central islands of radii less than 75 feet require speed reduction to less than 25 mph [40 km/h].

A radius of 23 m (75 ft) plus a 7-m (24-ft) circulatory roadway would give an inscribed circle diameter of 60 m (198 ft) as the minimum size AASHO considered. In Britain, a modern roundabout of that size would be expected to handle about 6000 vehicles/h.

"A [section] length below 100 feet [30 m] resolves weaving movements into typical at-grade crossings," stated the 1965 AASHO Blue Book (p. 482). These rotaries have no weaving sections, and traffic does not weave—it intersects. In consequence, the passage quoted from the ITE report that "rotaries are undesirable because of the high number of weaving movements that must be accommodated within a short distance" should not be applied to a type of rotary that has no weaving sections.

"Where the distance [between adjacent entrances and exits] is so small that vehicles cross at an oblique angle without weaving, the intersection is not classed as a rotary but rather as a channelized intersection" (51, p. 514). The performance of this type of intersection has never been evaluated in the United States, and conclusions drawn from experience with large layouts of pre-1942 design, all of them operating under the yield-to-the-right rule until the early 1950s and many until this day, cannot simply be transferred to a type of rotary that was developed for the express purpose of overcoming the disadvantages previously encountered, a type that occupies far less space, is not designed for speed but for capacity, and lends itself to stage development far better than any other type of intersection.

As to the accident frequency mentioned by Hagenauer, in spite of much effort no data have been found that allow a comparison of unsignalized rotaries with other forms of control in the United States. Any information from readers would be welcome. Rotaries have been blamed for accidents that were, in fact, due to inadequate signposting, a lack of advance warning, a slippery pavement, or poor design. If more PDO collisions do occur, which is quite possible, they should be set off against any reduction in the more serious accidents, as well as reductions in accidents and operating costs due to less congestion and fewer stops in the road network as a whole.

In regard to Nolan's remarks on signalizing rotaries, this has also been done in Britain. A more positive solution would be to raise the output at the critical entry point by geometric changes, but this is not always possible. Signals are likely to increase overall delay, particularly if they are kept in operation during hours when the problem they were put up to cure does not occur. Several methods other than stop-go signals (for instance, pedestrian movements or peak-hour police control) could be used to generate gaps in a predominant traffic stream at a rotary or elsewhere, but their description is beyond the scope of this paper.

The timid, or defensive, driver will always reduce capacity at unsignalized intersections when he or she rejects the shorter gaps. At a rotary, this is minimized when more entry lanes are provided that other drivers can use while the timid driver is hesitating. Defensive drivers also reduce capacity at signals by leaving longer headways. It is the aggressive driver within the rotary who reduces capacity; he or she intimidates others into rejecting short gaps. Although it has often been shown that drivers accept shorter gaps in faster traffic, there are sound reasons to believe that an enforced speed reduction to about 25 km/h (16 mph) can raise gap acceptance—and capacity—at rotaries and elsewhere.

A road has been cut through the center of several circles in the Washington, D.C., area. The circle then no longer operates on the rotary principle, which requires that all entering drivers yield. It functions in a manner similar to two jug handles and needs signals.

Finally, like Leisch, I am interested in the philosophy of traffic control. Since the profession has for 50 years advocated the use, wherever possible, of controls less restrictive than signals (52, p. 13; 53, pp. 322-323), I do not think that anyone would wish to recommend signals (with more stops, delay, queueing, and congestion) in order to achieve better operational quality and greater driver satisfaction.

For the design engineer, the difference in philosophy is perhaps whether to build for speed or to build for capacity. A philosophy developed for the construction of high-speed roads in the days when money was plentiful is not necessarily the most suitable for treating bottlenecks in times of severe inflation. But this paper does not deal with philosophy. It confines itself to describing the potential of modern rotaries, and the reader can decide according to his or her own philosophy what to do with the information. Nevertheless, when a highway department proposes the widening of a road, the elimination of a street jog, or the construction of an interchange, it would be expected to submit other feasible, prudent, and less harmful alternatives, together with its own recommendations and reservations. The fact that past and present design standards do not deal adequately with rotaries should not deprive the public of the benefits of a highly cost-effective TSM alternative.

REFERENCES

49. Control and Design of Multiple Intersections. Proc., American Road Builders Association, Bull. 32, 1932, pp. 81-92.
50. A Policy on Rotary Intersections. AASHO, Washington, DC, 1942.
51. T. M. Matson, W. S. Smith, and F. W. Hurd. Traffic Engineering. McGraw-Hill, New York, 1955.
52. A Standard Code for Traffic Control Signal Installation and Operation. Massachusetts Department of Public Works, Boston, 1929.
53. L. J. Pignataro. Traffic Engineering: Theory and Practice. Prentice-Hall, Englewood Cliffs, NJ, 1973.

Publication of this paper sponsored by Committee on Operational Effects of Geometrics.

NOTICE

This paper was previously published in the ITE Journal (July 1979), without the three discussions and author's closure that accompany this version of the paper.

Highway Guide Signs: A Framework for Design and Evaluation

J. Paul Dean*, Reid, Crowther, and Partners Ltd., Calgary, Alberta, Canada

Inadequate guide-sign design and location practices are responsible for a great number of instances of inefficient and potentially hazardous traffic operations within the highway system. Many of these deficiencies are rooted in inadequate highway system planning. The great majority of these locations, however, have operational problems that are directly related to the failure to coordinate guide-sign design and functional design as part of the highway development process. A framework for the coordination of guide-sign design with functional design is proposed and illustrated; emphasis is on the guide-sign development, design, and evaluation tasks. A computer program to plot perspectives was modified to provide the means of accurately depicting proposed guide signs within the perspective of a highway facility from the position of the driver's eye. The view presented by the perspective provides the designer with the third dimension—the view of the sign from the road. Shifts in vertical, lateral, and longitudinal sign position can be studied accurately by using the perspective as a tool. Examples were taken from two existing highway locations and provide the means for evaluating the suggested guide-sign design procedure for both existing and proposed guide-sign installations. Recommendations for further research and development are aimed at the reduction of the manual aspects of the procedure, thereby providing the designer with maximum incentive for investigation of alternatives, graphic details, and variations thereof.

Highway design philosophy, procedures, tools, and techniques have undergone rapid change in response to a recognized need for safe, efficient, environmentally acceptable, and economic highway facilities (1-5). The major objective of these studies, taken together, is directed toward the provision of a highway facility on paper that clearly satisfies recognized needs before construction and before the facility is put into operation.

Until recently, the driver and the driving task have been largely neglected by highway engineers; the driver has been obscured by gross statistical descriptions contained in design manuals, handbooks, and policies under the label of "driver and traffic characteristics." Since current research in human factors engineering, as applied to the driver and the driving task, are gaining more acceptance among highway engineers, attitudes are changing in the designers' approach to planning and design of new facilities and in refurbishment of obsolete highways. This attitude can be described as an awareness and concern for the driver and for facility operation; it incorporates a design philosophy that attempts to include the driver.

The need for recognition of new tools and techniques for design by the designer is absolutely necessary in order for him or her to deal effectively with obvious past deficiencies (obvious after the facility is put into operation). Why should accident experience or inefficient traffic operations be the prime motive for change and evolution of design procedures?

An important component of the design process is the design of the formal communications system for a highway facility. Guide signs constitute the most expensive part of the formal communication system, and they are the most permanent feature, other than the right-of-way and roadway surface.

Present costs (in 1977 dollars) of reflectorized, illuminated overhead and shoulder-mounted guide signs are on the order of \$300-\$400/m^2 (\$30-\$40/ft^2). This cost includes engineering, fabrication, delivery to the site, erection, and structures to carry the sign-panel display, along with illumination requirements. A typical kilometer of suburban freeway that contains one interchange may require an expenditure of more than \$50 000-\$100 000 for guide-sign needs. A careful, soundly based engineering approach is obviously required.

More important, however, is the highway engineer's awareness of the interaction of the guide sign and its message with the driver and the roadway. Guide signs are used by the driver primarily in the driving subtask of navigation. Failure of the formal information system (and, hence, of the highway facility) may, at worst, produce erratic driver response leading to accidents, injuries, and fatalities; at best, failure of the formal communication system may increase trip time and driver frustration and produce lack of confidence by the driver in what he or she perceives. Because the driver builds on past experience in the driving task, his or her skill is directly related to the ability to process information (6). Guide signs therefore play an important role in the successful operation of a highway facility.

It is a fact of driving experience that guide-sign legends and directional information are inadequate at numerous locations within our highway systems. There are many reasons for these apparent inadequacies, but one major cause is lack of communication within highway agencies—among those personnel responsible for system planning, functional planning and design, and design of guide signs. Many agencies separate the planning and design function from traffic engineering tasks, which normally include responsibility for design of guide signs. Usually, those whose responsibility includes guide signs are the last to become involved in the design and construction process.

Even sophisticated and time-consuming design review procedures have not satisfied or completely eliminated the resulting lack of coordination between the highway design and driver information needs. When so much effort has been expended in reaching the design review stage, the resistance to change or modification is naturally much greater. What is needed, therefore, is the direct involvement of sign designers in the functional planning and design stage; after all, if the facility cannot be signed, it should be modified. The functional design stage is where modifications should occur.

The balance of this paper deals with a description or proposal for linking guide-sign planning, design, and evaluation to the functional planning and design stage of the highway development process. In addition, a tool for evaluating a proposed sign or series of signs is described—a perspective-plotting technique wherein scaled sign-panel mock-ups are positioned within perspective plots. The results of a limited application of the perspective technique to actual locations are illustrated and described. Suggestions for further research and development outlined are subject to practical acceptance and testing of the technique illustrated.

COORDINATION OF GEOMETRIC DESIGN AND SIGNING

Highway System Planning Prerequisites

Before discussing the link between geometric design and signing, it is important to recognize that the guide-sign legend begins to take shape in the first stage of highway system planning where broad aspects are dealt with—route numbering system and other coding aspects, cardinal direction assignment, route classification, corridor studies, interfacing existing and proposed routes, etc. Results of these studies (which may be national, provincial or state, regional, or municipal) are used for progressively more detailed development of highway system components, such as freeway or arterial highway segments and interchanges, until the facility is finally designed, constructed, and put into operation. The importance of highway system planning in guide-sign legend development cannot be understated, because many of the inadequacies in guide-sign legends are seated in deficient system planning procedures. For example, inconsistent classification leads to difficulties in determining interchange form and route continuity; the choice of destination names or control city may not satisfy the through driver; incompatible cardinal route direction and actual direction may cause driver confusion; a change in route number along a rural facility that penetrates an urban area is confusing.

Rational aspects of highway system planning that are consciously or unconsciously appreciated by the driver include a network of routes that has cardinal direction association; a hierarchy of numbered route systems that reflect national, regional, and local routes; and a hierarchy of control destination names that reflects national, regional, and municipal destinations and is used in conjunction with numbered routes. Official highway maps reflect the above and form the basis of pretrip and enroute planning by a large number of drivers who have a reasonable level of driving experience. For example, the existing system in the United States is readily perceived by successive study of a map of the Interstate highway system, an official highway map of a state, and (finally) a street map of an urban area within the state.

This overview is necessary for the highway engineer. It provides an appreciation of the broad picture and enables him or her to view progressively smaller components of the system with an understanding of how everything fits together, and it is a necessary prerequisite for the succeeding steps in the design process. This overview is comparable to the use of high-level aerial photos first and then progressively lower-level photos in a route location and design assignment.

The second major category of guide-sign legend and location deficiencies can be associated with a lack of coordination between the guide sign and the roadway.

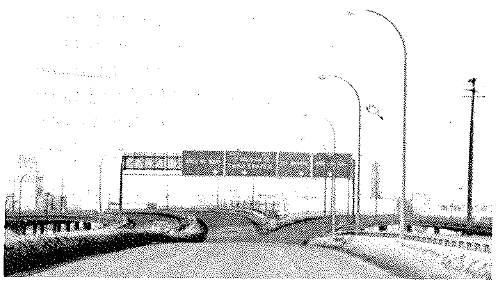
Function Planning and Design: The Logical Place for Guide-Sign Development

Lack of coordination of guide signs with the roadway result from a number of interrelated design deficiencies, such as lack of horizontal and vertical alignment coordination (sight losses and kinks), lane balance, route continuity, provision of basic number of lanes, lack of uniformity in successive interchange ramp configurations, and use of obsolete ramp geometrics and interchange types (5, 6). Other perceptual problems are more subtle and involve driver expectancies at all levels of the driving task: tangential exit ramps, dominant vertical elements adjacent to the roadside that create a wrong expectancy with respect to direction of major route (such as utility pole lines), large openings in the landscape at the end of long tangents, and railway-highway grade-separation structures parallel to the roadside (7).

Figures 1 and 2 illustrate several problems not anticipated even with careful design. These examples are drawn from an urban freeway project with which I was

Figure 1. Example of guide-sign panel design and location that need improvement: site 1.

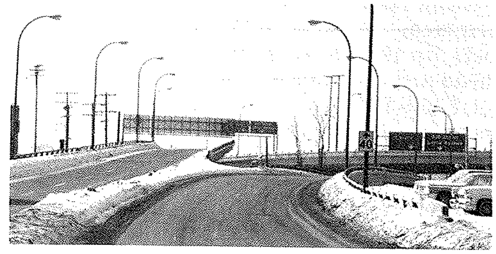
a



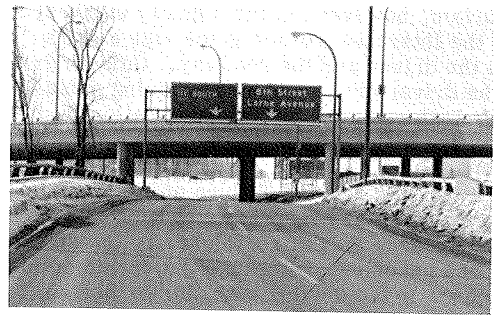
b

c

Figure 2. Example of guide-sign panel design and location that need improvement: site 2.



a



b

associated. In Figure 1a, the view at night, note the obstruction of the panel at far right by the light standards and the problems caused by the position of the lighting. Figure 1b, the view in daylight, raises questions of whether relocation upstream or a diagrammatic alternative would improve effectiveness and whether the arrows on the two right-hand panels are ineffective. The closer view in Figure 1c, from about 100 m (330 ft), indicates that there are problems with the design of the route shields and that the THRU TRAFFIC sign is not meaningful. Figure 2a shows how signs for one roadway, across the parapet, may be read from another roadway and that down arrows may be meaningless around a curve. In Figure 2b, the down arrow on the left panel does not relate to the view of the roadway, the route shield is too small, and the display might better have been placed on the overpass structure in the background.

Many of these problems can be detected in the functional planning and design stage; this stage is therefore the most effective time to develop the guide signs for a segment of a highway facility. Figure 3 is a simplified flow diagram illustrating the relationship of functional planning and design to highway system planning and detailed design. Guide sign planning, design, and evaluation are depicted as a loop that uses the results of functional planning studies. Figure 4 is an expansion of the guide-sign task, commencing with preliminary guide-sign development and proceeding, step by step, to the preparation of contract documents.

Description of the Framework

The functional drawings, in plan and profile, provide reasonable base drawings for the superposition of proposed guide signs located by station in the plan and shown adjacent to the roadways in each direction of travel. Graphics of the sign panels should not be too detailed at this time, because the major question to be addressed is related to whether the proposed signing will work and whether the proposed geometrics can be signed by using recommended signing practices (8). Alternative signs for particular locations should be considered and displayed.

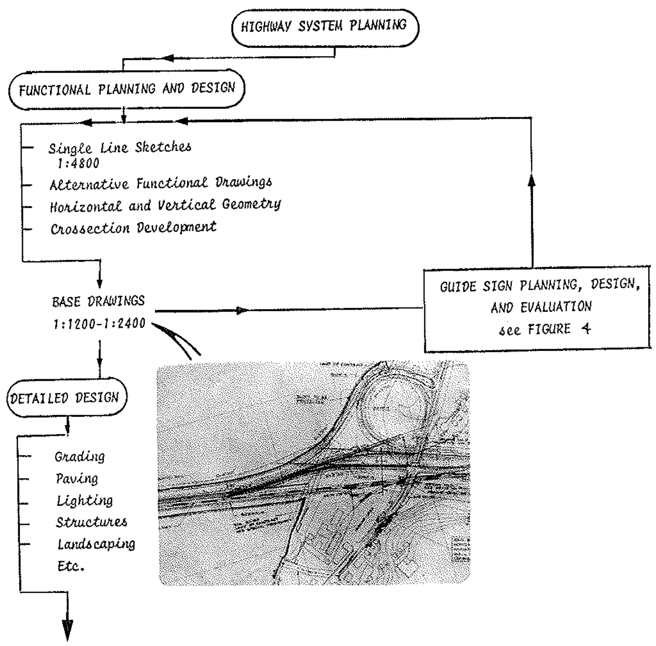
The designer then is in a position to acquire feedback by means of a design team review, in which other engineers and technicians involved in functional planning, system planning, and traffic engineering review separately, and as a group, the preliminary signing plan. Through the interaction of these individuals, revisions and adjustments can be readily made and a basic signing drawing agreed on, subject to detailed study of location, graphics, and coordination with other detail design tasks. The sign designer is then in a position to study in detail each sign panel and its alternatives by using a variety of signing principles and procedures recommended and described by King and Lunenfeld (9). Use of perspective plots in addition to these procedures provides the third dimension; the proposed guide sign is shown accurately within the perspective plot from the driver's eye. The designer is able to relate the proposed sign to the roadway much as the driver will in traveling along the proposed (or existing) highway facility.

USE OF PERSPECTIVE PLOTS

Perspectives of highway facilities, produced by means of a computer-driven plotter, have been used by designers to test for general visual quality of a proposed facility, detailed studies of horizontal and vertical alignment, aesthetics of bridges and overpasses, and safety-related improvements associated with fixed objects and modification to the roadside and median (4, 10, 11). Extension of this capability or tool for the study of guide signs was undertaken by using the perspective plotting program HWYPPLOT (11).

As Figure 4 showed, there are two main ingredients of a realistic perspective plot with proposed or existing guide signs superimposed: the first is an accurately scaled guide-sign mock-up; the second is knowledge of scale within the perspective at the location of the sign panel display.

Figure 3. Functional relationship of guide signs in the design process.



Scaled Guide-Sign Panel Mock-Up

Green card stock and pressure-sensitive white lettering are necessary to produce an accurate replica of a sign panel display. The dimensions of the Helvetica medium-letter alphabet and those of the series E alphabet used for guide sign legend are very close. Since there is an abundant selection of upper- and lower-case letters and numerals, a designer can use the normal range of letter and numeral heights found on guide-sign panels at a predetermined scale. Scales of 1:30 to 1:50 are common for detailed drawings of guide-sign panels, and experience has shown these scales to be suitable for the panel mock-ups. The legend of each panel is designed with proper letter spacing, interline spacing, and edge and border distances, and the white transfer lettering is overlaid on green card stock. Then, 35-mm slide or color film negatives of the mock-up form the basis for subsequent rear projection or direct mounting of prints of scaled guide-sign panel mock-ups on the perspective.

It should be noted that the scaled original mock-ups can be used for preparation of detailed drawings that form part of the contract documents for fabrication of sign panels (Figure 4).

Figure 4. A framework for guide-sign planning, design, and evaluation.

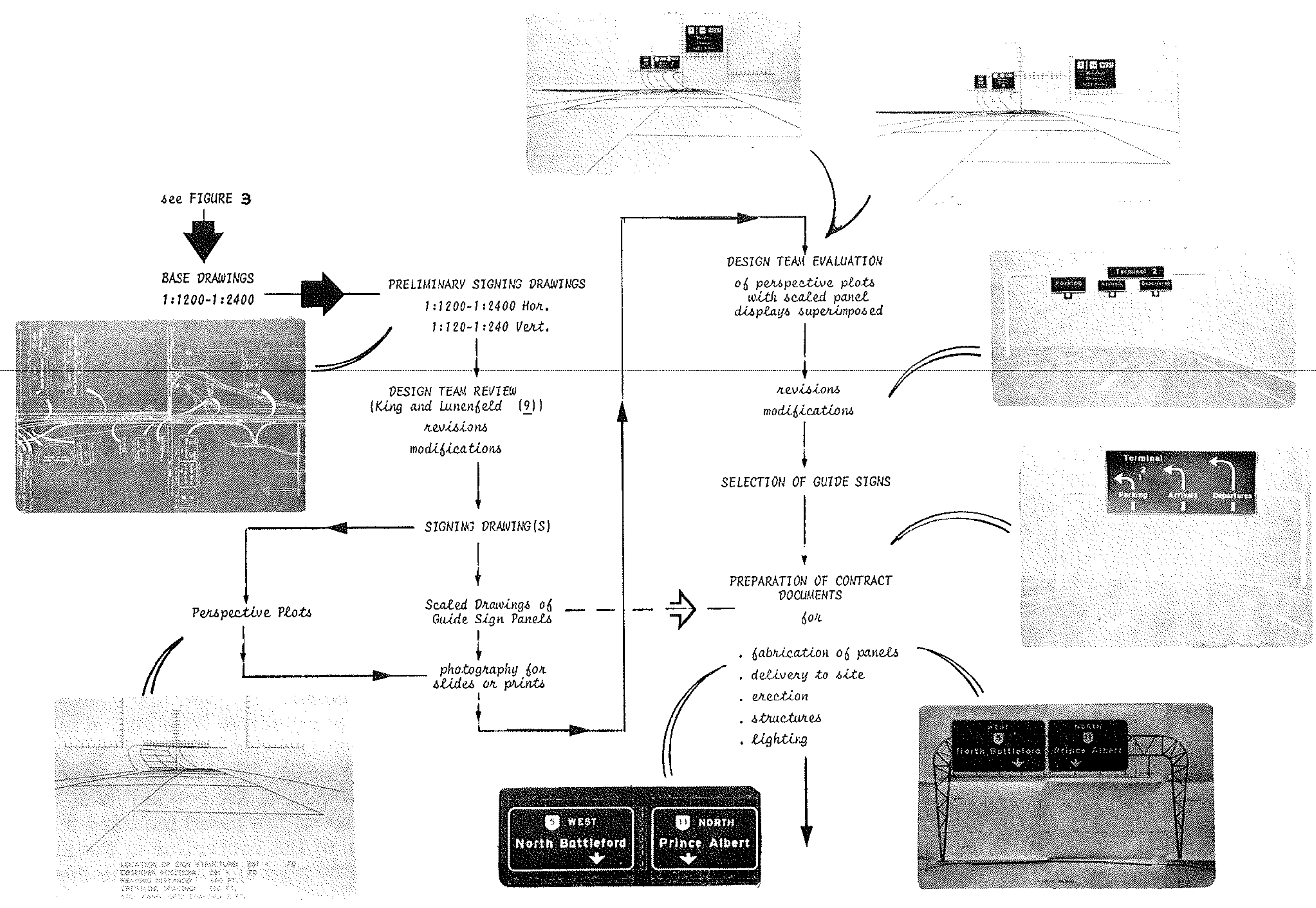


Figure 5. Typical perspective plot with shoulder-mounted and overhead guide-sign grids.

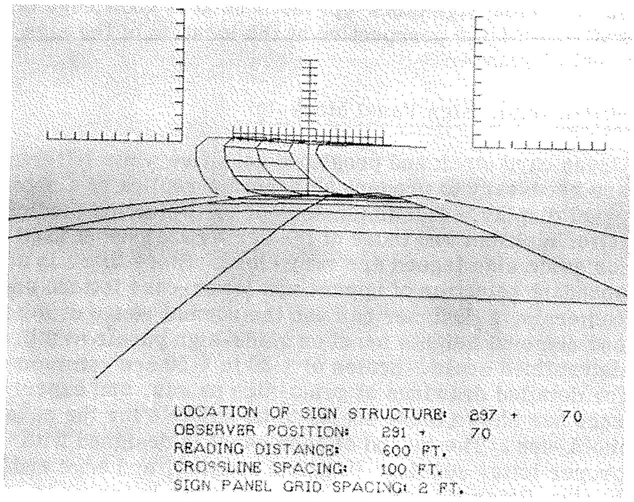


Figure 6. Perspective plot with exit-direction and gore signs mounted on sign grids.

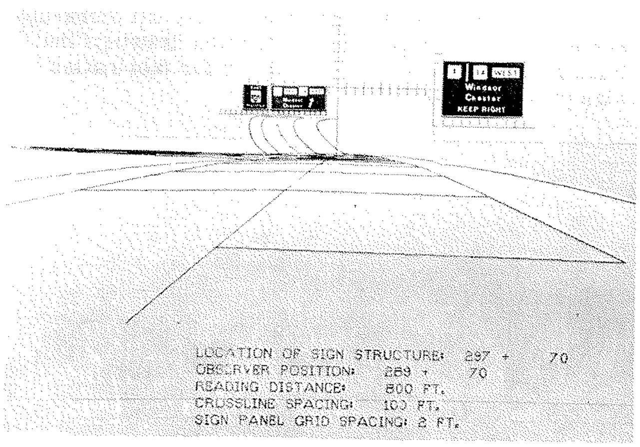


Figure 7. Functional plan of interchange.

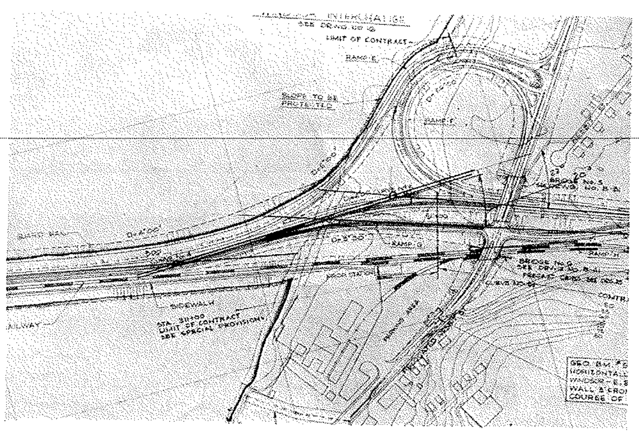
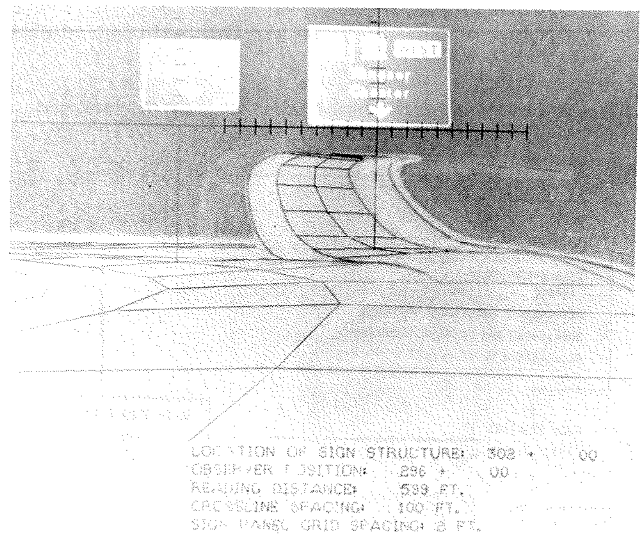
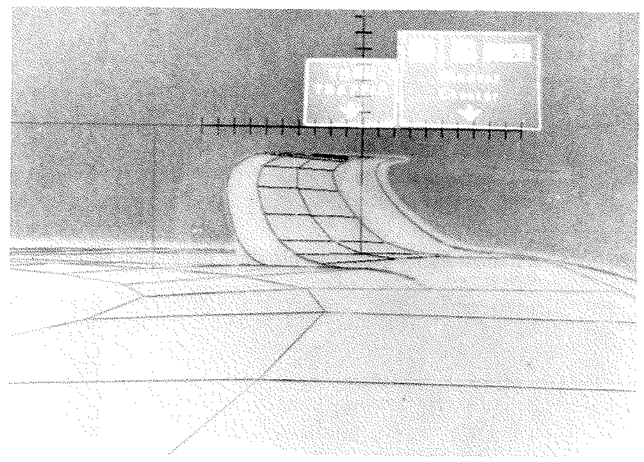


Figure 8. Existing gore sign installation and rear-projected guide-sign mock-ups.



a



b

The Perspective Plot

Accurate display of a guide-sign panel in a perspective of the highway facility requires knowledge of the horizontal and vertical scale at the location of the sign. A subroutine (12) was written for use with HWYPPLOT (11) to create, within the perspective, plotted grids that have a nominal size of approximately 0.67 m (exactly 2 ft).

Left and right shoulder-mounted sign grids, as well as overhead sign grids, can be specified separately or in combination for a single sign location or multiple sign locations that will appear from a particular driver-observer position. Figure 5 illustrates a perspective plot with sign grids.

The origin of the overhead grid is 5 m (16 ft) above the roadway centerline. Shoulder grids are located 2.5 m (8 ft) above the centerline and 6.7 m (22 ft) to the left or right. Selection of these origins leaves a portion of the grid visible after properly scaled guide-sign panel mock-ups are superimposed, because in practice a sign panel will normally be located away from the grid origin. Shifts in horizontal and vertical sign position can then be accurately studied.

Figure 6 illustrates a perspective with scaled mock-ups superimposed on the guide-sign grids. The arrangement of panels shows a shoulder-mounted exit-direction sign located 7.3 m (24 ft) right of the centerline of the roadway and the lower edge of the panel mounted 3 m (10 ft) above the elevation of the centerline at the sign position. An overhead gore sign appears in the distance and is mounted 5.5 m (18 ft) above the roadway centerline at the gore. A grid for an overhead exit-direction sign is also shown.

Use of the Perspective Plots to Evaluate Guide Signs

The following discussion is intended to expand on the use of the perspective plots within the framework of the process shown in Figure 4.

Knowledge of scale within the perspective allows the designer to mount properly scaled prints of panel mock-ups and, if 35-mm slides of the mock-ups are available, they may be rear-projected on the perspectives if the plot is made on a translucent paper.

By using these "complete" or composite perspectives as originals, the designer can rephotograph to produce a second set of slides or prints. These slides make it possible for a group or design team to review a proposed sequence of signs by projection on a screen. Panel shifts, sign location shifts, variation in panel legend graphics, and other problems can be studied with the assurance that the sign displays have been accurately depicted. Most important, however, is the added advantage of including the third dimension of the highway in the procedure where subtleties of geometry in three dimensions can be related to the proposed formal guide-sign information. Signs must relate to the view of the road beyond in order to confirm driver expectancy and provide positive guidance (13).

Up to this stage, the review of proposed signing is undertaken by highway engineers. Obviously, they are relatively biased evaluators and some degree of uncertainty remains, even though the resulting signs are based on sound design and operations principles.

A further test may remove some uncertainty and provide a more objective basis for acceptance or revision. A sample of drivers not involved with design of the facility would, most likely, be readily available to the design team from within the agency itself—clerks, stenographers, administrators, and technicians.

The perspective plots described above would form the basis for testing by the driver sample. While the design review team would not be bothered by the "raw" perspective plot, the test sample of drivers would probably find an embellished perspective more meaningful and realistic. Roadway color, texture, or joint lines, centerlines, edgelines, utility poles, and structures may be necessary to create a sense of realism for proper response evaluation. Tests could be designed to measure legibility, reaction time, and graphics.

