TRANSPORTATION RESEARCH RECORD

No. 1360

Highway Operations, Capacity, and Traffic Control

Traffic Operations

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1992

Transportation Research Record 1360
Price: $28.00

Subscriber Category
IVA highway operations, capacity, and traffic control

TRB Publications Staff
Director of Reports and Editorial Services: Nancy A. Ackerman
Senior Editor: Naomi C. Kassabian
Associate Editor: Alison G. Tobias
Assistant Editors: Luanne Crayton, Norman Solomon, Susan E. G. Brown
Office Manager: Phyllis D. Barber
Production Assistant: Betty L. Hawkins

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Traffic operations.
p. cm.—(Transportation research record, ISSN 0361-1981; no. 1360)
Research papers from the 71st Annual Meeting, Transportation Research Board, Washington, D.C., January 1992.
"A peer-reviewed publication of the Transportation Research Board."
ISBN 0-309-05223-8
1. Traffic engineering—Congresses. 2. Traffic engineering—United States—Congresses. I. National Research Council (U.S.). Transportation Research Board. II. Series: Transportation research record; 1360.
TE7.H5 no. 1360
[HE332]
388 s—dc20
[388.3'1] 92-32759
CIP

Sponsorship of Transportation Research Record 1360

GROUP 3—OPERATION, SAFETY, AND MAINTENANCE OF TRANSPORTATION FACILITIES
Chairman: H. Douglas Robertson, University of North Carolina-Charlotte

Facilities and Operations Section
Chairman: Jack L. Kay, JHK & Associates

Committee on Transportation System Management
Chairman: Wayne Berman, Federal Highway Administration, U.S. Department of Transportation
Paul N. Bay, Robert C. Blakey, Charles C. Crevo, J. R. Doughty, Gary R. Erenrich, Larry R. Goode, Thomas J. Higgins, Rajendra Jain, Harvey R. Joyner, J. Richard Kuzmyak, Steven Z. Levine, Herbert S. Levinson, Jeffrey A. Lindley, Timothy J. Lomax, Michael D. Meyer, Susan L. Moe, Jouko A. Parviainen, David C. Rhinard, George J. Scheuernstuhl, Susanne Pelly Spitzer, Andrzej B. Tomecki, Cy Ulberg, Douglas W. Wiersig

Committee on Freeway Operations
Chairman: Ronald C. Sonntag, Wisconsin Department of Transportation
Secretary: Jeffrey A. Lindley, Federal Highway Administration
B. Beukers, Peter M. Briglia, Jr., Donald G. Capelle, Glen C. Carlson, Robert F. Dale, Conrad L. Dudek, Walter M. Dunn, Jr., Jack L. Kay, P. R. Korpal, Walter H. Kraft, Steven Z. Levine, Louis E. Lipp, Robert E. Maki, Joseph M. McDermott, Nancy L. Nihan, Colin A. Rayman, James R. Robinson, David H. Roper, James F. Shea, Henry B. Wall III, Thomas C. Werner, Sam Yagar, Michael J. Zezeski, Philip Zove

Committee on Traffic Signal Systems
Chairman: Herman E. Haenel, Advanced Traffic Engineering
Secretary: Alberto J. Santiago, Federal Highway Administration
A. Graham Bullen, E. Ryerson Case, Edmond Chin-Ping Chang, David J. Clowes, Robert A. de Santo, Donald W. Dey, Gary Duncan, Gary W. Euler, Nathan H. Gartner, Peter Hakkesteegt, Robert David Henry, Dallas W. Hildebrand, Paul P. Jovanis, Les Kelman, Alfred H. Kosik, Joseph K. Lam, Feng-Bor Lin, David C. Powell, Raymond S. Pusey, Dennis I. Robertson, Lionel M. Rodgers, Stephen Edwin Rowe, Miroslav Supitar, James A. Thompson, Charles E. Wallace

Committee on High-Occupancy Vehicle Systems
Chairman: Donald G. Capelle, Parsons Brinckerhoff
Secretary: Dennis L. Christiansen, Texas A&M University
Richard D. Bauman, Paul N. Bay, John Bonsall, A. Frank Cechini, Jr., Donald J. Emerson, Charles Fuhs, David R. Gehr, Alan T. Gonseth, Leslie N. Jacobson, Theodore C. Knappen, James R. Lightbody, Timothy J. Lomax, Brian Pearson, Allan E. Pint, Lew W. Pratsch, Gary C. Price, William T. Roach, Morris J. Rothenberg, Sheldon G. Strickland, David L. Tollett, Gary K. Trietsch, Katherine F. Turnbull, Leonard J. Whitlock

Task Force on Transportation Demand Management
Chairman: Michael D. Meyer, Georgia Institute of Technology
Robert M. Baron, Paul N. Bay, Daniel E. Benson, Carolyn Di Mambro, Robert T. Dunphy, Gary W. Edson, Erik T. Ferguson, Gary Hawthorn, Thomas J. Higgins, Mary L. Macinnes, Paul F. Maxwell, G. Scott Rutherford, Eric N. Schreffler, Edward L. Thomas, Mary Lynn Tischer, Katherine F. Turnbull, Philip L. Winters

Richard A. Cunard, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1991.

Transportation Research Record 1360

Contents

Foreword

The papers in this Record are related by their focus on issues related to traffic operations. Readers with a specific interest in transportation systems management will find papers about traffic management during a bridge reconstruction project, prioritization and cost effectiveness analysis of signalized intersection improvements, and effectiveness of improving arterials as freeway reliever routes.

Readers interested in freeway operations or high-occupancy vehicle (HOV) facilities will find papers on simulation modeling for freeway design, improving freeway efficiency and capacity by removing bottlenecks, evaluating traffic-responsive ramp metering strategies, evaluating video image processing systems, reviewing the performance of incident-detection algorithms, the development of a traffic operations center simulator, comparison of HOV facilities between projects in North America and other parts of the world, analysis of the institutional arrangements that have resulted in successful HOV projects, and the role of HOV facilities in a multimodal transportation system.

Traffic signal systems design and timing developments are the focus of a series of papers. Included are papers on a method for comparative evaluation of alternative traffic control strategies; a forward progression–based optimization model in TRANSYT-7F; a proposed framework for a hierarchical design of a real-time traffic control system; an analytical process to determine appropriate actuation delay for right-turn-on-red locations; a new computerized traffic signal system design for real-time traffic management; enhancements to TRANSYT-7F for fuel consumption, emissions, and users costs; optimizing traffic signal coordination during unsaturated conditions; and the concept of a regional arterial for short and medium-length trips. An examination of the indicators of congestion level is also included.

Abridgment

Traffic Management During the I-195 Providence River Bridge Repair Project

STEPHEN A. DEVINE, JOSEPH A. BUCCI, AND DANIEL J. BERMAN

Many states in the Northeast no longer can overlook their aging, overly congested urban highways. However, rebuilding these highways is not simply design and construction. Today's major highway reconstruction projects require the skills of a wide range of transportation professionals, responsible not only for rebuilding the highway, but also for managing the transportation needs of the affected commuters. Realizing this, the Rhode Island Department of Transportation (RIDOT) called on traffic engineers, transportation planners, transit officials, local police officers, community liaison staff, and public relations consultants to develop a traffic-management plan for the I-195 Providence River Bridge Repair Project. To meet the needs of commuters in Rhode Island and Southeastern Massachusetts, traffic management strategies were compiled, selected, and implemented to move people through and around the project area. These strategies included education (public information and community liaison activities), traffic engineering (coordination of traffic control signals and emergency routes), enforcement (monitoring and enforcing speed limits), and ridesharing and transit (carpool matching, free bus service and park-and-ride lots). The traffic management strategies presented are not all-inclusive. They are intended to present the strategies and actions that were both successful and unsuccessful for RIDOT. Each reconstruction project is unique. Traffic-management strategies must be selected to conform to the project area and the project's characteristics.

In developing strategies for traffic management, it was particularly important to consider the packaging and compatibility of the action with other current highway construction projects and the tourism industry in both Rhode Island and Massachusetts. Already apparent for parking and traffic circulation in, around, and approaching the downtown Providence area was the adverse short-term impact of a major Capital Center construction project less than 1 mi away. Further impacts were unknown but could be anticipated from several other construction projects in the downtown area. The solution to these issues could result in major gridlock in the cities of Providence and East Providence.

As a result of these concerns, the Federal Highway Administration (FHWA) Division Office proposed an aggressive plan of mitigative measures described as "the four E's": engineering, enforcement, education (public relations effort), and emergency actions.

ENGINEERING STRATEGIES

The first step was to use existing traffic counts and run the Texas queuing model QUEWZ (*1*). The magnitude of the problem at this early design stage surprised all of the participants and demonstrated the need to formulate a transportation systems management (TSM) plan.

Existing peak-hour volumes were exceeding 5,100 vehicles per hour and were carried by three lanes with many on-off merges and weave points. Both inbound and outbound traffic volumes were nearly equal in this 1-mi section of the Interstate. If a major disruption occurred, traffic could potentially back up into Massachusetts and create downtown gridlock. In addition, during the height of the summer season, Cape Cod-bound traffic would make weekend construction almost as difficult as weekdays and could prove counterproductive.

For these reasons it was felt that the maximum diversion that could be expected would be 500 vehicles per hour during the 8:00 a.m. to 9:00 p.m. hours. The overall goal would be to divert 5 percent of the total daily volume or approximately 7,000 vehicles per day. Since the project would still have at least three open lanes in each direction, it would be essential to optimize the capacity of these three narrow lanes. To achieve this goal it would be necessary to remove many conflicting weaves and on-off movements as well as slow the traffic down to 30 mph. A type of "poor man's" ramp-metering would be used by police enforcement.

Three 11-ft lanes of traffic in each direction were maintained between 6 a.m. and 7 p.m. Two open lanes in each direction were maintained at night between 7 p.m. and 6 a.m. To facilitate increased traffic flow on the Memorial and Labor Day weekends, three lanes were open from 6 a.m. Friday morning to 7 p.m. Tuesday evening. For the July 4th holiday, three lanes remained open from 6 a.m. Friday, June 30, to 7 p.m., Wednesday, July 5th.

The project was reconstructed in four stages, which enabled the contractor to maintain three lanes of traffic in each direction during each stage. Each stage included an individual completion date with liquidated damages of $2,500 per day for work beyond that date.

Park-and-Ride Facilities

The Rhode Island Department of Transportation (RIDOT) maintains 21 park-and-ride lots on major routes throughout the state. During the planning phase, RIDOT identified a total of 21 additional sites, both in Rhode Island and Massachusetts, for consideration as commuter lots during the construction period. Most of the sites required RIDOT to lease private property. Because of the time constraints imposed for the planning of traffic mitigation alternatives, only five sites were selected and only one required leasing of private prop-

S. A. Devine and J. A. Bucci, Planning Division, Rhode Island Department of Transportation, Two Capitol Hill, Providence, R.I. 02903. D. J. Berman, Federal Highway Administration, 380 Westminster Mall, Providence, R.I. 02903.

erty. Free express-bus service was provided at four of the sites.

Free Express Bus Service

To entice as many East Bay and southeastern Massachusetts commuters as possible to park their vehicles and use mass transit, the Rhode Island Public Transit Authority (RIPTA) provided free express-bus service from four RIDOT park-and-ride lots that were designated free service lots.

Total cost to provide free bus service for 26 weeks was approximately $148,000. FHWA provided 90 percent of the cost for the first 4 weeks and the state matched the 10 percent remainder. For the remaining 22 weeks of service, the costs were split 50/50 between FHWA and the state, costing approximately $62,000 each.

RIPTA began the free express-bus service at three of the park-and-ride lots the week of May 15, 1989. During the first week the contractor conducted design improvements to one of the leased parking lots.

Total ridership for the service was 80,610 passenger trips, an average of 651 passenger trips per day. With a total service cost of $148,000, the cost per trip was $1.84.

ENFORCEMENT STRATEGIES

The enforcement program was one of the best "E's" in terms of maintaining capacity and reducing delay. The program generally consisted of two elements: (a) free towing and assistance on I-195 to keep the roadway clear; and (b) strict enforcement of the 30-mph speed limit to meter traffic and allow rapid passage of emergency vehicles.

In preparation for the start of bridge repairs, RIDOT included a stipulation in its contract that required two tow trucks on site to remove breakdowns and vehicles involved in collisions quickly. To respond immediately to accidents or breakdowns, two tow trucks were positioned on each side of the highway on a 24-hr per day basis.

Strict enforcement of the 30-mph speed limit proved to be the most effective TSM strategy. As noted in the Highway Capacity Manual (HCM), maximum capacity can be achieved at 30 mph for design speeds greater than 60 mph. RIDOT was able to take advantage of this speed-flow characteristic because of the short work-zone length. Ticketing took place in a coned-off breakdown lane before the work zone. This resulted in minimum traffic disruption, and also functioned as a "poor man's" ramp-metering by slowing the remaining traffic down to the posted speed.

The need for (and number of) officers required depended on the work operations underway. At the beginning of Phase II, this usually resulted in one additional officer during peak-hour traffic—total of three police cruisers, 24 hr per day. Approximately 16,000 hr were budgeted—592-hr per week—at an estimated cost of $216,720.

EDUCATION STRATEGIES

A Providence-based public relations agency developed a public communication campaign to increase public awareness and to foster the development of travel contingency plans. The Providence River Bridge project represented the first time an outside source had been contracted by RIDOT for communications assistance.

The public relations campaign primarily consisted of a public information program that included project briefings with elected and appointed officials, emergency medical technicians, and other special interest groups. Advertising in newspapers and on radio was also included. The agency also worked with companies in the Providence and East Bay areas to discuss flextime alternatives and to conduct focus groups to assess the impact of the project on the business and employee communities.

The public information effort cost $270,386. The FHWA funded 90 percent of this cost and RIDOT provided the balance.

EMERGENCY ACTION STRATEGIES

An emergency plan strategy was developed by RIDOT consisting of three basic elements that were designed to

- Map an emergency route in case of gridlock,
- Identify variable message–signing sites and standard messages, and
- Monitor rescue vehicle operations.

Since the choice of alternate routes was very limited, RIDOT encouraged off-peak travel on I-195. To minimize rush-hour traffic and ensure travel when three-lane passage was available, RIDOT suggested that I-95 to I-195 be driven during the daylight hours of 9:30 a.m. to 3:30 p.m.

With the work zone in place, the available width of the roadway was reduced from 49 ft to 33 ft. If an accident or deck failure occurred in the work zone, it was possible that all three lanes in one direction could be closed. Therefore, at the suggestion of FHWA, RIDOT planned emergency detour routes for use when the roadway was closed.

TRAFFIC OBSERVATIONS

After 6 months, traffic barriers were removed and I-195 reconstruction was completed. In general, little change was observed in the driving habits of local commuters. Backups during peak hours were kept to ½ mi with a maximum of 1 mi on some occasions. Automobile commuters did change their departure times, but not in significant numbers. Traffic volumes compared before and during the months of construction indicated that approximately 13.5 percent of the traffic shifted during the morning and afternoon peak periods. During construction there was a drop in hourly traffic volumes in the peak travel period of the day because of increased carpooling, flextime, and traffic diversion.

The RIDOT Planning Division monitored traffic at 16 locations before, during, and after construction. The data collected showed no major traffic diversion through any phase of construction. There was some minor diversion over the Point Street and Henderson bridges accounting for a total of approximately 7 percent diversion during the project.

Overall, as a direct result of the bridge deck replacement project, there were no major changes in travel patterns.

Staff from RIDOT's Planning Division conducted visual vehicle occupancy counts on I-195 before, during, and after construction to determine whether commuters altered their traveling behavior. The data for the morning downtown commute showed that the number of high-occupancy vehicles increased during construction. Because the high-occupancy vehicle rate did not decrease in the months following the construction, it appears that several of the commuters continued carpooling after construction.

The average number of persons per vehicle increased slightly during construction. Automobile occupancy rate averages are as follows:

Month	*Average Rate*
May	1.12
June	1.18
August	1.19
November	1.16

SUMMARY AND CONCLUSIONS

The experience gained in mitigating traffic congestion during reconstruction of the I-195 viaduct using various TSM strategies was analyzed as it applied to this high-volume urban interstate facility. This particular work-zone area was less than ½ mi in length. Similar, longer projects, up to several miles, may not work out in the same manner.

Based on the measured traffic volumes, occupancy counts, and transit usage counts, it appears that there was very little change in commuter driving habits. This does not imply that commuters should not be given opportunities to change their commuting habits through various transportation alternatives. What it does mean is that the alternative transportation opportunities offered should be proportional to the cost of providing the necessary TSM strategies. If TSM planning is sufficiently flexible, then future needs can be addressed as demand increases. Of the $5.9 million project cost, $635,000 (11 percent) was spent on TSM activities, not including the cost of constructing the park-and-ride facilities that were retained permanently.

Each reconstruction project has its problems and institutional framework for mitigating impacts. Project planners and engineers must ensure that the selected strategies are compatible and that they do not compete with each other. The lessons learned on this project were as follows:

- TSM strategies to reduce construction impacts can be very costly and produce only minor benefits.
- Despite its minimal effect on traffic reduction, carpooling was a low-cost traffic management strategy for RIDOT to undertake.
- Strict enforcement of a reduced speed limit (by issuing citations) can have a dramatic impact on capacity of short-length work zones. For this project, the enforcement program was the most cost-effective strategy for maintaining capacity and reducing delay.

REFERENCE

1. C. Dudek and S. Richards. *Traffic Capacity Through Work Zones on Urban Freeways*. Report FHWA/TX-83/20-292-1. Texas Transportation Institute, Texas A&M University, College Station, Tex. 1982.

Publication of this paper sponsored by Committee on Transportation System Management.

Prioritizing Signalized Intersection Operational Deficiencies

JAMES M. WITKOWSKI

A two-level screening process is described for evaluating short- to medium-term improvements for signalized intersections, and a procedure is developed for evaluating and ranking intersection operational deficiencies. A deficiency index (DI) is developed using a linear utility function. A detailed description is provided of the criteria evaluation and selection process used to screen 21 candidate criteria and to select the final formulation of the deficiency index. Also provided is a description of the analysis procedure used to determine the final weights applied to the factors in the DI operation. The use of the DI is demonstrated through the rating of operational deficiencies for all 286 signalized intersections in Tucson, Arizona. The DI is used to identify the 30 intersections most in need of operational improvements.

An essential element of transportation planning and traffic engineering is knowledge of the existing conditions of the roadway system. This knowledge supplies the basis for decisions regarding highway system improvement, improvement priorities, and the staging of improvement implementation. The knowledge of the existing capacity and level of service (LOS) of the elements of the roadway system also supplies a basis for measuring the impact of land development and community growth. The capability of an element of the roadway system (for example, an intersection or highway segment) to accommodate an increase in demand resulting from nearby land development can only be accurately assessed with a clear understanding of the current vehicle demand and roadway capacity. The proper assessment of highway improvement needs requires the knowledge of current and anticipated deficiencies. This knowledge is particularly important with respect to signalized intersections, which typically establish urban arterial system capacity and operating conditions.

Local jurisdictions typically maintain a process by which highway system improvement needs are identified, prioritized, and included in an annual capital improvement program (CIP). The impetus for this study was a concern of the city of Tucson (COT), Arizona, that its existing process for identifying, evaluating, and prioritizing arterial improvements (particularly at signalized intersections) did not provide sufficient information for rational technical decisions regarding improvement needs and priorities. This was of particular concern in light of the city's inability to fund large-scale, long-term transportation improvements because of lack of funding. Also of concern were recently adopted policies limiting roadway widening and establishing a LOS D threshold for the initiation of a planning study for urban arterials. Therefore, the city has established a position of attempting to maximize short-term congestion relief with the available funds, at the same time attempting to identify arterial corridors exceeding the LOS threshold, permitting study for long-term improvement implementation.

JHK & Associates, 110 South Church Avenue, Suite 470, Tucson, Ariz. 85701.

The primary goal of this project was to provide COT with a comprehensive information data base and evaluation procedure in order to assess the existing operating conditions of the city's signalized intersections and to evaluate existing intersection improvement needs and priorities. This study focused on individual congestion hot spots and low-cost improvement alternatives for providing short- to medium-term relief. The procedures developed and presented were not intended to replace the long-range comprehensive planning process or the implementation of long-term transportation improvements. Instead, these procedures were intended to supplement long-range planning and to provide direction in the selection of shorter-term improvements in lieu of factors that prevent the immediate implementation of a long-range system plan.

The goal was reached in part through the satisfaction of the following objectives:

1. Provide an accurate and quantified assessment of the current operating status of the city's signalized intersections.
2. Develop a rating system for prioritizing intersection improvements on the basis of criteria that reflect the existing improvement needs, and establish this rating system in a microcomputer-based software program.
3. Establish a data base management system to enable the city to maintain an up-to-date assessment of intersection improvement priorities using the developed software.
4. Develop alternative concept designs to alleviate problem conditions for the worst 30 intersections identified and prioritize these improvements based on cost-effectiveness.

A detailed description of the elements of this entire study is provided in the final report (*1*). The following discussion details the procedures developed to identify and prioritize signalized intersection operational improvement needs. This procedure was developed to establish a short list of 30 intersections most in need of operational improvements. This procedure was intended to provide a focus for the analysis of improvement alternatives at these 30 locations. The discussion of the improvement alternatives analysis and cost-effectiveness ranking of improvements is described in a companion paper by Witkowski in this Record.

EXISTING COT PROCEDURES

A summary of the previous COT process for identifying intersection and roadway segment improvement needs is shown in Figure 1. The initial screening of signalized intersections was based on intersection accident history using intersection accident rate stratified by the functional classification of the intersecting roadways as the evaluation criterion.

Information on intersection geometric or operational conditions was not explicitly included in the previous analysis procedure. These intersection characteristics were evaluated after the intersections of concern had been identified on the basis of accident history. Therefore, intersections with operational deficiencies (i.e., long delays or poor level of service) were not identified as intersections of concern unless they had a high accident rate. Poor intersection operating conditions may not necessarily result in a high intersection accident rate, and intersection safety problems may not necessarily be alleviated through improvements in intersection operations. Therefore, the previous evaluation process failed to provide information vital to the assessment of intersection operational improvement needs.

The city's previous procedure for determining roadway segment improvement needs, including major widening, was based on a sufficiency rating analysis. This sufficiency rating is based on the physical condition of the street, considering pavement structural condition, maintenance needs, traffic congestion (present and forecast), and accident history. A point system is used to quantify each criterion, and the points are combined to provide an overall assessment of roadway condition. Sufficiency rating systems of this type have been commonly used by state and local transportation agencies for many years (*2*).

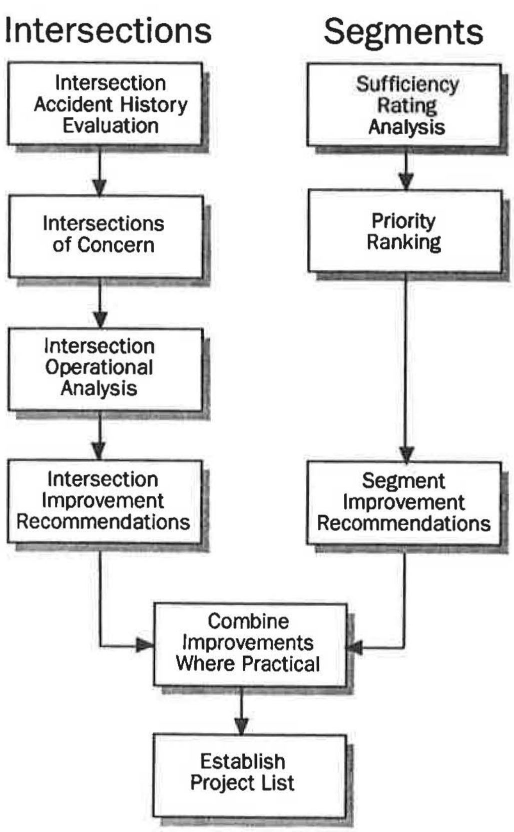

FIGURE 1 Overview of current COT intersection and roadway segment evaluation process.

OPERATIONAL VERSUS SAFETY IMPROVEMENTS

There appears to be a clear dichotomy in the evaluation of intersection improvement needs. This dichotomy arises from the need to identify both safety and operational deficiencies and the improvements that are specifically designed to address each type of problem.

The need for this dichotomy also becomes apparent with the consideration of the potential liability created by evaluating safety and operational deficiencies together and evaluating improvement needs on the basis of a combined deficiency index (DI). There is a potential for a needed safety improvement not being identified because it is somehow overshadowed by intersections with high operational deficiencies. The analysis procedure must be capable of identifying both safety and operational improvement needs separately. Therefore, a procedure was developed to identify intersection operational and safety deficiencies separately and to combine improvement recommendations when an analysis indicates that the combination is practical.

The recommended analysis procedure for establishing intersection improvement priorities is shown in Figure 2. The prioritization procedure is a two-screen process. In the first screen, all intersections under analysis are evaluated separately for both safety and operational deficiencies using selected evaluation criteria and given a separate rating for both safety and operational improvements. The deficiency rating is an indication of the overall need for improvement at each location.

After the first screen, intersections with the highest deficiency rating are selected for a more detailed assessment of problems and potential solutions. Requirements for safety and operational improvements are compared in order to determine where improvements should be combined because they address related problems. In addition, intersection safety and operational improvement needs are compared and coordinated with other system improvements. The comparison with other system improvements identifies where intersection improvements can be combined with planned major facility upgrades. This comparison also provides for an evaluation of the continuity of improvements in a systemwide context.

The second screen in the analysis procedure is an evaluation of cost-effectiveness. The cost-effectiveness analysis is used to establish the final improvement priorities for operational and safety improvements.

The cost-effectiveness evaluation of safety and operational improvements should be performed separately when the improvement requirements cannot be combined at a given location. The rationale for this is that operational improvements typically generate much higher cost-effectiveness values than safety improvements. Therefore, it would be difficult for purely

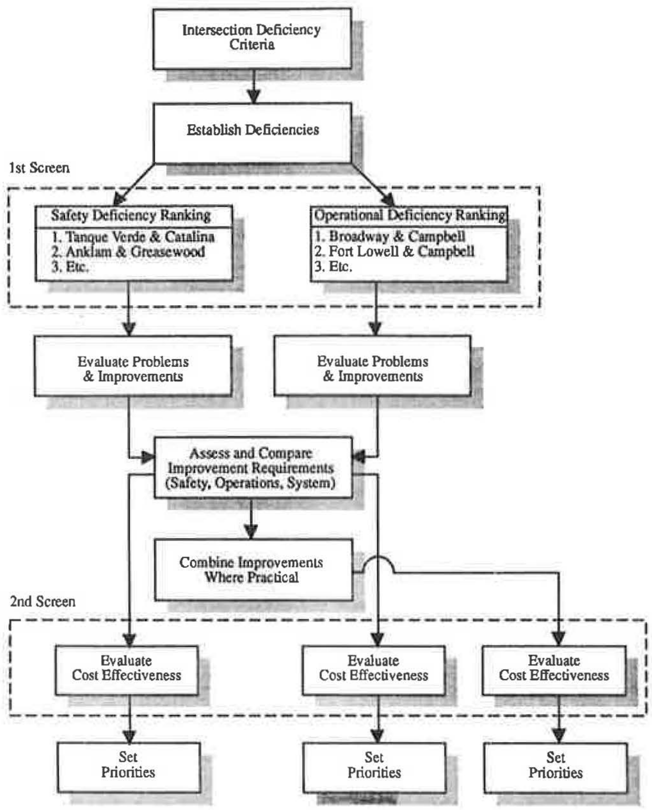

FIGURE 2 Recommended intersection analysis procedure.

safety-related improvements to compete for improvement funds. However, making no safety improvements at locations with identified problems places municipalities in a poor position relative to potential liability in accident cases that occur at these locations. Potential liability is not a parameter that has been included traditionally in the cost-effectiveness of safety improvements, but it must be considered as an important element in the justification of the dichotomy of safety and operational improvement categories.

OPERATIONAL DEFICIENCY CRITERIA

The criteria used in the analysis of operational deficiencies are of primary importance in the successful identification of improvement needs. The criteria must possess several important characteristics. For the purposes of this study, these characteristics were defined as follows:

- Technical reliability—The level of each criterion must vary with the operational condition of the intersection. The level of the criteria must be obtained with sufficient measurement-estimation accuracy to provide a useful and reliable evaluation tool.
- Importance—The criteria must convey a measure of importance in the evaluation of improvement needs. It must be related meaningfully to the operational condition of the intersection.
- Availability—The measure or estimate of each criterion should be available and updated periodically without unreasonable expense or level of effort.
- Independence—The measure of each criterion should be unique in terms of the operational condition it represents relative to the operational condition represented by other criteria. This avoids double-counting or the weighting of a particular operational condition too heavily.

The selection of evaluation criteria for operational improvements was focused on five major categories:

- Traffic volume,
- Present peak-hour traffic conditions,
- Safety,
- Air quality, and
- Transit operations.

Conformance with design standards was considered an additional category but was eliminated early in the review process. This category was eliminated because it was considered a primary factor in the evaluation of improvement alternatives for both operational and safety problems after the problem had been identified through other criteria.

Twenty-one individual evaluation criteria from these five categories were evaluated for inclusion in the procedure for establishing intersection operational deficiencies. These criteria are presented in the first column of Table 1.

Each of the criteria discussed was evaluated on the four characteristics of technical reliability, importance, availability, and independence. A subjective evaluation was performed and each criterion was rated on a scale of 1 to 5 for each of the first three characteristics. A value of 1 was considered the lowest and 5 the highest in each of the criteria categories. Criteria were only rated against other criteria in the same category. This rating is also presented in Table 1.

The independence characteristic was evaluated by noting whether a criterion was related to any of the other criteria and a notation was made as to what measure the criterion represented. This information is also presented in Table 1.

The results of the criteria assessment are presented in the last column of Table 1. The criteria screening process led to the initial indication of those criteria that would be suited for inclusion in an intersection operational deficiency rating model. This assessment also indicated criteria that might not be suited for use together in the same deficiency rating formulation.

Two sets of preliminary operational DI criteria resulted from the criteria assessment. The criteria included in each of the sets are presented in Table 2. These criteria were subjected to further, more detailed evaluation. The final recommended criteria were selected based on a numerical assessment of each criterion measure for each intersection and an analysis of the impact of each criterion on the intersection deficiency rating.

CRITERIA VALUE ESTIMATION

The criteria in Table 2 were estimated for each of the 286 signalized intersections within the jurisdiction of COT. Afternoon peak-hour turning movement counts were taken for each intersection during the peak travel months (September through April) of 1989 and 1990. Using the operational analysis procedures for signalized intersections contained in the 1985

TABLE 1 Operational Analysis Evaluation Criteria

Evaluation Criteria	Formulation	Technical Reliability	Importance	Availability	Independence	Measure of	Preliminary Criteria Set(2)
Traffic Volume							
1. Present Average Daily Traffic (PADT)	ADT Entering	5	3	4	Related to 2., 7.	Total demand	
2. Present Peak-Hour Volume (PHV)	PHV Entering	5	5	4	Related to 1., 7.	Peak-hour demand	A
3. Forecast Average Daily Traffic (FADT)	FADT Entering	3	3	3		Future total demand	
4. Forecast Peak-Hour Volume (FPHV)	FPHV Entering	3	3	3	Factor of 3.	Future peak-hour demand	
Present Peak-Hour Traffic Operations							
5. Intersection Level of Service (LOS)	Capacity Analysis	3	4	3	Function of 2.,6.,7.	Intersection operations	
6. Intersection Critical Volume to Capacity Ratio(Xc)	Capacity Analysis	4	4	3	Function of 2.	Capacity utilization	A, B
7. Intersection Stopped Delay per Vehicle (SD)	Capacity Analysis	4	4	3	Function of 6.,2.	Intersection operations	A
8. Intersection Total Stopped Delay (TSD)	7. X 2.	4	4	3	Function of 6.,2.,7.	Operations and demand	B
9. Time Duration of LOS (TLOS)	Capacity Analysis	3	4	1	Function of 1.,2.	Operations and demand	
10. Volume per Through Lane	2./Number of Lanes	3	3	4	Function of 2.	Capacity Utilization	
Safety							
11. Total Accidents (last three years (1))	Number of Accidents	5	3	4	Related to 1.	Total Accidents	
12. Accident Rate (last three years (1))	10./1.	5	5	4	Function of 11.,1.	Accidents/unit demand	A,B,S
13. Pedestrian Accidents (last three years (1))	Number of Ped Accidents	5	4	4	Related to 1.,2.	Pedestrian safety	S
14. Bicycle Accidents (last three years (1))	Number of Bk Accidents	5	4	4	Related to 1.,2.	Bicycle safety	S
15. Accident Severity	Severity Index	4	2	4		Seriousness of Accidents	S
Air Quality							
16. Pollution Added	Grams per min. X (8.)	3	5	3	Function of 8.	Total pollution added	
17. Peak-Hour Stopped Delay per Vehicle (SD)	(7.)	4	2	3	Same as 7.	Pollution added/vehicle	
18. Peak-Hour Total Stopped Delay (TSD)	(8.)	4	5	4	Same as 8.	Total pollution added	A
Transit Operations							
19. Peak-Hour Stopped Delay per Transit Vehicle(SD)	Same as (7.)	4	3	3	Same as 7.	Transit delay per vehicle	
20. Peak-Hour Total Transit Delay (TTSD)	(7.) X # of Transit Vehicles	4	4	3	Function of 7.	Total transit vehicle delay	
21. Peak-Hour Total Transit Person Delay (TTPSD)	(20.)X Transit Load Factor	4	5	3	Function of 7.	Total transit pass. delay	A,B

1. Three years of accident data should be used to establish the accident rate unless a major intersection reconstruction has occurred during that period. Under that condition, the accident history since the reconstruction should be used, and noted as a period of less than three years.
2. A = Set A B = Set B. S = Suitable for safety deficiency index.

TABLE 2 Preliminary Operational Deficiency Rating Criteria

Evaluation Criteria Category	Set A Criteria	Set B Criteria
Traffic Volume	Present Peak-Hour Volume	
Present Traffic Operations	Critical Volume to Capacity Ratio Average Stopped Delay per Vehicle	Critical Volume to Capacity Ratio Peak-Hour Total Stopped Delay
Safety	Accident Rate	Accident Rate
Air Quality	Peak-Hour Total Stopped Delay	
Transit Operations	Peak-Hour Total Transit Person Delay	Peak-Hour Total Transit Person Delay

Highway Capacity Manual (HCM) (*3*) additional data were collected to provide the information necessary to assess the existing level of service for each intersection. The intersection stopped-delay estimates and critical volume-to-capacity ratios calculated from the HCM procedure were used as the measures of these criteria. Peak-hour total stopped delay was estimated in vehicle-hours from the average stopped delay per vehicle and the total peak-hour volume entering such intersection. Peak-hour total transit person delay was estimated in person-hours from data representing the average number of transit passengers entering each intersection during the peak-hour (provided from COT transit system records) and average stopped delay per vehicle (estimated from the HCM analysis). The constant value of 1.3 to convert stopped-delay estimates to total-delay estimates was not applied in this analysis because, as a constant multiplier, it would have no effect on the relative values of the delay estimates between intersections.

The most current 3 years of complete accident data (1986 through 1988) along with estimates of the average daily traffic entering each intersection were provided by the city. These data were used to calculate the accident rate for each intersection in accidents per million vehicles entering the intersection.

DEFICIENCY INDEX FORMULATION

A key element in the development of a prioritization process is the methodology used to combine the various criteria into a single index of operational deficiency. The purpose of the index is to identify locations that are most in need of operational improvements. The index must be technically sound, easily understood, and easily implemented, and it must generate results that can be logically supported.

The method selected for development of operational and safety deficiency indexes was a linear utility function. Linear utility functions combine weighted measures of the evaluation criteria into a single index, which is the basis for identifying improvement needs. This is the same type of procedure that is used in sufficiency rating schemes. The DI is described in Equation 1.

$$DI = W_1X_1 + W_2X_2 + \ldots + W_nX_n \quad (1)$$

where X_i is the normalized value of criterion i and W_i is the weight applied to criterion i.

Criteria normalization precludes any single criterion from dominating the DI because of its sheer magnitude relative to the other criteria values. It also allows the criteria weights to be truer reflections of the overall importance of each criterion in determining intersection deficiencies.

There are two basic ways in which normalization can occur. The first is to normalize the value of each criterion based on its largest value for a given set of intersections. With this scheme the criteria in the DI formulation would be expressed on a zero-to-one scale, and the criteria weights would be an expression of the relative importance of each criterion in the ranking formula. The disadvantage of this procedure is that, because of possible future changes in the base for normalization, there is no way of tracking the change in intersection deficiency over time. Also, there is no way of using the DI to determine whether or not an intersection exceeds some threshold condition requiring improvement.

The second means of normalizing the criteria is to use a preselected threshold value as the base for normalization. The threshold value would be used year after year and would allow changes in intersection deficiency to be traced over time on the basis of the DI. This would also supply a basis for assessing the impact of improvements using the DI as a measure of effectiveness. The disadvantage of this procedure is that the range of normalized criteria values is not controlled as well as the previous procedure. The zero-to-one range for the criteria values cannot be maintained unless the threshold value is selected so that it cannot be exceeded. Threshold values must be set so that the integrity of the relative magnitude of the normalized criteria is maintained. The zero-to-one range for the normalized values of the criteria is not necessary as long as the relative magnitude of each criterion is maintained at a reasonable level in comparison to the weights used to value each criterion in the DI.

The threshold values for normalizing the criteria can be established in at least two ways. One possibility would be to determine the desirable maximum level, or standard, for each of the criteria keeping in mind that the range of values for the normalized values for each criterion should be approximately the same. In this way, normalized criteria values that exceed a value of one would be indicative of a condition that exceeds the desirable maximum.

Another method of establishing the threshold values would be to use the maximum values from the present condition as the threshold. This would provide a direct comparison of each succeeding year to the worst conditions that presently exist. The normalization of criteria for this study was based on the maximum value of each criterion for the existing condition.

CRITERIA ANALYSIS

The evaluation of the criteria and the selection of the final parameters to be included in the DI equations were based on an assessment of the relative interdependence of the criteria and the sensitivity of the ranking of the intersections to the criteria. The interdependence of the criteria was judged using linear regression analysis techniques. The impact of the criteria on the ranking of the intersections was based on a sensitivity analysis.

Regression Analysis

A matrix of the simple linear coefficients of determination values (r^2) was developed using linear regression analysis procedures applied to both the actual and normalized criteria values. There was virtually no difference between the coefficients generated using the actual and normalized values. For brevity only the results using the normalized values are presented here.

The linear regression analysis results are presented in Table 3. The mean and standard deviation of each of the variables are presented along with the r^2 values.

In general, variables that are highly correlated should not be used together in the relationship for the DI because they represent a redundant explanatory power and would double count for the same effects. Therefore, stopped delay per vehicle and total stopped delay were judged to be too highly correlated to appear in the same DI formulation, as were total stopped delay and peak-hour volume. Further refinement of the DI criteria was based on a sensitivity analysis.

Sensitivity Analysis

A sensitivity analysis was performed to determine the sensitivity of the deficiency rating of the intersections to the operations criteria. The base-case rating that included each of the criteria with an equal weight in determining the DI was established. Systematically, one variable at a time was removed from the DI equation, and the intersections were rated with the remaining variables having equal weight. The ratings with the deleted variable were compared to the base case, and the changes in rank of the 30 highest-rated intersections in the base condition were determined. An overall sensitivity index was calculated as the sum of the absolute value of the change in rank for the 30 intersections rated highest in the base condition. The results of the sensitivity analysis are presented in Table 4.

The rating of intersections showed very little sensitivity to accident rate and the critical volume to capacity (v/c) ratio. These variables added little explanatory power to the analysis. In order for accident rate and the critical v/c ratio to affect the results of the DI rating to any significant degree, the weights applied to these parameters in the DI equation would have to far exceed their relative importance as operations analysis parameters. Therefore, these variables were excluded from the DI.

TABLE 3 Normalized Operations Data Statistics

	Coefficient of Determination (r^2)				
Variable	Average Delay per Vehicle	Peak Volume	Accident Rate	Critical v/c	Total Delay
Peak Volume	.350				
Accident Rate	.008	.005			
Critical v/c	.210	.400	.004		
Total Stopped Delay	.852	.600	.005	.279	
Transit Person Delay	.482	.257	.002	.093	.474

Variable	Mean	Standard Deviation
Average Delay per Vehicle	0.2502	0.2029
Peak Volume	0.4649	0.2047
Accident Rate	0.1555	0.0975
Critical v/c	0.2390	0.1016
Total Stopped Delay	0.1790	0.2053
Transit Person Delay	0.1357	0.1764

TABLE 4 Operational Criteria Sensitivity Analysis

		Variable Deleted						Variable Deleted					
Int #	Base	PV	Acc Rate	V/C	Ave Delay	Total Delay	Trans Delay	PV	Acc Rate	V/C	Ave Delay	Total Delay	Trans Delay
		Rank						Change in Ranking					
123	1	1	1	1	1	1	1	0	0	0	0	0	0
234	2	3	2	2	3	4	3	-1	0	0	-1	-2	-1
584	3	5	3	3	2	3	5	-2	0	0	1	0	-2
575	4	2	9	4	5	2	16	2	-5	0	-1	2	-12
596	5	9	5	7	4	10	2	-4	0	-2	1	-5	3
343	6	8	4	6	6	6	8	-2	2	0	0	0	-2
348	7	10	6	8	7	11	7	-3	1	-1	0	-4	0
488	8	4	7	5	15	5	18	4	1	3	-7	3	-10
346	9	11	8	10	11	12	9	-2	1	-1	-2	-3	0
492	10	14	11	11	10	9	12	-4	-1	-1	0	1	-2
489	11	12	10	9	8	7	17	-1	1	2	3	4	-6
101	12	7	13	12	18	15	6	5	-1	0	-6	-3	6
681	13	15	12	13	14	19	4	-2	1	0	-1	-6	9
230	14	13	14	15	13	14	11	1	0	-1	1	0	3
483	15	19	18	16	9	8	20	-4	-3	-1	6	7	-5
582	16	16	15	17	20	18	19	0	1	-1	-4	-2	-3
401	17	18	16	19	16	17	23	-1	1	-2	1	0	-6
262	18	26	17	20	17	21	10	-8	1	-2	1	-3	8
341	19	21	20	18	12	16	22	-2	-1	1	7	3	-3
709	20	6	19	14	31	13	34	14	1	6	-11	7	-14
219	21	20	21	21	22	20	13	1	0	0	-1	1	8
47	22	17	22	22	32	24	21	5	0	0	-10	-2	1
600	23	31	23	23	19	23	15	-8	0	0	4	0	8
603	24	22	28	24	36	31	14	2	-4	0	-12	-7	10
504	25	32	24	25	24	26	25	-7	1	0	1	-1	0
416	26	24	25	26	28	27	26	2	1	0	-2	-1	0
223	27	28	27	29	21	22	33	-1	0	-2	6	5	-6
496	28	35	29	28	23	25	30	-7	-1	0	5	3	-2
263	29	25	26	27	38	34	24	4	3	2	-9	-5	5
408	30	27	30	30	30	29	36	3	0	0	0	1	-6
							Total Absolute Change	102	32	28	104	81	141

The most significant variables in the DI were peak-hour volume, average stopped delay, total stopped delay, and transit person delay. Because of the high correlation between total delay, average delay, and the peak-hour volume, it was recommended that all three of these variables not be contained in the same DI relationship. Two relationships were subject to further testing in order to determine the criteria weights for the DI. These relationships were (a) an equation containing total delay and transit person delay and (b) an equation containing peak-hour volume, average delay, and transit person delay. A summary of the recommendations for the criteria to be used in the DI is provided in Table 5.

Weight Analysis

The analysis of the weights to be used in the DI equation proceeded in a manner similar to that employed for the sen-

TABLE 5 Operational Criteria Recommendations

Included Criteria

- Total Stopped Delay and Transit Person Stopped Delay
  or
- Peak-Hour Volume, Average Stopped Delay, and Transit Person Stopped Delay

Rationale

- Rating is sensitive to these parameters
- Average delay and peak volume are logical operations measures
- For intersections with same average delay, higher volume should be ranked higher
- For intersections with same volume, higher delay should be rated higher
- Total delay combines peak volume and average delay in appropriate manner
- Average delay and total delay are highly correlated
- Transit delay adds a significant rating parameter

Excluded Criteria

- Accident Rate
- Critical v/c

Rationale

- Accident rate adds no explanatory power to operations analysis -- not correlated to operations parameters
- Accident rate included in separate safety analysis
- Rating is insensitive to accident rate
- Rating is insensitive to critical v/c

sitivity analysis. The change in intersection rankings as observed in relation to a base condition for various weights applied to the criteria in the equation. The sensitivity of the rankings to the change in the criteria was used to focus the recommendations for the final criteria.

The evaluation of the operations criteria required three separate analyses. Analysis A evaluated the criteria weight for an index comprised of total delay and transit person delay using ratings based solely on total delay as the base condition. Analysis A indicated that the intersection rankings had a low sensitivity to the inclusion of transit person delay in the relationship with a weight of 10 percent or less. These results are presented in Table 6. The rankings were moderately sensitive to transit person delay with a weight of 20 percent and exhibited a high level of sensitivity to transit person delay with a weight of 30 percent. An additional test was performed using a 15 percent weight on transit person delay and resulted in a moderate level of sensitivity in the rankings that was less than that using the 20 percent weight.

Analysis B compared the rankings using peak-hour volume and average delay as the criteria in the index with a base condition using only total delay recall. The total delay is the product of peak-hour volume and average delay. Therefore, the equation with only total delay contains both the peak-hour volume and average delay in a different form. The purpose of this analysis was to evaluate which equation provided a better overall index to be used in the ranking process. The results using various weighting schemes for peak-hour volume and average delay are presented in Table 7.

The results of Analysis B indicated that using peak-hour volume and average delay, each at a 50 percent weight, produced results very similar to those generated using only total delay in the equation. The ranking of the first 13 intersections remained unchanged, with only minor changes for the remaining intersections. Deviations from the 50 percent weights used for peak-hour volume and average delay resulted in increased change in the rankings compared to the base condition. There was no clear rationale for weighting the peak-

TABLE 6 Operational Criteria Weight Analysis A

	Test						Test				
Intersection	1A	2A	3A	4A	5A	6A	2A	3A	4A	5A	6A
		Rank					Change In Rank				
123	1	1	1	1	1	1	0	0	0	0	0
234	2	2	2	2	2	2	0	0	0	0	0
596	3	3	3	4	7	3	0	0	-1	-4	0
584	4	4	4	3	3	4	0	0	1	1	0
348	5	5	5	6	6	5	0	0	-1	-1	0
681	6	7	7	9	12	9	-1	-1	-3	-6	-3
343	7	6	6	5	5	6	1	1	2	2	1
346	8	8	8	8	8	7	0	0	0	0	1
101	9	9	10	10	13	10	0	-1	-1	-4	-1
262	10	11	12	15	18	14	-1	-2	-5	-8	-4
488	11	10	9	7	4	8	1	2	4	7	3
492	12	12	11	12	11	11	0	1	0	1	1
230	13	13	14	14	14	13	0	-1	-1	-1	0
603	14	16	17	20	25	18	-2	-3	-6	-11	-4
219	15	15	15	17	19	17	0	0	-2	-4	-2
489	16	14	13	11	10	12	2	3	5	6	4
263	17	18	19	21	23	21	-1	-2	-4	-6	-4
582	18	17	16	16	15	16	1	2	2	3	2
47	19	19	21	19	21	20	0	-2	0	-2	-1
600	20	22	22	26	27	24	-2	-2	-6	-7	-4
401	21	21	20	18	16	19	0	1	3	5	2
210	22	24	26	28	33	28	-2	-4	-6	-11	-6
416	23	23	23	24	24	23	0	0	-1	-1	0
575	24	20	18	13	9	15	4	6	11	15	9
504	25	26	25	27	26	25	-1	0	-2	-1	0
341	26	25	24	22	20	22	1	2	4	6	4
350	27	27	29	30	35	29	0	-2	-3	-8	-2
99	28	29	31	34	36	32	-1	-3	-6	-8	-4
579	29	30	30	33	34	31	-1	-1	-4	-5	-2
483	30	28	27	25	22	27	2	3	5	8	3
						Total Absolute Change	24	45	89	142	67

Weight Factors Used:
Test:
1A --> Total Delay = 1.000
2A --> Total Delay = 0.950, Trans. Person Delay = 0.050
3A --> Total Delay = 0.900, Trans. Person Delay = 0.100
4A --> Total Delay = 0.800, Trans. Person Delay = 0.200
5A --> Total Delay = 0.700, Trans. Person Delay = 0.300
6A --> Total Delay = 0.850, Trans. Person Delay = 0.150

TABLE 7 Operational Criteria Weight Analysis B

	Test							Test				
Intersection	1B	2B	3B	4B	5B	6B	7B	2B	3B	4B	5B	6B
		Rank						Change In Rank				
123	1	1	1	2	3	1	1	0	0	-1	-2	0
234	2	2	2	1	2	2	2	0	0	1	0	0
596	3	3	3	3	1	3	3	0	0	0	2	0
584	4	4	4	4	4	4	4	0	0	0	0	0
348	5	5	5	5	5	5	5	0	0	0	0	0
681	6	6	6	6	8	6	6	0	0	0	-2	0
343	7	7	7	7	7	7	7	0	0	0	0	0
346	8	8	8	9	9	8	9	0	0	-1	-1	-1
101	9	9	10	10	12	10	8	0	-1	-1	-3	1
262	10	10	9	8	6	9	11	0	1	2	4	-1
488	11	11	12	13	15	11	10	0	-1	-2	-4	1
492	12	12	11	11	11	12	12	0	1	1	1	0
230	13	13	15	15	14	14	13	0	-2	-2	-1	0
603	14	16	17	17	21	17	14	-2	-3	-3	-7	0
219	15	17	16	16	17	16	16	-2	-1	-1	-2	-1
489	16	15	14	14	13	15	15	1	2	2	3	1
263	17	18	19	19	26	19	19	-1	-2	-2	-9	-2
582	18	19	20	20	25	20	20	-1	-2	-2	-7	-2
47	19	20	21	26	27	21	18	-1	-2	-7	-8	1
600	20	14	13	12	10	13	17	6	7	8	10	3
401	21	22	22	21	22	22	25	-1	-1	0	-1	-4
210	22	23	28	28	31	24	21	-1	-6	-6	-9	1
416	23	28	30	29	28	29	26	-5	-7	-6	-5	-3
575	24	24	29	30	34	27	23	0	-5	-6	-10	1
504	25	21	18	18	16	18	24	4	7	7	9	1
341	26	26	24	24	23	23	27	0	2	2	3	-1
350	27	30	27	27	24	26	29	-3	0	0	3	-2
99	28	33	34	35	35	33	34	-5	-6	-7	-7	-6
579	29	32	33	32	32	32	33	-3	-4	-3	-3	-4
483	30	29	25	23	18	25	30	1	5	7	12	0

Weight Factors Used:
Test:
1B --> Total Delay = 1.000
2B --> Peak Vol. = 0.500, Ave. Delay = 0.500
3B --> Peak Vol. = 0.530, Ave. Delay = 0.470
4B --> Peak Vol. = 0.550, Ave. Delay = 0.450
5B --> Peak Vol. = 0.600, Ave. Delay = 0.400
6B --> Peak Vol. = 0.470, Ave. Delay = 0.530

hour volume or the average delay more or less than the other. Therefore, there appears to be no advantage to using peak-hour volume and average delay over an equation containing only total delay. Total stopped delay also provides the advantage that it can be used as a surrogate for, or directly in computations of, vehicle emission levels. It also provides a good effectiveness measure for use in the economic analysis of improvement alternatives.

Analysis C evaluated the inclusion of transit person delay in the equation with peak-hour volume and average delay. The equation with peak-hour volume and average delay weighted equally at 50 percent was used as the base condition. The results are presented in Table 8. The results indicate that the rankings are considerably more sensitive to the inclusion of transit person delay in this relationship than in the relationship with total delay. In each case where total volume and average delay were weighted equally, the inclusion of transit person delay had a much greater impact on the rankings at a given weight than it did in the relationship with total delay at the same weight (Analysis A). Because of the large shifts in the rankings when transit person delay was included in the equation, Analysis C rankings were considered overly sensitive to transit delay.

TABLE 8 Operational Criteria Weight Analysis C

	Test								Test						
Intersection	1C	2C	3C	4C	5C	6C	7C	8C	2C	3C	4C	5C	6C	7C	8C
		Rank							Change In Rank						
123	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
234	2	2	2	2	3	2	2	2	0	0	0	-1	0	0	0
596	3	7	7	7	6	7	3	4	-4	-4	-4	-3	-4	0	-1
584	4	3	3	3	2	3	4	3	1	1	1	2	1	0	1
348	5	6	6	6	5	6	5	5	-1	-1	-1	0	-1	0	0
681	6	14	13	12	14	12	8	9	-8	-7	-6	-8	-6	-2	-3
343	7	5	5	4	4	5	6	6	2	2	3	3	2	1	1
346	8	8	8	8	8	8	7	8	0	0	0	0	0	1	0
101	9	13	14	16	20	14	10	10	-4	-5	-7	-11	-5	-1	-1
262	10	17	15	13	12	15	11	13	-7	-5	-3	-2	-5	-1	-3
488	11	4	4	5	7	4	9	7	7	7	6	4	7	2	4
492	12	11	11	11	10	11	12	12	1	1	1	2	1	0	0
230	13	15	16	17	16	16	14	15	-2	-3	-4	-3	-3	-1	-2
600	14	23	21	21	18	22	15	18	-9	-7	-7	-4	-8	-1	-4
489	15	10	9	9	9	9	13	11	5	6	6	6	6	2	4
603	16	27	29	29	31	29	17	21	-11	-13	-13	-15	-13	-1	-5
219	17	21	22	22	24	21	16	19	-4	-5	-5	-7	-4	1	-2
263	18	26	27	27	29	26	22	26	-8	-9	-9	-11	-8	-4	-8
582	19	16	18	19	19	17	18	17	3	1	0	0	2	1	2
47	20	22	24	25	26	23	20	22	-2	-4	-5	-6	-3	0	-2
504	21	24	23	23	23	24	24	25	-3	-2	-2	-2	-3	-3	-4
401	22	18	19	18	15	18	23	20	4	3	4	7	4	-1	2
210	23	34	36	38	41	36	27	29	-11	-13	-15	-18	-13	-4	-6
575	24	9	10	10	11	10	19	14	15	14	14	13	14	5	10
709	25	12	12	14	21	13	21	16	13	13	11	4	12.	4	9
341	26	20	20	20	17	20	25	24	6	6	6	9	6	1	2
335	27	38	38	36	34	40	31	32	-11	-11	-9	-7	-13	-4	-5
416	28	29	28	28	28	28	28	28	-1	0	0	0	0	0	0
483	29	19	17	15	13	19	26	23	10	12	14	16	10	3	6
350	30	31	30	30	27	30	30	30	-1	0	0	3	0	0	0
								Total Absolute Change	154	155	156	167	154	44	87

Weight Factors Used:
Test:
1C --> Peak Vol. = 0.500, Ave. Delay = 0.500
2C --> Peak Vol. = 0.400, Ave. Delay = 0.400, Trans. Person Delay = 0.200
3C --> Peak Vol. = 0.424, Ave. Delay = 0.376, Trans. Person Delay = 0.200
4C --> Peak Vol. = 0.440, Ave. Delay = 0.360, Trans. Person Delay = 0.200
5C --> Peak Vol. = 0.480, Ave. Delay = 0.320, Trans. Person Delay = 0.200
6C --> Peak Vol. = 0.400, Ave. Delay = 0.360, Trans. Person Delay = 0.200
Total Delay = 0.040
7C --> Peak Vol. = 0.475, Ave. Delay = 0.475, Trans. Person Delay = 0.050
8C --> Peak Vol. = 0.450, Ave. Delay = 0.450, Trans. Person Delay = 0.100

The recommended relationship for use in the rating of intersections based on operations parameters was to use total delay in combination with transit person delay. The evaluation of the assignment of weights to total delay and transit delay suggested that the use of a weight of 10 percent or less for transit person delay provided rankings that were basically insensitive to the inclusion of transit delay. A 30 percent weight on transit person delay affected the ranking results more than was deemed appropriate. The 15 and 20 percent weights on transit delay provided reasonable impacts on the ranking of the intersections. After review by city staff, the 15 percent weight for transit delay was selected in combination with an 85 percent weight on total delay. The results of the operational DI analysis are presented in Table 9 for those 30 intersections considered most operationally deficient. Note that these results differ slightly from the results generated during the evaluation of the criteria and weighting factors as a result of the final review and update of the data used in the analysis.

RESULTS AND CONCLUSIONS

The procedures developed throughout this study provide a useful element in a comprehensive congestion-management program. The identification of existing intersection operational deficiencies is a key element in establishing an effective program to reduce urban congestion, improve automobile and transit travel time, reduce vehicle emissions, and improve air quality. These procedures are intended to supplement the long-range regional transportation planning process and to provide assistance in the selection of short- to medium-term congestion relief measures by identifying those signalized intersections most in need of operational improvements. This will allow local transportation agencies to focus their manpower and financial resources on problems that will benefit the most from improvement.

It should be emphasized that the identification of hazardous intersections is an important element in the overall assessment of improvement needs. The safety analysis should be conducted separately to ensure that intersections of safety concern are properly identified and not overshadowed by the operational deficiencies. This is particularly important because the intersection accident rate was shown to be unrelated to the estimated congestion levels. Therefore, it cannot be assumed that identifying operational deficiencies will concurrently identify safety deficiencies.

The analysis procedures used to screen the criteria for inclusion in the DI provided for the rational selection of the final criteria used in the index formulation. The analysis procedures used to evaluate the weighting factors applied to each criterion provided a logical quantitative assessment.

The application of the 1985 HCM procedures for the analysis of signalized intersections was extremely valuable in the assessment of the existing operating conditions of the COT arterial system. This analysis provided the basis for the establishment of the DI and the development of the city's CIP. Future applications of the deficiency analysis will require that the capacity analysis be updated on a periodic basis with a reasonable level of effort. This can be accomplished by monitoring traffic growth trends and establishing a program to update turning movement counts as dictated by traffic growth.

TABLE 9 Thirty Intersections with the Highest Operational DI

ID #	East/West Street	North/South Street	DI	Peak Vol. (Veh/Hr)	Ave. Delay (Sec/Veh)	Total Delay (Veh-Hrs)	Transit Person Delay (Person-Hrs)
483	BROADWAY BLV	CAMPBELL AV	0.947	6228	75.4	130.442	8.755
123	FORT LOWELL RD	CAMPBELL AV	0.929	5680	88.1	139.002	4.625
234	GRANT RD	CRAYCROFT RD	0.893	6415	77.5	138.101	2.799
596	22ND ST	WILMOT RD	0.802	7224	63.8	128.025	1.134
584	22ND ST	ALVERNON WY	0.759	7167	57.2	113.876	3.654
348	SPEEDWAY BLV	WILMOT RD	0.740	6729	60.9	113.832	2.538
343	SPEEDWAY BLV	SWAN RD	0.729	6354	62.0	109.430	3.479
681	GOLF LINKS RD	CRAYCROFT RD	0.696	6052	66.6	111.962	0.703
346	SPEEDWAY BLV	CRAYCROFT RD	0.692	6408	58.7	104.486	3.065
223	GRANT RD	CAMPBELL RD	0.609	5705	53.5	84.783	5.261
262	TANQUE VERDE RD	GRANT RD	0.594	7227	47.2	94.754	0.839
219	GRANT RD	01ST AV	0.547	5456	56.3	85.326	1.486
575	22ND ST	06TH AV	0.534	3736	68.3	70.880	5.862
412	05TH ST	SWAN RD	0.513	4607	60.8	77.807	2.145
401	06TH ST	CAMPBELL AV	0.504	5603	46.9	72.995	3.400
600	22ND ST	KOLB RD	0.473	7165	37.4	74.436	1.018
416	05TH ST	CRAYCROFT RD	0.468	4627	55.5	71.333	1.865
341	SPEEDWAY BLV	ALVERNON WY	0.466	5888	41.0	67.058	3.291
335	SPEEDWAY BLV	CAMPBELL AV	0.463	6440	33.5	59.928	5.658
504	BROADWAY BLV	KOLB RD	0.447	6520	37.7	68.279	1.738
350	SPEEDWAY BLV	KOLB RD	0.405	6080	37.2	62.827	1.209
579	22ND ST	KINO PKWY	0.395	5242	41.1	60.283	1.507
582	22ND ST	COUNTRY CLUB RD	0.394	4713	44.2	57.865	2.357
99	PRINCE RD	ORACLE RD	0.391	5033	43.1	60.256	1.341
408	05TH ST	ALVERNON WY	0.385	4549	44.1	55.725	2.573
496	BROADWAY BLV	CRAYCROFT RD	0.380	6600	30.3	55.550	2.348
489	BROADWAY BLV	ALVERNON WY	0.345	6172	28.8	49.376	2.504
747	AJO WY	PARK AV	0.344	3928	49.0	53.464	0.994
587	22ND ST	SWAN RD	0.331	5904	31.6	51.824	0.790
338	SPEEDWAY BLV	COUNTRY CLUB RD	0.324	5259	30.0	43.825	3.258

Weight Factors Used:
Total Delay: 0.850 Trans. Person Delay: 0.150

The basic data base for the update of the capacity analysis was developed through the initial effort to establish the existing operating conditions. Traffic volume, intersection geometry, and traffic signal parameters must be updated periodically to facilitate future application of the developed procedures.

In situations where the duration of peak-period congestion varies between intersections, it is advisable to include a factor in the deficiency ranking that accounts for this phenomenon. A measure of the time duration of the estimated congestion levels could be used to factor the delay values used in the DI calculation.

In addition to the evaluation procedures described, a comprehensive data base management procedure was developed for COT to store information and to provide statistical analysis for both operational and safety improvement evaluation. This data base management procedure computes the operational DI and several safety-related indexes and provides numerous data-reporting and summarizing utilities. Such a data base management procedure is a key element application of these procedures as well as an application for updating the analysis in the development of future CIPs.

The procedure presented was intended to supplement long-range improvement implementation through the provision of direction for implementation of short-term improvements. An additional element that was not included in the analysis but that could prove important is the systemwide implications of improvements on the basis of the deficiencies identified. Similar procedures could be used to evaluate and rank corridors needing improvement. Also, consideration could be given to the addition of a factor in the DI to reflect a measure of systemwide importance.

ACKNOWLEDGMENTS

The work reported in this paper was conducted under contract with the city of Tucson, Arizona, Department of Transportation. The author wishes to thank Ms. Jill Merrick, city of Tucson Department of Transportation, for her assistance and cooperation throughout the conduct of the study.

REFERENCES

1. JHK & Associates. *Tucson Transportation System Planning Study—Final Report*. Transportation Planning Division, City of Tucson, Arizona, April 1991.
2. J. M. Witkowski. Methods Used To Evaluate Highway Improvements. *Journal of Transportation Engineering*, ASCE, Volume 109, No. 6, Nov. 1983.
3. *Special Report 209: Highway Capacity Manual*. TRB, National Research Council, Washington D.C., 1985.

Publication of this paper sponsored by Committee on Transportation System Management.

Cost-Effectiveness Analysis for Signalized Intersection Improvements

JAMES M. WITKOWSKI

A procedure is presented to prioritize operational improvements for signalized intersections based on a cost-effectiveness analysis. The measure of effectiveness used is the estimated reduction in person-hours of stopped delay resulting from the improvements. The procedure was developed to provide technical assistance in the evaluation of short- to medium-term improvement alternatives and to aid Capital Improvement Program (CIP) development. This procedure represents the second level in a two-level analysis process. The first-level analysis is a deficiency index to identify the intersections most in need of improvement. The application of the procedure is demonstrated through the evaluation of improvement alternatives for 30 intersections in Tucson, Arizona. These intersections were identified as being the most operationally deficient based on the first-level deficiency analysis. The procedure is easily applied through the aid of a computer spread sheet. The procedure facilitates the selection between alternatives at a single location and the ranking and selection of improvements for inclusion in a local CIP considering budget constraints.

Figure 1 shows a flow diagram of an analysis process for the identification and prioritization of signalized intersection improvement needs. The process shown in Figure 1 was developed for the city of Tucson (COT), Arizona to support the development of their annual Capital Improvement Program (CIP). The analysis process is designed to treat intersection operational improvements and safety improvements separately to ensure that the need for both improvement types is given proper consideration.

The analysis procedure shown in Figure 1 is a two-level screening process. The first level is designed to identify those intersections most in need of improvement and thus focus the attention of transportation agency resources on a manageable number of locations where improvements can have a significant impact on either congestion or safety. Separate operational and safety deficiency indexes were developed to provide the first-level screen and rank signalized intersections based on measures of operational and safety deficiency. The operational deficiency procedure is the subject of a companion paper in this Record by Witkowski. Details of the safety deficiency index analysis can be found in the project final report (*1*).

The research will define the development and application of a method to provide a cost-effectiveness analysis of signalized intersection operational improvements to aid in the technical evaluation of alternatives for a given location and facilitate the selection of projects for inclusion in the city's CIP. The cost-effectiveness analysis provides the second-level screen shown in Figure 1 and provides a method of establishing improvement priorities with consideration of funding constraints.

It was the primary goal of this project to provide COT with a comprehensive information database and evaluation procedure to assess the existing operating conditions of the city's signalized intersections and to evaluate existing intersection improvement needs and priorities. This study focused on individual congestion hot-spots and low-cost improvement alternatives for providing short- to medium-term congestion relief. The procedures developed and presented in this study were not intended to replace the long-range comprehensive planning process or the implementation of long-term transportation improvements. These procedures were intended to supplement long-range planning and provide direction in the selection of shorter-term improvements in lieu of factors that prevent the immediate implementation of long-range system plan elements.

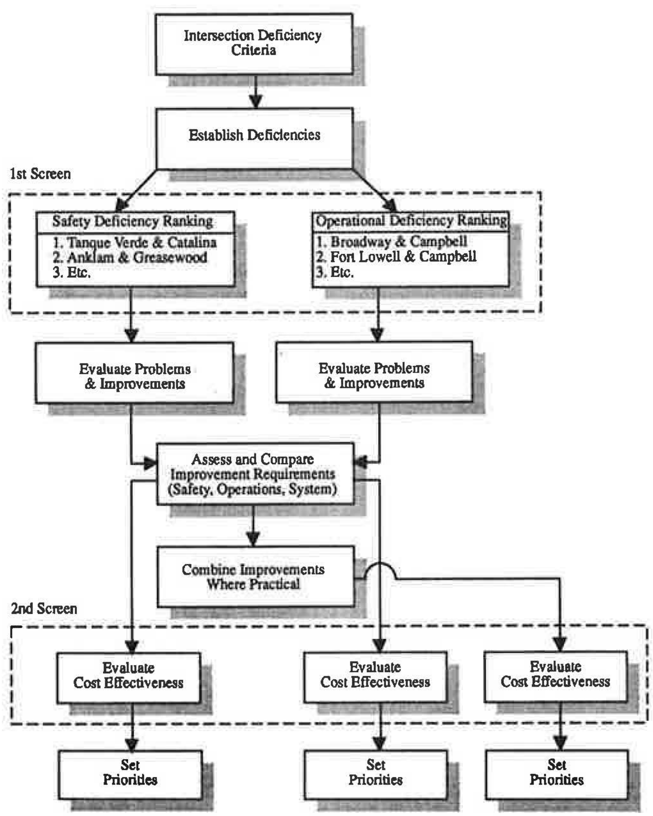

FIGURE 1 Recommended intersection analysis procedure.

JHK & Associates, 110 South Church Avenue, Suite 470, Tucson, Ariz. 85701.

The cost-effectiveness analysis will provide sufficient information for rational decision making when not all relevant factors can be evaluated or quantified. The use of a cost-effectiveness index to establish improvement priorities accepts the fact that not all potential evaluation factors have been entered into the analysis. The goal is to incorporate the most significant evaluation factors in the analysis so that the final improvement priorities provide a sound return for the required investment. The cost-effectiveness analysis is not intended or designed to be a detailed economic analysis or benefit-cost assessment.

PRIORITIZATION OBJECTIVES

Three basic objectives are used when applying economic analysis to roadway decision making (*2*):

- Maximize net benefits (i.e., the amount that benefits exceed costs),
- Minimize the amount of resources required to achieve a given level of service and meet other requirements demanded of the particular situation, and
- Maximize the level of service or other system performance measures from a given level of investment and operating cost.

Typically, one or more of these objectives are used to select between alternatives for a given location. In this case, the individual locations have already been screened using a deficiency index to identify the need for improvement based on delay. This, in effect, attempts to maximize the level of service for a given level of investment, because only the intersections most in need of improvements are subject to further analysis.

The minimization of resources required to obtain a desired level of service has been addressed, in part, by identifying improvement alternatives that provide a minimum acceptable level of service (LOS) for each movement on each intersection approach for the 30 intersections judged to be most operationally deficient. Except for four cases, this resulted in an estimated LOS C operating condition at each intersection from at least one of the improvement alternatives proposed.

The cost-effectiveness index (CEI) developed can be used to maximize the return for the investment by establishing improvement priorities based on the CEI and selecting improvements from the top of the list until the available funding is expended. The objective is to maximize the total cost-effectiveness within the constraints of the capital budget.

MEASURE OF EFFECTIVENESS

The measure of effectiveness used in this analysis was the change in the value of the estimated peak-hour person-hours of stopped delay derived from each of the intersection improvement alternatives. This was selected as the measure of effectiveness because peak-hour stopped delay was the primary criterion incorporated in the operational deficiency index used to identify intersection improvement needs and because reductions in delay are easily translatable to reduced user cost and vehicle emissions. (See the companion paper by Witkowski in this Record for a detailed description of the criteria selection process for the deficiency index.) The change in value of the peak-hour person-hours of stopped delay was divided by the estimated cost of each of the improvements (including the cost of right-of-way and construction). The measure of effectiveness includes the peak-hour vehicle-hours of stopped delay, estimated by using the 1985 *Highway Capacity Manual* (HCM) (*3*) operational analysis procedures for signalized intersections, factored by an average vehicle occupancy rate of 1.2 and the person-hours of transit rider stopped delay. Transit passenger delay was estimated on the basis of an estimate of the average transit passengers entering each intersection during the peak hour factored by the stopped delay per vehicle estimated using the 1985 HCM procedures.

COST-EFFECTIVENESS INDEX

The cost-effectiveness index used in this analysis is provided in Equation 1.

$$CEI = (D_1 - D_2) * \Big[\big(1.2V_1 * \{A + [(B - 1) * C]\}\big) + \big(T * \{A + [(E - 1) * C]\}\big)\Big] * 253 \text{ days/year} * \$4/\text{hr} * (1/3{,}600 \text{ sec/hr})/X \qquad (1)$$

where

- D_1 = the estimated average stopped delay per vehicle in seconds for the existing traffic volume and geometric conditions;
- D_2 = the estimated average stopped delay per vehicle in seconds for the existing traffic volume and improved geometric conditions;
- V_1 = the existing peak-hour traffic volume entering the intersection;
- T = the average hourly transit passengers entering the intersection on all approaches;
- A = (P/A, i, N) (i.e., the present worth factor for a uniform series at interest rate i for N years). This factor accounts for the assumption that the change in delay for the existing traffic volume is a constant amount for each year of the analysis period of the improvement (N years) and brings the benefits of the improvement back to their present worth based on interest rate i. The interest rate i was assumed to be 8 percent in this analysis. Values of A are presented in Table 1;
- B = (F/P, 3, N) (i.e., the single payment compound amount factor used to compute the maximum peak-hour total entering traffic volume at the end of the analysis period (N years) of the improvement, assuming a uniform annual growth rate of 3 percent). The 3 percent growth rate was based on an assessment of the forecast traffic growth and represents an average for the 30 intersections. Values of B are presented in Table 1;
- C = (P/G, i, N) (i.e., the gradient series present worth factor at interest rate i, for N years). This factor adjusts the annual change in benefits experienced by the growth in traffic volume to the present worth

of these benefits. As a simplification in the analysis process, it was assumed that the reduction in delay resulting from the improvement, and experienced by the increase in traffic volume, is the same as that for the existing traffic volume. This assumption was necessary due to limitations in the 1985 HCM procedures to produce reliable estimates of delay for conditions where the v/c ratio exceeds 1.2, which resulted when future traffic volumes were used with the existing intersection geometry in the analysis. The interest rate i was assumed to be 8 percent in this analysis. Values of C are presented in Table 1;

E = (F/P, i, N) (i.e., the single payment compound amount factor used to provide an estimate of the increase of the transit ridership for future years based on an average annual increase of i percent for N years where N represents the analysis period of the improvement). Transit passenger growth rates were assumed based on an analysis of local historical data. Values of E are presented in Table 1;

X = the estimated construction cost plus right-of-way cost of the improvement in thousands of dollars, prorated to account for the unequal service life spans of each improvement;

253 = a constant value representing the number of typical weekdays per year. The delay estimates and traffic volumes are for typical weekdays;

\$4/hr = the assumed average hourly value of an individual's travel time; and

3,600 = a constant representing the number of seconds per hour. This factor converts the estimated change in person delay from seconds to hours.

The CEI represents a partial assessment of one element of a traditional benefit-cost analysis, that is, the benefits associated with a reduction in travel time. Other roadway-user factors that are not directly included in this analysis are vehicle operating costs, safety, and the economic benefits associated with reductions in travel time for truck traffic. It is considered only a partial assessment of travel time benefits because the index considers the improvement impacts for only a single hour of the day. Therefore, the CEI cannot be construed as a benefit-cost ratio. However, reductions in vehicle operating cost are typically the single most significant factor in a benefit-cost assessment and have been shown to account for as much as 72 percent of the road-user benefits derived from intersection improvements (*4*). In addition, the change in vehicle operating cost through intersection improvements is directly related to the change in delay, so that the effect of this change is generally accounted for in the CEI.

IMPROVEMENT LIFE SPAN

The life span of an improvement for this project was interpreted to mean the number of years that the proposed improvement could be expected to function before requiring additional geometric improvements. LOS C is the current city standard for the desirable operating condition for signalized intersections. However, based on the evaluation performed as part of this study, intersections are not likely to be programmed for improvement unless the existing operating condition is LOS E or F. That is, intersections with operating conditions of LOS E or F accounted for 22 of the 30 most deficient cases. Also, the intersections with existing LOS D generally require only minor geometric improvements to achieve LOS C operation in comparison to intersections with a lower level of service.

It is not appropriate to consider LOS F operation as the threshold of the useful life on an intersection improvement. LOS F operation implies failure of the intersection to accommodate the demand with resulting severe traffic congestion. Improvements should be programmed before LOS F conditions occur. Therefore, LOS E was chosen as the threshold for determining the intersection life span.

Based on the 1985 HCM procedures, the range of stopped delay for LOS E conditions is 40.1 to 60.0 sec per vehicle. Using the lower range of the LOS E condition as the threshold implies a much shorter life for each improvement and is inconsistent with the existing conditions where intersections are allowed to function at a higher delay before improvements

TABLE 1 Values for Cost-Effectiveness Factors (*7*)

Improvement Life Span Years	A (P/A,8,N)	B (F/P,3,N)	C (P/G,8,N)	E (F/P,5,N)	E (F/P,8,N)	E (F/P,10,N)
1	0.926	1.030	0.000	1.050	1.080	1.100
2	1.783	1.061	0.857	1.102	1.166	1.210
3	2.577	1.093	2.445	1.158	1.260	1.331
4	3.312	1.126	4.650	1.216	1.360	1.464
5	3.993	1.159	7.372	1.276	1.469	1.611
6	4.623	1.194	10.523	1.340	1.587	1.772
7	5.206	1.230	14.024	1.407	1.714	1.949
8	5.747	1.267	17.806	1.477	1.851	2.144
9	6.247	1.305	21.808	1.551	1.999	2.358
10	6.710	1.344	25.977	1.629	2.159	2.594
11	7.139	1.384	30.266	1.710	2.332	2.853
12	7.536	1.426	34.634	1.796	2.518	3.138
13	7.904	1.469	39.046	1.886	2.720	3.452
14	8.244	1.513	43.472	1.980	2.937	3.797
15	8.559	1.558	47.886	2.079	3.172	4.177

are programmed. The upper range of the LOS E condition is approaching severe congestion conditions, and because of the lag between problem identification and improvement implementation, it does not seem reasonable to use the upper limit as the threshold for improvement life.

A value of delay in the mid-range of the LOS E condition was selected for use in defining the threshold for improvement life span. The existing intersection stopped delay from the 30 most deficient intersections was averaged and found to be just over 50 sec. This lends support to the use of 50 sec as the LOS E threshold in that this represents the average value for the intersection delay for those intersections that made the list of 30 and were evaluated for improvements, indicating that the existing condition had exceeded an acceptable condition for these intersections.

Fifty sec was used as the average intersection delay for the evaluation of the improvement life span. To determine the number of years of anticipated life span for each intersection, the existing traffic volumes were increased uniformly for all movements on all approaches by a factor of 3 percent per year until the average intersection delay equaled or exceeded 50 sec. The 3 percent factor was selected based on a historical review of traffic growth for the intersections and projected traffic growth rates. A uniform lane distribution was assumed in this analysis for the through lanes on all intersection approaches, that is, the lane distribution factor in the HCM analysis was set to a value of 1.0 for the through lanes. This is consistent with typical lane usage when traffic volume increases to near-capacity conditions.

ADJUSTING FOR UNEQUAL TIME PERIODS

Economic analysis and decision making theory dictate that alternatives must be examined over a common planning horizon. The planning horizon is the period of time over which the prospective consequences of various alternatives are assessed. Typically, the problem of a common planning horizon is dealt with in one of three ways:

1. The assumptions of a replacement cost or repetition of an expenditure is used to provide equal planning horizons for each alternative. In this case, this would require the evaluation of the improvement needs and life spans for successive improvements over a common time period for each of the independent locations. This was impractical because of the uncertainty regarding the need for improvement and the type required following the proposed improvements and because the establishment of a common time period for 30 locations with different expected life spans for each improvement is virtually impossible.

2. To account for the difficulties of defining future improvement needs and common planning horizons for multiple locations, it is typical for the analysis to be performed over an extended time period (e.g., 20 years) with the improvements required to satisfy a minimum service standard (e.g., LOS C or D) at the end of the planning horizon. This approach is used in the development of long-range transportation plans and the evaluation of long-range transportation improvement alternatives. This approach was considered impractical for this study because the intent was to identify low-cost improvement alternatives to address existing problems, whereas the long-range analysis typically results in the requirement for high-cost major capital improvements.

3. A third approach for adjusting the analysis to account for unequal time periods is described by Winfrey (*5*).

> The adjustment in procedure to equalize the analysis periods between the alternatives is to use an analysis period equal to the shorter lived alternative and to allow a terminal value for the remainder of the life of the longer lived alternative. This terminal value would be equal to the value of the unexpired service, based upon a pro rata share of the original investment.

The latter approach for dealing with unequal time periods was selected for use in this study for the development of the final CEI. This approach was implemented by setting the analysis time period (N) in Equation 1 equal to the life span of the shortest-lived alternative and prorating the cost of each alternative by adjusting the denominator in Equation 1 as shown in Equation 2:

$$X = (C_c + C_R) * N/Y \tag{2}$$

where

X = the prorated construction plus right-of-way cost of the improvement,
C_c = estimated construction cost of the improvement,
C_R = the estimated right-of-way (ROW) cost of the improvement,
N = the analysis period in years (i.e., the life span of the shortest-lived alternative), and
Y = the estimated life span of the improvement in years.

It could be argued that the right-of-way life exceeds the project life, in which case the ROW cost would be prorated over a longer time period in the cost-effectiveness analysis. However, this would result in projects with shorter life spans and high ROW cost moving up in the final rankings because of the longer ROW cost proration period. This was believed to negatively affect the final results and was rejected in favor of the relationship in Equation 2.

IMPROVEMENT ALTERNATIVES

Operational improvement alternatives were developed for each of the 30 most deficient intersections. These alternatives were developed through an assessment of the following factors:

- Existing intersection geometry,
- P.M. peak-hour level of service by intersection approach and movement,
- Desirable intersection LOS of C,
- Intersection approach lane balance, and
- Current reconstruction plans.

The geometric improvement alternatives included adding or improving exclusive turn lanes, adding through lanes, reversible lanes (continuous exclusive left-turn lanes during the off-peak time periods), and the use of grade-separated intersections. The improvement alternatives are presented in Table 2. In most cases, two alternatives were evaluated for each

TABLE 2 Alternative Intersection Improvement Cost Assessment

Intersection Name	Alt.	Int. No.	DI (4)	Improvements (3) EB	WB	NB	SB		Construction Cost (1) ($ X 1,000)	Right-of-Way Cost (2) ($ X 1,000)	Total Cost ($ X 1,000)
Broadway/Campbell		483	1	No improvement needed at this time due to short duration of congestion problem.							
Ft. Lowell/Campbell	A	123	2	R	R	R	R	Reversible Lane N/S	200	700	900
Ft. Lowell/Campbell	B	123	2	T	T	T	T		300	1080	1380
Grant/Craycroft	A	234	3	L,R	R	R	L,R		330	1030	1360
Grant/Craycroft	B	234	3	L,R	R	R	R	Reversible Lane N/S	310	1030	1340
Grant/Craycroft	C	234	3	L,T	R	T	L,T		420	1130	1550
22nd/Wilmot	A	596	4	L,IR	IR	IR	L,IR		210	1100	1310
22nd/Wilmot	B	596	4	GRADE SEPARATION					9000	5000	14000
22nd/Alvernon	A	584	5	IL		L	R		145	470	615
22nd/Alvernon	B	584	5	L		L,T	L		300	750	1050
Speedway/Wilmot	A	348	6	L,R		L,T,R			275	1120	1395
Speedway/Wilmot	B	348	6	L,R	L	L,R	L		390	770	1160
Speedway/Swan	A	343	7	R	R			Reversible Lane N/S	110	350	460
Speedway/Swan	B	343	7	R		T	T		235	1460	1695
Golf Links/Craycroft	A	681	8	L			L		150	320	470
Golf Links/Craycroft	B	681	8				L		75	160	235
Speedway/Craycroft	A	346	9	R	IL,R		R		160	790	950
Speedway/Craycroft	B	346	9	R	IL,R		R	Reversible Lane N/S	180	790	970
Speedway/Craycroft	C	346	9	R	L	T	L,T		345	1600	1945
Grant/Campbell	A	223	10			R	R	Reversible Lane N/S	110	380	490
Grant/Campbell	B	223	10	GRADE SEPARATION					9000	5000	14000
Tanque Verde/Grant	A	262	11		L,IR	IR	R		150	1350	1500
Tanque Verde/Grant	B	262	11	GRADE SEPARATION					9000	5000	14000
Grant/First	A	219	12	R	R	R	L,R		255	920	1175
Grant/First	B	219	12	R	R	R	R	Reversible Lane N/S	200	920	1120
22nd/Sixth	A	575	13	R	R	R	R	Reversible Lane E/W	200	930	1130
22nd/Sixth	B	575	13	T	T				150	770	920
Fifth/Swan	A	412	14	R	R				90	270	360
Fifth/Swan	B	412	14			T	T		150	750	900
Sixth/Campbell	A	401	15	IR	R	R	L,R		225	1760	1985
Sixth/Campbell	B	401	15				L		75	160	235
22nd/Kolb	A	600	16	L	L		L		225	480	705
Fifth/Craycroft	A	416	17	R	R	T,R	T,R		280	2900	3180
Fifth/Craycroft	B	416	17			T	T		150	400	550
Speedway/Alvernon	A	341	18	L	L				150	620	770
Speedway/Campbell		335	19	Improvements to be provided by Phase II reconstruction.							
Broadway/Kolb	A	504	19	L,IR	L,IR	L,R	L,IR		360	1740	2100
Broadway/Kolb	B	504	20	L	L				150	320	470
Speedway/Kolb	A	350	21			L	L		150	320	470
22nd/Kino	A	579	22	R	R				100	270	370
22nd/Kino	B	579	22	T	T				150	540	690
22nd/Country Club	A	582	23	R			L,R		165	430	595
22nd/Country Club	B	582	23	R			L		120	390	510
Prince/Oracle	A	99	24	L,IR	IR	L,IR	L,IR		285	1675	1960
Prince/Oracle	B	99	24	L		L	L		225	480	705
Fifth/Alvernon	A	408	25	R	R	R	R		180	700	880
Fifth/Alvernon	B	408	25	R		R	R		135	860	995
Broadway/Craycroft	A	496	26				L		75	160	235
Broadway/Alvernon	A	489	27		L	L	IL		175	370	545
Broadway/Alvernon	B	489	27		L				75	880	955
Ajo/Park	A	747	28	R	R	R	R		180	620	800
Ajo/Park	B	747	28	R		R			90	270	360
22nd/Swan	A	587	29	L	R		L		195	600	795
22nd/Swan	B	587	29	L			L		150	590	740
Speedway/Country Club		338	30	Improvements to be provided by Phase II reconstruction.							

1. These cost figures are based on estimates provided by the City of Tucson for use in this study.
2. These values do not constitute appraisals, but are estimates only for use in this study. These estimates were provided by the City of Tucson and are based on the following assumptions: a) Land at $15/sq. ft. b) Lost parking at $5,000/space. c) Lost Buildings at $60/sq. ft. for older retail and $125/sq. ft. for new office. d) Previous severance amounts from prior roadway widenings.
3. L = left turn lane, R = right turn lane, T = through lane, IL = improved left turn lane, IR = improved right turn lane
4. DI = deficiency index rank

location. In two cases, three alternatives were evaluated, and in three cases no alternatives were proposed. No alternatives were proposed for the intersection of Broadway and Campbell because the traffic congestion problem was found to exist for a very short duration (approximately 15 min) and coincided with the end of the workday at the nearby University of Arizona. No improvements were proposed for the intersections of Speedway and Country Club or Speedway and Campbell because these intersections had already been scheduled for reconstruction under an arterial improvement project.

INTERSECTION IMPROVEMENT COSTS

The cost estimates for intersection improvements were developed on the basis of data provided by the city of Tucson from recent construction projects. These data consisted of average construction costs for various improvement types, which are presented in Table 3. In addition to the construction costs, the ROW costs for each individual intersection were evaluated by the city based on the individual right-of-way requirements for each proposed improvement and the existing geometric configuration of the intersection. The construction and right-of-way estimates for each intersection used in the cost-effectiveness analysis are presented in the alternative improvement summary in Table 2.

The cost estimates provided here were not developed based on a detailed assessment of each individual location, but on averages for each type of improvement. These estimates are designed to provide a planning level estimate of costs for use in the cost-effectiveness analysis. More detailed and precise cost estimates would be required prior to actual construction of any intersection improvements.

DECISION MAKING UNDER UNCERTAINTY

Several assumptions and simplifications in this analysis are a direct result of uncertainties encountered in developing the procedure and limitations in the technical analysis tools available. These assumptions and simplifications are described here with a brief discussion of how each may affect the results of the cost-effectiveness analysis.

Uniform Traffic Growth

The assumption of uniform traffic growth was a direct result of the uncertainty associated with traffic forecasts at any individual location. Traffic growth is affected by various factors, and the actual growth rate experienced at any single location could vary significantly from the assumed 3 percent growth rate. Also, the assumption of uniform growth for all traffic movements will most likely not occur.

The traffic growth rate directly affects the estimated life span of each improvement. Lower growth rates increase the life span and generally improve the cost-effectiveness of any improvement. Higher growth rates shorten the life span and lower the cost-effectiveness. For improvements with equal life spans and costs, a higher traffic growth rate will result in a faster accumulation of benefits and a higher cost-effectiveness.

Delay Estimation

The delay estimates for the existing and improved condition for each intersection were generated using the procedures in the 1985 HCM for signalized intersections. Typically, the delay estimates should be generated for the first year and the last year of the planning horizon for both the existing and improved conditions. However, this was not possible due to a limitation in the 1985 HCM procedures. Once the volume to capacity (v/c) ratio for any movement on any approach exceeds a value of 1.2, the HCM delay estimate increases geometrically and is considered unreliable.

To adjust for this problem in computing delay for the existing condition under the increased traffic volume scenario, it was assumed that the difference in the delay between the existing and improved condition remained constant as traffic volume increased. This assumption is most likely conservative because the delay experienced on the existing intersection condition would generally increase at a faster rate than on

TABLE 3 Construction Cost Estimates for Various Intersection Improvements

Improvement Type	Construction Cost Per Approach
Left-Turn Lane	\$40,000
Right-Turn Lane	\$45,000 (2)
Through Lane (1)	\$75,000
Reversible Lane	\$20,000 (per mile of treatment)

(1) Assumes construction 1,000 ft. upstream and 500 ft. downstream of intersection.

(2) Assumes right-turn acceleration lane included.

NOTE: For some intersections, the construction and right-of-way cost estimate includes the cost of opposing left-turn lane improvements when one of the left-turn lanes is considered to be "stripped out" and not used operationally. This condition was considered necessary for safety reasons.

Source: Estimates provided by the city of Tucson based on averages for each improvement type.

the improved condition assuming an equal rate of traffic growth. Therefore, the cost-effectiveness of each improvement would most likely be higher than estimated. The impact of this assumption on the final rankings of the improvements by the CEI is unknown but depends on how each individual location is affected.

Transit Passenger Growth

The impacts of the assumptions for transit ridership have very little impact on the results of this analysis because the number of automobile passengers is so much higher than transit passengers. Excluding transit ridership from Equation 1 results in an average 3.5 percent decrease in the CEI for each intersection with a maximum change of approximately 8 percent. The impact of excluding transit passenger delay in the CEI is very minor.

The assumed rate of increase in transit passengers does have a minor impact on the ranking of intersections. For intersection improvements that have approximately equal cost-effectiveness excluding transit passengers, the higher rate of transit passenger increase will result in a higher CEI value. The differential rate of increase for transit passengers for each intersection does not have a significant impact on the overall ratings.

Cost Estimates

The single most influential factor in establishing the CEI rankings is the cost estimate of the improvements. These values should be reviewed carefully to determine if they are indeed realistic for the improvements being proposed. Changes in the cost estimates may change the overall CEI rankings dramatically.

RESULTS AND CONCLUSIONS

The cost-effectiveness analysis was performed at two levels. The first-level analysis was a comparison of the alternatives at each individual location to aid in the selection of the preferred alternative. This comparison of the grouped alternatives is presented in Table 4. This information was used in conjunction with other information on geometric lane balance and right-of-way availability to select the preferred alternative from each group.

The description of the preferred alternative for each location is presented in Table 5. The cost-effectiveness analysis results for the preferred alternatives and the ranking of alternatives by the cost-effectiveness index are presented in Table 6.

In general, the application of reversible lanes is very cost-effective, and these alternatives tend to rise to the top of the cost-effectiveness ranking. The elimination of left turns and the use of the left-turn green time for the through movement is a major factor in the effectiveness of the reversible lane alternatives.

The addition of exclusive left-turn lanes also tends to be a very cost-effective improvement. The use of dual left-turn lanes reduces the green time required for the left-turn movement and allows this time to be given to the through movement, which increases the effectiveness of the alternative.

A CIP can be established from the information in Table 6 that will approximate a maximum return in operational improvements by starting at the top of the list and selecting projects until the maximum budget has been reached. For example, with a $2 million budget, the first four projects plus project six would be selected at a cost of $1,780,000. With a $3 million budget, the first seven projects would be selected at a cost of $2,715,000.

The urban grade-separated intersections do not fare well in this analysis. This is primarily because of the large cost associated with these projects. This should not be construed to imply that grade-separated intersections are not cost-effective. In fact, they have been shown to be very cost-effective for high volume intersections in comparison to at-grade treatments (*4,6*). This analysis simply shows that, in the short-term, for intersections that have not yet reached a need for a grade separation, other types of improvements can be more cost-effective. In general, intersections requiring grade separations to achieve a desirable level of service should not be included in an analysis of intersections that can be improved through low-cost treatments. They should be compared only to other intersections requiring similar treatments and levels of investment, and the benefits should be assessed on a systemwide rather than on an isolated location basis.

Another important issue regarding the cost-effectiveness of transportation investments requires exploration. The results developed in this study and used to recommend short-range improvement projects must be viewed within the context of the long-range transportation plan development and implementation. The true maximum return for the transportation investments can only be achieved through the analysis of long-range transportation system needs and investment opportunities. The use of the short-range procedures developed in this study will not necessarily provide the best investment opportunities for the long term. In fact, the results in Table 6 could lead one to the erroneous conclusion that grade separations should not be implemented until all signalized intersections in the city were built to the maximum at-grade geometrics. In addition, projects selected for implementation from the short-range project list should support the goals and objectives of the long-range plan.

The correct conclusions to be derived from this study are that under the annual budget limitations for the provision of short-range spot improvements, candidate projects vary considerably in their cost-effectiveness. The procedures developed in this project can provide useful information for determining which short-range projects provide the most return, measured in reduction in peak-hour stopped delay, for the investment.

Other factors should also be considered when applying this procedure. A factor accounting for the duration of the delay should be included in the analysis in those areas experiencing extended peak periods. This concept could be expanded to include estimates of peak and off-peak period delay if there was significant variation in these time periods between intersections. This was not considered a major issue for this study.

It may be desirable to include vehicle operating cost and the value of a reduction in accidents in the cost-effectiveness

TABLE 4 Signalized Intersection Cost-Effectiveness Analysis Grouped Alternatives

Intersection Name	Alt	Int. No.	DI	Proj. Years	Anal. Years (1)	Exist. Delay (secs)	Improve Delay (secs)	Exist. Vol. (vph)	Trans Pass.	Effect (2)	Const. Cost	ROW Cost	Total Cost	CEI
Ft. Lowell/Campbell	A	123	2	8	7	88	21	5680	189	1121085	200	700	900	1424
Ft. Lowell/Campbell	B	123	2	7	7	88	28	5680	189	1003956	300	1080	1380	728
Grant/Craycroft	A	234	3	8	8	78	25	6415	130	1231489	330	1030	1360	906
Grant/Craycroft	B	234	3	10	8	78	22	6415	130	1301196	310	1030	1340	1214
Grant/Craycroft	C	234	3	8	8	78	25	6415	130	1231489	420	1130	1550	795
22nd/Wilmot	A	596	4	1	1	64	51	7224	64	29550	210	1100	1310	23
22nd/Wilmot	B	596	4	20	1	64	14	7224	64	113652	9000	5000	14000	162
22nd/Alvernon	A	584	5	2	2	57	47	7167	230	45586	145	470	615	74
22nd/Alvernon	B	584	5	8	2	57	25	7167	230	145874	300	750	1050	556
Speedway/Wilmot	A	348	6	5	5	61	31	6729	150	359483	275	1120	1395	258
Speedway/Wilmot	B	348	6	10	5	61	21	6729	150	479311	390	770	1160	826
Speedway/Swan	A	343	7	8	8	62	22	6354	202	959263	110	350	460	2085
Speedway/Swan	B	343	7	8	8	62	23	6354	202	935281	235	1460	1695	552
Golf Links/Craycroft	A	681	8	4	4	67	35	6052	38	256377	150	320	470	545
Golf Links/Craycroft	B	681	8	6	4	67	27	6052	38	320471	75	160	235	2046
Speedway/Craycroft	A	346	9	3	3	59	44	6408	188	93261	160	790	950	98
Speedway/Craycroft	B	346	9	7	3	59	23	6408	188	223826	180	790	970	538
Speedway/Craycroft	C	346	9	11	3	59	19	6408	188	248695	345	1600	1945	469
Grant/Campbell	A	223	10	11	11	54	18	5705	354	1402874	110	380	490	2863
Grant/Campbell	B	223	10	20	11	54	19	5705	354	1363905	9000	5000	14000	177
Tanque Verde/Grant	A	262	11	4	4	47	34	7227	64	124476	150	1350	1500	83
Tanque Verde/Grant	B	262	11	20	4	47	14	7227	64	315979	9000	5000	14000	113
Grant/First	A	219	12	6	6	56	24	5456	95	399546	255	920	1175	340
Grant/First	B	219	12	12	6	56	16	5456	95	499432	200	920	1120	892
22nd/Sixth	A	575	13	4	4	68	32	3736	309	190241	200	930	1130	168
22nd/Sixth	B	575	13	6	4	68	29	3736	309	206095	150	770	920	336
Fifth/Swan	A	412	14	5	5	61	27	4607	127	280336	90	270	360	779
Fifth/Swan	B	412	14	7	5	61	25	4607	127	296827	150	750	900	462
Sixth/Campbell	A	401	15	10	9	47	20	5603	261	694171	225	1760	1985	389
Sixth/Campbell	B	401	15	9	9	47	22	5603	261	642751	75	160	235	2735
22nd/Kolb	A	600	16	9	9	37	23	7165	98	443333	225	480	705	629
Fifth/Craycroft	A	416	17	12	8	56	19	4627	121	624132	280	2900	3180	294
Fifth/Craycroft	B	416	17	8	8	56	23	4627	121	556659	150	400	550	1012
Speedway/Alvernon	A	341	18	8	8	41	23	5888	289	413461	150	620	770	537
Broadway/Kolb	A	504	20	10	8	38	18	6520	166	481239	360	1740	2100	286
Broadway/Kolb	B	504	20	8	8	38	23	6520	166	360929	150	320	470	768
Speedway/Kolb	A	350	21	8	8	37	24	6080	117	291037	150	320	470	619
22nd/Kino	A	579	22	7	7	41	30	5242	131	168400	100	270	370	455
22nd/Kino	B	579	22	11	7	41	21	5242	131	306181	150	540	690	697
22nd/Country Club	A	582	23	10	7	44	16	4713	192	391773	165	430	595	941
22nd/Country Club	B	582	23	7	7	44	27	4713	192	237862	120	390	510	466
Prince/Oracle	A	99	24	13	11	43	17	5033	112	880229	285	1675	1960	531
Prince/Oracle	B	99	24	11	11	43	20	5033	112	778664	225	480	705	1104
Fifth/Alvernon	A	408	25	7	7	44	22	4549	210	298776	180	700	880	340
Fifth/Alvernon	B	408	25	7	7	44	25	4549	210	258034	135	860	995	259
Broadway/Craycroft	A	496	26	11	11	30	20	6600	279	455061	75	160	235	1936
Broadway/Alvernon	A	489	27	9	8	29	21	6172	313	189546	175	370	545	391
Broadway/Alvernon	B	489	27	8	8	29	24	6172	313	118467	75	880	955	124
Ajo/Park	A	747	28	14	12	49	17	3928	73	967832	180	620	800	1411
Ajo/Park	B	747	28	12	12	49	20	3928	73	877097	90	270	360	2436
22nd/Swan	A	587	29	11	9	32	19	5904	90	339832	195	600	795	522
22nd/Swan	B	587	29	9	9	32	22	5904	90	261410	150	590	740	353

(1) Based on least value of project years for group.
(2) Present value of travel time saved.

TABLE 5 Intersection Improvement Summary—Preferred Alternatives (Based on CEI Only)

Intersection Name	Alt.	Int. No.	DI (4)	Improvements (3) EB	WB	NB	SB		Construction Cost (1) ($ X 1,000)	Right-of-Way Cost (2) ($ X 1,000)	Total Cost ($ X 1,000)
Ft. Lowell/Campbell	A	123	2	R	R	R	R	Reversible Lane N/S	200	700	900
Grant/Craycroft	B	234	3	L,R	R	R	R	Reversible Lane N/S	310	1030	1340
22nd/Wilmot	B	596	4	GRADE SEPARATION					9000	5000	14000
22nd/Alvernon	B	584	5	L		L,T	L		300	750	1050
Speedway/Wilmot	B	348	6	L,R	L	L,R	L		390	770	1160
Speedway/Swan	A	343	7	R	R			Reversible Lane N/S	110	350	460
Golf Links/Craycroft	B	681	8				L		75	160	235
Speedway/Craycroft	B	346	9	R	IL,R		R	Reversible Lane N/S	180	790	970
Grant/Campbell	A	223	10			R	R	Reversible Lane N/S	110	380	490
Tanque Verde/Grant	A	262	11		L,IR	IR	R		150	1350	1500
Grant/First	C	219	12	R	R	R	R	Reversible Lane N/S	200	920	1120
22nd/Sixth	B	575	13	T	T				150	770	920
Fifth/Swan	A	412	14	R	R				90	270	360
Sixth/Campbell	B	401	15				L		75	160	235
22nd/Kolb	A	600	16	L	L		L		225	480	705
Fifth/Craycroft	B	416	17			T	T		150	400	550
Speedway/Alvernon	A	341	18	L	L				150	620	770
Speedway/Campbell		335	19	Improvements to be provided by Phase II reconstruction.							
Broadway/Kolb	B	504	20	L	L				150	320	470
Speedway/Kolb	A	350	21			L	L		150	320	470
22nd/Kino	B	579	22	T	T				150	540	690
22nd/Country Club	A	582	23	R			L,R		165	430	595
Prince/Oracle	B	99	24	L		L	L		225	480	705
Fifth/Alvernon	A	408	25	R	R	R	R		180	700	880
Broadway/Craycroft	A	496	26				L		75	160	235
Broadway/Alvernon	A	489	27		L	L	IL		175	370	545
Ajo/Park	B	747	28	R		R			90	270	360
22nd/Swan	A	587	29	L	R		L		195	600	795
Speedway/Country Club		338	30	Improvements to be provided by Phase II reconstruction.							

1. These cost figures are based on estimates provided by the City of Tucson for use in this study.
2. These values do not constitute appraisals, but are estimates only for use in this study. These estimates were provided by the City of Tucson and are based on the following assumptions: a) Land at $15/sq. ft. b) Lost parking at $5,000/space. c) Lost Buildings at $60/sq. ft. for older retail and $125/sq. ft. for new office. d) Previous severance amounts from prior roadway widenings.
3. L = left turn lane, R = right turn lane, T = through lane, IL = improved left turn lane, IR = improved right turn lane
4. DI = deficiency index rank

analysis. This would provide an analysis more closely related to a benefit-cost assessment. These factors were not included in the analysis for several reasons. The purpose of the cost-effectiveness analysis was to provide a simple mechanism for ranking intersection improvements. This included the factors contributing the most to the benefits associated with the improvement and resulted in a reasonable list of improvement priorities. Previous work (*4*) has shown that the value of the reduction in delay is the single most significant benefit factor resulting from operational improvements at intersections, accounting for between 60 and 75 percent of the benefits achieved. The vehicle operating cost could be added to the cost-effectiveness analysis and could be computed as a function of the additional fuel consumption resulting from the stopped delay. Because stopped delay is already included in the analysis, the inclusion of another term based on stopped delay would probably not change the relative ranking of improvement alternatives to a large extent.

The results from the initial phase of this study (see companion paper by Witkowski in this Record) indicated that there was no linear correlation between accident rate and level of congestion as measured by stopped delay. Therefore, reducing the delay through capacity improvements could not be easily associated with a reduction in accidents. Accident reduction at congested intersections has been shown to represent only a small portion (on the order of 10 percent) of the benefits associated with capacity improvements when this benefit is actually realized (*4*). The majority of the benefits are from the reduction in travel time. A detailed accident analysis would have been required to estimate the potential accident reduction resulting from the improvements evaluated in order to include the typically small additional benefit in the analysis. A separate procedure was used in this study to identify intersections with high accident rates. None of the intersections with high accident rates appeared on the list of 30 intersections with high operational deficiencies.

ACKNOWLEDGMENTS

The work reported in this paper was conducted under contract to the city of Tucson, Arizona, Department of Transportation. The author wishes to thank Ms. Jill Merrick, city of Tucson Department of Transportation, for her assistance and cooperation throughout the conduct of the study.

TABLE 6 Signalized Intersection Cost-Effectiveness Analysis for Preferred Alternatives (Based on CEI Only)

Intersection Name	Alt	Int. No.	DI	Proj. Years	Anal. Years (1)	Exist. Delay (secs)	Improve Delay (secs)	Exist. Vol. (vph)	Trans Pass.	Effect (2)	Const. Cost	ROW Cost	Total Cost	CEI
Golf Links/Craycroft	B	681	8	6	5	67	27	6052	38	425571	75	160	235	2173
Sixth/Campbell	B	401	15	9	5	47	22	5603	261	255208	75	160	235	1955
Grant/Campbell	A	223	10	11	5	54	18	5705	354	379575	110	380	490	1704
Speedway/Swan	A	343	7	8	5	62	22	6354	202	462289	110	350	460	1608
Ft. Lowell/Campbell	A	123	2	8	5	88	21	5680	189	684773	200	530	700	1565
Ajo/Park	B	747	28	12	5	49	20	3928	73	202135	90	270	360	1348
Broadway/Craycroft	A	496	26	11	5	30	20	6600	279	120882	75	160	235	1132
Grant/Craycroft	B	234	3	10	5	78	22	6415	130	638489	310	1030	1340	953
Grant/First	B	219	12	12	5	56	16	5456	95	386830	200	920	1120	829
Speedway/Wilmot	B	348	6	10	5	61	21	6729	150	479311	390	770	1160	826
22nd/Country Club	A	582	23	10	5	44	16	4713	192	239123	165	430	595	804
Fifth/Craycroft	B	416	17	8	5	56	23	4627	121	272906	150	400	550	794
Fifth/Swan	A	412	14	5	5	61	27	4607	127	280336	90	270	360	779
Prince/Oracle	B	99	24	11	5	43	20	5033	112	207917	225	480	705	649
22nd/Alvernon	B	584	5	8	5	57	25	7167	230	412218	300	750	1050	628
Broadway/Kolb	B	504	20	8	5	38	23	6520	166	175680	150	320	470	598
Speedway/Craycroft	B	346	9	7	5	59	23	6408	188	413557	180	790	970	597
22nd/Kino	B	579	22	11	5	41	21	5242	131	187175	150	540	690	597
Speedway/Kolb	A	350	21	8	5	37	24	6080	117	141397	150	320	470	481
22nd/Kolb	A	600	16	9	5	37	23	7165	98	177164	225	480	705	452
Speedway/Alvernon	A	341	18	8	5	41	23	5888	289	197149	150	620	770	410
22nd/Swan	A	587	29	11	5	32	19	5904	90	135761	195	600	795	376
22nd/Sixth	B	575	13	6	5	68	29	3736	309	274386	150	770	920	358
Broadway/Alvernon	A	489	27	9	5	29	21	6172	313	91308	175	370	545	302
Fifth/Alvernon	A	408	25	7	5	44	22	4549	210	182265	180	700	880	290
22nd/Wilmot	B	596	4	20	5	64	14	7224	64	634989	9000	5000	14000	181
Tanque Verde/Grant	B	262	11	20	5	47	14	7227	64	419266	9000	5000	14000	120

(1) Based on least value of project years for all alternatives.
(2) Present value of travel time saved.

REFERENCES

1. JHK & Associates. *Tucson Transportation System Planning Study—Final Report.* Transportation Planning Division, Tucson, Ariz., April 1991.
2. B. Campbell and T. F. Humphrey. Methods of Cost-Effectiveness Analysis for Highway Projects. *NCHRP Synthesis of Highway Practice 142*, TRB, National Research Council, Washington, D.C., 1988.
3. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
4. J. M. Witkowski. Benefit Analysis for Urban Grade Separated Interchanges. *Journal of Transportation Engineering*, ASCE, Vol. 114, No. 1, Jan. 1988.
5. R. Winfrey. *Economic Analysis for Highways.* International Textbook Company, Scanton, Pa., 1969.
6. Johnson-Brittain and Associates. *Campbell-Grant Advance Planning Study.* Tucson, Ariz., 1990.
7. G. A. Fleischer. *Engineering Economy—Capital Allocation Theory.* Brooks/Cole Engineering Division, 1984.

Publication of this paper sponsored by Committee on Transportation System Management.

Making Intelligent Vehicle/Highway Systems Really Work: A Status Report on the Congestion Avoidance and Reduction for Autos and Trucks Project

DAVID T. HARTGEN AND EDWARD L. BRUCE

Most Intelligent Vehicle/Highway Systems (IVHS) studies focus on the characteristics of selected technologies operating in narrow limited environments, consisting of a few selected corridors or markets in the largest cities. If successful, such applications would have a small overall effect on congestion. A different approach—placing full-service technologies into the hands of all drivers in medium-sized cities—holds much greater promise for real impact. One such proposal for the Charlotte, North Carolina, metropolitan area, a region of 1.6 million people, is described. Termed CARAT (Congestion Avoidance and Reduction for Autos and Trucks), the concept is an information provision (ATIS) element of a broad congestion and traffic management strategy being developed by the North Carolina Department of Transportation. CARAT envisions a large information network based on middle-tech devices, serving consumers, truckers, fixed sites, and emergency services. Telephone, radio, and other selected audio systems are used to transmit data directly to clients from a central office using straightforward geographic information system–based information storage technology. Only marketable and cost-effective services are offered, at low subsidy. CARAT would be partially supported by private-sector advertising and subscriptions. A careful market survey is being used to gauge the demand for services before implementation. A parallel effort to outfit the greater Charlotte freeway and expressway system with modern ATMS traffic operation technology is also in the development stage. The CARAT approach is not only feasible, but more cost-effective and more widely applicable than its high-tech IVHS cousins.

Intelligent vehicle highway systems (IVHS) are modern communications systems that give the driver real-time information on highway conditions (*1*). A variety of technologies are now available or emerging that permit the driver to receive data or warnings on traffic congestion, incidents, weather, and safety, as well as to transmit information on the vehicle's location to others. The devices can be as simple as radio or cellular telephone systems with electronic roadside sensors, or more complex with on-board electronic maps, in-street traffic detectors, satellite location systems, and on-board information systems. Other companies are experimenting with high-speed crash-avoidance technologies that would permit closer spacing on freeways and higher traffic volumes (*2*). Drivers use the information to avoid or reduce congestion or incidents, thereby saving time, fuel, operating costs, and accidents. These systems have been given much attention in the last 5 years, particularly in Europe and Japan, where they are seen as providing significant congestion relief, fuel savings, and safety improvements. In Europe, governments and businesses have combined under a broad project known as Prometheus, the goal of which is to reduce accidents and congestion by 40 percent (*3*). In Japan, a similar project would control traffic flows for a 300-mi^2 area of Tokyo (*4*). Each of these projects is very large, proposing \$700 million to \$800 million in expenditures.

D. T. Hartgen, Department of Geography and Earth Sciences, University of North Carolina, Charlotte, N.C. 28223. E. L. Bruce, Federal Highway Administration, Atlanta, Ga.

In the United States several demonstrations of these technologies have been initiated on a smaller scale. In Los Angeles, the Pathfinder project is using about 25 cars to test motorist information use in the congested Santa Monica corridor. In San Francisco, a tourist-oriented map system is being tested at airports and hotels. In Orlando, a demonstration of providing tourist information to drivers, TRAVTEK, is being tested. In New York, traffic flow coordination on the Long Island Expressway is being implemented (*4*). The federal government has initiated limited funding for IVHS research: \$150 million is being suggested for each of the fiscal years 1992 to 1996 (*5*). The U.S. General Accounting Office (*6*) has concluded that the technology can be useful but must be pursued aggressively.

However, the present studies focus primarily on large cities, with limited general application to the congestion problems of most U.S. cities. Systems concentrating on one large corridor in a metropolitan area are likely to have very limited local effect if not applicable beyond that corridor. One study by the authors suggests that overall IVHS adoption and use rates in major cities would have to be at least 50 percent before even slight (5 percent) changes in congestion could be noticed (*7*). More generally available technologies capable of more extensive use in a variety of settings are needed if IVHS is to ultimately be a cost-effective way of reducing traffic congestion. There is a need to develop and test systems that cover wider applications of congestion problems in many more places. Mid-sized cities such as Charlotte and its surrounding communities are more typical of the U.S. situation and provide an opportunity for tests and demonstrations that can be transported to other sites.

The fundamental problem that needs to be addressed in IVHS technology is not whether the technology will work (it has generally been established that the technology will work), but rather its cost and value to the average user. Supply and demand, balanced by costs and the characteristics of these technologies in a mix of marketplace services, will ultimately determine whether or not the demand for these technologies is sufficient to justify investments in their development and implementation. Unless IVHS supply and demand can be balanced at a reasonable price to consumers, services must be cut back or made more specialized or, alternatively, subsidized with the intent of gaining additional benefits to users and nonusers. Such assessments cannot be undertaken using small demonstrations of technology, but rather must be subjected to the hard spotlight of choices made by consumers in the course of everyday behavior.

The intent of this paper is to describe a recent effort to develop and implement a full-service IVHS program for the Charlotte, North Carolina, metropolitan region. The genesis of the effort, Congestion Avoidance and Reduction for Autos and Trucks (CARAT) is described as an outgrowth of concerns about increasing traffic congestion in several wedges of the metropolitan area. The paper then describes the steps taken to develop the CARAT concept, increase community support and involvement in its evolution, secure funding, evaluate alternative technologies, and conduct preliminary market assessments. The paper ends with a status report on CARAT at the present time.

INITIAL STEPS

To begin the process of reviewing IVHS technologies, the lead in the Charlotte metropolitan region was taken by the University of North Carolina at Charlotte. The university's goal was to encourage the community to evaluate carefully the potential for IVHS technology in the region and to reach a consensus concerning whether such technology should be explored and adopted. In March 1990, the university hosted a well-attended lecture by a General Motors executive on "Smart Vehicles." This was followed in April 1990 by a workshop attended by about 40 analysts to explore the opportunities further (*8*). In the summer of 1990, a lecture was given by Daimler-Benz (DB) executives working on the Prometheus project in conjunction with Freightliner Corporation, a local DB subsidiary.

The overall conclusion of these discussions was that Charlotte and the surrounding 100-mi region (Figure 1) have a number of attractive features to develop an IVHS demonstration:

- The road system is well suited to large IVHS demonstrations, particularly two crossing Interstates, multiple-path commuting corridors, and a basically radial street pattern. Travelers in the region cannot avoid congestion.
- Traffic congestion is significant and pervasive. The Charlotte region was initially projected by FHWA as the nation's No. 1 city for congestion in the year 2000 (*9*). Even though this forecast has since been discounted (*10*), congestion is still a problem. In particular, the southeast portion of the city has no crosstown circumferential routes; radial routes are congested and outmoded. These are under improvement, and the numerous parallel streets allow for route diversion tests during construction.
- A strong trucking and shipping presence ensures private-sector involvement. Charlotte is a major distribution center, and over 185 trucking companies, 2 major railroads, and numerous shippers and distributors have operations there.
- Numerous agencies and corporations are involved in preparing the proposal and are interested in the CARAT demonstration. These include communications developers and suppliers, computer companies, truck manufacturers and operators, chambers of commerce, electronics companies, and shippers.
- The government, particularly the city of Charlotte, Mecklenburg County, the North Carolina Department of Transportation (NCDOT), and the Carolinas Transportation Compact, is active and interested. The Carolina Transportation Compact is a 13-county organization of elected officials whose mission is to advocate the transportation needs of the region. The city of Charlotte has recently installed a modern traffic-signal coordination system that could form a key part of the in-road traffic detection system.
- Test sites are good, particularly the Charlotte Motor Speedway, the Independence Boulevard Corridor (a major congested arterial undergoing rehabilitation and reconstruction), several reversible lane sections, a number of integrated signal systems, and Interstates 85 and 77.
- The metropolitan region, in many ways a good microcosm of medium-sized U.S. cities, contains over 500,000 people in Charlotte proper, 1.6 million in a 50-mi radius, and over 5 million within 100 mi (Figure 1). The region is large, and development is spread quite thinly throughout the area, generating strong long-distance radial commute patterns overlaid on dense local traffic patterns.
- The area contains the University of North Carolina at Charlotte, a full-service institution with advanced degree programs in engineering, business, and numerous other technical subjects, and a high-tech industrial research park. The university's new Applied Research Center and its Ben Craig Business Incubator Center provide ideal homes for necessary research and development.

COMMUNITY SUPPORT

Early in the review of IVHS technology, the analysts recognized that a successful project would require strong cooperative effort from government, businesses, agencies, and the university. The university is not a manager, operator, or constructor of transportation systems. Therefore, responsibility for the development and implementation of CARAT would need to lie with some other organizational entity. The question of how to develop such an entity and encourage it to "own" the CARAT proposal was seen as critical to the project.

Therefore, CARAT was developed as a cooperative effort among government, businesses, the university, and others. Among the major actors are likely to be the city of Charlotte, Mecklenburg County, surrounding cities and counties, communications and electronics companies, truckers, vehicle manufacturers, shippers, the university, the state of North Carolina, other agencies, chambers of commerce, trade organizations, and, of course, appointed and elected officials.

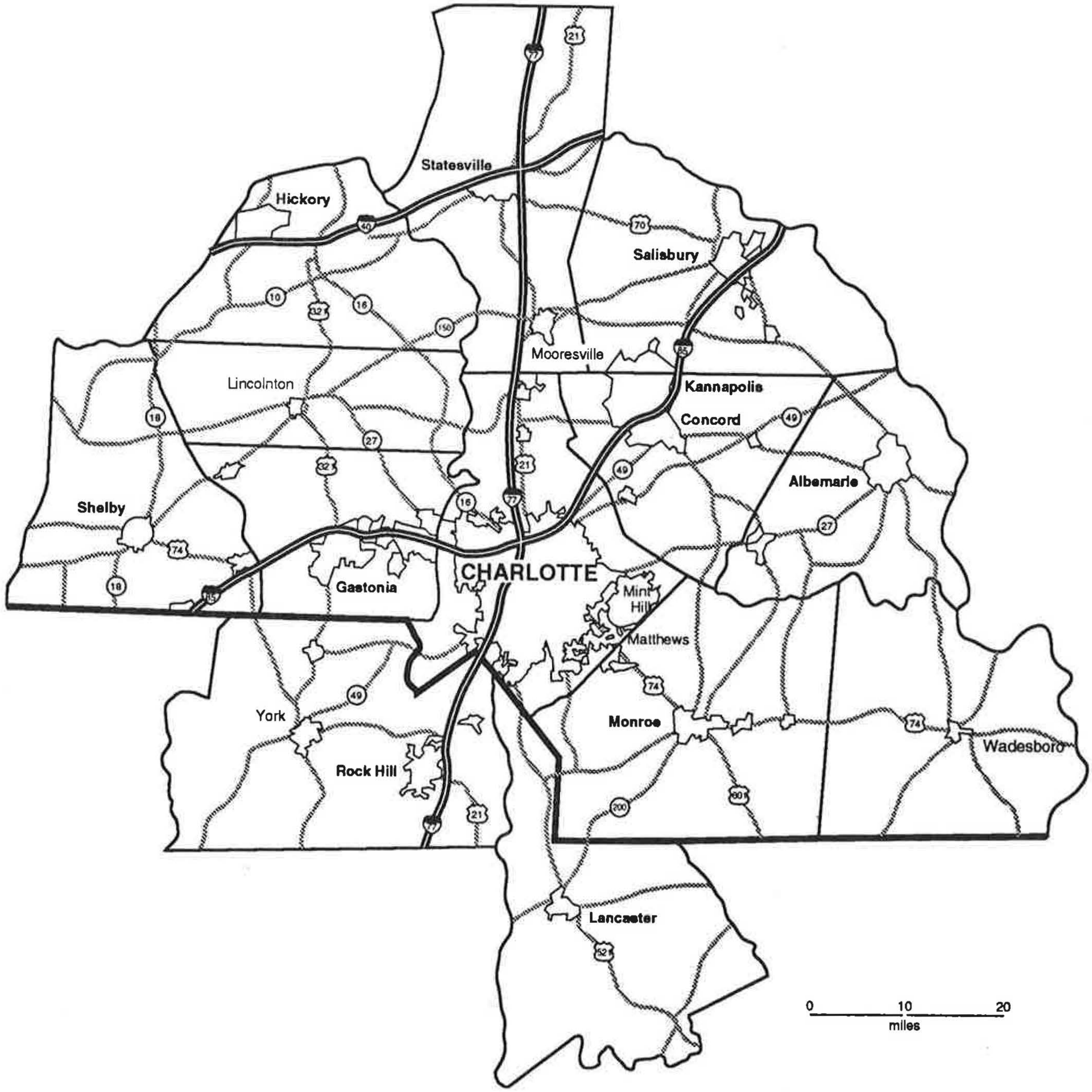

FIGURE 1 Charlotte, North Carolina, region.

Among the organization now regularly attending CARAT meetings are Alltel Mobile (mobile communications); American Automobile Assoc. (travel services); Ben Craig Center (University business incubator center); Carolina Freight (trucking); Charlotte Coliseum (fixed-site operator); Charlotte Convention and Visitors Bureau; Charlotte Department of Transportation; Charlotte/Douglas International Airport; Charlotte Emergency Medical Service; Charlotte Fire Department; Charlotte/Mecklenburg Chamber of Commerce; Charlotte Motor Speedway (race-test track); Charlotte Police Department; ESRI (GIS systems); Federal Highway Administration; Freightliner Corp. (truck manufacturer); Governor's Highway Safety Program; Harris Teeter, Inc. (large grocery chain); IBM; ITRE (state transportation research agency); Intergraph Corp. (GIS systems); JHK & Associates (consultants); Kimley-Horn (consultants); Moss Trucking Company (trucking); North Carolina Department of Transportation; North Carolina State Senate (elected officials); Parson Brinkerhoff (consultants); Sandoz Chemical Company; Scientex Corp. (consultant); Southern Bell Corp. (telecommunications); Southeastern Freight Corp. (trucking); UNC Charlotte, departments of Civil Engineering, Marketing, Geography, and Psychology; and UNC Charlotte Urban Institute (policy group).

To manage the development process, the university has organized a CARAT working group consisting of numerous organizations throughout the metropolitan region and other interested agencies and companies. The membership of the working group has been left open purposely with organizations being added on a steady basis. Initially the purpose of the working group was to assist in the development of an appropriate design for IVHS service for the metropolitan region. More important, the aim of the working group was to develop a community consensus on what the region needs and a sense of ownership by the region for the concept. Particular care was taken to include members of the media, other similar services and systems such as mobile phone companies and telecommunication organizations, and a number of elected and appointed officials. To encourage general awareness of the proposal, reports were developed for use in local newspapers, television, and radio. These reports emphasized the exploratory nature of the effort but encouraged the community to learn about "smart car" technology and how it might affect the region.

Particularly important participants in this study are the Charlotte Department of Transportation and NCDOT. Strong positive support and participation by these two organizations is viewed as absolutely critical to the ultimate success of the concept. In fact, the project will die if these organizations are not in strong support of the concept or are reticent to encourage its study. To develop interest in the project and an awareness of the study, analysts from the state and city transportation departments met with university organizers to express their interest in the effort and to suggest ways by which the project could be adapted to meet their concerns.

DESCRIPTION OF THE CARAT CONCEPT

Overview

The community support activities and learning process described permitted the development of a draft preproposal describing a suggested system for the Charlotte metropolitan region. The following description is for the ATIS portion of the CARAT concept. A parallel ATMS portion by NCDOT is now in the concept stage.

The ATIS portion of the system is envisioned as a multiagency operation—a sort of command center—that would gather data on the highway system from many sources on a continuing basis and package and provide that information to subscribers. Ideally the service would be self-sustaining financially, combining subscriptions with advertising and other revenues. Figure 2 shows the overall structure.

The central "command center" would gather and process information from a variety of sources on the status of the road network. The job of the command center would be to package the information for users and transmit it to them in real time. Key sources of information would include air surveillance via planes and helicopters; truck and delivery vehicle radio; selected field cameras; commuters, through cellular telephones; in-road traffic detectors; police, service, and emergency vehicles; and some fixed-site monitors, such as businesses and towers.

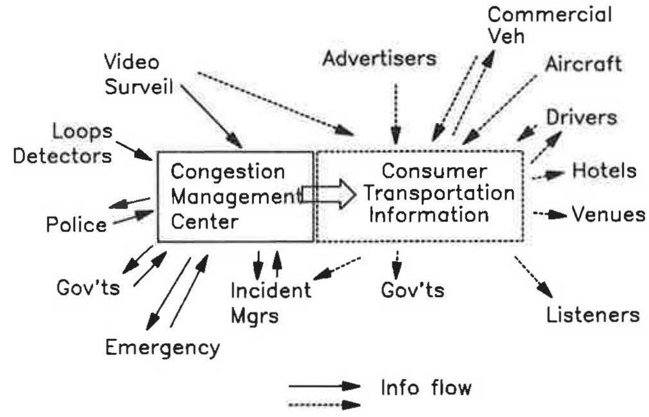

FIGURE 2 CARAT operation.