ILLUSTRATIVE APPLICATIONS

An Awkward Curvilinear Exit Ramp

The first example concerns an exit ramp for an interchange of a two-lane, undivided highway that has grade-separated entrance and exit ramps. A portion of the functional plan is shown in Figure 7. Eastbound roadways are controlled by a causeway and gate structure. A rail line is located parallel to the highway. Restrictive geometry exists, largely because of the railway and causeway, in the form of 4° curves at each end of the causeway. Back-to-back 4° reverse curves exist in the interchange area.

At the advance reading position, the main line appears to have a slight kink to the left and then sweeps upward to the right. In reality, the main line curves left and the ramp continues on more or less the same tangent as the causeway approach roadway (see Figure 8). It is not until the driver is 100 m (300-350 ft) from the exit-direction sign that the main line and ramp begin to form in the driver's field of view. The main-line direction is not distinct because of its flat vertical alignment and the relatively sharp 4° curve.

The questions to be addressed here are (a) how realistic are the series of perspectives produced with mock-ups of the existing guide sign installation and (b) how might the existing guide signs be altered to improve the information presented to the driver? The reader will have to judge these questions in the light of the following illustrations and commentary.

Figure 8a projects the existing gore sign installation on the perspective plot at about 200 m (600 ft). Note the position of the left-hand panel. Since the main line is an undivided highway, the eastbound through-lane arrow appears to point downward to the westbound, or oncoming, traffic lane. Figure 8b illustrates the effect of moving the sign panels together and shifting them to the right by approximately 4.3 m (14 ft). The entire sign display now appears to lie directly above the approach roadway, and a better visual association of sign panel display with the highway results.

Figure 9 illustrates a more conventional (8) overhead gore sign installation in which the obsolete THRU TRAFFIC message is replaced by a distinctive through-traffic route shield and destination. Changes in lateral sign position are illustrated. In 9a the exit-direction arrow is closely associated with the ramp appearing below; 9b and c show the loss of association with a shift to the right. The sign positions shown in 9b and c seem to relate best to the approach roadway that appears below the sign display.

Figure 10 indicates a diagrammatic gore sign alternative. The importance of arrow geometry as it relates to the roadways below is evident. Figure 10a illustrates a first attempt; 10b and c show second and third trials for a diagrammatic arrow (shown in black) more closely related to the roadway below the sign panel. The sign appears to be located properly with respect to the approach roadway below the guide sign.

Airport Access Roadway

This example is concerned with a major airport's access roadway, the guide signs of which were the subject of a reaction-time and glance-legibility study by Dewar, Ells, and Cooper (14). The location of, and format for, advance guide signs for approach roadways to one terminal are illustrated here. There is a three-lane approach that splits into eight lanes immediately beyond a 90° turn. This creates an extremely short reaction time for drivers who find themselves in the wrong lane, not to mention the awkward visual appearance and relationship of this sign to the roadway below at a normal reading distance.

Properly scaled prints of the guide-sign mock-ups were positioned on the perspective sign grids. The alternatives in Figures 11b and c seem to be more effective in transmitting the message than the one proposed by the agency, 11a. Note the apparent increase in effect of the down arrows in 11b and c. Obviously the alternatives shown would be more costly but, on the other hand, they are probably more effective. (By comparing the unembellished perspectives shown in the first example, one can appreciate that the addition of pavement texture and pavement markings greatly improves the realism of the perspective.)

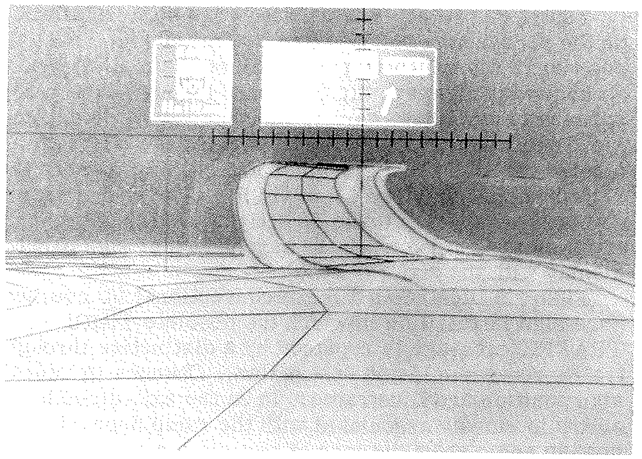
Other variations of alternative guide signs are illustrated in Figure 12, as well as variations in the driver's lane position. The diagrammatic alternative illustrated in 12c appears to relate better to the roadway and to be easier to read than the more conventional alternatives.

CONCLUSION

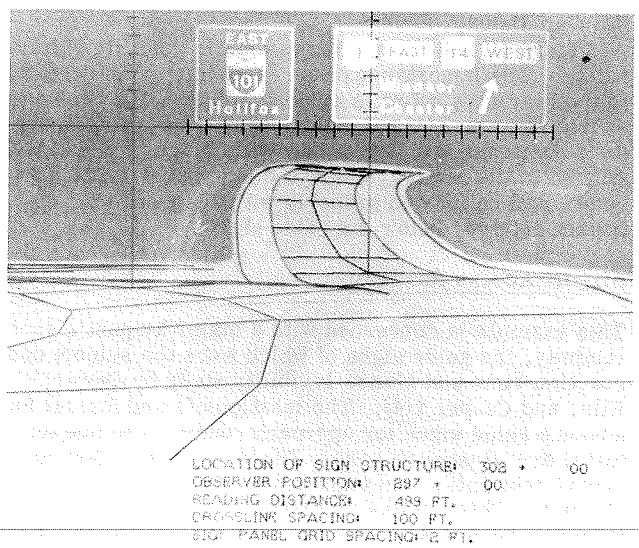
The usefulness of the perspective techniques for the evaluation of proposed guide signs and their position, legend, and graphics appears to hold promise and should be considered for use by highway agencies in new and reconstruction projects, particularly in locations that have extreme geometrics and in known problem spots. One positive feature of the use of the perspective plot is that it is an accurate and therefore objective portrayal of the roadway and sign, not an artist's perspective that is subject to individual artistic interpretation.

The illustrations show the connection between functional design and guide-sign position and, therefore, provide a strong argument for linking the functional design and guide-sign design processes. In the case of the tangential exit ramp, minor modification of horizontal and vertical alignment would have served to create a more positive guidance situation for the eastbound

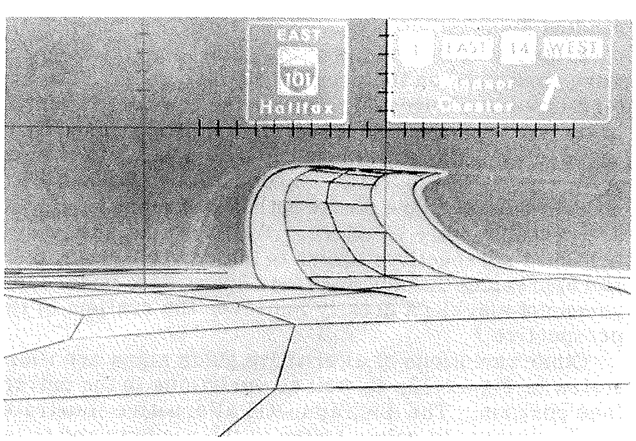
Figure 9. Alternative rear-projected gore sign panels and mock-ups.



a

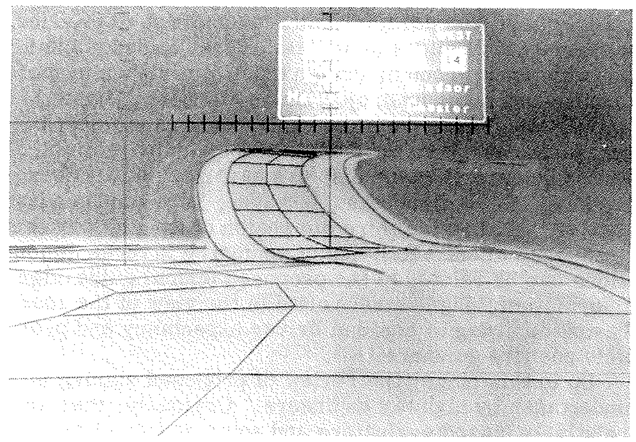


b

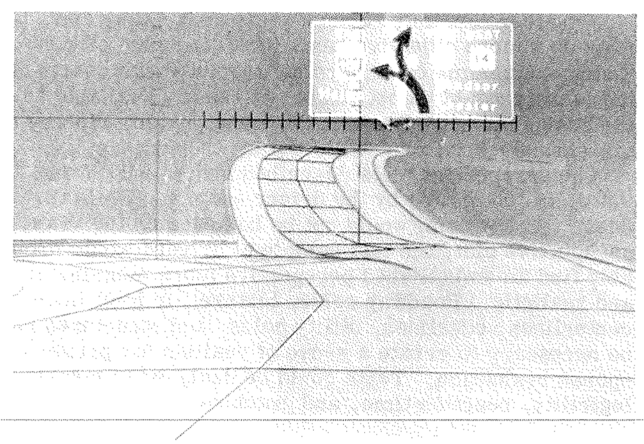


c

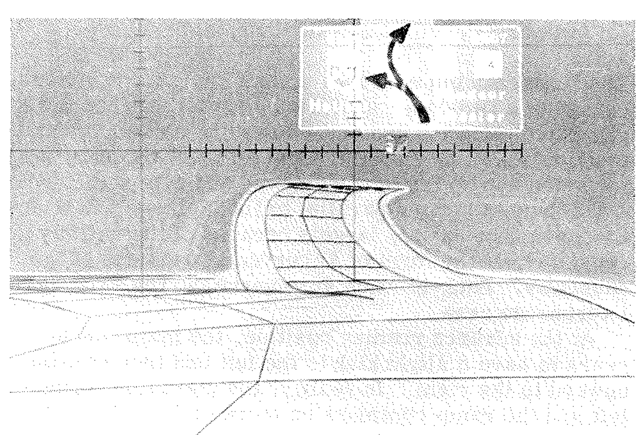
Figure 10. Diagrammatic gore sign alternatives.



a

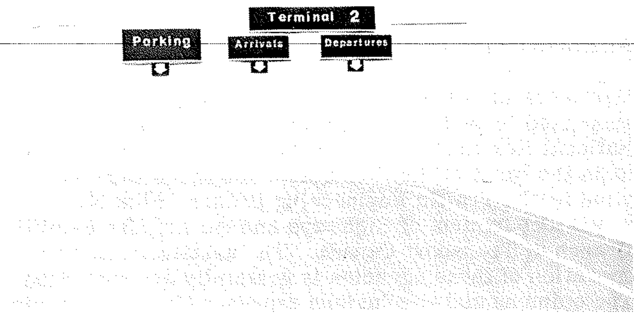


b

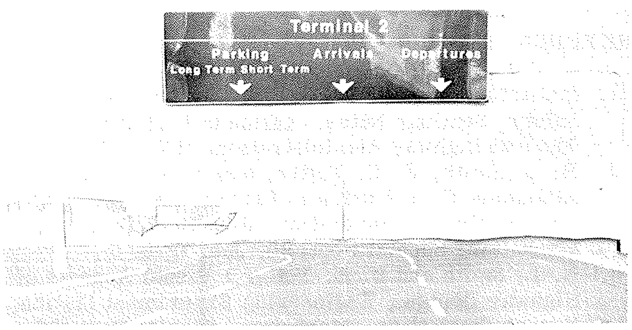


c

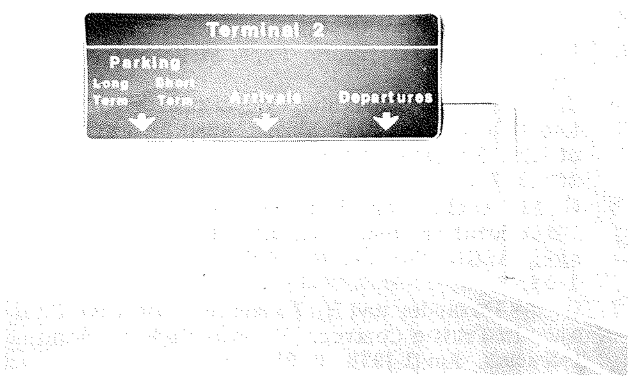
Figure 11. Existing and alternative advance signs.



a

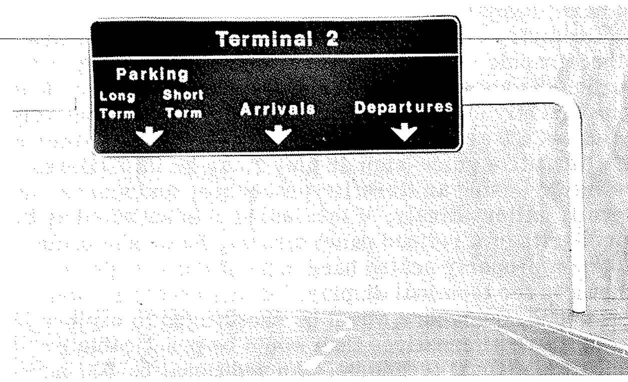


b

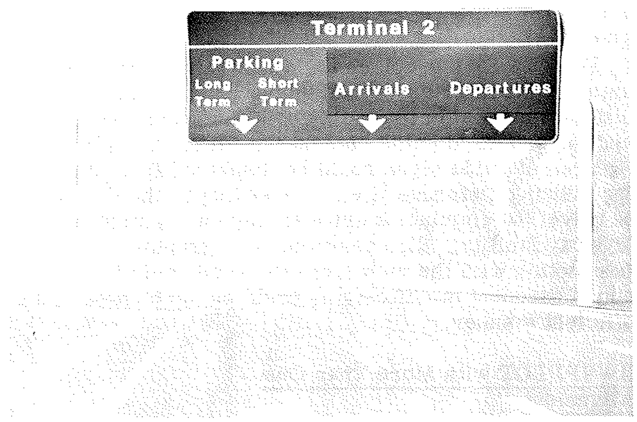


c

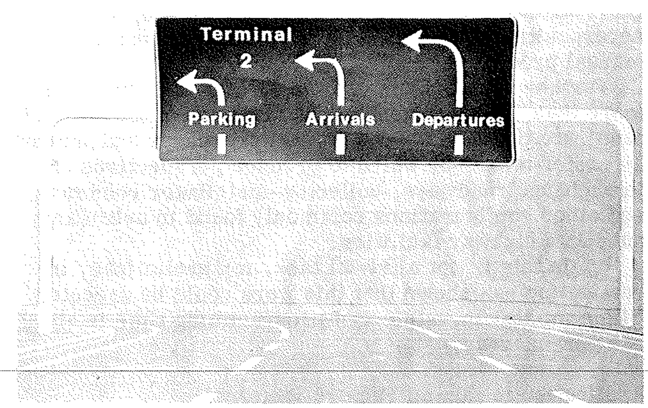
Figure 12. Alternative advance signs displayed from three driver positions.



a



b



c

driver; an overhead exit direction sign is better than the shoulder-mounted sign, and a diagrammatic gore sign might have merit.

The example of the airport access roadway provides definite evidence of a needed sign position shift, as well as a realistic view of a proposed sign and more visually pleasing alternatives thereto. The effect of embellishing perspectives with the addition of pavement markings and pavement texture is apparent.

Finally, the perspective view provides the designer with what Forbes (15) has referred to as "an intangible communication from the roadway and the general aspect it presents the motorist . . . the complex interaction that may subtly, yet profoundly, influence motorists as they move along any given section of roadway." Further experiment and research, combined with practical use of the technique by highway agencies, is warranted and practically feasible.

SUGGESTIONS FOR FURTHER RESEARCH AND DEVELOPMENT

Should the procedures, techniques, and output described and illustrated here prove to be useful, the following suggestions should be explored with the objectives of minimizing manual aspects of the procedure and providing maximum incentive to the designer to explore a great number of sign display alternatives.

Toward an Automated, Scaled Guide-Sign Mock-Up

The most expensive aspect of the technique described is that associated with preparation of scaled guide-sign mock-ups. Even with substantial expenditure of time in the preliminary planning stage (see Figure 4), not many alternative sign legend and graphic possibilities can be practically studied; this is particularly evident in the graphics associated with diagrammatic signing. Recent developments in low-cost, interactive, computer-driven cathode-ray-tube (CRT) graphics terminals (16) appear to hold promise.

Geometry of alphabets, route shields, arrows, and other graphic variables (such as interline spacing, border distances, and letter spacing) can be stored and subsequently manipulated by the designer, interactively using a CRT graphics terminal. Thus, a large number of alternative guide-sign displays may be investigated, refined by using an iterative procedure, and stored for recall. Alternatively, if the designer is satisfied as to the merits of a refined panel display, he or she could obtain a properly scaled hard copy of the sign panel shown in the terminal display. If this capability were developed, designers would be encouraged to explore many more alternatives than would be practicable by using the manual technique. An additional benefit may lie in further research and investigation of the details of sign design: "optimum arrangement of message; advantage of symbolic or schematic signing; ... letter design details" (9).

Use of a CRT graphics terminal to interactively design and store guide-sign displays leads to the possibility of being able to display both the perspective plot and the proposed guide sign on the CRT. Complete sequences of guide signs could be displayed from decreasing viewing distances (i.e., proceeding in the direction of travel) to simulate a drive through a segment of a highway facility. Sign position, sign graphics, and their interaction with the view from the road could be studied, and subsequent modifications could be performed in an efficient manner.

HWYPPLOT with More Than One Alignment

The examples have described how both ramp and mainline roadway perspective plots are created by using a superposition of separate plots—a time-consuming, manual procedure.

Further development of HWYPPLOT to provide for more than one alignment in one direction of travel would eliminate the need for the manual procedure and provide an easier means by which to produce perspectives of interchange roadways, collector-distributor roadways, and other configurations commonly found in urban and suburban highway facilities.

While this is not a trivial task, my preliminary investigation has shown that this work could be executed with a modest investment of programming time in about one person-year.

Application in Human Factors Research

The preceding suggestions are essentially from the viewpoint of a highway engineer involved in planning, design, construction, and traffic operations.

A potential application of the techniques described is in the area of human factors research. In particular, would study of driver expectancy at the guidance and navigation level be enhanced by using perspectives with both formal information and highway geometry displayed? Would animation, using successive perspective plots, be useful? Would perspectives, as described herein, be a superior tool in glance-legibility and reaction-time studies (14)? Finally, would the techniques and products suggested be suited to study of the psychophysiological elements of visual syntax in a dynamic driver-oriented context? See, for example, a discussion of sign composition by Donis (17), which bears directly on the graphic consequences of message units and their position within a composition (or sign).

I hope that our colleagues in human factors engineering research will address themselves to these questions.

ACKNOWLEDGMENT

Development of the modified HWYPPLOT program was supported in part by my operating grant from the National Research Council of Canada. Design data that form the basis of the illustrative examples were supplied by the Traffic Engineering Branch of the Nova Scotia Department of Highways and the Airport Facilities Branch of Transport Canada. The assistance and cooperation of these agencies is gratefully acknowledged. The opinions and conclusions expressed herein are mine and do not necessarily reflect the views or policies of the cooperating agencies.

REFERENCES

1. Institute of Traffic Engineers. Dynamic Design for Safety: Seminar Notes. Office of Highway Safety, Federal Highway Administration, 1975, 548 pp.
2. B. L. Smith, E. E. Yotter, and J. S. Murphy. Alignment Coordination in Highway Design. Engineering Experiment Station, Kansas State Univ., Manhattan, Special Rept. 90, Jan. 1970.
3. B. L. Smith and E. E. Yotter. Visual Aspects of Highway Design. Engineering Experiment Station, Kansas State Univ., Manhattan, Special Rept. 80, June 1968.
4. L. J. Feeser. Computer-Generated Perspective Plots for Highway Design Evaluation. Federal Highway Administration, Rept. FHWA-RD-72-3, Sept. 1971.
5. J. E. Leisch. Designing Operational Flexibility into Urban Freeways. Proc., 33rd Annual Meeting of ITE, Toronto, Ontario, Aug. 26-29, 1963, pp. 209-237.
6. G. M. Webb. Correlation of Geometric Design and Directional Signing. Journal of the Highway Division, ASCE, Vol. 84, No. HW2, May 1958, pp. 1-31.
7. G. J. Alexander and H. Lunenfeld. Positive Guidance in Traffic Control. Federal Highway Administration, April 1975, p. 25.
8. Manual on Uniform Traffic Control Devices for Streets and Highways. Federal Highway Administration, 1971, pp. 136-171.
9. G. F. King and H. Lunenfeld. Development of Information Requirements and Transmission Techniques for Highway Users. NCHRP, Rept. 123, 1971, pp. 123-125.
10. Texas Highway Department. The Roadway Design System: RDS. Federal Highway Administration, 1972.
11. D. J. Mellgren. Highway Perspective Plotting by Computer: A User's Guide for HWYPPLOT. Department of Civil Engineering, Kansas State Univ., Manhattan, M.S. thesis, 1972.
12. W. Mersereau and J. P. Dean. HWYPPLOT: Guide Sign Subroutine—A User's Guide. Department of Civil Engineering, Univ. of New Brunswick, Fredericton, 1977.
13. G. F. King and H. Lunenfeld. Urban Area Highway Guide Signing. NCHRP, Research Results Digest 52, 1972.
14. R. E. Dewar, J. G. Ells, and P. J. Cooper. Evaluation of Roadway Guide Signs at a Large Airport. Transportation Engineering, Vol. 47, No. 6, June 1977, pp. 19-23.
15. T. W. Forbes, ed. Human Factors in Highway

Traffic Research. Wiley-Interscience, New York, 1972, p. 130.
16. G. Dallaire. The CRT Computer Graphic Terminal: Indispensible Design Aid for Some Structural Engineers. Civil Engineering (ASCE), Vol. 46, No. 2, Feb. 1976, pp. 47-53.
17. D. A. Dondis. A Primer of Visual Literacy. MIT Press, Cambridge, MA, 1973, p. 20 ff.

Publication of this paper sponsored by Committee on User Information Systems.

**J. P. Dean was formerly with the Department of Civil Engineering, University of New Brunswick.*

Abridgment

Formation and Dissipation of Traffic Queues: Some Macroscopic Considerations

Panos G. Michalopoulos, Vijay B. Pisharody, and George Stephanopoulos, University of Minnesota, Minneapolis
Gregory Stephanopoulos, California Institute of Technology, Pasadena

Queue lengths at signalized intersections are state variables that are frequently used for optimal control of traffic signals particularly at high-volume intersections. In the absence of a reliable macroscopic model that describes queue lengths as a function of the demands, intersection capacity, and the control decisions, existing control schemes are using effective queue size rather than queue length for optimal control. Effective queue size is defined as the actual number of automobiles waiting for service on a particular approach to the intersection at an instant. Queue length, on the other hand, is the distance immediately behind the stopline within which traffic conditions are on the right side of the flow-versus-concentration curve (i.e., they range from congested to capacity).

Anyone familiar with traffic signal control problems will recognize that queue length rather than effective queue size is the parameter that should be used to describe the state of the intersection. This is because an efficient signal control policy should prevent upstream intersection blockage; it should effectively control queue lengths rather than queue sizes. This criterion for optimal operations is somewhat relaxed when traffic demands are relatively low and queue lengths do not pose any immediate threat to adjacent intersections.

In this paper, a rigorous mathematical model shows the evolution of queue length in time at any approach to the intersection as a function of the demands, the intersection capacity, and the signal control policy. Due to space limitations, the results of the simplest possible model are given here. More detailed (therefore, more realistic) models, along with stability analysis and numerical examples, can be found in Michalopoulous and Stephanopoulos (1). The mathematics of queue dynamics discussed here can be used for optimal control of traffic signals. It is believed that, in light of these results, the traffic signal control problem can be placed on a new, more realistic, and rigorous framework of analysis.

BACKGROUND

Consider distance L behind the stopline of a particular approach to the intersection without entrances or exits. Further, assume that L is long enough so that queues never extend beyond this section. Within L, the following equation of continuity applies (2):

$$(\partial K/\partial t) + (\partial q/\partial x) = 0 \tag{1}$$

where

K = density,
q = flow,
x = space, and
t = time.

Assuming that flow is a function of density, that is, $q = f(K)$, it can be seen that Equation 1 is a first-order partial differential equation in which K is the dependent variable and x and t are independent variables. Solution of this equation allows the estimation of density at any point in the time-space domain. Although space limitations preclude a detailed presentation here, the solution of Equation 1 (also known as the continuity equation) leads to these conclusions.

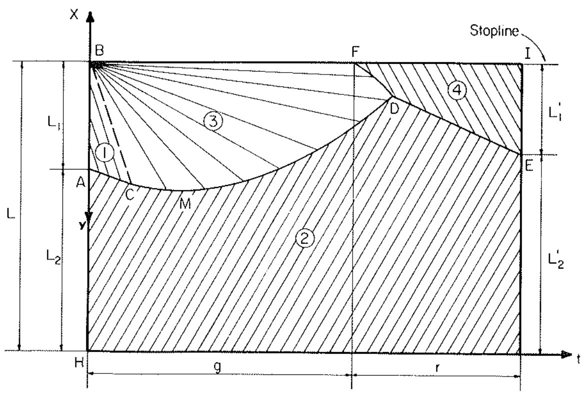
1. Density K is constant along a family of curves called characteristic lines or characteristics.
2. The characteristics are straight lines emanating from the boundaries—$x = L$ (stopline), $x = 0$ (end of the section), and $t = 0$—that have a slope tangent to the flow-concentration curve:

$$(dx/dt) \equiv h(k) = (dq/dK) \tag{2}$$

3. The characteristics carry the value of density at the point from which they emanate.

These findings suggest that density at any point that has the coordinates x and t is found by drawing the appropriate characteristic line emanating from one of the boundaries and passing through the point. The value of density at the boundary is carried through the characteristic line (i.e., it is maintained constant), and it corresponds to the density of the point of interest. If k_1 is

Figure 1. Queue length developments behind the stopline during a saturated cycle.



ZONE	DENSITY
1 & 4	$K = K_j$
2	$K = K_a$
3	$K_j \geq K \geq K_m$

the boundary value of the density of the characteristic passing through points x and t, then density at this point is also k_1 and the slope of the characteristic line is (dq/dK) evaluated at point k_1.

When two characteristic lines of different slope intersect, then density should have two different values at the point of intersection that is physically unattainable. The discontinuity of density at the point of intersection is explained by the generation of a shock wave moving upstream or downstream of the highway with speed u_w as given in the well-known equation first proposed by Lighthill and Whitham (2):

$$u_w = [(k_2 u_2 - k_1 u_1)/(k_2 - k_1)] = [(q_2 - q_1)/(k_2 - k_1)] \quad (3)$$

where

k_1 = upstream section concentration,
k_2 = downstream section concentration,
u_1 = speed of the upstream section,
u_2 = speed of the downstream section,
q_1 = upstream flow, and
q_2 = downstream flow.

THEORETICAL RESULTS

Based on this preliminary discussion, Figure 1 was prepared to show the shock wave developments behind the stopline. This figure assumes that at the stopline the discharge rates are those as suggested by Webster (3). Briefly, Webster's model suggests that some time is lost due to the driver's response time, acceleration, and deceleration. During the remaining green interval, automobiles are discharged from the intersection at saturation flow as long as a queue exists, and they depart with no delay after the queue dissipates. Of course, during the red interval, no automobiles are discharged. This modeling leads to the conclusion that the entire cycle can be divided into two intervals called effective green, denoted by g, and effective red, denoted by r, in Figure 1.

In Figure 1, it should be noted that along the x axis, point B corresponds to the stopline and point A to the tail end of the queue at the beginning of the effective green interval. Thus, t = 0 corresponds to the start of the effective green. Within AB, jam density and zero flow conditions prevail. Upstream of A and in the remaining portion L_2 of section L, automobiles arrive at an average flow rate q_a. Thus, density within L_2 is K_a. Assuming an average arrival flow q_a and density K_a during the cycle, then flow and density at the beginning of section L (point H) are q_a and K_a during the period g + r = c, where c is the cycle length. Finally, assuming that the cycle is saturated, capacity flow and density conditions q_m and K_m prevail at the stopline during g (i.e., from point F to point I).

From this definition of initial and boundary conditions, the characteristic lines emanating from t = 0, x = 0, and x = L were drawn. These lines are tangent to the flow-versus-density curve evaluated at the flow and density conditions corresponding to the point of origin. For example, within AB, the slope of the characteristics is negative, and it is the same as the tangent at the point $0_1 K_j$ of the flow-density curve. At point B, density changes instantaneously from K_j to K_m and, therefore, the characteristics at B fan out, i.e., they take all possible slopes from $(dq/dK)_0, K_j$ to zero.

The characteristic lines emanating from the boundaries divide the entire time-space domain $[0 \leq x \leq L, 0 \leq t \leq c]$ into four distinct zones of different flow and density conditions as shown in Figure 1. When the characteristics intersect, a shock wave is generated. The shock wave developments that result are shown in Figure 1. At the tail end of the queue, shock wave ACMDE is generated during the period of one cycle, and this line represents the trajectory of the tail end of the queue. Therefore, its vertical distance to the stopline represents queue length denoted as y(t). The slope of line ACMDE at any point represents the speed at which this shock wave (or, equivalently, the tail end of the queue) propagates upstream or downstream of the highway. At the end of the effective green, shock wave FD is generated and meets the tail end of the queue at point D. Finally, at the end of the cycle, the distance L_1' represents the final queue length of the existing cycle or, equivalently, the initial queue length of the next cycle.

It should be noted that, if the cycle is undersaturated, line ACMD intersects the stopline during green and point D falls on the stopline. After point D, queue length is zero. For the remainder of the green interval, automobiles depart without delay. At point F, the queue length starts increasing again linearly until the end of the cycle.

ANALYTICAL RESULTS

Each segment of line ACMDE and the coordinates of points C, M, D, and E can be described analytically. Analytical expressions are, of course, needed for the purpose of developing a control policy that restricts the queue lengths within predetermined upper bounds for each approach to the intersection. In order to obtain analytical results, however, one must assume a specific relation between flow and density or, equivalently, between speed and density. For simplicity, we assumed the linear speed-density model (4), but it should be noted that similar results can be obtained for any other model if the guidelines given here are followed. Because space limitations do not allow presentation of detailed proofs, only the final results are given. [See Michalopoulos and Stephanopoulos (1) for further details and for more elaborate models that take into account gradual transitions at point B (Figure 1) and capacities to the left or right of the theoretical capacity.]

Analytical expressions for the trajectory of the queue length (Figure 1) were obtained by using the following notation:

$y(0) = L_1$ = initial queue length at $t = 0$,
$y(c) = L_1'$ = final queue length at $t = c$,
$y(t)$ = queue length at any time point t,
g = effective green interval,
r = effective red interval,
$c = g + r$ = cycle length,
g_{min} = minimum green time required for undersaturation,
X_{IJ} = equation of any line IJ,
u_f = free-flow speed of the approach under consideration,
k_j = jam density of the approach under consideration,
q_a, k_a = arrival flow and density conditions,
x_i, t_i, y_i = coordinates of point i, and
y_{IJ} = equation of line IJ with respect to the y axis (Figure 1).