The information would come in to the command center in a variety of forms, including automated data retrieval, verbal messages, and visual signals. The information collected would consist of traffic and road situation information deemed of interest to the subscribers. An initial list would include traffic congestion at intersections, freeways, and arterials; accidents and tie-ups; fires, crimes, and other emergency events; major community or sports/concert events; average speeds; road construction and utility work; location of free-flowing routes; anticipated delay times; best route options; specialized advertising; and special services or sales by businesses.

As the data flowed in, they would be entered on regional digitized road maps in a form suitable for summarizing by area, "windowing" for transmission to moving vehicles in map or other form or "routing" data for use in congestion avoidance. A regional geographic information system (GIS) such as that being developed at UNC Charlotte would be used for information storage.

Packaged data would be delivered in a variety of ways, including home and office phones, cellular phones in cars, low-band subscriber radios in vehicles, two-way radios in vehicles, on-board maps or other digitized forms, subscriber and cable TV, closed-circuit TV, office announcement systems, drive-by electronic messages, and local-station network TV.

Users would include individual subscribers, trucking and delivery companies, shippers, police and other emergency service agencies, departments of transportation, the media, and distributors. Individual subscribers could receive messages prepackaged by route or location in whatever form is suitable for their needs. As an example, an individual subscriber might elect to purchase the "route congestion service" by telephone. The subscriber would provide the system with data on the normal or usual trip to and from work, including origin and destination, departure times, routes usually taken, and carpool names. Each morning, the subscriber would call the service and be recognized by an incoming number. The service would have prepackaged data for the subscriber in its files by recording all events on the subscriber's path. Information given would then consist of real-time monitoring of intersection delays and an "exception report" noting tie-ups on the route, incidents, other on-going events, and so forth. The data could be given in audio form or hard-copy readout, either at home, office, or on board the vehicle. Essentially, the system would provide a personalized route-planning service, updated continuously but activated only when contacted by the client. As technology advances, the telephone could be replaced by some kind of on-board map such as that developed for Pathfinder, but this is not necessary for initial operation.

As another example, over-the-road trucking companies might also supply the system with data from radio reports and events in and around the region, as far away as 100 mi. As trucks approached the region, the command center could supply directions on routing to avoid congestion, locations for fuel and other stops, and off-Interstate routes for local deliveries. Participating companies would be able to tailor the information to their own needs, rather than attempt to gather and report it on their own, as is common now.

For incident management, an emergency response company might subscribe to data on police and fire or crime calls by location, as well as best-route data from basing locations to the event site and the event site to area hospitals. Presently, some companies compile and make such data available in partial form; these companies might wish to participate in or provide some of the component services for CARAT.

Large fixed-site operators might use the system to direct patrons to use certain routes or park at selected sites for venues. Costs could be added to ticket prices.

The region could also serve as a test site for crash avoidance and other on-board vehicle control technology as it develops. The Charlotte Motor Speedway, a 1.8-mi NASCAR track with high-speed banked turns, might be used for vehicle control and guidance tests before road tests are tried. Since over one-half of all congestion is caused by incidents, many of which are accident-related (*11*), the ability to avoid accidents is a key element in the ability to reduce congestion.

System Benefits

Systems such as CARAT provide a significant opportunity to reduce traffic congestion in medium-sized cities. Among the probable benefits of a comprehensive congestion reduction program would be

- Reduced accident rates and savings in accident costs;
- Savings in travel time;
- Reductions in vehicle operating costs, fuel use, and air pollution;
- Increases in trucking productivity;
- Better, more effective use of the existing street system and a possible savings in future road construction;
- Increased attractiveness for jobs and living; and
- A higher overall regional quality of life.

The exact amount of benefits derived would be a function of adoption, usage, and the impact of a particular technology, but clearly general-use systems would have greater impact than limited-market services (*6,7*). With today's high road-construction costs, benefits such as these can quickly offset the relatively small cost of the study. Construction of 1 mi of freeway in an urban area can easily cost $20 million to $40 million; if the CARAT demonstration leads to a reduction in construction of just 1 mi, its costs will be recouped many times over.

Incremental Phasing

Because CARAT is a complex service, it is not possible to plan, design, and implement the systems in one phase. Rather, implementation is envisioned as taking place over several years. In the early years, work will concentrate on maximizing existing technology by implementing middle-tech solutions that can provide benefits in the short run. Examples include telephone- and radio-based communications systems, coordinated emergency services, and audio response and data gathering. Figure 3 shows an example of some systems that might be included in this first phase. In later phases, services will plan and integrate more advanced high-tech solutions such as tying into in-road traffic detectors, on-board digitized maps, automated vehicle communications, and crash avoidance systems. Most in-road detectors, at signals for instance, could be used to feed the command center on data about traffic density and congestion. Newer technologies under development are expected to improve this capability in the near future. Similarly, on-board maps are available but are essentially still experimental. These maps would be added to the system when their use is perfected.

Organizational Structure

In the start-up years, it is envisioned that CARAT would be developed and operated by a private nonprofit entity, perhaps structured as a "business incubator company." In its initial years the company would have organizational and structural support from the university. As the organization grows, a different, perhaps fully private, structure could evolve.

In summary, the CARAT concept is different from other IVHS concepts in a number of very important ways.

1. It proposes a comprehensive solution to the traffic issues of most metropolitan cities large and small, rather than a specialized solution for particular markets in just large cities.
2. It uses existing available middle-tech technology rather than futuristic high-tech solutions.
3. It proposes incremental development starting with most easily implemented technologies first, adding features later.
4. It relies heavily on the private sector to drive the process.
5. It proposes the use of advertising and subscriptions to offset costs.

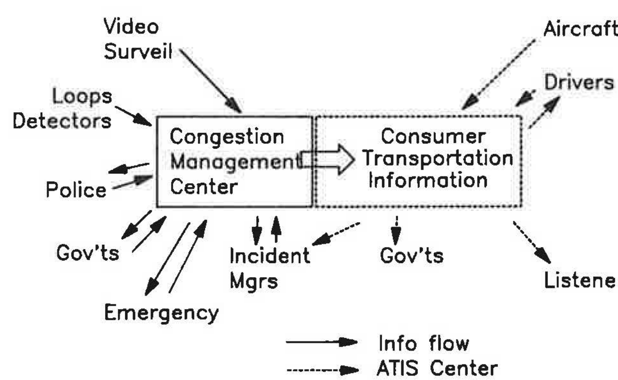

FIGURE 3 CARAT operation: initial phase.

6. It proposes the use of a multiagency command center that would manage the graphical information systems provided by the service.
7. It can be implemented at a reasonable cost.

Costs of Development

Estimates of the cost to develop and operate a CARAT system are presently in preparation. The multiyear program is proposed to ensure continuity and progressive building toward a full working system. Short-range systems planning will not produce the results needed for convincing demonstrations of applicability.

An important element of the evaluation is advertising. It is envisioned that services would be provided along with advertising within the communication channel as a means to increase the revenues to the managing agency and to encourage subscribers to obtain additional useful information about community services and products. It is expected that advertisers themselves would also contribute, thereby reducing the cost to subscribers. It is hoped that revenues through advertising and subscriptions will pay for the system, enabling it to be self-sufficient. For the immediate term, we have been modest in our expectations about the magnitude of potential revenues.

DEVELOPMENT OF THE FULL PROPOSAL

Structure

Efforts to develop the proposal necessary to describe the CARAT study in full detail are under way. NCDOT's involvement in the project has increased, particularly in the development of the parallel ATIS component of CARAT. The tasks involved in preparing the full proposal are as follows:

1. Organizational—the CARAT task force has been expanded and structured to target specific services intended to provide for the needs of different client groups.
2. Data gathering—the study team will prepare a review of IVHS technologies presently under way, assess the circumstances surrounding the Charlotte metropolitan region, and conduct a marketing study to evaluate the demand potential for IVHS technologies in the Charlotte area.
3. Component description—specific IVHS proposals will be described and evaluated extensively. Numerous elements of the proposal will be assessed and an integrated set of components will be developed.
4. Consolidation of elements—numerous IVHS technologies share common requirements. For instance, many technologies require the existence of an electronic or digitized mapping system describing the features of the urban region and highway network. Communication devices that transmit information to vehicles from a central point are another common element of many systems. The CARAT proposal will identify such common elements among the various systems and develop specifications for their use.
5. System description—the various elements of the proposed IVHS system for the Charlotte region, with all of its components, will be identified in detail. The description would include a discussion of how the elements work together, as well as organization structure, timing for implementation, and a plan for evaluation.
6. Preparation of proposal—the preceding materials will be organized into a formal proposal necessary for submission to appropriate organizations for potential funding.

Major Potential Markets

Working with its initial task force committee members, the CARAT study team has identified four markets likely to be most affected by IVHS technologies in the Charlotte metropolitan region:

1. Consumers, particularly commuters in selected high-density and high-congestion corridors in the region, are likely to be affected. There are approximately three such corridors, each of which contains several major commuter streets and freeways.
2. Charlotte is a major warehouse and distribution hub for the southeast and has many trucking organizations and distribution companies likely to be affected. Services provided to these companies might include routing, congestion, incident management, scheduling, delivery and pickup.
3. Several major arterials in the Charlotte region operate in a congested fashion on a continual basis. These arterials are particularly vulnerable to reduction in capacity by emergencies and other incidents.
4. Major hotel and motel operators, coliseums and amusement facilities, parks and lakes, shopping malls, and major employers all need information related to traffic congestion and routing in and around their locations.

Within each of these component areas, the study team has begun to evaluate what services are needed and how they might be organized and funded. The next step of the process will be to evaluate these services with the goal of selecting a smaller number of survivors, that is, the services which pass the feasibility (market versus cost) screen and can be financed.

Market Survey

A particularly important element of the CARAT study is the conduct of an accurate market survey to gauge the potential for different services in each of the four component areas. To undertake this effort, the university has designed a four-part market study concentrating on each of the preceding groups. In each group, representative members will be interviewed by telephone concerning the need for the service, its features, willingness to pay, and overall value compared with other services. The output of the market study analysis will be an estimate of the potential demand for different services at various price levels. Sampling rates for the consumer segment will be designed to yield about 500 completed market responses drawn from the metropolitan region, focusing on

Mecklenburg County and the city of Charlotte. For the other components, the universe of respondents is finite and generally less than 300 members. Therefore, a more extensive sampling effort will be undertaken with sampling rates somewhere in the 30 to 40 percent range. Preliminary market findings show that about 15 percent of consumers would use a general congestion avoidance system at a price of $12 per month.

Evaluation

The CARAT services, if implemented, would not necessarily lead to significant changes in travel patterns or in consumer acceptance. To evaluate the overall effectiveness of this service, it is anticipated that a carefully structured statistical design will be used. The initial design will be a before-after study with a test group and control group. These designs are common in the social services (*12*) and take the following form:

$$\frac{\text{test } O_1\ XO_2}{\text{control } O_3\ O_4}$$

where O's are "observations" of behavior, and X is the "treatment," in this case the IVHS service. Tests for X's effect are made by determining the significance of the "difference in the differences," for example:

$$t = \frac{(\overline{O}_1 - O_2) - (\overline{O}_3 - O_4)}{\frac{1}{n}\sqrt{S_1^2 + S_2^2 + S_3^2 + S_4^2}}$$

A number of studies in transportation (*13*) have used these designs to isolate the effect of targeted services.

In this manner, both the internal (intraservice) and external validity of the treatment can be estimated. Internal evaluations using such data as number of inquiries, number of consumers and subscribers, revenues, costs, and operational features will also be maintained, but these alone will not allow for a complete evaluation of the external effectiveness of the system. An analogy may be drawn to the carpool programs in common use around the country. These programs have extensive records on internal activities, such as telephone calls received, inquires made, list provided, and so forth, but very few of them have hard data on the actual number of carpools formed or the effect of those carpools on regional vehicle miles traveled. A quasi-experimental design that would permit the evaluation of this type of program is extremely rare in transportation studies; they need to be developed more extensively.

NEXT STEPS

The full CARAT proposal is expected to be prepared during spring 1992. It will detail additional issues that must be resolved before the project can be undertaken, including

1. Organizational structure,
2. Specifics on kinds of services to be offered,
3. Revenues and costs,
4. Financing,
5. Timing,
6. Public/private partnership and mix, and
7. Implementation plan.

In summary, initial efforts are well under way to develop the CARAT proposal as the first major effort to implement a general-use IVHS system in a medium-sized metropolitan region.

REFERENCES

1. *Discussion Paper on Intelligent Vehicle-Highway Systems*, Department of Transportation, May 1989.
2. Radar Control Systems. *Vehicle Radar* (videotape). July 1990.
3. J. F. Gabard. CALIF: An Adaptive Strategy. *INRETS Review*, Dec. 1990.
4. Operation Tests and Research Activities. *Intelligent Vehicle Highway System (IVHS) Projects in the United States*, FHWA, March 1991.
5. Senate Bill S. 965. IVHS language. May 1991.
6. U.S. General Accounting Office, *Smart Highways: An Assessment of the Potential To Improve Travel*. May 1991.
7. E. Bruce and D. Hartgen. IVHS: How Much Can It Really Help? Presented at Symposium on Vehicle Information Systems, Ann Arbor, Mich., Oct. 1991.
8. *Smart Vehicles and Roads for Charlotte*. Report 38, Transportation Academy, UNC Charlotte, April 1991.
9. J. A. Lindley. Urban Freeway Congestion: Qualification of the Problem and Effectiveness of Potential Solutions. *ITE Journal*, Jan. 1987, pp. 27–32.
10. T. Lomax et al. *Impact of Declining Mobility in Major, Texas and Other U.S. Cities*. Report 431-1F. Texas Transportation Institute, College Station, Tex., 1988.
11. U.S. General Accounting Office. *Traffic Congestion: Trends, Measures, and Effects*. 1989.
12. D. T. Campbell and J. C. Stanely. *Experimental and Quasi Experimental Designs for Research*. Rand McNally, 1963.
13. J. Brusno and D. T. Hartgen. Statistical Controls in Ridesharing Demonstration Programs. In *Transportation Research Record 914*, TRB, National Research Council, Washington, D.C., 1982.

Publication of this paper sponsored by Committee on Transportation System Management.

FREEVU: A Computerized Freeway Traffic Analysis Tool

B. R. HELLINGA AND J. H. SHORTREED

FREEVU (Freeway Evaluation with Visual Understanding) is a personal computer simulation model intended for freeway design and analysis. It allows the user to specify a freeway section, including lanes, grades, exits and entrances, posted speed limits, and detector locations. The section can then be viewed to confirm the proposed design. A variety of traffic situations can be specified, including percentage trucks, distribution of car and driver characteristics, and entrance and exit percentages. The user can simulate the traffic situation for various freeway design alternatives and then evaluate the design through two methods. First, the information from the specified detector locations can be used to evaluate average volumes, speeds, and densities over time. Second, an animation of the simulation results can be viewed to evaluate weaving sections, stability of traffic flow, impacts of trucks, and so forth. The model is a descendent of the simulation models INTRAS and FOMIS. As in these two models, vehicle movement is based on classic car-following theory and collision-avoidance restrictions. However, FREEVU also incorporates behavioral lane changing algorithms and vehicle performance constraints. FREEVU is user-friendly with extensive menus and default values. The simulation model has been evaluated using different sites and was generally found to represent simulation traffic flow accurately.

Freeway systems have typically been the major infrastructure element used to meet traffic demands in large urban centers. However these freeway systems are becoming more heavily congested for longer periods of the day. Engineers and designers require more-sophisticated tools to help them analyze and evaluate freeway segments and understand the dynamics of traffic flow on these segments.

The *1985 Highway Capacity Manual* (HCM) is widely used for design and analysis. However, because the HCM is based primarily on aggregated empirical results, it often lacks the ability to provide an understanding of the dynamic nature of the traffic flow.

Simulation models can be used to provide additional understanding. FREEVU was developed as a first attempt at providing engineers and designers with such a simulation tool.

The simulation model FREEVU is presented here. Four questions are presented and answered: What is FREEVU? What is the simulation logic basis? How can FREEVU be used? How well does FREEVU perform?

WHAT IS FREEVU?

FREEVU (pronounced "free view") is a stochastic, microscopic, freeway traffic simulation program, for use on a personal computer. It stands for Freeway Evaluation with Visual Understanding. It combines a user-friendly interface with a simulation core to produce an effective freeway traffic analysis tool.

Data inputs are minimal. Data entry is facilitated by a menu system and on-screen input forms. Error checking is carried out on data input. High-resolution graphics are used to display the freeway section as well as portray simulation results in the form of a movie, with individual vehicles depicted as they traverse the freeway section.

Model Capabilities

Specifically, FREEVU can model the following freeway components:

- Unidirectional, multilane freeway segments of two to eight lanes in width;
- Lane adds and drops;
- On-ramps—single lane from either the right or the left side of the freeway;
- Off-ramps—single and multiple lanes from either the left or the right side of the freeway;
- Posted speed limit or other speed restrictions; and
- Vertical alignment with the ability to specify unique grades for individual lanes or ramps.

FREEVU does not explicitly model all factors affecting traffic flow (i.e., lane width, horizontal curvature, passing sight distance, weather, road surface conditions, incidents, and rubber-necking); however, these factors tend to inhibit traffic speed, so many of these effects can be represented in the simulation by specifying a reduced speed limit for the affected lane and section. In this manner, speed can act as a surrogate means for simulating these other effects.

Because of computer hardware and software limitations, the restrictions presented in Table 1 have currently been selected for FREEVU.

History of Development

FREEVU's simulation core is a descendent of the INTRAS model (*1*). The INTRAS model itself is a stochastic, microscopic model created primarily for studying freeway incidents (*2*). Developed in 1975, the model was designed to represent traffic and traffic control elements in a freeway and surrounding surface street environment.

B. R. Hellinga, Department of Civil Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6. J. H. Shortreed, Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada N2I 3G1.

TABLE 1 Current Restrictions Applicable to FREEVU

Item	Maximum Permitted
Section length	10 km
Traffic lanes	8
Entrances, including on ramps	20
Exits	20
Detector Locations	20
Geometric segments	15
Speed limit zones	20
Vertical alignment segments	20
Different vehicle types	100
Simultaneous vehicles	2500

INTRAS has been widely applied, with reported uses ranging from freeway reconstruction design evaluation (*3–5*), conflict analysis for weaving areas (*6*), energy conservation studies (*7*) and as a benchmark for validation of other models (*8,9*).

However, according to Van Aerde et al. (*10*), users of INTRAS have reported problems with some aspects of traffic behavior such as merging behavior and vehicle behavior at off-ramps (*11*).

In response to some of these criticisms and as an attempt to provide the unique capabilities of INTRAS in a more compact and structured form, FOMIS was developed (*11*). FOMIS restricts the simulation process to the freeway only, eliminates the link structure of INTRAS, and reduces vehicle processing to a single scan.

In the course of a study of the impact of large trucks on traffic flow, a microscopic simulation model was required. FOMIS was evaluated, in light of the study's requirements, and found to be lacking in the following three areas:

- The model was not simple to understand.
- The model was difficult to use.
- The model did not simulate many important components of the freeway system, including grades, trucks, driver's lane-changing decision-making process, speed limits, and truck restricted lanes.

FREEVU consists of four distinct program modules: traffic performance, freeway specification, data translation, and simulation (Figure 1). The first three modules are part of the integrated design shell; the fourth module is the simulator.

Integrated Design Shell

The user can directly interact with two modules in the design shell, freeway specification and traffic performance. The third interface module, which is invisible to the user, translates the data input by the user into the correct format required by the simulator.

The program's interface structure is constructed around the common menu. It is thought that this approach provides a simple, familiar, easy-to-understand appearance to the user with the minimum of complex program code. The structure consists of a menu tree; each menu presents the user with a number of possible alternatives.

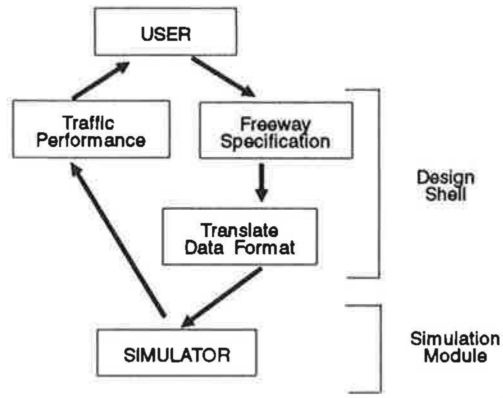

FIGURE 1 Modular structure of FREEVU.

Data input is highly structured and controlled by internal checking routines. The user is informed of the data required and only permitted to enter data of the specified type (i.e., integer or alphabetic) and within specific ranges dependent on previous input. Speed of input is facilitated by default values given when possible, allowing the user to simply accept the values and move on to the next input cell. Logical errors in the user's definition of the freeway section are checked. If found, the user is informed and given the opportunity to modify the data.

Simulation Module

The freeway segment is structured as a single continuous unit, with elements (i.e., vehicles or fixed objects) located by their distance from the upstream boundary and the lane number. Fixed objects are used to define the geometry and characteristics of the section, including lane adds and drops, on- and off-ramps, weaving sections, speed limits, grades, and detector locations.

Vehicles are processed in order of their location, regardless of lane, from downstream to upstream. A single pass is made for each time interval, during which each vehicle is processed.

WHAT IS THE SIMULATION LOGIC BASIS?

Within FREEVU, vehicle movements are governed by fou controlling elements:

- Classic car-following theory,
- Collision-avoidance restrictions,
- Behavioral lane-changing algorithms, and
- Vehicle performance.

The first two elements are unchanged from INTRAS. However, the use of extensive algorithms to represent drivers' lane-change decision making is an innovation. A decision algorithm is hypothesized and subsequently tested using observed data, as is the variation in vehicle performance on grades.

Car Following

The car-following algorithms are essentially unchanged from FOMIS and INTRAS. A full derivation of the car-following algorithm is available in the literature (*1*).

The car-following model is built on the underlying assumption that drivers will attempt to maintain their spatial headway as a function of the driver's characteristics, the speed of the leading and following vehicle, and the length of the leading vehicle. A driver will modify his acceleration in order to maintain this desired spatial headway.

Collision Avoidance

In addition to the basic car-following relationship, an emergency constraint exists that overrides the basic car-following algorithm in order to prevent collisions. Again, this constraint is the same as that developed for INTRAS.

Behavioral Lane Changes

It was observed from data gathered from the Queen Elizabeth Way near Toronto, Ontario, and from general observance of freeway traffic that significant numbers of lane changes occur for reasons other than origin and destination requirements and that these lane changes are not random, as was previously assumed.

Both INTRAS and FOMIS represent these lane changes by simply specifying a constant probability that any vehicle will at any time make a lane change. It was believed that this did not adequately represent reality, and because no algorithms could be found in the literature, lane-changing algorithms were hypothesized, implemented into the model, and later tested using observed data. It now appears that similar algorithms were simultaneously developed for the FRESIM model (*12*).

On the basis of the assumption that many driver actions are governed by a driver's self-interest, it was hypothesized that there exist two types of behavioral lane changes:

- Passing to increase speed, and
- Yielding to following traffic.

Passing

Passing is predominately governed by a driver's self interest—that of maintaining or increasing speed. The process of a driver deciding to make a passing maneuver was divided into four phases.

- Driver becomes dissatisfied,
- Driver evaluates alternatives,
- Driver decides whether the improvement is significant enough to warrant the maneuver, and
- Driver attempts to change lanes.

Phase 1 To determine if a driver will attempt a lane change, it must first be established that the driver is unhappy with the present state. A driver is assumed to be dissatisfied with the present state if the vehicle is not accelerating and is traveling at a speed less than the desired speed. The assumption is that if the driver is not traveling at his or her desired speed and is not accelerating, then the driver is being impeded by some other downstream vehicle and may be able to improve the situation by making a lane change. Because a driver traveling slower than desired may be undergoing a small acceleration and still not be satisfied, a threshold acceleration value of 0.3 m/sec^2 (1 ft/sec^2) was chosen. Any driver accelerating at greater than this value is assumed to be satisfied with the current state and will not consider a passing maneuver.

Because it is not realistic for every dissatisfied driver to consider passing, some measure of the driver's dissatisfaction or frustration is required.

It was hypothesized that two factors affect a driver's frustration: the vehicle's acceleration potential and the difference between the driver's desired free speed and the vehicle's current speed. The frustration index (FI) reflecting the driver's dissatisfaction is the product of the two factors.

It was assumed that the probability that a driver is frustrated enough to consider passing, during any scanning interval, is directly proportional to FI. However, as each vehicle is scanned each second, a maximum probability of 40 percent was imposed to reflect the fact that drivers are not likely to make passing decisions every second (Equation 1).

$$P[\text{frustrated}] = \begin{cases} \text{FI} & \text{if FI} \leq 40 \\ 40 \text{ percent} & \text{otherwise} \end{cases} \tag{1}$$

Phase 2 If the driver is frustrated enough to consider passing, a more favorable state must exist, or else no lane change is warranted. A more favorable state is one that gives the driver the longest period of unimpeded travel. This is measured by the time required for the driver to overtake the next downstream vehicle in the lane being considered. This time is a function of the speeds of the two vehicles, their respective positions, length, and accelerations. Because it is assumed that acceleration is constant over the time interval, simple equations of motion can be used to describe the vehicles' movements.

Phase 3 If the current state is found to be more favorable than the available alternatives, no lane change is made. However, if the perceived improvement of either adjacent lane is greater than that of the current state, a lane change may be attempted. It is assumed that there must be a significant advantage to making the lane change. A value of 13 sec was chosen as an initial threshold value.

Having evaluated the alternatives, the driver decides which one provides the greatest improvement. It is assumed that drivers tend to prefer passing on the left to passing on the right. Therefore, an improvement of five times is required for lane changes to the right over the left.

Having decided on a potential alternative, there is a probability that a driver will choose to change lanes. This probability is based on the amount of improvement over the current state the driver will receive. Equations 2 and 3 provide the relationship for passing to the left and right, respectively.

The parameter values in these equations have been assumed as initial estimates. Note that a maximum probability of 40 percent was again imposed.

$$P[\text{left}] = \min\begin{cases} 0.15(T_L - T - 13) \\ 40 \text{ percent} \end{cases} \quad (2)$$

$$P[\text{right}] = \min\begin{cases} 0.73(T_R - T - 13) \\ 40 \text{ percent} \end{cases} \quad (3)$$

where

T_L = left-lane unimpeded travel time,
T_R = right-lane unimpeded travel time, and
T = current-lane unimpeded travel time.

Phase 4 For a lane change to occur, it must be physically possible for the vehicle to complete the lane change (i.e., a sufficient gap must exist). This criterion is checked in the same manner as for all lane changes.

Yielding

Yielding maneuvers, changing lanes to the right to benefit the following vehicle, are governed more by courteousness to other drivers than by self-interest. Yielding maneuvers can occur in two situations: if a vehicle cannot maintain its speed on a grade and if the driver simply desires to travel more slowly than other vehicles. In either case, for a driver to consider yielding, the driver's vehicle must be impeding the following vehicle.

$$P[\text{yield}] = \min\begin{cases} 1.6[A_{\max}(S_c - S_d)] \\ 80 \text{ percent} \end{cases} \quad (4)$$

where

$A_{\max}$ = maximum possible acceleration,
S_c = current speed, and
S_d = desired speed.

For the second yielding situation, in which a vehicle desires to travel much slower than other vehicles, there is a 5 percent probability that it will yield during any simulation interval.

The vehicle considering a yielding maneuver will only attempt the maneuver if it will not be impeded by a downstream vehicle after the lane change is made.

These algorithms and probability functions were hypothesized and then implemented in the model. During model evaluation, a quantitative analysis of the hypotheses was made and found to be reasonable for the cases studied.

Vehicle Performance

The last element that controls vehicle movement is vehicle performance. Up to 100 vehicle types can be defined for use in FREEVU. In defining these types, the vehicle's horsepower, gross vehicle mass, and frontal area are required, because these characteristics dictate vehicle performance.

Each second, the maximum possible acceleration a vehicle can undergo is computed using the concept of tractive effort and tractive resistance. Tractive resistance is the sum of the grade, rolling, and drag resistances. Tractive effort is dependent on the vehicle's power and current speed. The equations and coefficients required for these computations are available in the literature (*13*). This process does not incorporate momentum effects.

The performance of most cars today is not significantly affected, except on very steep grades. However, heavy truck configurations regularly experience substantial degradation in performance on even moderate grades. It is primarily for the realistic modeling of trucks that vehicle performance has been addressed in FREEVU at this level of detail.

HOW CAN FREEVU BE USED?

FREEVU is intended to complement, not replace, existing methods of evaluation. FREEVU is intended to provide the engineer with qualitative and quantitative results regarding a proposed alternative. This alternative can take the form of a geometric improvement (i.e., a truck climbing lane), an expected increase in traffic demand, an anticipated change in traffic composition (i.e., more heavy trucks), or the implementation of a new policy (i.e., truck restricted lanes).

Input Data Requirements

The user is required to input data on section geometry, detector locations, vertical alignment, and posted speed limits for each freeway section.

Geometry

FREEVU represents freeway sections linearly, with all positional data taken as the distance, in meters, along the centerline from the upstream boundary. The upstream boundary is the upstream end of the simulated section at which vehicles are generated.

To illustrate, consider, in Figure 2, the freeway section to be modeled. Figure 2 also shows this section represented linearly as it appears on the computer screen. The user inputs the total length of the section to be modeled and the maximum

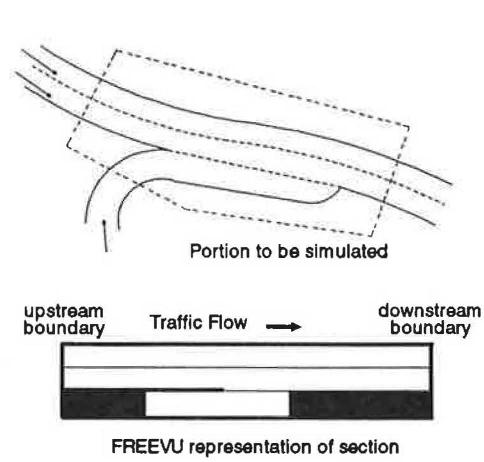

FIGURE 2 Freeway segment and portion to be modeled.

number of lanes. Having defined the area of the section, the lane type must be defined over each lane's total length. This is accomplished for each lane, by defining a number of segments of lane homogeneous in lane type. Both the upstream and downstream end of each segment must be specified. Each segment may contain only one lane type. Permissible lane types are presented in Table 2. For example, in Figure 2, the middle lane has three segments; a left lane, a center lane and a right lane. Permissible physical features are also given in Table 2.

Detectors

The detectors provided in FREEVU represent paired induction loop detectors. Defining a detector location places detector loops in all traffic lanes across the freeway section at that point.

Vertical Grades

In defining vertical grades, segments of the freeway having a consistent grade are defined. Each lane of the freeway may have a unique integer grade defined for each of the defined segments.

Speed Limits

The primary purpose for allowing the user to define speed limits is to enable the model to more realistically reflect the effects of reduced speeds of vehicles on ramp sections. However, speed limits can also be used to reflect other factors that affect speed but can not be explicitly modeled using FREEVU. For example, if a section of the freeway has very poor pavement surface conditions, it may be desirable to designate this area as having a posted speed limit of 90 km/hr to reflect the effect the poor surface has on speed.

Figure 3 shows how the different type of lane segments are independent of each other and are layered by FREEVU to define the characteristics of the freeway section.

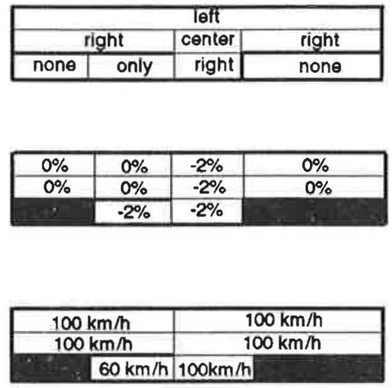

FIGURE 3 Segment types used to define a freeway section: *top*, geometric; *middle*, vertical alignment; *bottom*, speed limit.

Traffic Performance

Graphical Output

The on-screen output consists of a movie of the freeway that can be viewed. Detector information is displayed at the top of the screen. Average 30-sec volume, speed, or density is displayed for each lane at each detector location.

This movie feature allows for instant visual feedback to the designer about the microscopic level of interaction occurring on the freeway. For example, at the development level, this feature was invaluable for debugging of the program, reducing the time required by an estimated 80 percent. During the movie animation, the user can pause the display, advance a single frame at a time, change the displayed detector information, and speed up or slow down the movie.

This display mode is extremely useful to gain an immediate understanding of how well the freeway is operating. This feature prompted the name FREEVU, Freeway Evaluation with Visual Understanding. Visually, one can immediately identify queues forming, platoons existing behind slow trucks, effects of lane changes, and the amount of disruption in the vicinity of merging areas.

Numerical Output

Currently four output files can be generated:

- Standard detector information—speed, volume, and density;
- Record of the number and type of lane changes that occurred during the simulation;
- Record of the number and type of maneuvers by vehicle class; and
- The average travel times, in seconds, for each origin-destination pair for cars and trucks; the standard deviation and number of vehicles observed is also given.

It is expected that as design experience with FREEVU is obtained, the output will be refined.

TABLE 2 Permissible Physical Features and Lane Types

Physical Features	Lane Types
Lane Drop	Center
Lane Add	Right
Off Ramp End	Left
On Ramp Start	Only
System Start	None
Lane Type Change	
Bull Nose	

HOW WELL DOES FREEVU PERFORM?

Scope of Evaluation

In addition to validations conducted during development of FREEVU, the ancestor programs, INTRAS and FOMIS, have been validated and evaluated by a number of different users

(*1,3,11,14–16*). These validations of FREEVU's ancestors can be referred to as the first level of validation of the model. Care has been taken not to change parameters unless there was considerable evidence from two or more sites, to support new values.

Initial modifications were made to FOMIS to permit the modeling of trucks on grades. This enhanced FOMIS version was evaluated using data from the Queen Elizabeth Way (QEW) on the Burlington Skyway in Hamilton, Ontario. The model was found to perform reasonably well, but areas for improvement were identified. These improvements were made and can be considered to be second-level enhancements. They include the interface design shell and the behavioral lane-changing algorithms. This new model, FREEVU, was evaluated and validated using a data base obtained from FHWA and data from the QEW in Mississauga, Ontario. Conclusions and recommendations from this evaluation form the basis for recommended future third-level enhancements.

A more detailed reporting of these evaluation results can be found in the literature (*17*).

Validation Results

Data from the QEW eastbound between Highways 403 and 427 were obtained from the Ministry of Transportation of Ontario (MTO). A section of the QEW eastbound, corresponding to one of the loop detector stations, was videotaped during the morning peak period.

The FHWA data set (*18*) consists of digitized aerial photography taken at different sites across the United States. Each site had been filmed for approximately 1 hr. From the film, complete vehicle trajectories were produced, recording the vehicle's position for each time interval. Data from four sites describing different geometric configurations were selected (Table 3).

TABLE 3 FHWA Sections Used for Validation

Site Number	Location	Section Type
2	I-95 S.B. at Backlick Road (Route 617), Fairfax County, Virginia	Ramp
3	I-395 S.B. (Shirley Highway) at Duke Street (Route 236), Alexandria, Virginia	Off-Ramp
4	I-405 N.B. at Mulholland Drive, Los Angeles, California	Tangent
5	I-405 S.B. at Santa Monica Blvd., Los Angeles, California	On-Ramp

Mesoscopic—Lane-Changing Rates

Normalized nonmandatory lane-change rates were determined from the data for each of the five sections. Mandatory lane changes are those that have a ramp lane as either the origin or destination lane during a lane change.

Data had been assembled from five separate sites, four from the FHWA data, and the QEW section in Mississauga. Each site was simulated using FREEVU. Lane-change data were recorded. Normalized lane-change rates for nonmandatory lane changes were calculated and compared with those presented in Table 4.

Total simulation lane-change rates compared well with those observed; the discrepancies were within 14.3 percent, except for the QEW section and FHWA Site 3. FHWA Site 3 had a major interchange about 300 m downstream of the site. This interchange affected lane-changing maneuvers, but it was not known in what way. As such, it was difficult to determine the nonmandatory lane-change rate with accuracy.

For the QEW section, the distance over which lane changes were recorded was estimated to be 200 m, based on an observed queue of 20 cars at near-jam density. However, it is possible that this distance was underestimated and that the distance per vehicle was in the range of 15 to 17 m. This would result in the lane-change rate error of only −6.1 percent.

The lane-changing algorithms implemented in the model had been hypothesized. The required parameters had been selected intuitively and subjectively before any validation. From the results presented in Table 4, the hypothesis regarding drivers' lane-changing decision process is reasonable, and the parameters chosen, appropriate, given the available data.

Macroscopic—Section Flow and Speed

To more fully evaluate FREEVU, a macroscopic analysis of the FHWA data sets was conducted. From results presented in Table 5, FREEVU reproduces observed section flows within 13.7 percent of those observed. FREEVU produces weighted average section speeds that are within 16.5 percent of the observed weighted speeds.

Qualitative Evaluation

In addition to the model features evaluated previously, the distribution of total volume across lanes, merging behavior,

TABLE 4 Comparison of Simulated and Observed Nonmandatory Lane Changes Per Vehicle-km

Site number	Observed			Simulated			Error (%)
	Left	Right	Total	Left	Right	Total	
QEW	0.091	0.048	0.139	0.015	0.062	0.077	-44.6
2	0.217	0.135	0.352	0.285	0.091	0.376	6.8
3	0.082	0.140	0.222	0.195	0.254	0.449	102.3
4	0.186	0.234	0.420	0.307	0.173	0.480	14.3
5	0.386	0.098	0.484	0.416	0.094	0.510	5.4

TABLE 5 Comparison of Simulated and Observed Section Flows and Speeds

Site #	Number of Lanes	Average Section Flow (vph)			Average Section Speed (km/h)[a]		
		Simulated	Observed	Error (%)	Simulated	Observed	Error (%)
2	3	4,581	5,115	10.4	60.3	57.4	-5.1
3	3	5,066	5,824	13.0	63.6	66.3	4.0
4	5	7,634	7,357	-4.5	65.8	78.8	16.5
5	4	5,884	6,820	13.7	50.6	56.4	9.7

[a] Speed is average lane speed weighted by volume

TABLE 6 Qualitative Evaluation of Simulation Features

Evaluation Feature	Validation Site				
		FHWA			
	QEW	2	3	4	5
Lane Distribution	P	F/P	F	F	F
Merging Behavior	F	F	n.a.	n.a.	F
Breakdown Process	G	G	n.a.	n.a.	G

G = Good; F = Fair; P = Poor; n.a. = Not Applicable

and the flow breakdown process were also evaluated. A qualitative assessment was made and is summarized in Table 6.

The validation exercises undertaken in this work have been primarily based on nonaverage traffic conditions. All of the FHWA sites were specifically chosen for their high levels of congestion and poor operating conditions. Thus, the evaluation of FREEVU was based on demanding situations, well beyond the scope of existing average methods of analysis.

Observations on Driver Behavior

During the course of the evaluation exercise, the observed data files were converted into a format such that they could be displayed graphically using FREEVU's interface. This permitted the unique opportunity to see visually what the numerical data files represented.

It appears that drivers have the ability to anticipate downstream traffic conditions and respond to observed downstream events. FREEVU determines each vehicle's actions on the basis of the next downstream vehicle in the current lane. This results in a traffic stream that behaves in a nature more reactive than anticipatory.

Driver behavior varies significantly between commuter and noncommuter traffic streams: commuter streams seem to be more aggressive and homogeneous in nature than noncommuter streams. At times, drivers appear to be insensitive to short separation distances and accept, at least temporarily, unsafe following distances.

CONCLUSIONS

FREEVU is a first attempt at a freeway analysis tool that goes beyond the capabilities of existing methods to meet designers' needs.

FREEVU is very designer-friendly and represents the mechanics and dynamics of traffic flow reasonably well. The ability of FREEVU to provide a view of the dynamic behavior of the traffic flow greatly improves understanding of traffic flow and the traditional measures of effectiveness.

The incorporation in a simulation program of extensive lane-changing decision algorithms based on drivers' self-interest is a unique innovation.

Evaluations of FREEVU's simulation capabilities at the mesoscopic level indicate that nonmandatory lane-change frequencies are within 15 percent of those observed for three of the five sites investigated. It is concluded that the hypothesized lane-changing algorithm can be accepted as realistically reflecting a driver's decision process. However, it is recognized that as further work is carried out, the parameters used in the lane-changing algorithms may require further calibration.

FREEVU simulated many traffic situations well; however, it had difficulties in realistically simulating the actions of merging and, to some extent, diverging vehicles under some moderate- to high-volume conditions. However, it was found that these situations also vary widely in reality.

The implementation of a driver's desired free speed based on the posted speed limit was useful. This permitted the control of overall speeds, particularly for ramp sections.

The aspects of FREEVU that appear to inadequately reflect reality—such as merging and diverging behavior, drivers' anticipation of downstream events, and vehicle's traveling at unsafe following distances—all require extensive, accurate microscopic data bases in order to understand each event fully.

A number of potential third-level enhancements can be identified:

- Enhance merging behavior at high volumes;
- Improve modeling of vehicles accessing off-ramps;
- Incorporate concepts of driver anticipation;
- Investigate how and if the distribution of driver sensitivity to desired distance headway changes over time; and
- Determine if effects of momentum should be modeled.

ACKNOWLEDGMENTS

This work was carried out under contract with MTO. The authors would like to thank Alex Ugge of the Research and Development Branch and Rye Case of the Transportation and Energy Branch for their contributions.

REFERENCES

1. D. A. Wicks and E. B. Lieberman. *Development and Testing of INTRAS, a Microscopic Freeway Simulation Model*, Vol. 1–4. Final Report. FHWA, U.S. Department of Transportation, Oct. 1980.
2. A. D. May. Models for Freeway Corridor Analysis. In *Special Report 194: The Application of Traffic Simulation*. TRB, National Research Council, Washington, D.C., 1981.

3. A. M. Skabardonis, A. D. May, and S. Cohen. Application of Simulation To Evaluate the Operation of Major Freeway Weaving Sections. In *Transportation Research Record 1225*, TRB, National Research Council, Washington, D.C., 1989, pp. 91–98.
4. S. L. Cohen. Application of Traffic Simulation to Analysis of Freeway Reconstruction Alternatives—A Case Study. *Proc., Engineering Foundation Conference* (S. Yagar, ed.), New England College, Henniker, N.H., June 1987.
5. S. L. Cohen and J. Clark. Analysis of Freeway Reconstruction Alternatives Using Traffic Simulation. In *Transportation Research Record 1132*, TRB, National Research Council, Washington, D.C., 1987.
6. J. Fazio and N. M. Rouphail. Conflict Simulation in INTRAS: Application to Weaving Area Capacity Analysis. In *Transportation Research Record 1287*, TRB, National Research Council, Washington, D.C., 1990.
7. J. F. Torres. *Freeway Control and Management for Energy Conservation.* Final Report. JFT Associates, Los Angeles, Calif., Jan. 1982.
8. A. K. Rathi and Z. J. Nemeth. FREESIM: A Microscopic Simulation Model of Freeway Lane Closures. In *Transportation Research Record 1091*, TRB, National Research Council, Washington, D.C., 1986.
9. P. G. Michalopoulos and J. Lin. Integrated Modeling of Freeway Flow and Application to Microcomputers. In *Transportation Research Record 1091*, TRB, National Research Council, Washington, D.C., 1986.
10. M. Van Aerde, S. Yagar, A. Ugge, and E. R. Case. A Review of Candidate Freeway-Arterial Corridor Traffic Models. In *Transportation Research Record 1132*, TRB, National Research Council, Washington, D.C., 1987.
11. A. G. R. Bullen. Development of Compact Microsimulation for Analyzing Freeway Operations and Design. In *Transportation Research Record 841*, TRB, National Research Council, Washington, D.C., 1982.
12. A. Halati, J. F. Torres, and B. Mikhalkin. *Freeway Simulation Model Enhancement and Integration—FRESIM Technical Report.* FHWA, U.S. Department of Transportation, 1990.
13. W. S. Homburger, L. E. Keefer, and W. R. McGrath. *Transportation and Traffic Engineering Handbook*, 2nd ed. ITE. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1982, pp. 162–169.
14. P. Waurick. *Evaluation of the INTRAS Microscopic Freeway Simulation Model Using P.M. Peak Congestions on Q.E.W. Westbound at Hwy 427.* Internal Report. Traffic Systems Research, Research and Development Branch, Ministry of Transportation and Communications, Downsview, Ontario, Canada, Sept. 1985.
15. R. B. Goldblatt. *Development and Testing of INTRAS, A Microscopic Freeway Simulation Model—Vol. 3, Validation and Application.* Report FHWA/RD-80/108. FHWA, U.S. Department of Transportation, 1980.
16. A. G. R. Bullen and S. L. Cohen. Car Following and Lane Changing Laws in INTRAS Model. *Proc., National Meeting of the Institute of Management Science and Operation Research Board*, Washington, D.C., 1980.
17. B. Hellinga and J. Shortreed. *FREEVU—A Computerized Freeway Traffic Analysis Tool.* Report TDS-91-01. Research and Development Branch, Ministry of Transportation, Ontario, Canada, Jan. 1991.
18. S. A. Smith. *Freeway Data Collection for Studying Vehicle Interactions.* Technical Report. Report FHWA/RD-85/108. FHWA, U.S. Department of Transportation, 1985.

Publication of this paper sponsored by Committee on Freeway Operations.

Abridgment

Recapturing Capacity by Removing Freeway Bottlenecks

CAROL H. WALTERS, CHRISTOPHER M. POE, AND DOUGLAS A. SKOWRONEK

Congestion on urban freeways affects safety, air quality, energy consumption, and motorist delay. In areas where travel demand far exceeds capacity, some level of congestion is inevitable. However, when imbalances in the freeway system exist, bottlenecks restrict the use of available capacity. If these bottlenecks are removed, valuable capacity can be recaptured, congestion reduced, and impacts diminished. Detection of bottleneck locations requires knowledge of traffic volumes and travel speeds and observations of traffic patterns and driver behavior. Bottleneck relief can be found in many cases by restriping lanes, by modifying weaving areas, or by converting some shoulders to driving lanes. Implementation of low-cost improvements, even on an interim basis before permanent improvements, can have an excellent result. Because removal of congestion on one segment of the freeway system may move congestion and impacts to other segments, consideration must also be given to the system implications. In addition, issues such as state and federal policies, public acceptance, and funding for any improvement must be addressed before implementation. However, if these obstacles can be cleared, low-cost improvements to remove bottlenecks will provide significant benefits. In a time of fiscal constraints and with public attention focused on urban congestion, every attempt must be made to get the most efficiency out of our existing freeway systems.

Most urban freeway facilities are planned and designed for traffic volumes forecast 25 years into the future. During the life of the facility, unanticipated changes in travel patterns can occur because of increased growth rates, unforeseen urban land development, or the failure to implement other transportation facilities previously assumed. Whatever the cause, the result is more demand than capacity on some elements of the facility. When these imbalances occur in the system bottlenecks may arise. These sections differ from overcapacity freeway corridors in that often a low-cost improvement can be implemented over a short section of the freeway to significantly relieve congestion and return the system to balance.

WHY TARGET FREEWAY BOTTLENECKS?

One bottleneck in a corridor can create severe congestion for the entire corridor. Considering the high volumes during the peak periods of many urban freeways, the annual delay cost to motorists can be substantial. In addition, capacity is wasted because the bottleneck is effectively "metering" traffic volumes downstream and causing stop-and-go conditions upstream.

Texas Transportation Institute, Texas A&M University, 1600 East Lamar Boulevard, Suite 120, Arlington, Tex. 76011.

There are also environmental and safety effects of bottlenecks. Vehicles operating in congested sections of freeways emit more hydrocarbons and carbon monoxide into the atmosphere. Vehicles caught in stop-and-go traffic consume more fuel. Accidents and vehicle breakdowns (with associated safety impacts) tend to increase in freeway sections with severe congestion. In addition, motorists make erratic maneuvers in attempts to bypass congestion and save time, causing further hazard.

DETECTIVE WORK: WHAT'S HAPPENING HERE?

The first step in attempting to remove freeway bottlenecks is to determine the cause. It is easy to identify the congested freeway sections within an urban area. However, detective work is required to determine which congested freeway sections are caused by bottlenecks and why.

Four different data collection efforts are necessary at a suspected bottleneck site:

1. Peak-period traffic volume counts by 15-min periods,
2. Peak-period travel time runs (every 15 min),
3. Traffic demand patterns (if weaving is involved), and
4. Videotaping of freeway operations (at the bottleneck and drive-through).

Each of these data collection efforts provides clues to the existence of a true bottleneck and its cause.

Peak-Period Traffic Volume

It is recommended that traffic volumes be collected in the peak direction by 15-min time intervals. The small time increment allows detection of patterns in the volume profiles; for instance, capacity may be reached in one 15-min period and then followed by a drop in volume from stop-and-go operations that result from over-capacity conditions. As an example Figure 1 shows plots of peak-period traffic volumes for the southbound section of IH-35E Freeway in Dallas, Texas, at the merge with the westbound-to-southbound ramp from IH-30. The peak-hour (4:30 p.m. to 5:30 p.m.) volume on IH-35E in this section is below the theoretical capacity of the two-lane roadway. At first glance this would not indicate a problem. However, the expected traffic pattern (under free-flow operation) is peaking during the peak period. In Figure

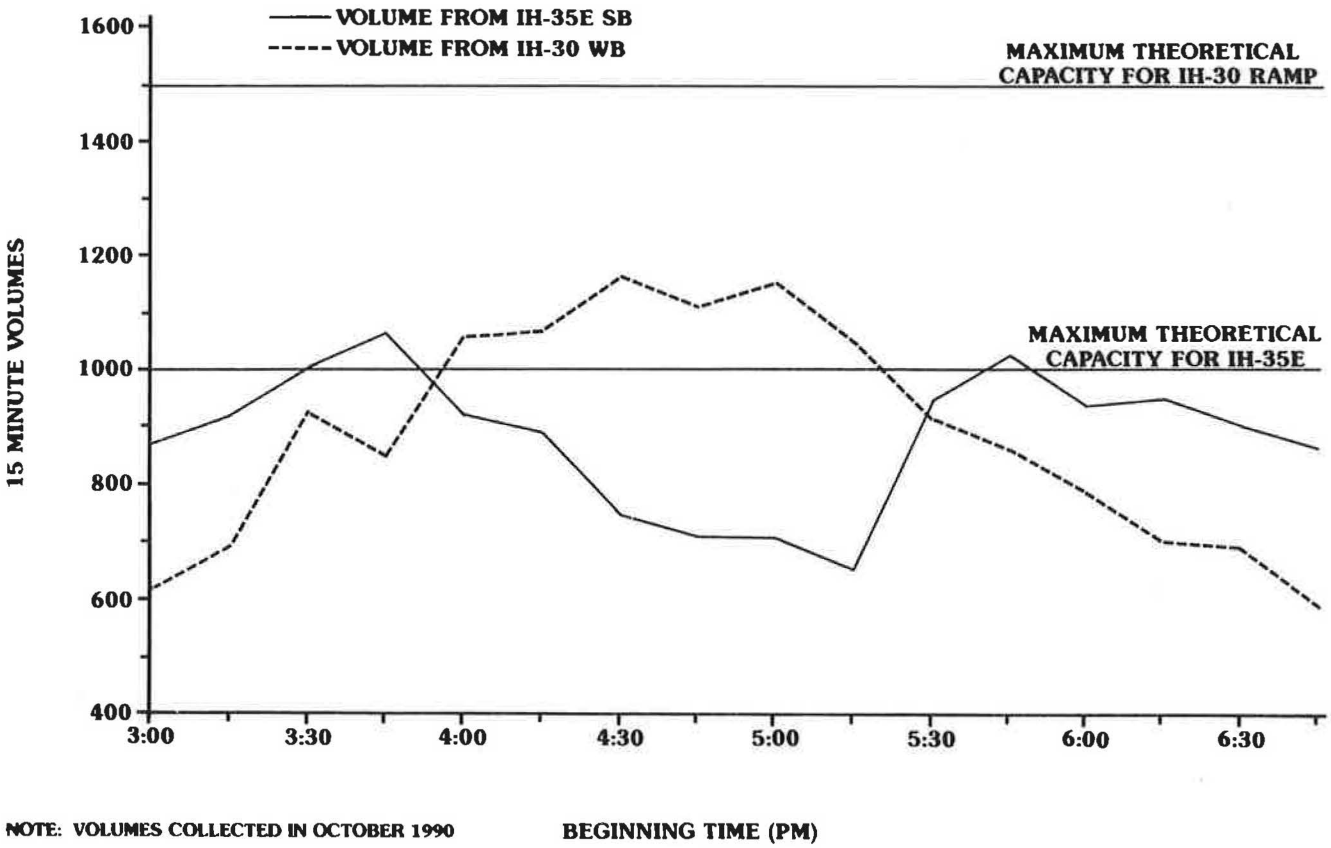

FIGURE 1 15-min volumes for IH-35E southbound and IH-30 ramp (p.m., October 1990).

1, the traffic volumes on IH-35E at 3:00 p.m. are low and start to increase in the following 15-min intervals. However, from 3:45 p.m. to 5:15 p.m. the traffic volumes take a downward trend, indicating constrained operation due to the increase in volume on the competing IH-30 entrance ramp. The traffic pattern on the IH-30 ramp, which is three lanes, indicates that the demand is better served by this facility because of the increase in volumes during the peak hour of the peak period. Traffic on the two approaches fills the capacity of the downstream section (four lanes), but does not accurately reflect the demand from the two approaches. Thus, one of the approaches is experiencing severe congestion from a bottleneck, whereas the other is not.

Peak-Period Travel Times

Peak-period travel times along a section of freeway can give considerably more information about the impacts of a bottleneck. For the example discussed above, peak-hour travel time runs are presented in Figure 2 for both IH-35E and the IH-30 ramp to IH-35E. The speeds along IH-35E immediately upstream of the bottleneck reach a low of 10 mph and speeds below 30 mph last for about 10 min of travel time. Congestion extends for 2 mi and is considerably more severe than that along IH-30 upstream of the merge. The lowest speeds observed on the IH-30 ramp upstream of the merge are 25 mph. Speeds below 30 mph last only a minute or two of the total trip time and congestion extends for only ¼ mi.

Using peak-period traffic volumes and speeds, the delay to motorists can be estimated compared with travel at free-flow speeds. For example, with a 2-hr peak period, the estimated annual delay from this bottleneck on IH-35E is 110,000 vehicle-hr/year and for the IH-30 westbound ramp, only 11,000 vehicle-hr. Clearly, the imbalance in congestion shows that the demands are not proportional to the design capacity of this section.

Traffic Patterns

Often important to understanding the causes of a freeway bottleneck are the origins and destinations of the vehicles in the traffic stream. If the problem is a weaving section that is over capacity, the entire freeway upstream can experience congested conditions. Even major capacity improvements may be ineffective if the problem lies in excessive weaving volumes.

Videotaping of Freeway Operations

Unusual traffic patterns that may be contributing to, or even causing, bottlenecks cannot be detected by simply examining the traffic volumes and travel speeds. Field observations and videotaping are essential in discovering traffic patterns. Observations sometimes also reveal dangerous erratic maneuvers made by motorists as a result of extreme congestion on freeways. It is uncommon to observe motorists driving on shoul-

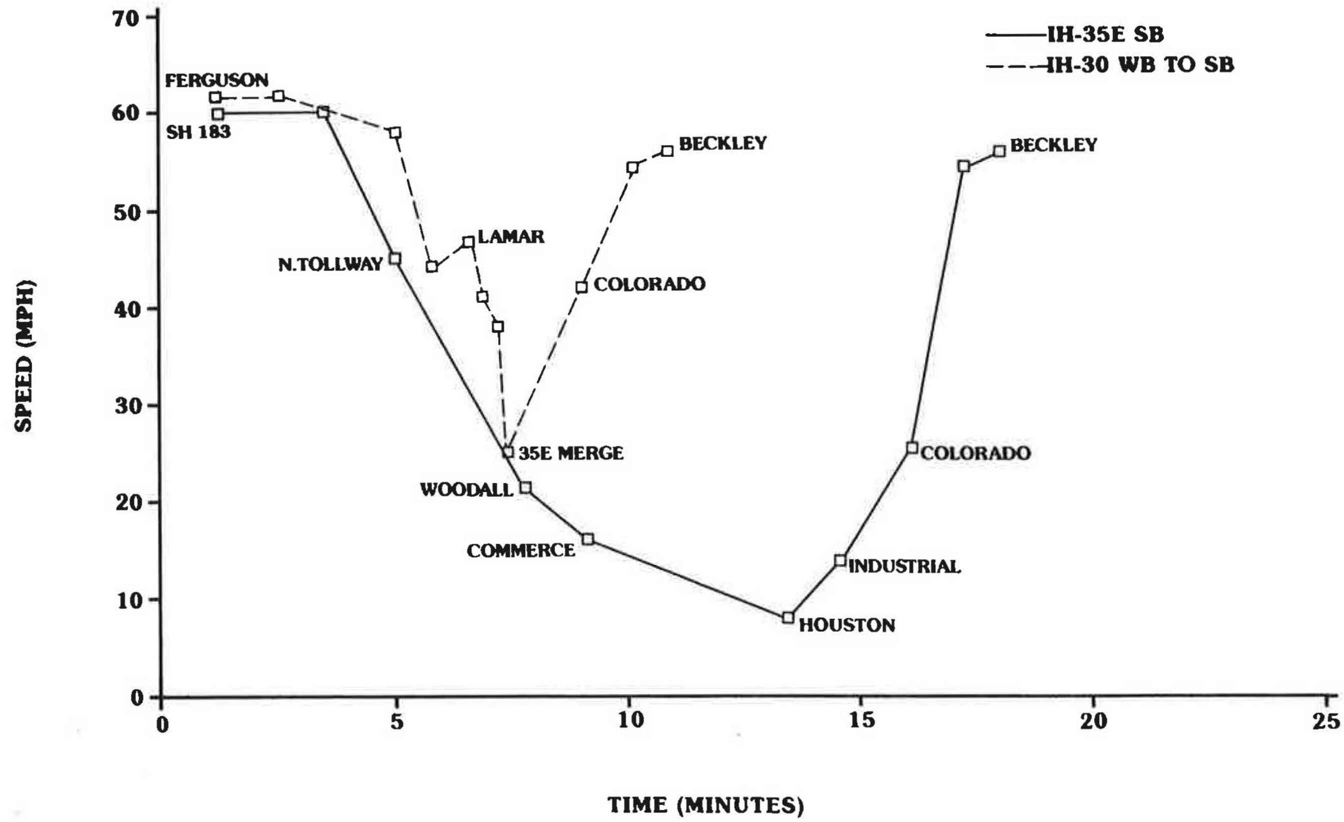

FIGURE 2 **Typical travel speeds on IH-30E southbound and IH-30 ramp, p.m. peak period.**

ders or driving two and three abreast in wide ramp gore areas. From an operational standpoint, these maneuvers sometimes result in stretching capacity, and they may even point toward the solution needed for the problem. However, in general, they simply provide an advantage for those making the erratic maneuver. These motorists are a distinct disadvantage to the overall freeway operation and frustrate those driving legally. The safety implications can be enormous.

FOUND CAPACITY: THE LOGICAL SOLUTION

Once a bottleneck has been identified and its causes have been determined, solutions within the given right-of-way should be sought. Ideally, a bottleneck improvement will be a low-cost improvement designed to provide more capacity to a freeway section temporarily until some permanent capacity addition to the freeway can be constructed. Because the improvement is temporary in nature, minimum design standards may need to be considered. Future reconstruction of the freeway would allow for the minimum designs to be upgraded to more desirable standards. In the meantime, the potential safety impacts of allowing minimum design standards must be weighed against the safety benefits of decreasing congestion.

The following are some commonly implemented low-cost improvements that have been used in Texas:

1. Use a short section of shoulder as an additional lane,
2. Restripe merge and diverge areas to balance capacity and demand,
3. Modify weaving areas,
4. Add lanes for short segments, and
5. Use peak-period ramp metering or ramp closures.

The use of these improvements often depends on the approval of local, state, and federal agencies in a given area.

BENEFITS AND COSTS: WHAT'S THE PAY-OFF?

First, the most obvious benefits from removing bottlenecks are the increased peak-period speeds and the reduced delay to motorists. These benefits can considerably outweigh the cost of bottleneck improvements.

However, estimating the extent of delay reduction can be difficult. Theoretical analysis using the *Highway Capacity Manual* (*1*) does not always yield useful results in these cases. In sections with constrained volumes, theoretical analysis shows an adequate level of service when the volume is below capacity. Demand upstream of the bottleneck must first be estimated in order to quantify the congestion on the freeway. This estimated demand can be used to analyze the operation with the proposed improvement. The time savings with the improvement can be quantified and the motorists' benefits calculated. Common benefit-cost ratios of projects in Texas have ranged from 2.3 to 16.6, using \$10/vehicle-hr (*2*).

Safety benefits are also likely to occur. Accident data collected at three bottleneck locations in Dallas, Texas, were analyzed to determine the types of accidents occurring during the peak period under congested conditions. The types of accidents attributed to congestion were rear end or single vehicles that ran off the road or into a barrier to avoid hitting a slow-moving vehicle. Almost 80 percent of all accidents that occurred during the peak period were attributable to congestion. If the congestion conditions were relieved, there would be fewer accidents. In addition, these accidents in themselves contribute to further delay, which would also be relieved.

CONCLUSION

Congestion on urban freeways increases delay to motorists, pollution to the environment, and accidents and incidents. Many freeways have simply too much demand for their capacity, but some short freeway sections are bottlenecks that may yield to low-cost solutions within the freeway right-of-way. With some detective work to uncover the causes of the problems, low-cost improvements can be designed to reduce the severe impacts of congestion and recapture the freeway capacity that bottlenecks deplete.