Thus, the following analytical expressions can be obtained by following the guidelines that are offered in Michalopoulos and Stephanopoulos (1):

$$X_{BC} = L - u_f t \quad (4)$$

$$X_{AC} = (L - L_1) - [(k_a u_f/k_j)] t \quad (5)$$

$$X_C = L - [k_j L_1/(k_j - k_a)] \quad (6)$$

$$t_C = [k_j L_1/u_f(k_j - k_a)] \quad (7)$$

$$y_C = [k_j L_1/(k_j - k_a)] \quad (8)$$

$$y_{CMD} = [u_f + h(k_a)]\,(t \cdot t_C)^{1/2} - h(k_a)t \quad (9)$$

where

$$h(k_a) = u_f[1 - (2k_a/k_j)] \quad (10)$$

$$t_M = [u_f + h(k_a)]^2\, t_C/4[h(k_a)]^2 \quad (11)$$

$$y_M = [u_f + h(k_a)]^2\, t_C/4[h(k_a)] \quad (12)$$

$$y_{FD} = u_f t - u_f(tg)^{1/2} \quad (13)$$

$$t_D = \{(t_C)^{1/2} + [u_f(g)^{1/2}/u_f + h(k_a)]\}^2 \quad (14)$$

$$y_D = u_f\{t_C - [u_f h(k_a) g]/[u_f + h(k_a)]^2 + [u_f - h(k_a)]/[u_f + h(k_a)]\,(g t_C)^{1/2}\} \quad (15)$$

$$y_{DE} = y_D + [u_f k_a(t - t_D)/k_j] \quad (16)$$

$$y_E = L'_1 = L_1 + [(k_a u_f c)/k_j] - [k_j u_f g]/4(k_j - k_a) \quad (17)$$

$$t_E = c \quad (18)$$

In an undersaturated cycle, the queue dissipates in time:

$$g_{min} = [(y_C/t_C) + h(k_a)]^2\, t_C/[h(k_a)]^2 \quad (19)$$

This is the minimum green time to dissolve the initial queue L_1. In such a cycle, the final queue length L_1' is independent of the initial L_1 and is given by

$$y_E = L_1' = (c - g)(k_a u_f)/k_j \quad (20)$$

QUEUE LENGTH STABILITY

The analytical relations between the initial and final queue developed in the preceding section of this paper can be used for stability analysis. Equation 17 can be rewritten as

$$L_1' = L_1 + b \quad (21)$$

where

$$b = (k_a u_f c/k_j) - [k_j u_f g/4(k_j - k_a)] \quad (22)$$

If c and g are given, b is constant, i.e., it is independent of the initial queue L_1. Thus, Equation 22 can be generalized for any cycle N and rewritten as

$$L_{N+1} = L_N + b \quad (23)$$

where L_N and L_{N+1} are the queues at the beginning of cycle N and N+1. Clearly, a steady state exists if $L_N = L_{N+1}$ or if $L_N = L_N + b$, i.e., if $b = 0$. Therefore, for steady state

$$(k_a u_f c/k_j) - [k_j u_f g/4(k_j - k_a)] = 0 \quad (24)$$

and to solve for g/c

$$g/c = [k_j g/4(k_j - k_a)] = \lambda \quad (25)$$

Since λ is positive, it is easily seen that if $g/c < \lambda$, the queue length at the end of the cycle will be growing for as long as this situation persists. Otherwise, if $b < 0$ or, equivalently, $g/c > \lambda$, the queue at the end of the cycle will decrease. It should be noted that Equations 23 and 25 are meaningful for saturated cycles, i.e., for green times less than the ones given by Equation 19. Otherwise, L_{N+1} is not related to L_N and it is given from Equation 20. A final note concerning the stability of the steady state is worthy of emphasis. As Equation 23 reveals, the steady state is metastable. If $b = 0$, a small variation of the demands will change the steady state to a nearby value that is also metastable. Therefore, the queue length at the beginning of each cycle will change according to the fluctuating values of b, which depend on the demands.

CONCLUSION

The approach taken here to a new and rigorous mathematical model and analysis that show the formation and dissipation of traffic queues at signalized intersections is macroscopic in nature in the sense that automobiles are treated as platoons rather than as single units. The shock wave analysis is fairly realistic for moderate to high demands where the shock waves can be clearly realized. In fact, it is this volume range in which the traffic signal control problem is more pronounced. The analytical results given here have been used for the derivation of a real-time control policy that minimizes total intersection delays subject to queue length constraints at isolated critical intersections (1, 5). It is important to note that the traffic queue dynamics noted here are valid for isolated intersections at which the assumption of constant average flow rates is fairly realistic. If traffic arrivals are affected by an upstream signal, the analysis becomes more complex. Extension of the basic theory to a system of two or more intersections in succession is not trivial due to the side streets and sinks or sources between the intersections. The fact that the input flows to the downstream intersection is a function of the output of the upstream intersection for any pair of intersections further complicates the analysis. We have also studied the queue dynamics for a system of intersections as well as the related optimal traffic signal control problems (1).

It should be recognized that the analysis presented here does not include all possible combinations of shock wave developments that can occur at signalized intersections. Rather, the ones that are most likely to occur are noted. However, analysis for the cases not discussed can be easily performed by following similar guidelines. A number of other cases and numerical examples are treated in Michalopoulos and Stephanopoulos (1).

ACKNOWLEDGMENT

We would like to acknowledge the support of the National Science Foundation and the Program of University Research of the U.S. Department of Transportation.

REFERENCES

1. P. G. Michalopoulos and G. Stephanopoulos. Development and Evaluation of Optimal Control Strategies for Large-Scale Urban Transportation Networks. Department of Civil and Mineral Engineering, Univ. of Minnesota, Minneapolis, Final Rept., 1976.
2. M. H. Lighthill and G. B. Whitham. On Kinematic Waves: Part 2, A Theory of Traffic Flow on Long Crowded Roads. Proc., Royal Society, London, Series A229, No. 1178, 1955, pp. 317-345.
3. F. V. Webster. Traffic Signal Settings. Her Majesty's Stationery Office, London, Road Research Laboratory Paper No. 39, 1958.
4. B. D. Greenshields. A Study of Traffic Capacity. Proc., HRB, Vol. 14, 1934, pp. 448-477.
5. G. Stephanopoulos and P. G. Michalopoulos. Real Time Control of Critical Intersections Based on Queue Dynamics. Proc., IEEE Conference on Decision and Control, San Diego, 1978, pp. 1018-1024.

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

Abridgment

Discomfort Glare: A Review of Some Research

Corwin A. Bennett, Kansas State University, Manhattan

Extensive research on discomfort glare as applied to roadways has been done in Europe by De Boer (1) and Hopkinson (2). However, discomfort glare research in the United States, as has most lighting research, has focused on interior applications. In the past few years, discomfort glare research conducted at Kansas State University under the sponsorship of the Illuminating Engineering Research Institute has been aimed primarily at fixed-roadway lighting. This paper surveys this research and briefly discusses its applications.

SINGLE-SOURCE STUDY

An initial major study was conducted with a single glare source.

Method

Putnam and others (3-5) did what might be considered pilot studies for this experiment by selecting the variables and the range of variation and by running a few subjects. The study summarized here is described in detail by Bennett (6).

Glare source size, position, and background luminance were independent variables; glare source luminance at the borderline between comfort and discomfort (BCD) was the dependent variable.

Glare source size was varied in five equal steps from 10^{-6} to 10^{-3} steradian. At arm's length, these vary from pinhole size to that of a quarter and were selected to cover the range of practicable sizes of the luminous parts of roadway luminaires. Source position varied in five equal steps from 0° (along the horizon) up to 30° above the line of sight (above the occluding angle of windshield tops). Background luminance was varied in five equal steps from 0.0034 cd/m^2 (0.001 footlambert) to 34 cd/m^2 (10 footlamberts). According to Kaufman (7), the former represents an overcast horizon night sky with the moon and the latter, an horizon sky on a very dark day.

Observers adjusted the luminance of a glare source to the BCD, which has been the common North American criterion for discomfort glare for about 30 years. The long instructions say that somewhere between a dim comfortable light and a bright uncomfortable one is a point of change or threshold called BCD. They further state that this threshold is neither the one that distinguishes pleasantness and comfort nor the one that distinguishes tolerable and intolerable. Rather, at BCD, if the glare source was made just slightly brighter, it would be uncomfortable.

The 97 paid participants in this study—primarily college students—adjusted (with replication) the incandescent glare source to BCD for 23 of the 125 possible combinations of the three variables in a confounded design. The observer looked at the pole of a 0.6-m (2-ft) radius hemisphere sitting on edge. By using a combination of a transformer and several neutral density filters (to reduce the voltage range and, hence, the lamp color variation), the glare source was set to BCD.

Multiple regression analysis, which involved some trial and error on transformations of variables, was performed. However, this work was largely guided by previously published research.

Results and Discussion

The selected multiple-regression model is as follows:

$$BCD = 200\,(L_B)^{0.3} \times e^{0.05A}/S^{0.6} \qquad (1)$$

where

BCD (in cd/m^2) = borderline between comfort and discomfort,
L_B (in cd/m^2) = luminance of the background ($r = 0.26$),
A (in degrees) = source angle above the line of sight ($r = 0.12$), and
S (in steradians) = source size ($r = -0.41$).

Comparisons of the results of this study to those of Putnam (3-5), Hopkinson (2), and others show essential similarities but with some differences in empirical constants. The BCDs in the Kansas State experiment tend to be higher than earlier results.

A later experiment with 24 new observers compared the source at 22.5° above the line of sight to one at 22.5° to the right of line of sight for the intermediate size and background luminance. No significant difference in BCD was found for these conditions.

Although the relative size of the correlations of the independent variables with BCD might be thought to indicate their relative importance, it is a function of the range of variation included in the experiment. Thus, originally there was an angle as large as 60°. This was eliminated because observers frequently could not achieve BCD at this high angle. However, to the extent that the ranges of variation are ecologically correct (i.e., they properly simulated the real world), the relative correlations are meaningful.

The coefficients of determination for the regression analysis show that more (0.54) of the predicted variation in BCD is associated with observers (individual differences) than the three independent variables (0.28). This is based on the inclusion in the multiple regression analysis of observers as dummy variables. That is, each observer was called "0" or "1" depending on whether he or she was currently in the equation or not (i.e., treated as a measure). The substantial coefficient of determination reflects the fact that there were large, consistent (over experimental conditions) individual differences in glare sensitivity or BCD. This may also be seen in the variation in the multiplier of the regression equation. This had a median of 217 (rounded to 200 in the equation), with a range from 0.52 to 8800—a ratio of almost 17 000:1. Wide individual variations in glare sensitivity have been long known to glare researchers and are discussed in the next section.

INDIVIDUAL DIFFERENCES

Although Fisher and Christie (8) found a relation between age and disability glare sensitivity, no other such finding and very little research on individual differences and discomfort glare have occurred. In 1972, 162 visitors during a Kansas State University open house made glare adjustments and filled out personal information forms. In 1974 and 1975, 199 open house visitors did the same. These were then interrelated. This work is described in more detail by Bennett (9).

Observers adjusted a 2.2×10^{-4} steradian (one degree) incandescent lamp at 0° with a 5.5-cd/m^2 (1.6 - footlambert) background luminance to BCD. Significant small correlations were found between BCD and age (-0.31), eye color (0.16), and indoor versus outdoor occupations (0.17). Older observers were more glare sensitive (had lower BCDs), and the empirical relation was found to be

$$\text{BCD, cd/m}^2 = 86\,000/\text{age, years} \qquad (2)$$

Light-eyed observers and those with indoor occupations have lower BCDs (are more sensitive). The size of one's residential community, a person's sex or hair color, whether glasses are worn, type of occupation, and the sunniness of one's residential community were not significantly or consistently related to BCD.

The extreme conditions that produce disability glare almost always produce discomfort, but the more moderate conditions that may produce discomfort need not produce disability. Discomfort may be viewed as a warning reaction that could lead to the avoidance of disability-producing conditions or worse. Similarly, although discomfort and disability apparently have different physiological mechanisms, it makes sense for older people who are more sensitive to disability to be more sensitive to discomfort effects. People with more melanin have darker eyes, and the melanin filters out light so that they are less sensitive to glare.

Lane (10) found that people who have recently done more detailed close work (are visually fatigued) are more sensitive to discomfort. Thus, one might expect indoor workers to be more sensitive also.

All in all, despite a few significant correlations, this research has been rather unsuccessful in accounting for differences in sensitivity to glare in terms of demographic variables. Consequently, M. M. Babiker in a 1977 master's thesis done at Kansas State University studied personality and attitude variables.

A Personal Enlightment Test, given after preliminary screening, was developed that consisted of 64 personality, attitude, and demographic items. This test had items such as, "I am in just as good physical health as most of my friends" (true or false) and, "In your opinion, can headlight glare be prevented" (yes or no). This test was given in 1977 to 101 open house visitors. The visitors adjusted to BCD in the hemisphere with a 1.76×10^{-4} steradian source at 0° with a 34-cd/m^2 (10-footlambert) background luminance. In addition, the observers made similar white-noise adjustments to the borderline between comfort and annoyance (BCA).

Stepwise regression analysis to predict BCD resulted in a 22-item questionnaire with an R of 0.81. The total sample was arbitrarily subdivided into halves. A new nine-item question set based on stepwise regression of one-half of the observers predicted the other set of data with an R of only 0.21 (nonsignificant at 0.05).

Thus, again, the attempt to predict who will be sensitive to glare has proven elusive. It seemed plausible that personality items (largely anxiety items) and glare attitude items would be useful. Ostberg and others (11) did find a significant correlation (0.53) between a test of neuroticism and glare sensitivity. Also, the correlation between BCD and BCA was a nonsignificant 0.19, thus discouraging the idea of general personal sensitivity to environmental stimuli.

APPLICATION TO ROADWAYS

In one sense, the Roadway Lighting Committee of the Illuminating Engineering Society is the customer for the ongoing research. The Standard Practice produced by this committee has not included discomfort glare in its considerations. Although European standards do consider discomfort glare, U.S. and Canadian engineers have encountered difficulties in tests of these procedures. Consequently, North American research is under way.

The single-source experiment gives some results that should be applicable to roadway lighting. In some cases, however, such as at interchanges, vast arrays of many light sources appear in the field of view. For a line of lights along the driver's roadway, two counteracting effects take place. The near lights are larger but higher above the line of sight. The far lights are smaller but closer to the line of sight. Research is

Table 1. Estimated BCDs for roadway lighting for a single source.

		BCD (cd/m² 000s)					
		Mounting Height at 3-m Lateral Distance (m)			Mounting Height at 9.1-m Lateral Distance (m)		
Longitudinal Distance (m)	Background Luminance (cd/m²)	9.1	11	12	9.1	11	12
27	0.0034	11	14				
	0.034	23	25				
	0.34	45	51				
	3.4	89	100				
	34	180	200				
55	0.0034	17	17	21	21	21	24
	0.034	34	38	38	41	45	45
	0.34	69	72	79	82	86	93
	3.4	137	140	160	160	170	180
	34	270	290	310	330	340	380
82	0.0034	24	24	27	27	27	31
	0.034	48	51	51	55	55	58
	0.34	96	99	110	110	110	120
	3.4	190	200	210	220	220	230
	34	380	400	410	450	450	480

Notes: 1 m = 3.3 ft; 1 cd/m² = 0.292 footlambert.
Empty cells indicate < 20° cutoff to the top of the windshield.

under way to study such multiple-source effects.

In the meantime, Merle Keck of Westinghouse and Ramkumar Viswanathan of Kansas State each did analyses that related some representative roadway-lighting conditions to those involving a single light source. Table 1 shows BCDs estimated from an analysis based on the regression equation from the single-source experiment.

It was assumed that a varying visible portion of a 0.13-m² (200-in²) cobrahead luminaire was mounted at 9.1 m (30 ft), 11 m (35 ft), or 12 m (40 ft). The driver's line of sight was assumed to be 1.2 m (4 ft) above the ground. The lights were assumed to be either 3 m (10 ft) or 9.1 m (30 ft) to the side of the driver's track. The BCD was examined at 27 m (90 ft), 55 m (180 ft), and 82 m (270 ft) longitudinally from the light.

In some cases, the light at closer distances was above the occluding windshield top. The BCDs may be appraised by observing that Viswanathan and I made a few luminance measurements of roadway lights in our neighborhood that ranged from 21 000 cd/m² (6000 footlamberts) to 86 000 cd/m² (25 000 footlamberts) (for mercury, high- and low-pressure sodium). If such an actual source luminance was viewed in a position where a lower BCD luminance was expected, one might expect at least half the observers to be uncomfortable. Thus, some analyzed conditions will be problems, some will not. Generally, most discomfort problems can be avoided by raising the mounting height. Other analysis is under way to figure out how to cope with skewed distributions of observers within conditions so as to specify various percentages of observers who would be discomforted.

REFERENCES

1. J. B. De Boer, ed. Public Lighting. Cleaver-Hume Press (Phillips Technical Library), London, 1967.
2. R. G. Hopkinson. Discomfort Glare in Lighted Streets. In Transactions of the Illuminating Engineering Society, Illuminating Society, London, 1940, p. 5.
3. R. C. Putnam and K. E. Faucett. The Threshold of Discomfort Glare at Low Adaptation Levels. Illuminating Engineering, Vol. 46, No. 10, 1951, pp. 505-511.
4. R. C. Putnam and W. F. Gillmore. Discomfort Glare at Low Adaptation Levels: Part 2, Off-Axis Sources. Illuminating Engineering, Vol. 53, No. 4, 1957, pp. 226-232.
5. R. C. Putnam and K. D. Bower. Discomfort Glare at Low Adaptation Levels: Part 3, Multiple Sources. Illuminating Engineering, Vol. 53, 1958, pp. 174-184.
6. C. A. Bennett. Discomfort Glare: Parametric Study of Angularly Small Sources. Journal of the Illuminating Engineering Society, Vol. 7, No. 1, 1977, pp. 2-15.
7. J. E. Kaufman, ed. IES Lighting Handbook, 5th ed. Illuminating Engineering Society, New York, 1972.
8. A. J. Fisher and A. W. Christie. A Note on Disability Glare. Visual Research, 1965, pp. 5, 565.
9. C. A. Bennett. The Demographic Variables of Discomfort Glare. Lighting Design and Application, Vol. 7, No. 1, 1977, pp. 22-24.
10. B. C. Lane. Lowering of Functional Photophobia Threshold with Daily, Substained Closework Accommodation. Paper presented at Optometric Academy Meeting, Birmingham, AL, Oct. 1977.
11. O. Ostberg, P. T. Stone, and R. A. Bensen. Free Magnitude Estimation of Discomfort Glare and Working-Task Difficulty. Goteborg Psychological Reports, Vol. 5, No. 15, 1975.

Publication of this paper sponsored by Committee on Visibility.

Economic Models for Highway and Street Illumination Designs

Richard W. Slocum and Daniel R. Prabudy, Research and Development Division, Ministry of Transportation and Communications, Ontario

A key issue in the field of illumination is energy conservation. At the same time, the application of economic resources should be optimized. For instance, to save energy, roadway illumination lamps can be replaced by more efficient lamps that provide the same light for less wattage. However, such energy savings may be offset by other cost elements. For these reasons, detailed cost calculations are needed to ensure lighting investment optimization. The cost-effectiveness of lighting systems can be established by the discounted total cost or annual equivalent cost models described in this report. These economic models allow various cost items, such as capital outlay, maintenance, and operational and en-

ergy costs, to be compared realistically by taking into account discount rates and inflation rates. Consideration of inflation is particularly important here because some costs, notably energy, may inflate at higher rates than other costs. The methods given in this paper provide for the inclusion of projected dollar cost inflation in the analysis. A computer program for the models has been developed, copies of which may be obtained from the author, and can be used independently or in conjunction with the overall illumination program. This computer program allows a comparison of different strategies by varying the input values of costs, as well as inflation and discount rates. The most effective design solution can then be identified and considered in the overall evaluation of the alternatives. The program should be a useful tool in selecting designs of light sources and other features of roadway lighting systems.

Efficient resource allocation is an ever-present objective of government spending. Cost-effectiveness modeling can assist in reaching this objective. Such models provide a cost comparison between various means of providing a given level of service. This paper describes methodologies for cost-effectiveness modeling of highway and street-lighting design alternatives. One can test different lighting system design parameters and their arrangements to obtain equal level of service (illuminance or luminance). The onus is on the lighting design engineers to test the parameters with different inputs and to compare how they satisfy a prescribed level of service. Because this area is discussed in Jung and Blamey (1), such comparisons are not included in this paper.

The key considerations in cost-effectiveness modeling are as follows:

1. Assumption of equal level of service for investment alternatives obviates the need for benefit estimation.
2. Cost elements are analyzed over a given time period or life cycle.
3. Inflated dollar or constant dollar figures are important.
4. Salvage value is considered to be insignificant (as a simplifying assumption).

The models discussed herein were developed because of the need to consider the effect of inflated costs (especially energy costs) on alternative highway and street-lighting designs. The rates of inflation are different for energy, material, and labor. The two models are the discounted total cost (DTC) and the annual equivalent cost (AEC). Both models will provide an economic cost comparison. Preference is given to the DTC method because it considers each year in the analysis period and does not rely on averaging some costs that may occur, for example, only every four years. The AEC method does allow the user to make more rapid calculations, albeit with somewhat rougher results than the total cost approach.

A number of costs are incurred in the provision of lighting. These costs change over time, some more so than others. The models presented here attempt to consider the effect of changing future costs and to discount them by some rate of discount. The rate of discount used can vary. One could choose, for example, the rate of interest on long-term government securities, the social time preference rate, or the opportunity cost rate of interest (2). Notwithstanding the rate chosen, a range of rates should be tested to determine the sensitivity of the decision to the selected rate. Note that the discount rate will often include an implicit factor for inflation.

Although the standard approach would use constant dollars with relative inflation, in actual experience inflation is rarely computed and included in cost-effectiveness analysis. The models presented here allow the reader to use projected inflation rates for each factor of cost, with no sacrifice in the relevance of the resulting cost comparison between lighting designs.

HIGHER-PERFORMANCE FACTORS

A given standard of illumination can be provided along a roadway by various design alternatives with respect to pole spacing, mounting height, lamp type, and luminaire type. A given lamp type has its own performance curve over time (lamp lumen depreciation curve) and its own burn-out rate (lamp mortality curve). Figure 1 (3) illustrates examples of these curves.

For instance, out of any given group of lamps, usually more than 10 percent will burn out during the first 16 000 h of operating time. This is equivalent to about four years of operation. The number of operating hours varies, depending on the type of lamp. During the same time, the lamp lumen output of remaining lamps may have decreased to a point where it gives illuminance near the minimum design level at the roadway surface. This usually occurs when the lumen output is about 85 percent of the initial lumens. At this point, a total group relamping of the lamps should be done.

Lamps burn independently of each other and at different rates. Their burn-out time may be weeks, months, or even years apart. Because it is not desirable to have even one lamp out at any given time, spot relamping must be carried out.

The amount of dirt that accumulates on the luminaire hardware surfaces greatly depends on environmental factors. The ambient category of several selected localities (4) is shown below:

Ambient Category	Description	Suspended Particulates ($\mu g/m^3$)
1	Clean, average for areas remote from pollution sources	0-150
2	Heavy traffic, light industrial	150-300
4	Moderate industrial, some smoke- or dust-generating activities nearby	300-600
8	Dirty, numerous smoke- or dust-generating sources nearby	600-1200
16	Very dirty, heavy smoke at luminaire elevation	1200-2400

Table 1 (4) shows luminaire dirt depreciation values for the ambient categories against cleaning intervals. Regular luminaire cleaning will contribute to higher illuminance on the road surface.

Given the above parameters and maintenance methods and by using known unit costs, the annualized and/or DTCs of alternative lighting systems can be calculated.

DISCOUNTED TOTAL COST

The following formulas describe the discounted total cost method (5). The general formula is

$$DTC = C_o + \sum_{i=1}^{n} C_i/(1+r)^i \qquad (1)$$

where

DTC = discounted total cost,
C_o = initial cost,
C_i = cost in year i,
r = discount rate, and
n = analysis period (years).

When constant dollars are used, DTC is the economic value of future expenditures expressed as if spent today. Since there is a benefit to delaying costs, future costs are discounted in computing their "present value".

The use of inflated dollars will also capture the relative resource value of alternative investments; DTC will approximate present value, depending on the discount rate used. If inflation is used, it must be included in all of the cost elements. If inflation is not used, the models revert to traditional constant dollar comparisons.

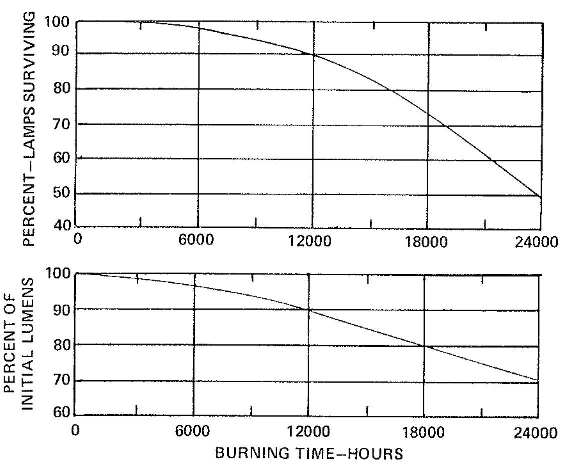
By using the general framework above, the following highway and street-lighting specific formula is presented:

$$\begin{aligned}
DTC = {} & [1000B_1(C_1 + C_4 + C_{11})/B_2 + 1000(C_2 + C_3)/B_2 + C_5 + C_6] \\
& + \sum_{i=1}^{n} [1/(1+r)^i](1+I_1)^i (B_1 B_3/B_2)(T_6 \cdot C_7 + 12C_8) \\
& + \sum_{i=1}^{n} [1/(1+r)^i] C_{10}(1+I_3)^i \\
& + \sum_{i=1}^{n} [1/(1+r)^i] B_{4i} [C_4(1+I_2)^i + C_9 \cdot T_4 (1+I_3)^i] \\
& + \sum_{\alpha=1}^{\alpha=n/T_1} [1000B_1/(1+r)^{\alpha T_1} B_2] [C_4(1+I_2)^{\alpha T_1} + C_9 \cdot T_3 (1+I_3)^{\alpha T_1}] \\
& + \sum_{\alpha=1}^{\alpha=n/T_2} [1000B_1/(1+r)^{\alpha T_2} B_2] [C_9 \cdot T_5 (1+I_3)^{\alpha T_2}] \qquad (2)
\end{aligned}$$

where

DTC = discounted total cost per kilometer of roadway;
B_1 = number of luminaires per pole;
B_2 = pole spacing (m);
B_3 = lamp wattage;
B_{4i} = number of burnt-out lamps per kilometer in year i for annual spot-relamping purposes;
C_1 = cost of one luminaire;
C_2 = cost of one pole;
C_3 = cost of one foundation;
C_4 = cost of one lamp;
C_5 = cost of equipment, wire, switching, and so forth per kilometer of roadway;
C_6 = total initial labor cost per kilometer of roadway;
C_7 = energy cost per kilowatt hour;
C_8 = demand charge per kilowatt per month;
C_9 = cost of labor and vehicle per hour;
C_{10} = miscellaneous maintenance cost per kilometer of roadway per year;
C_{11} = cost of one bracket;
T_1 = relamping period (years);
T_2 = cleaning period (years);
T_3 = relamping time per luminaire (h);
T_4 = spot-relamping time (h);
T_5 = cleaning time per luminaire (h);
T_6 = hours of operation per year;
I_1 = expected annual rate of inflation for electricity rates;
I_2 = expected annual rate of inflation for lamps or materials;
I_3 = expected annual rate of inflation for maintenance labor;
r = discount rate; and
n = investment life (years).

Figure 1. Examples of lamp lumen depreciation curve and lamp mortality curve (250-W Lumalux).



The above formula represents the discounted total cost per kilometer of roadway, which is the initial cost plus discounted total energy costs and the remaining operating costs throughout the investment life n, usually 20 years.

The initial cost includes materials and labor for installation of foundations, poles, brackets, lamps, luminaires, switchings, wirings, equipment, and so forth. Operating cost is cost incurred for energy and

Table 1. Luminaire dirt depreciation values for ambient categories against cleaning intervals.

Luminaire Category	Ambient Category	Cleaning Interval (years) 1	2	3	4	5	6	7
1 (open)	1	0.95	0.90	0.81	0.64	0.48	0.39	0.33
	2	0.89	0.80	0.64	0.40	0.29	0.19	0.13
	4	0.78	0.63	0.40	0.19	0.10	0.07	0.05
	8	0.55	0.37	0.18	0.06	0.04	0.03	0.02
	16	0.32	0.16	0.05	0.02	0.02	0.01	0.01
2 (ventilated)	1	0.97	0.94	0.89	0.79	0.70	0.62	0.53
	2	0.92	0.88	0.80	0.63	0.47	0.38	0.32
	4	0.86	0.78	0.62	0.39	0.28	0.19	0.13
	8	0.73	0.58	0.38	0.18	0.09	0.06	0.04
	16	0.50	0.32	0.16	0.05	0.03	0.02	0.02
3 (semi-sealed)	1	0.97	0.94	0.91	0.83	0.76	0.69	0.62
	2	0.93	0.90	0.83	0.71	0.58	0.46	0.39
	4	0.87	0.81	0.70	0.47	0.35	0.27	0.20
	8	0.76	0.65	0.44	0.26	0.15	0.09	0.06
	16	0.56	0.38	0.23	0.08	0.05	0.03	0.02
4 (tightly sealed)	1	0.98	0.96	0.95	0.91	0.87	0.83	0.80
	2	0.95	0.94	0.91	0.85	0.80	0.75	0.70
	4	0.91	0.89	0.84	0.75	0.67	0.58	0.50
	8	0.83	0.79	0.72	0.57	0.42	0.35	0.29
	16	0.67	0.61	0.49	0.31	0.22	0.15	0.10

maintenance. Maintenance includes materials and labor for spot and group relamping, cleaning, and miscellaneous repairs for items other than lamps. Pole repairs or replacement would be a function of the accident rate. For simplicity, only the labor inflation rate I_3 is included in miscellaneous maintenance C_{10}.

The number of burnt-out lamps in any given year is given by the lamp mortality curve of the particular type of lamp. Spot relamping requires a cumulative function within a group-relamping period to allow for the probability of the replaced lamps themselves burning out. Then the same cycle is repeated for the next group-relamping period.

Demand charge per kilowatt per month depends on the total demand of wattage per month in a certain area. For example, an area may require an estimated consumption of 100 MW energy per month; the power company must make that amount available monthly. The cost incurred for mobilization of equipment, labor, installation, and other operations varies according to the wattage requirements.