The following summarizes the key points to bottleneck removal:

1. Traffic volumes alone will not detect (but may suggest) locations of bottlenecks. Vehicle speeds, local traffic patterns, and field observations are needed to detect both the existence and the cause of freeway bottlenecks.
2. The amount of congestion on different approaches at freeway-to-freeway interchanges can be very imbalanced. Distributing the capacity to reflect the demand can greatly reduce the overall congestion in the system.
3. Improvements such as restriping lanes, using shoulders, and modifying weaving areas produce primary benefits of reduced congestion and improved safety. These improvements also produce secondary benefits in reductions in emissions, vehicle operating costs, and congestion on alternative routes.
4. These benefits can be accomplished at minimal cost compared to overall capacity improvements because they amount to recapturing existing capacity within the freeway system, smoothing out the problems.
5. Additional checks must confirm that implementation of a bottleneck improvement will not simply move the congestion to another location (which may cause further safety problems).

If design elements can be approved and if all involved agencies cooperate, low-cost improvements to remove freeway bottlenecks can provide significant benefits. In a time of fiscal constraints and public attention on urban congestion, every attempt must be made to get the most efficiency out of our existing freeway systems.

REFERENCES

1. *Special Report 209: Highway Capacity Manual*, TRB, National Research Council, Washington, D.C., 1985.
2. W. F. McFarland and M. Chui. *The Value of Travel Time: New Estimates Developed Using a Speed Choice Model.* Research Report 396-2F. Texas Transportation Institute, College Station, 1985.

Publication of this paper sponsored by Committee on Freeway Operations.

Abridgment

Control Emulation Method for Evaluating and Improving Traffic-Responsive Ramp Metering Strategies

YORGOS J. STEPHANEDES, EIL KWON, AND KAIKUO CHANG

A method is developed for evaluating traffic-responsive ramp metering strategies and improving freeway performance in intelligent vehicle-highway systems operations. The method emulates real-time metering, as currently practiced in the United States and Canada, and traces the interactions between automatic rate-selection metering strategies and freeway performance through time. The method can facilitate traffic management by aiding in the design of new traffic-responsive ramp strategies, and in the evaluation of existing real-time control systems and selection of the best metering schemes. The method was tested successfully in emulating volume and occupancy thresholds, rate tables, and automatic rate-selection control strategies, and in assessing and selecting ramp strategies based on freeway performance measures (such as total volume and delay) on I-35W in Minneapolis.

The most advanced concept in freeway management and control includes a hierarchical traffic control structure; overall freeway control is decomposed into several components—demand prediction, network optimization and local control—in order to achieve computational and practical feasibility of optimal control strategies (*1*). However, the state of the art in freeway ramp metering has not reached the point at which comprehensive, networkwide control strategies are automatically generated and implemented through on-line optimization. As a result, most traffic-responsive metering systems still use automatic rate-selection procedures. These procedures select the most appropriate metering rates for a ramp by using predetermined thresholds and rate tables using the information received from loop detectors on the main freeway, upstream and downstream from the ramp. Although this method provides a degree of self-adjustment to prevailing traffic conditions, the lack of an efficient tool for updating the key components of the control system (i.e., thresholds, rate tables, and locations of the detector stations) significantly restricts the effectiveness of control.

Developing an efficient and cost-effective tool for evaluating "what if" control policies is an essential element in improving intelligent vehicle-highway system (IVHS) freeway traffic management. In this study, a method is developed that can simulate various traffic-responsive ramp metering strategies. The method emulates real-time metering systems based on a "mapping" of the control logic developed by experienced traffic engineers, at state departments of transportation, and by simulated freeway performance. The rest of this paper describes the structure of the method and an example application in the Minneapolis metering system.

Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis, Minn. 55455.

CONTROL EMULATION METHOD

The control emulation method developed in this study emulates real-time rate-selection metering using simulated freeway traffic performance when the freeway demand pattern and geometrics are known. Figure 1 illustrates the key components of the method (i.e., ramp control and freeway performance modules) interacting continuously through time. The ramp control module emulates the on-line automatic rate-selection process and determines metering rates using the traffic information provided by the freeway performance module. The key elements of the process include a set of

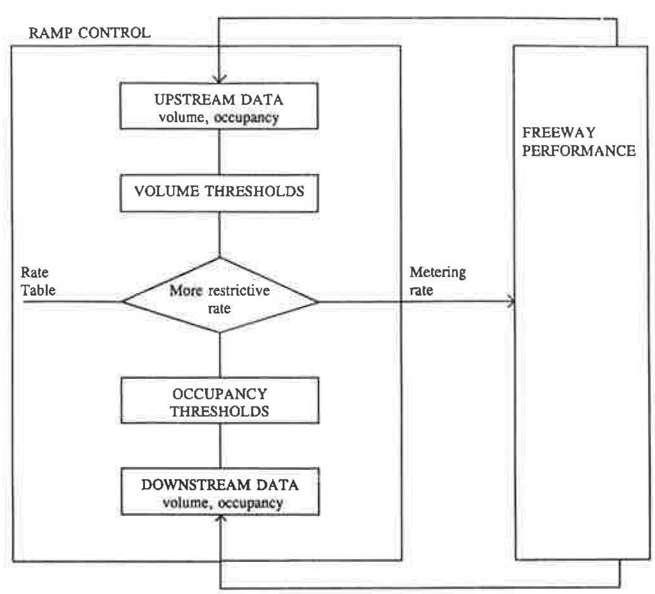

FIGURE 1 Automatic rate-selection process.

volume-occupancy thresholds, a rate table, and the locations of detector stations associated with each ramp. First, traffic volume and occupancy at predefined detector locations are determined by the freeway performance module, using dynamic simulation based on continuum modeling. Upstream volume and the highest downstream occupancy among a set of stations are compared with the predetermined volume-occupancy thresholds. Using the rate table, two rates are identified and the more restrictive rate is selected. Following a time delay, the freeway performance module uses that rate for determining the number of vehicles entering the freeway system. This rate-selection process is currently operating at the Minnesota Department of Transportation; variations of it are operating elsewhere in the United States and Canada.

The dynamic simulation developed in this study uses a finite difference scheme, which was specifically designed for one-dimensional, time-dependent, compressible flows containing strong shocks; the simulation has been applied to modeling traffic flow (*2,3*). The method determines traffic density, speed, and flow in each 100-ft segment of freeway every second, so that the effects of short-term (e.g., 6- to 30-sec) changes in metering rates on the traffic performance at the detector station locations can be assessed accurately. The detailed description of the modeling methodology for the performance module is reported elsewhere (*4*).

APPLICATION

The features of the control emulation method are illustrated through application to a freeway in the Minneapolis metropolitan area. In Minneapolis, metering rates at a ramp are determined from volume data from a mainline station upstream and occupancy data from five mainline stations downstream of that ramp. For each ramp, Minnesota Department of Transportation traffic engineers have used historical data to derive a set of volume-occupancy thresholds and a rate table consisting of six metering rates. In addition, rate selection every 30 sec is based on two measurements: the 1-min upstream volume and the highest 1-min occupancy among five downstream detector stations.

The preceding metering policy can be further described using the volume-occupancy (V-O) diagram. To illustrate, consider a typical volume-occupancy threshold policy described by curve AA' in Figure 2, where the whole space is subdivided into four regions, (i.e., no-control, occupancy-dominant, volume-dominant, and common volume-occupancy control regions). If a V-O measurement from the freeway falls into either the volume- or the occupancy-dominant region, the metering rate is determined by volume or occupancy respectively. In the common V-O control region the resulting metering rate is the same, whether it is determined by volume or occupancy. In Figure 2, the six metering rate levels are indicated by number, ranging from 1 to 6, where Rate 6 is the most restrictive. For instance, if upstream volume is 40 vehicles per minute and highest downstream occupancy is 30 percent per minute, control will be occupancy-based and Rate 5 will be activated at the ramp meter. If occupancy falls to 18 percent and volume remains the same, control will activate volume-based Rate 4, a less restrictive rate.

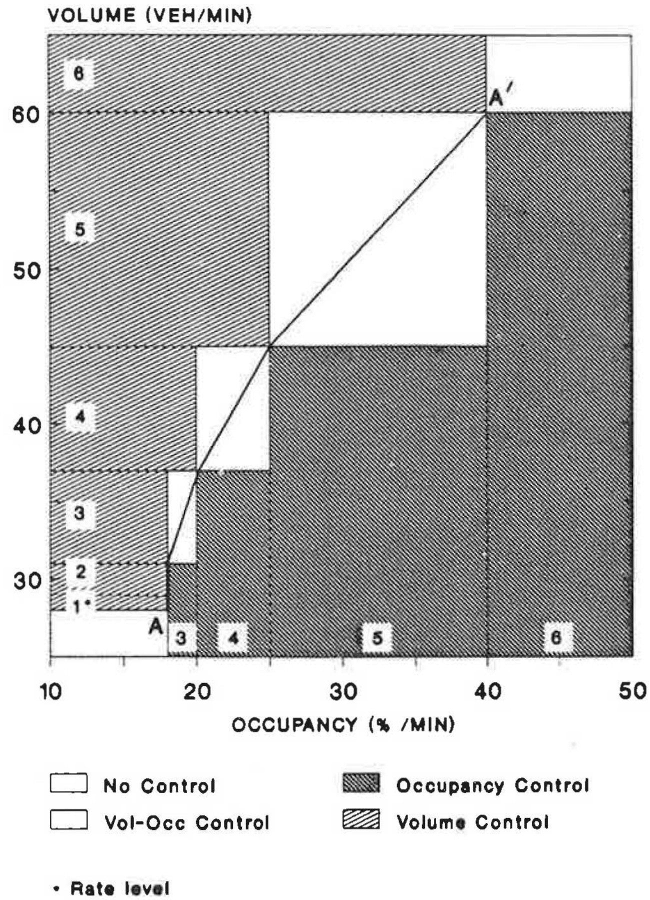

FIGURE 2 Typical volume-occupancy threshold policy.

Test Site and Model Testing

A 6-mi section of the I-35W freeway south of Minneapolis was selected as the test site (see Figure 3). This section contains 18 entrance-exit ramps and a variety of geometric types such as merging, diverging, and weaving. The test section also includes 13 mainline loop detector stations, at which volume and occupancy measurements can be obtained every minute.

The control emulation method was first tested by simulating the test site with the current metering strategy and the real volume data collected from the upstream boundary and each ramp of the test section. The simulated volume was compared with the real data collected from the mainline loop detectors. Results from the volume comparison indicated an error ranging from 5 to 12 percent. Performance could improve with improvements in modeling, an issue currently being addressed by the authors.

Evaluation of Occupancy Thresholds

As an example, the control emulation method was applied to determine the effects of occupancy threshold changes on traffic performance in a sample freeway section. Four alternative threshold policies were formulated by increasing or decreasing occupancy thresholds by 2 and 4 percentage points above or below their current values, respectively, while keeping vol-

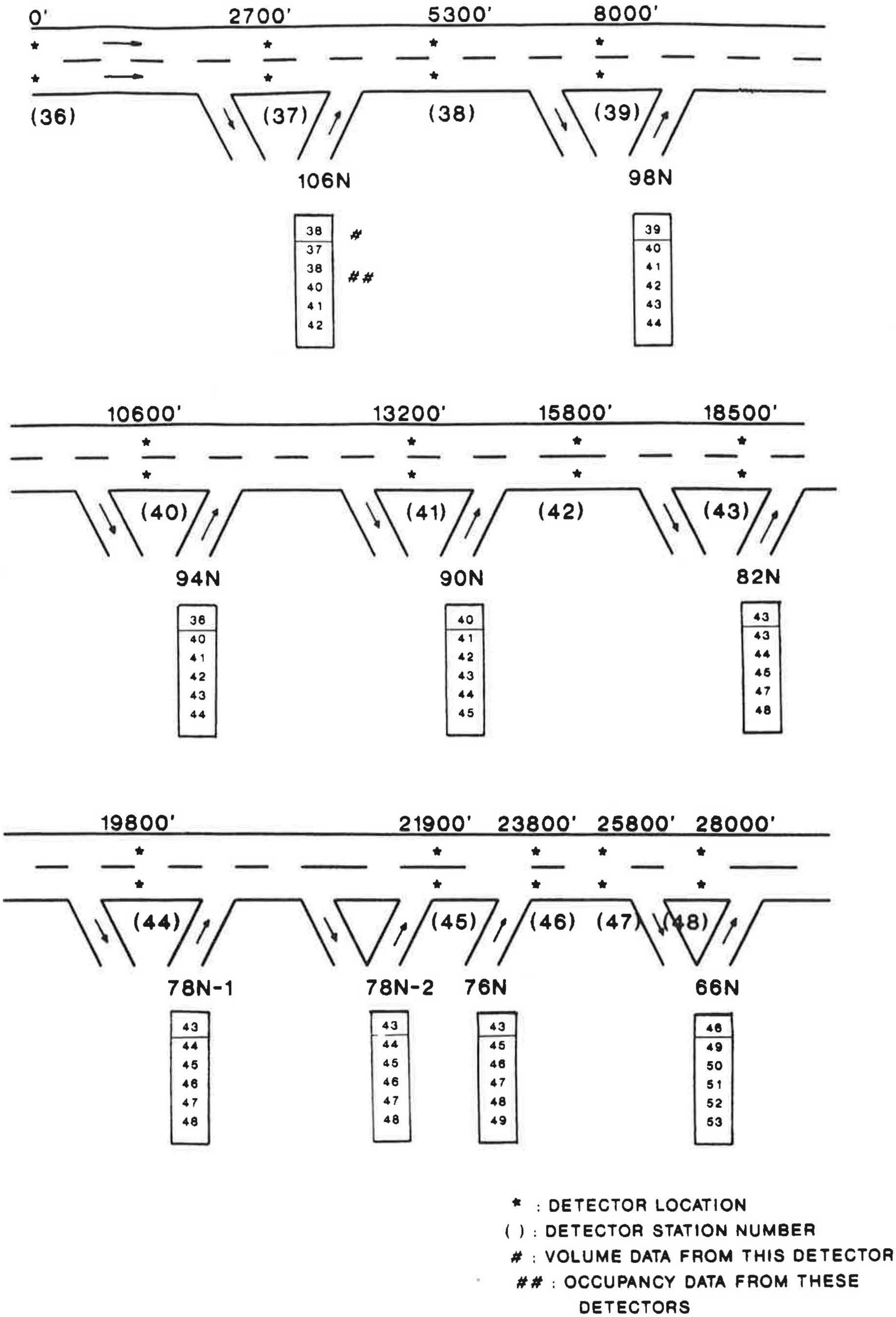

FIGURE 3 Freeway test section, I-35W, Minneapolis.

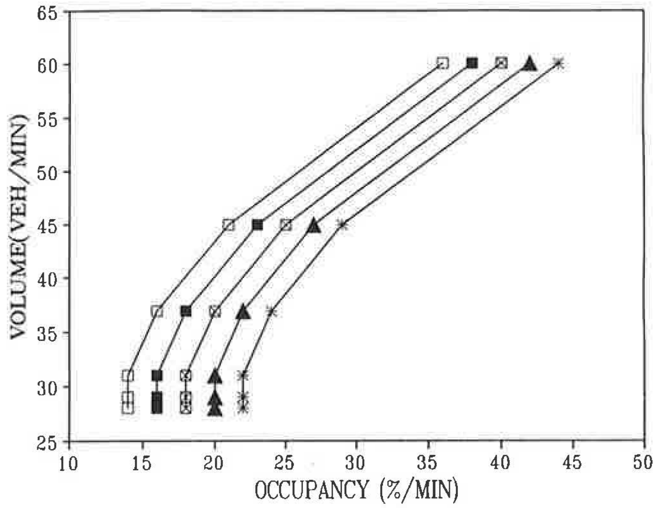

FIGURE 4 Occupancy threshold strategies.

ume thresholds unchanged. The four policies and the current one are presented in Figure 4. The demand, consisting of 5-min measurements of traffic volume entering and exiting each ramp and traffic volume arriving at the upstream boundary, was the average Thursday demand measured in November 1990. Two performance indexes were selected and their values assessed.

- Total volume entering the freeway during the simulation period, and
- Total delay, defined as the total travel time of all vehicles at a speed less than 45 mph.

Higher-occupancy thresholds resulted in higher 3-hr volume and delay on the freeway as expected (see Figure 5). The results also suggest that volume is more sensitive to occupancy threshold decreases below the current value than to threshold increases. For example, with existing demand, increasing the current 18 percent threshold by 2 percentage points results in a 0.6 percent increase of freeway volume and a 4.8 percent increase of total delay. However, shifting down the current threshold by 2 percentage points results in a 1.1 percent volume decrease and a 8.9 percent delay reduction. By considering the trade-offs between volume increase and delay reduction, a desirable threshold policy can be determined for a freeway section.

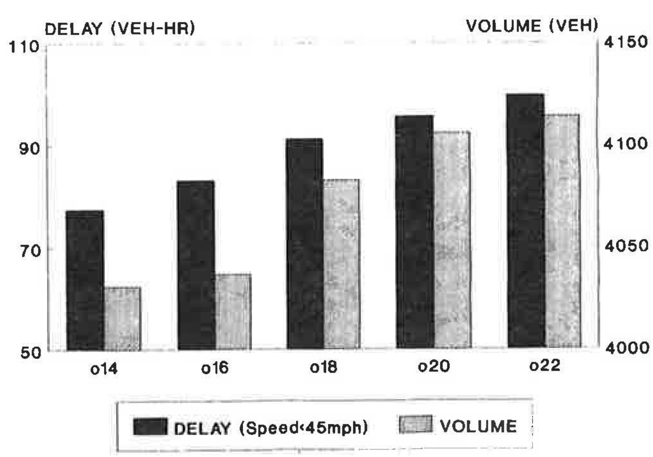

FIGURE 5 Performance of occupancy threshold strategies: existing demand.

CONCLUDING REMARKS

The proposed control-emulation method can be a practical traffic management tool by aiding in the design of new traffic-responsive ramp strategies as well as in the evaluation of existing control systems and selection of the best metering schemes. Using this tool for consideration of trade-offs between performance indicators, such as volume increase and delay reduction, a desirable threshold policy can be determined for a freeway section before implementation. Although the new method improves on the performance of conventional systems, it is still restricted by modeling and hardware limitations. In order to achieve an adaptive metering performance in real-time, large IVHS freeway networks, the authors are researching the need for improved traffic prediction, data collection, and parallelization of large-scale computations. Research also continues for automating the selection of optimal strategies in arterials. It is anticipated that the two branches of work will lead to improved selection of demand responsive control strategies in urban networks.

ACKNOWLEDGMENTS

This study was supported by the National Science Foundation and the Center for Transportation Studies, Department of Civil and Mineral Engineering, University of Minnesota. The Minnesota Supercomputer Institute provided partial support. The Traffic Management Center, Minnesota Department of Transportation, cooperated in the study by providing the necessary data and information on current freeway control strategies in the Twin Cities.

REFERENCES

1. M. Papageorgiou. A Hierarchical Control System for Freeway Traffic. *Transportation Research*, Vol. 17b, 1983, pp. 251–261.
2. P. Lax. Weak Solution of Non-Linear Hyperbolic Equations and Their Numerical Computation. *Communications on Pure and Applied Mathematics*, Vol. 7, 1954, pp. 159–173.
3. P. Michalopoulos, E. Kwon, and J. Kang. Enhancements and Field Testing of a Dynamic Freeway Simulation Program. In *Transportation Research Record 1320*, TRB, National Research Council, Washington, D.C., 1991.
4. Y. Stephanedes, E. Kwon, and K. Chang. *Development and Application of Demand Responsive Ramp Metering Control To Improve Traffic Management in Freeway Corridors: Phase I*. Final Report. Minneapolis, Minn., 1991.

Publication of this paper sponsored by Committee on Freeway Operations.

Abridgment

Evaluation of Video Image Processing Systems for Traffic Detection

CARL A. MACCARLEY, STEPHEN L. M. HOCKADAY, DANIEL NEED, AND SAMUEL TAFF

Recent advances in microprocessor technology have made possible the cost-effective computer processing of video images of highway traffic for measurement of traffic speed, volume, and density. Video image processing systems (VIPS) are now being considered as key components of advanced traffic management systems (ATMS). Eight commercially available or prototype VIPS were recently evaluated for effectiveness at the request of the California Department of Transportation. Test procedures were intended to duplicate conditions typically encountered on California freeway systems. A summary of the test results is reported. Most systems performed well under optimum conditions but degraded significantly under nonoptimum conditions, which varied from system to system. On the basis of the VIPS systems studied, this technology is considered feasible for specific traffic monitoring problems. Limitations and areas for further development are identified.

Using computer vision technology to determine traffic flow data from video images is considered an important enhancement and valuable component of advanced traffic management system (ATMS) strategies. Several commercial or near-commercial systems are now available. Our study examined the current (1990–1991) state of the art in traffic VIPS technologies.

At the time of the study, 10 commercial or prototype VIPS were found to be available in the United States, Europe, and Japan. These are presented in Table 1. All manufacturers listed, except Hitachi Ltd. and Sumitomo Ltd., participated in the study.

These systems were designed to detect some subset of the basic traffic flow parameters: instantaneous and time-average vehicle speed, vehicle density per mile, mean headway, lane occupancy, and accumulated vehicle count. Traffic volume and mean headway may be inferred from the other metrics. These data are usually reported on a per-lane basis, sometimes with full-roadway averages provided. Our study targeted the ability of the systems to count vehicles and determine individual vehicle speeds as the primary metrics of performance.

All systems were software-based. Some required specialized hardware platforms, others ran on IBM PC-compatible platforms requiring only video digitizing cards for the camera interface. The core of the software task involved some algorithm for detecting each vehicle and measuring its velocity (*1–5*).

C. A. MacCarley, D. Need, Electronic and Electrical Engineering Department, California Polytechnic State University, San Luis Obispo, Calif. 93407. S. L. M. Hockaday, S. Taff, Civil and Environmental Engineering Department, California Polytechnic State University, San Luis Obispo, Calif. 93407.

Although the objectives of the project did not include analysis of detection algorithms, setup and operation of each system clearly identified two fundamental algorithmic approaches. We designate these as Types 1 and 2 and segregate the systems on the basis of algorithm type, as indicated in Table 1.

Type 1 algorithms involve detection of the time difference of light-level changes between two virtual gates in the image, spaced a known physical distance apart. A vehicle moving down each lane causes an intensity change at the first gate, then at the second gate. This pair of events is interpreted as the passage of a single vehicle. The vehicle velocity is determined by measurement of the time between the two gate-crossing events.

Type 2 algorithms might be referred to as vehicle tracking algorithms since they first detect the presence of a cohesive object moving in the image and then measure the velocity along its trajectory. Type 2 algorithms are generally more sophisticated and require significantly greater computer processing power. They are similar to military target tracking algorithms.

All algorithms tested were designed to handle oncoming traffic, although most could handle departing traffic also. Detection of departing traffic is possibly a more reliable approach.

All systems used monochrome video images and were designed to operate with standard monochrome surveillance video cameras. Systems manufactured in the United States and Japan used the Electronic Industry Association (EIA) 170 video format. Systems manufactured for use in Europe conformed to the International Radio Consultive Committee (CCIR) video format; some were provided in EIA format for the U.S. market.

TEST PROCEDURES

The use of a real-time video feed of freeway traffic was unfeasible for the nature and scope of the test performed. Therefore, videotaped traffic images were used.

TABLE 1 VIP Systems Evaluated

Company	Status	Algorithm Type	Video Format
ASPEX	prototype	1	EIA
CRS	commercial	1	EIA
Devlonics	commercial	1	EIA
Eliop	prototype	2	EIA
Hitachi	prototype	N/A	EIA
ISS	commercial	1	EIA
INRETS	prototype	2	CCIR
Sense & Vision Systems	prototype	2	EIA
Sumitomo	prototype	N/A	EIA
University of Newcastle	prototype	1	CCIR

Two self-contained mobile video data acquisition units were constructed. Each contained two super-VHS format professional video recorders interfaced with two video cameras, frame number encoders, and monitors, providing concurrent EIA and CCIR format recording and indexing capability. Solid-state metal oxide semiconductor/charge coupled device (MOS/CCD) cameras were used, with variable focal length lenses and mechanical aperture adjustments. Certain night tests substituted specialized MOS cameras, which were less prone to a problem associated with CCD cameras known as vertical smearing, in which vertical white lines appear to extend from vehicle headlights. The two mobile units were operated simultaneously at each data collection site, recording concurrent traffic images from two camera positions.

A suite of 28 test conditions was defined for evaluation of the systems, described in Table 2. This collection of images was intended to emulate actual field conditions that would be encountered during 24-hr, year-round service on California urban freeways. Parameters included day and night illumination levels, variable numbers of lanes (2 to 6), various camera elevations and angles to the roadway, rain and fog conditions, camera vibration and sway, traffic conditions ranging from free flow through heavy congestion, shadows from vehicles or stationary objects, and the effects of simulated ignition noise and 60 Hz electromagnetic noise combined with the video signal. Tests were performed on approaching and departing traffic. As a practical matter, included in the test suite were only those combinations of variables most representative of standard deployment scenarios. Table 2 indicates the parameter or combination of parameters emphasized in each of the 28 standard tests.

The test suite was created by editing several hundred hours of raw video collected over the course of a year. Each test segment is 20 min long. This includes a 10-min initial period to permit the system under test to cancel the background and adapt to the ambient light level. Actual traffic counts and vehicle velocities on a per-lane basis were determined by inspection of the videotaped images over the duration of each segment. Frame-by-frame inspection was employed for accurate velocity measurements and vehicle counts.

Most systems were designed for camera placement directly above the roadway centerline at a height of 10 to 15 m. A high camera position minimizes vehicle occlusion but is more prone to sway and vibration. A centered camera minimizes perspective distortion. A roadside placement is easier to install and maintain and provides a greater field of view.

TABLE 2 Summary of Test Suite

Test #	Parameter Tested
1	Large # of Lanes
2	Small # of Lanes
3	Day to Night Transition
4	Shallow Camera Angle
5	Steep Camera Angle, Departing Traffic
6	Shallow Camera Angle, Departing Traffic
7	Night, Steep Camera Angle, Approaching Traffic
8	Night, Shallow Camera Angle, Approaching Traffic
9	Night, Steep Camera Angle, Departing Traffic
10	Night, Shallow Camera Angle, Departing Traffic
11-18	Same as 3-10(above), Side Camera Mounting
19	Weather - Fog
20	Weather - Rain, Daytime
21	Weather - Rain, Night time
22	Unstable Camera Mount - Sway
23	Heavy Traffic - Capacity Operation
24	Congested traffic
25	Heavy Shadows from Vehicles
26	Heavy Shadows from Environment
27-28	Ignition and Electromagnetic Noise

All test images were acquired from freeway overpasses, with cameras placed on both the roadway centerline and off to the side. This duplicates the placement of surveillance cameras anticipated by the California Department of Transportation (Caltrans) in the Los Angeles area. Camera heights for the 28 test conditions varied from 8.3 to 14.2 m above the roadway surface, measured with an ultrasonic range finder.

Lenses for both camera placements were selected to permit all traffic lanes (in one direction) to be contained in the field of view.

All systems were set up and calibrated to the manufacturer's specifications; manufacturer representatives were present in most cases.

A qualitative evaluation of system human factors was also performed, considering issues of ease-of-setup and use, quality of graphical interface, and sensibility of data display. System reliability was not specifically evaluated, but observations relating to it were recorded and reported.

EVALUATION RESULTS

Figure 1 summarizes the average performance of the systems, classified by algorithm type. Shown are group average detection accuracies relative to each test condition and in performance categories. For all systems, we observed error rates usually less than 20 percent for vehicle count and speed measurements over a mix of low, moderate, and high traffic densities, with optimum camera placement and clear, daylight, nonshadow conditions. Complete numerical results of all tests are included in the project final report (*6*).

Systems designed for very high camera placement were usually intolerant of partial occlusion of vehicles, yielding high error rates for tests with lower camera heights.

Tests with slow-moving high traffic densities usually yielded reduced accuracy and occasionally complete detection failure, probably because of the action of the particular background subtraction algorithm employed. These situations were emphasized in Tests 23 and 24 (Table 2).

Light-level changes at sunrise and sunset caused reduced accuracy. During these periods, most systems make a transition from daytime algorithms, which detect entire vehicles, to nighttime algorithms, which detect headlight groups. To some degree, this was a notable area of weakness for all systems tested as this area is suggested for further study, because peak traffic periods usually coincide with sunrise and sunset.

Tests 21, 25, and 26 (Table 2) emphasized two aberrant conditions that caused particularly high error rates for most systems: rain at night and long vehicular and stationary shadows. Long shadows are a particular problem at sunrise and sunset, adding to the transition difficulties just mentioned. Headlight reflections from a wet roadway, especially in low-light conditions, cause similar detection errors. These problems are related in the sense that they challenge the ability of the systems to discriminate actual vehicles from other moving areas of high contrast (either light or dark) in the camera image.

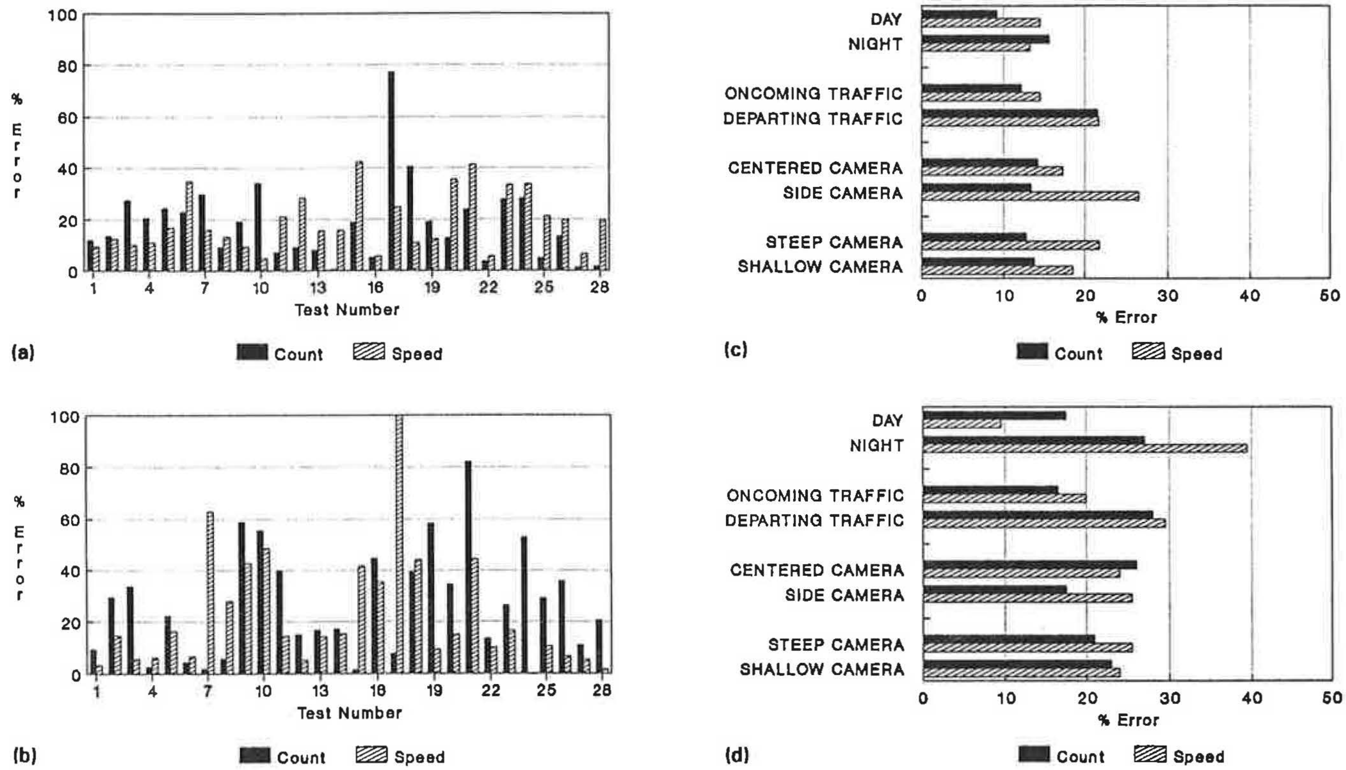

FIGURE 1 Summary of test results: average absolute error by test for (*a*) Type 1 systems and (*b*) Type 2 systems; average absolute error by grouping for (*c*) Type 1 systems and (*d*) Type 2 systems.

Type 1 algorithms attempt to cancel headlight reflections and vehicle shadows by rejecting detection events that occur in too brief a time interval. Type 2 systems attempt to correlate a shadow or reflection with an associated vehicle. However, the source of the shadow or light may be outside the field of view, for example, a car off the detected area of roadway, aircraft overhead, or the shadow of a tall object or tree. In these situations, both algorithm classes usually fail.

The effects of added electronic (ignition) noise were also studied in Tests 27 and 28. Generally, low noise levels had little effect on count or speed accuracy up to a threshold at which detection failed completely. A similar observation for atmospheric fog was made in Test 19.

At the time of the study, only three of the systems were available commercially for "immediate installation." All used Type 1 algorithms.

CONCLUSIONS AND RECOMMENDATIONS

Under optimum daytime conditions, the Type 1 systems generated more accurate vehicle counts and the Type 2 systems generated more accurate speed measurements. Under optimum conditions, no system was clearly superior to the others. Aberrant conditions yielded high error rates for both algorithm classes. Overall, Type 1 systems showed somewhat lower error rates in both vehicle count and speed measurements. However, it should be noted that the Type 2 systems studied were prototype versions at the time.

Conditions that degraded the performance included the following:

1. Nonoptimum camera placement,
2. Transition from day to night,
3. Headlight reflections on wet pavement,
4. Shadows from vehicles or objects outside the detection area, and
5. Obscured atmospheric conditions (fog or heavy rain).

System costs have fallen significantly from 1989 to 1992; quantity prices for commercial systems are now projected at under $5,000 (U.S.). As more commercial products become available, costs should approach that of the processing hardware.

Specifications should be developed for field installations to simplify deployment and maintenance activities. Links to traffic operation center (TOC) technologies should also be designed into systems to avoid telecommunications problems. Compatibility with video data compression algorithms is also recommended as an area for development, because signals from field-deployed cameras may be transmitted to a TOC in compressed format.

ACKNOWLEDGMENT

The support of Caltrans and Randall Ronning, the project technical monitor, is gratefully acknowledged.

REFERENCES

1. J. M. Blosseville et al. Titan: A Traffic Measurement System Using Image Processing Techniques. *Proc., IEEE 2nd International Conference on Road Traffic Monitoring*, London, England, 1989.

2. L. Cypers et al. CCATS: The Image Processing-Based Traffic Sensor. *Traffic Engineering and Control*, June 1990.
3. R. M. Inigo. Application of Machine Vision to Traffic Monitoring and Control. *IEEE Transactions on Vehicular Technology*, Aug. 1989.
4. P. Michalopoulos. Automated Extraction of Traffic Parameters Through Video Image Processing. *ITE 1990 Compendium of Technical Papers*, Aug., 1990.
5. M. Takatoo et al. Traffic Flow Measuring System Using Image Processing. *Automated Inspection and High-Speed Vision Architectures III*, SPIE Vol. 1197, 1989.
6. S. Hockaday et al. *Evaluation of Image Processing Technology for Application on Highway Operations*, Final Report. TR 91-2. June 1991.

Publication of this paper sponsored by Committee on Freeway Operations.

Comparative Performance Evaluation of Incident Detection Algorithms

YORGOS J. STEPHANEDES, ATHANASIOS P. CHASSIAKOS, AND PANOS G. MICHALOPOULOS

A critical review of the most widely accepted conventional incident detection algorithms is presented. An improved algorithm, employing short-term time averaging to filter the traffic data, is proposed and compared with the best of the existing algorithms. The filtering addresses the problem from false alarms that are due to short-term traffic inhomogeneities. On the basis of data collected from a typical freeway in the Twin Cities metropolitan area, evaluation results suggest that the new algorithm achieves a 30 to 80 percent false alarm reduction compared with existing algorithms.

Fast and reliable freeway incident detection is instrumental in reducing traffic delay and increasing safety. In particular, with the information from incident detection, optimal control strategies guide the traffic flow toward smooth operation by preventing additional vehicles from entering the freeway upstream of the incident and by communicating relevant information to travelers. In addition, incident detection constitutes the cornerstone for prompt incident management and safety improvement near the incident location.

Existing techniques for the detection of freeway incidents do not provide the necessary reliability for freeway operations. Because they generate a high level of false alarms, conventional automated techniques based on computerized algorithms are less effective than is desirable for operational use. Operator-assisted methods minimize the false alarm risk, but they suffer from missed or delayed detections, they are labor-intensive, and they restrict the potential benefits from advanced, integrated traffic management schemes.

Responding to the need for effective and reliable detection of freeway incidents, an essential element for improved traffic management and control in freeway corridors (*1*), the authors initiated this research to investigate the performance limitations of conventional automatic incident detection systems and define the specifications for a new algorithmic logic that can lead to improved detection performance. The research initially focused on assessing the ultimate detection performance that can be accomplished with existing and new incident detection systems that use traffic data from presence detectors. A new algorithm was developed and tested against the major existing ones with promising results towards the development of a more-sophisticated detection structure. All tests employed a unified system of performance assessment (*2*), suitable for direct algorithm evaluation.

Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis, Minn. 55455.

BACKGROUND

Automatic incident detection (AID) involves two major elements: a traffic detection system that provides the traffic information necessary for detection and an incident detection algorithm that interprets the information and ascertains the presence or absence of a capacity-reducing incident. Presence detectors imbedded in the freeway pavement are used extensively to obtain traffic data, primarily on occupancy and volume. Video detectors can also be used for the same purpose.

A number of AID algorithms can be found in the literature. Their structure varies in the degree of sophistication, complexity, and data requirements. The most important include the comparative algorithms (California logic) (*3–5*), the type employing statistical forecasting of traffic behavior (time-series algorithms) (*6–8*), and the McMaster algorithm (*9*). These algorithms operate on typical detector outputs of 30- to 60-sec occupancy and volume data. Equally important is the HIOCC algorithm (*10*), which uses 1-sec occupancy data to detect stationary or slow-moving cars; however, such data are not always available with existing surveillance systems. Additional detection methods (*11–13*) involve macroscopic traffic flow modeling to describe more fully the evolution of traffic variables. Table 1 presents the major AID algorithms, including the proposed algorithm, on the basis of their data requirements. For instance, the table indicates that the McMaster algorithm employs volume occupancy, and (optionally) speed data, averaged over 30-sec periods from one or two adjacent stations.

Algorithms of the comparative type, developed by Payne et al. (*3*) and later by Levin et al. (*4*), rely on the principle that an incident is likely to increase significantly occupancy upstream while reducing the occupancy downstream. To detect incidents, actual values of occupancy are compared against preselected thresholds. A typical algorithm includes tests to distinguish between incident and bottleneck congestion, compression wave tests to isolate compression shock-waves, and a persistence test to ensure that exceeding a threshold is not due to random fluctuations in the data. In an attempt to improve the significance of incident alarms, Levin and Krause (*5*) observed historical probability distributions of traffic variables under both incident and incident-free conditions; then they proposed the use of Bayes' rule to derive the optimal thresholds.

The California algorithm (the most widely known comparative algorithm) consists of three simple comparisons to preset thresholds (*3*). An incident is detected when upstream oc-

TABLE 1 Summary of Automatic Incident Detection Algorithms

Algorithm	Traffic Variables			Time (sec) Discretization			Number of Stations	
	Vol.	Occ.	Spd.	1	5–15	30–60	Single	Adjacent
PROPOSED ALG		x				x		x
COMPARATIVE		x				x		x
TIME SERIES	x	x				x	x	x
McMASTER	x	x	x(1)			x	x	x
HIOCC		x		x(2)		x	x	
WILLSKY	x	x			x(2)			x(3)
CREMER	x		x		x(2)			x

(1)Optional
(2)Not typically obtained from existing loop detector systems
(3)Requires closely spaced stations along the freeway

cupancy is significantly higher than downstream occupancy both in absolute value (Test 1) relative to upstream occupancy (Test 2) and at the same time, downstream occupancy has adequately decreased during the past 2 min (Test 3). Test 3 distinguishes an incident from a bottleneck situation by indicating that a reduction in downstream occupancy has occurred over a short period of time as a result of the incident. An alarm persistence and an incident termination test may be added to improve the algorithm performance. The persistence test requires the relative spatial occupancy difference (Test 2) to remain larger than the corresponding threshold for two consecutive time periods before signaling an incident. Moreover, a termination test seeks to signal the end of the incident. In particular, after an incident has been detected, the relative spatial occupancy difference is tested during the succeeding time periods; a drop in the value of this parameter below its threshold indicates that the incident effect has terminated.

Within the family of comparative algorithms, Algorithm #7 (*3*) has attracted the attention of previous reviewers. It is similar to the California algorithm, but it replaces the temporal downstream occupancy difference in the third test of the California model with the present downstream occupancy measurement. This replacement seeks to reduce the false alarms produced by compression waves, a common feature of the Los Angeles freeway traffic, where the algorithm was originally tested.

A second general class of algorithms considers the recent history of a traffic variable and employs a time-series model to provide short-term traffic forecasts. Significant deviation between observed and forecast values are attributed to incidents. The first of three algorithms in this class, the standard normal deviation algorithm (*6*), calculates the mean and standard deviation of occupancy for the last 3 to 5 min and detects an incident when the present value differs significantly from the mean in units of standard deviation.

The double exponential algorithm (*7*) performs a double exponential smoothing of traffic occupancy to forecast occupancy and identifies as incidents the calibrated deviations (the algebraic sum of all previous-estimate errors divided by the current estimate of the mean absolute deviation). The same idea characterizes the autoregressive integrated moving average (ARIMA) algorithm, in which an ARIMA model provides short-term forecasts of the state variable (traffic occupancy) and the associated 95 percent confidence limits (*8*). An incident is detected when the observed occupancy values lies outside the confidence limits.

Unlike the other algorithms that use mainly occupancy data, the McMaster algorithm (*9*) is based on a two-dimensional analysis of the traffic data. In particular, it proposes separating the flow-occupancy diagram into four areas corresponding to different states of traffic conditions. Incidents are detected after observing specific changes of the traffic state in a short time period. This approach requires calibration of the boundaries separating different traffic conditions—algorithm thresholds—individually for each station, as volume-occupancy characteristics vary across stations. Simplicity of design and a potential for improved detection performance are the major advantages of the algorithm.

Besides the previous approaches, which use aggregate traffic data averaged over 30 to 60 sec, Collins et al. (*10*) developed the HIOCC algorithm for the British U.K. Transport and Road Research Laboratory (TRRL) on the basis of 1 sec instantaneous occupancy data. The algorithm seeks several consecutive seconds of high detector occupancy in order to identify the presence of stationary or slow-moving vehicles over individual detectors. A weakness of this method is the lack of an effort to distinguish incidents from other congestion-producing traffic phenomena (e.g., compression waves).

The above algorithms use only detector output to make a decision, but other methods take advantage of insights gained from research in traffic flow modeling. Willsky et al. (*11*) proposed using macroscopic traffic modeling to describe the evolution of spatial-average traffic variables (velocities, flows, and densities), thus capturing the dynamic aspect of the traffic phenomena to alleviate the false alarm problem. Although scientifically appealing, the two methods resulting from that research did not attract the interest of practitioners. The lack of interest was owed to the complexity of the methods and the strong data requirements, in terms of type of variable (density, space mean speed) and short time-space discretization. These restrictions limited testing of the methods to a small number of simulated incident patterns.

Cremer has proposed a similar approach applicable to congested cross-country freeways in Europe, where detectors are located several kilometers apart (*12*). Whereas Willsky models an incident as having a capacity reduction effect, Cremer proposes that detection can be improved by modeling the attenuation of the road capacity with an additional (fictitious) volume input at the location of the incident.

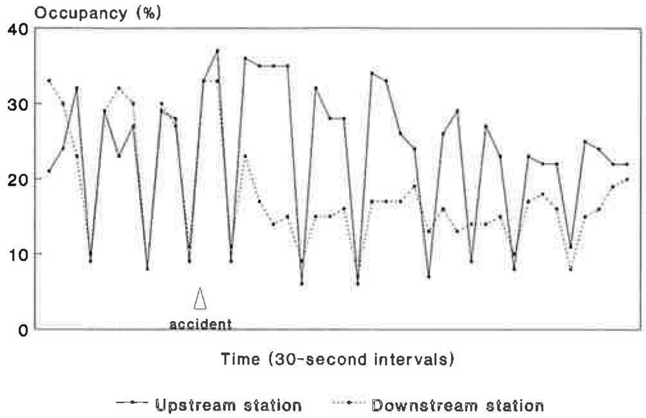

FIGURE 1 Noisy detector occupancy data.

Kühne (*13*) proposed the use of high-order continuum models to calculate the standard deviation of the speed distribution and found that the deviation broadens when density approaches the critical value in which stop-start traffic movement is observed. He concludes that detecting such broadening could be the basis for an early warning strategy and incident detection. However, his work has not produced a practical incident detection algorithm.

Despite several appealing features, existing algorithms are characterized by certain limitations. These limitations result mainly from two sources: (a) the use of raw data with only limited filtering and (b) the lack of effort, or effectiveness of effort, in distinguishing incidents from incident-like traffic situations.

More specifically, our experience suggests that raw detector data are often inappropriate for incident detection if care is not taken to filter out the noise before use—a weakness characterizing comparative algorithms. In particular, when corrupted by noise (see Figure 1) incident patterns in the traffic data may not be detected easily by an algorithm. Similarly, fluctuations produced by noise sources can be detected as incidents.

In addition, comparative algorithms are not very effective in detecting incident patterns. In an attempt to increase the success of differentiating incidents from recurrent congestion, these algorithms use tests that are somewhat restrictive. As a result, the only traffic patterns easily identified are those occurring under severe incident conditions and satisfying every test of an algorithm.

Statistical forecasting algorithms employ filtering as dictated by their design specifications. However, either they use simplified models or their models depend on excessive parameter estimation at each new site; this limits algorithm transferability. The major weakness of these algorithms lies in their lack of effort in distinguishing incidents from similar traffic patterns. In particular, the algorithms focus on detecting abrupt changes in the traffic stream without involving more-specific tests that classify the source of the abrupt change.

ALGORITHM DESCRIPTION

The proposed logic seeks to reduce the number of false alarms that can result from short-term traffic inhomogeneities. More specifically, it uses a moving average to filter the raw data before algorithm implementation and performs detection on the basis of spatial occupancy difference between adjacent stations. Attention is given to keeping the new algorithm structure as simple as possible. In particular, the algorithm does not use highly specific tests that would seek to account for given traffic abnormalities. From the authors' review of existing algorithms, highly specific tests are strongly data-related and can lead to decreased algorithm performance when transferred across data sets.

According to the new algorithm, an incident is likely to create congestion in the upstream and reduce flow in the downstream detector station; this leads to a high spatial occupancy difference between the two stations. Unlike existing algorithms, which consider such difference at single time intervals, the new algorithm uses a 3-min average (y_t, corresponding to six 30-sec measurements) of the spatial occupancy difference, x_t, between adjacent stations for the period following the incident occurrence time t:

$$y_t = 1/6 \sum_{k=0}^{5} x_{t+k} = 1/6 \sum_{k=0}^{5} (\mathrm{OCC}^u_{t+k} - \mathrm{OCC}^d_{t+k}) \tag{1}$$

where OCC^u_t is the upstream station occupancy at time t, and OCC^d_t is the downstream station occupancy at time t.

The averaging technique ensures that high values of y_t are not random but reflect congestion in the freeway segment.

If used alone, the preceding variable may lead to detection errors, since it does not consider the past flow condition. Incorporating the recent past information is crucial, especially in locations with special geometrics. For instance, in a lane-addition location, upstream occupancy is typically higher than the downstream occupancy (even under normal traffic conditions). To account for such permanent traffic abnormalities, y_t is compared with the average spatial occupancy difference, z_t, observed in the 5 min before the incident:

$$z_t = 1/10 \sum_{j=1}^{10} x_{t-j} = 1/10 \sum_{j=1}^{10} (\mathrm{OCC}^u_{t-j} - \mathrm{OCC}^d_{t-j}) \tag{2}$$

Significant difference between y_t and z_t indicates a temporal change in the state of traffic, signaling an incident.

Finally, because the values of y_t and z_t depend not only on the incident severity but also on the local flow level at the time of the incident, the foregoing variables are normalized to reflect relative changes with respect to existing traffic conditions. The normalization is done by m_t, the maximum of upstream and downstream station occupancy averaged over the most recent 5-min period before the incident:

$$m_t = 1/10 \max\left(\sum_{j=1}^{10} \mathrm{OCC}^u_{t-j}; \sum_{j=1}^{10} \mathrm{OCC}^d_{t-j}\right) \tag{3}$$

In summary, the algorithm logic includes two tests: first, when

$$\frac{\sum_{k=0}^{5} x_{t+k}}{m_t}$$

exceeds a given threshold, congestion is detected. If the congestion test is satisfied, a second test examines whether

$$\frac{\sum_{k=0}^{5} x_{t+k} - \sum_{j=1}^{10} x_{t-j}}{m_t}$$

exceeds a second threshold. If true, the test indicates an incident rather than recurrent congestion.

The selection of the pre- and post-incident period lengths for data averaging at 5 and 3 min, respectively, reflects our findings from preliminary analysis of the data. In particular, 5-min averaging removes high-frequency fluctuations of traffic flow without modifying the overall preincident traffic pattern. However, a shorter postincident averaging period of 3 min is needed to prevent high delays in the algorithm response time.

The proposed logic, while retaining simplicity, offers certain advantages over conventional logic. First, because of data filtering, the new algorithm substantially reduces the risk of false alarms from random fluctuations in traffic flow. The filtering also provides a more efficient persistence test than the tests employed by the comparative algorithms, since it accounts for six time periods instead of two. In addition, unlike traditional algorithms of the comparative type, this algorithm does not require the formation of a highly specific incident pattern to ensure detection. Nevertheless, a more sophisticated algorithm could benefit from considering the characteristics of each individual station. Such an algorithm could also benefit from a more effective effort in distinguishing incidents from other traffic patterns (e.g., compression waves) that resemble incidents, an issue currently being addressed by the authors.

DATA DESCRIPTION

All data for algorithm testing were collected from Interstate 35W, a heavily traveled and often congested freeway in Minneapolis, Minnesota (Figure 2). The selected freeway segment is fully instrumented with television cameras (camera locations are shown in Figure 2 with a circle) that allow detailed incident information to be gathered. The study was confined to the afternoon peak period (4:00 p.m. to 6:00 p.m.), since incident detection under moderate to heavy traffic conditions is of greatest importance for advanced freeway management. Traffic data, routinely collected by the traffic management center, consist of 1-min volume and occupancy counts that are updated every 30 sec and averaged over all lanes.

Information on actual incidents is available from the incident logs filed by the traffic management center operator daily. These logs report the time and location of each incident, its type, duration, severity, and impact on traffic, as well as the roadway condition and other incident-related information. Detection of incidents is mainly accomplished through observation of television monitors by on-duty traffic personnel, map displays, scanners, and state patrol reports.

Algorithm development and testing used 140 hr of traffic data from 72 weekdays. The data were obtained from 14 single detector stations along a 5.5-mi segment of I-35W. This segment includes most types of geometric configuration usually found in a freeway: entrance and exit ramps (often carrying heavy volumes) lane drop and addition, and freeway intersection. The complete data set consists of about 240,000 data points.

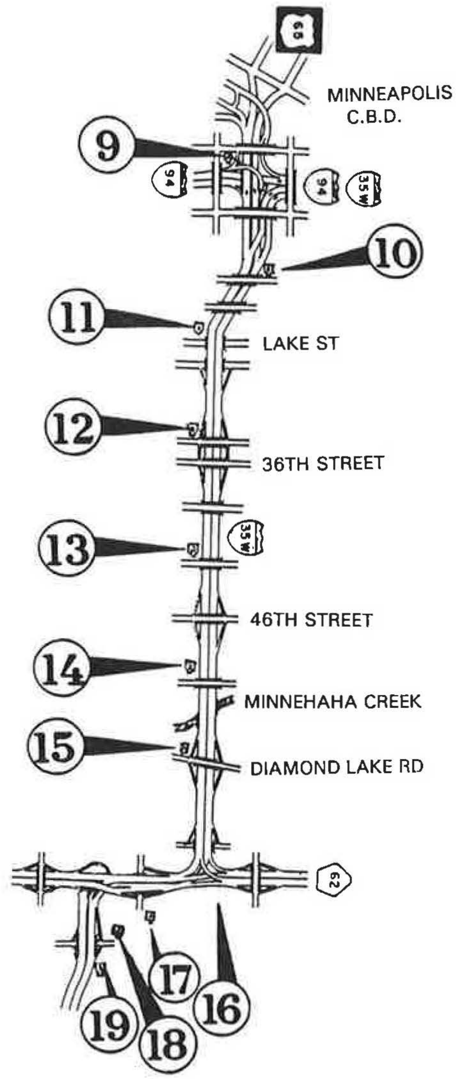

FIGURE 2 Study site in Minneapolis I-35W (Source: Traffic Management Center, Minnesota Department of Transportation).

Detailed incident data included 27 capacity-reducing incidents, primarily accidents and stalls in the moving lanes and shoulders, reported during the study period by the traffic management center on the I-35W segment. The data impact on traffic operations varied from limited to very severe. Two additional incidents (stalls on the shoulder) classified by traffic operators as causing no impact on traffic were excluded from the analysis, because detection is based solely on observable changes in the traffic flow stream.

ALGORITHM EVALUATION

Evaluation of incident-detection algorithm performance generally uses three major indicators: detection rate, false alarm rate, and mean time to detect.

- Detection rate is the ratio of incidents detected out of all incidents that occur during a specified time period.
- False alarm rate is the ratio of false alarms out of all decisions (incident and nonincident) made by the system during a specific time period. Certain authors also employ an online definition of false alarm rate as the percent of false decisions out of all incident decisions during a certain period of time (*5*).
- Mean time to detect is the average amount of time required by the system to make a detection.

The preceding measures of effectiveness are related because, at least in the single-test algorithms, increasing the detection rate causes the false alarm rate to increase. No standards have yet been adopted for determining the best combination of detection and false alarm rates. The lack of standards lies primarily in the difficulty of reconciling the difference between the consequences of a missed incident and those of a false alarm, so that the average cost of incident detection is minimized.

The need for evaluating existing algorithms is evident in the literature, which does not address conclusively the potential for transferability of evaluation results. Our findings indicate that evaluation results are not always transferable within an acceptable error range, probably because of varying traffic conditions, weather, and driver characteristics across application sites. Differences across sets of incident data are an additional reason for the lack of transferability. These differences result primarily from varying assumptions on whether incidents with very limited impact on the traffic should be included. Therefore, unbiased evaluation of a new algorithm requires concurrent evaluation of major existing ones on a common data set.

We can increase the potential for robustness and transferability of comparative-evaluation results by employing a unified system for the evaluation through an operating characteristic curve (*2,3,7,8,14*). Such a curve is a plot of detection probability (or rate) P_D versus false alarm probability (or rate) P_F that an algorithm can achieve. To construct this curve, the threshold parameters are allowed to vary over a wide range of values. Every threshold (or threshold set) produces a performance point (P_D, P_F) on the curve. The main advantage of this technique for comparing detection algorithms lies in its independence from algorithm structure. Consequently, algorithms can be compared to each other without regard to the number of tests, traffic variable involved, or type of algorithm. Additionally, the operating characteristic curve covers the whole range of detection rate (0 to 100 percent), allowing the algorithm user to establish the thresholds that best meet the requirements for a traffic operation.

Existing Algorithms

In this work, we performed detailed testing of two general algorithm types: comparative (California logic) and time-series. We did not include other approaches in the evaluation because they either require information currently not obtained routinely from loop detectors at most U.S. test sites (HIOCC algorithm) or do not provide explicit algorithms to be implemented with real data (Willsky, Kühne). We are in the process of testing the McMaster algorithm and plan to incorporate its most appealing features in more sophisticated detection algorithms under development.

The performance of two comparative algorithms—the California algorithm and Algorithm #7—is shown in Figure 3 in terms of their operating characteristics. A common feature of these algorithms is that each employs three traffic variables in its structure. Thus, increasing some thresholds while reducing others can maintain the same detection rate at different false alarm rates and vice versa. The lack of one-to-one correspondence between detection and false alarm rates in the case of multiple variables leads to the scattered performance pairs (P_D, P_F) as shown in Figure 3. In practice, high sensitivity to threshold selection reduces the algorithm attractiveness, since performance can easily deteriorate if the threshold set is poorly selected.

Three time-series algorithms, the standard deviation, the double exponential, and the ARIMA, are compared in Figure 4 employing station occupancy data. The standard deviation algorithm uses a 5-min time base to calculate averages and standard deviations and requires persistence of alarm for two consecutive time periods to signal an incident. The values for the smoothing constants in the double exponential and the ARIMA algorithms were taken as recommended (*7,8*). Updated values of these parameters could lead to improved algorithm performance. The evaluation reveals that at low detection rates the standard deviation exhibits the best performance, whereas at high detection rates the best performance is exhibited by the double exponential algorithm.

Test results also indicate that spatial occupancy difference (the key variable in comparative algorithms) is more sensitive to incidents than single station occupancy and leads to improved algorithm performance. This sensitivity differential, exhibited by all tested algorithms, is shown in Figure 5 for the double exponential algorithm.

The performance evaluation results reveal a superiority of the comparative algorithms to time-series. Figure 6 shows the operating characteristic curves, summarizing the performance of algorithms from both classes (California, Algorithm #7, standard deviation, and double exponential), as well as the new algorithm; its detection performance is discussed in the next section. The curves of the time-series algorithms are based on spatial occupancy difference data so that only their improved performance is considered. For both comparative

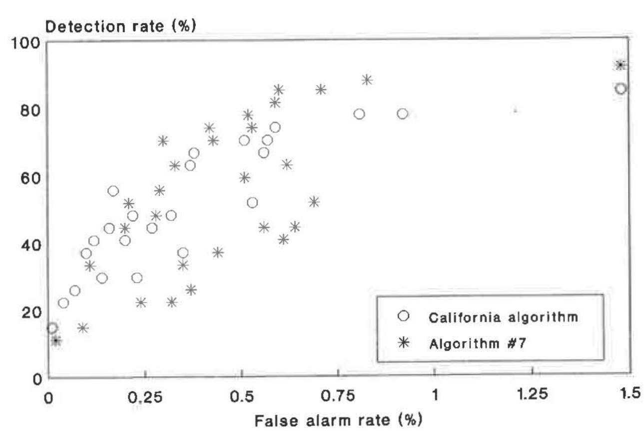

FIGURE 3 Operating characteristics of comparative algorithms.

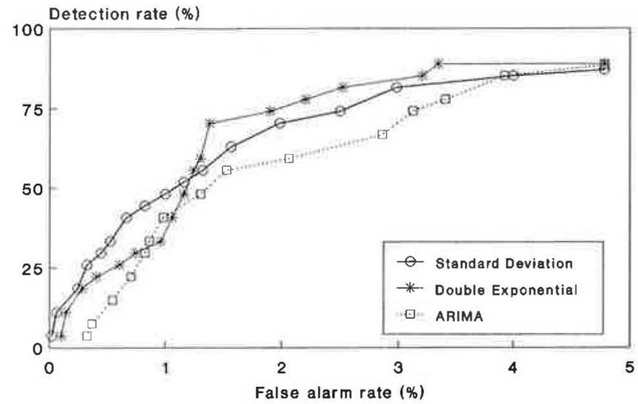

FIGURE 4 Operating characteristic curves: time-series algorithms.

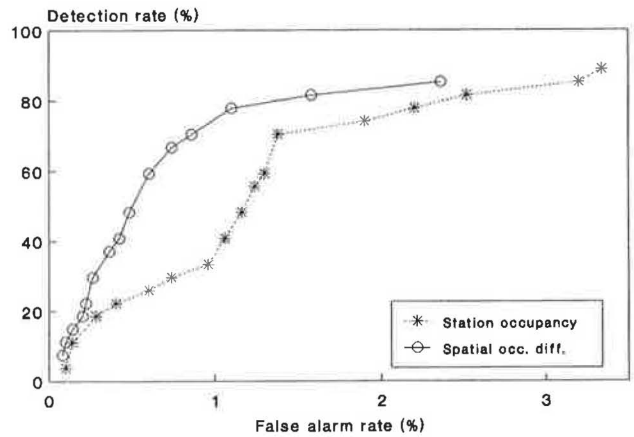

FIGURE 5 Performance comparison between two detection parameters.

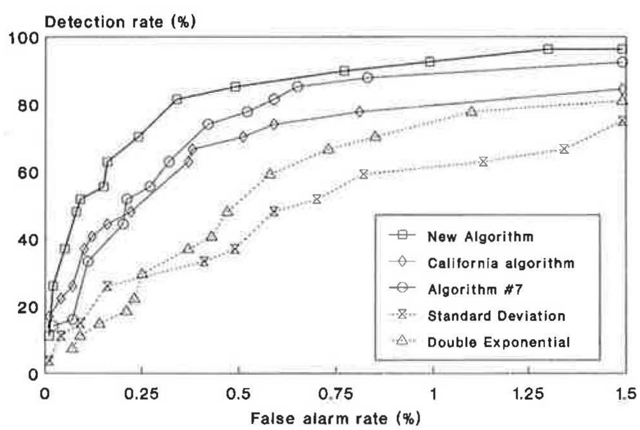

FIGURE 6 Algorithm performance comparison.

algorithms, when a specific detection rate can lead to several false alarm rates (see scattered plots in Figure 3), only the superior operating point is included in Figure 6; the assumption is that, in practice, the thresholds that lead to superior performance can be obtained through testing. As the figure illustrates, the false alarm rates of the comparative algorithms are 30 to 60 percent lower than those of the time-series algorithms.

All conventional algorithms exhibit performance that at first glance may appear to be acceptable for implementation. However, a better feeling of this performance can be gained by considering the absolute number of false alarms instead of the false alarm rate. For example, among conventional algorithms, at 50 percent detection rate, the lowest false alarm rate equals 0.21 percent—achieved by Algorithm #7. Such a false alarm rate in a sample of 140 hr of 30-sec traffic data from 14 detector stations corresponds to approximately 3.5 false alarms per hour for the relatively small 5.5-mi freeway segment of our application.

Figure 7 shows the mean time-to-detect performance of the above algorithms. The standard deviation presents the lowest response time in signaling incidents, 1 to 1.5 min earlier than comparative algorithms. The double exponential tends to respond slowly to traffic changes; this causes excessive delays in detecting incidents. The negative time-to-detect values in the figure result from the fact that time of incident is the instant when the operator identifies that incident, usually a few minutes after its occurrence. For instance, a negative value of mean time to detect implies that the algorithm detects the incident before the operator does.

For examining the transferability of evaluation results across application sites, we compared our findings to those from previous performance studies. In particular, we considered earlier findings on the California algorithm and Algorithm #7. The analysis revealed that the operating characteristic curves of the California algorithm produced by the earlier (*3*) and present studies almost coincide, indicating a potential for transferability. However, the performance of Algorithm #7 in the original study was different from that in our tests. Algorithm #7 was designed specifically with emphasis on eliminating false alarms resulting from compression waves, as, for instance, in the Los Angeles freeway system. Performance may be expected to deteriorate in the Minneapolis–St. Paul freeways on which compression waves are fewer and less severe.

New Algorithm

The new algorithm is compared with the best-performing existing ones from the previous section via their operating char-

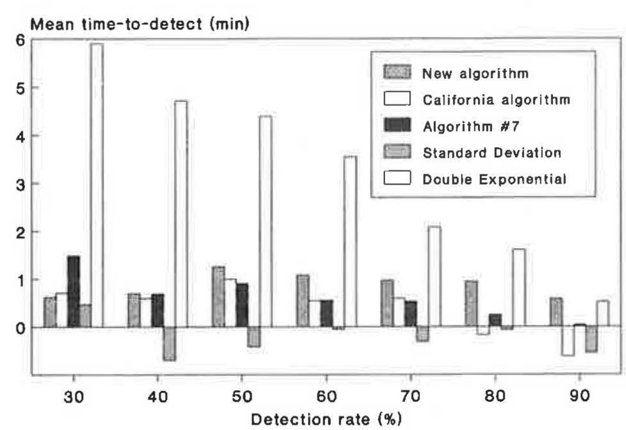

FIGURE 7 Mean time-to-detect performance.

acteristic curves (see Figure 6). At all levels of detection rates, the new algorithm is superior to all existing algorithms. As shown in the figure, the new algorithm yields detection rate of 50 percent at a 0.1 percent false alarm rate, whereas best-performing Algorithm #7 produces more than twice the number of false alarms at the same detection rate. Similarly, the new algorithm achieves a 0.3 percent false alarm rate at 80 percent detection rate, whereas Algorithm #7 results in a 0.6 percent false alarm rate at the same detection rate. As the figure suggests, the new algorithm produces 50 to 70 percent fewer false alarms than the California algorithm, 30 to 60 percent fewer than Algorithm #7, and 70 to 80 percent fewer than the standard deviation and double exponential. In practical terms, although better than that of the existing algorithms the performance of the new algorithm may not be operationally acceptable. For instance, at a 50 percent detection rate the new algorithm produces about 1.5 false alarms per hour within the 5.5-mi segment. The false alarm rate could decrease further through improved traffic flow modeling that could distinguish more effectively incidents from incident-like traffic patterns, an issue being addressed by the authors.

The new algorithm also exhibits acceptable mean time to detect. As Figure 7 indicates, although its structure imposes a 3-min detection delay, its mean time to detect remains close (by less than 1 min) to the detection times required of the comparative algorithms. The competitive response time of the new algorithm can be expected, as the algorithm begins sensing an incident from the onset of congestion (with a 3-min delay); on the other hand, comparative algorithms initiate an alarm only after the incident-generated congestion reaches a level that is sufficiently high for the test variables to exceed the thresholds.

CONCLUSIONS

Conventional automatic incident detection algorithms and the new algorithm were evaluated on the basis of their operating characteristic curves. The evaluation revealed that comparative algorithms, employing three test variables, can distinguish incidents from other traffic phenomena more effectively than single-variable time-series algorithms that use statistical forecasting of traffic. At all detection levels, the comparative algorithms produce 30 to 50 percent fewer false alarms than time-series algorithms.

Test results with the new algorithm indicate a decrease of 30 to 70 percent in false alarm rates compared to comparative algorithms and a 70 to 80 percent reduction compared to time-series algorithms. In addition, the new algorithm presents a mean time to detect comparable to that of existing algorithms.

To avoid false alarms that are primarily due to short-term traffic inhomogeneities, the new algorithm uses filtered detector output (i.e., values averaged over short time periods). It is simple to implement and less sensitive to random fluctuations of traffic than existing algorithms, and it requires only the minimum amount of data routinely provided by current presence detector systems.

Comparison of our results with findings from the literature indicates that algorithm performance may exhibit varying degrees of transferability across test locations. Whereas transferability potential increases for algorithms designed with general traffic behavior in mind, it deteriorates when algorithms involve specific tests to account for traffic phenomena (e.g., compression waves) that do not exhibit the same frequency of occurrence and severity across test sites.

Although the new algorithm improves on the performance of conventional detection methods, it is still restricted by modeling and hardware limitations. In order to achieve a lower false alarm rate appropriate for operational use, further research is being pursued by the authors, addressing the need for improved traffic modeling, more effective distinction between incident and nonincident alarms, and use of machine vision techniques for the derivation of traffic data.

ACKNOWLEDGMENTS

This study was supported by the National Science Foundation. The Minnesota Supercomputer Institute provided partial support. The Center for Transportation Studies, Department of Civil and Mineral Engineering, University of Minnesota is acknowledged for its support. The Traffic Management Center, Minnesota Department of Transportation, cooperated in the study by providing the necessary data.

REFERENCES

1. Y. J. Stephanedes and K. Chang. Optimal Ramp-Metering Control for Freeway Corridors. *Proc., 2nd International Conference on the Applications of Advanced Technologies in Transportation Engineering*, Minneapolis, Minn., Aug. 1991, pp. 172–176.
2. Y. J. Stephanedes and A. P. Chassiakos. A Low Pass Filter for Incident Detection. *Proc., 2nd International Conference on the Applications of Advanced Technologies in Transportation Engineering*, Minneapolis, Minn., Aug. 1991, pp. 378–382.
3. H. J. Payne and S. C. Tignor. Freeway Incident Detection Algorithms Based on Decision Trees with States. In *Transportation Research Record 682*, TRB, National Research Council, Washington, D.C., 1978, pp. 30–37.
4. M. Levin and G. M. Krause. Incident Detection Algorithms, Part 1: Off-Line Evaluation; Part 2: On-Line Evaluation. In *Transportation Research Record 722*, TRB, National Research Council, Washington, D.C., 1979, pp. 49–64.
5. M. Levin and G. M. Krause. Incident Detection: A Bayesian Approach. In *Transportation Research Record 682*, TRB, National Research Council, Washington, D.C., 1978, pp. 52–58.
6. C. L. Dudek and C. J. Messer. Incident Detection on Urban Freeways. In *Transportation Research Record 495*, TRB, National Research Council, Washington, D.C., 1974, pp. 12–24.
7. A. R. Cook and D. E. Cleveland. Detection of Freeway Capacity-Reducing Incidents by Traffic-Stream Measurements. In *Transportation Research Record 495*, TRB, National Research Council, Washington, D.C., 1974, pp. 1–11.
8. S. A. Ahmed and A. R. Cook. Discrete Dynamic Models for Freeway Incident Detection Systems. *Transportation Planning and Technology*, Vol. 7, 1982, pp. 231–242.
9. B. Persaud, F. Hall, and L. Hall. Development and Testing of the McMaster Incident Detection Algorithm. Presented at the 69th Annual Meeting of Transportation Research Board, Washington, D.C., Jan. 1990.
10. J. F. Collins, C. M. Hopkins, and J. A. Martin. *Automatic Incident Detection—TRRL Algorithms HIOCC and PATREG*. TRRL Supplementary Report 526. U.K. Transport and Road Research Laboratory. Crowthorne, Berkshire, England, 1979.
11. A. S. Willsky, E. Y. Chow, S. B. Gershwin, C. S. Greene, P. Houpt, and A. L. Kurkjan. Dynamic Model-Based Techniques for the Detection of Incidents on Freeways. *IEEE Transactions on Automatic Control*, Vol. AC-25, No. 3, June 1980, pp. 347–360.

12. M. Cremer. Incident Detection on Freeways by Filtering Techniques. *Proc., 8th IFAC Congress*, Vol. 17, Kyoto, Japan, 1981, pp. 96–101.
13. R. D. Kühne. Freeway Control and Incident Detection Using a Stochastic Continuum Theory of Traffic Flow. *Proc., 1st International Conference on the Applications of Advanced Technologies in Transportation Engineering*, San Diego, Calif., Feb. 1989, pp. 287–292.
14. A. P. Chassiakos. *An Improved Time-Space Incident Detection Algorithm*. M.Sc. thesis. Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis, 1990.

Publication of this paper sponsored by Committee on Freeway Operations.

Abridgment

Caltrans–Cal Poly Traffic Operations Center Simulator

Stephen Hockaday, Edward Sullivan, Samuel Taff, and Clinton Staley

The California Department of Transportation (Caltrans) is developing traffic operations centers (TOCs) in many urbanized areas throughout the state. To help in planning and operating these centers, Caltrans has contracted with the California Polytechnic State University, San Luis Obispo (Cal Poly), to build and operate a TOC simulator. The project is organized in three phases. In Phase I a prototype simulator was built to clarify the type of facility required. The permanent facility was constructed in Phase II. Testing and training using the permanent simulator is currently underway in Phase III. The design, functions, and principal considerations that guided its development are described.

Urban congestion is one of the critical issues of the 1990s. Transportation professionals recognize that meeting the growing travel demand cannot be achieved by adding capacity to the system. By the year 2000, population in California is predicted to grow by nearly 26 percent, vehicle registration by 55 percent, and miles traveled by 105 percent. This is in a state that during the last decade increased new roadways by only 2 percent (*1*, p. 6).

Emphasis is being placed on improving the operating efficiency of existing highway systems. One strategy is the use of centralized control to monitor highway conditions and respond appropriately to problems as they occur. The California Department of Transportation (Caltrans) has been implementing centralized control by developing traffic operations centers (TOCs).

To effectively develop TOCs, Caltrans decided that it needed a facility for testing alternative TOC configurations, methods for presenting and processing information, and training. Simulation was the method selected to meet these needs.

TOC SIMULATOR PROJECT

The California Polytechnic State University, San Luis Obispo (Cal Poly), has created a TOC simulator as a three-phase project funded by Caltrans. Phase I developed a facility plan, equipment specifications, and computing requirements and constructed a prototype simulator facility. On the basis of experience gained with the prototype facility, Phase II constructed and calibrated the permanent simulator. In Phase III, the research team is using the simulator to represent a wide range of traffic control situations in training and strategy testing.

California Polytechnic State University, San Luis Obispo, Calif. 93407.

During each simulation session, workstation and communication equipment locations in the simulator facility can be configured to resemble the layout of particular TOCs. This can provide a training or test environment unique to each urban area in an off-line setting.

TOC CONCEPTS

Most transportation systems operate with a combination of centralized and distributed control. Centralized control systems have three necessary elements:

1. A means of detection to determine the status of the system,
2. A means of responding to system conditions and modifying the system configuration and operation, and
3. A centralized control facility.

Estimates of the potential benefits of centralized control in the highway environment are limited. New and improved methods of acquiring traffic data are becoming available rapidly. Advances in sensors, communications, and information-processing technologies offer an increased capability in centralized highway traffic control. Advances in computer support for human decision making and automated control offer additional possibilities for significant improvements.

Centralized control requires rapid analysis of system status. Often it is not clear which control strategies are most effective in responding to specific conditions. The TOC simulator provides an environment in which system control alternatives can be tested and evaluated under various scenarios.

TOCs have a wide range of beneficial control measures for urban congestion. Examples include dispatching incident response teams, diverting traffic, providing motorist information through changeable message signs (CMSs), radio, and television; and eventually directing transmission of information to on-board navigation-display units.

FUNCTIONS OF TOC SIMULATOR

The TOC simulator provides an off-line facility on which the environment and operations of an actual TOC can be modeled without affecting real traffic. The activities within the Simulator can be characterized as input, output, and processing. Each of these activity sections has specific functions associated with it.

Flow Monitoring

The flow monitoring function is part of the input activity section and allows the TOC to detect, verify, and monitor incidents. Input sources for the simulator represent a variety of methods including loops, closed-circuit television (CCTV), calls from the California Highway Patrol (CHP) and Caltrans field units, and calls from the general public. Loop input is provided by a traffic simulation program developed at Cal Poly. Outputs from this simulation are fed to a graphical display program that presents the freeway system as a network of colored sections. As different display parameters are selected, levels of speed, flow, or density are shown through changes in color.

The simulator has a computer console that replicates a direct two-way communications link to the CHP incident data base. CCTV cameras deployed in the field are represented by video tape recorders that show footage selected from a library of scenes of actual traffic conditions and incidents. Telephone and radio communications simulate exchanges with Caltrans field personnel, tow service patrols, enforcement personnel, transit operators, and so forth.

Traffic Management

Traffic management is part of the output activity section. In the simulator, TOC personnel respond to current conditions on a real-time basis. In the event of major incidents, the TOC can communicate directly with simulated response teams, media representatives, and others.

Motorist Information Services

Motorist information is another function of the TOC output activity section. TOC personnel can send information to drivers using CMSs, highway advisory radio, highway advisory telephone, computer dial-up services, and other media. As technology advances, the TOC can also simulate data transfer to on-board traveler information and navigation systems.

Traffic Analysis

The last function of the TOC is traffic analysis, part of the processing activity section. The TOC provides an environment conducive to evaluating the effectiveness of control measures and other system changes. The traffic analysis function also provides the means for archive storage and retrieval of data used in trend analysis.

DESCRIPTION OF TOC SIMULATOR FACILITY

The Cal Poly TOC simulator facility is the result of an in-depth investigation into TOC research and training needs and a review of actual TOC designs across the nation. Several layouts were evaluated, leading to the permanent facility plan shown in Figure 1. Construction of the facility began in March and was completed in November 1990.

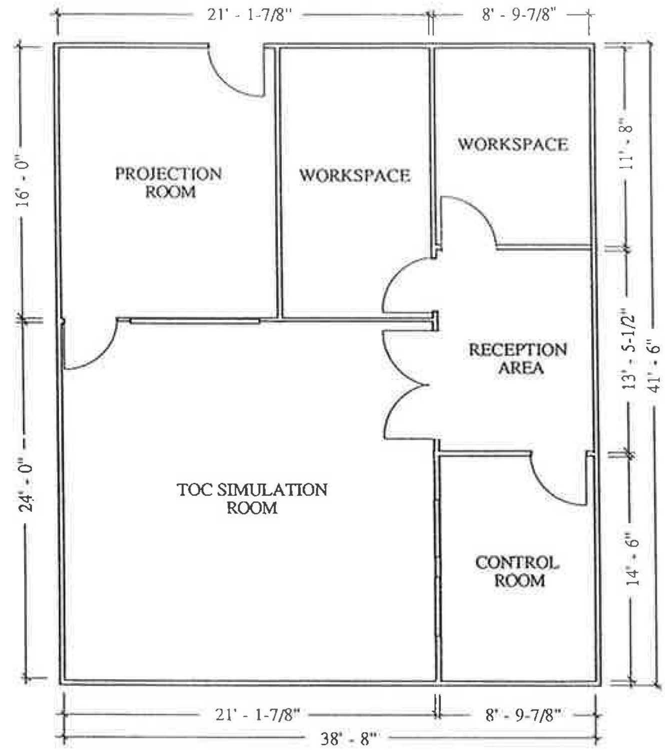

FIGURE 1 TOC facility dimensions and layout.

Two main rooms are the heart of the simulator: the simulation room and the control room. Several support rooms complete the facility.

In the simulation room, operators simulate activities that take place in a real TOC. This 24 × 21-ft room has a suspended acoustic ceiling and raised modular floor to facilitate rapid reconfiguration and placement of cables and conductors. The consoles provide equipment for flow monitoring, radio and telephone communications, computer data entry and retrieval, and CCTV network monitoring; there is also a writing surface and storage space.

Other important features of the simulator room are a CHP communications workstation that represents a link to the CHP data base for incoming information from field officers, call boxes, citizens band radios, cellular telephones, and calls by citizens. A Caltrans dispatch workstation represents a direct line to the Caltrans field dispatch center, and a rear projection display is linked to several video input sources, including the traffic conditions display, log programs from the communications console, and simulated CCTV images.

The control room, adjacent to the simulation room, contains communications, computer, and display equipment used to manage simulation activities. Audio input from recorded tape or live voice input can be transmitted to the operators in the simulation room using the public address system.

A Sun SPARC Station I workstation in the control room runs the simulation program that provides traffic input to the simulation room. The simulation supervisors can manipulate the traffic simulation program in real time to influence highway network conditions observed by operators in the simulation room. Four VCRs in the control room feed video images to the monitors in the simulation room. Finally, a "masterslave" configuration of three IBM-compatible microcomputers handles the transmission of all other information used during simulation.

COMPUTER SIMULATION OF TRAFFIC CONDITIONS

The traffic simulation is performed on a computer network that handles three overlapping tasks: (a) modeling traffic flows on the transportation network, (b) displaying traffic conditions to the TOC controllers, and (c) allowing for the interactive exchange of information between the model and the TOC. Three UNIX (Sun SPARC I) workstations with high-resolution color displays, linked via Ethernet, perform the traffic simulation and graphics processing. The UNIX environment provides substantial processing power, the opportunity for increased processing power when needed, and a high degree of portability.

The simulation and graphics software runs best on workstations with at least 10 M of memory, RISC processor performance (1 to 2 MFLOPS, 5 to 10 MIPS), a megapixel 8-bitplane color display, and 500 M of hard disk space. A black and white display can be used, but it results in a considerably degraded user interface.

The needs of the TOC simulator facility placed heavy demands with respect to software requirements. Good programs are available for both graphics and traffic simulation modeling, but none met the performance criteria and none was available for modification. Consequently, new graphics and simulation software were developed.

The graphics program displays traffic data in a maplike format and allows an operator to obtain detailed information about traffic activity. The server program generates simulated traffic data and coordinates activities between different operators. Communication between the server and graphics programs is done via UNIX Internet stream sockets, so the communication code will work properly on any network of computers running Berkeley-style UNIX.

The graphics program provides a maplike display of conditions on the freeway system. The program uses the X Windows system (X), one of the most portable of window systems available. The graphics program permits the screen to be divided into rectangular windows each showing a different portion of the freeway system. The graphics program also supports operator communication using shared bulletin-board windows. A special supervisor mode adds features so that the simulation supervisor can alter network characteristics and introduce incidents.

The server program provides simulated freeway traffic statistics on flows, speeds, and densities. It employs a standard link-node highway network representation and a macroscopic fluid-flow traffic model. Link traffic variables are recalculated at fixed time intervals of less than 1 sec.

The current first-generation traffic simulation includes only freeways with links corresponding to sections of ramps, connectors, and basic freeway sections. This program, iterating each ½ sec, can simulate more than 5,000 links in real time on an otherwise idle Sun SPARC I workstation. The model is now being reformulated to include major arterials, signalized intersections and special facilities such as high-occupancy vehicle lanes.

The freeway-only version represents link performance by rudimentary functions interrelating speed, flow, and density. Although the initial simulation for training purposes uses the Greenshield's formulation (linear speed–density relationship), any speed-density-flow relationships may be used, including tabulated functions. The simulation logic uses a simple conservation-of-flow rule and carefully accounts for how queues grow and shrink within subnetworks of congested links.

Conservation-of-flow equations such as the following simplest case are used to calculate the changes in link densities for each time interval:

$$dk_i = (q_{i-1} - q_i) * \frac{T}{L_i}$$

where

q_i = flow from a link i to the next downstream link $i + 1$ during the time interval (veh/sec),
T = length of simulation time interval (sec), and
L_i = length of link i (mi).

The average link density k_i after each time interval is obtained by adding dk_i to the k_i value from the previous interval. Unless constrained by link capacity or the presence of a queue, the flow exiting link i during the next time interval is calculated from k_i based on the link's characteristic $q - k$ function. Link speed is calculated directly from density and flow.

The calculations for actual links may be complicated by the presence of junctions where route choice is determined from turning ratios. Turning ratios are exogenous data in the first-generation simulation and will be based on a trip matrix and route-choice algorithm in later versions.

Link calculations may also be complicated by the link capacity constraint and the presence of forced flow (queuing) within the current link or within links immediately downstream. This logic, too intricate to describe here, is fully documented in the Phase II technical report (*2*).

Besides running the traffic simulation model, the server program coordinates the activities of different operators, answers requests for information from operator workstations, and ensures that bulletin-board messages are transmitted to all operators.

The TOC software was designed to be portable, modular, and easily modifiable with the anticipation of changing any portion as experience and needs dictate. The current version (Graphics 2.2 and Server 2.2) consists of 11,000 to 12,000 lines of traditional C code, in 40 to 45 files. The software utilizes X11 Release 4, the Sun equivalent of Berkeley 4.3 UNIX, and makes heavy use of the X1ib interface to X. Some portions of the code use the Athena widget set.

NEW DIRECTIONS

The trend in California toward developing primitive traffic operations centers (PTOCs) has influenced the direction of the TOC simulator project. A PTOC uses existing resources and technology to begin center operations quickly at minimum cost. PTOCs are now operational in several urban areas and will continue to be implemented throughout California.

One of the primary benefits of the TOC simulation project is to provide an off-line test bed for evaluating the components of a functioning TOC. Another related role is the development and testing of new traffic management technology.

SUMMARY

New methods to improve highway traffic operations are needed. This paper examines one specific tactic, the traffic operations center, and one specific tool, simulation. The Cal Poly TOC simulator project has achieved its stated goal of providing design information, software, and a permanent facility to develop control methods and train TOC personnel.

ACKNOWLEDGMENTS

The authors acknowledge the support of the California Department of Transportation. Although the work described in this paper was funded by the California Department of Transportation, the statements and conclusions expressed are those of the authors and not necessarily those of the State of California or Caltrans. This paper does not constitute a standard, regulation, or specification. The mention of commercial products, their sources, or their uses in connection with materials reported herein is not to be construed as an actual or implied endorsement of such products.

REFERENCES

1. *California 2000: Gridlock In the Making.* Assembly Office of Research. Sacramento, Calif., 1989.
2. *Traffic Operation Center Simulator Phase II Project Report.* ARDFA, California Polytechnic State University, San Luis Obispo, 1991.

Publication of this paper sponsored by Committee on Freeway Operations.

Evaluation of INFORM: Lessons Learned and Application to Other Systems

STEVEN A. SMITH AND CESAR PEREZ

INFORM (Information for Motorists, formerly known as the Integrated Motorist Information System, or IMIS) is a corridor traffic management system designed to obtain better utilization of existing highway facilities in a 40-mi highway corridor on Long Island, New York. The system includes integrated electronic traffic monitoring, variable message signing, ramp metering, and related strategies to optimize traffic flow through a heavily congested corridor. The evaluation of INFORM was conducted using extensive field data, perception surveys, and data collected through the system. The overall results of the evaluation are presented, including comparisons of vehicle miles of travel, average speeds, ramp delays, motorist perceptions, and other congestion-related measures for the a.m. and p.m. peak periods. Incident case studies were used to evaluate motorist response to and effectiveness of variable message signing strategies. In addition to a summary of the quantitative results, an overview of the many lessons learned in the design, implementation, operation, and evaluation of INFORM is provided.

INFORM (Information for Motorists, formerly known as the Integrated Motorist Information System, or IMIS) is a corridor traffic management system designed to obtain better utilization of existing highway facilities. INFORM has been implemented in a 40-mi (64-km) highway corridor on Long Island, New York, as an operational demonstration. This demonstration was developed in accordance with a cooperative agreement between FHWA, the New York State Department of Transportation (NYSDOT), and the transportation agencies of local governments on Long Island.

The INFORM corridor contains two major freeway facilities, the Long Island Expressway (LIE, which is Interstate 495) and the Northern State Parkway/Grand Central Parkway (NSP/GCP), as well as a number of parallel and crossing arterial streets and freeways, for a total of 128 mi (206 km) of controlled roadways. The corridor extends east from the Queens borough of New York City through Nassau County and into Suffolk County. The system consists of electronic surveillance, communications, signing, and control components, all providing motorist information for warning and route diversion, ramp control, and signal control.

The various INFORM control elements and their functions include

- Overall supervision, provided by operators in a control facility at the State Office Building in Hauppauge, New York. Three minicomputers assist with traffic flow monitoring, traffic control, and response to traffic incidents.
- Traffic monitoring, consisting of 2,400 in-roadway vehicle presence detectors and 20 roadside citizens band radio monitor units. A limited number of closed-circuit TV cameras have been installed since late 1989 to monitor traffic in construction areas. A 160-mi (258-km) coaxial cable communications network connects equipment at more than 400 roadside locations with the control facility.
- Ramp metering, provided by traffic signals at 50 freeway ramps. Roadside hardwired digital controllers operate these ramp traffic signals, which are under the supervision of one of the control center computers (or independently, in case of communications failure).
- Variable message signs (VMSs) at 74 locations to provide information to motorists about congestion and delays. The controllers for these signs are roadside microcomputers, operating under the supervision of a control center minicomputer.
- Traffic signals at 110 arterial street intersections under INFORM control. New York's Model 170 signal controllers are used at these intersections, with supervision of coordinated signal indications by one of the INFORM control center computers.

This paper presents summary information on the results of various aspects of the INFORM evaluation. The emphasis is on the overall evaluation of INFORM, lessons learned, and guidance that can be provided in the design, operation, and evaluation of traffic surveillance and control systems such as INFORM. It also presents specific information on the evaluation of the variable message signs and the ramp metering subsystem. In addition, it documents perceptions of INFORM by the public and by those responsible for its planning and implementation.

One of the difficulties in conducting a time-series type evaluation is determining which time segments to compare. In the evaluation of time periods for INFORM, it was determined that two comparisons would be of most value: March 1990 metering versus March 1990 nonmetering, and March 1990 metering versus spring 1987 metering. The comparison of the two March 1990 data sets best reflects the traffic restraint impacts of metering. The comparison of March 1990 metering with spring 1987 reflects more of the long-term change. The long-term changes could have been brought about by several factors, including change and redistribution of volume, change in commuting patterns due to metering, and motorist response to variable message sign information.

As indicated earlier, INFORM was designed as an operational demonstration of corridor traffic control technology. INFORM has broken ground, but not without difficult encounters with the reality of building a system of this scale.

S. A. Smith, JHK & Associates, 1650 Iowa Avenue, Riverside, Calif. 92507. C. Perez, Federal Highway Administration, IVHS Research Division, 6300 Georgetown Pike, McLean, Va. 22101.

INFORM continues to change and improve, but experience with the system has taught many lessons that are important in designing, constructing, and operating corridor traffic control systems. Some of the lessons have been learned the hard way—by trial and error. INFORM can also lay claim to some legitimate successes. This summary draws from both sides: it highlights the major findings, and it presents a variety of lessons learned in several areas of design, construction, and operation of INFORM.

MOTORIST INFORMATION

INFORM is one of the most advanced VMS-based motorist information systems in the United States. In addition to the benefits it has given the motoring public, it has been and will continue to be a testing ground for further improvement of motorist information strategies. Some of the specific findings and lessons learned are discussed in the following.

Impact on Delay

Variable message signs are an effective part of INFORM. The incident case studies have indicated that drivers, in fact, do modify their routes if they are consistently given accurate information. Estimated delay savings for the peak period incidents analyzed ranged up to 1,900 vehicle-hr. The estimated annual delay savings for the incident-related effects of the variable message signs is 300,000 vehicle-hr. Delay savings are also attributable to INFORM involvement in recurring traffic congestion, construction activity, and special events, but they are difficult to quantify. The availability of the signs for certain functions also eliminates the need to perform that same service in another more expensive way (e.g., nighttime closure of the LIE or NSP/GCP for construction and maintenance).

Automated Sign Message Selection

Automated sign message selection is an important part of INFORM. It is accurate within the limitations of the detector data provided by the surveillance system and is essential in allowing the operators to keep up with the information demands in a corridor the size of INFORM, particularly in the peak periods.

Commitment to VMS Operation and Information Quality

Significant operational staff time is invested in maintaining the timeliness and accuracy of the INFORM sign information. A commitment to the installation of variable message signs must be backed by a commitment to their operation. Monitoring and controlling the INFORM VMS information takes an estimated 80 percent of the operator's time (based on operator interviews), even with an automated message-selection algorithm. Advances have been made in sign control algorithms, but one cannot expect the system to run itself and maintain the quality of information that the motoring public expects. INFORM produces more than 14,000 sign messages a month in an attempt to maintain the quality of information.

Level of Diversion

Clearly, diversion is taking place in response to the sign messages. For typical incident conditions using passive messages (i.e., no recommended alternative route), 5 to 10 percent of mainline traffic in the INFORM corridor could be diverted over several upstream off-ramps (typically 3 to 4 percent at an individual ramp). Data from incident reconstructions indicated that, for a typical incident using passive messages, volume at upstream off-ramps (up to three ramps) increased by 40 to 70 percent. However, this can vary widely on the basis of location and severity of the incident, availability of alternative routes, and other factors. This suggests that motorists have some degree of faith that the INFORM information is accurate and that motorists believe faster travel can take place on an alternative route.

Diversion and Alternative Route Traffic Control Schemes

The development of effective corridor diversion schemes is heavily dependent on the ability of alternative routes to absorb traffic from the mainline. Parallel freeways such as the LIE and NSP/GCP offer an ideal opportunity for such diversion to take place, and such diversion has been identified from INFORM. The lack of traffic-responsive capabilities on parallel arterials is the greatest detriment to the potential overall effectiveness of diversion strategies. Several incident reconstructions indicated a high initial diversion to arterials, followed by arterial breakdown when capacity was exceeded.

Importance of Information Quality

Maintaining the quality of the information displayed by the signing system must be a top priority of system operation. Signing is a passive method of control that relies on an informed, voluntary decision by drivers. Motorist confidence in the system is difficult to earn and easy to lose.

Ramp Detectorization Strategies

Detectorization of all on-ramps and off-ramps, as was done on INFORM, is an important part of the signing and diversion strategy. On-ramp and off-ramp volumes are often referred to by operators to determine whether the signing messages are having an effect (or too much of an effect). Even if on-ramps are not metered, they should still be detectorized. Under budgetary constraints, this could be done selectively, with emphasis on important diversion-related ramps.

RAMP METERING

Overview of Ramp Metering Results

A.m. peak period freeway speeds for the March 1990 metering increased 3 to 8 percent over speeds for March 1990 non-metering and 13 percent over speeds for spring 1987. Certain subsections showed higher increases, and others showed lower increases or no change. Vehicle miles of travel (VMT) were either higher or remained stable for the metering case. Changes were statistically significant at the 95 percent confidence level.

P.m. peak period freeway speeds for March 1990 metering were unchanged from those for March 1990 nonmetering; they increased 13 percent over speeds for spring 1987. VMT increased approximately 1 percent over the March nonmetering case and approximately 5 percent over the spring 1987 case.

The maximum increase in throughput in a bottleneck section for the metering scenario was 7 percent. Other bottleneck sections increased by 2 to 3 percent, and still others were unchanged. Thus, ramp metering may produce marginal increases in throughput through bottleneck sections but not likely more than 2 to 3 percent, on average.

Average queues at metered ramps throughout the metering periods are relatively short, ranging from 1.2 to 3.4 vehicles over the typical 2-hr period. This represents only about 0.1 percent of the total VHT on the LIE and NSP/GCP. Contributing factors to this low number are several low-volume ramps as well as the propensity for metering to be shut off on the higher-volume ramps to avoid surface street impacts. Later versions of the ramp metering algorithm have enabled the metering operation to be preserved more frequently.

Limitations in Ramp Metering Effectiveness

Ramp metering resulted in a slight increase in average speed, but the potential effectiveness of ramp metering on INFORM is constrained by the limitations in the number of ramps metered, in the storage areas to manage queues, and in the maximum metering rates for single-lane metering. Ramp metering proved to be not as effective as was anticipated in the feasibility study. INFORM does not have sufficient ramp metering control over enough traffic to produce a noticeable, sustained change in freeway speeds. Some of the potential ramp meters were eliminated from the design, and others were eliminated by construction projects. Significant use of two-lane metering is needed to exercise greater control over high-volume on-ramps. More ramps also need to be metered, including selected freeway-to-freeway ramps, before adequate control can be established.

PUBLIC PERCEPTION

As part of the evaluation, a survey of households in the INFORM corridor was performed to measure public perception of the system. A brief summary of the results follows.

Driver Awareness of VMSs

Ninety-six percent of the residents surveyed in the INFORM area stated that they had seen the variable message signs.

Usefulness of Information

Overall, 29 percent of the respondents rated the sign information very useful; an additional 46 percent rated it moderately useful.

Accuracy of Information

Seven percent of the respondents indicated that the information was always accurate; an additional 56 percent indicated that it was usually accurate.

Changes in Route

Approximately 45 percent of the drivers stated that they sometimes change their route in response to the sign messages. Slightly more than 25 percent have never changed their route in response to a message.

Perceived Wait Time at Ramp Meters

Waiting time at the ramp meters is perceived to be 1 min or less by 80 percent of the drivers who have had to wait. This seems to correspond to the findings of the ramp delay studies.

Diversion to Avoid Ramp Meters

Some 15 percent of those encountering a red merge light indicated that they frequently use the service road or another roadway to avoid waiting. Another 27 percent indicate that they do this occasionally. Thus, ramp metering does produce some diversion effects.

Overall Perception of Ramp Metering

Approximately 40 percent of respondents viewed ramp metering to be a good idea, and another 40 percent viewed it to be a poor idea. The rest had no opinion.

Overall Perception of INFORM

Twenty-five percent of respondents viewed INFORM to be quite helpful. Forty percent indicated that the system helps once in a while. Overall, it can be concluded that drivers view the variable message signs positively, but their reaction to ramp metering is mixed.

OTHER OBSERVATIONS ON DESIGN, OPERATIONS, AND EVALUATION

Ongoing Traffic Engineering Involvement in All Phases of Operations

A corridor traffic control system cannot be expected to run itself. A commitment must be made to traffic engineering involvement in all phases, including ramp metering initiation, VMS operations, refinement and modification of metering operations, tuning of incident detection algorithms, traffic signal operations, communications with emergency services, and communications with the media.

Commitment from Top Management

Commitment from top management and constant provision of information to them is needed to sustain continuity over time. Several commissioners were involved over the course of INFORM implementation, and state personnel had to keep each one of them informed. Transition in leadership is inevitable; operators of traffic control systems should have the mechanisms in place to keep upper management and elected officials informed about the system: what it is, how it operates, and what benefits it provides.

Designing for System Evaluation Needs

There is a need for including plans for operation and maintenance in the system design phase. This should be expanded to include provisions for evaluation. In fact, a strong case can be made that surveillance needs for operation and evaluation are highly correlated, if not identical. What the system evaluator knows after the fact should also be known by the system operator as input into control decisions. For example, little information was available for the evaluation concerning arterial traffic performance. INFORM operators are also, in effect, blind to what is occurring on the arterial system. This knowledge is essential for obtaining the best use of VMSs for diversion, and the lack of detectorization undoubtedly results in the underutilization of INFORM's capabilities. Designing for evaluation needs should cost no more than designing for effective operation and, in the long run, will limit outlays for extensive field evaluation. Traffic simulation may also play a future role in bridging gaps in real-time data.

Publication of this paper sponsored by Committee on Freeway Operations.

Comparative Evaluation of Alternative Traffic Control Strategies

NATHAN H. GARTNER AND DENNIS L. HOU

A growing variety of computer models are being developed for the design of various traffic control strategies. Appropriate tools are needed to compare the performance of new strategies with old ones, as well as to determine which alternative design options work best for any given strategy. In most cases, simulation experiments are used for this purpose; in some cases field tests are also performed. The application of the Gafarian-Halati statistical estimation method for the comparative evaluation of alternative traffic control strategies is described. The method is based on the use of ratio estimation techniques. Such techniques are needed for the analysis of typical simulation model output parameters as well as for the analysis of field test data results. Application of the methodology is illustrated in a special case by comparing alternative arterial traffic signal control strategies.

A growing variety of computer models are being developed for the design of advanced traffic control strategies. Appropriate tools are needed to evaluate the performance of such strategies and to refine them before their implementation. Most important is the need to assess the potential benefits that they offer in comparison with existing strategies. In most cases, simulation models are used for this purpose, since they are the most economical method for evaluation. However, in some cases, it may be desirable to conduct field experimentations, which are the ultimate proof of performance.

To analyze the output from simulation trials as well as field tests, suitable statistical tools are needed. The application of a statistical methodology that is useful in comparing alternative traffic control strategies is described. As an example, two methods for arterial traffic signal optimization are evaluated. The methodology, of course, has general applicability and can be used to compare any alternative strategies.

SIMULATION OUTPUT ANALYSIS

Simulation methods are widely used to analyze the performance of urban traffic networks. Typically, they are applied to evaluate proposed operational changes, such as a new control or management strategy. The most widely used traffic simulation model in the United States is NETSIM. NETSIM is a microscopic, stochastic simulation model that enables the engineer to analyze and evaluate a wide range of traffic control and surveillance concepts for complex signalized street networks. The basic input requirements for this model are the network geometry, signalization information, and flow data, which consist of input rates and turning movements.

Department of Civil Engineering, University of Massachusetts, Lowell, Mass. 01854.

The simulation procedure consists of a warm-up period and the actual simulation period during which statistical data are accumulated. During the warm-up time, traffic generators feed vehicles into the empty simulated network until equilibrium conditions are reached, that is, the rate at which vehicles are fed into the network equals the rate at which vehicles are discharged from it. Only then are the simulated data valid. The NETSIM model also requires a random seed number for the random number generator that governs the flow of vehicles into the network. The simulation results depend on the choice of the seed number, so multiple replications of NETSIM runs with different seed numbers are necessary.

The standard output of the NETSIM model includes estimates of many important traffic parameters, such as

- Total vehicle minutes of travel time,
- Number of vehicles discharged,
- Total vehicle miles of travel distance,
- Total delay time,
- Average travel time per vehicle,
- Average number of stops per vehicle,
- Average speed, and
- Average delay time per vehicle.

The estimates of the traffic parameters are provided both on a link-by-link basis (links represent a one-way direction of flow on a street between two adjacent signals) and on a networkwide basis. To calculate the parameters involving average values per vehicle, NETSIM uses the ratio of sample means of observations that are, in fact, autocorrelated and cross-correlated. This may produce erroneous results. To make valid statistical statements about these parameters, one must apply statistical techniques to the model outputs that are based on ratio estimators. In this paper, we use a technique of Gafarian and Halati that provides a measure of the accuracy of the estimates in terms of confidence intervals (*1*). The technique is then applied, as an example, to compare the performance of two optimization schemes for arterial traffic signal control.

RATIO ESTIMATORS

Each of the last four measures of effectiveness (MOE) generated by NETSIM (and described earlier) is actually the ratio of the means of two random variables X and Y. For example, the average delay per vehicle is obtained by dividing the total delay time accumulated on each link by the total number of vehicles discharged from the link. The average speed is calculated by dividing the number of vehicle miles traveled on each link (equal to link length times number of exits) by the vehicle minutes on the link (the amount of time spent by all vehicles traversing the link). Similar ratios are obtained for the other measures. Consequently, the MOE itself is the ratio m_X/m_Y, that is, the actual mean of X divided by the actual mean of Y. However, the NETSIM output provides only the ratio of the sample means of X and Y random variables, and it is well known that in general $E[X/Y] \neq m_X/m_Y$.

Gafarian and Halati (*1*) developed a statistically valid method for using the ratio $\overline{X}/\overline{Y}$ as a point estimate for the ratio m_X/m_Y that provides, with a confidence interval, a measure of its accuracy. They assessed the efficacy of the method through a Monte Carlo experiment and an *M/M/1* queueing system analysis, and they discuss its applicability in the analysis of NETSIM output variables. They show in a simple example how the coverage probability of the desired results is degraded when observations from independent replications are not treated as a ratio of random variables. The coverage probability is defined as the probability that the confidence interval produced by a certain sample size will cover the true value of the ratio of the means. For the sake of completeness, the essence of the Gafarian-Halati estimation method is demonstrated. The ratio estimation techniques are also described in the general statistics literature (*2*).

The development of a confidence interval for the estimate $R = m_X/m_Y$ is based on the observations $\{(X_i/Y_i), i = 1, \ldots, n\}$ of all the independent replications of the simulation model. We define a variable Z_i as follows:

$$Z_i = X_i - RY_i$$

Taking means we obtain

$$Z = \overline{X} - R\,\overline{Y}$$

Assuming that X_i and Y_i are normally distributed, then Z_i and $\overline{Z}$ are also normally distributed. Because $E(Z_i) = E(\overline{Z}) = 0$, then the expression

$$\overline{Z}/\{(1/n)[1/(n-1)] \sum_{i=1}^{n} (Z_i - \overline{Z})^2\}^{1/2}$$

has a Student t-distribution with $(n-1)$ degrees of freedom. It can be shown that

$$[1/(n-1)] \sum_{i=1}^{n} (Z_i - \overline{Z})^2 = (S_X^2 + R^2 S_Y^2 - 2RS_{XY})$$

where

$$S_X^2 = [1/(n-1)] \sum_{i=1}^{n} (X_i - \overline{X})^2 = \text{sample variance of } X_i\text{s},$$

$$S_Y^2 = [1/(n-1)] \sum_{i=1}^{n} (Y_i - \overline{Y}]^2 = \text{sample variance of } Y_i\text{s},$$

and

$$S_{XY} = [1/(n-1)] \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$$

$$= \text{sample variance of } (X_i, Y_i)\text{s}.$$

For the $(1 - a)$ level,

$$\Pr\{|\overline{X} - R\,\overline{Y}|/[(1/n)(S_X^2 + R^2 S_Y^2 - 2RS_{XY})]^{1/2}\} \leq t_{1-(a/2),n-1} = 1 - a$$

When both sides of the argument in the probability statement are squared, the result is a quadratic inequality in the unknown $R = m_X/m_Y$, the known estimates X, Y, S_X^2, S_Y^2, S_{XY}, and $t_{1-(a/2),n-1}$. The roots of this quadratic function are

$$\left[[\overline{X}\,\overline{Y} - g(a)S_{XY}] \pm \left([\overline{X}\,\overline{Y} - g(a)S_{XY}]^2 - \{[\overline{Y}^2 - g(a)S_Y^2][\overline{X}^2 - g(a)S_X^2]\}\right)^{1/2}\right]/[\overline{Y}^2 - g(a)S_Y^2]$$

where $g(a)$ equals $(t_{1-(a/2),n-1}^2)/n$.

The $(1 - a)$ 100 percent confidence interval is then (r_1, r_2), where r_1 is the smaller root and r_2 is the larger root. Further details are given elsewhere (*1,2*).

PILOT STUDY

The validity of this method depends on how well the assumptions made in its derivation are met by the NETSIM model, namely, system in steady state, normality of numerator and denominator observations, and independent replications. Steady state in NETSIM is achieved by the warm-up procedure. To obtain independent replications, each run is started with a different seed number. The only assumption that is approximately met is that of normality, and it can be claimed that the method is not sensitive to this requirement.

We illustrate the implementation of this procedure in a pilot study of NETSIM simulation runs for a nine-signal arterial street. We performed 30 replications with a different seed number for each of two different signal control strategies, MAXBAND and MULTIBAND (which will be described further). The purpose of the pilot study was to check the normality of the random variables, the stability of the simulation results, and the sample-size requirements. The numerical results for all the replications are presented in Figures 1 and 2. Two ratio parameters were calculated: average speed (mph) and average delay per vehicle (sec). The bounds of the 95 percent confidence intervals were obtained by the procedure outline. We observe that the intervals are rather narrow; this may indicate that the number of replications can be significantly reduced and still produce meaningful outputs. The results can also be compared graphically, as shown in Figures 3 and 4. They indicate the superior performance of the particular MULTIBAND option that was chosen in this case: an increase of 5.6 percent in average speed and a reduction of 15.2 percent in average delay per vehicle, compared with MAXBAND. We also plotted the frequency histogram for both travel distance and travel time. If a sequence of observations is normally distributed, then the frequency histogram is bell-shaped. Figure 5 illustrates that this assumption is approximately true in this case. In the next section we show that we can obtain meaningful results with a much smaller sample size. This is particularly helpful when we need to compare a large number of alternative strategies.

Simulation Result from NETSIM with 95% confidence interval t(0.975;29)=2.045
CANAL STREET

METHOD WEIGHT	TOTAL MILES [MILES]	TOTAL TIME [HOURS]	TOTAL DELAY [SEC]	TOTAL VEHICLES	AVERAGE SPEED M.P.H.	AVERAGE DELAY [SEC/VEH]
MULTIBAND	651.77	38.44	43896	1776	16.96	24.72
(VOLUME/CAP.)4	653.76	38.39	43362	1780	17.03	24.36
RATIO	653.48	38.56	43344	1778	16.95	24.38
	651.14	38.04	42510	1774	17.12	23.96
	653.78	38.48	43560	1780	16.99	24.47
	651.43	38.35	43764	1777	16.98	24.63
	652.61	38.04	42594	1780	17.16	23.93
	652.91	38.81	44718	1784	16.83	25.07
	652.91	38.81	44718	1784	16.83	25.07
	652.25	37.78	42348	1780	17.27	23.79
	652.61	38.32	43530	1777	17.03	24.50
	652.11	38.46	43656	1778	16.96	24.55
	652.80	38.24	42882	1783	17.07	24.05
	652.80	38.24	42882	1783	17.07	24.05
	652.45	38.07	42636	1775	17.14	24.02
	652.80	38.24	42882	1783	17.07	24.05
	652.61	38.22	43020	1777	17.08	24.21
	652.17	38.35	42984	1777	17.00	24.19
	653.19	38.23	42504	1778	17.09	23.91
	653.44	38.04	42942	1777	17.18	24.17
	652.18	38.40	43566	1775	16.98	24.54
	653.16	38.43	43158	1779	16.99	24.26
	652.11	38.06	43080	1776	17.13	24.26
	652.71	38.31	43320	1777	17.04	24.38
	652.96	38.61	44268	1778	16.91	24.90
	653.20	38.27	43734	1780	17.07	24.57
	654.12	38.55	43776	1784	16.97	24.54
	652.80	38.22	43044	1778	17.08	24.21
	652.16	38.38	43212	1779	16.99	24.29
	654.06	38.02	42762	1778	17.20	24.05
AVERAGE	652.75	38.31	43288.40	1778.83	17.04	24.34
STD	0.72	0.23	605.41	2.83		
95% CONFIDENCE INTERVAL						
LOWER BOUND					17.00	24.21
UPPER BOUND					17.08	24.46

FIGURE 1 MULTIBAND NETSIM runs (30 replications).

Simulation Result from NETSIM with 95% confidence interval t(0.975;29)=2.045
CANAL STREET

METHOD WEIGHT	TOTAL MILES [MILES]	TOTAL TIME [HOURS]	TOTAL DELAY [SEC]	TOTAL VEHICLES	AVERAGE SPEED M.P.H.	AVERAGE DELAY [SEC/VEH]
MAXBAND	651.02	40.20	50790	1767	16.19	28.74
TOTAL VOLUME	651.52	40.24	50142	1765	16.19	28.41
RATIO	652.20	40.43	50802	1768	16.13	28.73
	652.81	40.69	51366	1771	16.04	29.00
	651.56	40.35	50880	1768	16.15	28.78
	650.47	40.41	51012	1767	16.10	28.87
	651.47	39.94	49758	1766	16.31	28.18
	652.02	40.33	50754	1765	16.17	28.76
	650.47	40.41	51012	1767	16.10	28.87
	651.64	40.50	51114	1767	16.09	28.93
	650.93	40.39	51168	1765	16.11	28.99
	650.53	40.48	51732	1767	16.07	29.28
	651.07	40.14	49842	1765	16.22	28.24
	650.47	40.41	51012	1767	16.10	28.87
	650.41	40.70	51426	1766	15.98	29.12
	649.43	40.24	51144	1764	16.14	28.99
	650.92	40.38	50742	1763	16.12	28.78
	649.62	40.20	50676	1765	16.16	28.71
	650.39	40.36	51174	1765	16.12	28.99
	650.39	39.90	49314	1766	16.30	27.92
	650.56	40.30	50640	1767	16.14	28.66
	650.50	40.26	49908	1768	16.16	28.23
	651.80	40.38	50640	1767	16.14	28.66
	650.27	40.25	50574	1765	16.16	28.65
	649.32	40.57	51708	1765	16.00	29.30
	651.90	40.38	50496	1769	16.15	28.54
	650.66	40.35	50982	1767	16.13	28.85
	651.73	40.16	50346	1767	16.23	28.49
	648.44	39.91	49380	1764	16.25	27.99
	650.10	40.39	50712	1766	16.10	28.72
AVERAGE	650.82	40.32	50708.20	1766.30	16.14	28.71
STD	0.94	0.19	607.58	1.64		
95% CONFIDENCE INTERVAL						
LOWER BOUND					16.03	28.58
UPPER BOUND					16.25	28.84

FIGURE 2 MAXBAND NETSIM runs (30 replications).

CASE STUDY

To illustrate the application of the methodology described, it is used in the comparative analysis of two optimization schemes for arterial traffic signal control. Since the two schemes have multiple design options, each of these options is evaluated as an alternative control strategy. The purpose of this comparison is to evaluate the benefits that can accrue from the implementation of an improved design alternative. A similar approach can be used in the comparative evaluation of other types of operational changes, alternative management strategies, or new control designs.

The base strategy that we evaluate is a traditional bandwidth maximization method. Such methods are widely used by traffic engineers to optimize signal settings in urban arterial streets (*3,4*). The method we are using is MAXBAND (*3*). This method is based on mixed-integer linear programming optimization and calculates the cycle time, offsets, link progression speeds, and left-turn phase patterns that maximize

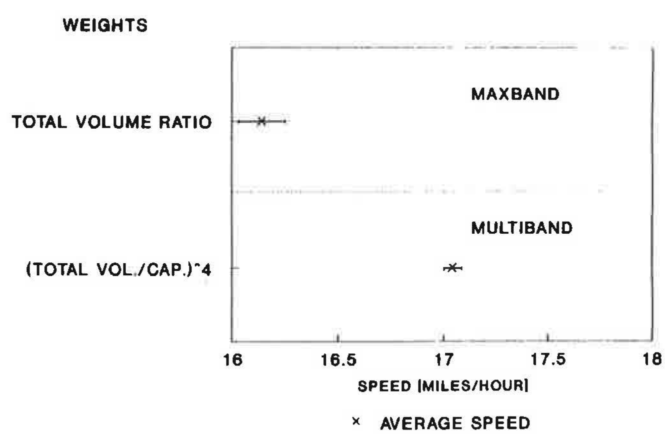

FIGURE 3 Average speed comparison, Canal Street, pilot study (30 NETSIM runs).

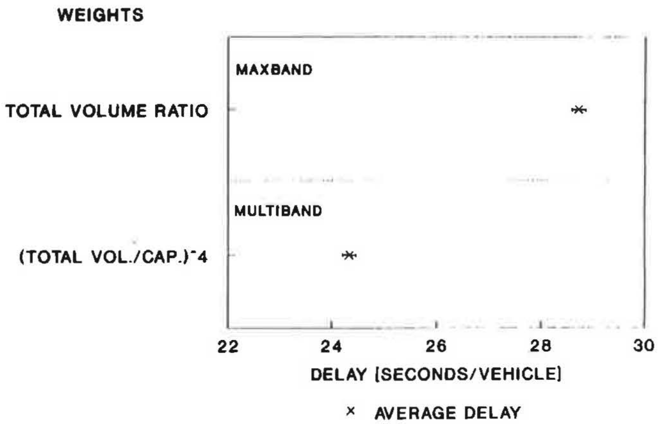

FIGURE 4 Average delay comparison, Canal Street, pilot study (30 NETSIM runs).

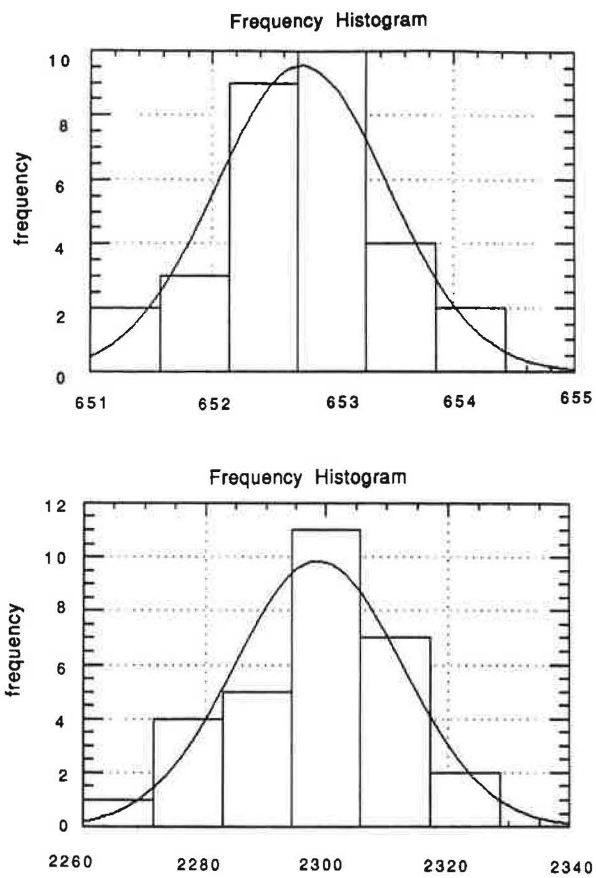

FIGURE 5 Normality check via histogram plot: *top*, total travel distance (miles); *bottom*, total travel time.

a weighted combination of the bandwidths b, $\bar{b}$ in the two directions of travel along the artery. Continuous bands of equal width are produced by this method. The MAXBAND optimization program is summarized in Figure 6. Similar designs can be produced by the PASSER II optimization program (*4*).

A basic limitation of existing bandwidth-based programs is that the progression design criterion does not depend on the

MAXBAND	Given:	splits, queue clearances, target ratio of arterial bandwidths, and for each section limits on: cycle time, link speeds, and changes in speeds.
	Find:	cycle time, offsets, interferences, arterial bandwidths b and $\bar{b}$, link progression speeds and left-turn phase patterns
	To:	maximize $b + k\bar{b}$
	Subject to:	cycle time constraint,bandwidth ratio constraints, interference constraints, loop integer constraints, andspeed and speed-change constraints

FIGURE 6 MAXBAND optimization program (*3*).

actual traffic flows on the arterial links and, therefore, is insensitive to variations in such flows. The total bandwidth obtained for the arterial can be allocated in any desired ratio among the two directions of travel. A common practice is to apportion it according to the directional volume ratio k, taken as the ratio between the average (or total) link volumes in each direction.

However, because of turn-in and turn-out traffic we generally do not have constant volumes along each direction of the arterial.Consequently, the idea of a uniform platoon moving through all the signals in one direction, which forms the conceptual basis for the bandwidth approach, does not always hold. Moreover, the ratio of volumes on opposing road sections between each pair of adjacent signals is also varying. It is, therefore, inconceivable that a single parameter for the entire arterial (k) can adequately reflect this diversity and guarantee the best progression (in terms of delays and stops) for different traffic flow patterns.

This has led to the development of a new optimization approach named MULTIBAND (*5*), which is designed to remedy such deficiencies. The approach places the arterial bandwidth optimization concept on a more solid foundation by incorporating into the calculation procedure a systematic traffic-dependent criterion. The volume on each link of the artery, along with other traffic parameters (such as capacity and speed), will have an effect on the optimization outcome through suitably chosen link-specific weighting factors, as contrasted with a single weight for all links in existing programs. Thus, the objective of providing volume-weighted progression while reducing delay (or travel time) and stops can be achieved.

The MULTIBAND model is derived from the MAXBAND model and also uses mixed-integer linear programming optimization. In MULTIBAND, we define a different bandwidth for each directional road section of the arterial:

$b_i(\bar{b}_i)$ = outbound (inbound) bandwidth between signals S_i and S_{i+1}; there is now a specific band for each directional road section or link.

This band can be individually weighted with respect to its contribution to the overall objective function. (The band is continuous; only the width can vary). The method is sensitive to varying traffic conditions, and the progression scheme can be tailored to different traffic flow patterns. Users can still choose uniform bandwidth progressions if they so desire, but this is now only one of many user options.

The most important change with respect to MAXBAND or PASSER II occurs in the objective function. The bands are link-specific, so they can be weighted disaggregately to achieve desirable traffic objectives for each link. The MULTIBAND objective function has the following form:

$$\text{MAX } B = \left[1/\left(n-1\right)\sum_{i=1}^{n}\left(a_i b_i + \bar{a}_i \bar{b}_i\right)\right]$$

where $a_i(\bar{a}_i)$ are the link-specific weights in the two directions. A multitude of options are available for determining the weighting coefficients. To illustrate the comparative evaluation methodology in this paper, we investigate the following options:

$$a_i = \left(\frac{V_i}{S_i}\right)^p$$

$$\bar{a}_i = \left(\frac{\bar{V}_i}{\bar{S}_i}\right)^p$$

where

$V_i(\bar{V}_i)$ = directional volume on section i, outbound (inbound); either total volume or through (platoon) volume can be used;

$S_i(\bar{S}_i)$ = saturation flow on section i, outbound (inbound); this is the capacity volume (vphg);

p = exponential power; the values used were $p = 0$ (unit coefficients), $p = 1$ (volume/capacity ratio), $p = 2$ [(vol/cap)^2], $p = 4$ [(vol/cap)^4].

The MULTIBAND program is summarized in Figure 7. Since we have four options for p and two options for V, we have $2(4) - 1 = 7$ different options ($p = 0$ produces the same coefficients for the two volume options). Furthermore, we can choose centered or noncentered bands for a total of 14 MULTIBAND alternatives.

MULTIBAND	Given:	splits, queue clearances, target ratios of bandwidths, and for each section limits on: cycle time, link speeds, changes in speeds, and allowed left-turn phase patterns.
	Find:	cycle time, offsets, interferences; link-specific bandwidths b_i, $\bar{b}_i$; link progression speeds; and left-turn phase patterns
	To:	$\text{MAX } B = \frac{1}{(n-1)}\sum_{i=1}^{n-1}(a_i b_i + \bar{a}_i \bar{b}_i)$
	Subject to:	cycle time constraint bandwidth ratio constraints interference constraints loop integer constraints speed and speed-change constraints

FIGURE 7 MULTIBAND optimization program (5).

NUMERICAL RESULTS

We compare the performance of the two methods with respect to Canal Street, an arterial street in New Orleans, Louisiana. The NETSIM input for this street is illustrated in Figure 8. The default distance from starting node to intersection node is 394 ft. Distances between intersections are shown to the right of the right starting nodes; volume ratios are to the left of the left starting nodes. The inbound direction is that from Intersection 1 to Intersection 9, and outbound is the reverse. The parameter k_i represents the inbound/outbound ratio of

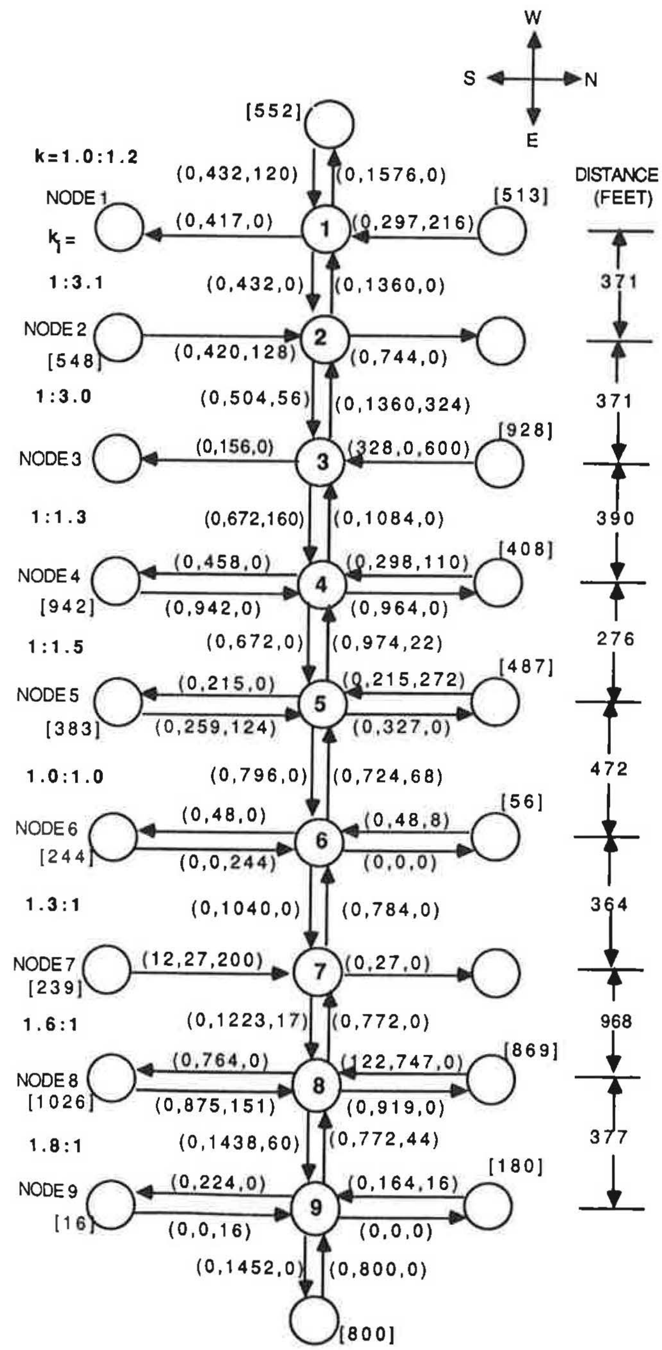

FIGURE 8 Geometry of NETSIM input, Canal Street.

volumes on opposing links between adjacent signals. There is a wide variation in the volume ratios along the arterial. However, the aggregate volume ratio for the entire arterial is 1:1.2. Clearly, conventional progression methods such as PASSER-II or MAXBAND cannot capture the full extent of information offered by this data set. MULTIBAND, on the other hand, is specifically designed to handle such a case. Sample designs produced by the MAXBAND and MULTIBAND programs are shown in Figures 9 and 10. Whereas MAXBAND generates uniform bandwidth progressions, MULTIBAND produces variable-bandwidth designs that are adapted to the variable volume ratios.