ANNUAL EQUIVALENT COST

The annual equivalent cost method uses the capital recovery factor to annualize initial costs (6, 7). This annualized initial cost, added to the annual operating cost, is the AEC for that particular year. The general formula is as follows:

$$AEC = C_0 \cdot CRF + C_a \tag{3}$$

where

$CRF = [r(1 + r)^n]/[(1 + r)^n - 1]$,
CRF = capital recovery factor,
n = investment life (years),
AEC = annual equivalent cost,
C_o = initial capital cost, and
C_a = annual operating cost for the reviewed year.

$C_o \cdot CRF$ gives the annualized initial cost that is fixed for every year of investment life. It is also known as the annual fixed charges for the initial cost. Annual operating cost C_a varies from year to year, depending on the actual cost incurred for the year in review. The initial annual operating cost is usually known at the time of investment and will be assigned to C_a. For the following years, an assumed annual inflation rate I is used to reasonably predict the annual operating costs. Applying I to C_a, Equation 3 becomes

$$AEC_y = C_o \cdot CRF + C_a (1 + I)^Y \tag{4}$$

where

I = general annual inflation rate;
y = year in review, 1, 2, ... n; and
AEC_y = inflated annual equivalent cost at year y.

The other notations are as previously defined. However, this general inflated annual equivalent cost (given in Equation 4) does not distinguish differing rates of inflation for the factors of cost.

By adapting Equation 4 specifically to highway and street lighting and considering the different annual inflation rates for energy, lamps, materials, and maintenance labor, the following formula results:

$$\begin{aligned} AEC_y = {} & [1000\,B_1(C_1 + C_4 + C_{11})/B_2 + 1000(C_2 + C_3)/B_2 + C_5 + C_6] \\ & \times [r(1+r)^n]/[(1+r)^n - 1] \\ & + [B_1 \cdot B_3(T_6 \cdot C_7 + 12C_8)/B_2](1 + I_1)^y + C_{10}(1 + I_3)^y \\ & + B_{4a}[C_4(1 + I_2)^y + C_9 \cdot T_4(1 + I_3)^y] \\ & + (1000\,B_1/B_2)[C_4(1 + I_2)^y + C_9 \cdot T_3(1 + I_3)^y]/T_1 \\ & + (1000\,B_1/B_2)[C_9 \cdot T_5(1 + I_3)^y]/T_1 \end{aligned} \tag{5}$$

where

$B_{4a} = (B_{41} + B_{42} + B_{43} + \ldots + B_{4T_1})/T_1$, and
$B_{41} \ldots B_{4T_1}$ = numbers of burnt-out lamps in year 1, 2 ... T_1.

If the group-relamping period T_1 = 4 years (which is usually so), $B_{4a} = (B_{41} + B_{42} + B_{43} + \ldots + B_{4T_1})/T_1$ becomes $B_{4a} = (B_{41} + B_{42} + B_{43} + B_{44})/4$.

The value B_{4a} is the average annualized number of burnt-out lamps per kilometer in a relamping period T_1. This number is assumed to be a fixed number of burnt-out lamps for each year of the relamping period. Since the pattern of lamp burnouts is the same for the following periods, it is also fixed for each year throughout the investment life.

Costs for group relamping and cleaning must also be annualized. The assumed cost of lamps and labor at the time of investment is averaged throughout the relamping period T_1. (See the last two factors of Equation 5.)

The value y can be chosen for any future year to test the effect of inflation on annual total cost.

COMPUTER PROGRAM

A computer program (Illum III) has been developed for use by the highway and street-lighting design engineer and may be used independently or in conjunction with Illum I and II. The Illum I program contains calculations of illuminance, luminance, and glare. The Illum II program is for preliminary analysis for efficient roadway lighting design. The overall program is called the Illumination Design Systems Program. By inputting the design and cost parameters, the user will be able to compare the cost-effectiveness of alternative designs and to test the sensitivity of the result under various assumptions with regard to inflation.

The Illum III program is in conversational mode and includes both DTC and AEC. It calculates the annual equivalent cost, computed at 10, 15, and 20 years. It is available on request for a nominal charge. A manual example problem for the DTC and AEC formulas is presented in the next section of this paper.

EXAMPLE PROBLEM

The example problem that follows for discounted total cost and annual equivalent cost formulas uses metric units and manual calculations. The problem illustrates the practical use of the formulas to determine current and future costs.

The input figures in this example problem have been chosen to be as close as possible to typical values. They are listed as follows:

B_1 = 2,
B_2 = 53.34 m (175 ft),
B_3 = 250 W high-pressure sodium,
B_{41} = 0.375,
B_{42} = 1.500,
B_{43} = 1.875,
B_{44} = 3.750 (see explanation below for B_{41} to B_{44}),
C_1 = \$150,
C_2 = \$300,
C_3 = \$100,
C_4 = \$30,

C_5 = \$15 000,
C_6 = \$750,
C_7 = \$0.025,
C_8 = \$2.50,
C_9 = \$30,
C_{10} = \$350,
C_{11} = \$50,
T_1 = 4 years,
T_2 = 4 years,
T_3 = 1 h,
T_4 = 1 h,
T_5 = 0.50 h,
T_6 = 4000 h,
I_1 = 10 percent = 0.10,
I_2 = 8 percent = 0.08,
I_3 = 6 percent = 0.06,
r = 8 percent = 0.08, and
n = 20 years.

The above input figures are valid only for the reviewed system. These figures may vary with the size of the system.

Refer to Figure 1. The curves shown are for typical 250-W high-pressure sodium lamps. The group-relamping period for 250-W high-pressure sodium lamps is 16 000 h. At this point, the lamp lumen output has decreased to 85 percent of the initial lumens. After the first year, or 4000 h of operating time, the lamp mortality is 1 percent. Thus, $B_{41} = 0.01 \times 1000 \times B_1/B_2$. Values of B_1 and B_2 are 2 and 53.34, respectively. Therefore, $B_{41} = 0.01 \times 1000 \times 2/53.34 = 0.375$. This real number is the average number per kilometer of burnt-out lamps in the first year of the whole system in review. For example, the number of burnt-out lamps in a 16-km (10-mile) highway system during the first year is $B_{41} = 0.01 \times 1000 \times 2 \times 16/53.34 = 6$. The average number of burnt-out lamps per kilometer is $6/16 = 0.375$.

By using the preceding mathematical procedure and by inputting a lamp mortality of 4, 5, and 10 percent for the second, third, and fourth years, respectively, the B_{42}, B_{43}, and B_{44} values are calculated and quoted in the input figures. Values of B_{41} to B_{44} are repeated for every four-year relamping cycle—that is, when i = 5, $B_{4i} = B_{41}$; when i = 6, $B_{4i} = B_{42}$; and so on.

In this example, the probability of the new spot-replaced lamps burning out is neglected because they are quantitatively insignificant in comparison to the number of burnt-out lamps of the originally installed or group-installed lamps. However, if desired, they can be included in the calculation.

By using the input figures, DTC and AEC can be calculated from Equations 2 and 5, respectively, as follows:

$$\begin{aligned}
\mathrm{DTC} = {} & [1000 \times 2(150 + 30 + 50)/53.34 + 1000\,(300 + 100)/53.34 \\
& + 15000 + 750] \\
& + \sum_{i=1}^{20} (1/1.08^i)\,(1.1)^i\,(2 \times 250/53.34)\,(4000 \times 0.025 + 12 \times 2.50) \\
& + \sum_{i=1}^{20} (1/1.08^i)\,350(1.06)^i \\
& + \sum_{i=1}^{20} (1/1.08^i)\,B_{4i}\,[30(1.08)^i + 30 \times 1(1.06)^i] \\
& + \sum_{a=1}^{a=5} (1000 \times 2/(1.08)^{4a} \times 53.34)\,[30(1.08)^{4a} + 30 \times 1(1.06)^{4a}] \\
& + \sum_{a=1}^{a=5} (1000 \times 2/(1.08)^{4a} \times 53.34)\,[30 \times 0.50 \times (1.06)^{4a}] \qquad (6)
\end{aligned}$$

Substituting values of B_{4i} according to the number of years i and the calculated factors results in DTC = 31 872.98 + 29 716.38 + 5784.80 + 2042.73 + 10 142.86 + 2259.28 = 81 819.03. Thus, the discounted total cost is \$81 819.03/km.

$$\begin{aligned}
\mathrm{AEC}_y = {} & [1000 \times 2 \times (150 + 30 + 50)/53.34 + 1000(300 + 100)/53.34 + 15\,000 + 750]\,(0.08 \times 1.08^{20})/(1.08^{20}-1) + \\
& [2 \times 250/53.34)\,(4000 \times 0.025 + 12 \times 2.50]\,1.1^y + \\
& 350 \times 1.06^y + 1.875\,(30 \times 1.08^y + 30 \times 1 \times 1.06^y) + \\
& (1000 \times 2/53.34)\,(30 \times 1.08^y + 30 \times 1 \times 1.06^y)/4 + \\
& (1000 \times 2/53.34)\,(30 \times 0.50 \times 1.06^y)/4
\end{aligned}$$

$\mathrm{AEC}_y = 3246.54 + 1218.60\,(1.1)^y + 912.43\,(1.06)^y + 337.46\,(1.08)^y$.

Substituting number of years in review gives annual equivalent cost for that year. Following are the annual equivalent costs for years 10, 15, and 20, respectively:

$\mathrm{AEC}_{10} = 3246.54 + 1218.60\,(1.1)^{10} + 912.43\,(1.06)^{10} + 337.46(1.08)^{10} = 3246.54 + 3160.73 + 1634.02 + 728.55 = 8769.84$

$\mathrm{AEC}_{15} = 3246.54 + 1218.60\,(1.1)^{15} + 912.43\,(1.06)^{15} + 337.46(1.08)^{15} = 3246.54 + 5090.39 + 2186.69 + 1070.48 - 11\,594.10$

$\mathrm{AEC}_{20} = 3246.54 + 1218.60\,(1.1)^{20} + 912.43\,(1.06)^{20} + 337.46(1.08)^{20} = 3246.54 + 8198.13 + 2926.29 + 1572.89 = 15\,943.85$

The annual equivalent costs for years 10, 15, and 20 are \$8769.84, \$11 594.10, and \$15 943.85, respectively.

CONCLUSION

This paper has addressed the question of how to treat lighting-system cost components that are subject to varying inflation rates. A computer program has been developed that is generally applicable, with or without inflation, and DTC and AEC models offer the lighting design engineer the opportunity to include unadjusted projected inflation in a system-cost calculation. As with most models, their usefulness is only as good as the quality of engineering and economic data used. Although the reliability of economic predictions of inflation is open to question, the computer model allows the user to test for the sensitivity of a wide range of possible inflation rates and mixes of costs.

This paper addressed the straightforward world of equal system benefits, but additional research is required to develop models for the economic comparison of lighting systems that yield differing levels of service.

ACKNOWLEDGMENT

We wish to thank the following for their contributions to this paper: P. Mok, Computer Systems Branch, and A. Maghsoudi, Highway Wayside Equipment Research Section, for developing the computer program mentioned in this paper; V. McCullough and F. Tung, Electrical Design Section, for their constructive comments and criticism on aspects of lighting engineering; F. W. Jung, Program Manager, Highway Wayside Equipment Research Section, for his time, valuable comments, and guidance that made the writing of this paper possible; M. Hastay, Department of Economics, Washington State University, for discussion and insights on theoretical aspects of the economic models; and GTE Sylvania, Rexdale, Ontario, for permission to reproduce the lamp curve shown in Figure 1. While we acknowledge the valuable assistance of these individuals and GTE Sylvania, any errors and omissions are our sole responsibility.

REFERENCES

1. F. W. Jung and C. Blamey. Computer Program for Roadway Lighting. TRB, Transportation Research Record 628, 1977, pp. 25-32.
2. P. D. Henderson. Investment Criteria for Public Enterprises. In Public Enterprise (R. Turvey, ed.), Penguin Books, London, 1968.
3. Coated 250 Watt Lumalux. In New Product Information, GTE Sylvania, Ontario, July 1977.
4. H. A. Van Dusen, Jr. Maintenance and Adjustment Factors in Street Lighting Design Calculations. Journal of the Illuminating Engineering Society, Oct. 1971.
5. O. Eckstein. Water Resource Development: The Economics of Project Evaluations. Harvard Univ. Press, Cambridge, MA, 1958.
6. R. H. Baldock. The Annual Cost of Highways. HRB, Highway Research Record 12, 1963, pp. 91-111.
7. A. Cassel and D. Medville. Economic Study of Roadway Lighting. NCHRP, Rept. 20, 1966.

Publication of this paper sponsored by Committee on Visibility.

Abridgment

Test of 400-W High-Pressure Sodium Vapor Lighting

Michael C. Belangie, Research and Development Unit, Utah Department of Transportation, Salt Lake City

At present, 400-W high-pressure sodium vapor (HPS) lamps are advertised as having an average life expectancy of 24 000 h, the equivalent of six years of operation. The cost in labor and equipment to replace a single lamp in Utah is about 90 percent of the initial cost of a 400-W HPS lamp.

Since both lamp and maintenance costs are rising rapidly, inferior lamp performance will place an unnecessary, unprogrammed cost on lighting maintenance programs. The results of this test indicate that manufacturer-supplied lamp mortality data should not be considered reliable (performances of between 35 percent and 50 percent were typical). These findings strongly suggest that users should develop and implement contractual mechanisms for enforcing performance specifications in regard to the advertised life of the lamp.

QUALIFICATIONS ON THE USE OF THIS REPORT

Substantial technological modifications have occurred since these lamps were manufactured. Therefore, the lamp mortality data described here are only applicable to the lamps and fixtures available from the manufacturers in late 1971 and early 1972 and in the lamp-and-fixture combinations used in this test. The direct application of these findings to lamps and fixtures currently being manufactured is therefore strongly discouraged.

However, until the manufacturers can unequivocally demonstrate that their products will perform as advertised, the user is fully justified in assuming that manufacturer-supplied lamp mortality data may be questionable.

BACKGROUND

HPS lighting was first used in Utah in 1968. During that year, contracts were let for more than 1400 lighting units; these were the initial elements in an installation that would number more than 6600 operational luminaires by 1977.

In 1977, 81 percent (5533 luminaires) of this system was composed of 400-W HPS units; the remainder of the system was composed of 250-W HPS (13 percent) lamps and either 400-W or 250-W mercury vapor units (6 percent). Since 1968, the advertised average life expectancy of the 400-W HPS lamp has increased from 6000 h (1.5 years) to 15 000 h (3.75 years) in 1971 and, most recently, in 1976 to 24 000 h (6 years).

In 1971, substantial concern over the costs resulting from the terms of utility maintenance agreements had surfaced within the Utah Department of Highways. The lack of hard data on the validity of the advertised increases in lamp life compounded the problem. This study was initiated to address both problems.

FINDINGS

Four test groups of 12 luminaires each were operated on a 10-h on, 2-h off continuous cycle for two years (14 360 h of operation) (1). Replacement of burnouts occurred during the first year of operation only. Figure 1 depicts the performance of each test group. Performance of each group has been divided into three categories:

1. Original: performance of the initial 12 operating lamps;
2. Original and replacement: performance of the original 12 operating lamps plus any replacements made during the course of the test;
3. Original, replacement, and infant: performance of the original 12 operating lamps plus any replacements, including lamps that burned out the first time they were energized (infant mortality).

Only one test group (Lamp B in Fixture X) performed without a single failure during the two-year test. The other three test groups all had failures. Lamp A in Fixture W had 9, 5 of which occurred in the first year and were replaced. Lamp D in Fixture Z had 12 failures, 4 of which occurred in the first year and were replaced. Lamp C in Fixture Y had 15 failures in the first year

Figure 1. Summary of lamp performance.

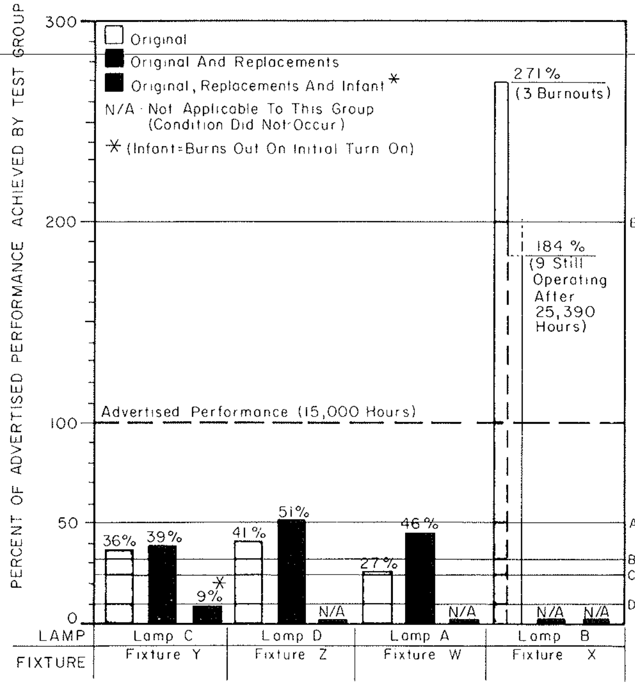
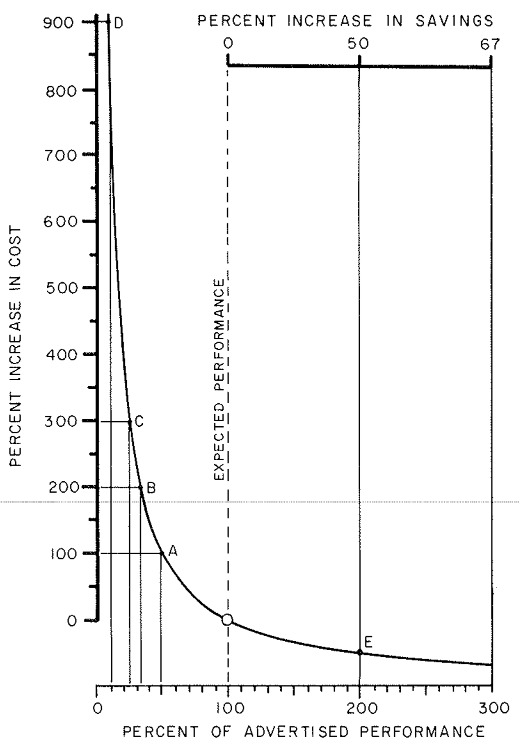


Figure 2. Estimated effect of lamp performance on expected costs (excluding cost of replacement lamp).



(7 infant mortalities and 8 during the course of the year). This test group was removed after 8380 h of operation. The other test groups, with the exception of Lamp B in Fixture X, were removed after two years. The remaining test group was placed on a regular 10-h on and 14-h off burning cycle and left in place. On April 1, 1977, 9 of the original 12 lamps were still burning. The total number of hours of operation for each of the 9 operating lamps on that date was 25 390 h.

Consistency of performance is a factor. Cycling (lamp going off and then restarting) occurred in all four test groups:

Lamp	Fixture	Occurrences
B	X	1
A	W	18
D	Z	51
C	Y	97

Most of the cycling occurred during the first summer of the test, during July and August. Both months were exceptionally hot; temperatures rarely fell below 27°C (80°F) and often exceeded 38°C (100°F). The second summer was comparatively mild. Cycling during this period was consistent with observed winter operation, and cycling principally occurred prior to failure.

Energy use remained fairly constant for the four test groups over the course of the two-year operating period; the average daily consumption per lamp was 0.45 kW·h (90 percent within 0.03 kW·h). For the first 18 months following the removal of the other three test groups, the remaining group consumed an average of 0.48 kW·h/day (95 percent within 0.02 kW·h). In the 19th month, consumption jumped 20 percent to an average of 0.58 kW·h/day (94 percent within 0.03 kW·h) and remained at that level until April 1977, when readings were discontinued.

Peak draw (power) increased steadily over the entire period. The average high 15-min draw per month (power billing) began at 0.46 kW. After 61 months of operation, it had increased to about 0.62 kW, an increase of 25 percent.

Observed light intensity (apparent lumen output) did not decline significantly over the course of the observations.

CONCLUSIONS

Lamp Mortality Data

Figure 1 is sufficient justification to question the reliability of the lamp mortality data supplied by the various manufacturers.

Group Replacement

Group replacement of lamps is a function of economics and performance. As the lamps age, the rate of burnouts increases. Since group replacement is less costly per unit than individual replacement, there is a threshold point (determined by rate of burnout and the alternative replacement costs) beyond which it is cheaper to replace by group than to continue with individual replacements.

The HPS lamp has a very small decrement in light output over the course of its life. This aspect of the lamp, in combination with an expected average lamp life of 24 000 h and the apparently unreliable manufacturer-supplied lamp mortality data, makes group replacement at this time a highly questionable practice.

Energy Use

Energy use increased modestly over the course of this study. However, it should be noted that this was a small installation that had an overhead power supply. The rates shown here should be considered to be somewhat lower than those that would be expected in a field installation.

Demand

Peak energy demand (power) may play a role in lighting costs in areas where penalties are placed on users who contribute to peak demand.

IMPLEMENTATION

Figure 2 indicates the effect of lamp performance on the expected cost of maintenance. Points A through E refer to the equivalent reference lines drawn on the right vertical of Figure 1. In a lighting system composed of lamps performing at 50 percent of advertised capability (reference line A in Figure 1) that were replaced at burnout with lamps of equivalent operational (as opposed to advertised) capability, the cost of maintaining the system would be 100 percent greater (point A in Figure 2) than the amount budgeted for the time period. In relating Figure 1 to Figure 2, it can be seen that the performance of the first three groups in Figure 1 could have raised the expected maintenance cost of a lighting system composed exclusively of those lamps and fixtures by 100-900 percent of the expected cost.

In considering the fourth group in Figure 1 (Lamp B, Fixture X) and relating reference line E in Figure 1 to point E in Figure 2, it would appear that this combination would yield substantial savings. However, if a group replacement policy that was based on the manufacturer's lamp mortality data had been in effect, then the indicated savings would not have been realized; the lamps would have been replaced in groups at approximately 92 percent of their advertised life (13 800 h). It must be noted that in 1977 no major lamp manufacturer recommended group replacement of HPS lamps.

It is apparent that the users must find some method to protect themselves from manufacturers of inferior products. The low-bid requirement, in an area such as lamp purchases, can cause a user to incur a substantial maintenance liability through the mandatory purchase of a less than satisfactory product.

The city of Seattle (2) and the state of Idaho (3) have used a life-cycle costing approach to implement a partial solution to this problem. Whether their approach is adequate is yet to be determined.

In order to achieve the savings indicated in Figure 2, group replacement that is based on manufacturer-supplied data must be eliminated in favor of either replacement after individual burnout or a group replacement system that is based on in-house data. The economic feasibility of the latter approach is questionable and needs to be evaluated before serious thought is given to its implementation.

The potential benefits of this study will be largely determined by the effectiveness of the steps taken by users to require performance at the level advertised.

REFERENCES

1. M. C. Belangie. 400-Watt High-Pressure Sodium Vapor Lighting Test (15 000-Hour Lamp Rating), Executive Summary. Utah Department of Transportation, Salt Lake City, Final Rept., July 1977.
2. B. Bermack. Cost Analysis of High-Intensity Discharge Mercury Vapor and High-Pressure Sodium Lamps, 1975 Annual Contract. City of Seattle, Seattle, Washington, 1975.
3. Bid Invitation for Electric Lamps for State of Idaho Agencies. Idaho Division of Purchasing, Boise, May 27, 1977.

Publication of this paper sponsored by Committee on Visibility.

Abridgment

Pavement Inset Lights for Use During Fog

Frank D. Shepard, Virginia Highway and Transportation Research Council, Charlottesville

Reduced visibility as a result of fog presents a very hazardous condition on the highway because motorists are unable to readily observe pavement markings and signs and the movement of traffic. Afton Mountain, which is traversed by I-64, is often the site of such reduced visibility because of the low cloud cover on the mountaintop during rainy periods.

An acute awareness of the fog problem on Afton Mountain led to a decision by the Virginia Department of Highways and Transportation to install a lighting system that consists of pavement inset lights and low-level illumination lights to aid motorists during periods of fog. The installation was made on a 9.3-km (5.8-mile) section of highway that encompasses the top of Afton Mountain. Since fog often occurs on only a portion of the mountain, the installation was divided into three sections that represent the points observed to most often correspond with the fog patterns. Each section is controlled by two fog detectors, located at or close to the endpoints of the section, that are capable of detecting five levels of fog density. The intensity of the guidance lights within each section is controlled by the density of the fog at each detector.

The fog guidance system consists of unidirectional airport runway lights installed in the pavement edge line along each side of the roadway in both directions, spaced at 61-m (200-ft) intervals on tangent sections and 30.5-m (100-ft) intervals on curved sections. In addition to the white inset lights on the main line, amber lights are installed on one side of the off-ramps. Also, low-level illumination lights are installed on a short section of an on-ramp.

It was felt that the lighting system would help delineate the highway and thus lead to an improvement in traffic operations. However, it was not known how the system of lights would affect vehicle speeds, headways, and placement. Also, there was some concern that the system might promote a false sense of security and lead to higher speed and greater differentials in speed, which would increase the possibility of accidents. It was, therefore, the purpose of this research to investigate the traffic flow characteristics during fog within the system of pavement inset lights installed on Afton Mountain.

The scope of the study included the collection and analysis of traffic flow data before and after the lights were installed. The parameters evaluated were vehicle speeds, headways, queues, and lateral placement. A before-and-after accident analysis was also made. Because of the variability of fog densities for the period of study and the limitations on the time, personnel, and funding allocated for the study, it was not possible to collect traffic flow data for all fog conditions.

METHOD

The effect of the system of lights on the traffic flow characteristics was determined by comparing data collected before and after the system was installed. Data were collected for all fog conditions (day and night) occurring during the time allowed for the project. The collection of data was limited to weekdays and to off-peak hours.

For the accident study, data were obtained from the Division of Motor Vehicles.

Site

One location on the two westbound lanes of I-64 over Afton Mountain just east of the Blue Ridge Parkway was chosen for data collection. This location, which had an annual average daily traffic count of approximately 5000 vehicles for the period 1973-1976, is on the level section of the mountaintop and has a slight horizontal curve to the right. There are no interchanges in advance of the site for a distance of 11.3 km (7 miles). Data were collected at only one site.

Collection of Data

Data were collected before (11/27/73—5/1/75) and after (1/26/76—9/30/76) installation of the lighting system by using tape switches placed on the highway. Under clear, dry conditions the tape switches were attached to the road surface with double-faced tape. During adverse weather, they were attached to 0.56-mm (0.022-in) metal ribbons stretched across the highway and fastened to metal anchors in the shoulder and the median.

Data from all the tape switches were recorded simultaneously on a four-channel chart recorder; the switches were identified by assigning different voltages to each. Since the distance between the tape switches on the road and the speed of the chart recorder were known, vehicular speeds and headways were determined by measuring the distances between impulses on the chart.

Vehicle placements were obtained by installing tape switches of different lengths on the right edge of the traffic lane and noting which switch combinations were activated.

The chart recorder was placed in a vehicle parked approximately 304.8 m (1000 ft) past the site to eliminate any influence the parked vehicle may have had on traffic flow. Also, the switches were not conspicuous to motorists.

Weather Conditions

The fog problem considered is one caused by low cloud cover in the mountainous areas. Although the fogs are relatively dense and uniform, variable fog conditions, fog banks, and so on do occur at the edges of the cloud cover and in areas under broken clouds that result from clearing weather. For the evaluation, data were taken only during uniform fog conditions extending at least 152.4-304.8 m (500-1000 ft) in advance of the test site.

Fog Density

It was very important that a relative measure of fog density be obtained, because this influences traffic flow characteristics. The density was determined by noting the number of visible centerline stripes on the pavement during daylight hours and the number of reflectorized shoulder delineators during hours of darkness. These distances were used to identify relative fog densities in the analysis of the data.

Since the tape switches were approximately 304.8 m from the data recorders, the fog densities and the uniformity of the fog were monitored by driving through the site at regular intervals.

RESULTS

The numerous variables associated with the project, problems encountered in data collection during adverse weather conditions, and the time frame in which it was possible to collect the data limited the quantity of data collected. Therefore, the study results reflect only those data for which comparable before-and-after conditions were available.

Speeds

The installation of the pavement inset lights has resulted in a decrease in daytime speeds and an increase in nighttime speeds. For all the sight distances tested in fog during daytime and nighttime, there was a significant increase in the variability of speeds for automobiles in the traffic lane in the after period. Results for the remainder of the comparisons showed no differences for automobiles or tractor-trailers, with the exception of automobiles in the passing lane for a sight-distance interval of 33.5-45.7 m (110-150 ft) ahead of the data collection point. The decrease in daytime speeds is explained by the fact that during daylight hours, within the range of sight distances encountered in fog, the motorist can see at least two or three centerline skips [12.2-m (40-ft) intervals; 4.6-m (15-ft) painted line]; however, the uniqueness of the system, in addition to some glare associated with the lights, caused the motorist to slow down. It is felt that when the data were accumulated the inset lights were a little brighter than they needed to be, which caused the glare and deceleration, and that, once experience allows a realistic coordination between the sight distances in fog and light intensities, speeds will tend to be closer to those found in the before conditions.

Daytime fog creates restrictive driving conditions; however, it is during nighttime fog that driving becomes difficult, primarily because of the driver's inability to see pavement markings and delineators whose retroreflectivity is significantly reduced under night or wet-weather conditions. Also, the reflection of the vehicle's headlights from the fog seriously restricts the motorist's visibility. The significant increases in speeds during nighttime fog give strong support to the contention that the inset lighting system provides additional delineation for guidance. However, it should be noted that the increase in speed at night for the after condition increases the accident potential, since even the nighttime speeds before installation exceed those required for a safe stopping distance.

Safe Stopping Distance

For all conditions investigated (before and after), the actual sight distances were less than the safe stopping distances, a finding that indicates that drivers tend to go too fast for sight distances in fog. It is noted above that the increase in nighttime speeds since the installation of the lights raises the accident potential as a result of the increased stopping distances required. However, the low traffic volume encountered on this rural section of Interstate highway, especially at night, leads to a decrease in vehicle interaction that lessens the significance associated with the safe stopping distance. Also, the improved delineation is thought to help prevent vehicle stoppages along the roadway.

Headways

A review of the headway data showed little difference in headways between the before and after conditions during the hours of daylight. However, available data showed a decrease in nighttime headways (below 3 s) after the inset lights were installed. This finding, coupled with results that showed less vehicle queuing in the after condition, indicates that motorists were using the inset lights for guidance rather than relying on car following.

Queuing

There was a decrease in daytime vehicle queuing for the sight distance of 33.5-45.7 m (110-150 ft) but little difference within the range of 45.7-61.0 m (150-200 ft). At night, for both sight distances considered, there was a decrease in vehicle queuing. There was little difference in the numbers of vehicles in the queues. The increase in headways and decrease in vehicle queuing at night might indicate a reduction in the potential for accidents under the lighting system. However, it should be noted that, for severely restricted sight distances before the system was installed, vehicles tended to form queues for the purpose of being led through the fog, which may be thought of as being safer than having no one to follow.

Lateral Placement

During daylight, the lateral placement of automobiles was farther from the right edge line after the lights were installed. Also, the placement was farther from the edge line during fog for both the before and after periods than it was during clear weather conditions. Both automobiles and tractor-trailers were positioned farther from the edge line for nighttime fog conditions than for clear conditions.

Accidents

It would be difficult to surmise what, if any, increase in accident potential would result from the differences noted in traffic flow parameters. There has been only one accident during fog conditions since the system of lights was installed. Also, in a recent subjective evaluation of the system, more than 95 percent of the motorists interviewed indicated that they were aided by the system and 90 percent reported that the lights reduced their anxiety while driving in fog (1).

REFERENCE

1. M. F. Creech and T. A. Crisman. Evaluation of the 5.8-Mile Highway Light System for Guidance in Fog on Afton Mountain. Virginia Highway and Transportation Research Council, Charlottesville, Sept. 1976.

Publication of this paper sponsored by Committee on Visibility.

Driver Performance with Right-Side Convex Mirrors

Ronald R. Mourant, Oakland University, Rochester, Michigan
Robert J. Donohue, General Motors Corporation, Warren, Michigan

The mirror-use behavior of drivers was investigated as they gathered information from rearview mirrors in order to execute freeway lane changes and merges. Nine drivers (three novice, three experienced, and three mature) drove a 1973 Buick LeSabre with and without a right-side fender-mounted convex mirror along a 22.5-km (14-mile) freeway route. The total time to obtain information per maneuver was the same for both cases. In a subsequent study, the mirror-use behavior of five subjects who drove a 1976 Nova without a right-side convex mirror was compared with that of 12 subjects who drove the same vehicle with a right-side door-mounted convex mirror. Again there were no differences in total time to obtain rear-vision information. Experienced drivers (mean age = 24) took less time to obtain information when a right-side convex mirror was available than when it was not; older drivers (mean age = 61) took more time. Also, experienced drivers required about 10 h of driving experience to become efficient users of a right-side convex mirror, while older drivers required considerably more driving experience. Finally, a comparison of right-side door- and fender-mounted convex mirrors indicated that the drivers' total time to obtain information was the same for each mounting location, but drivers who had the fender-mounted mirror made a greater number of direct looks to the rear.

Some drivers may find it difficult to obtain the proper information necessary to execute lane changes and merges to the right. Factors that contribute to this difficulty include the following: (a) plane mirrors located on the right door do not always provide an adequate field of view, (b) sail panels located at the right rear of the vehicle can obstruct vision, (c) high head restraints can restrict the vision of short drivers, and (d) physical afflictions and old age can restrict turning one's head to

Table 1. Subjects' ages and distances driven, by groups.

	Group 1		Group 2		Group 3		Group 4	
Subject	Age (years)	Distance Driven (km 000s)	Age (years)	Distance Driven (km 000s)	Age (years)	Distance Driven (km 000s)	Age (years)	Distance Driven (km 000s)
1	16	4.0	20	120	17	55	21	39
2	21	0.8	23	80	18	40	21	320
3	26	5.6	24	240	20	24	22	97
4	23	113	47	480	20	48	23	48
5	24	130	69	480	21	48	23	97
6	25	72			21	320	54	80
7	42	320			43	390	55	160
8	46	400			44	160	58	72
9	51	560			49	48	60	80
10					54	80	61	140
11					60	80	62	160
12					63	800		

Note: 1 km = 0.6 mile.

make direct looks to the rear.

One way to improve right-rear vision is to mount a convex mirror on the right side of the vehicle. A convex mirror of relatively small size [13.3 cm (5.25 in) wide] and moderate radius of curvature [102 cm (40 in)] will provide an adequate field of view (19-20°) for detecting vehicles located at the right rear of the vehicle. For calculation of field-of-view requirements, see Sugiura and Kimura (1).

Since the radius of curvature of a convex mirror is less than infinity, objects appear smaller than they would in a plane mirror. The viewing of these minified objects may result in erroneous distance judgments; that is, drivers may judge vehicles to be located further to the rear than they actually are. Several investigators have studied this problem (2-4) and have found that drivers can misjudge distances when they use convex mirrors. Walraven and Michon (4) reported that experienced drivers would accept smaller traffic gaps than inexperienced drivers, suggesting that judgment would improve with training. Also, Mortimer (5) found that, when drivers use the inside mirror in conjunction with an outside convex mirror, misjudgments in distance do not occur.

Furthermore, when a convex mirror is used as a "go or no-go" device (6), distance judgments are irrelevant. In a "go or no-go" situation, a driver simply uses the convex mirror to detect a vehicle's presence. If no vehicle is present, the driver proceeds with the maneuver. If a vehicle is present, the driver checks its actual location by using the inside mirror or by a direct look to the rear.

Burger and others (7) have collected data on drivers' glance behavior with convex mirror systems while the drivers executed left and right maneuvers in freeway and city traffic. They found that when a right-side fender-mounted convex mirror was available drivers made fewer direct looks to the side or rear of the vehicle than when it was not available.

Questions to be answered by means of the present research include the following:

1. How does the information-gathering behavior of drivers change when a right-side convex mirror is available? Will drivers use the convex mirror in actual traffic situations, and will they take more or less time to obtain information? A survey by Kaehn (8) found that drivers of government vehicles who used right-hand convex mirrors were impressed by the improved field of view.

2. Does convex-mirror-use behavior differ according to the age of the driver? As Mourant and Rockwell (9) reported, young novice drivers make very little use of the left-side mirror and hence are likely to make little use of the right-side convex mirror. Older, mature drivers have ingrained information-gathering habits and therefore will probably take a longer time to become accustomed to the convex mirror. Even though older drivers, in general, have poorer vision than younger drivers, the location of the convex mirror on the right side of the vehicle places it far enough away from the driver so as to eliminate the necessity for large changes in eye accommodation (as well as reducing the need for a head turn to the rear). Thus, as noted by Seeser (10), older drivers will not need corrective lenses to view it.

3. How long does it take drivers to learn to use convex mirrors? Since little is currently known about how drivers use a right-side convex mirror as a detection device, especially how it is used in conjunction with inside mirror glances, it may be difficult to tell when drivers have become highly skilled convex-mirror users. However, since skill learning improves rapidly during the early part of training, it will be possible to detect when a learning plateau has been reached.

4. Does the location of the right-side convex mirror (door or fender mounted) affect the method and duration of information gathering? A door-mounted convex mirror requires a greater head turn from straight ahead, and it permits the use of peripheral vision to obtain more information. A fender-mounted mirror, on the other hand, requires only slight eye movement from straight ahead. If the fender- and door-mounted mirrors are of the same size and convexity, their fields of view will be about equal.

5. Does information obtained prior to the decision to execute a lane change affect subsequent information-gathering strategy? Realization that a vehicle is moving through a constantly changing environment suggests that the "memory effect" of information obtained prior to making a decision would be fairly short. This question was addressed mainly for methodological considerations. If there is no memory effect, different mirror systems may be compared by having drivers sample their mirrors whenever desired (the natural driving method). However, if there is a memory effect, then mirrors should be sampled only after the decision to make a lane change has been reached.

METHOD

Drivers

Table 1 shows the ages and distances driven for the four subject groups. All drivers had valid Michigan licenses and at least 20/40 visual acuity.

Design

The subjects performed in four groups:

Group	Condition
1	Exploratory
2	Training
3	Door-mounted mirror
4	Fender-mounted mirror

Group 1 subjects drove a 1973 Buick LeSabre first with conventional mirrors and then with a 102-cm (40-in) convex mirror mounted on the right front fender. The conventional mirror system consisted of the inside and left outside mirrors supplied by the manufacturer. Subjects practiced using the convex mirror for about 200 km (125 miles) before data collection. In this exploratory study, subjects 1-3 were novice drivers, subjects 4-6 were experienced drivers, and subjects 7-9 were mature drivers.

Group 2 subjects drove a 1976 Nova Concours twice with conventional mirrors, six times with a door-mounted 102-cm convex mirror, and finally twice again with conventional mirrors. Before each convex-mirror run, subjects practiced using the convex mirror for about about 4 h or 320 km (200 miles). Thus, every subject had more than 1600 km (1000 miles) of driving experience with the convex mirror before the start of the sixth convex-mirror run. Subjects 1-3 were experienced drivers, subject 4 was mature driver, and subject 5 was an older driver.

Subjects in groups 3 and 4 also drove the 1976 Nova Concours. For group 3 the 102-cm convex mirror was door mounted, and for group 4 it was fender mounted. Both mirrors provided about the same field of view (20° horizontal) as measured by Society of Automotive Engineers (SAE) recommended practice J1050. After each subject had driven the Nova for about 12 h or 970 km (600 miles), four data-collection runs were made. Group 3 subjects 1-6 were experienced drivers, subjects 7-9 were mature drivers, and subjects 10-12 were older drivers. Group 4 subjects 1-5 were experienced drivers and subjects 6-11 were older drivers.

Procedure

Data were collected as subjects executed lane changes on a 22.5-km (14-mile) freeway route that had moderate to heavy traffic. Data collection for each maneuver was initiated 15 s prior to the execution of the maneuver. Execution of the maneuver occurred when the leading edge of the test vehicle crossed the lane marking.

Drivers were instructed to align both left and right outside mirrors so that a small portion of the vehicle was visible in the inboard edge of the mirror. Drivers were told to use the right-side convex mirror as a "go or no-go" device. If a vehicle was present in the convex mirror, the driver was to check its location by a glance to the inside mirror or a direct look. All subjects drove the test vehicle for at least 30 min before data were collected, so that they could become familiar with the route and vehicle-handling characteristics.

Before the start of data collection, subjects were read the following instructions:

> You are to execute either left or right lane changes upon command of the experimenter. However, please check for possible traffic in the adjacent lane before proceeding to make the lane change. Please use the convex mirror as a "go or no-go" device. That is, if a vehicle is present, check its location by a glance to the inside mirror or a direct look to the rear. If no vehicle is present in the convex mirror, you may want to proceed with the maneuver. Do you have any questions?

Data Collection Equipment

A closed-circuit television system consisting of four television cameras, special-effects electronics, an electronic counter, a video monitor, and a videotape recorder was installed in the rear passenger compartment of the test car. Power was supplied from an inverter connected to the vehicle's battery.

One camera monitored the driver's eyes through a front surface mirror mounted on the instrument panel, and a second camera monitored the road scene ahead of the vehicle. Two additional cameras separately monitored the road scene on the left rear and on the right rear. The video display from each camera appeared simultaneously as one of four separate sections of each television frame. Each frame was also numbered by the electronic counter; these numbers were then used to calculate glance durations. An experimenter seated out of view of the driver in the rear passenger compartment could view the scenes from all four cameras simultaneously on the monitor while the information was being videotaped.

RESULTS

Mirror Use With and Without a Right-Side Convex Mirror

The top part of Figure 1 contains the information-gathering data for nine subjects (group 1) who drove a 1973 Buick LeSabre with and without a 102-cm-radius convex mirror mounted on the right front fender. To obtain the necessary information for a right lane change, drivers averaged 2.65 s/maneuver without the use of a right-side convex mirror and 2.84 s/maneuver with the use of a right-side mirror. The increase in time with the use of a convex mirror was not statistically significant ($t = 0.11$; $df = 8$). Note that the number of glances made to the vehicle's inside mirror decreased when the convex mirror was available. None of the nine subjects had any previous driving experience with convex mirrors, which probably accounts for the approximately equal amount of time spent making direct looks with (0.38 s/maneuver) and without (0.36 s/maneuver) the use of the right-side convex mirror.

The lower part of Figure 1 contains the information-gathering data for 5 subjects (group 2) who drove a 1976 Nova Concours without a right-side mirror and 12 subjects (group 3) who drove the same vehicle with a 102-cm-radius convex mirror mounted on the right door. Group 2 drivers averaged 2.95 s/right lane change and group 3 drivers averaged 2.86 s. The decrease in time with the use of a convex mirror was not statistically significant ($t = 0.46$; $df = 22$). However, drivers' mirror-use behavior when a convex mirror was available was dramatically different from that observed when the vehicle had no right-side convex mirror.

Group 2 drivers averaged 0.52 s/maneuver for direct looks when the vehicle had no right-side convex mirror, and group 3 drivers averaged only 0.04 s/maneuver when the convex mirror was available. Group 3 drivers made "combination looks." That is, they sampled the inside mirror and then moved their eyes directly to the convex mirror before returning to look at the road scene ahead. Other combination looks were from convex mirror to inside mirror and from inside mirror to convex mirror to inside mirror. These combination looks resulted in the inside mirror being sampled 1.98 times/maneuver and the convex mirror being sampled 1.39 times/maneuver.

Mirror Use Behavior as a Function of Age

Figure 2 contains the information-gathering data for experienced, mature, and novice drivers of a 1973 Buick LeSabre. The mature drivers took 3.47 s to obtain information when the convex mirror was available and only 2.55 s when no right-side convex mirror was available. The additional time with the use of the convex mirror was due to the fact that the mature drivers sampled the convex mirror on the average of 0.93 times/maneuver. Both the experienced and novice drivers took less time to obtain information when the convex mirror was available than when it was not. However, the novice drivers averaged sampling the convex mirror every fourth maneuver, while the experienced drivers averaged sampling it every second maneuver.

Figure 1. Information-gathering data for right lane changes with and without a right-side convex mirror.

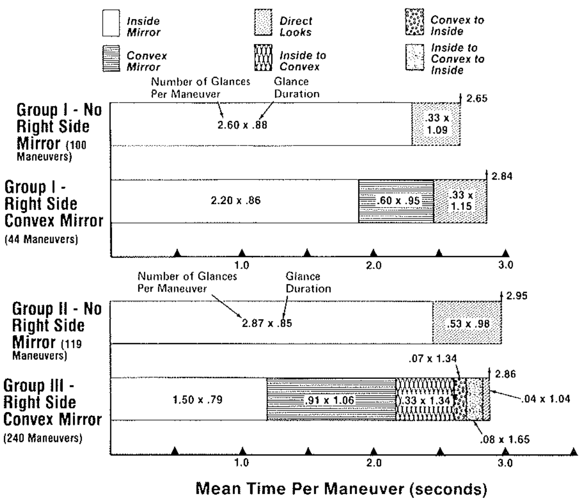


Figure 3 contains the information-gathering data for experienced, mature, and older drivers of a 1976 Nova Concours. The experienced drivers who had a right-side convex mirror (group 3) averaged less time per maneuver than the experienced drivers who did not have a right-side convex mirror (group 2). However, the mature and older drivers who had a right-side convex mirror (group 3) averaged more time per maneuver than the mature and older drivers who did not have a right-side convex mirror (group 2). These results agreed with those for the Buick drivers. Note also that drivers made very few direct looks when a right-side convex mirror was available.

Training Results

Figure 4 contains the information-gathering data averaged over all group 2 drivers for four control runs (no right-side convex mirror) and six convex-mirror runs. After the first convex-mirror run, the frequency of inside-mirror use alone decreased dramatically. On convex-mirror runs 3, 4, 5, and 6, the average total time per maneuver was very close to that of the control runs. Note also that when the convex mirror was available drivers made very few direct looks to the rear.

As there are in all training experiments, there were large individual differences in behavior. Figure 5 contains the information-gathering data for a 23-year-old male and a 69-year-old male. The younger driver's average time per maneuver on convex-mirror runs 4, 5, and 6 (3.20 s) was considerably longer than the average time on four control runs (2.65 s). Thus, this was an atypical young driver. Yet his behavior was different

Figure 2. Age and information-gathering data for right lane change with and without a right-side convex mirror: 1973 Buick.

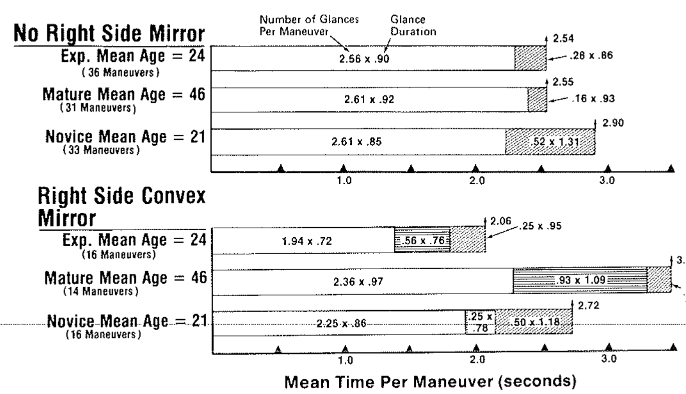


Figure 3. Age and information-gathering data for right lane changes with and without a right-side convex mirror: 1976 Nova Concours.

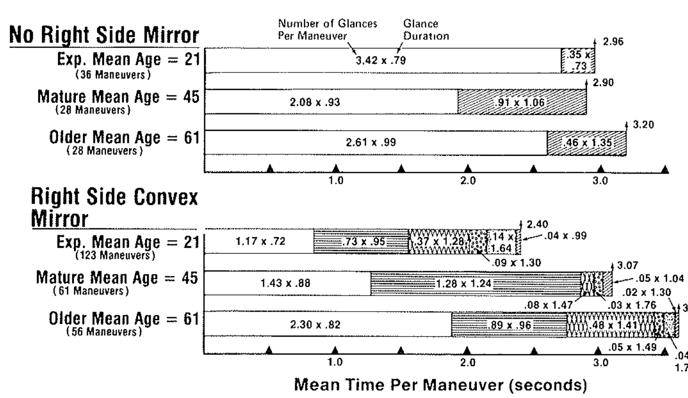


Figure 4. Effect of training on information gathering by group 2 for right lane changes.

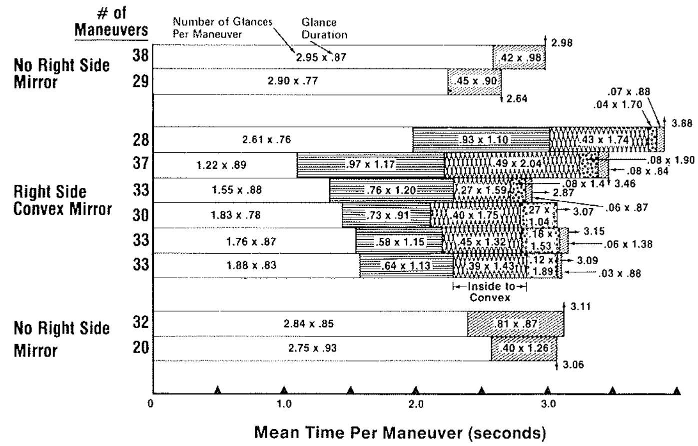


Figure 5. Effect of training for two individual subjects.

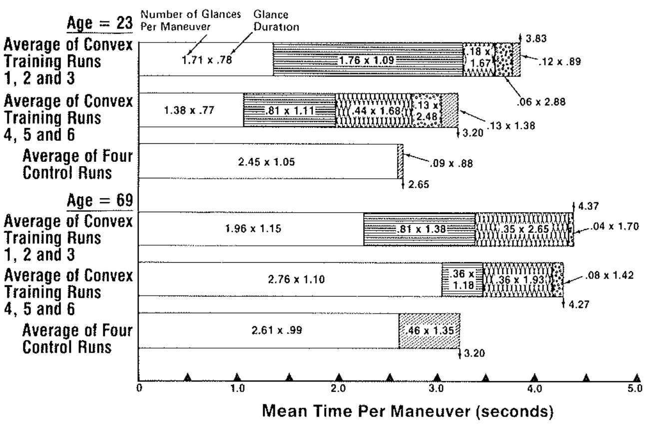
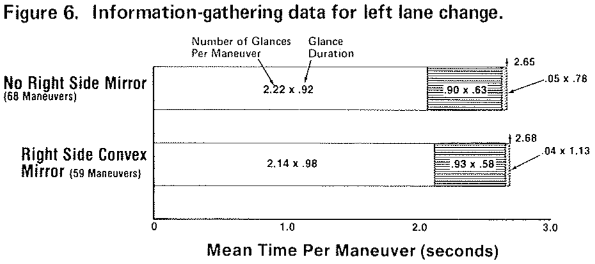


Figure 6. Information-gathering data for left lane change.



from that of the older driver, since his total time on convex-mirror runs 4, 5, and 6 was shorter than that on convex-mirror runs 1, 2, and 3. Apparently older drivers take a much longer time to learn sampling with the convex mirror than do younger drivers.

As an experimental control, driver mirror-use behavior in making left lane changes with the standard left-side mirror was also analyzed. These results are shown in Figure 6. Driver mirror-use behavior on the control runs was the same as that on convex-mirror runs 4, 5, and 6. Thus, such uncontrollable factors as traffic, weather, and driver motivation either had a negligible effect on the control and convex-mirror runs or they were averaged out.

Comparison of Door- and Fender-Mounted Convex Mirrors

Figure 7 contains the information-gathering data for run 4 with the door-mounted convex mirror (group 3) and with the fender-mounted convex mirror (group 4). Subjects who had the door-mounted mirror averaged 2.86 s to obtain information, while subjects who had the fender-mounted mirror averaged 2.89 s. This result might be expected, since the door- and fender-mounted mirrors provided about the same field of view.

However, there were differences between the two groups in their information-gathering behavior. While subjects who had the door-mounted mirror averaged only 4 direct looks/100 maneuvers, subjects who had the fender-mounted mirror averaged 40 direct looks/100 maneuvers. This may be due to the fact that use of the door-mounted convex mirror requires a larger turn of the eye or head from straight ahead and permits the use of peripheral vision on the right side. Use of the fender-mounted convex mirror requires a smaller eye turn from straight ahead but limits peripheral vision to the right side.

Memory-Effect Results

Figure 8 contains the information-gathering data for left lane changes of a restricted group of drivers (group 4) and a normal group of drivers (group 3). The restricted group of drivers was not permitted to sample any mirrors between commands to execute left lane changes. This behavior may be considered "memoryless" in that rear-vision information was gathered only after the command to execute a left lane change had been made. The normal group of drivers was permitted to sample all mirrors continuously. Thus, information gathered just prior to a command to execute a left lane change may have been useful in actually completing the lane change.

The restricted group of drivers averaged more time per maneuver to obtain information on runs 1 and 2, but on runs 3 and 4 both groups averaged about the same amount of time per maneuver to obtain information. On all trials, the average glance duration to the left outside mirror was longer for the restricted group of drivers than for the normal group of drivers, which may indicate that a more concentrated effort was being made to detect vehicles by means of the left outside mirror. The restricted group of drivers also made more direct looks to the rear than did the normal group of drivers, perhaps because they were unsure whether other vehicles were present in the area to their left side.

CONCLUSIONS

1. A right-side door-mounted convex mirror is an acceptable visual aid for drivers in obtaining informa-

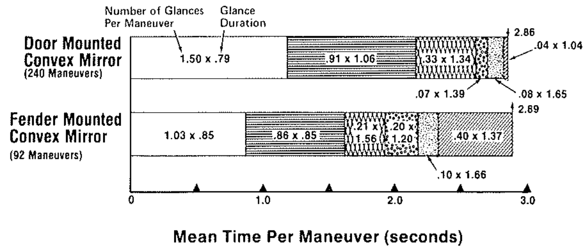


Figure 7. Information-gathering data for right lane changes with door-mounted and fender-mounted convex mirrors.

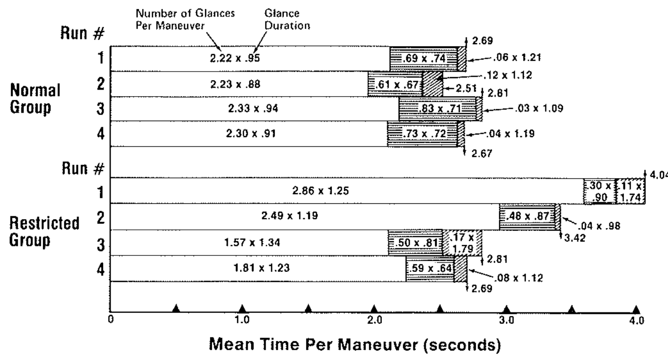


Figure 8. Information-gathering data for left lane changes for the normal and restricted group.

tion for right lane changes. The convex mirror enables drivers to substantially reduce the number of direct looks to the side and rear of the vehicle. Thus, short drivers and drivers who have physical afflictions that prevent head turns can effectively obtain information to execute right lane changes.

2. Older drivers spent more time sampling mirrors (plane and convex). This is supported by Kretovics (11), who found that older drivers turned their heads farther toward the mirror than younger drivers.

3. The use of the right-side convex mirror does not affect the use of the left-side plane mirror. Drivers took the same amount of time and used the same techniques to acquire left lane-change information whether the vehicle was equipped with a right-side convex mirror or not.

4. A fender-mounted convex mirror resulted in more direct looks to the side and rear of the vehicle than did a door-mounted convex mirror. Even though the fields of view for the fender- and door-mounted mirrors were about equal, drivers who used the door-mounted mirror apparently obtained information from their peripheral vision and therefore did not need to make as many direct looks to the right rear.

5. It appears that drivers do not rely on information obtained before a command to make a lane change. Since freeway traffic is so dynamic—i.e., vehicles move at 88 km/h (55 mph) and there are multiple lanes—drivers must obtain information immediately prior to the execution of a lane change. This makes previously obtained information of little value.

REFERENCES

1. S. Sugiura and K. Kimura. Outside Rearview Mirror Requirements for Passenger Cars: Curvature, Size, and Location. Society of Automotive Engineers, Technical Paper 780339, 1978.
2. T. S. Bowles. Motorway Overtaking with Four Types of Exterior Rearview Mirrors. IEEE-ERS International Symposium on Man-Machine Systems, Cambridge, England, 1969.
3. R. G. Mortimer and C. M. Jorgeson. Drivers' Vision and Performance with Convex Exterior Rearview Mirrors. Society of Automotive Engineers, Technical Paper 740961, 1974.
4. P. L. Walraven and J. A. Michon. The Influence of Some Mirror Parameters on the Decisions of Drivers. Society of Automotive Engineers, Technical Paper 690270, 1970.
5. R. G. Mortimer. The Effects of Convex Exterior Mirrors on Lane-Changing and Passing Performance of Drivers. Society of Automotive Engineers, Technical Paper 710543, 1971.
6. R. R. Mourant and R. J. Donohue. Acquisition of Indirect Vision Information by Novice, Experienced, and Mature Drivers. Journal of Safety Research, Vol. 9, 1977, pp. 39-46.
7. W. J. Burger, J. D. Beggs, R. L. Smith, and J. W. Wulfeck. Studies in Motor-Vehicle Rear Vision: II—Evaluation of Innovations in Passenger Car and Truck Rear-View Systems. U.S. Department of Transportation, DOT-HS-801-258, Nov. 1974.
8. C. Kaehn. Evaluation of a New Automotive Plane- and Convex-Mirror System by Government Drivers. Society of Automotive Engineers, Technical Paper 76006, 1976.
9. R. R. Mourant and T. H. Rockwell. Strategies of Visual Search by Novice and Experienced Drivers. Human Factors, Vol. 14, 1972, pp. 325-335.
10. J. W. Seeser. Automotive Mirror Size Requirements Based on Field-of-View Considerations. Society of Automotive Engineers, Technical Paper 76005, 1976.
11. T. J. Kretovics. Analysis of Frequency and Amplitude of Driver Head Movements in Mirror Sampling. Ohio State Univ., Columbus, Master's thesis, 1976.

Discussion

Thomas H. Rockwell, Department of Industrial and Systems Engineering, Ohio State University, Columbus

The authors are to be commended for their approach to this problem; indeed, the use of field studies for evaluation purposes is essential, albeit at some sacrifice of experimental control. Several of their experimental findings are consistent with my own work in mirror use. Glance durations are about right, although I would also like to see individual subject means and variances, since aggregation of the rate removes intrasubject and intersubject variation.

The paper is silent about the explicit instructions given to the subject and about the effect of traffic on individual maneuvers. I suspect that the velocity and density of the following traffic in relation to the subject's car would influence glance frequency and duration, as would the headways of automobiles directly ahead of the subject driver. These variables are difficult to measure, let alone control. I would like to have seen a components-of-variance analysis of key dependent measures as they were affected by such independent variables as lead-automobile headway, relative velocity and spacing of traffic following in the right lane, age, sex, and subject strategy.

The lack of sample sizes probably prohibited good estimates of intersubject and intrasubject variation, although both of these give insights into the quality of experimental control and interpretation of results.

The idea of using the left-side mirror sampling as an experimental control was an excellent one, particularly for trip-duration effects and driver attentiveness and less so for traffic effects. The authors' concern that peripheral vision may account for the sharp differences in mirror sampling for the fender- versus door-mounted mirrors is probably valid (40/100 versus 4/100 maneuvers). I understand that the door-mounted position was about 80° from line of sight. Data from an Ohio State University (OSU) study of 48 drivers (using their own cars in freeway lane changes and merges) that involved 7000 mirror samples indicated that head turns cover 50 percent and eye movements 50 percent of the angle between line of sight and the mirror (11). A 40° head turn in this case (to use the door-mounted mirror) might have allowed peripheral detection of target automobiles.

The OSU study also showed that age effects are significant in glance durations and head turns, but we found older drivers making greater head turns than younger drivers. (Older in this case was defined as over 45.) Previous research at OSU indicated larger eye-search patterns among older drivers, perhaps indicating less confidence in peripheral detection. We also found that sex, driver height, and automobile size had little effect on left-side and inside mirror sampling.

The issue of memory probably is more important when drivers overtake vehicles and return to the right lane and may account for certain mirror sequence sampling. In that regard, were the differences in certain combinations of mirror sampling found in group 3 (and in group 4) and not in group 1 the result of automobile type or experience?

The authors are encouraged to plot glance frequency against glance duration; i.e., do successive glances produce shorter durations? Finally, we need to probe for the driver's strategy in mirror use; i.e., does the driver depend on the convex mirror for detection, evaluation, and decision information or does he or she use the convex mirror for detection and use other mirrors or direct looks and memory for the evaluation and decision processes?

As to the final judgment on convex mirrors, I believe that this must wait until we ascertain the quality of resultant maneuvers and can be assured of minimal acclimation time. I presume from the paper that the experiments were not conducted in high-density traffic with its attendant stress. This condition might be the ultimate test of convex mirrors as aids to information acquisition in driving.

Robert L. Henderson, National Highway Traffic Safety Administration

My remarks are directed, by request, at providing additional data on the general subject of convex mirrors rather than discussing the authors' paper.

I would like to describe briefly a National Highway Traffic Safety Administration (NHTSA) contract I am managing that will eventually examine a rather large number of rearview mirror systems. It is being conducted for us by Bill Burger at Vector Research. The basic objective of the study is to evaluate driver mirror-use behavior while performing in-traffic maneuvers. The study uses a variety of mirror systems designed to meet the requirements of a proposed new federal motor vehicle safety standard. Measures of mirror and direct glance frequency and duration similar to those reported by Mourant and Donohue will be the major criteria measures. A repeated-measures design will be used in which the same 12 subjects will use all mirror systems.

Data collection on this main portion of the study is now in progress. The data I will present today were collected in the first phase of the study for the express purpose of selecting the radius of curvature for the convex-mirror component of those systems that use convex mirrors.