Our study compared the two possible options of MAXBAND with seven design options of MULTIBAND (only the centered options are shown). If we were to take 30 replications for each case, we would need to perform 270 different NETSIM runs of 30 min each (540 runs if noncentered options were also included). This is a formidable undertaking that may not be necessary in order to draw appropriate conclusions. Gafarian and Halati have shown that in the worst nonnormal case, five replications still provide more than 80 percent coverage probability by 95 percent confidence intervals. This coverage probability was deemed satisfactory for our purposes, therefore we limited our comparisons to five replications each. A more detailed discussion of the sample size

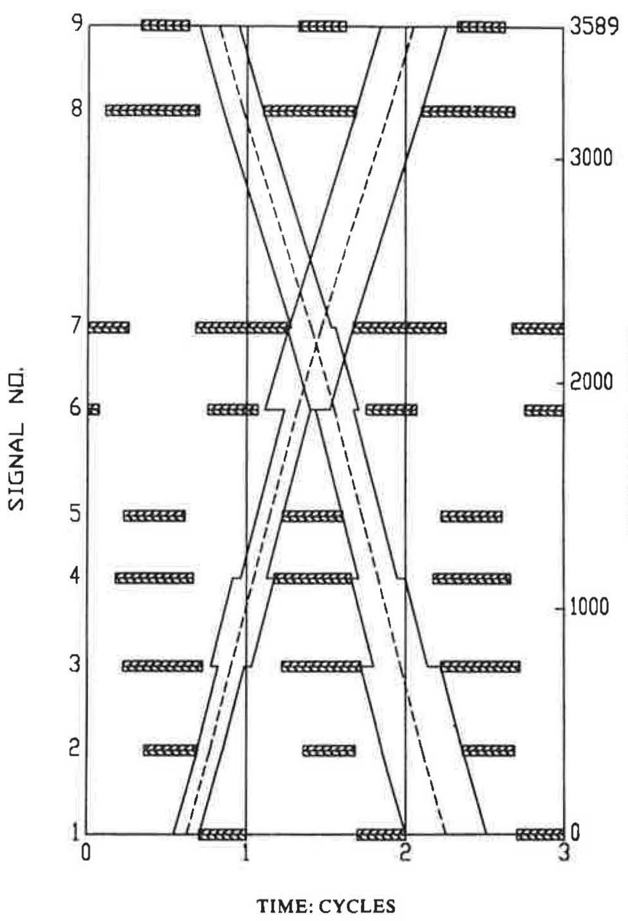

FIGURE 9 MULTIBAND time-space diagram for Canal Street: band weights = (total volume/ratio) ** 4; cycle time = 68 sec.

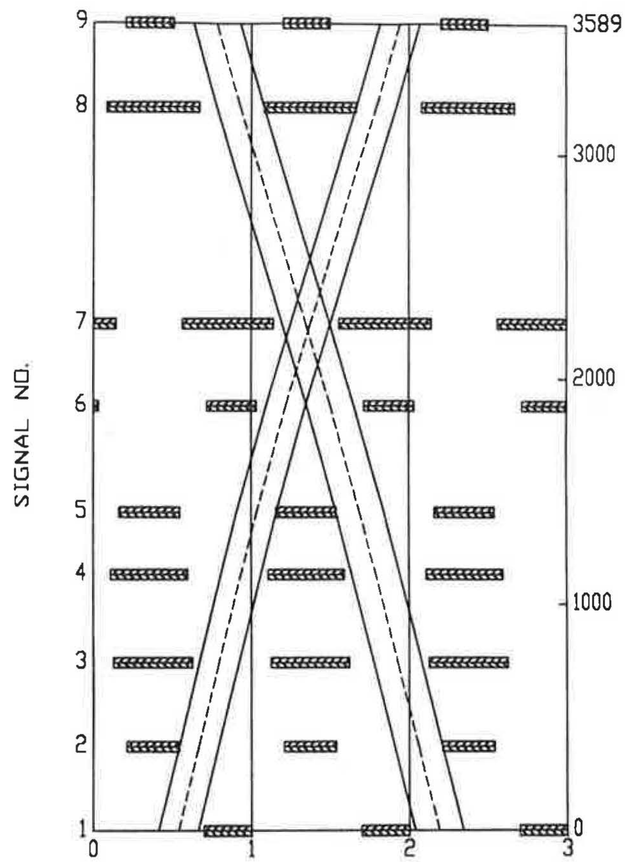

FIGURE 10 MAXBAND time-space diagram for Canal Street: band weights = total volume ratio (k = 1:1.2); cycle time = 60 sec.

requirements is given in the next section. The results of the comparisons are given in Figure 11 and illustrated in Figures 12 (average speed) and 13 (average delay). It can be easily seen that had we not used this method of analysis, there was a definite chance that we would have drawn incorrect conclusions—that is, that the advantages of the MULTIBAND options with respect to MAXBAND would not be identified. Only by identifying the appropriate confidence intervals of the different options can we make valid statistical statements on the alternatives that we compare.

Additional statistical analysis techniques can be applied for the comparison of ratio values using hypothesis testing (*2*). However, from a traffic engineering standpoint the results obtained so far indicate a clear advantage of all the MULTIBAND options relative to the MAXBAND strategies. This is particularly evident when comparing average delays, because the signal controls directly affect the delay time on the arterial. On the other hand, the overall travel time (or the speed) also includes that portion of travel time on the link that is unaffected by the signal settings.

SAMPLE SIZE REQUIREMENTS

According to the Gafarian-Halati study, when normality is reached and the number of replications exceeds 40, then the

Simulation Result from NETSIM with 95% confidence interval

CANAL STREET

METHOD	WEIGHT	TOTAL MILES [MILES]	TOTAL TIME [MINUTES]	TOTAL DELAY [MINUTES]	TOTAL VEHICLES	AVERAGE SPEED M.P.H.	AVERAGE DELAY
MAXBAND	1.0	651.41	2441.40	876.70	1762.00	16.01	29.85
		651.71	2454.90	880.60	1762.00	15.93	29.99
		651.75	2448.90	872.20	1767.00	15.97	29.62
		651.06	2429.00	844.60	1765.00	16.08	28.71
		651.06	2453.20	889.70	1762.00	15.92	30.30
AVERAGE		651.40	2445.48	872.76	1763.60	15.98	29.69
STD		0.34	0.18	1020.55	2.30		
95% CONFIDENCE INTERVAL							
UPPER BOUND						16.18	30.59
LOWER BOUND						15.78	28.79
MAXBAND	TOTAL	650.71	2431.80	859.30	1765.00	16.06	29.21
	VOLUME	650.47	2424.80	850.20	1767.00	16.10	28.87
	RATIO	650.35	2401.40	825.70	1766.00	16.25	28.05
		651.12	2420.20	853.70	1766.00	16.14	29.00
		648.48	2432.50	855.20	1765.00	16.00	29.07
AVERAGE		650.23	2422.14	848.82	1765.80	16.11	28.84
STD		1.02	0.21	799.82	0.84		
95% CONFIDENCE INTERVAL							
UPPER BOUND						16.75	29.55
LOWER BOUND						15.47	28.14
MULTIBAND	1.0	650.69	2311.40	735.90	1745.00	16.89	25.30
		652.29	2305.20	732.80	1753.00	16.98	25.08
		652.92	2325.60	744.50	1748.00	16.85	25.55
		651.18	2345.20	764.30	1745.00	16.66	26.28
		653.33	2340.50	753.10	1747.00	16.75	25.86
AVERAGE		652.08	2325.58	746.12	1747.60	16.82	25.62
STD		1.12	0.29	772.89	3.29		
95% CONFIDENCE INTERVAL							
UPPER BOUND						17.58	26.31
LOWER BOUND						16.06	24.93

Simulation Result from NETSIM with 95% confidence interval

CANAL STREET

METHOD	WEIGHT	TOTAL MILES [MILES]	TOTAL TIME [MINUTES]	TOTAL DELAY [MINUTES]	TOTAL VEHICLES	AVERAGE SPEED M.P.H.	AVERAGE DELAY
MULTIBAND	TOTAL	653.19	2309.30	733.00	1741.00	16.97	25.26
	VOLUME/	653.60	2292.00	716.00	1742.00	17.11	24.66
	CAPACITY	652.92	2299.90	725.80	1742.00	17.03	25.00
		651.18	2331.20	747.80	1745.00	16.76	25.71
		651.90	2311.60	736.20	1741.00	16.92	25.37
AVERAGE		652.56	2308.80	731.76	1742.20	16.96	25.20
STD		0.99	0.25	711.47	1.64		
95% CONFIDENCE INTERVAL							
UPPER BOUND						17.65	25.84
LOWER BOUND						16.26	24.57

FIGURE 11 MULTIBAND versus MAXBAND comparison (reduced sample set); T.V. CAP. = (total volume)/(total capacity) and P.V./CAP. = (platoon volume)/(platoon capacity). ***(continued in next column)***

MULTIBAND (T.V./		649.28	2304.20	728.60	1768.00	16.91	24.73
	CAP.)[a]	649.62	2322.40	739.00	1769.00	16.78	25.07
		651.48	2325.30	765.90	1775.00	16.81	25.89
		649.66	2343.30	759.70	1769.00	16.63	25.77
		649.19	2307.50	746.90	1771.00	16.88	25.30
AVERAGE		649.85	2320.54	748.02	1770.40	16.80	25.35
STD		0.94	0.26	907.90	2.79		
95% CONFIDENCE INTERVAL							
UPPER BOUND						17.43	26.15
LOWER BOUND						16.17	24.55
MULTIBAND (T.V./		652.06	2294.70	717.40	1777.00	17.05	24.22
	CAP.)[4]	652.92	2283.20	711.10	1780.00	17.16	23.97
		652.92	2311.40	722.90	1778.00	16.95	24.39
		650.36	2284.70	711.50	1775.00	17.08	24.05
		651.67	2280.40	707.70	1775.00	17.15	23.92
AVERAGE		651.99	2290.88	714.12	1777.00	17.08	24.11
STD		1.06	0.21	361.31	2.12		
95% CONFIDENCE INTERVAL							
UPPER BOUND						17.81	24.43
LOWER BOUND						16.34	23.80

Simulation Result from NETSIM with 95% confidence interval

CANAL STREET

METHOD	WEIGHT	TOTAL MILES [MILES]	TOTAL TIME [MINUTES]	TOTAL DELAY [MINUTES]	TOTAL VEHICLES	AVERAGE SPEED M.P.H.	AVERAGE DELAY
MULTIBAND	PLATOON	653.28	2308.10	730.50	1746.00	16.98	25.10
	VOLUME/	653.21	2318.20	734.60	1742.00	16.91	25.30
	CAPACITY	652.69	2318.90	734.70	1745.00	16.89	25.26
		653.53	2294.60	724.40	1747.00	17.09	24.88
		651.43	2287.70	724.40	1748.00	17.09	24.86
AVERAGE		652.83	2305.50	729.72	1745.60	16.99	25.08
STD		0.84	0.23	308.62	2.30		
95% CONFIDENCE INTERVAL							
UPPER BOUND						17.57	25.36
LOWER BOUND						16.41	24.81
MULTIBAND (P.V./		652.04	2306.60	728.40	1744.00	16.96	25.06
	CAP.)[4]	653.67	2301.30	730.70	1749.00	17.04	25.07
		652.69	2296.30	717.60	1739.00	17.05	24.76
		653.78	2347.60	762.00	1745.00	16.71	26.20
		650.69	2298.80	729.90	1741.00	16.98	25.15
AVERAGE		652.57	2310.12	733.72	1743.60	16.95	25.25
STD		1.28	0.35	1000.23	3.85		
95% CONFIDENCE INTERVAL							
UPPER BOUND						17.82	26.14
LOWER BOUND						16.08	24.36

FIGURE 11 ***(continued)***

coverage probability for the 95 percent confidence intervals of the estimator will approach .95. As indicated above, MULTIBAND has 14 options and MAXBAND has 4 options. To evaluate all these options with 30 replications would take 540 NETSIM runs. Practically, it may not be necessary to pursue this degree of precision; a smaller sample size may be a good alternative. Statistically, a small sample size may lead to a biased estimation, though it may still be acceptable. In the worst case, five replications of NETSIM still provide 80 percent coverage probability of 95 percent confidence intervals. For our study, this coverage probability is already good enough to determine the best design option for each method. After this decision has been made, additional replications can be performed for a more accurate comparison of the two best options.

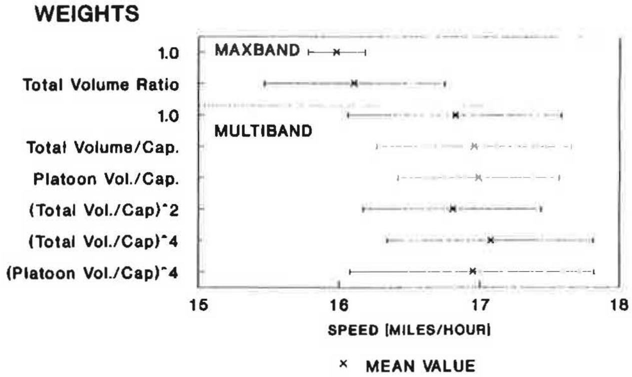

FIGURE 12 Average speed comparison, Canal Street, MULTIBAND versus MAXBAND.

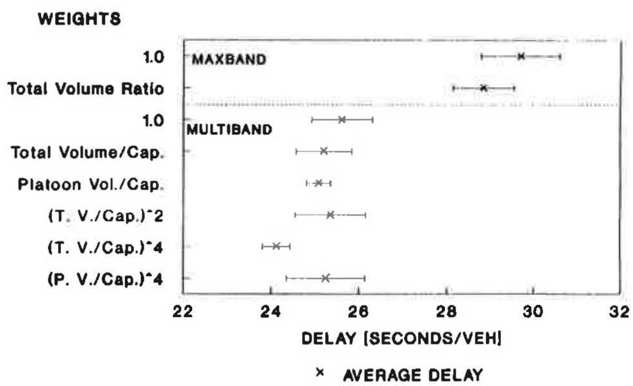

FIGURE 13 Average delay comparison, Canal Street, MULTIBAND versus MAXBAND.

In our comparison study, we took five replications for each option of both methods (shown in Figures 11–13). It shows clearly that the best option for MULTIBAND in terms of average delay is when the weights equal to the total volume/capacity ratio with the fourth power; for MAXBAND, it is when the weight equals to the total volume ratio. In both cases, the band's location option is centered.

Comparing Figures 3 and 4 with Figures 12 and 13, it is clear that the 95 percent confidence intervals shown in Figures 12 and 13 are larger. The reason is that a larger sample size reduces the variation caused by the different random seeds or, in case of a field test, the unknown factor of variation. It is, therefore, very important to decide what is the desired sample size, because a proper sample size not only saves time and money but also provides a good conclusion for the comparison study. A logical way to decide on the sample size is to use as input the desired variation and the results of a small pilot study.

Sometimes, according to varying requirements, the engineer may specify the desired variation and then choose the number of replications on the basis of the specified variance. For example, if the desired result is for the variance of the number of vehicle-trips to be smaller than desired variance (DVAR), the required number of replications according to a sample size of n_1 in a pilot study is $n = s_1^2\,(1 + 2/n_1)/\mathrm{DVAR}$ where s_1^2 is the variance of the pilot sample.

Assume a pilot study is conducted with a sample variance of total vehicle trips of $s_1^2 = 64$ and $n_1 = 5$. The sample variance of the final study could be reduced to 4, yielding a total sample size for the final study of $n = 64(1 + 2/5)/4 = 22.4 \approx 23$, that is, an additional 18 replications can reduce the variance of the vehicle trips to 4. Thus, a small pilot study can be used as a guide in obtaining the final desired accuracy.

CONCLUSIONS

Many important traffic MOEs are calculated as the ratio of two random variables. This is also so when simulation models are used for traffic system evaluation. To obtain a valid statistical analysis, ratio estimation techniques must be used. In this paper we have shown how such techniques can be applied to compare the effectiveness of alternative traffic control strategies. Guidelines are provided for determining the number of simulation samples that should be produced to achieve a desirable accuracy.

The methodology described in this paper can be used to evaluate any control or management strategy using simulation data. The analysis techniques have wider applications: they may also be applied to the estimation of parameters, the assessment of the accuracy of estimates from field data, and the comparison of field data with simulated data for validation studies.

REFERENCES

1. A. V. Gafarian and A. Halati. *Statistical Analysis of Output Ratios in Traffic Simulation.* In *Transportation Research Record 1091*, TRB, National Research Council, Washington, D.C., 1986, pp. 29–36.
2. W. G. Cochran. *Sampling Techniques*, 3rd ed. John Wiley & Sons, New York, N.Y., 1977.
3. J. D. C. Little, M. D. Kelson, and N. H. Gartner. *MAXBAND: A program for Setting Signals on Arteries and Triangular Networks.* In *Transportation Research Record 795*, TRB, National Research Council, Washington, D.C., 1981, pp. 40–46.
4. E. C. P. Chang, C. J. Messer, and R. U. Garza. Arterial Signal Timing Optimization Using PASSERII-87. *ITE Journal*, Nov. 1988, pp. 27–31.
5. N. H. Gartner, S. F. Assmann, F. Lasaga, and D. Hou. MULTIBAND: A Variable Bandwidth Arterial Progression Scheme. In *Transportation Research Records 1287*, TRB, National Research Council, Washington, D.C., 1990, pp. 212–222.

Publication of this paper sponsored by Committee on Traffic Signal Systems.

Progression-Based Optimization Model in TRANSYT-7F

MOHAMMED A. HADI AND CHARLES E. WALLACE

The forward progression opportunities (PROS) concept has been developed as an alternative for design and evaluation of arterial signal timing. The concept expands upon the maximal bandwidth approach by considering time-space progression opportunities that do not necessarily extend throughout the length of the route. Further, PROS can be available for traffic outside the through bands of the traditional time-space diagram. PROS can be used alone or in conjunction with system disutility measures (such as stops, delay, and fuel consumption) for a more complete progression design. The implementation of the PROS concept for application on multiple arteries within networks, using TRANSYT-7F, is presented. Data are presented to demonstrate the performance of the model relative to other optimization strategies. Experiments with the model indicate that it can produce significant improvements over both the traditional bandwidth and disutility methods for designing arterial and network signal timing plans.

Optimizing traffic signal timing is one of the most cost-effective ways of improving the quality of traffic flow and reducing fuel consumption (*1*). There are two basic approaches for off-line optimization of progressive signal system timing: (a) maximizing bandwidth efficiency, and (b) minimizing a disutility index that has generally been a function of a combination of delay, stops, fuel consumption, and sometimes queue spillover.

The first approach has traditionally been appropriate only for arterial streets and includes models such as PASSER II (*2*) and MAXBAND (*3*). MAXBAND has been extended for application to multiarterial networks, but this version is not widely used, primarily because of excessive computer run time and its limitation to mainframe computers (*4*).

The second approach can be used for arterial streets as well as two-dimensional networks; it includes models such as TRANSYT-7F (*5*). TRANSYT-7F is one of the most powerful computer programs for traffic signal timing and traffic flow analysis; however, many traffic engineers prefer bandwidth-based solutions for designing arterial timing plans to ensure perceived progression. Designs based on minimizing disutility may not produce the wide through progression bands. On the other hand, maximal bandwidth solutions do not necessarily result in minimum delay and stops or, more significantly, fuel consumption. This is because these solutions do not explicitly recognize traffic demand as a function of time on individual links. That relationship is only implicit.

Several studies have investigated the benefits of combining the disutility optimization approach and the bandwidth optimization approach. One approach was to use maximal bandwidth optimization models such as PASSER II and MAXBAND to develop initial timing plans for TRANSYT-7F (*6*–*8*). Chang et al. used an estimate of link delays to determine the directional bandwidth ratio (*9*). Cohen and Liu developed an approach that constrains the TRANSYT-7F optimization to preserve the band computed by a maximal bandwidth program (*10*). Gartner et al. developed a method that generates a variable bandwidth progression in which each through link can obtain an individually weighted bandwidth (*11*). Both volume and saturation flow rates on a through link were used to obtain the link weight.

It was also attempted to give more priority to the arterial links in a TRANSYT-7F optimization. TRANSYT documentation has suggested the use of link-to-link flow weighting, stop weighting factors, and delay weighting factors to ensure time-space progression (*5*,*12*). Moskaluk and Parsonson suggested absolute prioritization of arterial through links while controlling minor movement performance degradation by specifying a maximum degree of saturation for these movements (*13*).

All of these approaches have been simple manipulations of the model's inherent capabilities, which can be replicated through data coding. None has resolved the fundamental need to combine the *modeling* of progression explicitly in TRANSYT or to combine a progression-based objective with disutility.

Wallace (*14*) and Courage (*15*) developed the forward progression opportunities (PROS) model as an alternative design and evaluation approach for traffic progression. This model considers not only the through bands, but also short-term progression opportunities within the system. The concept of forward progression opportunities simply recognizes the ability to travel through two or more adjacent intersections at the desired progression speed without stopping. When such opportunities are obtained, both in time and space, a disaggregate measure of progression is available that is much more flexible than traditional through bands. Traditional through bands are severely bounded by absolute physical rules. PROS are less constrained.

When PROS are optimized solely, the problem experienced by Moskaluk and Parsonson exists—minor (generally cross-street) movements may be driven to their minima in an optimization. To overcome this, an extension of the PROS concept was developed to maximize a combinational objective function of PROS and the TRANSYT-7F disutility index (DI). In the past the TRANSYT objective function has been a performance index (PI), which was a function of stops and delay and was always minimized. This is commonly referred to as performance optimization. The current model uses new definitions of objective functions (*16*). The older stops and delay function is now referred to consistently as the DI. The com-

Transportation Research Center, University of Florida, 512 Weil Hall, Gainesville, Fla. 32611.

binational objective function, defined as PROS/DI, combines the benefits of the two design approaches. An early experimental version of the TRANSYT model was modified to implement the PROS model. Experience with this model (*14*) suggested that the concept had merit.

Recently, the PROS model was incorporated into TRANSYT-7F, Release 7. The basic principle of using PROS and, more significantly, PROS/DI as an objective function in TRANSYT-7F remains the same. However, other improvements and new features were added. These improvements include extending the PROS concept for application to multiarterial networks, allowing a cycle range evaluation, adjusting timing to reduce the DI in a simple PROS optimization, using PROS-versus-DI weighting, and employing PROS directional weightings for individual arteries.

Other general improvements to TRANSYT-7F that support this are important, but they are not reported explicitly herein. In short, they include

- Improved modeling of stops at degrees of saturation near or over 100 percent,
- Better split optimization based on degrees of saturation, and
- Explicit handling of overlap phases.

This paper describes the incorporation of the PROS concept into TRANSYT-7F and investigates the results obtained using the model.

TRANSYT-7F MODEL

The Traffic Network Study Tool (TRANSYT) model was originally developed by Dennis I. Robertson in England (*17*). Since the development of the original model, many improvements have been made in Great Britain, the United States, and elsewhere. TRANSYT has been extensively tested and used throughout the world for design and evaluation of traffic signal timing. The model most widely used in the United States is TRANSYT-7F; its latest version is Release 7, into which the PROS model discussed in this paper was incorporated.

TRANSYT consists of two main parts:

1. A traffic flow model that is a deterministic macroscopic time-scan simulation. It simulates traffic flow in a street network of signalized intersections to compute a disutility index for a given signal timing plan.
2. An optimization procedure based on an iterative gradient search technique, known as hill-climbing, makes changes to the signal timing (offsets and splits) to determine whether or not the PI is improved. By adopting only those changes that improve the PI, the optimizer tries to find a set of timing that optimizes the PI, subject to any limits placed on the process.

These submodels are inexorably intertwined in TRANSYT-7F—particularly in the new version, which has considerably more complex traffic simulation and optimization models.

PROGRESSION OPPORTUNITIES CONCEPT

A forward progression opportunity is defined as the ability presented at a given point in time to enter one intersection on green (including the change period) and expect to travel through the next downstream intersection without stopping, independent of other traffic.

Each such opportunity available during a given period of time (for example, a "step" in TRANSYT-7F) is tabulated as one progression opportunity. Multiple opportunities, both in time and space, are accumulated as PROS.

The number of PROS in a given direction for a given time period (or step) is the number of successive green signals that will be encountered at the design speed without stopping. The aggregate PROS is found by summing the PROS over all time periods in both directions, or

$$\text{PROS} = \sum_{k=1}^{2} \sum_{j=1}^{N} \sum_{t=1}^{C} \text{PROS}_{kjt} \tag{1}$$

where

k = direction of travel;
j = intersection number, of which there are N; and
t = time in units common to the model, up to the cycle length, C.

A subset of the above is to consider only the PROS_{kjt} where t ranges from T_{kl} to T_{kt} and these limits respectively represent the leading and trailing edges of the arterial through green bands in each direction—for example, from PASSER II. Optimizing this subset would have the effect of explicitly optimizing maximal bandwidth.

Another measure related to the PROS concept is also defined as cycle progression opportunities (CPROS). In the absence of any signal, the full cycle would be available as progression opportunities. This hypothetical measure is obtained for both directions as follows:

$$\text{CPROS} = CN(N - 1) \tag{2}$$

where the variables are as defined before. The ratio PROS/CPROS, called effective PROS (PROS_e), is analogous to the bandwidth efficiency of PASSER II and MAXBAND. By the nature of this definition, PROS_e will always be less than unity.

MODEL IMPLEMENTATION

The procedure used to optimize signal timing on the basis of PROS or PROS/DI is functionally the same as that used by the TRANSYT-7F model to minimize the normal DI. To implement this model, TRANSYT-7F was modified to calculate, and to optimize optionally, PROS or PROS/DI instead of DI.

Simple PROS Optimization

In a simple PROS optimization, the main objective is to maximize PROS of the arterial through movements. However, to deal with other movements, the model also tries to minimize the DIs elsewhere in the system while not reducing the arterial PROS.

If this option (referred to as "PROS" and "DI") is selected, TRANSYT-7F performs two steps of the traditional hill-climbing procedure as follows:

1. The PI is initially based purely on the $PROS_e$ on the arteries. A change in timing is retained if it increases the $PROS_e$. When computing $PROS_e$ for a multiarterial system, the PROS and CPROS are computed first for each artery using Equations 1 and 2. Then the aggregate PROS and CPROS for the whole system are determined by summing their values over all designated arteries. The $PROS_e$ for the system is determined on the basis of these two values. (The ratio $PROS_e$, rather than the raw value of PROS, is used to allow for a cycle search, because the raw value always increases with cycle length.)
2. In Step 2, the hill-climbing procedure is used to minimize the DI without reducing the $PROS_e$ value achieved in Step 1. This allows for some adjustment of the offsets and splits on the arteries. In addition, the timing for intersections (nodes) not considered in Step 1 (because they are not on any designated artery) are optimized entirely in Step 2.

The model also allows for a cycle search to select the best cycle length within a user-specified range. In this process, only the first step of the optimization can be performed, because there can be only one objective function when different cycles are compared. The comparison is based exclusively on the $PROS_e$ values.

When PROS are optimized in a grid network, the default is to give the same weight to all nodes on all arteries; however, directional weightings can be used to give priority to specific arteries or directions.

As noted before, PROS-only optimization suffers from the same disadvantages as the maximal bandwidth approach in that the actual traffic demand is not explicitly considered. For example, split optimization generally forces the green times of minor movements at nodes where split optimization is permitted to their minima. This is because PROS-only optimization does not provide criteria for setting green times for these movements. One of the following procedures can be used to avoid oversaturating minor movements and ensure an equitable distribution of green times when optimizing the PROS.

1. For pretimed controllers, initial splits that equalize the degrees of saturation can be requested. For actuated controllers, initial splits are automatically calculated to achieve a desired degree of saturation. All splits are then held constant in the first optimization step.
2. The splits can be coded by the user and fixed during the first optimization step. These splits should be calculated externally.

Another problem with simple PROS optimization is that it may cause the PROS to increase in one direction (not necessarily the critical direction) at the expense of the PROS in the other direction. This problem can generally be avoided by using proper directional weightings or by coding initial timings from a maximal bandwidth optimization program.

PROS/DI Optimization

An extension of the PROS concept was developed to redefine the objective function as the PROS/DI ratio (*14,15*). The purpose of this formulation was to combine the advantages of maximizing the PROS with those of minimizing the DI.

In Release 7 of TRANSYT-7F, an option allows the user to select the PROS/DI ratio as the PI. The model calculates the effective PROS on the arteries and the DI for the entire system after each timing shift. The shift is retained if it increases the PROS/DI ratio.

Unlike simple PROS optimization, this policy considers the PROS and DI at the same time; thus, it eliminates the need for the second optimization step discussed before.

Because the minor movements are accounted for in calculating the DI, splits can be optimized in addition to offsets. This policy tries to maximize progression, subject to maintaining sufficient green times for the minor movements. In addition, the policy attempts to find the set of offsets that clears the existing queue before the platoon's arrival. Nodes not on the designated arteries are explicitly considered in the traditional DI calculation; thus, their offsets and splits are optimized in concert with the arterial progression.

The relative weighting of the $PROS_e$ in the objective function can be varied by

$$PI = (100 \times PROS_e)^{WP}/DI \qquad (3)$$

where WP is the relative weight of PROS to the DI.

This allows for fine-tuning the relative importance of the PROS on the arteries versus the DI for all links. Experience with the early version of the model indicated that, in some cases, optimizing splits based on the PROS/DI strategy without weighting tended to discriminate against minor turning movements. Cohen and Liu likewise noted that the earlier optimization strategy resulted in shorter side-street greens (*10*).

To correct this anomaly, it was determined that a WP value of 0.5 would generally reduce the weight of PROS relative to the DI. This increases the weight of minor movements in the optimization and generally results in a better optimization of splits while it maintains good progression on the arteries. It should be noted, however, that this value of WP (0.5) is not likely to be the ideal value for all networks.

Selecting cycle length on the basis of the PROS/DI ratio is also allowed. In this process, the model employs the same procedure as that used by TRANSYT-7F to select the cycle length that produces the best PI.

Directional Weighting

From the traffic engineering point of view, it may be desirable to favor one direction of travel on an artery over another, such as during peak periods. In addition, it may be preferable to give different weights to different arteries. Thus, a weighting strategy is employed by modifying the formulation of Equations 1 and 2 for a specific artery. First, define

$$PROSR_i = \frac{PROS_{i1}}{PROS_{i1} + PROS_{i2}} \qquad (4)$$

and

$$WDR_i = \frac{WD_{i1}}{WD_{i1} + WD_{i2}} \tag{5}$$

where

$PROSR_i$ = relative PROS in the "forward" direction (k = 1) on artery i,
$PROS_{ik}$ = actual PROS in direction k on artery i,
WDR_i = relative weighting for the forward direction (k = 1) on artery i, and
WD_{ik} = weighting factors for direction k on artery i.

Now define a desired directional factor for artery i (DDF_i):

$$DDF_i = \frac{\min(PROSR_i, WDR_i)}{\max(PROSR_i, WDR_i)} \tag{6}$$

and the resulting definition of the effective PROS ($PROS_e$) is

$$PROS_e = \frac{\sum_{i=1}^{A} DDF_i \sum_{k=1}^{2} WD_{ik} \sum_{j=1}^{N_i} \sum_{t=1}^{C} PROS_{ikjt}}{\sum_{i=1}^{A} \sum_{k=1}^{2} WD_{ik} \sum_{j=1}^{N_i} \sum_{t=1}^{C} CPROS_{ikjt}} \tag{7}$$

where A equals the number of arteries.

By weighting the arteries relative to each other, the two directions of travel along each artery, or both, the engineer may be able to influence the resultant design to achieve a desired policy.

MODEL APPLICATIONS

The PROS optimization strategy implemented in TRANSYT-7F was evaluated using two real-world traffic systems: an artery and a two-dimensional network.

Timing plans were designed for the two systems, using the PROS, PROS/DI, and standard DI objective functions in TRANSYT-7F. Only the offsets were optimized initially. Then both offsets and splits were optimized. The DI was consistently defined as excess fuel consumption due to stops and delay.

The resultant designs were compared on the basis of perceived progression and fuel consumption as reflected by the $PROS_e$ and DI values, respectively.

Although macroscopic measures of effectiveness are necessarily reported, the individual runs were examined to ensure that no links were seriously oversaturated such that the results are unfairly biased. In several PASSER II and TRANSYT-7F final solutions, there was some minor oversaturation, but it was more to the disbenefit of the TRANSYT-7F results.

Arterial Application

Cape Coral Parkway, in the city of Cape Coral, Florida, was used as the study artery. This is an east-west artery with seven intersections. The configuration of the artery is presented in Figure 1. The numbers between nodes are the intersection spacings.

The PROS and PROS/DI optimization strategies were compared with the performance optimization strategy as described. The study was performed for the existing phase sequences and the optimal phase sequences selected by PASSER II-90.

For the existing phase sequences, the comparison was repeated using two different initial timing plans. These plans were calculated using the internal initial timing routines of PASSER II and TRANSYT-7F, respectively, to illustrate the point that TRANSYT-7F's optimal solution generally results in a superior plan if its initial timing plan is good. TRANSYT computes the initial splits for pretimed controllers based on equalizing the degrees of saturation on the critical conflicting links. The routine sets all initial offsets to zero. For the optimal phase sequences, the initial timing was that optimized by PASSER II. The TRANSYT-7F internally generated plan was not examined, because it is known that it would result in an inferior design.

The results of the comparative study are summarized in Table 1. As mentioned earlier, the default value used for the relative weight of PROS in the PROS/DI optimization (*WP* in Equation 3) was 0.5. The results used to represent the PROS/DI strategy in the comparative analysis are based on this value unless otherwise stated.

The results of Table 1 indicate that the PROS and PROS/DI solutions were clearly superior to the minimal DI solutions in terms of perceived progression. For the existing phase sequence and the two initial timing plans used (TRANSYT-7F and PASSER II), the PROS/DI optimization strategy increased the $PROS_e$ by 54 and 31 percent, respectively (that is, 39.5 versus 25.6 percent and 39.8 versus 30.4 percent) compared with the DI policy. (All percentages used for comparisons in the test are based on results obtained from optimizing both offsets and splits unless otherwise specified.)

The improvement was less significant when using the PASSER II optimal sequence solution as the initial timing (41.6 versus 40.3 percent, which is only 3 percent). Similar trends were observed when comparing the solutions from the PROS and DI policy with those from the DI policy. For the optimal phase sequence, the hill-climbing process was able to reach a good local optimum, in terms of progression, in the performance optimization. Thus, PROS-based solutions could not improve much over the DI solutions in this case.

In terms of excess fuel consumption, optimization based on PROS or PROS/DI did not result in a serious increase in the DI compared with the performance optimization strategy. In

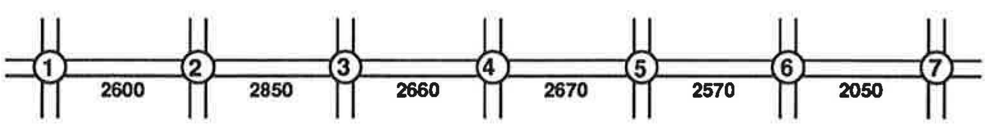

FIGURE 1 Configuration of Cape Coral Parkway.

TABLE 1 Comparison of DI, PROS/DI, and PROS Optimizations for Cape Coral Parkway

Initial Timing/ Sequence	Optimization	Objective Function	$PROS_e$ (%) Right[a]	Left[a]	Total	Bandwidth (sec) Right[a]	Left[a]	Artery DI	Total DI
TRANSYT/ Existing	Offsets only	DI	23.4	24.4	23.9	16	0	94.5	248.6
		PROS	21.4	57.4	39.4	3	73	104.6	252.1
		PROS/DI	31.0	43.2	37.1	12	43	95.0	243.9
	Offsets and Splits	DI	27.0	24.3	25.6	23	0	90.8	246.9
		PROS	24.6	54.3	39.5	5	62	95.1	245.0
		PROS/DI	35.3	43.7	39.5	18	46	92.6	242.7
PASSER II/ Existing	PASSER II	---	23.8	42.2	33.0	24	44	110.4	252.0
	Offsets only	DI	21.2	38.3	29.7	0	22	103.5	243.9
		PROS	28.6	40.3	34.4	23	41	106.8	249.3
		PROS/DI	27.4	40.4	33.9	13	39	104.2	245.5
	Offsets and Splits	DI	20.7	40.2	30.4	6	28	94.9	234.0
		PROS	25.5	45.1	35.3	0	44	97.1	233.1
		PROS/DI	35.5	44.2	39.8	29	47	95.2	240.8
PASSER II/ Optimal	PASSER II	---	31.2	50.3	40.7	36	58	104.6	244.3
	Offsets only	DI	28.7	52.0	40.3	14	54	99.8	237.8
		PROS	30.1	54.8	42.4	27	68	102.7	241.9
		PROS/DI	28.3	54.9	41.6	18	65	100.2	238.4
	Offsets and Splits	DI	28.7	52.0	40.3	14	54	99.8	237.8
		PROS	29.5	55.5	42.5	25	69	103.4	241.9
		PROS/DI	28.3	54.9	41.6	18	65	100.2	238.4

[a]Right and Left refers to the right-bound and left-bound travel on the artery.

fact, in some cases, they produced a lower DI. In these cases, the PROS-based optimization led the hill-climbing process to a better local optimum in terms of the DI. Varying the relative weight of the PROS to DI (as will be shown later) or directional weightings of the PROS may reduce the DI further.

When the PROS or PROS/DI optimization policy was applied to the PASSER II solutions, some improvements in the PROS values were realized. For example, the PROS/DI optimization could increase the $PROS_e$ values by 21 percent (39.8 versus 33.0 percent) and 2 percent (41.6 versus 40.7 percent), respectively, for the existing phase sequence and optimal phase sequence compared with PASSER II solutions.

In terms of systemwide traffic operation as measured by the DI, solutions based on all three optimization strategies in TRANSYT-7F produced improvements over PASSER II solutions in terms of both PROS and DI. It is recognized that PASSER II, by the nature of its objective function, sometimes gives wider through bands.

Generally, when the PROS/DI strategy was used, split optimization resulted in an increase in the PROS and a decrease in the DI. The maximum increase in the $PROS_e$ was 17 percent (39.8 versus 33.9 percent), and the maximum decrease in the DI was 2 percent (240.8 versus 245.5).

When the splits in PROS and DI optimization are adjusted, the process tries to minimize the DI while not reducing the PROS. Considerable reduction in the DI could be realized during this adjustment. Although PROS were not explicitly considered in this process, they might increase due to improvements in the DI on the artery. The maximum improvement achieved in the $PROS_e$ value was 3 percent (35.3 versus 34.4 percent).

As expected, split optimization using the DI objective function generally reduced the DI. It also produced some increases in the PROS values.

For the data set investigated, optimizing the phase sequence using PASSER II improves PROS for all optimization strategies considered, however, its effect on fuel consumption was varied.

Table 1 presents numbers that demonstrate that increasing the PROS did not necessarily result in a decrease in the DI of arterial links (i.e., those most affected by coordination). It is believed that the effect of PROS on the DI of arterial links is a function of many factors, including link lengths, turning movements from cross streets, and degrees of saturation on the artery.

In all cases, the arterial link-only DIs were within a few percentage points for all three optimization strategies (i.e., DI, PROS and DI, and PROS/DI); but in both comparisons with PASSER II solutions, the DIs were lower with PROS/DI or PROS optimization.

Next, an investigation was conducted to assess the benefits of weighting the PROS relative to DI in the PROS/DI formulation. As noted, the PROS/DI policy produced very good progression solutions. Thus, it was decided to use the PROS relative weighting to try to reduce the DI. That implied putting less weight on the PROS relative to the DI.

The results, presented in Table 2, indicate that the DI could be decreased without causing serious reductions in PROS by using a proper weighting. In a few cases, putting less emphasis on PROS by reducing its weight increased the PROS. This suggests a need for more work to assess the stability of this weighting strategy.

Network Application

The network chosen for this study is a 26-intersection network in Flint, Michigan. Its configuration is shown in Figure 2. The PROS-based model implemented in TRANSYT-7F was used in an attempt to improve the perceived progression on two arteries within the system. These two arteries are Dupont and Detroit streets, both of which are north-south arteries with eight intersections.

The results obtained for this network are summarized in Table 3. These results demonstrated significant improvements in the PROS on both arteries when either of the PROS-related functions is used instead of the DI alone.

Compared with the DI minimization policy, PROS/DI maximization increased the $PROS_e$ by 117 percent (46.6 versus 21.4 percent). However, it also increased the DI by 9 percent (193.3 versus 177.6).

The PROS/DI solution was superior to the PROS and DI solution in terms of both PROS and DI. The PROS/DI optimization produced a 46 percent (46.6 versus 31.8 percent) increase in the $PROS_e$ but a minimal 0.8 percent (that is, 193.3

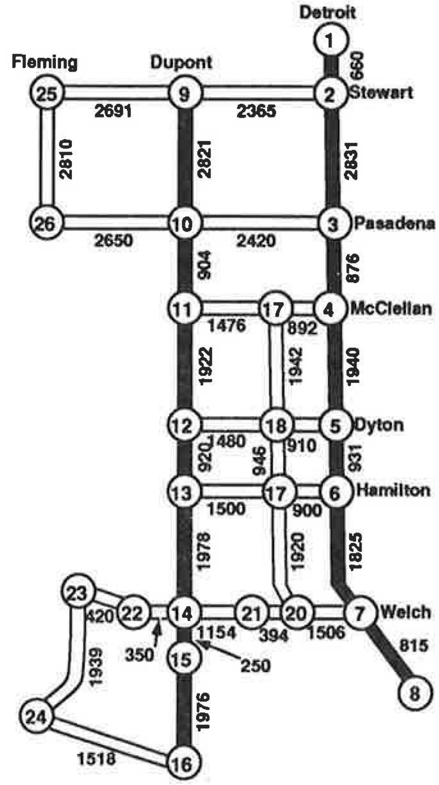

FIGURE 2 Flint network configuration (distances in feet).

TABLE 2 Effect of Changing Relative Weight of PROS to DI

Phase Sequence	Initial Timing	PROS Weighting Factor	$PROS_e$ (%) Right[a]	Left[a]	Total	Bandwidth (sec) Right[a]	Left[a]	Artery DI	Total DI
Existing	TRANSYT	1.00	36.4	46.6	41.5[b]	29	50	88.3	248.3
		0.60	30.0	51.0	40.5	20[b]	60	94.5	246.5
		0.50	35.3	43.7	39.5	18	46	92.7	242.7
		0.40	36.7	43.2	39.9	28	46	93.7	244.4
		0.30	36.1	42.4	39.2	27	38	93.5	249.4
		0.20	28.4	43.7	36.1	13	46	94.0	241.3[b]
		0.10	19.7	43.4	31.6	6	33	90.6	246.6
Existing	PASSER II	1.00	30.9	53.3	42.1[b]	19[b]	59	91.4	246.5
		0.60	36.1	43.6	39.9	31[b]	47	95.8	241.5
		0.50	35.5	44.2	39.9	29	47	95.2	240.8
		0.40	33.0	43.8	38.4	13	37	91.4	238.2
		0.30	29.6	48.2	38.9	17	49	91.4	237.6
		0.20	30.8	45.4	38.1	27	51	96.6	235.0
		0.10	31.9	42.3	37.1	28	45	95.7	234.2[b]
Optimal	PASSER II	1.00	35.6	53.3	44.5[b]	26	54	93.1	256.9
		0.60	29.6	53.7	41.6	20	62	100.8	238.4
		0.50	28.3	54.9	41.6	18[b]	65	100.2	238.4
		0.40	33.6	48.0	40.8	22	56	98.8	237.5
		0.30	32.6	47.9	40.3	18	58	97.6	236.3[b]
		0.20	29.4	53.3	41.4	17	58	99.8	237.5
		0.10	28.9	53.7	41.3	16	58	100.4	238.2

[a]Right and Left refers to the right-bound and left-bound travel along the artery.
[b]Indicates the "best" solution for each measure. In the case of bandwidth, the "best" applies to both directions.

TABLE 3 Comparison of DI, PROS/DI, and PROS Optimizations for Flint Network

Optimization	Objective Function	PROS Weight (WP)	Artery Number	$PROS_e$ (%) Right[a]	Left[a]	Total	Bandwidth (sec) Right[a]	Left[a]	Artery DI	Network DI
Offsets Only	DI	__[b]	1	20.6	19.4	21.0	0	0	36.4	
			2	14.0	30.0		0	11	35.7	200.3
	PROS	__[b]	1	47.7	14.8	31.9	27	0	40.9	
			2	26.8	38.2		12	19	34.7	207.5
	PROS/DI	0.5	1	24.0	29.4	29.3	1	9	38.6	
			2	25.1	38.6		11	20	34.4	198.9
Offsets and Splits	DI	__[b]	1	13.9	17.1	21.4	0	0	33.5	177.6
			2	21.1	33.3		0	9	27.9	
	PROS	__[b]	1	31.0	29.0	31.8	14	1	31.1	194.9
			2	29.9	37.2		11	17	26.3	
	PROS/DI	0.5	1	38.7	45.2	46.6	10	19	30.4	193.3
			2	39.8	62.7		18	36	32.0	
	PROS/DI	0.2	1	15.4	30.7	33.1	0	11	32.2	181.2
			2	30.5	55.9		7	30	32.6	

[a]Right and Left refers to the right-bound and left-bound travel on the artery
[b]"__" Means that the WP is not applicable in this case.

versus 194.9) decrease in the DI relative to the PROS and DI optimization.

By using the proper directional and arterial weightings or relative PROS–to–DI weighting, the resultant designs might be further improved. For example, using a *WP* of 0.2 in the PROS/DI optimization produced a good design from both the PROS and DI perspectives. This strategy increased the $PROS_e$ by 55 percent (33.1 versus 21.4 percent) with only a small increase in the DI (181.2 versus 177.6, which is 2 percent) compared with the performance optimization solution.

CONCLUSIONS AND RECOMMENDATIONS

From the results presented in this study, it can be concluded that the PROS concept has been successfully implemented in TRANSYT-7F and that it is a practicable design tool for networks. The PROS and PROS/DI optimization strategies significantly improve progression over the normal performance optimization in TRANSYT-7F. These improvements are realized for single arteries as well as multiarterial networks.

Designs based on either of the PROS-related objective functions for individual arteries compare favorably with the solutions obtained by a bandwidth optimization program. The new version of TRANSYT-7F can now deal with multiple arteries within a network.

The two PROS-related strategies combine the advantages of maximizing progression with that of minimizing the DI. It appears that these functions can be used instead of the traditional performance optimization strategy of TRANSYT when the objective is to optimize both progression and fuel consumption.

The use of proper relative PROS–to–DI weighting, directional weightings, and arterial weightings appears to have merit. More work is required to improve these strategies.

The macroscopic simulation model in TRANSYT-7F was used to assess the effectiveness of different optimization strategies in reducing fuel consumption. Although the TRANSYT-7F simulation model is realistic and widely accepted, there is a need to validate the PROS-based model by field testing, or at least by using a microscopic simulation model such as TRAF-NETSIM (*18*). Further research is suggested to ensure that all aspects of the PROS-based optimization are stable, particularly with respect to weighting factors.

ACKNOWLEDGMENTS

The research described in this paper was conducted as a part of the Release 7 development of TRANSYT-7F. This release was developed by the Center for Microcomputers in Transportation (McTrans) of the University of Florida Transportation Research Center. Charles Jacks of McTrans adapted the source code for the PROS model in TRANSYT-7F, Release 7. The authors thank him for his contribution to this work.

REFERENCES

1. C. E. Wallace and K. G. Courage. *Methodology for Optimizing Signal Timing—The (MOST) Reference Manual*, Vol. 1. Courage and Wallace, Gainesville; FHWA, U.S. Department of Transportation, 1991.
2. E. C. P. Chang and C. J. Messer. *Arterial Signal Timing Optimization Using PASSER II-90—Program User's Manual.* Texas Transportation Institute, Texas A&M University, College Station, 1991.
3. C. E. Wallace. *MAXBAND User's Manual.* Transportation Research Center, University of Florida, Gainesville, 1987.
4. E. C. P. Chang, S. L. Cohen, C. Liu, N. A. Chaudhary, and C. J. Messer. MAXBAND-86: Program for Optimizing Left-Turn Phase Sequence in Multiarterial Closed Networks. In *Transportation Research Record 1181*, TRB, National Research Council, Washington, D.C., 1988, pp. 61–67.
5. C. E. Wallace, K. G. Courage, D. P. Reaves, G. W. Schoene, G. W. Euler, and A. Wilbur. *TRANSYT-7F User's Manual (Release 6).* Transportation Research Center, University of Florida, Gainesville, 1988.
6. A. Skabardonis and A. D. May. Comparative Analysis of Computer Models for Arterial Signal Timing. In *Transportation Research Record 1021*, TRB, National Research Council, Washington, D.C., 1985, pp. 45–52.
7. R. O. Rogness and C. J. Messer. Heuristic Programming Approach to Arterial Signal Timing. In *Transportation Research Record 906*, TRB, National Research Council, Washington, D.C., 1983, pp. 67–75.
8. S. L. Cohen. Concurrent Use of MAXBAND and TRANSYT Signal Timing Programs for Arterial Signal Optimizing. In *Transportation Research Record 906*, TRB, National Research Council, Washington, D.C., 1983, pp. 81–84.
9. E. C. P. Chang, C. J. Messer, and S. L. Cohen. Directional Weighting for Maximal Bandwidth Arterial Signal Optimization Programs. In *Transportation Research Record 1057*, TRB, National Research Council, Washington, D.C., 1986, pp. 10–19.
10. S. L. Cohen and C. C. Liu. The Bandwidth-Constrained TRANSYT Signal-Optimization Program. In *Transportation Research Record 1057*, TRB, National Research Council, Washington, D.C., 1986, pp. 1–8.
11. N. H. Gartner, S. F. Assman, F. Lasaga, and D. L. Hou. A Multi-Band Approach to Arterial Traffic Signal Optimization. *Transportation Research*, Vol. 25B, No. 1, 1991, pp. 55–74.
12. M. R. Crabtree. *TRANSYT/9 Users Manual.* U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1988.
13. M. J. Moskaluk and P. S. Parsonson. Arterial Priority Option for the TRANSYT-7F Traffic Signal Timing Program. In *Transportation Research Record 1181*, TRB, National Research Council, Washington, D.C., 1988, pp. 57–60.
14. C. E. Wallace. *Development of a Forward Link Opportunities Model for Optimization of Traffic Signal Progression on Arterial Highways.* Ph.D. dissertation. University of Florida, Gainesville, 1979.
15. C. E. Wallace and K. G. Courage. Arterial Progression—New Design Approach. In *Transportation Research Record 881*, TRB, National Research Council, Washington, D.C., 1982, pp. 53–59.
16. C. E. Wallace and K. G. Courage. *TRANSYT-7F Users Guide*, Vol. 4. Courage and Wallace, Gainesville, Fla., 1991.
17. D. I. Robertson. *TRANSYT: A Traffic Network Study Tool.* TRRL LR 253. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1969.
18. *TRAF-NETSIM Users Manual.* FHWA, U.S. Department of Transportation, 1988.

Publication of this paper sponsored by Committee on Traffic Signal Systems.

Hierarchical Framework for Real-Time Traffic Control

K. Larry Head, Pitu B. Mirchandani, and Dennis Sheppard

With the availability of faster computers and communications systems in the traffic control environment and more reliable monitoring and control hardware, better real-time control of traffic should be possible. The intelligent vehicle-highway system program now being proposed and implemented by transportation agencies, practitioners, and researchers will (a) need better real-time control methods for effectiveness in dealing with vehicle traffic and (b) allow the implementation and effectuation of better real-time traffic control. A framework for a hierarchical design of a real-time traffic control system is presented. The goal of the design concept is to respond to and monitor the various stochastic components of the traffic process with appropriate controls, frequencies, and sampling rates. The design is based on the decomposition of the traffic control problem into decision subproblems defined over different time and distance horizons. At the highest level of the hierarchy, the component process is considered that describes how over extended periods of time, travelers become aware of travel times and delays associated with the routes of a network and equilibrate into making routine route choices. At the middle level of the hierarchy, the faster dynamics are considered, for example, those dealing with traffic flows and queues during rush hours or traffic accidents. At the lowest level, the second-by-second dynamics in the traffic process are considered: the stochastic behavior of individual drivers and their responses to traffic controls at individual intersections. The conceptual design of RHODES, a prototype hierarchical traffic control system being developed for the city of Tucson, Arizona, and a comparison of its envisioned attributes with existing systems are described.

Advances in electronic control and communication technologies, coupled with significant increases in computer computational power and improvements in systems engineering and operations research methodologies, present an opportunity for significant improvements in traffic control systems. Computers have the ability to process information at rates that were only dreamed of 20 years ago. For example, telecommunication systems have utilized these technological and methodological advances to produce high-speed reliable communication between points separated by long distances. Integrated services digital network—or ISDN—technology allows the high-speed communication of voice, data, and video information over a single communication network among large groups of customers. Innovative large-scale distributed routing and flow control algorithms (based on advances in queueing theory, stochastic processes, optimization methods, and control theories) have been developed to address the problems associated with the utilization of modern telecommunication technologies.

Traffic control system design now benefits from these technological and theoretical advances in control and communication systems. Continued growth in travel demand without similar growth in new infrastructure has forced the traffic engineer to design traffic control systems that provide a higher level of performance without reducing safety and comfort. Modern communication systems allow the utilization of more information for traffic control than used by most existing traffic control systems. Synergistically, modern computers have the increased capacity required to process existing amounts of information. Furthermore, methodological advances in systems engineering and operations research can be used to design algorithms to improve system performance. The challenge for the traffic researcher is to design traffic control systems that integrate these advances.

Traditionally, advances in traffic control have resulted from extending existing models and control methodologies. This approach has been somewhat successful, but continued research in this direction does not use the available technological and methodological state of the art. The intelligent vehicle-highway system (IVHS) program proposed by transportation agencies, practitioners, and researchers presents a new structure for the traffic control system of the future (*1*). The integration of the advanced traffic management system (ATMS), the advanced traveler information system (ATIS), and the advanced vehicle control system (AVCS) components within IVHS will provide improved prediction of traffic volumes and flows and better control of the associated traffic. Under the IVHS umbrella, the solution of the traffic control problem requires new and innovative methods of information utilization and the generation of signal controls.

HIERARCHICAL DESIGN

The design of an ATMS requires a systems viewpoint of the problem, in which the entire system function must be considered. The function of a road network is to provide users a conduit for traveling from an origin to a destination. The function of the traffic control system is to manage the network so that travelers can traverse the network in a timely, safe, and efficient manner. Together these two functional components must satisfy the traffic demand placed on the system.

Within the framework of the IVHS structure, the ATMS must accept, as input, the available data on (a) travelers' origins and destinations, (b) the present and predicted traffic on the network, (c) the ATIS information provided, (d) the

K. L. Head, P. B. Mirchandani, Systems and Industrial Engineering Department, College of Engineering and Mines, University of Arizona, Tucson, Ariz. 85721. D. Sheppard, City of Tucson, P. O. Box 27210, Tucson, Ariz. 85726.

AVCS signals suggested, and (e) the geometrics of the road network. In turn, it must produce, as output, control parameters that can be communicated to both the traffic control signals and AVCS. It is within this complex structure that a hierarchical system design concept for real-time traffic control is proposed. We focus only on the traffic control of an urban street network. The developed architecture will be extended to include freeway and corridor control later.

We refer to our traffic control concept as RHODES: a Real-Time, Hierarchical, Optimized, Distributed, and Effective System for traffic control. It is intended to provide a foundation that can be implemented independently and before the full realization of IVHS and allow for an evolution of an effective ATMS component within IVHS. The design concept for RHODES is based on the consideration of the characteristics of the traffic control problem. The direct synthesis of these considerations leads to a hierarchical control structure.

The goal of RHODES is to respond to the natural stochastic behavior of traffic. This stochasticity, which is both spatial and temporal, results from independent trip generations between spatially distributed origins and destinations, driver route selections, transit traffic, pedestrians, distribution and fluctuation of vehicle speeds, network events (accidents, road closures, etc.), and driver and vehicle characteristics (headway, speed, size, etc.). The spatial and temporal response characteristics of these stochastic sources are best described on different time and distance scales. Generation of origin-destination trips and response to network events such as road closures for construction evolve in time periods of days, weeks, and months. Transit traffic and transients in traffic due to accidents and scheduled events (e.g., rush hours and sporting events) affect the network within hours and minutes. Drivers and vehicles respond in time scales of minutes and seconds to events such as phase changes at signals, moving vehicle traffic, pedestrians, and queues at intersections. Together all of these sources result in the evolution of a complex stochastic system. A real-time traffic control system must respond to the various stochastic events in the system with appropriate time constants.

Identifying the appropriate performance criteria and response time constants to events is crucial in structuring the traffic control system. When the traffic network is lightly to moderately loaded, it may be more appropriate to control traffic so that vehicles are allowed to flow as freely as possible, without stops—that is, with the objective of accommodating individual vehicles. Under heavy loads (congestion) it may be more appropriate to control traffic for better network performance, that is, to make vehicles flow to accommodate the entire network traffic instead of individual travelers.

Most existing traffic control approaches respond to the stochastic nature of the traffic by attempting to statistically smooth the data and respond to average characteristics. Whereas such approaches may be appropriate for responding to long-term, slowly varying characteristics, they fail to realize that the data also represent actual traffic fluctuations that statistical computations cannot smooth. This reduction of information use can be understood by considering an analogous problem in speech processing. Speech is a stochastic process that when measured, recorded, or coded, is corrupted with electronic and sampling noise. Both the speech and the noise that affects it are processes that contain a high degree of variance. The sound made by a single letter "s" and a sample realization of Gaussian noise are indistinguishable. The goal in speech processing is to separate the information in the speech from the measurement noise. If statistical smoothing is used, the natural variance in the speech and in the noise will be reduced. But it is the variance in the speech—the words, the notes, the tones, the silent periods—that contains the useful information. Speech-processing methods, such as linear predictive coding and adaptive noise cancellation, have been developed to address the problem of eliminating noise from the signal. Furthermore, the choice of sampling rate used in speech processing depends on the variance (frequency content) of both the signal and the noise. Similarly, for the time/distance scales corresponding to the traffic characterization at each level of the RHODES hierarchy, the sampling rates need to be chosen and estimation methods developed that eliminate the measurement noise from the corresponding signal (traffic characterization) at each level. At the lower levels in the hierarchy, at which decision time scales are in seconds and minutes, estimation problems are more difficult because both the signals and the noise may have large variances.

Figure 1 shows the functional block diagram for RHODES. This hierarchical control architecture consists of four levels of control and real-time monitoring of vehicle flow. The hierarchical decomposition of the total traffic control problem considers the problem at the highest level in an aggregate fashion as well as with a long-term systems perspective. At the lowest level, the problem is decomposed, spatially and temporally, with the resulting subproblems considering short-term details with a local (intersection) viewpoint (see Figure 2). The highest-level problem is referred to as the "network loading problem," the second level as the "network flow control problem," the third as the "intersection control problem," and the fourth as the "traffic signal actuation problem."

At the highest level of the hierarchy we envision a stochastic traffic equilibrium module for network loading, in which the decision time horizons are in hours, days, and weeks. The premise for this model is that over this period, travelers become somewhat aware of the travel times, delays, and the associated statistical characteristics of links and routes (e.g., "during peak periods it takes between 15 to 20 min to go from the intersection of Swan Road and Sunrise Drive to the intersection of Campbell Avenue and Speedway Boulevard"), and they make route choices accordingly. This results in a stochastic equilibrium, which in essence provides an estimate, in a probabilistic sense, of the predicted loads on the links of the network. Mirchandani and Soroush provide a detailed discussion of this model and the approaches to find this stochastic equilibrium (*2*). Changes in network design and land-use patterns, trends in traffic flow, and ATIS information provided to the travelers may be fed back to this decision-making function to adjust the model parameters and predict near-future loads. Essentially, this planning level of the hierarchy provides (a) a priori estimates of link loads and (b) a posterior prediction of the trends in the change of loads from real-time data. This process constitutes the outer feedback loop for RHODES, as shown in Figure 1.

Level 2 of the hierarchy represents the high-level decision making for setting signal timings to optimize vehicle flow in the network. If flows were perfectly uniform and predictable, optimal timing plans could be downloaded in an open-loop

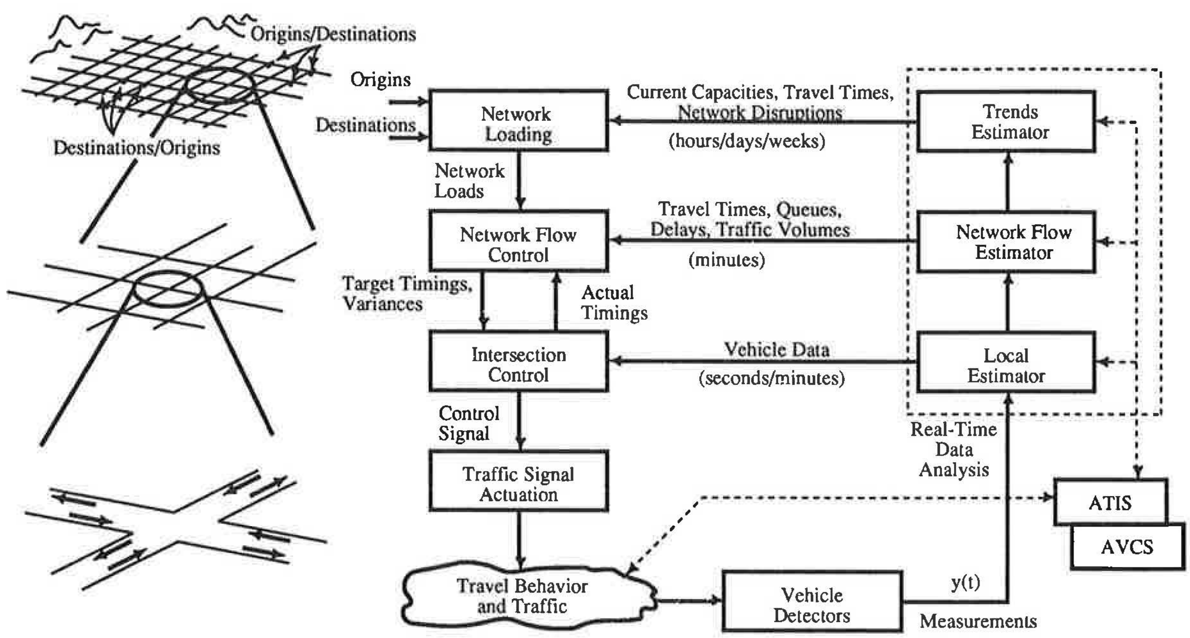

FIGURE 1 Functional block diagram of RHODES real-time traffic control system.

fashion. The computational requirements for such a fixed-time control system are not stringent; timing plans can be generated off-line using, for example, TRANSYT. This is the assumption and the process by which many current systems are set in the United States. However, flows are stochastic, and to be real-time responsive, trends in traffic volumes must be monitored, traffic volume time profiles must be estimated, and, if necessary, new timing decisions must be implemented. It is envisioned that the network flow control function at Level 2 will continually update the estimates of the traffic volumes and flow profiles with a decision horizon in the range of several minutes. Because of the potential computational complexity of this problem, it is necessary to apply methodological advances in (a) problem decomposition, (b) parallel computation and, (c) good heuristics and approximations to develop solution methods. The network flow control function forms the middle feedback loop for RHODES, as shown in Figure 1.