The literature indicates that a 51-cm (20-in) radius of curvature is probably the maximum usable convexity because of image minification and distortion. At the other extreme, when the radius exceeds about 203 cm (80 in), the limited gain in the field of view over a less expensive plane mirror does not justify the added cost. We have selected four specific values from within this range for this study—51-cm, 102-cm (40-in), 140-cm (55-in), and 203-cm radii of curvature. Unlike some past studies in which mirror size is held constant as convexity varies, we held field of view constant for all convexities by varying the size of the mirrors. The question we wished to answer was whether there is an optimum convexity and, if there is, whether that optimum value varies as a function of the presence of other mirrors in the system.

METHOD

Three vehicles were used; each vehicle had a mirror system designed specifically to meet the requirements of the proposed new mirror standard. The vehicles were a passenger automobile, a light truck, and a van. Two types of mirror systems were employed on the automobile. The first type used two plane mirrors: one inside and the other on the left outside. Each of these mirrors was slightly larger than that found on passenger automobiles today. The second type used a left outside plane mirror somewhat larger than standard, a standard inside plane mirror, and a right outside convex mirror. The convex mirror (radius of curvature of 51, 102, 140, or 203 cm) was systematically varied.

The second type of automobile mirror system was also used on the van and the pickup. In addition, these two vehicles were tested with a combination plane and convex mirror on each side and with a single convex mirror on each side. The convexity was again varied among the four levels described earlier.

A semidynamic test condition was used to obtain distance, speed, and gap-acceptance judgments. Subjects viewed a target automobile approaching their own static vehicles through the mirror system of interest. Each of 12 subjects was tested by using each of 45 rearview systems properly counterbalanced in a repeated-measures design. These same 12 subjects will be used later in our on-the-road studies. A total of 11 880 trials were run. Each subject has two blocks of eight speed and distance judgments and six lane-change judgments for each mirror system. For all trials, subjects also performed a forward tracking task for perceptual loading.

RESULTS

1. Learning: Significant learning effects were noted for all mirror systems. Because of the counterbalanced order of presentation, no clearcut and consistent learning curves were noted for specific mirror systems. In spite of the relatively large number of trials, it appears that learning is still taking place after approximately 900 trials, indicating that learning to use convex mirrors is a long-term phenomenon.
2. Relative distance error: Examination of the overall relationship between convexity and overestimates or underestimates of distance indicates some trends, but the magnitude of the differences is so slight as to indicate no practical significance.
3. Absolute distance error: There was little difference in average error as a function of mirror convexity. The same is true for the standard deviation. Although some of the differences were statistically significant, the practical differences are trivial.
4. Gap acceptance: We found small but fairly consistent trends toward better gap-acceptance performance from larger-radius mirrors. Best performance at all convexities was found in mirror systems in which a plane mirror was in close proximity to the convex mirror.
5. Subjective ratings: Perhaps the clearest and most consistent trends were found in the subjective ratings. Highly curved mirrors were consistently judged to be more difficult to learn and to produce less confidence than less highly curved mirrors.

These data provide very little guidance for selecting the convex-mirror component of the systems we are using in our on-the-road study. Although certain trends are evident, in most instances the magnitude of the observed differences is so small as to have little practical significance. Consequently, we selected the convex-mirror components after considering the general literature, the trends observed in our data, the extent of the obstruction to direct field of view, and judgments concerning what manufacturers might prefer for aesthetic reasons or for overall standardization. For your information, we are using 140-cm radius of curvature for the right outside mirror for all systems that involve plane mirrors on the inside and left outside. For truck and van systems that use combination plane and convex mirrors on both sides of the vehicle, we are using 51-cm-radius convex mirrors. For the truck and van systems that use a single convex mirror on each side, the 102-cm radius of curvature was selected.

I appreciate the opportunity to describe this project and hope that next year we will be able to report the results of the on-the-road study.

Rudolf G. Mortimer, Department of Health and Safety Education, University of Illinois at Urbana-Champaign

Historically, U.S. automobiles have been fitted with plane interior and exterior mirrors. The field of view available from a plane mirror is determined by the size of the mirror and its distance from the eyes of the driver. With respect to the exterior left-side mirror, the lateral field of view can be increased by increasing the size of the mirror along its horizontal axis. While there are some limitations on the extent to which the mirror can be extended (due to limitations on the overall width of motor vehicles), it is possible to obtain improvements in the horizontal field of view to the left of the vehicle by improving the design location of the exterior left-side mirror and increasing its width. However, this is not feasible with a plane mirror mounted on the right side of the vehicle, because of its much greater distance from the eyes of the driver.

A simple solution to increasing the field of view from exterior mirrors is to use convex mirrors. Convex mirrors have been used by European and British drivers for years, mounted on the door or fender; a significant precedent therefore exists to suggest that they can be used safely. However, U.S. manufacturers have not felt that American drivers would willingly accept convex mirrors and have been concerned with the effects that such mirrors can have on distance judgments because of the minification of objects.

To this end, a number of studies have been completed during the last decade concerning the ability of drivers to make various kinds of judgments related to the use of rearview mirrors in order to compare performance with plane versus convex exterior mirrors. These studies have used various types of measurements, such as the frequency and duration of glances in interior and exterior mirrors (12), detectability of vehicles in dusk or dawn illumination (3), subjective evaluations of the effects of headlight glare in rearview mirrors (3), distance judgments (13), and gap-acceptance judgments (2-5).

As Mourant and Donohue have pointed out in their paper, the balance of these studies indicated that drivers overestimate the distance of a vehicle seen in a convex mirror, i.e., judge it to be further away than it really is, although, in a normal driving situation, when a driver also has available the plane interior mirror, the judgments made when a convex exterior mirror or a plane exterior mirror were used were the same. Thus, it would appear that drivers can make safe judgments with exterior convex mirrors, though their eye-fixation patterns were somewhat different than when a plane exterior mirror was used.

The present study was concerned with eye-fixation behavior of drivers who used plane and convex exterior mirrors mounted on the right side of the vehicle. Two locations were selected for mounting of the mirror: on the right door and on the right fender. Other variables investigated were the effect of driver experience and age, effects of learning, and the effect of information obtained from rearview mirrors shortly before the decision was made to execute a lane-change maneuver, which the authors termed the memory effect.

The authors used a video recording technique to ascertain the frequency and duration of glances made by drivers in lane-change maneuvers and measured the frequency and duration of glances in each of the rearview mirrors. The emphasis in the study was on the comparison of mirror use with and without a convex mirror on the right side.

It is important to note one aspect of the instructions given to the subjects. The subjects were told to use the right-side convex mirror as a "go or no-go" device. Furthermore, if a vehicle was noted in the convex mirror, the driver was to check its location by a direct glance to the rear or by use of the inside mirror. Thus, any findings from this study will be limited because of the nature of the instructions given to the subjects about how they should use the convex right-side mirror. Clearly, the instructions were such that subjects were discouraged from using the right-side convex mirror to make estimates of distance or relative velocity of a vehicle seen in the right lane before making a lane change. One effect of this type of instruction would naturally be to increase the proportion of direct glances the drivers would make when using a right-side convex mirror in relation to driving without it. Based on the data presented (for example, in Figures 3 and 4 there were relatively few direct looks to the rear when a convex mirror was used and there were relatively few glances from the convex mirror to the inside mirror), it appears that there were few occasions when there was a vehicle visible in the convex mirror at the time that a decision to make a right lane-change maneuver was made. There is no indication given in the paper as to the proportion of maneuvers in which a vehicle was reasonably close to the subject's vehicle when the command was given to execute the maneuver, either when the right-side mirrors were in use or without them. It would seem that this could have been an important piece of data that should have been available from the video recordings and might have been taken into account in explaining some of the findings.

I was somewhat concerned that there were relatively few subjects used in some of the conditions. For example, the findings in Figure 2 are based on the result of only three subjects in each of the age and experience groups. So few subjects would scarcely be enough to allow reasonable differences to be discovered between categories of this variable.

However, it was refreshing to see the findings in Figure 6, which indicated that, when mirror-use behavior was observed in left lane changing for the five subjects in group 2, the performance in left lane-change maneuvers was very similar when the vehicle was equipped with a right-side convex mirror and on a set of control runs when it was without the right-side mirror. This indicates that the overall procedure appears to have a good degree of reliability, at least when 60 or more maneuvers are involved.

The authors also address the effect of the age of the drivers. In this case, there were substantially more subjects available in the two groups (groups 2 and 3) that were used for this comparison, which showed that experienced drivers required less time to obtain mirror information than mature and older drivers when a right-side convex mirror was available, whereas without the right-side convex mirror all three groups required approximately the same amount of mirror information-gathering time.

A comparison is also shown, in Figure 7, between door- and fender-mounted right-side convex-mirror performance. This figure appeared to be based on the comparison of the performance of the 12 subjects in group 3 and the 11 subjects in group 4 who had a mean time to obtain rear-vision information of 2.86 s and 2.9 s, respectively, for the door-mounted and fender-mounted convex mirror. It appeared to me, however, that this comparison was confounded by differences between these two groups. Group 3 was described as carrying out the experiment in the normal mode of mirror scanning, whereas group 4 was restricted in scanning behavior; i.e., the subjects were not permitted to scan the mirrors until the command had been given by the experimenter to begin a lane-change maneuver. The purpose was to permit evaluation of the memory effect. It seems to me that these two groups should not be used to compare another variable. The effect of fender mounting versus door mounting of the convex mirror could perhaps be obtained from the data obtained on group 1, which carried out the normal task with a fender-mounted convex mirror on the right side, compared with group 3, which performed the task with the convex mirror mounted on the right door. In that comparison, the mean time to obtain information was 2.84 and 2.86 s for the fender- and door-mounted convex mirrors, respectively.

The concept of evaluating the memory effect is certainly an interesting one and shows that the authors were sensitive to many subtle variables that could affect driver behavior with rearview mirrors. This effect was evaluated by comparing a group operating in the normal mode with a group operating in the restricted mode in making left lane-change maneuvers. While the information-gathering time was substantially longer on the first two (out of four) trials in the restricted group, indicating a potential memory effect, there were no differences in the subsequent two trials. While this might indicate that the memory effect had been erased by the third and fourth trials, I feel that this is not likely. Is it possible, for example, that a motivational effect was operating, due to a slightly greater stress imposed on the subjects and the interaction with the experimenter, who called out the command to begin the lane-change maneuver?

Based on the viewing behavior that was measured in this study, it does seem reasonable that convex mirrors could be used on the right side of the vehicle. The total amount of time spent viewing mirrors was not increased significantly by the addition of this third mirror on the right side. Certainly, a convex right-side mirror of adequate size and moderate radius of curvature, such as that used in these studies, would greatly improve the field of view to the right of the vehicle, where the present use of a plane mirror or no mirror at all provides a potentially hazardous blind spot.

Although this study indicated that the fender-mounted convex mirror resulted in more direct looks to the side and rear of the vehicle than did a door-mounted convex mirror, this should not be taken to imply that a fender-mounted location is undesirable. Location on the fender has some additional advantages, such as actually providing a greater field of view to the right. This is particularly valuable in dense traffic when another vehicle is close by on the right side and potentially outside the field of view of a door-mounted mirror but visible in a fender-mounted mirror. Secondly, a fender-mounted mirror requires less divergence of the eyes from the forward field of view, which is the most important location to be scanned by the driver in normal driving situations.

This study has been useful in taking a look at age and maturational factors. Certainly, more data of this type would be valuable.

In this context, the authors have studied the memory effect, which was of considerable interest and requires further study. The memory effect could be particularly important in those emergency situations where there is little time to obtain additional visual information and a rapid decision must be made as to the most appropriate evasive maneuver to make. In such cases, there may be little or no time for adequate mirror scanning to determine the locations of other potential vehicles, and the memory effect could be of critical importance. Eye fixations of experienced drivers indicate that they scan the environment all around the vehicle frequently; it may be premature to suggest that drivers do not retain an appraisal of the location of other vehicles around them in short-term memory.

Finally, one might legitimately ask what the relevance of obtaining frequency and duration of glance behavior in mirrors is in relationship to accidents in which rear-vision information may have been inadequate? The overall duration of mirror glances would be important in those situations where little time is available to make a decision in an evasive maneuver. On the other hand, the quality of the information that is obtained is also critical. Thus, one might ask whether drivers obtain more accurate information concerning the location, distance, and relative velocity of other vehicles with a right-side convex mirror than with no mirror on the right side and how safe the resulting lane-change maneuvers are. The latter question was not addressed in this study. This raises the issue of the relevance of performance criteria in rear-visibility studies. Perhaps such criteria cannot be structured properly until more information becomes available as to the underlying causes of crashes that involve inadequacies in visibility to the rear.

REFERENCES

12. C. R. Kelley and D. J. Prosin; Dunlap and Associates, Inc. Motor Vehicle Rear Vision. National Highway Safety Bureau, U.S. Department of Transportation, 1969.
13. G. E. Rowland, C. A. Silver, S. C. Volinsky, J. S. Behrman, N. F. Nichols, and W. Clishman. A Comparison of Plane and Convex Rear-View Mirrors for Passenger Automobiles. National Highway Traffic Safety Administration, U.S. Department of Transportation, 1970.

Authors' Closure

Both Mortimer and Rockwell had positive comments on the use of left lane-change data for experimental control purposes. These left lane-change data indicated that the effects of traffic, weather, and driver motivation were averaged out when computed over 60 left lane-change maneuvers. These data may also serve as a standard reference when considering driver search behavior with reference to left lane changes. In this regard, the data show that drivers made only six direct looks to the rear while executing 127 left lane-change maneuvers. However, drivers did rely heavily on their left outside mirrors; they looked at them more than twice per maneuver.

Concerns were voiced about the small number of subjects in the study of age effects. We agree that more data should be collected on this very important variable. The data could be used to develop aids and countermeasures for the older driver. Perhaps older drivers will find convex mirrors useful, in that they will partially eliminate the need for head turns to the right rear.

It should be noted that all data were collected while driving on freeways in the city of Detroit in moderate to heavy traffic. We considered this to be a very demanding task for most drivers. Thus, instructions to the subjects probably had very little effect on driver performance. Many times, when the experimenter gave the command to execute a lane change, other traffic in adjacent lanes prevented the subject from immediately executing the maneuver. Thus drivers had to search by using their mirrors or by making direct looks to determine when to proceed with the lane change. Because the traffic flow on the freeways was always moderate to heavy, it had little effect on comparisons between door- and fender-mounted mirrors.

Since the data in this study have shown that the use of a door-mounted convex mirror has reduced the frequency of direct looks per 100 maneuvers from 40 (with no right-side mirror) to 4, we believe that automobile drivers will have no problem in using a right-side convex mirror.

Publication of this paper sponsored by Committee on Visibility.

Human Factors Considerations for In-Vehicle Route Guidance

Conrad L. Dudek, Texas Transportation Institute, Texas A&M University, College Station

This paper considers the development and maintenance of credibility in the design, implementation, and operation of a route guidance system. Because drivers will have positive attitudes about a system that provides them with relevant, reliable, and accurate information, all precautions must be taken to ensure that these driver expectations are met. Messages must be presented clearly and must allow ample time for the driver to respond to a given situation. Factors that affect reading time of displays include driver work load, message load, message length, message familiarity, and display format. In order to maintain driver credibility, surveillance must be an integral part of a route guidance system. Such surveillance must be able to detect adverse conditions, validate the adverse conditions, and determine the nature and scope of the problem. Electronic sensor surveillance, however, has some limitations. Because it is a blind system, (a) some form of visual validation and assessment of incidents to ensure the accuracy of displayed messages and (b) some guarantee of adequate system maintenance are necessary. A "forgiving" system—one that alerts the driver and provides instructions about how to return to a scheduled route after a diversion—must also be considered.

An important consideration in a successful route guidance system for the United States (where success is measured by achieving desirable driver response) is to develop and maintain credibility, that is, driver faith in the system. The quickest way to fail is to lose driver confidence. The most elaborate and costliest system can deteriorate into an operational headache if the confidence of the motoring public is lost (1).

My remarks, therefore, center on credibility. In particular, this paper highlights software and hardware issues that I view as important human factors considerations during the development, design, implementation, and operation of a route guidance system in order to achieve and to maintain driver credibility.

My understanding of the system proposed for the United States is that it is a two-level system (2). The first level, referred to as a static routing system, provides instructions to drivers based on the best route information derived from historical data. The second level, referred to as a dynamic routing system, gives instructions to drivers based on current traffic conditions obtained from surveillance of the highway system. This paper essentially considers the dynamic system, although many of the same principles may be applied to the static-type system. A constraint in preparing this presentation was that the exact configuration of the U.S. in-vehicle display (i.e., whether it will give only routing information or traffic information in addition to routing information, or use both visual and audio displays) was not known.

DISPLAYS

First, it is necessary to consider the in-vehicle displays themselves and a system in which only routing information is visually displayed. Will this type of information suffice? Recent research has indicated that under certain applications it will suffice, provided drivers are convinced that the system is indeed directing them to the best route. One application is during the routing of traffic for a special event (e.g., football game, baseball game, or state fair). Recent experience with changeable message signing indicated that drivers will be responsive (3, 4). Since most drivers will be expecting some form of congestion on the primary highway near the major generator and many drivers will be unfamiliar with alternate routes, they will respond to BEST-ROUTE information. Another application is the routing of unfamiliar through-drivers along alternate freeways through and between metropolitan areas (5).

When dealing with local drivers, particularly commuters, there is a strong indication that drivers wish to know the location of a problem ahead; they desire diversion information and at least one good reason for being diverted (1)—for example, a major delay. They may have to be convinced that diversion is in their best interests. Therefore, it may be necessary to use word messages that describe the incident, location, and degree of the problem in addition to the routing instructions.

The question as to whether traffic information should be presented on a visual display or an auditory display requires further research.

CREDIBILITY

Drivers will have positive attitudes about a system that provides them with relevant, reliable, accurate, up-to-date, and timely information displayed in a manner that is clear and can be read in ample time to make a decision (1). All precautions must be taken to ensure that these driver expectations are met. Drivers will have negative attitudes about a system that does not meet these expectations. Each of these factors is examined briefly.

Clarity and Reading Time

As I understand the guidance system, information will not be continuously displayed but will be presented only at key intersection areas or other important locations. In reality, then, there will be constraints on message presentation time similar to those of on-road changeable message signs. At a given driving speed, factors that affect the reading time of displays include driver work load, message load, message length, message familiarity, and display format (1).

Messages must be viewed within a distance that allows sufficient exposure time for drivers to attend to the complex driving situation and glance at the display a sufficient number of times to read and comprehend the message. The higher the driver work load, the longer time the message must be legible.

With the enormous capability and flexibility of the types of visual displays that I envision for use with a route guidance system, there may be a temptation to display more information than drivers can read and comprehend during the available reading time. This has been a problem in some cases with the use of on-road changeable message signs (5). Hopefully, the lessons learned from this experience will help establish optimum message lengths for in-vehicle displays.

When too much information is displayed, drivers can become overwhelmed to the extent that they have difficulty in scanning and reading the message. There is evidence, for example, that for on-road changeable message signs, no more than four units of information should be displayed when all four units must be recalled by drivers (1). The informational unit refers to each separate data item given in a message that a motorist could recall and that would form the basis for making a decision. On proposed in-vehicle displays, definitive turning movements designated by arrows in contrast to word messages should enhance information transfer.

There is also evidence that an eight-word message (excluding prepositions) is approaching the processing limits of drivers traveling at high speeds. A recent Human Factors Design Guide (1) recommends minimum message exposure time of 1 s/short word (up to eight words) or 2 s/unit of information, whichever is larger. Research in the form of in-situ and field operational studies is necessary to better assess optimum message lengths for route guidance systems.

Another factor that influences message reading time is driver expectancy and familiarity with what will be displayed. Commuters, for example, who have seen several messages on the in-vehicle display, will develop expectations of message classes and types. Based on previous experience, they will more than likely tend to gloss over familiar elements of the message and concentrate on those elements that change from one situation to another. (This assumes that standard message formats are used consistently.) Unfamiliar drivers, on the other hand, who see the message for the first time, must read the entire message. Their reading times will thus be longer than those required for familiar drivers (1).

With respect to display formating, there are several human factors concepts that can be used (1). I will discuss two. First, when word messages are used, it is best to arrange the message somewhat proportionally within the horizontal and vertical dimensions. Messages that extend considerably on either the horizontal or the vertical scale are more difficult to read.

Second, redundancy can be used either in the form of repetition or coding. Color coding of critical messages, for example, would enhance recognition and reading times. With the advent of solid-state digital television image generation and anticipated technology advances in the future, this is not outside the realm of reality.

There are various messages that can be used to describe traffic conditions or so-called traffic states. Recent studies have indicated that drivers can only distinguish about three, or at most four, levels of traffic conditions (6). The studies have also shown that drivers from small cities perceive congestion at a much lower traffic density than drivers from large cities. Words used to describe traffic conditions must be carefully selected. For example, the descriptors STOP AND GO TRAFFIC and NORMAL TRAFFIC are very vague and should not be used. The traffic state descriptor, CONGESTION, is vague and should not be used (1).

Drivers who are given delay durations (or time saved) will base their decisions to divert or to continue on the delay information more than information about the particular type of incident. If a quantitative indication of delay (in minutes) is displayed, it must be reasonably accurate because its validity can be easily checked by the drivers. Studies conducted in four U.S. cities revealed that more than 50 percent of the surveyed drivers indicated that they would divert to avoid a delay of 20 min or more. Only 8 percent stated that they would divert to a delay duration of less than 5 min (7). MAJOR ACCIDENT implies to the average driver a delay of at least 20-25 min and MINOR ACCIDENT, a delay of not more than 15 min (7).

In areas where most drivers are primarily commuters, cross-street names should be used to identify the location of the incident. The incident location may also be referenced to a well-known landmark if available. For unfamiliar drivers, the incident location should be expressed in terms of distance (1). When names of cities are used on the display, they should be identical to those used on existing static signs (1).

Even with the flexibility of the in-vehicle displays, it will be necessary in many cases to use abbreviations because of the physical size limitations. Driver interpretations of several types of abbreviations more commonly used with traffic condition and routing information are not fully known at this time (5). More research is needed in order to ensure that proper, well-understood abbreviations are used.

Timely Information

Routing and lane assignment information must be given in ample time for drivers to respond to instructions. Lane assignments must be made far in advance of the turning point and far enough upstream of any possible traffic queues so that drivers can maneuver into the appropriate lane. Therefore, a route guidance system must have the capability for measuring traffic queues and adjusting itself accordingly so that drivers are not trapped. For example, under light flow conditions, the lane assignment information can be made rather close to the turning point; but, under heavy traffic conditions, it must be made much sooner.

We also need to consider those cases where the driver has two choice points in close proximity. The situation at a freeway-to-freeway interchange is an example. Information may have to be given far in advance of the interchange. If so, it must be presented clearly so that the driver does not incorrectly exit at a ramp upstream of the major interchange. It may be necessary, therefore, to display route shields (1).

Without question, the information displayed must be compatible with existing static guide signs. Therefore, the relative time frame or spatial frame with respect to the static signs is important. It is yet to be determined if routing instructions should be displayed in advance of interchange signs, in conjunction with advance interchange signs, or beyond advance interchange signs.

Relevant Information

Telling drivers that they are in congested traffic or repeatedly telling commuters of recurrent congestion when such information is obvious decreases confidence in the capability of the system to provide useful information. What is obvious should never be displayed (1).

The route guidance system should not be used to balance demands with available capacity during recurrent congestion conditions. There have been attempts to balance on-ramp demands during daily peak periods by suggesting that drivers use other ramps on the freeway. This approach was found to be ineffective and is a good way to lose credibility. Drivers are concerned with their individual travel times and are unconcerned with optimizing flow in a corridor. It is important that they realize a significant reduction in their personal travel time if they are to continuously abide by the information presented (1).

Reliable and Accurate Information

The specific messages that can be displayed with confidence to maintain driver credibility are influenced by the operating agency's ability to detect a traffic problem and to determine the nature and scope of the problem. Therefore, surveillance will be an integral part of a route guidance system and must serve the following functions (5):

1. Detection of adverse conditions,
2. Validation of the adverse conditions, and
3. Determination of the nature and scope of the problem.

Knowing what is occurring on the affected freeways and streets has an impact on the messages one can display. Driver credibility is at stake. Repeated display of erroneous information or route recommendations that are not the best in the driver's viewpoint are ways of losing driver confidence.

Electronic surveillance with in-place detectors does not always ensure that the information received is accurate. Detection systems seem to be a problem in several existing changeable message sign installations.

Electronic sensor surveillance has other inherent limitations. First, there could be considerable delay in recognizing that an incident has occurred (8). This is due to several reasons:

1. Because of cost, sensors are normally placed at long intervals—0.8 km (0.5 mile) or longer on freeways.
2. Sensors are a point source of information.
3. Current incident detection algorithms are not 100 percent accurate. (Changing the algorithm parameters to increase the percentage of incidents detected also increases the chance of false alarms. Thus, there is a need for additional work in this area.)

Another limitation of electronic sensor surveillance is that it is a blind system. The agency must rely on other means to assess the nature of the incident in order to display appropriate messages.

Because electronic surveillance is a blind system, some form of visual validation and assessment of incidents is necessary to ensure that accurate messages are displayed. Highway agencies are now emphasizing the need for closed-circuit television as an integral part of urban traffic advisory and incident management changeable message sign systems—not as a primary surveillance technique but as a means for rapidly validating incidents and determining the nature and scope of the problem (5).

One facet that is often overlooked during the design, development, and implementation of a traffic control system is hardware maintenance. Relative to the route guidance system, it is quite apparent that hardware will be installed in various local and state jurisdictions. The question that arises concerns the guarantee of adequate system maintenance. Assuming that each city will have the responsibility for maintaining the system within its own local jurisdiction, some cities may perform better than others because of attitude, availability of money, and better-qualified personnel. A motorist traveling crosscountry and following a given route would expect to have continuous information. What safeguards can be provided so that the motorist does indeed have continuous information, or what can be built into the system to at least make the driver aware that the system may be inoperative within a given stretch of highway?

Even with our on-road changeable message sign systems, highway agencies are concerned with the lack of funds to purchase replacement parts. Another concern is the long delay in having components repaired by the manufacturer. Still another concern is the unavailability of replacement parts. Incorporating more off-the-shelf components into the route guidance system may help somewhat, but this is not the total solution to the potential maintenance problem.

What safeguards and provisions will be made during partial system failures, either at an intersection point area or along a segmented link of the primary or alternate route? How will the driver be routed back to the primary facility if there is a system failure while the motorist is on the alternate route? The system design will have to address these questions.

THE "FORGIVING" SYSTEM

A "forgiving" system must also be considered. For example, assume that a driver who is being directed along an unfamiliar alternate route did not make the turn according to the display. It would appear that there would be a need to alert the driver and provide instructions about how to return to the scheduled route.

Before a driver diverts, he or she needs the assurance that the route will lead to the final destination or return the driver to a primary facility (1). One approach to provide this assurance is to continuously display the final destination while the driver is traversing the primary or alternate route. Thus, the driver will be assured that this route will lead to his or her destination. When the destination name disappears from the display, it could be an indication that the driver has failed to follow the instructions given.

REFERENCES

1. C. L. Dudek and others. Human Factors Requirements for Real-Time Motorist Information Displays: Vol. 1, Design Guide. Texas Transportation Institute, Texas A&M Univ., College Station, Rept. No. FHWA-RD-78-5, Sept. 1978.
2. B. W. Stephens, D. A. Rosen, F. J. Mammano, and W. L. Gibbs. Third-Generation Destination Signing: An Electronic Route Guidance System. HRB, Highway Research Record 265, 1968.
3. C. L. Dudek, G. D. Weaver, D. R. Hatcher, and S. H. Richards. Field Evaluation of Messages for Real-Time Diversion of Freeway Traffic for Special Events. TRB, Transportation Research Record 682, 1978, pp. 37-45.
4. S. H. Richards, W. R. Stockton, and C. L. Dudek. Analysis of Driver Responses to Point Diversion for Special Events. TRB, Transportation Research Record 682, 1978, pp. 46-52.
5. Changeable Message Signs. NCHRP, Synthesis of Highway Practice 61, 1979.
6. C. L. Dudek, R. D. Huchingson, R. J. Koppa, and M. L. Edwards. Human Factors Requirements for Real-Time Motorist Information Displays: Vol. 10, Human Factors Evaluation of Traffic State Descriptor Variables. Texas Transportation Institute, Texas A&M Univ., College Station, Rept. No. FHWA-RD-78-14, Feb. 1978.
7. R. D. Huchingson and C. L. Dudek. Delay, Time Saved, and Travel Time Information for Freeway Traffic Management. TRB, Transportation Research Record 722, 1979, pp. 36-40.
8. H. J. Payne, E. D. Helfenbern, and H. C. Knobel. Development and Testing of Incident Detection Algorithms. Technology Service Corp., Silver Spring, MD, Rept. No. FHEA-RD-76-20, April 1976.

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

Feasibility Study of Route Guidance System

Fumihiko Kobayashi, Corporate Planning Office, Toyota Motor Company, Ltd., Aichi, Japan

The Comprehensive Automobile Traffic Control (CAC) Project was started in 1973 as a six-year project sponsored by the Agency of Industrial Science and Technology, Ministry of International Trade and Industry. The aim of this major project is to develop a comprehensive system for controlling the flow of vehicular traffic through the use of the most up-to-date computer and other technology for monitoring and controlling traffic in order to improve overall traffic conditions by reducing accidents, congestion, and air pollution and to relieve drivers of unnecessary mental stress. This paper presents some results of the feasibility study of this CAC system. Since route guidance plays the most important role in the CAC

system, a detailed explanation is given of the route guidance subsystem and its guidance algorithm, which is based on the results of computer simulation. This report also discusses the effects that might be expected from introducing the CAC system in an area of Tokyo that contains 1500 intersections. In particular, this report closely scrutinizes the important route guidance subsystem, providing a quantitative analysis of the possibilities for easing congestion and achieving various other effects. A general review is made of the costs and benefits of such a system.

The growth of large Japanese urban centers in recent years has resulted in a greater volume of traffic, and a number of social problems have emerged—traffic congestion, an increased number of traffic accidents, and air pollution—whose resolution has by now become a priority task. The Comprehensive Automobile Traffic Control (CAC) system is a new development in traffic technology that seeks a solution to these problems.

The CAC technology project, sponsored by the Agency of Industrial Science and Technology of the Ministry of International Trade and Industry, was initiated in 1973 to cover a six-year research period with a budget of 7 billion yen [¥220 = $1.00 (August 1979)]. The basic function of the CAC system is to link individual motor vehicles to a central traffic control facility through communication devices that provide drivers with pertinent information concerning surrounding road conditions and that serve to regulate the general flow of traffic.

The CAC system is broken down into five subsystems for route guidance, driving information, traffic incident information, route display board, and public vehicle priority. The pilot test (1) was carried out to evaluate these subsystems' functions. In parallel with this pilot test, the feasibility study focused on route guidance of the CAC system.

STUDY OF ROUTE GUIDANCE METHODS

In the CAC system, the route guidance subsystem that provides information to drivers on the optimum route to their destinations plays the major role. The driver will input a destination code number into a unit mounted in the vehicle (see Figure 1) and, on approach to major intersections, will receive a visual display of route guidance instructions to turn left or right or to proceed straight ahead. By reacting appropriately to this information, the driver will avoid areas of traffic congestion and reach his or her desired destination by the optimum route.