Conceptually, the network flow control problem can be further decomposed into two sublevels, Levels 21 and 22. The framework of the decision problem at every level is depicted

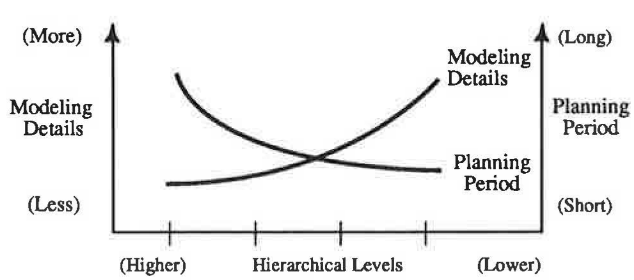

FIGURE 2 Level of modeling details and length of planning horizon considered at each hierarchical level.

in Figure 3. Here the estimator and optimizer functions explicitly take into consideration a dynamic traffic model of the form

$$x(t + 1) = f[x(t), u(t)] + \xi(t) \tag{1}$$

which states that $x(t + 1)$, the state of the system (volumes, queues, travel times) at time $t + 1$, is a function of $x(t)$, the state at time t, $u(t)$ the controls at time t, and a stochastic exogenous noise $\xi(t)$. (Equation 1 represents a discrete dynamic traffic model; a corresponding differential equation represents a continuous model.)

At Level 21 of the network flow control, the decision subproblem (referred to as the capacity allocation problem) is as follows:

> Given, at time t_0, the predicted exogenous inputs $\lambda_i(t)$ and outputs $\gamma_i(t)$, at each node i, the capacities c_{ij} on flows from node i to node j, the current travel times l_{ij} on link $[i,j]$, determine the fraction of time that "green light" should be allocated to each flow movement.

This problem can be solved as a linear programming model of the decision problem.

At Level 22, the decision subproblem (referred to as the network coordination problem) is as follows:

> Given, at time t_0, the platoon movements within the network, and approximate target allocations of green time, what should be the phase sequences and approximate green and red periods for each flow movement?

This problem could be modeled as a discrete network flow problem and should be solvable in 2 to 3 min.

In describing the subproblems at Levels 21 and 22, the performance criteria for the corresponding optimization models

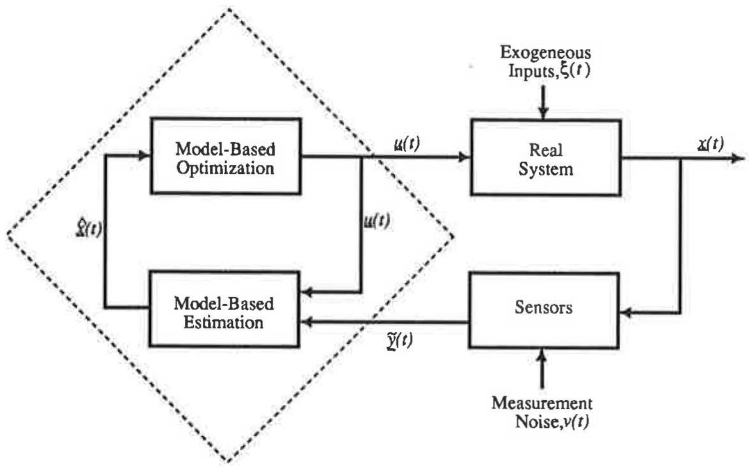

FIGURE 3 Framework of network flow control decision model at each level of hierarchy.

have been purposefully left out. The dominant optimization criterion (or criteria) at each level depends on the state of the network. Perhaps, it is most appropriate to minimize average queues when the network is saturated and congested and to minimize stops when the network use is very low. The RHODES framework allows the use of different criteria for different traffic conditions. The criteria that are most suitable must be determined through experimentation, field testing, and experience.

The intersection control at the third level can also be decomposed into the two Sublevels 31 and 32. The decision subproblem at Level 31 pertains to the determination of target timings at each intersection. The corresponding estimation-optimization subproblem (referred to as the signal scheduling problem) may be stated as

> Given, at time t_0, the traffic flow profiles entering the intersection, the phase sequences and approximate green and red periods for each flow movement, what are the optimal light change epochs for the next phase sequence?

This problem can be solved as a dynamic program in a distributed fashion (for each intersection). These local estimation-optimization problems should be solved within a 1-min time frame.

Levels 21, 22, and 31 provide target timings (phase sequences, phase times, splits, offsets) and allowable variances to the Level 32 subproblem. The allowable variances are to inform the local controllers of the sensitivity of the network flow to variations in the actual timings. These allowable variances will generally decrease as intersection saturation increases. The intersection controllers will use these timings and variances to respond to the stochastic fluctuations in traffic flow. The decision subproblem (referred to as the intersection dispatching problem) at Level 32 is a simple one:

> Based on the observable traffic on the approaches to the intersection and vehicles in the queues, should the current phase be shortened or extended?

Since the enormous number of factors that produce the short-term fluctuations in the observed traffic either are unknown or cannot be modeled, a model-based exact optimization method is not suitable for Level 32. However, concepts from artificial intelligence and learning theory may be used to develop a solution method that learns the characteristics of individual intersections and responds in real-time to the short-term traffic fluctuations that occur in a time frame of seconds and minutes.

Developments of the intersection control module and the establishment of appropriate vehicle detectors provides the inner feedback loop for RHODES for real-time local (distributed) control (see Figure 1).

The fourth control level, referred to as traffic signal actuation, is the interface with the local controller equipment. At this level, phase sequences, phase times, and offsets are passed to the controller. The key element to the success of these decisions is that field data regarding vehicles on all approaches are provided with sufficient advance notice to affect local timing decisions through control actuation.

The RHODES hierarchical structure provides a general framework for the design of a real-time traffic control system to react and affect the stochastic nature of vehicle traffic on a network. The hierarchical structure addresses the different decision and estimation problems that have different time-distance scales and different response time characteristics at each hierarchical level. Existing systems for traffic control address issues at one or two levels of the hierarchy, but none directly addresses the entire traffic control problem. In the following section several existing traffic control systems are discussed within the RHODES hierarchical framework.

COMPARISON WITH EXISTING APPROACHES

Existing traffic control systems include signal timings based on both fixed- and real-time control. The following categorization, used by many researchers and practitioners, distinguishes the mechanisms whereby signal timing adjustments are made (*3*):

- First-generation control (1-GC) involves off-line optimization and subsequent manual– or time-of-day–based implementation of new signal timing plans.

- Second-generation control (2-GC) involves generation of timing plans based on predicted trends in traffic condition and stepwise transitions among timing plans.
- Third-generation control (3-GC) involves on-line optimization (i.e., in real time) with very short (1- to 2-min) sampling periods between updates. The cycle lengths, offsets, and splits change continuously.
- One-and-a-half-generation control (1.5-GC) is a strategy with some of the features of both 1-GC and 2-GC. It involves automatic development of signal timing plans, but implementation requires operator approval.

First- and second-generation systems can be found throughout the world (*3*). In the United States, the most commonly used control system has been the Urban Traffic Control System (UTCS). Developed by FHWA during the 1970s, UTCS is capable of 1-GC (*4*) and 2-GC (*5*) control but not 3-GC control. The UTCS structure depends on time-of-day (TOD) plans that are developed off-line and on the basis of average conditions on the network during corresponding time periods and downloaded automatically at the corresponding time of day. To develop time-of-day plans for UTCS, a number of signal optimization programs have been developed and enhanced over the years, such as SIGRID, TRANSYT-7F, SIGOP, and PASSER II.

SCOOT is a notable example of 3-GC control (*6–11*). From available literature and personal communications, it appears that SCOOT makes incremental adjustments to the current signal timing plan (including cycle lengths, phase lengths, and offsets) for the next cycle, in response to changing traffic demands and suggestions by TRANSYT optimizations that are continually being performed "in the background." Red Deer (Alberta) was the first North American installation of SCOOT. Currently, installations are underway in Toronto, Ontario; Halifax, Nova Scotia; and Oxnard, California. The original prototype installations were made in Glasgow, Scotland, and Coventry, England, in 1984 (*12*); the associated evaluations indicate that it performed better than fixed-time control.

There are two real-time network control schemes developed and implemented in Australia: SCATS (Sydney Coordinated Adaptive Traffic System), and TRAC (Traffic Responsive Area Control). The more widely used system is SCATS, originally developed by Sims (*13*) for Sydney but now implemented in Melbourne, Adelaide and other cities in Australia, as well as in New Zealand and several major cities in Asia. As it is for SCOOT, very few technical details are published on SCATS. From the data available and personal communications, it appears that SCATS uses a hierarchical control architecture. At the local level, each subsystem (a set of intersections prespecified by a traffic engineer) makes independent decisions on its timing parameters (cycle, offsets, and phase lengths) on the basis of the degree of saturation in the subsystem. Adjacent subsystems "marry" and get coordinated by a higher-level regional computer when their cycle times are equal or nearly equal. Likewise, when the degrees of saturation and the consequent desired cycle lengths become different, the two subsystems "divorce." It is not clear how the timing parameters are adjusted on-line, but from observing SCATS' operations it is clear that the parameters are incrementally adjusted to varying traffic conditions to provide stability and damping in the overall control system.

TRAC is a system developed by the Main Roads Department, Queensland, that combines aspects of SCOOT and SCATS but does not perform incremental optimization (*14*). Each subsystem can have up to 12 stored plans, and the best plan is downloaded by a regional computer for implementation. The plan may be selected by average detector occupancy, time of day, manually, or, in principal, by any performance measure observable by detector data. The plans stored in each subsystem may be developed off-line, using TRANSYT for example. In personal communication, Lees indicated that plans may be continually updated depending on recent detector data and associated derived measures (*14*).

Each of the existing systems works well and addresses some of the issues that RHODES design attempts to address. The major drawback is that these systems are not proactive and, therefore, cannot easily accommodate the commonly occurring significant transients. The stochastic traffic equilibrium component at top of the hierarchy, as well as the model-based traffic predictions at each hierarchical level, allows RHODES to be proactive. A proactive system attempts to predict future demand to be placed on the network and to accommodate this demand as it evolves. Typically, control signals at each level respond to predictions over several time constants for the level. For example, at Level 31 of the intersection control, we consider predictions over time periods that may be equivalent to several cycles, as opposed to a single cycle generally used in most systems.

The intersection dispatching component at the bottom of the hierarchy is intended to make RHODES reactive to second-by-second random fluctuations in traffic and is implemented as a distributed control system. A reactive system responds to both predicted and unpredicted demand as it evolves. A distributed control system allows local control decisions at spatially separated locations.

For proper decision making at local controllers and at higher levels of the RHODES hierarchy, appropriate interconnection and communication is necessary among the levels, both through hardware-connecting processing and monitoring units and software for passing inputs and outputs to various decision-making algorithms. We cannot overemphasize the importance of modern computer and communication technologies and novel algorithmic methods to make the hierarchy of RHODES work effectively.

Table 1 presents a comparison of the preceding systems in terms of whether the system is (a) proactive, (b) reactive, (c) distributed, or (d) hierarchical. In addition, each of the systems is classified according to the timing decisions method used.

Implemented 1-GC UTCSs generally use fixed-time plans and sometimes allow for time-of-day plan selection. Implemented 1.5-GC and 2-GC UTCSs allow on-line selection of timing plans responding to time of day or detected traffic conditions, or plans are generated on-line on the basis of predicted smooth traffic flows. SCOOT and TRAC are closer to 3-GC control as characterized by McShane (*3*), in which plans can be either selected on-line or generated (and incrementally adjusted) on-line, in a time scale of a few minutes.

It is important to note that the preceding systems try to come up with timing plans for the whole network in terms of cycle times, offsets, phase lengths, and so forth. In considering whether such a strategy could be optimum, an optimum timing plan implicitly assumes the existence of steady-state condi-

TABLE 1 Comparison of Existing Traffic Control Systems and RHODES Framework

	Proactive	Reactive	Hierarchical	Distributed	Timing Decisions[a]
1-GC UTCS					I
2-GC UTCS		•			II
SCOOT		•	•		III
SCATS		•	•	•	IV
TRAC		•	•	•	III
RHODES	•	•	•	•	IV

[a]Timing Decisions are Classified as: (I) Fixed-Time Plan; (II) On-Line Plan Selection with Off-Line Plan Generation; (III) On-Line Plan Selection and/or Plan Generation; (IV) On-Line Timing.

tions at the time the plan comes into effect. Considering the very fact that a transition from one plan to another occurs and that traffic flow has some inertia associated with it, some time must lapse before steady-state may be attained. Such a plan selector and generator system cannot respond to accidents or traffic transients. Such events may introduce traffic impacts that slowly propagate through portions of the network and eventually, either leave the system or result in a new steady-state.

SCATS and RHODES attempt to provide on-line timing decisions for the given traffic loads and not on-line timing plans. In addition, SCATS does not include the consideration of predicted loads, whereas RHODES is supposed to predict appropriate traffic variables for the corresponding hierarchical levels and make optimal decisions accordingly. OPAC (Optimized Policies for Adaptive Control), a traffic control approach not discussed in this study, also provides on-line timing decisions (like SCATS and RHODES) and allows for proactive control based on predicted traffic loads (like RHODES), but the present model is suitable only for a single isolated intersection, not for a network (*15,16*).

DISCUSSION OF RESULTS

This paper has introduced a control hierarchy obtained from a system perspective of the real-time traffic control problem. Although the proposed system has not been fully developed for simulation and demonstration purposes, it has developed a conceptual design that responds to deficiencies in existing and available real-time control systems. This approach allows for an evolution of hardware and software developments. For example, constraints posed by existing signal controllers may be incorporated in the decision algorithms for determining optimal timings.

A research team at the University of Arizona and the city of Tucson, Arizona, together with the Pima County Association of Governments, is further investigating the viability of the RHODES concept. The Arizona Department of Transportation has provided research funding, and the city of Tucson has agreed to consider the implementation and field testing of the RHODES system once it has been developed and tested through computer simulation.

The hierarchical control architecture developed for RHODES parallels similar approaches used in modern manufacturing and production control systems. A factory is loaded with jobs; jobs are routed through processing centers, scheduled at the centers, and dispatched to the processing units in a hierarchical fashion to optimize appropriate measures of performance (*17*). The processes within a factory are also stochastic, and the production control problem is complex. The hierarchical decomposition of the production control problem, on the basis of time and distance scales for the various manufacturing processes, allows us to understand and solve this problem—and it is hoped that it will also aid in solving the real-time vehicle traffic control problem.

ACKNOWLEDGMENTS

The research discussed has been supported by the Arizona Department of Transportation. The authors wish to acknowledge the stimulating discussions with their colleagues who have been working on this research effort: Jeff Goldberg, Julia Higle, Paul Sanchez, Suvrajeet Sen, and Bob Wortman of the University of Arizona; George List of Rensselaer Polytechnic Institute; Tom Buick and David Wolfson from the Transportation Planning Division of the Pima Association of Governments; and Louis Schmitt from the Arizona Department of Transportation. These discussions have contributed greatly in crystallizing the RHODES architecture.

REFERENCES

1. R. K. Jurgen, ed. Smart Cars and Highways Go Global. *IEEE Spectrum*, May 1989.
2. P. B. Mirchandani and H. Soroush. Generalized Traffic Equilibrium with Probabilistic Travel Times and Perceptions, *Transportation Science*, Vol. 21, 1987, pp. 133–152.
3. W. R. McShane and R. Roess. *Traffic Engineering*. Prentice-Hall, New York, N.Y., 1990.
4. *Application of UTCS First Generation Control Software in New Orleans*. Report FHWA-RD-78-3. FHWA, U.S. Department of Transportation, 1978.
5. R. W. Kessman. *Overland Park Traffic Control System Software Documentation*. FHWA Contract DTFH61-82-00004. FHWA, U.S. Department of Transportation, July 1983.
6. D. I. Robertson. Optimum Control of an Intersection for Any Known Sequence of Vehicle Arrivals. *Proc., 2nd IFAC/IFIP/IFORS Symposium on Traffic Control and Transportation Systems*, North-Holland, Amsterdam, 1974, pp. 3–17.
7. D. I. Robertson. Traffic Models and Optimum Strategies of Control: A Review. *Proc., International Symposium on Traffic Control Systems*, 1979, pp. 262–288.
8. D. I. Robertson. Research on the TRANSYT and SCOOT Methods of Signal Coordination. *ITE Journal*, Jan. 1986, pp. 36–40.
9. G. D. Robertson. Handling Congestion with SCOOT. *Traffic Engineering and Control*, 1987, pp. 228–230.
10. P. B. Hunt and T. R. Holland. The Effect of an Incident in a SCOOT System—In Simulation and on Street. *Traffic Engineering and Control*, Feb. 1985.
11. P. B. Hunt, D. I. Robertson, R. D. Bretherton, and R. I. Winton. *SCOOT—A Traffic Responsive Method of Coordinating Signals*. Report LR 1014. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
12. J. Y. K. Luk. Two Traffic-Responsive Area Traffic Control Methods: SCAT and SCOOT. *Traffic Engineering and Control*, 1984, pp. 14–20.

13. A. G. Sims. The Sydney Coordinated Adaptive Traffic System. *Proc., ASCE Engineering Foundation Conference on Research Priorities in Computer Control of Urban Traffic Systems*, 1979, pp. 12–27.
14. J. J. Lees. The TRAC System. ARR 160, Australian Road Research Board, in J. Y. K. Luk, ed., *Proceedings—14th ARRB Conference, Urban Traffic Control Workshop*, Canberra, Australia, Aug. 1988.
15. N. H. Gartner. OPAC: A Demand-Responsive Strategy for Traffic Signal Control. In *Transportation Research Record 906*, TRB, National Research Council, Washington, D.C., 1983, pp. 75–81.
16. N. H. Gartner, P. J. Tarnoff, and C. M. Andrews. Evaluation of the Optimized Policies for Adaptive Control (OPAC) Strategy. In *Transportation Research Record 1324*, TRB, National Research Council, Washington, D.C., 1991.
17. A. Vasquez and P. Mirchandani. Concurrent Resource Allocation for Production Scheduling. *Proc., 1991 IIE International Conference on Robotics and Automation*, Sacramento, Calif., April 1991, pp. 1060–1066.

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Arizona Department of Transportation or FHWA. This report does not constitute a standard, specification, or regulation.

Publication of this paper sponsored by Committee on Traffic Signal Systems.

Detector Delay for Right Turn on Red

H. GENE HAWKINS, JR.

Delayed actuation at a right-turn-on-red (RTOR) location allows a driver to turn right onto the cross street without requiring green time. This increases the efficiency of the signal and reduces delay to cross-street motorists. However, there are no analytically based guidelines to assist the traffic engineer in selecting the appropriate length of delay for RTOR applications. A questionnaire sent to city transportation departments indicated that 100 percent based the length of RTOR delay on engineering judgment, field observation, or both. An analytical process used to determine the appropriate values of RTOR delay is described. Equations of motion and statistical relationships are applied to the three maneuvers of an RTOR. The calculations indicate that RTOR delay settings are primarily dependent on the loop length and the traffic volume in the outside lane of the cross street. The speed of cross-street traffic also impacts the length of delay but to a lesser extent than loop length or traffic volume. The resulting guidelines for RTOR settings are based on total loop length, traffic volume in the outside lane of the cross street, and speed of the cross-street traffic. The guidelines indicate that 7 to 10 sec is appropriate for locations with low cross-street traffic volumes and 9 to 13 sec, for locations with higher cross-street traffic volumes. The higher delay values in these ranges should be used for longer loop lengths and higher cross-street speeds.

The first traffic-actuated signal was installed in Baltimore in February 1928; it used a sound box as the detector (*1*). A driver sounded the vehicle's horn to obtain a green indication. Fortunately for those living near signalized intersections, detector technology has advanced considerably in the intervening years. Today, traffic engineers use the many advantages offered by actuated signals to improve traffic operations at signalized intersections.

The induction loop detector is currently the most common type of detector used with actuator traffic signals; it provides a great deal of flexibility in operating traffic signals. The loop detector consists of a loop of wire placed in the pavement and connected to a detector unit. The detector unit sends either a pulse or presence actuation call to the controller when a vehicle is detected in the space above the loop. In pulse detection, a short-duration actuation call is sent when a vehicle first occupies the loop. In presence actuation, the actuation call is maintained for as long as the vehicle occupies the loop.

One of the most useful features of the standard NEMA detector unit is the ability to delay a presence actuation call. With delayed actuation, the actuation call is dropped if the vehicle leaves the loop before the delay time expires. The delay value can be set at 1-sec increments between 0 and 15 sec and at 2-sec increments between 16 and 30 sec (*2*). Delayed actuation is especially effective at locations where a vehicle may turn right on red within a few seconds of arriving at the intersection. Delaying the call allows a right-turning vehicle to stop, select a gap, and turn right without calling an unnecessary green indication. This increases the efficiency of the signal and results in reduced delay to cross-street traffic. Heavy right-turn movements will bring up a green indication anyway, because the loop will continue to be occupied by the following vehicles.

Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843.

Optimizing the use of the delay feature in a right-turn-on-red (RTOR) location depends on selecting the appropriate delay setting. However, there are no published guidelines to assist the traffic engineer in determining the value of the delay setting. Therefore, a questionnaire was sent to city transportation departments in Texas to identify how they determined RTOR delay settings. A total of 23 questionnaires were mailed, and 15 were returned. Thirteen of the 15 respondents utilized delay at RTOR locations. Table 1 presents the values of RTOR delay identified in the survey. Delay settings ranged between 2 and 20 sec; 5 to 10 sec was the most common. In general, cities with longer loop lengths use longer delay settings, but there were exceptions. For instance, RTOR delay settings in one city ranged from 3 to 6 sec for a 50-ft-long loop, whereas the delay settings in another city ranged from 12 to 20 sec for the same size loop. Of the 13 agencies that use delay at RTOR locations, all cited field observation, engineering judgment, or both as the basis for selecting the delay setting. The size of this sample is too small to draw any statistically significant conclusions about the appropriate value of delay, but the results do indicate the variation in delay values and the lack of analytically based guidelines for determining the appropriate delay.

Although field observation and engineering judgment are vital to effective traffic signal operations, it is possible to determine mathematically the appropriate RTOR delay setting from the time needed to execute the maneuver. There are three distinct movements or elements in the RTOR maneuver, and the time needed to complete each one can be calculated from readily available information. The three movements and the associated delay (in seconds) are

1. Decelerating while in the detection area (referred to as deceleration delay, t_d),
2. Finding and accepting an adequate gap in the cross-street traffic stream (referred to as waiting delay, t_w), and
3. Accelerating out of the detection area (referred to as acceleration delay, t_a).

The deceleration and acceleration delays can be calculated easily from equations of motion and typical values for deceleration and acceleration rates. Additionally, there is little variation in these values for a given situation. However, determining the waiting delay is more complicated because it

TABLE 1 Summary of Loop Delay Timings in Texas

	Average	Minimum	Maximum	Number
Size of RTOR Loops	6' × 47'	6' × 20'	6' × 100'	13
Avg. Length of RTOR Delay (sec)	8.6	2.0	20.0	13
Delay for 6'×20' Loop (sec)	7.5	5.0	10.0	1
Delay for 6'×30' Loop (sec)	10.0	10.0	10.0	1
Delay for 6'×40' Loop (sec)	8.8	4.0	15.0	5
Delay for 6'×50' Loop (sec)	9.9	3.0	20.0	4
Delay for 6'×60' Loop (sec)	6.0	2.0	10.0	1
Delay for 6'×100' Loop (sec)	5.0	5.0	5.0	1

depends not only on the occurrence of gaps in the traffic stream, but also on an individual driver's accepting or rejecting each gap that occurs. The calculations for determining the time needed to complete each of these maneuvers are described in the following sections. Adding the individual delays provides an analytical measure of the value of the total RTOR delay setting. As with most procedures, the following assumptions have been made to simplify the calculation of the RTOR delay setting:

- Level roads,
- No sight distance restrictions,
- Intersection isolated from platooning effects (random flow),
- Vehicle makes a complete stop before turning,
- Vehicle makes RTOR into outside lane of cross street, and
- Vehicle comes to a stop before the driver searches for an acceptable gap.

DECELERATION DELAY

The first element of the RTOR delay setting is the length of time that a vehicle occupies the detection area while decelerating to a stop. This period, the deceleration delay, begins when the front of the vehicle enters the detection zone and ends when the vehicle comes to a stop. The deceleration delay is based on the following parameters:

1. Deceleration rate of the vehicle in feet per second squared ($-a$).
2. Length of the loop in the deceleration area in feet (L_d).
3. Speed of the vehicle when it begins decelerating in feet per second (V_i).

The deceleration delay is calculated from the equations of motion shown in Equations 1 and 2. Setting V_f equal to 0 and solving for t_d in Equation 2 results in Equation 3.

$$V_f^2 = V_i^2 + 2 \times -a \times L_d \tag{1}$$

$$V_f = V_i + -a \times t_d \tag{2}$$

$$t_d = \frac{\sqrt{2 \times a \times L_d}}{a} \tag{3}$$

where V_f is 0, final speed (ft/sec).

The deceleration rate ($-a$) controls the time required to come to a stop. Previous research has indicated the mean deceleration rate for vehicles approaching a yellow signal is 9.5 ft/sec^2 (*3*). This represents a maximum comfortable deceleration rate. For this analysis, it is assumed that a driver approaching a red indication will decelerate at a more comfortable rate before attempting to turn right on red. Table 2 presents the passenger vehicle deceleration rates used in this analysis.

The loop deceleration length, L_d, is the distance from the back of the loop to the point where the vehicle should stop, typically the stop line, as shown in Figure 1. Often the loop will extend beyond the stop line. However, the part of the loop beyond the stop line should not be included in the loop deceleration length. Table 3 presents the deceleration delay based on Equation 3 for loop deceleration lengths of 5, 10, 15, 20, 30, 40, and 50 ft and speeds of 30, 40, and 50 mph. Figure 2 graphically shows this relationship.

TABLE 2 Passenger Vehicle Deceleration Rates ($-a$) For Level Terrain (ft/sec^2) (*4*)

Final Speed, V_f (mph)	Initial Speed, V_i (mph)					
	15	30	40	50	60	70
0	7.8	6.7	6.2	5.9	5.7	5.6

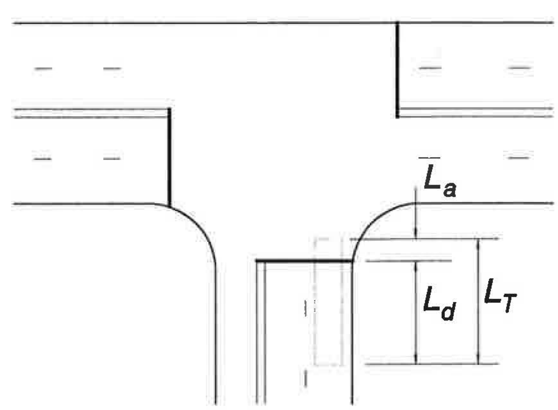

FIGURE 1 RTOR loop location.

TABLE 3 Deceleration Delay, t_d (sec)

Loop Deceleration Length, L_d, (feet)	Initial Speed, V_i, (mph)		
	30	40	50
5	1.2	1.3	1.3
10	1.7	1.8	1.8
15	2.1	2.2	2.3
20	2.4	2.5	2.6
30	3.0	3.1	3.2
40	3.5	3.6	3.7
50	3.9	4.0	4.1

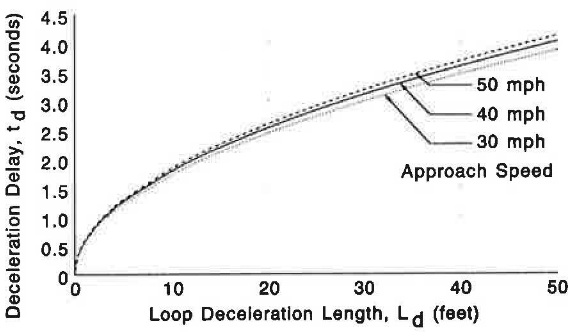

FIGURE 2 Deceleration delay, t_d.

Table 3 shows that there is little difference between the deceleration delay at speeds of 30, 40, and 50 mph for a given loop deceleration length. The deceleration delay for the 40-mph speed is ±0.1 sec from the 30- and 50-mph deceleration delay. Because the delay is set in 1-sec increments, a difference of 0.1 sec is not significant; the deceleration delay for the 40-mph speed is used to represent all speeds.

WAITING DELAY

The major and most complex element in the selection of an appropriate delay setting is the length of time that a vehicle occupies the loop while waiting for an acceptable gap in the cross-street traffic. This time is a function of two factors: the size of the critical gap and the length of time before the critical gap occurs in the cross-street traffic.

The critical gap for an RTOR maneuver at a signalized intersection can be approximated with the critical gap for a right-turn at a stop-controlled intersection. The 1985 *Highway Capacity Manual* shows the critical gap for a stop-controlled right turn from a minor to a major street to be 5.5 and 6.5 sec for major street speeds of 30 and 50 mph, respectively (*5*). From these, the critical gap for a major street speed of 40 mph can be interpolated as 6.0 sec.

The delay associated with waiting for the critical gap can be calculated from a statistical distribution of gap occurrence. Actuated signals are most effective at isolated intersections at which random arrivals are a reasonable approximation of the traffic distribution. Random vehicle arrivals can be approximated to the Poisson distribution, with the use of negative exponential distribution to represent the occurrence of gaps between the vehicle arrivals. The waiting delay is the length of time a driver must wait for the critical gap to occur in the cross-street traffic. It can be calculated from Equation 4 (*6*). Table 4 presents the waiting delay for major street outside lane volumes of 100, 300, and 500 vehicles per hour (vph) and speeds of 30, 40, and 50 mph as calculated from Equation 4. Table 4 also indicates the waiting delay for an unopposed RTOR maneuver. The unopposed waiting delay was calculated from a volume of 1 vph, because zero volume cannot be inserted into Equation 4. Figure 3 graphically shows the relationships of Equation 4.

$$t_w = \frac{3{,}600}{ve^{-vt/3{,}600}} - \frac{t}{1 - e^{-vt/3{,}600}} \tag{4}$$

where

t_w = waiting delay (sec),
t = critical gap size (sec), and
v = volume of traffic in outside lane (vph).

Gap acceptance of RTOR drivers has been addressed in a previous research study (*7*). A part of this study evaluated 359 drivers making an RTOR, of which only 202 rejected at least one gap before turning right onto the cross street. The behavior of these drivers indicated that a gap of fewer than 5 sec had little chance of being accepted and that a gap of greater than 15 sec was unlikely to be rejected. The critical gap for these drivers was found to be about 8.4 sec, as compared to the critical gaps between 5.5 and 6.5 sec used in this analysis. However, this study did not identify the effects of

TABLE 4 Waiting Delay, t_w (sec)

Major Street Volume, Outside Lane, v, (vph)	Major Street Speed (mph)			
	30	40	50	N/A
	Critical Gap, t, (seconds)			
	5.5	6.0	6.5	8.4
1	2.8	3.0	3.3	4.2
100	3.1	3.4	3.8	5.1
300	4.0	4.5	5.1	7.5
500	5.2	6.0	6.8	10.9

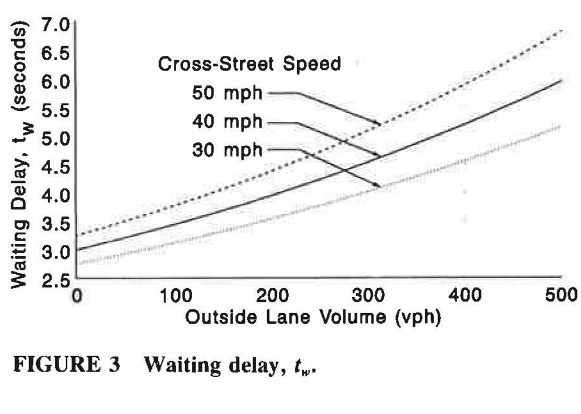

FIGURE 3 Waiting delay, t_w.

platoon flow and the cross-street speed on the critical gap. Therefore, the 8.4-sec gap is probably an overestimation of the critical gap for random flow. The waiting delay that results from a gap of 8.4 sec is shown in Table 4 for comparison purposes. Using a gap of 8.4 sec adds approximately 2 to 6 sec to the waiting delay.

ACCELERATION DELAY

The final movement in the RTOR maneuver is the acceleration off the loop into the cross-street traffic. The time required to accelerate off the loop is dependent on the length of the loop in front of the vehicle in feet (L_a), the length of the vehicle in feet (L_v), and the acceleration rate of the vehicle in feet per second squared (a).

Acceleration delay calculations are also based on Equations 1 and 2 with the following changes: the deceleration rate becomes the acceleration rate and changes from a negative value to a positive value, the initial speed becomes 0, the final speed becomes the speed of cross-street traffic, and the deceleration length becomes the acceleration length. The acceleration length is the distance from the stop line to the front of the loop. The vehicle length must be added to the loop acceleration length because the detection period will not end until the rear of the vehicle has cleared the loop. Solving for t_a yields Equation 5.

$$t_a = \frac{\sqrt{2 \times a \times (L_a + L_v)}}{a} \tag{5}$$

Acceleration rates from a stop can range from 3 to 15 ft/sec² depending on the vehicle type and the desired speed. Table 5 presents typical acceleration rates for passenger vehicles accelerating from a stop to various speeds.

If the loop extends beyond the stop line, then the vehicle must travel an additional distance to get off the loop. The distance from the stop line to the front of the loop is referred to as the loop acceleration length, L_a, as shown in Figure 1. In this analysis, loop acceleration lengths of 0, 5, and 10 ft are used. The length of the vehicle must also be included in the acceleration delay calculation. Although the design length of a passenger vehicle is 19 ft, actual vehicle lengths are considerably less (*8*). Vehicle lengths for domestic 1991 model passenger cars range from 11.7 to 18.2 ft, averaging 15.4 ft (*9*). The average of 15.4 ft is used in the following calculations. It should be noted that the total length of the loop (L_T) is the sum of the loop deceleration length (L_d) and the loop acceleration length (L_a) as shown in Figure 1 and Equation 6.

$$L_T = L_d + L_a \tag{6}$$

TABLE 5 Passenger Vehicle Acceleration Rates, a, for Level Terrain (ft/sec²)

Initial Speed, V_i, (mph)	Final Speed, V_f, (mph)				
	15	30	40	50	60
0	4.8	4.8	4.8	4.5	4.3

Inserting the values for acceleration (a), loop acceleration length (L_a), and vehicle length (L_v) into Equation 5 yields the delay associated with a vehicle accelerating out of the detection area. Table 6 shows the acceleration delay calculated from Equation 5 for loop acceleration lengths of 0, 5, and 10-ft and desired speeds of 30, 40, and 50 mph for a vehicle length of 15.4 ft. Figure 4 graphically shows the relationship. Table 6 presents that the acceleration delay for the 40-mph final speed is the same as the 30-mph speed and 0.1 sec less than the 50-mph speed. This difference is not significant because the total delay is set in 1-sec increments. Therefore, the acceleration delay for the 40-mph final speed is used to represent all speeds. If the vehicle length is increased to 19 ft, the acceleration delay increases by 0.2 to 0.3 sec.

DELAY SETTING FOR RTOR

The minimum delay setting for an RTOR location assumes that the driver selects a gap before stopping the vehicle and that the gap is immediately available. The minimum RTOR delay can be determined by adding the deceleration and acceleration delays as shown in Equation 7. Table 7 presents the minimum delay setting for various total loop lengths (L_T) and loop acceleration lengths (L_a). For loops between 5 and 50 ft, the minimum delay setting ranges between 3.8 and 6.9 sec.

$$t_m = t_d + t_a \tag{7}$$

where t_m is the minimum delay (sec).

Table 7 shows that although there is a difference of 0.8 sec between the acceleration delays for the various loop accel-

TABLE 6 Acceleration Delay, t_a (sec)

Acceleration Loop Length (feet)	Final Speed (mph)		
	30	40	50
0	2.5	2.5	2.6
5	2.9	2.9	3.0
10	3.3	3.3	3.4

t_a based on vehicle length of 15.4 feet.

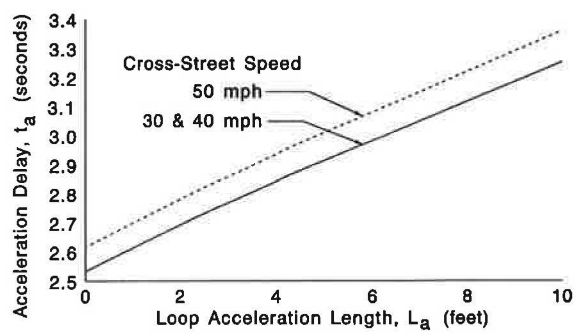

FIGURE 4 Acceleration delay, t_a.

TABLE 7 Minimum Delay Setting for RTOR (sec)

Total Loop Length, L_T (feet)	Loop Acceleration Length[a], L_a (feet)		
	0	5	10
5	1.3+2.5=3.8	N/A	N/A
10	1.8+2.5=4.3	1.3+2.9=4.2	N/A
15	2.2+2.5=4.7	1.8+2.9=4.7	1.3+3.3=4.6
20	2.5+2.5=5.0	2.2+2.9=5.1	1.8+3.3=5.1
30	3.1+2.5=5.6	2.8+2.9=5.7	2.5+3.3=5.8
40	3.6+2.5=6.1	3.4+2.9=6.3	3.1+3.3=6.4
50	4.0+2.5=6.5	3.8+2.9=6.7	3.6+3.3=6.9

[a]First number is the deceleration delay from Table 3 for 40 mph speed, second number is the acceleration delay from Table 6 for 40 mph speed.

eration lengths, the minimum delay setting exhibits a maximum difference of only 0.4 sec for the range of loop acceleration lengths. Therefore, the minimum delay setting for a loop acceleration length of 5 ft is used to represent all loop acceleration lengths for loops with a total length of 10 ft or more. A loop acceleration length of 0 ft is used for a 5-ft-long loop.

The RTOR study also evaluated the dwell or minimum delay of RTOR drivers (7). A total of 246 right-turn drivers who made a RTOR without any opposing volume were observed. The average dwell time of these drivers was 4.4 sec. This value compares favorably to the minimum RTOR delay for short loops shown in Table 7. The RTOR study did not address the increase in dwell time due to a longer loop. Therefore, it is not appropriate to compare the minimum delay for longer loops with the dwell time determined in the RTOR study. Approximately 40 percent of the drivers in the RTOR study executed the RTOR within 2 sec of their arrival at the turning position. This represents the proportion of drivers who violated the requirement to come to a stop.

The appropriate value of the delay setting for an RTOR detector can be determined by adding time to the minimum delay setting, as shown in Equation 8, to account for the waiting delay that a driver would encounter while selecting an acceptable gap after coming to a stop. The amount of waiting delay is obtained from Table 4 and is based on the cross-street speed and the volume of traffic in the outside lane of the cross-street. Values range from a minimum of 2.8 sec for a speed of 30 mph and no opposing volume in the outside lane to a maximum of 6.8 sec for a speed of 50 mph and a volume of 500 vph in the outside lane. When added to the minimum delay setting, the total delay for the given conditions ranges from 6.6 to 13.5 sec. Table 8 presents the total RTOR delay settings for various combinations of total loop length, cross-street speed, and traffic volume in the outside lane of the cross street.

$$t_{delay} = t_m + t_w = t_d + t_a + t_w \quad (8)$$

where t_{delay} is total delay (sec).

CONCLUSIONS

One of the useful features of loop detectors is the ability to delay an actuation. This feature is most often used at locations with heavy RTOR volumes. Delaying actuation allows a right-turning vehicle to complete the maneuver without calling an unnecessary green indication, thereby increasing the efficiency of the signal. A procedure for calculating the value of the RTOR delay on the basis of the length of the loop and the speed of cross-street traffic is described. The suggested values for RTOR delay apply to actuated traffic signals at isolated intersections where random flow exists.

The procedure described suggests that the delay setting for RTOR locations should be approximately 7 to 10 sec at locations with low volumes in the outside lane of the cross street. As the volumes increase, the delay setting should also increase. Delay settings in the range of 9 to 14 sec are appropriate for higher-volume cross streets. The higher values in these ranges should be used for longer loop lengths. The delay setting for 10-ft-long loops should be approximately 7 to 11 sec, and the delay setting for 50-ft loops should be about 10 to 14 sec.

Adjustments to the delay setting may be appropriate to account for factors such as grade, vehicle types and lengths,

TABLE 8 Right-Turn-On-Red Delay Settings (sec)

Total Loop Length (feet)	Volume[a] (vph) for 30 mph[b]				Volume[a] (vph) for 40 mph[b]				Volume[a] (vph) for 50 mph[b]			
	1[c]	100	300	500	1[c]	100	300	500	1[c]	100	300	500
5	6.6	6.9	7.8	9.0	6.8	7.2	8.3	9.8	7.1	7.6	8.9	10.6
10	7.0	7.3	8.2	9.4	7.2	7.6	8.7	10.2	7.5	8.0	9.3	11.0
15	7.5	7.8	8.7	9.9	7.7	8.1	9.2	10.7	8.0	8.5	9.8	11.5
20	7.9	8.2	9.1	10.3	8.1	8.5	9.6	11.1	8.4	8.9	10.2	11.9
30	8.5	8.8	9.7	10.9	8.7	9.1	10.2	11.7	9.0	9.5	10.8	12.5
40	9.1	9.4	10.3	11.5	9.3	9.7	10.8	12.3	9.6	10.1	11.4	13.1
50	9.5	9.8	10.7	11.9	9.7	10.1	11.2	12.7	10.0	10.5	11.8	13.5

[a]Volume in outside lane of cross-street traffic.
[b]Speed of cross-street traffic.
[c]Represents no opposing cross-street traffic.

vehicle acceleration rates, and driver patience and impatience, or to change the response time of the signal operation. In particular, RTOR locations with a large percentage of older drivers may require a longer delay setting. The RTOR delay setting should be increased up to 6 sec if an 8.4-sec critical gap is used, as suggested in the RTOR study (*7*). It is now common to use several short loops as a replacement for one long loop. When this type of segmented loop is used, the delay setting should be based on the length of the loop closest to the intersection. The use of the suggested RTOR delay values, combined with engineering judgment and field observation, will enable the traffic engineer to optimize the use of the loop detector delay option at RTOR locations.

REFERENCES

1. G. M. Sessions. *Traffic Devices: Historical Aspects Thereof.* ITE, Washington, D.C., 1971.
2. *Traffic Control Systems.* Standards Publication TS 1-1989. National Electrical Manufacturers Association, Washington, D.C., 1989.
3. M.-S. Chang, C. J. Messer, and A. J. Santiago. Timing Traffic Signal Change Intervals Based on Driver Behavior. In *Transportation Research Record 1027*, TRB, National Research Council, Washington, D.C., 1985, pp. 20–30.
4. A. May. *Traffic Flow Fundamentals.* Prentice-Hall, Englewood Cliffs, N.J., 1990.
5. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
6. D. L. Gerlough and M. H. Huber. *Statistics with Applications to Highway Traffic Analysis*, 2nd ed. Eno Foundation for Transportation, Inc., Westport, Conn., 1978.
7. F.-B. Lin. Right-Turn-on-Red Characteristics and Use of Auxiliary Right-Turn Lanes. In *Transportation Research Record 1010*, TRB, National Research Council, Washington, D.C., 1985, pp. 9–15.
8. *A Policy on Geometric Design of Highways and Streets.* AASHTO, Washington, D.C., 1990.
9. *Parking Dimensions, 1991 Model Year Passenger Cars*, Motor Vehicle Manufacturers Association of the United States, Inc., Detroit, Mich., 1991.

Publication of this paper sponsored by Committee on Traffic Signal Systems.

New Computerized Traffic Signal System Design

EDMOND CHIN-PING CHANG

The city of Taipei is developing a new generation of computerized signal control and traffic management systems to combat growing urban congestion. An international team has successfully formulated a new system hardware and software design adequate for long-term expansion. Design considerations, control system design, functional components, and implementation strategies are described. The new system will provide centrally located, distributed-network, real-time traffic management for a metropolis of 3½ million people into the 21st century.

Maintaining adequate traffic control and improving urban mobility has become the top priority item on the political agenda for officials and traffic managers in many countries undergoing rapid economic growth. This is especially true in Taiwan, Republic of China (ROC), since the national economy relies heavily on the transportation infrastructure to support both international commerce and trade activities. Significant development has been made in the past decade in designing and implementing computer-based signal control systems in order to upgrade urban traffic operations. Furthermore, recent system development in the city of Taipei has allowed engineers to take advantage of advanced computer network control architecture, on-line timing plan generation capabilities, and dynamic traffic control strategies.

BACKGROUND

Located along the western highway corridor in Taiwan, the city of Taipei has been upgrading its signalized intersections under central computer control within the inner city. Several development efforts began in the 1970s. The first project was awarded in 1976 to Hitachi, Japan; two subsequent phases were completed, led by the Philips group of the Netherlands. This early development effort put 119 signalized intersections and 11 arterials under centralized time-based and hardware-interconnected control. However, the signal hardware and system software are not adequate to handle dynamic traffic needs because of out-of-date control technology and inflexible timing selection for the dynamic traffic demands.

Taipei has approximately 1,000 signalized intersections within the metropolitan area, as shown in Figure 1. Among them, 119 intersections are under computerized control and 147 are under hardware-interconnected coordination. There are also 51 intersections under time-base coordination, 96 under two-way progression, 116 under partial synchronization control, and 600 under isolated, pretimed control. Local equipment includes 96 loop detectors, 34 supersonic detectors, 2 changeable message signs, and 18 sets of closed-circuit television cameras. Located in a central control center, the present central computer equipment includes a PDP11 central computer, 27 units of Hitachi control-based subsystems, 42 units of Philips subsystems, 42 domestic-made subsystems, a communication conversion system, control console, video display, communication system, and other peripherals.

In the past several years, many domestic controller manufacturers, research communities, and governmental agencies in Taiwan have successfully developed several generations of microcomputer-based signal control systems that are equivalent in function to the closed-loop systems used in the United States. These software systems, such as COMDYCS and TRUSTS, can affectively assist signal timing plan development during daily operations. Many city and county officials have implemented these systems in traffic responsive mode with on-line optimization software based on the modified PASSER, TRANSYT, MAXBAND, and other packages. However, these systems were developed for signal networks with fewer than 90 intersections. There is a need to design a large-scale computerized traffic-responsive system that can respond to the metropolitan traffic control needs for the city of Taipei.

Since 1989, an international team consisting of the Taipei City Traffic Bureau; China Engineering Consultants, Inc. (CECI); DelCan International, Inc.; and overseas Chinese consultants has successfully formulated a long-term system hardware and software development plan. This new computerized traffic control and management system is being implemented in three development phases to update the signal control capabilities for the entire metropolitan area. The overall system development involves designing controller hardware, system control software, and traffic management strategies. This control operation, similar to the Urban Traffic Control System (UTCS) 1.5-GC system concept, has already been implemented; initial system integration tests were completed successfully during October 1991. This system design can comfortably coordinate more than 1,000 signalized intersections for the operational staffs in the traffic bureau to implement various traffic management techniques in the future.

OBJECTIVES

The overall system development was focused on developing a comprehensive system hardware and software design to improve traffic control capabilities by upgrading the existing

Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843.

FIGURE 1 Taipei traffic signal control system.

traffic signal control systems used in the Taipei metropolitan area. Two studies—*Taipei City Signal Control System Upgrade Feasibility Study* and *Taipei City Signal Control System Manager Project*—were completed. The detailed study tasks included conceptual system design, functional specifications, staged development planning, and integration software. The studies also provided system manager supports during the system hardware contracting and implementation process.

The study had three important design objectives. First, the strategy development objective was to devise the timing design to increase arterial circulation of all signalized intersections in the inner city and outlying areas. This systemwide control strategy will improve travel time along major arterials, thereby encouraging more systemwide arterial progression than the previous operations. The new system will enable engineers to obtain quickly real-time traffic measures, identify congestion hot-spots, and develop feasible signal coordination schemes. The system architecture was designed with hardware and software flexible enough to accommodate a wide range of urban control strategies. These include time-of-day operations, traffic responsive operation, and dynamic timing plan generation, they also provided a system communication and execution hardware and software infrastructure sufficient for experimenting with various adaptive signal control schemes. The study developed an open-end, large-scale control system architecture, a new-generation traffic signal controller, and assistance to the domestic electronic industry and signal controller manufacturers in implementing the upgraded design.

The system development process, control strategy, communication schemes, computer network architecture, and other advanced design features being implemented are summarized. This traffic control system will provide city traffic engineering staffs with a centrally located, distributed-network, real-time traffic management infrastructure that can be used to battle growing urban congestion problems within a metropolis of 3½ million people.

DESIGN CONSIDERATIONS

City officials and traffic managers have provided three guidelines during the new system development. These design considerations are addressed individually and well understood as the new system being developed for the potential implementation throughout the metropolitan area. First, this control system provides "open-architecture" system expansion capabilities for easy adaptation to future generations of control hardware and advanced control strategy. Second, the system will provide dynamic, on-line signal timing plan generation so that control operators can make up-to-date timing adjustments to accommodate traffic pattern changes during reconstruction, maintenance operations, recurring, congestion, highway incidents, and special events. Third, the new computerized system will provide a powerful traffic management tool that can take advantage of both foreign and domestic hardware, as well as software systems already being developed.

Operational Features

Most domestic system developments emphasize upgrading the control hardware and software for the smaller systems that provide basic pretimed, time-of-day operations. The system will integrate effective traffic control strategies into the system hardware and software design to improve traffic management capabilities that can apply existing technologies more efficiently. This computerized traffic signal control system has the following traffic control and operational features:

1. An effective traffic management system for responding to both short- and long-term changes in traffic demands and roadway conditions within the urban traffic network;
2. The capability to allow city control center operators and traffic engineers to develop traffic signal timing optimization plans quickly from updated traffic information;
3. The ability to evaluate real-time traffic control system performance as affected by the present and proposed signal timing control strategies in various traffic operating conditions; and
4. An open-architecture and user-friendly signal system control software based on advanced computerized hardware configuration to accommodate advanced and future on-line signal control operations.

Implementation

The overall control system upgrade is divided into three phases, and a number of signalized intersections and detectors were planned for installation in 1989. These system implementation efforts include replacing intersection controllers, installing the central control computer, renovating the communication network, developing control system software, and, after development is completed, providing system integration supports to assure maximum technology transfer to city traffic engineers for future system maintenance and operations.

Phase 1 of the system development was to install all the field equipment and central facilities to initially integrate the system hardware and software on 301 controllers and 200 detectors in the central business district. Phases 2 and 3 have also been awarded and are currently being implemented. The initial system integration tests on the field equipment, control center, and system control software for Phase 1 were scheduled to be completed at the end of 1991. The system integration carries a 2-year warranty period during which all the primary contractors are responsible for overall system hardware and software maintenance. After the completion of technology transfer activities the system will be operated and maintained by the traffic engineering staffs of the Taipei City Traffic Bureau.

For the system currently being upgraded, all the signalized intersections are equipped with pre-timed operations (PT), time-of-day (TOD), and dynamic computation (DC) operations. Some intersections also employ critical intersection control (CIC), and one or two intersections will be under experimental adaptive control (AC). However, all intersection controllers will be equipped with identical software and hardware capable of upgrading with all the control features for the full-scale implementation.

Besides the traffic signal control system upgrade project, Taipei is also undergoing several citywide infrastructure improvement projects and rapid mass transit construction. A number of urban freeway and expressway reconstruction projects are also underway, along with the relocation of at-grade railroad crossings underground within the urban area. Therefore, the control system design must be flexible enough to accommodate locations of both temporary and permanent control centers, meet traffic control needs during the roadway construction, and provide maximum system future expansion capabilities.

Potential Benefits

Motorists in the metropolitan area will benefit substantially from the enhanced traffic control system features compared with the existing traffic control system. Through field observations and limited TRANSYT-7F network simulation, the feasibility study indicates that coordinated signal system operations, once completed and fully operational, can improve overall traffic operations with reduced travel time, delay, and a systemwide reduction in fuel consumption and air pollution.

Even with the conservative estimation, the potential reduction on the systemwide travel delay, improved travel time, and reduced fuel consumption can be converted to an approximate total dollar savings of nearly $1 billion (U.S.) a year. These system implementation benefits can be converted into an optimum rate of return on the overall investment of a 14:1 benefit-to-cost ratio during each year of system operation.

FEATURES

The following sections describe system operational features and unique control features available in the Taipei traffic signal control system design. These features include traffic control strategy, system control software, color graphics display, central communication unit, phone system, portable PCs, closed-circuit television system, and control center system interface to provide improved motorist information for eventual potential implementation throughout the entire city street network.

Control Strategy

Among the 1,000 intersections being upgraded, 301 traffic signals are inside the central business district and 349 are in the grid pattern. The remaining signals are in outlying areas. Phase 1 development focused on the city's central business district. Of the total 301 intersections, 5 will use queue control detectors, 4 will use actuated detectors, 6 will use critical intersection control, and 1 will use adaptive control. All installed controllers in the Taipei system will conform to the Ministry of Transportation and Communications Institute of Transportation (MOC/IOT) standard for solid-state, 16-bit–based microprocessors.

As shown in Figure 2, all intersections will be equipped with traffic system control features through either system automatic or operator selectable commands. The new system is designed with enhanced PT, TOD, dynamic table lookup (DTL), DC, and CIC modes with the capability of allowing AC in the future. Adaptive control software will initially be implemented at one or two selected intersections if successful, and may eventually be implemented throughout the entire metropolitan area.

Control System Architecture

The new control center will be upgraded in the existing location. A microcomputer-based interim control system was set up to monitor field equipment during the system hardware replacement. The permanent traffic control management center will be located in a building that also houses the central bus transit terminal in a city-owned joint development project. Along with the mass transit control center, the center will be used to coordinate the metropolitan transit system and traffic signal control center. In addition to accommodating all traffic signal system control hardware and software, this permanent control center will provide adequate office space for the traffic operations and signal maintenance staff. This arrangement provides better communication without requiring human operator intervention between those that optimize traffic operations and those that monitor the operating system around the clock. The target date for the completion of field implementation was the end of 1991.

As shown in Figure 3, the control system includes a communication network, dedicated 32-bit single-board computer (SBC), backup host computer, and a traffic control workstation (TCWS) with a Chinese-based operator interface. TCWS is a stand-alone, RISC-based system with graphics CPUs, keyboards, digitizing tablets, and high-resolution monitors (1,024 × 800 pixels, 19 in.). Besides providing traffic surveillance and real-time monitoring, TCWS can also display the control status of selected intersections, signal groups, arterial streets, and area operations. This real-time information includes error status display; red, yellow, and green indications; walks and don't walks; detector volumes and occupancies; vehicular types; intersection split-timing, signal phasing data; and many other advanced control features. TCWS will provide the operator command interfaces for initiating various control strategies. The TCP/IP- and NFS-based networked workstations can share real-time traffic and control information intelligently with any other units through VME Bus, Ethernet, SCSI, and RS-232C serial connections.

Based on the single-function, volume-distributed design concept, the control computer is configured with a series of Ethernet and VME NET connected 32-bit SBCs. Depending on the needed system surveillance, communication control, and display software functions, these SBCs will be divided further into regional traffic control computers (RTCCs), data gathering control computers (DGCCs), central communication units (CCUs), wall map control computers (WMCCs), and message display control computers (MDCCs). For the distributed data communication, two DGCCs will supervise six dedicated DGPs with 128 communication lines connected on each DGP. For the control function, RTCC can control 64 pieces of field equipment, such as intersections, detector units, and changeable message signs. In addition, one extra RTCC will serve as the on-line backup unit. It is important to recognize that all the targeted control functions are achieved mainly through these SBCs. These multi-CPU host computers are used mainly for system software development, operational monitoring, database server management, and system maintenance.

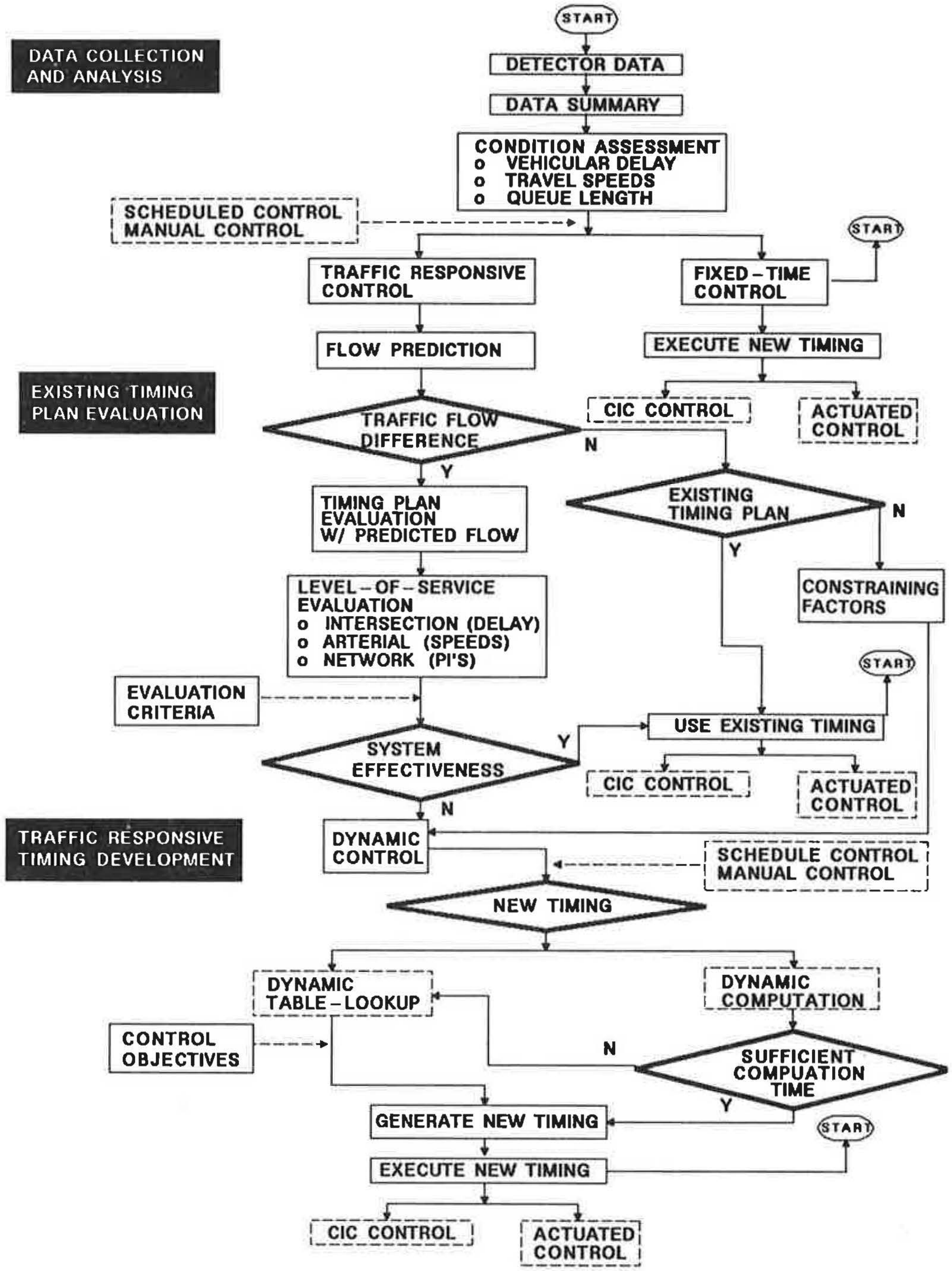

FIGURE 2 System traffic control strategy.

The control center is equipped with a large tile-assembled wall map to represent Taipei's street network currently under computerized signal control. In addition to display control area, the map will provide real-time operational and performance information to the control operator through assembled LED indicators, dot-matrix message boards, system measure of effectiveness (MOE) ranking scoreboards, and video screens. The high resolution RISC-based workstations, based on the Chinese graphics kernel system, can also display geographic-based traffic information. The operator can use multiwindow, zoom, and pan capabilities to modify the display screen to any selected scale down to one signalized intersection. The network communication speed of the TCWS color graphics computer will allow instant switch display from one TCWS to another.