The optimum route is defined as the one that, in terms of time, distance, or economy, offers maximum advantages given the existing traffic conditions. Thus, whenever traffic volumes on a specific route reach a certain level and travel efficiency along that route decreases because of congestion, that route is no longer considered optimum and drivers who normally use that route will be instructed to use alternative routes.

Providing drivers with route guidance information will help to create a more even distribution of traffic volumes on the road network, thereby increasing the efficiency of automobile transport. On the other hand, in order to give such information it is necessary to find the optimum route for each driver. This requires a route guidance algorithm.

One of the merits of route guidance is that, if the number of drivers following the guidance instructions given is small, those that follow the instructions will arrive at their destinations sooner than those who do not. A guidance algorithm for this situation is already established as the shortest-path method (2) with link cost as input.

As the percentage of drivers influenced by route guidance increases, however, the general traffic situation develops in ways that result from the guidance. Also, the aim of a route guidance system is not primarily to benefit the individual but to alleviate congestion, a social benefit. Thus, the aim of guidance is to satisfy the driver's origin-destination (OD) requirements by selecting the route that keeps congestion as low as possible. For one OD pair, therefore, the optimum route may not be only one route.

The control system, of course, has dynamic characteristics, and a need thus arises for considering algorithms that reflect time changes in the environment.

In considering the following route guidance algorithms, analyses were carried out by viewing as parameters the guidance percentage, the indicated route computation cycle, the OD pattern, and other factors on the basis of methods of equal-time allocation and shortest-route guidance, as well as partial corrections of those two methods.

Outline of Analyses and Tools

In order to consider a comparison of route guidance methods, traffic-related data were put into order, various models and simulators were prepared, and analyses were conducted. Figure 2 shows an outline of those activities.

Trip Structure Model

The trip structure model was a static traffic assignment model based on the shortest-path algorithm, and it treated the road network data, OD data, zoning data, and other data related to traffic in Tokyo systematically. The data were put into order for use as a data base, which provided estimated values of the present breakdown of OD trips and average traffic volume for any link. Through this model, therefore, it is possible to obtain OD, network, and other data related to Tokyo's 23 wards. A pilot area was selected for analysis in order to consider in detail the route guidance methods. Figure 3 shows a scissored network chosen for the analysis.

Route Selection Model for Nonguided Vehicles

In order to conduct a detailed study of route guidance methods, we must construct a model of the special characteristics of traffic flows when drivers choose their own routes.

First, by use of the kth path method (3), a route generator (4) that automatically picks out all the possible routes for each OD pair was constructed, and available routes that drivers may choose from the actual road network were searched.

In regard to these available routes, models that can be determined quantitatively by a multiple-regression analysis of measured data of the choices made by drivers were prepared. These models were based on the following formulation for route i ($i = 1 \sim n$):

$$d_i = w_1 r_{i1} + w_2 r_{i2} + \dots\dots\dots\dots + w_p r_{ip} + w_c \qquad (1)$$

where

d_i = selection standards values of route i,
w_j = weight of attribute j,
r_{ij} = selection standards values of attribute j of route i,
w_c = correction, and
n = number of routes being selected.

Figure 1. Components of route guidance subsystem.

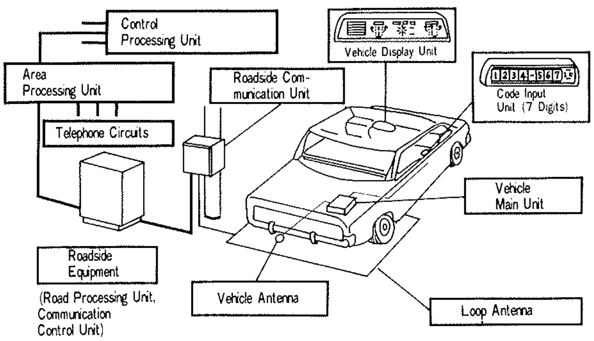


Figure 2. Flow of analysis.

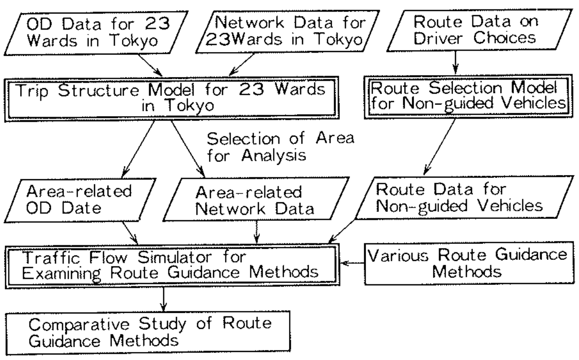
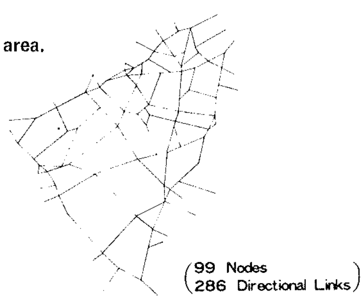


Figure 3. Scissored network of the pilot area.



In the analyses, we chose road length, number of lanes, percentage of trunk lanes, and number of left and right turns as the attributes influencing the route selection. Figure 4 shows the comparison between modeled traffic volumes and the ones actually measured. This result shows that our model fairly well represents the driver's route selection behavior.

Simulator for Examining Route Guidance Methods

In addition to the static model mentioned in the preceding sections, a simulation model is required for the investigation of route guidance algorithms for actual, dynamically varying traffic flows. This requirement has led us to the use of a simulator that incorporates vehicle movement logic on the basis of queuing concepts (5). This simulator consists of a route-computing module, a traffic-flow simulation module, and monitor control of these modules.

Queuing Logic

A vehicle is guided from the link queue toward the arc queue via the lane server and then to the following link queue via an arc server, as is shown in Figure 5. The link travel time is controlled by the lane server, while the arc travel time (intersection passing time) is controlled by the arc server.

Control of the Link Travel Time

The lane server takes into account constant-speed travel, safe head-to-head distances, and traffic signal conditions to determine the link departure time for each vehicle. The procedure for this determination is expressed by the following algorithm in which t_1 = departure time for the preceding vehicle and t' = departure time on the assumption of constant-speed travel for the vehicle in question:

1. If $t_1 > t'$, then $t = t_1$ + headway at start; otherwise $t = \max(t', t_1$ + safe head-to-head distance$)$.
2. If t falls within the red signal period, t = the next green time + delay in starting.

Control of the Arc Travel Time

The arc server determines the arc departure time by adding to the link departure time an average constant value according to whether the motion is straight travel, a left turn, or a right turn. The model thus obtained is event oriented, and it performs a calculation whenever a link or an arc event occurs.

In this model, the effect of route guidance is obtained by simulating the traffic flow when each vehicle follows the recommended route according to the guide table at each intersection.

Simulation Experiments and Results

Route guidance methods have been examined by conducting simulations according to the following conditions and by using the above-mentioned models.

Experimental Network

A simplified pilot area (Figure 3) containing 99 intersections and 286 directional links was constructed by extraction from a trip structure model of a network of some 1500 intersections in an area of Tokyo.

OD Traffic Volume

The patterns of OD traffic volumes associated with the pilot area were estimated to be those shown in Figure 6, based on the results of the trip structure model.

Route Guidance Methods

The following three algorithms were investigated:

1. Shortest-route guidance;
2. Exponentially smoothed shortest-route guidance, which gives multiple directional indications by overlapping at a certain percentage the guide table used previously with a new guide table computed by using the latest information from the shortest-route algorithm; and
3. Equal-time guidance, which uses the incremental assignment method (6) for all OD pairs to iterate computations of link cost and shortest routes 10 times in order to determine the guide table.

Nonguided Vehicles

The route selection model described above was used to determine the routes for nonguided vehicles.

Evaluation Functions

1. Total delay time expresses the degree of congestion for the specified area, and the pattern of the changes in this measure allows a comparison of the methods, thus: $c_i - c_{i_0}$, where c_i = average time required for traveling on link i and c_{i_0} = time required for traveling on link i when there is no congestion.

2. Total travel time expresses the degree of congestion for the specified area by a single value, thus: $\sum_i t_i$, when t_i = time required for travel by the ith vehicle. As Figure 7 shows, when all vehicles have

Figure 4. Comparison of measured values and modeled values of traffic volumes.

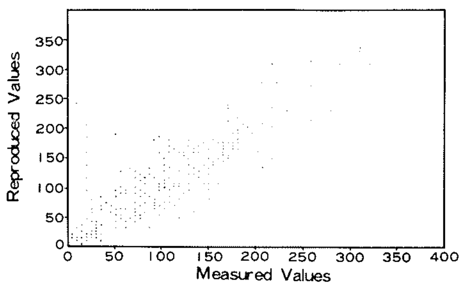


Figure 5. Model of vehicle movements.

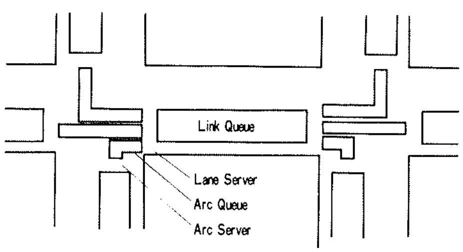


Figure 6. Pattern of OD traffic volumes.

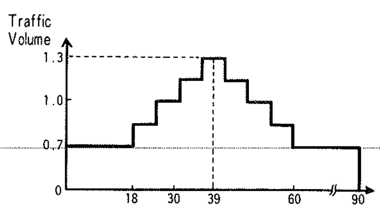
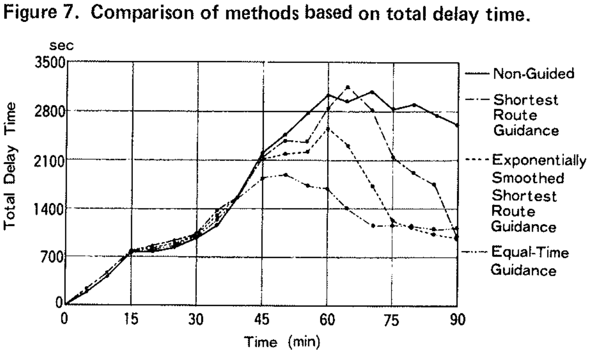


Figure 7. Comparison of methods based on total delay time.



on-board equipment, the most effective route guidance method is the equal-time method. The other methods range in order from the exponentially smoothed shortest-route guidance to the shortest-route guidance. From the point of view of practical application, the equal-time method has such problems as a large computation time and difficulty of obtaining the OD traffic volumes.

On the other hand, the exponentially smoothed shortest-route guidance and the shortest-route guidance methods are easily applied to actual situations because these two methods are mainly based on the link cost data that are directly obtained from our CAC system. We examined how the total travel time decreases as the guidance rate goes up when using these two practical methods. From the results (shown in Figure 8), we may conclude that it is most effective to use the shortest-route guidance method at a low guidance rate and to switch over to the exponentially smoothed shortest-route guidance method when the guidance rate is high.

EXAMINATION OF COST AND BENEFITS

Through the foregoing simulation experiments in the pilot area, the various route guidance methods were compared. Here a cost-benefit analysis of investments was performed on the supposition of improvements to the traffic flow if these systems were introduced to the whole Tokyo area.

Review of Benefits

A review was made of the benefits that would accrue from adoption of the route guidance subsystem, which is considered the most important subsystem in the CAC system.

Shortening of Required Travel Time

Areas A, B, and C (city center, city subcenter, and peripheral residential area, respectively) were selected as representative areas of Tokyo and the dynamic simulation was carried out to estimate the reduction in total travel time in those areas in the same manner as that mentioned above for the pilot area.

Based on these results and considering the macro characteristics of traffic flow in Tokyo, the overall reduction in total travel time in Tokyo was estimated to be about 6 percent on the average. If the time saving from this shortening of travel time is expressed in money terms, the total yearly benefit is computed to be approximately ¥80 billion.

Fuel Economy

By using the relationship between gasoline consumption and vehicle speed, the overall fuel economy for Tokyo's 23 wards was estimated to improve by approximately 3-7 percent according to the location and time of day. Therefore, the average fuel savings for the daytime period (7:00 a.m. to 7:00 p.m.) was taken to be 5 percent. Expressed in money terms, this improvement amounts to benefits of about ¥9 billion.

Reduced Air Pollution

According to the simulation tests, it can be expected that introducing the route guidance subsystem will result in reduced exhaust emissions of approximately 6.5 percent for carbon monoxide, 6.2 percent for hydrocarbons, and about 0.4 percent for nitrous oxides (7).

Figure 8. Dependence of total travel time on guidance rate.

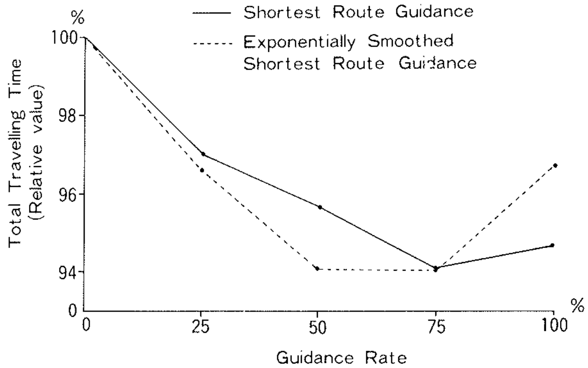
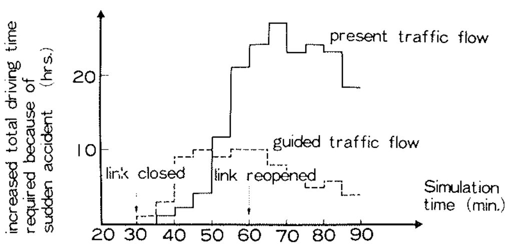


Figure 9. Effect of route guidance in case of sudden accident.



Keeping Close Schedules

Simulation test results indicate that use of the route guidance subsystem will lead to a tendency for reductions in average driving time required and reduction of variance in arrival times. For example, a random sample of 27 ODs (average of 220 trips per OD pair) was taken, and the distribution of required driving time for each OD pair was examined. In this case, the variation factor was found to decrease from 0.25 to 0.15 by introducing route guidance.

Reduction of Traffic Accidents

Based on Metropolitan Police Board data listing causes of traffic accidents in 1976, approximately 40 percent of the accidents that year resulted from factors that would be offset by using the route guidance and driving information subsystems. If 10 percent of these accidents could be prevented by using these two subsystems, the benefits could be expressed in money terms as follows [data on value of social loss obtained from Japan Research Society for Transport Policy (8)]:

[Traffic deaths (300/year) + traffic injuries (35 000/year)] × ¥2.46 million/person × [rate of guided roads (0.6)] × 0.04 = ¥2 billion/year

Preventing Traffic Congestion Caused by Sudden Accidents

When a sudden traffic accident blocks a roadway, the required driving time for vehicles in the vicinity increases sharply. It was ascertained from simulation that the route guidance subsystem has the effect of greatly preventing this type of traffic congestion. An example is shown in Figure 9, which depicts the closing of a link near Tokyo Station by an accident. The link is disconnected from the system and simulation is carried out. The graph shows the resultant increase in total driving time.

Review of Costs

1. Preconditions for estimates: Equipment needed for the route guidance subsystem for the 1500 intersections in Tokyo's 23 wards is shown below.

Item	Number of Units Needed
Route guidance roadside units	4 units/location x 1500 locations = 6000 units
Loop antennae	4 x 2 units x 1500 = 12 000 units
Main-area computers	1 set
Local-area computers	32 sets
Communications lines	4 x 1500 = 6000 lines

2. Cost: The initial investment costs and annual operating costs for ground facilities are projected from the pilot experiment data as shown below:

Item	Cost (¥000 000 000s)
Investment costs	
Construction: survey, roadside units, loop antennae, communications lines	4.74
Manufacture of equipment: roadside units	7.20
Central control unit	
Office remodeling	0.82
Computer purchase	5.45
Software preparation	0.20
Total	18.41
Annual operating costs	
Central control	
Office rental	0.44
Operations	0.2
Personnel	0.13
Roadside units: upkeep	0.6
Use of leased lines	0.86
Total per year	2.23

Besides these costs, there is also a need for on-board equipment in the route guidance subsystem, which is estimated to be about ¥30 000/vehicle.

Overall Evaluation

The following formula was used to calculate the expected costs (TC) and the possible benefits (TB) over x number of years following introduction of the route guidance subsystem.

$$TB = \sum_{i=1}^{x} B(i)/(1+r)i \qquad (2)$$

$$TC = IC(x) + RC(x) + CC(x) + RCC(x) \qquad (3)$$

in which

$$IC(x) = (IC_o) \sum_{N=0}^{\frac{x}{n_{ic}}} (1+r)^{-N \times n_{ic}} \qquad (3ai)$$

$$N = (i/n_{ic})\ i = 0 \sim x \qquad (3aii)$$

$$RC(x) = (RC_o) \sum_{i=0}^{x} (1+r)^{-i} \qquad (3b)$$

$$CC(x) = (\alpha \times S) \sum_{i=0}^{x} \{\alpha'(i) + [\alpha(i)/n_{cc}]\} \times (1+r)^{-i} \qquad (3c)$$

$$RCC(x) = (RCC_o \times S) \sum_{i=0}^{x} [\alpha(i)]\ (1+r)^{-i} \qquad (3d)$$

Figure 10. Rate of installation of in-vehicle units.

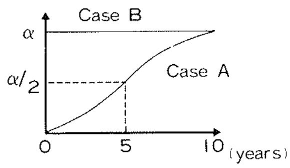


Figure 11. Annual trends in total costs and benefits.

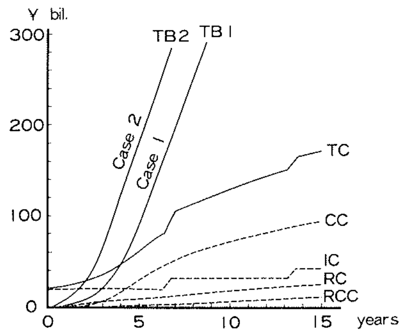
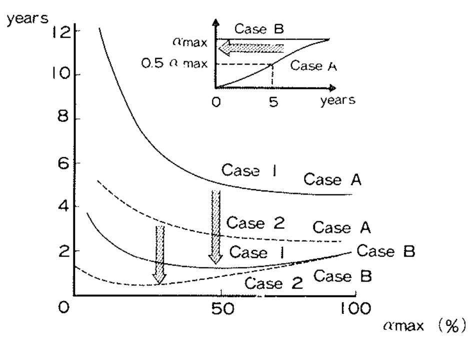


Figure 12. Social repayment years and the final diffusion rate of in-vehicle units.



where

IC_o = initial investment costs = ¥20 billion;
RC_o = annual operating costs = ¥2.5 billion;
RCC_o = annual cost of upkeep per in-vehicle unit = ¥1000/unit;
α = in-vehicle unit cost, including installation = ¥30 000/unit;
s = total number of vehicles in target area = 1.8 million;
r = social discount rate = 6 percent;
n_{cc} = durability of in-vehicle equipment = 6 years;
n_{ic} = durability of other than in-vehicle equipment = 7 years; and
$\alpha'(i)$ = rate of increase of automobiles that have in-vehicle units [$\alpha(i)$ after i years].

Two additional assumptions are made for this overall evaluation.

Reliance of Benefits on Guidance Rate (β)

The benefits are viewed only in terms of reduction of required driving time. Accordingly, by using the information from Figure 8 that expresses the reliance of the rate of reduction in required driving time on the guidance rate, the following formula will give us an approximation of the benefits:

$$B(\beta) = 80 \times (2-\beta) \times \beta(\text{¥}000\ 000\ 000\text{s/year}) \qquad (4)$$

Installation Rate of In-Vehicle Units (α) and Guidance Rate

Figure 10 shows two rates of installation of in-vehicle units. Case A shows a logistic curve that in five years reaches $\alpha/2$ and eventually reaches close to α. Case B shows a step function in which α is reached quickly in the first year.

Two other cases are examined: case 1, in which the percentages of β and α are equivalent, and case 2, in which in-vehicle units are installed in the target area beginning with vehicles that drive longer distances. The evaluation used Equations 2 and 3 and was based on the assumptions mentioned above. Figure 11 shows the annual trends for TC and TB in cases 1 and 2.

The number of years required for TB to equal TC is called the social repayment years. Figure 12 shows the relationship between x number of social repayment years and the final diffusion rate of in-vehicle units. If the final diffusion rate is supposed to be 100 percent, the social repayment period will be two to four years. Figure 12 shows that, even if the final diffusion rate of in-vehicle units is less than 100 percent, this system offers considerable benefits that meet the initial investment and operating costs within a couple of years.

ACKNOWLEDGMENT

This paper is contributed by permission of the Director General of the Agency of Industrial Science and Technology, Ministry of International Trade and Industry of Japan. The contents do not necessarily reflect the official views or policy of the Agency of Industrial Science and Technology.

REFERENCES

1. The Comprehensive Automobile Traffic Control System: A General Description of the Pilot Test System. Agency of Industrial Science and Technology, Ministry of International Trade and Industry, Tokyo, Japan, Oct. 1977.
2. S. E. Dreyfus. An Appraisal of Some Shortest-Path Algorithms. Bull. of the Operations Research Society of America, Vol. 1, Suppl. 1, Spring 1968, p. B-166.
3. Y. Anzai and others. Kth Shortest Path Algorithm in Time-Variant Network. Presented at 14th Symposium of the Society of Instrument and Control Engineers in Japan.
4. F. Kobayashi and M. Izumi. Comprehensive Automobile Traffic Control Project, Part 1: Improvement of Traffic Flow by Route Guidance System. 3rd Automotive Engineering Conference, Tokyo, Nov. 1977.
5. S. Stidham. Probabilistic Methods for Determining Travel-Time Functions in a Traffic Network. Cornell Univ., Ithaca, NY, Tech. Rept. 141.
6. K. Okudaira. Development Process and Application of Incremental Assignment Method. Operations Research of Japan, Vol. 22, No. 12, 1977, pp. 688-694.
7. S. Kokaji. Reduced Air Pollution by Route Guidance. Presented at International Federation of Automatic Control Symposium on Environmental Systems Planning, Design, and Control, Kyoto, Aug. 1977.
8. Social Loss of Road Accidents. Japan Research Society for Transport Policy, Oct. 1977.

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

Outline of the Comprehensive Automobile Traffic Control Pilot Test System

Nobuo Yumoto, Sumitomo Electric Industries, Ltd., Osaka, Japan
Hirokazu Ihara, Systems Development Laboratory, Hitachi, Ltd., Kawasaki, Japan
Tsutomu Tabe, Nippondenso Company, Ltd., Aichi, Japan
Masaru Naniwada, Central Research Laboratories, Nippon Electric Company, Ltd., Kawasaki, Japan

In the final phase of the Comprehensive Automobile Traffic Control Project, a pilot system was constructed and put into operation in October 1977. In the pilot area, which covers 28 km^2 in southwestern Tokyo, a two-way exchange of digital information occurs between instrumented vehicles and roadside equipment. The driver receives a visual display of routing information that is based on current traffic conditions and other information for safe driving. In order to provide the drivers of noninstrumented vehicles with the routing information, new roadside displays were developed. A roadside radio system, which gives the driver traffic information and provides instructions in an emergency, was also developed. This paper presents an overview of the pilot system and describes the procedure of optimum route finding, the specification of the digital communication link between road and vehicle, and the hardware features of the principal equipment. A preliminary evaluation of the system performance is also described.

The Comprehensive Automobile Traffic Control (CAC) Project (1, 2), sponsored by the Agency of Industrial Science and Technology of the Ministry of International Trade and Industry, was begun in 1973 for a six-year research period with a budget of seven billion yen [¥ 220 = $1.00 (August 1979)].

The objective of this project is to develop an integrated system that provides, by means of an improved communication medium between vehicles and the external world, such functions as optimum route guidance, diversion of traffic out of high-density pollution areas, priority for public service vehicles, advanced display of traffic regulations and alerts, and the potential for simultaneous communication in case of emergency. Such a system will contribute to an improved public acceptance of automobile transportation by reducing traffic congestion, traffic accidents, and air pollution.

The principal technical activities undertaken to achieve this objective were to develop (a) an optimum control strategy of traffic flow, as well as monitoring of vehicular movement and assignment of optimum route; (b) a high-speed and reliable two-way communication system between the individual vehicle and the ground; and (c) an excellent display method that considers human engineering factors. In 1974, prototype hardware was developed and evaluated in a preliminary experimental installation within the grounds of a factory.

Between 1975 and 1976, the software and hardware for the pilot system were manufactured and installed in a 28-km^2 test area in the southwestern part of Tokyo. The purposes of the pilot test include evaluating the effectiveness of the system in an actual urban situation, identifying potential problems regarding future adoption of the system, demonstrating the system to the public, and obtaining their opinions about it. The pilot system has been in actual operation since October 1977.

OUTLINE OF THE PILOT SYSTEM

Communication Between the Driver and the System

The pilot system has three means of communication between the driver and the ground:

1. Two-way digital communication link by inductive radio (when the vehicle's antenna passes over the road antenna, a data transmission of about 100 bits in each direction occurs via the magnetic field between both antennas),
2. Ground-to-vehicle oral communication by means of a roadside radio (radio transmission reaches only the vehicles that pass the roadside antenna, which extends along the street), and
3. Roadside visual display.

Functions of the Pilot System

The pilot system has five subfunctions:

1. Route guidance, which gives the driver a visual instruction at decision points and guides the driver to his or her destination in the same way that the Experimental Route Guidance System (ERGS) does (3) [the CAC system differs from ERGS in that the route is periodically optimized on the basis of the current traffic and road conditions and in that the system has functions other than route guidance];
2. Driving information, which gives the driver a visual display of information useful for safe driving and gives him or her a sonic alert when the vehicle exceeds the speed limit according to ground control;
3. Priority for public service vehicles, which gives priority to public service vehicles at the signalized intersections and improves the operating efficiency of such vehicles by continuous monitoring of their individual locations;
4. Urgent information, which gives the driver instructions in an emergency and provides audio information on localized traffic and road conditions in usual situations; and
5. Roadside route display, which gives drivers of noninstrumented vehicles the routing information.

The route guidance, driving information, and priority for public service vehicles subsystems use the digital communication link by inductive radio. The vehicle type, its identification number, and destination code are sent from the vehicle to the ground. Routing and driving information are sent in the opposite direction. At some locations where only driving information is necessary, downward data are neglected by the ground.

The urgent information subsystem uses the roadside radio. The audio medium of communication supplements the visual display of route guidance and driving information and provides the driver with more versatile information than digital communication does.

Configuration of the Pilot System

Arterial streets and expressways of about 100 km in total length within an area surrounded by Loop Road No. 7 and the national highways Route 1 and Route 246 were chosen as the test site.

Figure 1 shows the layout of the pilot test area. At 74 intersections on the arteries and 15 forks on the expressways, instrumented vehicles exchange digital data with roadside equipment for the route guidance and driving information subsystems. In addition, 15 locations are equipped with roadside equipment exclusively for the driving information subsystem. At 9 locations, 400-m-long leaky coaxial cables are installed as roadside radio antennas. Along an artery bound for the central business district, three roadside displays are installed.

The roadside equipment for the route guidance and driving information subsystems is connected via leased, duplex 1200-band lines to a communication control computer located at the control center. Routing information and driving information are sent from the center to the roadside equipment, and the data received from the vehicles are sent in the opposite direction. About 180 vehicle detectors are installed, 40 of which are connected to the communication control computer through leased 50-band lines and a central preprocessor; 130 others are connected to the roadside equipment, preprocessed, and then sent to the center with other information.

The data from the communication control computer are fed to the main computer, which estimates travel time along each link of the road network, based on the time of data exchange with each vehicle at each communication point and on the traffic volume and occupancy, measured by means of vehicle detectors. The main computer keeps a data base on the road network and interfaces with the operators through an operation control computer.

The main computer sends the travel time data to a pair of minicomputers or to a traffic network simulator that finds minimum travel-time routes and generates guide tables. The guide tables are sent to the roadside equipment through the communication control computer.

The contents of the roadside route display subsystem and the messages of the urgent information subsystem are manually controlled. The main computer makes suggestions about the contents and messages based on the data it keeps. The contents of the roadside route display subsystem are monitored at the traffic control center of the Metropolitan Police Department. An agreement between the CAC system operator and the police officer on duty is necessary to effect the display.

In-vehicle units for route guidance and driving information are mounted in 330 vehicles, 290 of which are owned by voluntary participants in the test. The remaining 40 are operated exclusively as research vehicles; these are also equipped with urgent information subsystem radio receivers. In order to obtain travel-time measurements, an additional 1000 vehicles are equipped with in-vehicle units that do not have display capability.

The pilot system was recently expanded to include a stretch of expressways between the new Tokyo International Airport in Narita and the City Air Terminal in downtown Tokyo (Figure 2). At seven places along the road, roadside equipment was installed and connected to a line concentrator and a minicomputer in Chiba Operations Center of the Japan Highway Public Corporation. Transceivers were installed on about 70 limousines that commute nonstop between the airport and the City Air Terminal. By means of these installations, travel time between each point is obtained about every 15 min. The information is transferred to the CAC system center via a leased line to monitor and analyze the data.

DISPLAY INSIDE VEHICLES

Figure 3 provides a conceptual schematic drawing of the route guidance and driving information subsystems (4). Figure 4 shows the inside view of an instrumented vehicle. The driver enters a seven-digit destination code into the encoder, informing the system of any option regarding expressway use by pressing the "expressway option" button. The expressway option, as well as the type of the vehicle, is taken into account in route selection.

The data received at a communication point are stored on board the vehicle and displayed step by step as shown in Figure 5. First, the display indicates the entrance lane to take before entering the intersection and the direction of turning movement at that intersection. Second, by keeping the indication of the turning direction, the display indicates the exit lane to take after departing from the intersection. This is intended to ease the maneuver at the next intersection. Third, the entrance lane and the turning direction to be taken at the next intersection are indicated, if necessary. The distance between the communication point and the places where the indications should be changed are controlled by data from the ground. A chime sounds when the indication changes.

The turning direction is indicated superimposed over the shape of the intersection (up to seven shapes exist). After departing from a route guidance intersection, a "go straight" indication is held until the vehicle approaches the next one. The final indication of the turning direction flashes when the vehicle approaches the destination. An up or down indication at a grade separation, entrance or exit indication at an expressway ramp, and indications of detour, impossibility of guidance, and transmission error are also displayed.

The driving information subsystem consists of warning information and speed-limit violation alert. The warning information is indicated in characters. The appropriate instructions from the following six are selected and displayed for about 10 s: (a) pedestrian crossing, (b) reduce speed, (c) changing road width, (d) stop and go, (e) priority lane, and (f) road work ahead.

An "alert" sound warns the driver when the speed of his or her vehicle exceeds the speed limit, according to the ground control.

ROUTING CALCULATION

Structure of Road Network Model

Each of the inbound and outbound links between adjacent route guidance intersections is called a point. A unique seven-digit destination code is given to each point (see Figure 6). Paths connecting adjacent points are called arcs. An arc is directional.

The whole Japanese road network is divided into about 100 regions as shown in Figure 7. Each region is subdivided into as many as 63 sections. Each section is subdivided into as many as eight zones. Each zone contains up to 63 points.

The pilot test area consists of five sections and contains 330 points and 680 arcs.

Approximation of Network Model

In order to reduce the computational load without an adverse effect on the quality of route selection, the network is approximated in a way that is accurate near the origin and becomes more simplified going away from the origin.

As shown in Figure 8, in an area composed of the section to which the origin belongs and the sections that circumscribe the origin section, the network is not approximated. We shall call this area the nonsimplified area.

Figure 1. Pilot test area.

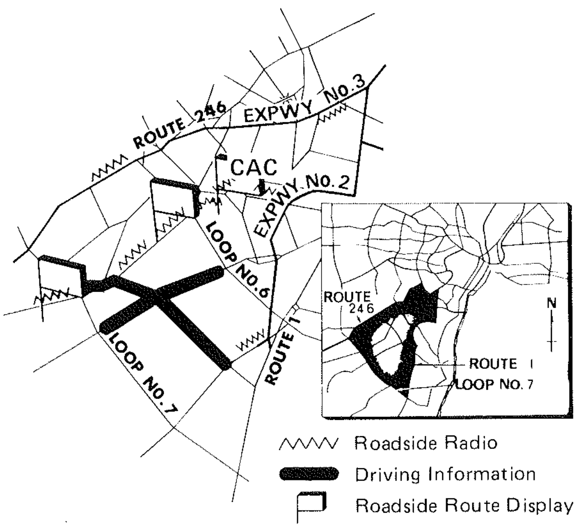
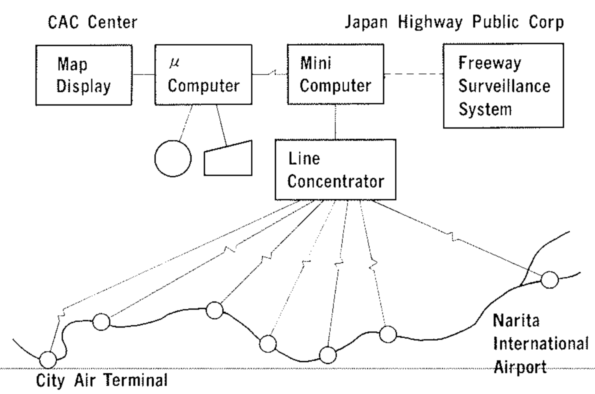


Figure 2. Travel time measurement between Narita Airport and City Air Terminal.



In the sections that circumscribe the nonsimplified area, structure of the network within a zone is neglected, and a zone is treated as a point and represented by a representative point in it. A representative point is tied to the neighboring representative points and to the border of the nonsimplified area by hypothetical arcs called regenerated arcs. Remaining sections in and regions outside the origin region are treated as points. The travel time along a regenerated arc between zones is determined by calculation of the travel time along the optimum route on the actual network.

Route Calculation

Travel time to a certain point from the origin point along the route is defined as the route travel time of the point. Branches are extended from the origin point to all of the succeeding points. From those points, branches are further extended successively. When two arcs run into the same point, the arc that gives the smaller route travel time is included in the route and the other arc is excluded from the route. The above procedure is repeated until all the points are reached.

Since certain arcs inhibit the types of vehicles that may pass, the optimum route depends on the vehicle type. Taking into consideration that the route from the origin to the inhibited arc is common to both vehicle types, the route calculation is carried out only for the portion of the network beyond the inhibited arc.

This principle is also applied to treatment of the expressway option, because avoidance of an expressway can be considered to block the arc entering an expressway.

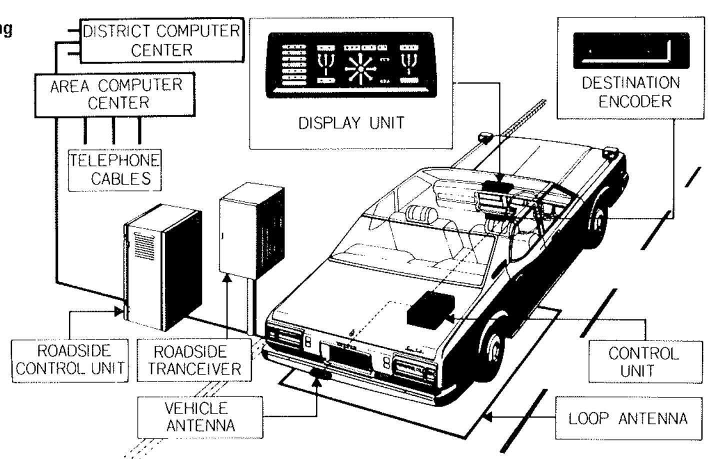
The pilot test system has two alternative means of route calculation: a software mode with a pair of minicomputers and a hardware mode with a traffic network simulator.

Guide Table

As a result of the routing calculation, a list of routing information for each destination is obtained for each intersection. The routing information is a series of exit links, each of which corresponds to a certain combination of the entrance link, expressway option, and vehicle type.

In respect to the routing information, destination codes are grouped and arranged into a tree structure called the destination table, in which the hierarchy of code corresponds to the structure of the tree. The terminal branch of the tree points entries to the entrance link table, which in turn points to the vehicle-type table.

Figure 3. Schematic drawing of route guidance and driving information subsystems.



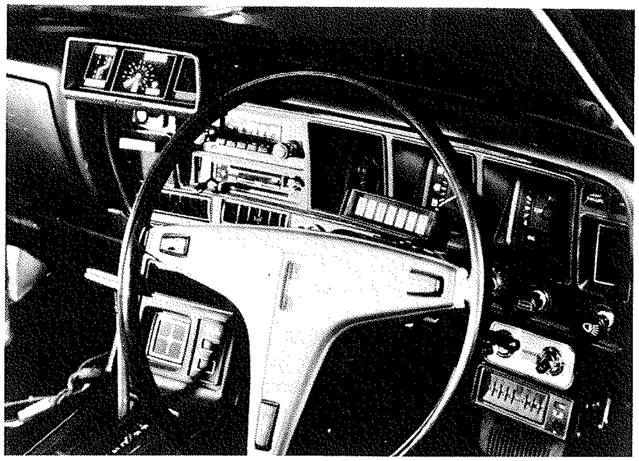


Figure 4. Display and encoder in the vehicle.

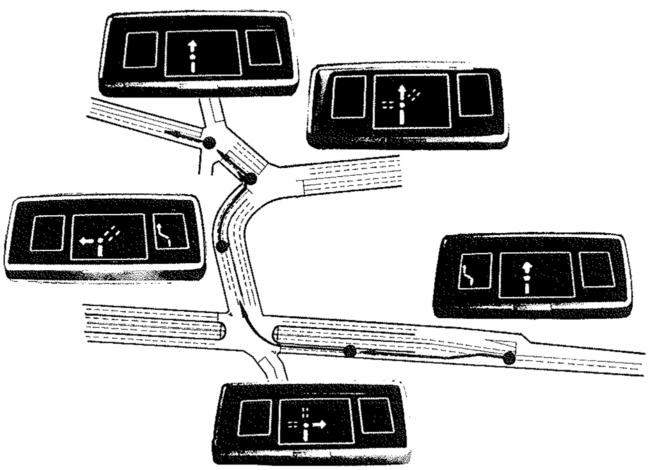


Figure 5. Display of routing instructions.

When the routing information differs by expressway option, the entrance link table is indirectly pointed via an expressway ramp table. A guide table, composed of the destination table, the expressway ramp table, the entrance link table, and the vehicle-type table, is generated for each intersection and sent to the roadside equipment (Figure 9).

DIGITAL COMMUNICATION BETWEEN VEHICLE AND GROUND

The vehicle transmits to the ground the request data, which consist of a 28-bit destination code, 1-bit expressway option, 31-bit vehicle identification number, and 8-bit vehicle-type code. In response, the ground sends a 58-bit route guidance message and a 12-bit driving message. The length of the message in each direction is 81 bits, including dummy and parity check bits (Figure 10).

The message is encoded into a modified NRZ code, which modulates the phase of the carrier. The carrier frequency is 172.8 kHz for vehicle to ground and 105.6 kHz for ground to vehicle. The transmission speed is 4800 bits/s. A parity check bit is inserted every 8 data bits.

The ground antenna is a 1-turn loop 2.5 m wide and 3 m long. The vehicle antenna is a 30-turn loop of 100x50 mm that is attached under the rear bumper to

Figure 6. Road network model.

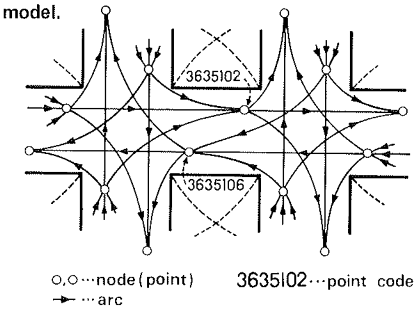


Figure 7. Nationwide region code.

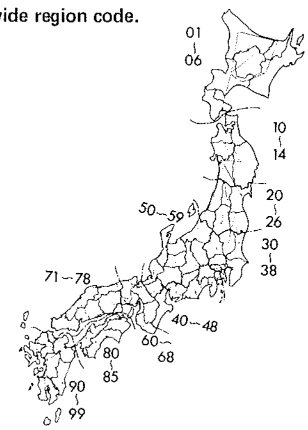
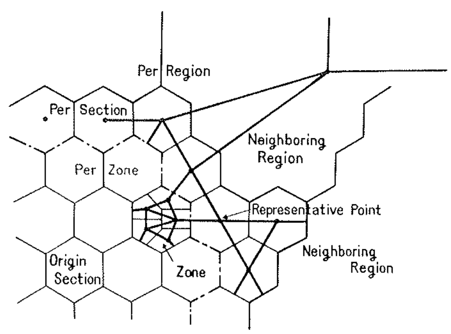


Figure 8. Approximation of network.



be parallel to and 45 cm above the ground loop.

The data transmission procedure is shown in Figure 11. The vehicle, having recognized the ground antenna by receiving the carrier wave from the ground, sends the request data, preceded by a header. The ground generates the response data and transmits to the vehicle. If the vehicle has received erroneous or invalid data, or if time is out, it sends the request data once more. If the ground detects an error in the data it received, it acknowledges the vehicle by sending invalid data. The longitudinal length of the ground antenna allows retransmission.

When the vehicle receives valid data, it locks itself and inhibits further communication until it travels a certain distance, which is prescribed by the data from the ground.

Figure 12 shows the pattern of coupling between the ground and vehicle antennas. The pattern is not symmetric because of the body effect of the vehicle; it varies vehicle by vehicle. The vehicle is given a higher receiving threshold (relative to the level at the center of the ground antenna) than the ground, in order to assure that the ground is ready to receive when the vehicle sends data after detecting the carrier from the ground. By this arrangement of level setting, we can do without the trigger loop that was employed in ERGS (3). The principal values of the transmission level are shown below:

Figure 9. Guide table.

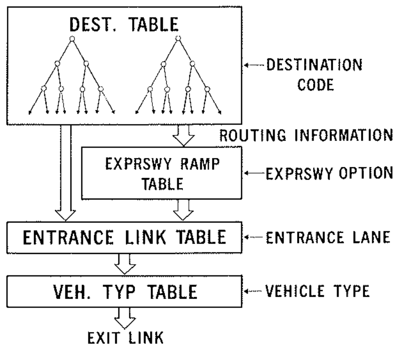
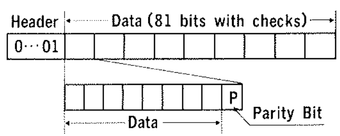


Figure 10. Frame configuration of vehicle-ground transmission.



Item	Vehicle to Ground	Ground to Vehicle
Antenna transmitting current (mA)	75	14
Field strength at receiving antenna (V/m)	0.2	1.9
Reception (mV)		
Induced voltage	2.5	0.6
Receiver threshold	0.5	0.3

A vehicle unit consists of a vehicle antenna, a transceiver and control, a display unit, an encoder, a power supply unit, and such accessories as wire harnesses and an odometer senser. The transceiver and control stores the data received and sends the messages to the display unit in accordance with the timing instruction sent from the ground.

Figure 13 shows a diagram of the roadside equipment. The roadside equipment receives geometric information about the intersection from the center at the initial stage and the guide table every 15 min (Figure 14). When the roadside equipment receives request data from a vehicle, it looks up the guide table, by indexing the destination, expressway option, vehicle type, and the entrance lane, and obtains the exit link. The instruction to the vehicle is obtained by indexing the entrance lane and exit link and looking up the intersection geometric table.

TRAFFIC NETWORK SIMULATOR SYSTEM

The Traffic Network Simulator System (TNSS) is a special-purpose computer to search optimal routes (minimum-time routes) in a road network. The TNSS

Figure 11. Procedure of vehicle-ground transmission.

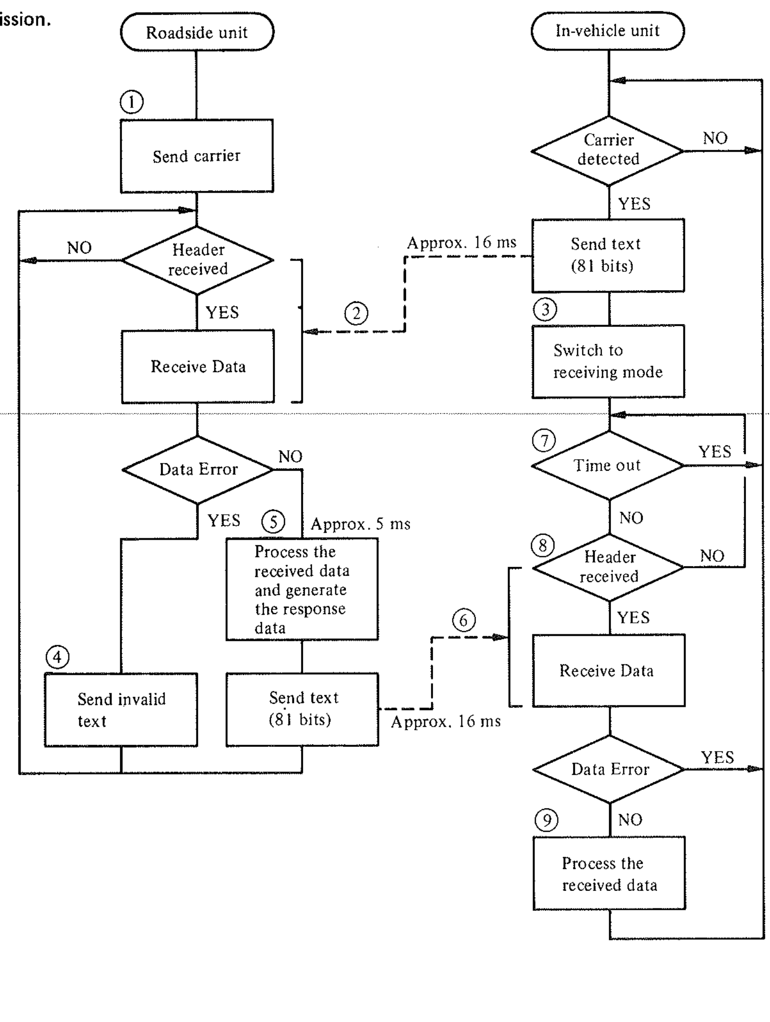


Figure 12. Coupling pattern between vehicle and ground antennas and setting of threshold level.

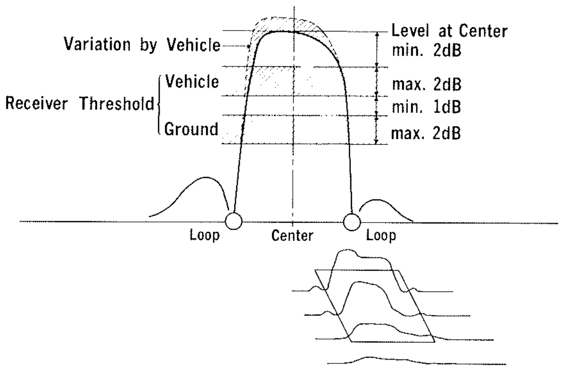


Figure 13. Diagram of roadside equipment.

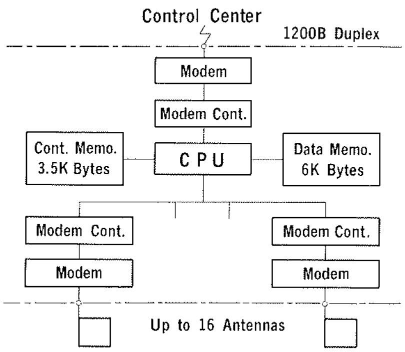
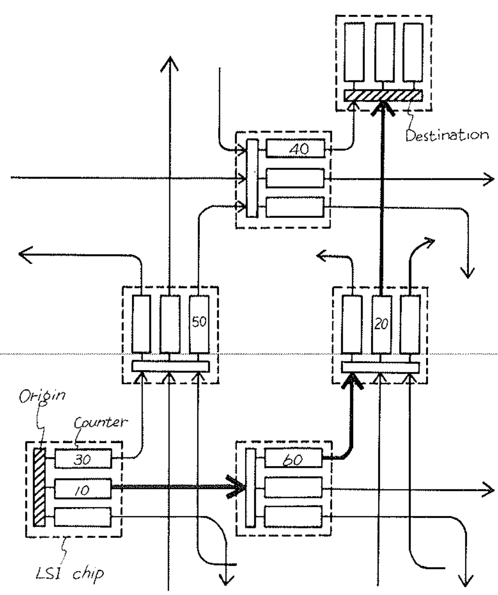


Figure 14. Principle of route search in TNSS.



has a large-scale integration (LSI) network in which newly developed LSIs are connected in a way analogous to the route guidance road network of the pilot area and its circumference. The optimal routes are computed by simulating trips in the actual road network by means of the propagation of electric pulses in the circuit, as is shown in Figure 14. TNSS executes the following information processing every 15 min:

1. Prediction of traffic flow: TNSS predicts the condition of traffic flow in the road network up to 1 h ahead based on past and recent traffic flow data (collected through the vehicle detectors). Then it computes the arc travel time for every arc of the road network.
2. Search for optimal routes: TNSS searches for the optimal routes from one origin to all destinations by using the travel-time data computed above. Repeating this process, the optimal routes are determined from all origins to all destinations.
3. Formation of guide tables: TNSS compiles the results of the optimal-route search in the form of the guide table for each intersection. Figure 15 shows the configuration of TNSS; the traffic network is an LSI network analogous to the actual road network.

In the pilot system, the route search can be performed either in the general-purpose computer mode, which adopts the algorithm described above, or in the special-purpose hardware mode, by means of TNSS.

ROADSIDE RADIO

Leaky coaxial cable (LCX) about 400 m long is wired at 5-m height along the street for broadcasting the urgent information subsystem. As a vehicle passes by this stretch of the street, the broadcast is received through an adaptor attached to the automobile's radio. Unlike regular commercial radio broadcast, the urgent information reaches only the vehicles in the vicinity of the LCX. It is thus possible to broadcast different messages in different places at the same time (Figure 16).

There are two classes of information priority. The class is identified by means of a pilot tone superimposed on the voice signal. When the class is of the first priority, the adaptor makes the vehicle's radio receive the urgent information, even if the radio has been turned off or tuned to a commercial broadcast. When the class is not of the first priority, the information can be heard only by drivers who have turned on a selecting switch on the adaptor.

The LCX antenna is divided into two sections. Different pilot tones are fed to both sections. The adaptor can recognize the direction of movement by the sequence of the pilot tones. In this way it is possible to provide the information only to the vehicles running in the direction the system specifies.

The voice message of broadcast is automatically synthesized according to the button operation by the operator and sent to the roadside transmitter via leased lines. The frequency used is 450 MHz, and the output power of the transmitter is 2 W. The roadside antennas are installed at nine locations.

ROADSIDE ROUTE DISPLAY

An approximated road network is shown on the roadside route display board (Figure 17). The information is displayed by changing the colors of the links of the network. There are three categories of display:

1. Display of congestion: the colors of the links change to yellow or red according to the degree of congestion,
2. Display of optimum route: the links along the recommended route are green, and
3. Display of an accident: the location of a traffic accident is indicated by a red link, and a red light flashes at the corner of the board.

Figure 15. Configuration of TNSS.

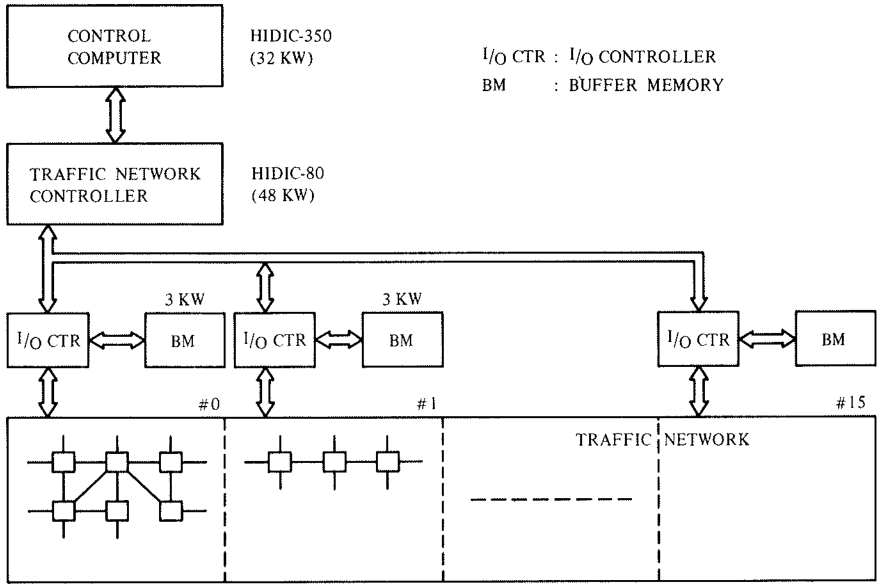


Figure 16. Configuration of roadside radio system.

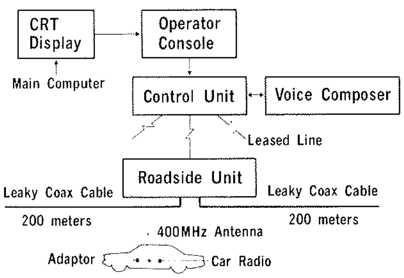


Figure 17. Roadside route display board.

The link on the board is a rotary drum that is divided longitudinally by four colors.

MILLIMETER-WAVE COMMUNICATION MODE

For communication between vehicles and roadside units, it is anticipated that a need to transmit more extensive information will arise in the future. Transmission of a large volume of information (2 M bits/s) over a millimeter-wave band (60 GHz) is being tested at one of the intersections in the pilot test area. By using the millimeter-wave communication mode, it is possible to transmit 160 kilobits of visual information to a vehicle traveling up to 100 km/h.

For this pilot test system, transmission data are two 128x128-dot pictures and eight 256x256-dot pictures compressed to $\frac{1}{4}$ in data length. An on-board television tube displays, by using two 128x128-dot pictures, a drawing of the approaching intersection on which a flashing arrow indicates the direction to be taken (Figure 18). At the same time, eight different images (256x256-dot pictures) can be reproduced on recording paper in an in-vehicle recording device. The vehicle antenna is fixed on the roof of the automobile. The ground antennas are installed under a pedestrian bridge (5, 6).

CONTROL CENTER

The control center has a control room (Figure 19), a machine room, and a pilot test headquarters. In the control room, equipment for display [including cathode-ray-tube (CRT) display] and operation are installed, and in the machine room are one large-scale computer (NEAC2200/375), four minicomputers (NEAC3200/70), one traffic network simulator with a control computer (HIDIC350), and some other control units.

Man-Machine Interface Equipment

The major equipment in the control room are the wall display and an operating console. The wall display presents information that all operators require and is controlled by the supervisory operator. The devices of the wall display and the information carried are shown below.

Device	Information
Central screen display (2.0x1.5 m)	Pilot area maps, with a video-projecter in the rear
Color character CRT	Roadside unit status, indicated by the color of the unit number
Flap-type indicator	Title of map projected on the screen

Figure 18. Picture received in the vehicle by millimeter-wave communication.

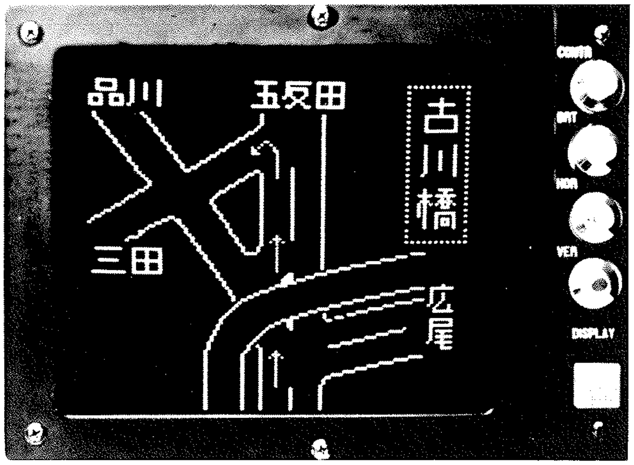
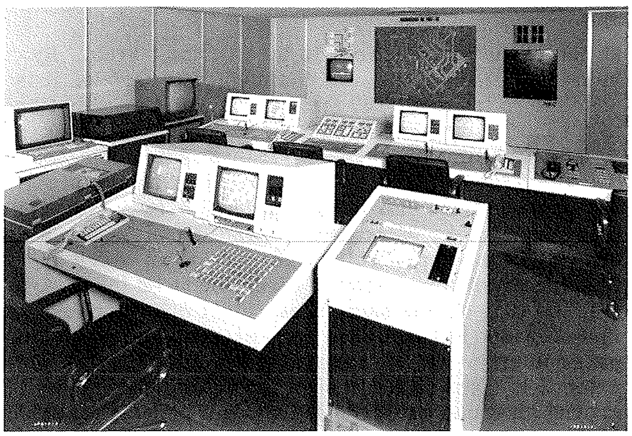


Figure 19. Control room.



Device	Information
Light-reflecting dot-matrix indicator	Date and time
Lamp indicator	Individual computer status
Lamp panels	Functions in operation, weather, and route-search mode

There are four operating consoles: one for the supervisory operator; one for the operator of the route guidance, driving information, and public-service-vehicle priority functions; one for the operator of the roadside route display; and the last for the operator of urgent information messages. These consoles are all equipped with a character CRT, function keys, typewriter keys, and lightpens. The first two are also equipped with a graphic CRT, joy sticks, flap-type indicators, telephones, and radio transceivers.

The maps that are shown on the central screen display and the graphic CRTs are listed below, along with the information they carry; each operator can display any of the maps on his or her own CRT:

1. Road map of the entire area: (a) degree of congestion (light = green, heavy = yellow, congested = magenta), (b) location of test vehicles (not present = green, present = magenta), (c) flow rate of test vehicles (low = green, medium = yellow, high = magenta), (d) vehicle tracking (vehicle 1 = red, vehicle 2 = yellow), and (e) optimal route (route 1 = red, route 2 = yellow);
2. Zoomed-up map: degree of congestion and the location and identification of test vehicles; and
3. Intersection maps: geometric figure of each intersection.

Control Operation

The control operation has a two-level hierarchical structure. The lower-level control operation is for the monitoring and control of each of the five functions of this system, e.g., route guidance, driving information. The higher-level one is for the supervisory control of the entire system. Operations performed on the supervisory console are

1. Initial system set-up,
2. Starting and stopping the overall operation or partial function of the system,
3. Changing system parameters and configuration,
4. Monitoring traffic condition, and
5. Monitoring system and device status.

Operations performed on the route guidance, driving information, and public-service-vehicle priority console are

1. System function interventions after a traffic accident,
2. Roadside units' status diagnosis,
3. Monitoring traffic condition, and
4. Monitoring system and device status.

Operations performed on the urgent information console are

1. Requesting recommendation for broadcast messages from computer,
2. Setting a message for each roadside unit, and
3. Monitoring system and device status.

Operations performed on the roadside route display console are

1. Requesting recommendation of displays,
2. Setting a display pattern for each roadside board,
3. Confirming the remote station's acknowledgment, and
4. Monitoring system and device status.

The supervisory console can be substituted for the route guidance, driving information, and public-service-vehicle priority console.

THE PILOT STUDY

The pilot test has been carried out since October 1977. The initial three months of the one-year test period were dedicated to long-term trial operation of the system and to collection of information in regard to the characteristics of traffic flow in the test area.

Besides the evaluation of route guidance performance, data were obtained on such items as capability of manual intervention in route selection, computing time for route calculation, and the effect of network approximation. The visibility and comprehensibility of the display were measured, and drivers' opinions were collected.

The performance of route guidance was measured by comparing the travel time between guided and nonguided

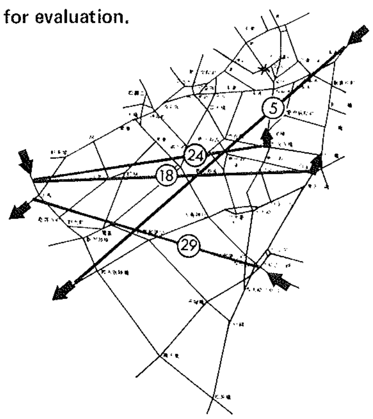


Figure 20. Selected OD pairs for evaluation.

Table 1. Comparison of travel time for guided and nonguided vehicles.

Pair	Mean Travel Time(s)	Advantage Factor	Reduction of Travel Time (%)
24		76.3	15.5
Guided	1239		
Nonguided	1467		
29		72.1	9.4
Guided	1327		
Nonguided	1464		
5		76.1	10.8
Guided	1763		
Nonguided	1976		
18		75.0	9.3
Guided	1657		
Nonguided	1837		

vehicles. About 1000 trials were made between seven origin-destination (OD) pairs (Figure 20). As is shown in Table 1, travel time has been reduced by 9-15 percent.

ACKNOWLEDGMENT

This paper is contributed by permission of the Director General of the Agency of Industrial Science and Technolog Ministry of International Trade and Industry of Japan. The contents do not necessarily reflect the official views or policy of the Agency of Industrial Science and Technology.

REFERENCES

1. M. Onda. Comprehensive Automobile Traffic Control Project, Part 1: Outlines of the Pilot Test System. 3rd Automotive Engineering Conference, Tokyo, Nov. 1977.
2. M. Onda. Comprehensive Automobile Traffic Control: Route Guidance and Other Subsystems. 8th International Road Federation World Meeting, Tokyo, Oct. 1977.
3. A Design for an Experimental Route Guidance System, Volume I: System Description. General Motors, GMR-815-I, Nov. 1968.
4. M. Onda. Safety Design in the Comprehensive Automobile Traffic Control System. International Federation on Automatic Control Symposium on Environmental Systems Planning, Design, and Control, Kyoto, Aug. 1977.
5. T. Usubuchi, K. Iinuma, and T. Ishiguro. Data Compression of Facsimile Signals by Cascade Division Coding. IEEE International Conference on Communication, 1975.
6. Y. Matsuo, Y. Akaiwa, and I. Takase. A Compact 60-GHz Transmitter-Receiver. IEEE Transactions on Microwave Theory and Techniques, Nov. 1976.

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