System Software

This new computerized control system applies design concepts similar to the public-domain UTCS software, as supported by FHWA, U.S. Department of Transportation. The completely

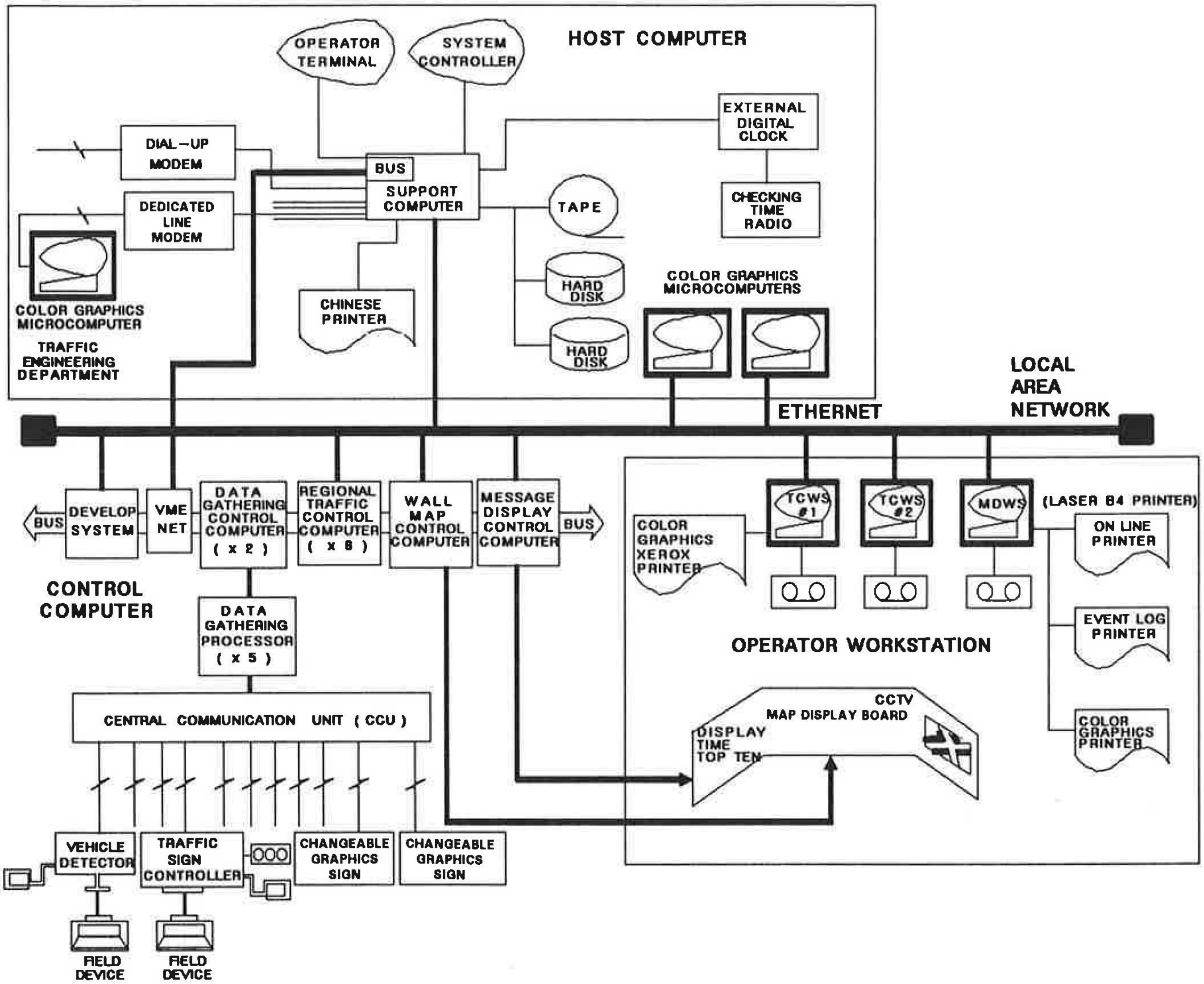

FIGURE 3 Control system network architecture.

new control software, developed jointly by CECI and the consulting team, is fully equipped with many advanced signal control features to suit Taipei's unique operational characteristics and traffic patterns. To support the most critical processing requirement adequate for future real-time adaptive control, the system uses advanced microcomputer electronics, network client-server communication schemes, fault-tolerant control hardware architecture, and simultaneous real-time system data base mirror-backup design. The new control system design has basic system functions similar to 1.5-GC UTCS system software and is equipped with on-line data base management in a distributed computer networking environment. The system can obtain all historical information stored in the controllers and local detector units. The control system will use either smoothed real-time or default historical volume and occupancy data for timing plan regeneration.

Detector System Design

The city of Taipei has more operational problems with heavily mixed traffic flows in the signal network than many other metropolitan areas in the world. These mixed-traffic problems result from a high percentage of bus transits, motorcycles, pedestrian crossings, and on-street parking. Since the geometric conditions and available roadway capacity are very limited at most signalized intersections, one major design objective is to identify the traffic volume and percentage of motorcycles presented through the intersection area at the beginning of the green light. The identification then would allow adjustment for the operational impacts and would provide a better assessment of available saturation flows for improving intersection signal timing design. Similar to the treatment of bus traffic, all motorcycle units will be converted to adjustable equivalent passenger car units (PCUs) in the signal timing optimization.

As shown in Figure 4, a double-loop detector was designed to collect detailed information, such as volume, occupancy, travel speeds, and vehicle classification. With this detector design at divided traffic lanes, the detector system can accurately detect the presence of a passenger car (95 percent), bus (80 percent), and motorcycle (90 percent). Because of the limits of roadway reconstruction, real-time data processing efficiency, and historical data storage requirements,

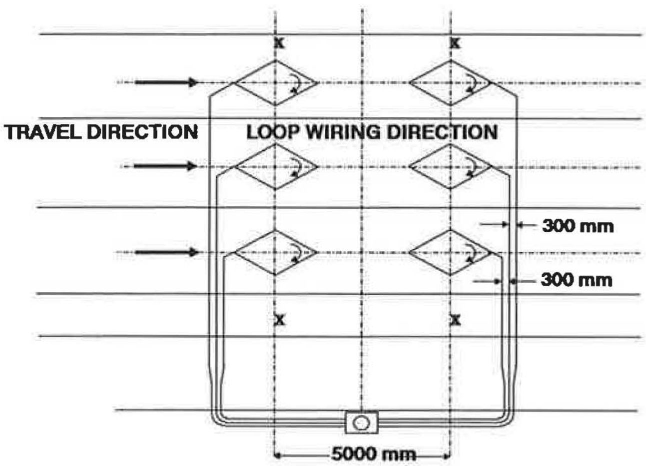

FIGURE 4 Double-loop detector system design.

a separate detector cabinet was designed. The detector unit will house the detector processing unit, provide a data consistency check, apply historical smoothing during detector failure, and use field-to-central communication schemes similar to those used in the traffic signal controller.

Central Communication Unit

Two communication systems were developed to collect and process traffic information, video images, and control commands between the CCU and field equipment. The first communication system uses the MOC telephone network through 1,200-baud dial-up modems and links data wires to the separate signal controller and detector units. The system will later be upgraded with twisted-pair cable and fiber optics during video surveillance upgrade. The system will continuously monitor all field devices through second-by-second or operator-selected, fixed-interval communication to each piece of field equipment. In turn, the control system will receive and confirm response messages. The majority of the response messages are based on the Communication Protocol Agreement developed by MOC/IOT for providing the needed communication interface between controllers from different signal manufacturers within the same traffic control system.

Basic communication message contents include current signal phases, pedestrian call locations, detector volumes, occupancy values, and so on. These transmitted commands and response messages will be updated for every signalized intersection and detector unit. In operation, the central computer will coordinate all signalized intersections by issuing needed communication commands. The local controller recognizes these commands as a prompt to either force out or leave the current signal timing interval and go to the next interval or to needed plan switching activities. These command sets will ensure that network coordination is adequately maintained in order to maintain signal coordination settings throughout the city.

Taipei's DGP unit uses one-to-one data communication modems to receive and transmit data between the CCU and field equipment. The DGP unit will also synchronize time-based control during an interconnect or central computer failure. Each unit will control up to a maximum of 64 separate units, whether local signal controllers, local detector units, or changeable message signs. Through a telephone service agreement, MOC is providing the traffic bureau with channels for the local controller and detector units. The city will pay only the actual cost to extend telephone lines, to drop connections to signal cabinets, and for special electronics transmission. This decision can significantly reduce the initial capital investment as opposed to developing the system's dedicated communication network. MOC maintains a very extensive telephone network. Using the telephone interconnect is relatively the most inexpensive solution, compared with other alternative communication schemes, before establishment of the city-owned fiber communication/fiber trunk network during the closed-circuit television (CCTV) video transmission upgrade.

Phone System and Portable PCs

Because the majority of the communication systems will be based on the local telephone network, two serial RS-232C modular communication ports are included in all traffic signal controller and detector cabinets to establish data communication with the control center. In this way, field technicians can log on the central host control computer and provide data upload and download conveniently to improve system operations in the field. This local system can also interface with portable laptop microcomputers and hand-held microcomputer units for field equipment interface. While at the controller cabin, the technician can plug these units into one of the two RS-232C serial communication ports as well as dial up the central computer for control status reports on the system. These portable units will also conduct systemwide signal-timing, implement immediate field changes, and update modifications.

Closed-Circuit Television System

The existing CCTV system will continue to provide basic traffic surveillance, monitor selected arterials, and offer visual confirmation of field operating conditions. With the CCTV, engineering staffs in the control center or other locations can quickly monitor traffic operations status, assess the operational impacts of special events, and detect and confirm unusual incidents, including emergency situations. The current CCTV system consists of 11 video cameras strategically located on buildings or poles for wide coverage of both major downtown streets and key access points to and from the freeway and expressway system.

Currently under design, a combined coaxial cable and fiber trunk communication network will initially improve video signals and eventually upgrade all the field devices to the control center at the high-speed, high-quality transmission. At the control center, four 19-in. color monitors will be mounted on a map display board to allow operator-selected display of the video cameras for condition assessment. One 13-in. monitor will be mounted on the control desk. Using a dedicated CCTV control panel, system operators can select which camera to view on which monitor and make adjustments on the remote video cameras. A video tape recorder will record special events or gather specific traffic information for traffic operations studies.

Motorist Information Service

Several prototype experiments are being developed worldwide to demonstrate the potential benefits of providing improved traffic information service to the general public and motorists before and en route to their final destination. In particular, several route guidance systems are being integrated to provide real-time traffic information using the dead-reckoning wheelbase, in-vehicle equipment, massive on-board digitized map storage, global position system, or automatic vehicle control systems. However, these complicated route guidance systems, traffic network optimization schemes, and the intelligent vehicle-highway system (IVHS) will not be able to function properly without reliable and updated traffic data base information collected from the traffic control systems.

Besides the various traffic signal control functions currently being developed, the up-to-date traffic operating condition and system status data base will also provide the basic infrastructure for improving travel assistance information using techniques similar to the computerized weather network service currently available in the United States. The selected motorist travel assistance information network may include network travel time information, traffic loading conditions, network congested hot-spots, scheduled roadway maintenance locations, and current incident locations. The potential users may include control centers, other city departments, television and radio stations, traffic service agencies, major office complexes, individual motorists, and research agencies.

Based on the electronic bulletin board system (EBBS) concept, a graphics-based software system design can allow any potential user of the highway network, using a remote PC terminal emulation program to be supplied by the city or third-party software vendors, to log on the computer system, retrieve partitioned data bases, and examine remotely the system situation map and current network performance. Since only restricted and summarized data base information will be made public, users can access the small, extracted data base through the EBBS using the telephone modem dial-up, direct communication lines, or videotext method. With the availability of various communication media, such as telephone, cable news network, and cellular telephone, the user can examine the network traffic status and adjust route selection in advance or enroute to their final destination.

Control Center Interface

Since the 1980s, the Taiwan Area National Freeway Bureau and MOC have jointly developed a comprehensive, computerized freeway traffic surveillance, communication, and control (SC&C) system to manage the freeway system in the northern section between the Keelung and Yangme section of No. 1 Freeway. The initial system concentrated on traffic surveillance in the northern freeway corridors. An emergency phone network is being extended to the southern portion of the freeway system for continuous traffic monitoring along the entire freeway corridor. The control system will also be expanded to the second freeway and other interconnected expressway systems on the island of Formosa.

Traffic information systems are an integral element of the freeway corridor management system that includes driver information system, traffic surveillance, vehicle detectors, and incident management. The driver information system applies freeway guide signing, changeable message signs, and AM/FM radio. Through vehicle detectors, CCTV surveillance, and emergency telephone network, traffic surveillance provides needed traffic control for the northern portion of the western freeway corridor. Through reserved voice communication channels, traffic data communication display, and related changeable message displays, control center coordination will enable city traffic control systems and the freeway traffic surveillance center to share essential real-time traffic information, status display, and coordinate control strategies. The computerized traffic control system will also provide the needed media interface to relay vital traffic information and recommended actions to motorists during freeway incidents or urban congestion within the metropolitan area.

The city traffic control system and freeway bureau's traffic surveillance center will work together through the improved system communication and operational status display to improve the traffic information service within the Taipei metropolitan area. The freeway bureau is primarily responsible for the freeway system, and the traffic bureau handles the urban street system. The two systems will be coordinated to form a complete highway information system through reserved voice communication channels, traffic data communication displays, and related changeable message displays. Successful implementation will depend on close agency cooperation between the traffic control centers of these two operating agencies.

In addition to the freeway systems, several municipalities are also developing computerized traffic SC&C systems to support freeway corridor communication and control activities. Often, the freeway agencies will handle the freeway system and the city traffic engineering department will handle urban street system with separate control objectives. The initial system development primarily concentrates on upgrading the traffic surveillance capability in the urban arterial segment, but work is still under way to integrate available information and operational management related to the entire corridor. The success of this integrated computerized traffic information system depends heavily on improved communication and cooperation among various operating agencies through control center interface capability.

CONCLUSIONS AND RECOMMENDATIONS

Increasing mobility and traffic service are vital to the economic growth and well-being of any nation. Urban transportation systems contribute significantly to overall economic growth through improved work-force productivity, business investment patterns, international competitiveness, and economic growth. The rapid expansion of urban areas, along with increasing population concentration, automobile ownership, freeway network development, and infrastructure reconstruction, has resulted in urban congestion in many developing countries. To sustain long-term economic growth and international commercial development, the transportation infrastructure must be able to move people and goods efficiently from origin to destination with minimum traffic delays and vehicular stops. Successful operations depend on the cooperation of reliable field operational equipment, control hardware platform, system control software, operating strategies, and the rapid identification of any equipment malfunction to assist operators in making system improvements. For the city of Taipei, this means expanding the existing traffic control system.

The new design begins with a system control strategy developed to improve the arterial circulation capability of signals on the city's major arterials, then the minor arterials and collector streets, and finally the signal intersections in outlying suburban areas. Now being implemented, an open-architecture system design will enable new subsystems to be added to the existing computer system at a relatively low cost. With this modular design concept, all traffic signals in the city may be connected eventually to the centrally located, distributed-network, real-time control system. Unlike previous development efforts, extensive, on-the-job training and technology transfer were achieved by the development team approach throughout the system design and management process. Thus, once the central system framework is established, city forces and domestic manufacturers can be responsible for traffic operations and system maintenance and expand the system according to self-contained system specifications.

Computerized traffic responsive strategies are in the experimentation stage as the system is being integrated. Staff training, operational capability, and response philosophy will continuously be enhanced as more operational experience is obtained. Nevertheless, the overall traffic control environment, including the advanced hardware platform, integrated system software, and control strategy design, has already given the city the necessary basic building blocks adequate for long-range expansion. The implementation of this system integration approach will result in positive traffic management for the city of Taipei and for other similar developing countries to meet growing urban mobility needs and sustain economical growth well into the 21st century.

REFERENCES

1. China Engineering Consultants, Inc. *Taipei Traffic Signal System Design*. Final Report. Taipei City Traffic Bureau, Taiwan, R.O.C., May 1990.
2. China Engineering Consultants, Inc. *Taipei Traffic Signal System Software Integration Specification*. System Design Document. Taipei City Traffic Bureau, Taiwan, R.O.C., Jan. 1991.
3. China Engineering Consultants, Inc. *Western Highway Information System Design*. Final Report. Institute of Transportation, Ministry of Communications, Taiwan, R.O.C., Dec. 1990.
4. *Traffic Control System Handbook*. FHWA-IP-85-11. FHWA, U.S. Department of Transportation, April 1985.
5. H. S. Tsay. A Microcomputer-Based On-Line Traffic Control System. *ITE Journal*, Vol. 59, No. 12, Dec. 1989, pp. 29–36.

Publication of this paper sponsored by Committee on Traffic Signal Systems.

TRANSYT-7F: Enhancement for Fuel Consumption, Pollution Emissions, and User Costs

MICHAEL A. PENIC AND JONATHAN UPCHURCH

Traffic Network Study Tool Version 7F (TRANSYT-7F) is a macroscopic deterministic traffic signal system simulation and optimization computer program. The program is used to optimize the performance of urban signal systems with respect to delay, number of intersection stops, and, most significantly, fuel consumption. The fuel consumption algorithm that has been used in the TRANSYT-7F program produces estimates that are not representative of the performance of the current U.S. vehicle fleet. A new set of interrelated algorithms that consider current vehicle fleet performance is described as well as a number of other factors for estimating fuel consumption. Besides considering fuel consumption, these algorithms have been expanded in scope to include estimates of air pollutant emissions and user costs. These additions and changes will result in the expansion of the TRANSYT-7F performance data summary to cover a wider range of measures of effectiveness describing the quality of traffic signal system performance.

TRANSYT (Traffic Network Study Tool) is a traffic signal system simulation and optimization computer program. TRANSYT uses a macroscopic deterministic platoon dispersion model to simulate the flow of traffic through a street network. Version 7F of TRANSYT (TRANSYT-7F) is used extensively throughout the United States to optimize the performance of urban signal systems with respect to delay and number of intersection stops.

Besides considering delay and stops, the traffic engineer analyzing a signal system with TRANSYT-7F may also be interested in system simulation or optimization with respect to other measures of effectiveness (MOEs) such as fuel consumption, pollutant emissions, and user cost. The estimates of fuel consumption currently produced by TRANSYT-7F are based on vehicle performance data obtained in the late 1970s; therefore, they are out of date. In addition, there are some deficiencies in the fuel consumption model that cause the estimates to be inaccurate. The TRANSYT-7F program currently does not produce estimates of pollutant emissions.

A user cost model was added in Release 6 of TRANSYT-7F. This elemental user cost model considers the cost of fuel, the cost of driver and passenger time, and the cost of operating the vehicle. However, the existing model does not consider the individual components of cost associated with the operation of a vehicle.

M. A. Penic, Johannessen & Girand Consulting Engineers, Inc., 6611 North Black Canyon Highway, Phoenix, Ariz. 85015. J. Upchurch, Department of Civil Engineering, Arizona State University, Tempe, Ariz. 85287.

This paper describes a research effort that was initiated to improve and update the TRANSYT-7F fuel consumption and user cost model, as well as add a new model for producing estimates of vehicle pollutant emissions. If incorporated into TRANSYT-7F, these models will produce

1. Fuel consumption estimates that are updated and more reliable than the estimates produced by the existing TRANSYT-7F fuel consumption model;
2. Reliable estimates of vehicle pollutant emissions not currently estimated by the program; and
3. User cost estimates based on fuel consumption, pollutant emissions, driver and passenger time, and vehicle depreciation and maintenance costs.

EXISTING TRANSYT-7F FUEL CONSUMPTION MODEL

The existing fuel consumption model for TRANSYT-7F was calibrated in 1981 (*1*). That model uses the following MOEs to compute fuel consumption:

1. Total travel (vehicle-mi/hr),
2. Total delay (vehicle-hr/hr),
3. Total stops (full stops/hr), and
4. Free speed on each link (mph).

Actual vehicle fuel consumption data was collected using one test vehicle. The data were normalized to represent the 1981 U.S. vehicle fleet before being applied to the existing fuel consumption model. The data collected that were applicable to the TRANSYT-7F model included the following:

1. Fuel consumption for uniform travel speeds,
2. Fuel consumption for normal stopped idling, and
3. Excess fuel consumption (in addition to free speed consumption) resulting from a full braking stop at a normal deceleration rate from free speed, and a normal acceleration back to the same free speed.

The data were applied to an elemental model using stepwise multiple regression. Some of the limitations of this model that were pointed out at the time of its creation are listed:

1. Only one test vehicle was used to derive the relational parameters before the parameters were adjusted to represent the U.S. vehicle fleet.

2. The model does not consider the effects of
   - Traffic congestion,
   - Trucks and diesel engines,
   - Road grade and curvature,
   - Pavement surface quality, or
   - Temperature.

REQUIREMENTS FOR NEW FUEL MODEL

To develop an improved fuel consumption model for TRANSYT-7F, it is necessary to do the following:

1. Obtain updated fuel consumption data that represent the current U.S. vehicle fleet. These data must be collected with appropriate consideration given to the wide range of engine types and performance capabilities available.
2. Evaluate the relative contributions of geometric and environmental factors to overall fuel consumption and determine a method of accounting for those factors that can be addressed within the data structure and framework of TRANSYT-7F.

Each of these tasks is addressed in the discussion of the development of the proposed fuel consumption model.

Source of Fuel Consumption and Pollutant Emissions Data

This section discusses the source of the fuel consumption, pollutant emissions, and user cost data used to calibrate the models developed. A 1985 study completed by R. McGill of Oak Ridge National Laboratory for FHWA summarizes the most recent fuel consumption and pollutant emissions data available (*2*). The test procedure used combined laboratory and on-road tests using 15 vehicles. Data were collected in tabular form as a function of both acceleration and velocity. All 15 vehicles were tested for fuel consumption; 6 of them were tested for pollutant emissions of carbon monoxide (CO), hydrocarbons (HC), and nitrogen oxides (NOx).

These data fulfill the first requirement previously defined for developing a new fuel model. First, the data were collected from a set of vehicles whose performance capabilities are representative of current automotive technology. Second, the vehicles tested were selected beforehand to be representative of the wide range of performance capabilities of the current U.S. fleet. These fuel consumption data also include a representative fraction of diesel engines.

Upon completion of the tests, the consumption and emission values from all of the vehicles tested were averaged in proportion to each vehicle's contribution to the January 1986 U.S. vehicle fleet.

Geometric and Environmental Factors

Besides producing an updated fuel and pollution data set, the Oak Ridge study produced equations that describe the variation in fuel consumption and pollutant emissions as a function of roadway grade and temperature.

Roadway Grade

The effect of roadway grade is accounted for by increasing or decreasing the engine acceleration rate to counter act the acceleration due to gravity. The equation used to calculate the change in acceleration rate due to grade can be derived easily using a force balance. The effective acceleration rate due to grade is computed using the following equation:

$$A_g = 32.2(G)/[(10)\ (100 + G)^{1/2}]$$

where A_g equals effective engine acceleration due to grade (ft/sec^2) and G equals roadway grade (percent).

Temperature

The research that generated the new fuel consumption and pollutant emissions data also generated an approximate relationship for the effect of temperature for small cars and for medium to large cars (*2*). The correction factor is computed using the equation

$$K_t = (-0.004334)(T - 78.258) + 1$$

where K_t equals a temperature correction factor and T equals the ambient temperature in degrees Fahrenheit.

FUEL CONSUMPTION AND POLLUTANT EMISSIONS DATA REDUCTION

The fuel consumption and pollutant emissions data developed by McGill were provided in a form suitable for use in a microscopic simulation model (*2*). Because TRANSYT-7F is a macroscopic model, it was necessary to reduce the data tables to a form suitable for use in a macroscopic model. It was also desirable to fit the data to curves, since computer table lookups are far more memory-intensive and time-consuming.

The fuel consumption and pollutant emissions data were manipulated into equations that describe consumption and emission rates for the four states of vehicle motion:

1. Acceleration from an initial speed to a final speed;
2. Deceleration from an initial speed to a final speed;
3. Travel over a defined distance at a constant speed; and
4. Engine idling while the vehicle is stopped.

The equations describing the rates for the acceleration and deceleration of vehicles were obtained by applying acceleration and deceleration velocity-time profiles to the fuel consumption and pollutant emissions data. The acceleration and deceleration models were selected on the basis of their ability to predict realistic speed change distances and times over the full range of speeds normally applied to the TRANSYT-7F program. The physical equations of motions for these models are described in the following.

Acceleration Model

The nonuniform acceleration model was used for an FHWA fuel consumption-vehicle emissions-user cost study performed by Zaniewski et al. (3). The model expresses acceleration as a linear function of velocity.

$$A = k_1 + k_2 (V)$$

where

A = acceleration (ft/sec²),
V = velocity (ft/sec),
k_1 = initial acceleration, and
k_2 = change in acceleration due to an increase in velocity.

Using this model, the time to change from speed V_o to V_f is given by:

$$t = \{\ln[k_1 - k_2 (V_f)] - \ln[k_1 - k_2 (V_o)]\}/(-k_2)$$

where t is time to change speed (sec). The distance traveled over the time interval t from initial speed V_o is then given by

$$x = (k_1/k_2)t - (k_1/k_2^2)\{1 - \exp[-k_2(t)]\} + (V_o/k_2)\{1 - \exp[-k_2(t)]\}$$

where x is distance (ft).

In the model, k_1 is the maximum acceleration of the vehicle, and k_1/k_2 is the maximum speed obtainable. The values for k_1 and k_2 were classified by vehicle type. Values are given here for the vehicle types relevant to the fuel consumption and emissions data to be used for updating TRANSYT-7F.

	Automobile Size		
Coefficient	*Small*	*Medium*	*Large*
k_1	7.2	8.6	7.9
k_2	0.060	0.076	0.055

Using vehicle fleet mix and car type data provided by FHWA, the January 1986 U.S. vehicle fleet may be described as having 18.5 percent small cars, 20.0 percent medium cars, and 61.5 percent large cars (1).

Using a weighted average of the coefficients against the vehicle fleet mix, the composite coefficients obtained were 7.91 for k_1 and 0.060 for k_2. The resulting acceleration-velocity relationship is given by

$$A = 7.91 - 0.060(V)$$

The equations of motion in this acceleration model are used to predict distances and times required to reach a range of final speeds.

Deceleration Model

The linear deceleration model is used in the TEXAS Model for Intersection Traffic (4). The model expresses the final deceleration at zero velocity as an empirical linear function of the initial velocity before braking.

$$D_f = (K)(-6 - V_o/44)$$

where

D_f = final deceleration (ft/sec²),
V_o = initial velocity (ft/sec), and
K = empirical constant to account for observed driver behavior.

A K-value of 5.25 is used in the TEXAS Model; this was found to produce reasonable deceleration distances and times over the range of speeds normally applied to TRANSYT-7F. The slope of the deceleration versus time curve is referred to as the deceleration jerk and is given by

$$S = \tfrac{1}{2}(D_f^2)/(V_o)$$

where S is the deceleration jerk (ft/sec³).

Using this model, the time to change from speed V_o to V_f is given by

$$t = [2(V_o - V_f)/S]^{1/2}$$

where t is time (sec) and V_f is final velocity (ft/sec).

The distance traveled over the time interval t from initial speed V_o is then given by

$$x = V_o(t) + S(t^3)/6$$

The equations of motion in this deceleration model are used to predict distances and times required to decelerate from a range of initial speeds.

Data Reduction Procedure

The fuel consumption and emission rate data are expressed as a function of velocity in increments of 1 ft/sec and as a function of acceleration in increments of 1 ft/sec/sec. The acceleration model was applied to the fuel consumption data by computing the amount of time required for a vehicle to accelerate 1 ft/sec. Using the initial speed, the change in speed, and the instantaneous acceleration rate of the vehicle, the fuel consumption rate in effect over that time was obtained from the table. This consumption rate was then multiplied by the speed-change time to obtain the amount of fuel required to accelerate the vehicle 1 ft/sec. This process was repeated for all speed changes from 0 to 88 ft/sec. The cumulative sum of the consumptions from zero to any speed was thus equivalent to the amount of fuel consumed when a vehicle accelerates from zero to that desired speed. Also, by subtracting the cumulative sums from two different speeds, it was possible to obtain an estimate of the amount of fuel consumed when a vehicle accelerates from the lower speed to the higher speed.

To account for the effect of grade, it was necessary to repeat the table integration process previously described for grades between −10 and +10 percent in intervals of 2 percent. The correction for grade is applied by increasing the acceleration rate of the vehicle to compensate for hill climbing, or by decreasing the rate to compensate for gravitational accelerations in effect on negative grades. This equivalent acceleration due to grade was previously described in this paper.

A curve-fitting computer program was then used to fit least-squares curves to each of the fuel consumption curves for each grade. The program tested six different regression models and presented the results, with goodness-of-fit parameters, so the best-fitting curve could be used. The models tested included linear, transformed (exponential, power, logarithmic, and reciprocal), and parabolic (*5*). The fuel consumption for accelerations occurring on grades between these curves are computed by linear interpolation.

The same table integration routine previously described was then applied to each of the three types of pollutant emissions tables to obtain emissions as a function of speed change. Then both the fuel consumption and the pollutant emissions data were applied similarly to the deceleration model. Once again, curves describing consumption and emissions as a function of speed were obtained for the same range of roadway grades.

To describe the fuel consumption and pollutant emissions for vehicles traveling at constant speed, curves were fitted to the consumption and emissions tables for the zero acceleration case. Since the performance of a vehicle traveling at a constant speed on a grade is related to the performance of an engine accelerating or decelerating at a constant speed, additional curves were fitted to the table data for accelerations from −3 to +3 ft/sec/sec. This range of accelerations due to grade simulates the effect on the performance of a vehicle traveling on grades from −10 to +10 percent at a constant speed. Consumption and emissions for idling during stopped delay were taken as the zero velocity–zero acceleration rate from the data tables. Variables for the N tables follow:

G = roadway grade (%),
A_g = acceleration on grade (ft/sec²),
V = travel speed (ft/sec),
V_i = initial speed (ft/sec),
V_f = final speed (ft/sec),
F = fuel consumption rate,
CO = carbon monoxide emissions rate,
HC = hydrocarbon emissions rate,
NO_x = nitrogen oxide emissions rate,
V_{mph} = travel speed (mph),
OIL = oil consumption rate,
TIRE = tire wear rate (% worn),
REP = repair visit (% visit),
DEP = vehicle depreciation rate (% initial cost).

Table 1 lists the group of fuel consumption equations obtained as the result of this data reduction procedure. Tables 2 through 4 list the group of equations obtained for carbon monoxide, hydrocarbons, and nitrogen oxide emissions.

VEHICLE MAINTENANCE AND USER COST DATA

Data for oil consumption, tire wear, vehicle maintenance and repair costs, and depreciation came from the 1982 study completed by Zaniewski et al. at the Texas Research and Development Foundation (*3*). Four passenger cars were tested, and the results were classified for small, medium, and large vehicles according to engine size. Then these tables were combined into one table with a proportional representation of each size of vehicle in the U.S. vehicle fleet.

The physical units for the four vehicle maintenance parameters were quarts of oil for oil consumption; percentage wear

TABLE 1 Fuel Consumption Rate

DELAY CONSUMPTION (Gallons per Vehicle Second)
$F_{delay} = 0.00013$

ACCELERATION SPEED CHANGE CONSUMPTION
(Gallons per Vehicle)
$A = 1.6570 + 0.0073607(G) + 0.0006955(G^2)$
$B = 0.30934 + 0.005019(G) - 0.001271(G^2)$
$F_{accel} = (10^B)(V_f^A - V_i^A)/100000$

DECELERATION SPEED CHANGE CONSUMPTION
(Gallons per Vehicle)
$A = 1.48922 + 0.0048494(G) - 0.000278(G^2)$
$B = -0.022132 - 0.029157(G) + 0.0010894(G^2)$
$F_{decel} = (10^B)(V_i^A - V_f^A)/100000$

CONSTANT SPEED FUEL CONSUMPTION
(Gallons per Vehicle Foot)

For Grades Less Than −1 Percent
$F_{const} = [14.074(e^{(0.0074057(V))})]/(100000(V))$

For Grades Between −1 and 0 Percent
$A = 14.074(e^{(0.0074057(V))})$
$B = 14.234(e^{(0.016779(V))})$
$F_{const} = [A + (B - A)(A_g + 1)]/(100000(V))$

For Grades Between 0 and 1 Percent
$A = 14.234(e^{(0.016779(V))})$
$B = 22.081(e^{(0.020016(V))})$
$F_{const} = [A + (B - A)(A_g)]/(100000(V))$

For Grades Between 1 and 2 Percent
$A = 22.081(e^{(0.020016(V))})$
$B = 23.351 + 1.0005(V) + 0.011864(V^2)$
$F_{const} = [A + (B - A)(A_g - 1)]/(100000(V))$

For Grades Greater Than 2 Percent
$A = 23.351 + 1.0005(V) + 0.011864*V^2$
$B = 23.006 + 1.9495(V) + 0.0093498*(V^2)$
$F_{const} = [A + (B - A)(A_g - 2)]/(100000(V))$

TABLE 2 Carbon Monoxide Emissions Rate

DELAY EMISSIONS (Grams per Vehicle Second)
$CO_{delay} = 0.003$

ACCELERATION SPEED CHANGE EMISSIONS
(Grams per Vehicle)
$A = 2.43738 - 0.005583(G) - 0.000760(G^2)$
$B = -0.49326 + 0.036210(G) + 0.0002949(G^2)$
$CO_{accel} = (10^B)(V_f^A - V_i^A)/1000$

DECELERATION EMISSIONS
(Grams per Vehicle)
$A = 1.54246 + 0.0051115(G) - 0.000288(G^2)$
$B = -0.751286 - 0.029621(G) + 0.0011968(G^2)$
$CO_{decel} = (10^B)(V_i^A - V_f^A)/1000$

CONSTANT SPEED CARBON MONOXIDE EMISSIONS RATE
(Grams per Vehicle Foot)

For Grades Less Than −1 Percent
$CO_{const} = 3.0741(e^{(0.0093192(V))})/(1000(V))$

For Grades Between −1 and 0 Percent
$A = 3.0741(e^{(0.0093192(V))})$
$B = 3.3963(e^{(0.014561(V))})$
$CO_{const} = [A + (B - A)(A_g + 1)]/(1000(V))$

For Grades Between 0 and 1 Percent
$A = 3.3963(e^{(0.014561(V))})$
$B = 4.6927(e^{(0.031454(V))})$
$CO_{const} = [A + (B - A)(A_g)]/(1000(V))$

For Grades Between 1 and 2 Percent
$A = 4.6927(e^{(0.031454(V))})$
$B = 5.5812(e^{(0.047365(V))})$
$CO_{const} = [A + (B - A)(A_g - 1)]/(1000(V))$

For Grades Greater Than 2 Percent
$A = 5.5812(e^{(0.047365(V))})$
$B = 6.5785(e^{(0.064392(V))})$
$CO_{const} = [A + (B - A)(A_g - 2)]/(1000(V))$

TABLE 3 Hydrocarbon Emissions Rate

DELAY EMISSIONS (Grams per Vehicle Second)
$HC_{delay} = 0.0003$
ACCLERATION SPEED CHANGE EMISSIONS
(Grams per Vehicle)
$A = 2.0145 - 0.26268/(G)$
$B = -0.56356 + 0.016475(G) + 0.0002783(G^2)$
$HC_{accel} = (10^B)(V_f^A - V_i^A)/10000$
DECELERATION SPEED CHANGE EMISSIONS
(Grams per Vehicle)
$A = 1.7741 + 0.0041659(G) - 0.000388(G^2)$
$B = -0.98898 - 0.028385(G) + 0.0012138(G^2)$
$HC_{decel} = (10^B)(V_i^A - V_f^A)/10000$
CONSTANT SPEED HYDROCARBON EMISSIONS RATE
(Grams per Vehicle Foot)
For Grades Less Than −1 Percent
$HC_{const} = 2.9262(e^{(0.020118(V))})/(10000(V))$
For Grades Between −1 and 0 Percent
$A = 2.9262(e^{(0.020118(V))})$
$B = 2.7843(e^{(0.015062(V))})$
$HC_{const} = [A + (B - A)(A_g + 1)]/(10000(V))$
For Grades Between 0 and 1 Percent
$A = 2.7843(e^{(0.015062(V))})$
$B = 3.7248(e^{(0.023644(V))})$
$HC_{const} = [A + (B - A)(A_g)]/(10000(V))$
For Grades Between 1 and 2 Percent
$A = 3.7248(e^{(0.023644(V))})$
$B = 4.2789(e^{(0.033437(V))})$
$HC_{const} = [A + (B - A)(A_g - 1)]/(10000(V))$
For Grades Greater Than 2 Percent
$A = 4.2789(e^{(0.033437(V))})$
$B = 5.2305(e^{(0.040708(V))})$
$HC_{const} = [A + (B - A)(A_g - 2)]/(10000(V))$

TABLE 4 Nitrogen Oxide Emissions Rate

DELAY EMISSIONS (Grams per Vehicle Second)
$NOX_{delay} = 0.0003$
ACCELERATION SPEED CHANGE EMISSIONS
(Grams per Vehicle)
$A = 1.6573 - 0.010139(G) + 0.0017341(G^2)$
$B = 0.51557 + 0.040046(G) - 0.002613(G^2)$
$NOX_{accel} = (10^B)(V_f^A - V_i^A)/10000$
DECELERATION SPEED CHANGE EMISSIONS
(Grams per Vehicle)
$A = 1.4293 + 0.0047751(G) - 0.000276(G^2)$
$B = -0.77628 - 0.029074(G) + 0.0010873(G^2)$
$NOX_{decel} = (10^B)(V_i^A - V_f^A)/10000$
CONSTANT SPEED NITROGEN OXIDE EMISSIONS RATE
(Grams per Vehicle Foot)
For Grades Less Than −1 Percent
$NOX_{const} = 1.7325(e^{(0.011815(V))})/(10000(V))$
For Grades Between −1 and 0 Percent
$A = 1.7325(e^{(0.011815(V))})$
$B = 1.5718(e^{(0.040732(V))})$
$NOX_{const} = [A + (B - A)(A_g + 1)]/(10000(V))$
For Grades Between 0 and 1 Percent
$A = 1.5718(e^{(0.040732(V))})$
$B = 4.2279(e^{(0.050231(V))})$
$NOX_{const} = [A + (B - A)(A_g)]/(10000(V))$
For Grades Between 1 and 2 Percent
$A = 4.2279(e^{(0.050231(V))})$
$B = 1.1096(V^{1.2624})$
$NOX_{const} = [A + (B - A)(A_g - 1)]/(10000(V))$
For Grades Greater Than 2 Percent
$A = 1.1096(V^{1.2624})$
$B = 3.0515(V^{1.1111})$
$NOX_{const} = [A + (B - A)(A_g - 2)]/(10000(V))$

of usable tread for tire wear; percentage of average maintenance or repair job for maintenance and repair costs; and percentage depreciation of initial new-car value for vehicle depreciation. The vehicle maintenance data were presented for constant speed vehicle motion and for speed changes. The constant speed data were presented as the amount of maintenance incurred when a vehicle travels 1,000 mi at a constant speed. The data covered the range of speeds from 0 to 60 mph. The speed change data were presented as the amount of maintenance incurred when a vehicle decelerates from a speed in the range of 0 to 60 mph to a full stop, then accelerates back to the initial speed. The data tables for both constant speed and speed change vehicle maintenance forms of motion were also presented as a function of roadway vertical grade.

Least-squares curves were fitted to the four vehicle maintenance parameters as a function of speed for a range of roadway grades from −10 to +10 percent. Then curves were fitted to the parameters of these curves as a function of roadway grade. Tables 5 through 8 present the equations obtained from the curve-fitting procedure.

TABLE 5 Oil Consumption

CONSTANT SPEED OIL CONSUMPTION
(Quarts per Vehicle Mile)
For Grades Greater Than 2 Percent
$A = -0.60125 - 0.26406(LOG_{10}(G))$
$B = 1.1060 + 1.0459(LOG_{10}(G))$
$OIL_{const} = [10^B(V_{mph}{}^A)]/1000$
For Grades Between 0 and 2 Percent
$A = 13.83557 + 14.86654(G)$
$B = 0.99220 + 0.136932(G)$
$OIL_{const} = [B + A/(V_{mph})]/1000$
For Grades Between −2 and 0 Percent
$OIL_{const} = [0.99272 + 13.836/(V_{mph})]/1000$
For Grades Less Than −2 Percent
$A = -27.693 - 16.637(G)$
$B = -0.40799 - 3.4550/(G)$
$OIL_{const} = [B + A/(V_{mph})]/1000$
SPEED CHANGE OIL CONSUMPTION
(Quarts per Speed Change Cycle)
$OIL_{a/d} = [0.0017727 + 0.0005531 * V_{mph}]/1000$

NEW DATA INPUT REQUIREMENTS

Implementing the new fuel, pollution, and user cost models will require new data that are not currently provided for in the existing TRANSYT-7F data input cards. Most of the new data are organized onto two new data input cards that fit into the existing series of cards required to run TRANSYT-7F. The roadway grade parameter is added to one of the existing TRANSYT-7F card types.

Roadway Grade Data

The grade of each roadway link in TRANSYT-7F may be designated using one of the unused fields from the link ex-

TABLE 6 Tire Wear Rate

CONSTANT SPEED TIRE WEAR
(Percent Tread Worn per Vehicle Mile)

For Grades Greater Than −1 Percent

$$A = 10^{(-1.1894)}(G + 2)^{(-0.73580)}$$
$$B = -4.3958 + 5.0811(\mathrm{LOG}_{10}(G + 2))$$
$$\mathrm{TIRE}_{\mathrm{const}} = [e^{(B)}e^{(A(Vmph))}]/100000$$

For Grades Between −4 and −1 Percent

$$A = e^{(-2.883183)}e^{(-0.045573(V\mathrm{mph}))}$$
$$B = -0.23931 - 0.12724(\mathrm{LOG}_{10}(V_{\mathrm{mph}}))$$
$$\mathrm{TIRE}_{\mathrm{const}} = [A + (B - A)(G + 2)]/100000$$

For Grades Less Than −4 Percent

$$A = 0.0061936 + 0.0027882(G)$$
$$B = -0.85561 - 0.27553(G)$$
$$\mathrm{TIRE}_{\mathrm{const}} = [B + A(V_{\mathrm{mph}})]/100000$$

SPEED CHANGE TIRE WEAR
(Percent Tread Worn per Speed Change Cycle)

$$\mathrm{TIRE}_{\mathrm{a/d}} = 10^{(-2.78104)}(V\mathrm{mph}^{(1.84284)})/100000$$

TABLE 7 Maintenance and Repair Rate

CONSTANT SPEED MAINTENANCE AND REPAIR
(Percent of Visit per Vehicle Mile)

For Grades Greater Than −2 Percent

$$A = e^{(-4.56393)}(e^{(0.040463(G))})$$
$$B = e^{(1.32637)}(e^{(0.0019682(G))})$$
$$\mathrm{REP}_{\mathrm{const}} = e^{(B)}(e^{(A(V\mathrm{mph}))})/100000$$

For Grades Between −4 and −2 Percent

$$A = -0.53013 + 0.13827(G)$$
$$B = -3.329245 - 25.10303(G)$$
$$\mathrm{REP}_{\mathrm{const}} = [B + A(V_{\mathrm{mph}})]/100000$$

For Grades Less than −4 Percent

$$A = -1.108532 + 0.0007879(G)$$
$$B = 7.9614 - 22.43846(G)$$
$$\mathrm{REP}_{\mathrm{const}} = [B + A(V_{\mathrm{mph}})]/100000$$

SPEED CHANGE MAINTENANCE AND REPAIR
(Percent of Visit per Speed Change Cycle)

$$\mathrm{REP}_{\mathrm{a/d}} = 10^{(-2.88889)}(V_{\mathrm{mph}}{}^{(2.01258)})/100000$$

TABLE 8 Vehicle Depreciation Rate

CONSTANT SPEED DEPRECIATION
(Percent Depreciation per Vehicle Mile)

$$\mathrm{DEP}_{\mathrm{const}} = [1.4704 - 0.54925(\mathrm{LOG}_{10}(V_{\mathrm{mph}}))]/100000$$

SPEED CHANGE DEPRECIATION
(Percent Depreciation per Speed Change Cycle)

$$\mathrm{DEP}_{\mathrm{a/d}} = 0.0002462(V_{\mathrm{mph}})/100000$$

tension card (Card Type 29). Fields 5 and 6 on Card Type 29 are not currently used for any purpose. The roadway grade is used to correct fuel consumption, pollution emissions, and user cost estimates for the effect of grade.

User Cost Data Input Card

This new card provides the user cost model with up-to-date unit cost information: consumer price index, gasoline price, oil price, tire price, new vehicle price, cost of an average shop maintenance and repair job, and value of travel time.

The unit cost data are applied to the user cost models so that an estimate of user cost associated with a given signalization scheme may be determined as an MOE.

Environment Data Input Card

The second new data input card is the environment data input card. This card would be required to provide correction factors for fuel consumption and pollutant emissions as well as the ambient temperature of the study area. The fuel and pollution correction factors may be used to correct fuel and pollution estimates to better represent current and local conditions. The ambient temperature is used to adjust fuel consumption and pollutant emissions for the effect of temperature.

COMPARISON OF OLD AND NEW FUEL CONSUMPTION MODELS

The fuel consumption estimates produced by the existing TRANSYT-7F fuel model were compared with those of the proposed fuel model. Table 9 shows a parametric comparison between fuel rates predicted by the two models as a function of speed. As expected, the idling fuel consumption estimate is lower for the new model. This occurred because the new fuel model was calibrated using a vehicle fleet mix with better fuel efficiency than those used to calibrate the old model.

A comparison of the constant speed fuel consumption between the two models indicates that the current TRANSYT-7F fuel model predicts fuel consumption values lower than the new model in ranges below 20 mph. The predicted fuel consumptions are nearly equal in the range of 20 to 40 mph;

TABLE 9 Comparison of Fuel Consumption Estimates

Idling Stopped Delay Fuel Consumption (Gallons per Vehicle-Hour)		
	Existing Fuel Model	Proposed Fuel Model
	0.732	0.468
Constant Speed Travel Fuel Consumption (Gallons per Mile)		
Speed (MPH)	Existing Fuel Model	Proposed Fuel Model
10	0.055	0.064
20	0.042	0.042
30	0.035	0.036
40	0.034	0.034
50	0.040	0.035
60	0.052	0.037
Extra Fuel Consumption for Stop-and-Go Speed Change Cycle (Gallons per Cycle)		
Speed (MPH)	Existing Fuel Model	Proposed Fuel Model
10	0.0014	0.0021
20	0.0052	0.0055
30	0.0120	0.0110
40	0.0210	0.0182
50	0.0330	0.0248
60	0.0480	0.0312

however, fuel efficiency for the new model is much better at speeds above 40 mph. This is consistent with trends toward fuel economy in this speed range through better vehicle aerodynamics. Figure 1 graphically displays the variation of fuel comparison with speed for both models.

A comparison of the speed change fuel consumption between the two models indicates that the newer vehicle fleet used to calibrate the new fuel model is more efficient than the older fleet only at speeds in excess of 35 mph. This improvement is also consistent with the trend toward better fuel economy at higher speeds. Figure 2 graphically displays the variation of fuel consumption with speed for both models.

IMPLEMENTATION OF NEW MODELS

This section describes considerations pertaining to the implementation of the new fuel, pollution, and user cost models as

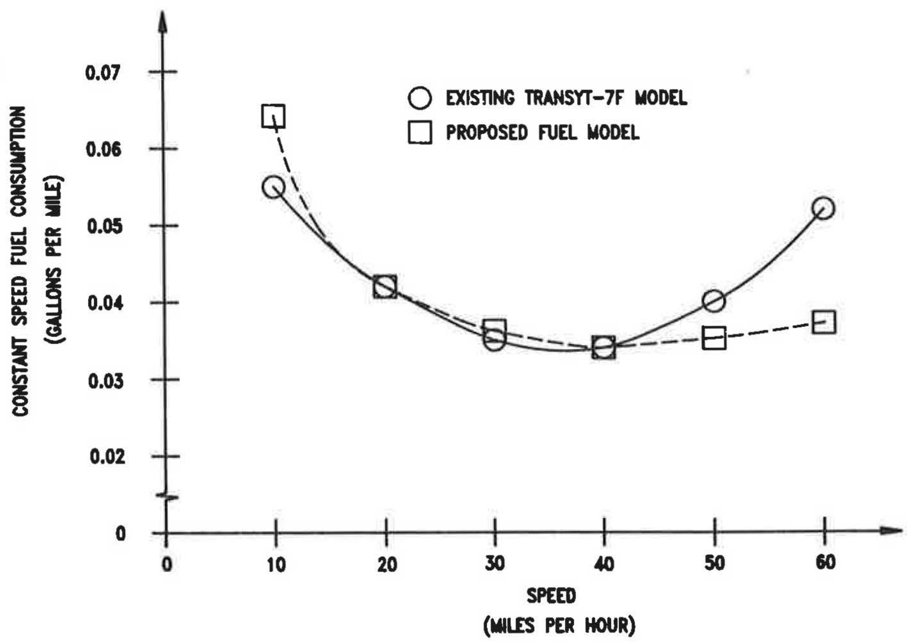

FIGURE 1 Comparison of constant speed fuel consumption.

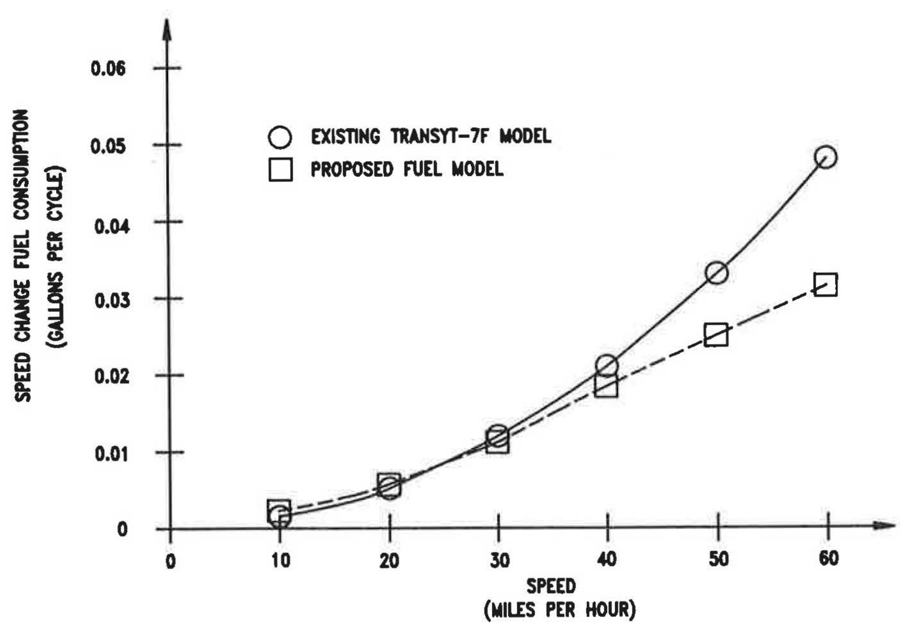

FIGURE 2 Comparison of speed change fuel consumption.

part of the TRANSYT-7F program. The new fuel, pollution, and user cost models may be easily implemented as part of the TRANSYT-7F program by coding the equations from Tables 1 through 8 into subroutines. This method of implementation would require that the following modifications be made:

1. Modify the TRANSYT-7F input utility to accept the two new data input cards and roadways grade. The input routine could be set up to generate automatically a set of default values for cost coefficients; all roadway grades could be set to a zero default unless otherwise coded by the user.
2. Modify the TRANSYT-7F program input routine to accept the two new data input cards and the roadway grade parameter.
3. Add the new fuel, pollution, and user cost equations as functions or subroutines to the TRANSYT-7F source code. Delete the existing fuel and user cost subroutines and functions.

CONCLUSIONS

New technology is leading to new patterns in fuel consumption and pollutant emissions. There is growing concern about energy conservation, urban air pollution, and economic justification of government public works improvements. Thus, the need for an updated fuel consumption model, a pollutant emissions model, and a user cost estimation model in TRANSYT-7F can be easily understood. With these models available, traffic engineers using TRANSYT-7F will have a means by which consumption, emissions, and user cost MOEs may be incorporated into signal systems timing design.

REFERENCES

1. H. C. Lorick. *Analysis and Development of Fuel and Platoon Models*. Transportation Research Center, University of Florida, Gainesville, 1981.
2. R. McGill. *Fuel Consumption and Emission Values for Traffic Models*. FHWA, U.S. Department of Transportation, May 1985.
3. J. P. Zaniewski, et al. *Vehicle Operating Costs, Fuel Consumption, and Pavement Type and Condition Factors*. FHWA, U.S. Department of Transportation, June 1982.
4. C. E. Lee, T. W. Rioux, and C. R. Copeland. *The TEXAS Model for Intersection Traffic—Development*. Research Report 184-1. Center for Highway Research, University of Texas, Austin, Dec. 1977.
5. J. L. Devore. *Probability and Statistics for Engineering and the Sciences*. Brooks/Cole Publishing Co., Monterey, Calif., 1982.

Publication of this paper sponsored by Committee on Traffic Signal Systems.

Compromise Approach To Optimize Traffic Signal Coordination Problems During Unsaturated Conditions

Chang-Jen Lan, Carroll J. Messer, Nadeem A. Chaudhary, and Edmond Chin-Ping Chang

A methodology to optimize the traffic signal coordination problem on an arterial network simultaneously considering delay minimization and progression bandwidth maximization criteria is presented. This approach generates a compromise solution to these two conflicting criteria and, in some cases, produces timing solutions with less delay and less bandwidth than the conventional MAXBAND solutions, although sometimes the outcome is reversed. In general, there is usually a trade-off between delay and bandwidth under the well-timed traffic signal system.

Two conventional approaches are available for coordinating traffic signals in an urban network: delay minimization and progression bandwidth maximization. Timing plans based on the bandwidth approach are preferred by drivers because of the ease in which progression bandwidth can be visualized. A traffic engineer also prefers the bandwidth-based approach because it provides dependable solutions. However, delay-based timing plans may produce better overall system performance and may be preferable from the perspective of traffic systems management. Therefore, traffic engineers may desire to optimize both signal timing objectives based on the conflicting nature of the delay and progression bandwidth criteria (*1*). Therefore, usually a trade-off is necessary.

Cohen and Liu (*2*) proposed a bandwidth-constrained delay minimization methodology that uses TRANSYT-7F to fine-tune offset and green split while preserving the progression bands generated from the MAXBAND program (*3*). This approach may improve system performance as compared with the centered bandwidth timing plans. However, in some situations preserving the progression bands only produces local optimal solutions from a systemwide viewpoint, since there is a trade-off between bandwidth and delay (see Figure 1). This diagram is conceptually constructed on the basis of an investigation of simulation results of real timing data collected from several study sites. The solution with least delay and best bandwidth is an ideal optimal solution used as a benchmark for evaluating the system performance of signal timing plans. Within the timing solution space (shaded area), the solutions along the frontier line are of major interest. Solutions S_1 through S_3 are bandwidth target solutions that carry the maximal bandwidth and relatively high delay. Solutions S_8 through S_{10} are solutions with lower bandwidth but the least delay. These timing solutions can be generated by existing technologies. In the bottom right corner, the general fashion of frontier line bypasses the ideal point. Solutions S_4 through S_7 follow a trend in which delay is reduced as progression decreases, depending on how the decision maker trades off O_1 against O_2. The evaluation of these timing solutions also depends on the preference of the decision maker. Therefore, it is suggested that reallocating the green time resource on the basis of the trade-off between lost bandwidth and delay savings would be beneficial to further reducing system delay and upgrading system performance.

A formal method is proposed to optimize arterial signal timing using both delay minimization and throughput maximization criteria. The enhanced arterial model called COMBAND follows the basic MAXBAND (*3*) formulation and concepts described in the original MITROP model (*4,5*). The results show that a trade-off between delay and progression objectives can render better signal timing plans while maintaining the sound features of both.

MODEL DESCRIPTION

The COMBAND model improves bandwidth-based timing plans by combining both bandwidth and delay/stops considerations. In the MAXBAND model, only the bandwidth decision variable is set up in the objective function, which implies that bandwidth maximization is the only criterion being considered. In most situations, such a single-criterion formulation may generate ineffective solutions because the sidestreet delay increases considerably when its traffic flow

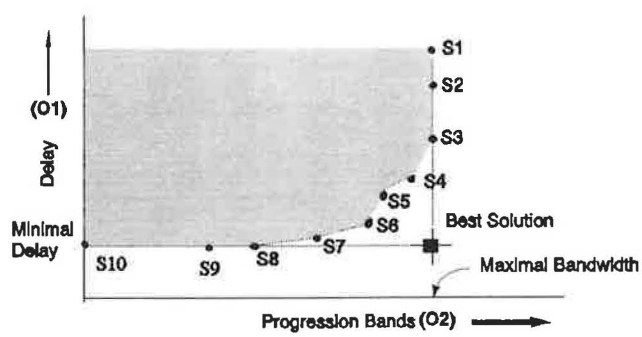

FIGURE 1 Signal timing solutions based on two criteria.

Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843.

increases. Some modifications were made to extend the capability of bandwidth-based methodology dealing with delay minimization, such as the directional weighting of the inbound and outbound bandwidth and multiweight bandwidth for each directional road link of the arterial (*6,7*); however, in terms of both timing objectives a global optimal solution still cannot be achieved.

The traffic signal optimization problem contains many variables that in turn affect the system performance, individually or collectively. These variables include delay, stops, progression bands, capacity (or throughput), fuel consumption, emissions, and journey time. Delay, stops, progression bands, and capacity have been analyzed often in previous research in that these variables are not strongly interrelated but sensitively responsive to traffic systems in most situations. Fuel consumption and emissions are normally modeled as secondary functions of delay and stops, hence, they are highly correlated with delay and stops as dependent variables to system performance. They need not be considered independently. As for journey time, it could be affected by several combinations of variables, rendering it difficult to be modeled. In this paper, delay/stops and throughput are chosen as the criteria to evaluate signal timing plans.

Essentially, delay/stops and progression bands become two conflicting criteria in well-timed systems. The defined problem is formulated as a multicriteria decision making (MCDM) problem and solved by linear multiobjective programming techniques. Another approach is to formulate the problem as linear programming (LP) formulation, combine the operational criteria in the LP objective function, and solve the problem using LINDO, MPSX/370E, or other LP optimization packages. Both approaches have to convert the nonlinear objective function into a linear form before applying the LP technique to solve the problem since there is no tool available to solve the mixed-integer nonlinear programming problem directly. Before proceeding further, several assumptions should be made. The following assumptions are also applied in the underlying model:

1. Prevailing traffic conditions are not saturated;
2. Traffic arrival flow rate and service rate remain constant. The time-stationary flow assumptions will be relaxed in the future work, with provisions of overflow queue or temporary oversaturation condition;
3. No platoon dispersion occurs on the coordinated arterial; and
4. No midblock flow occurs.

The COMBAND model uses notation similar to that of the MAXBAND model to promote ease of reference to existing technology. The complete model formulation and notation are provided in the Appendix. Figure 2 shows the notation in a time-space diagram.

Objective Function

Delay/Stops Criteria

From the preceding assumptions, we postulate that the queue accumulated during red dissipates during green and no overflow queue occurs. Thus, the total delay can be written without the overflow term and is defined as follows:

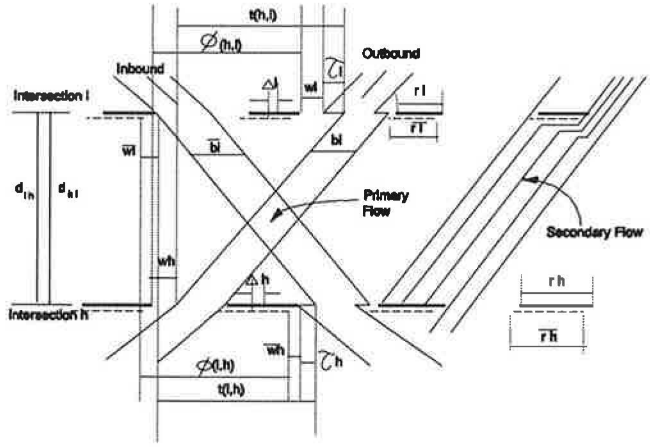

FIGURE 2 Time-space diagram for MAXBAND and COMBAND models.

$$TD = \frac{qR^2}{2C(1 - q/s)} = \frac{qCr^2}{2(1 - q/s)} \text{ (veh·sec/sec)}$$

where

q = arrival rate (vps),
R = effective red time (sec),
s = saturation flow rate (vps),
C = cycle time (sec), and
$r(g)$ = red (green) time in fraction of cycle.

Without considering overflow delay, the total delay is observed to be a quadratic function of red time as shown in Figure 3. The delay curve versus red time to be linearized is bounded by r_{min} and r_{max} and any amount of red time can be allocated within this range depending on the "steepness" of the delay curve. The first task is to determine how many components should be used to fit the curve accurately. Yagar indicated that three piecewise-linear components can fit the delay curve well enough to produce quite accurate results (*8*). Gartner et al. also split the curve into three piecewise linear

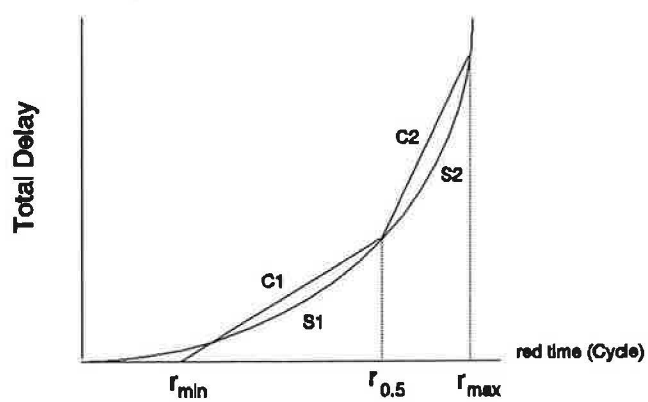

FIGURE 3 Total delay versus red time.

components. However, we have elected to apply only two piecewise linear components for the following reasons:

1. Under low to moderate flow conditions, the q/c value seldom exceeds 0.7. The delay curve is not too steep when (q/c) is below this level.
2. From the first reason, the two-component linearized approximation generates quite similar results to three-component linearized approximation.
3. The decision variables in the LP problem are reduced to enhance the computational efficiency, especially when dealing with a large-scale network problem.

Another question might be raised as to where the cutting line between two linear pieces lies. For each link, we have

$$x = \frac{q}{c} = \frac{q}{s \cdot g} = \frac{q}{s(1 - r)}$$

$$r = 1 - \frac{q}{sx}$$

Here $x = 0.5$ is selected as the cutting line after accomplishing several experiments; the associated r-value ($r_{0.5}$) is also calculated. Thus, the slope of each linear component can be determined as follows, respectively:

For C_1

$$\frac{Ar_{0.5}^2 - Ar_{\min}^2}{r_{0.5} - r_{\min}} = A(r_{0.5} + r_{\min})$$

For C_2

$$A(r_{0.5} + r_{\max})$$

where A equals $q/[2(1 - q/s)]$. On the other hand, it may be desirable to take into account stops along with delay in the disutility function. The number of vehicle stops per cycle is

$$\frac{qR}{(1 - q/s)} = \frac{qrC}{(1 - q/s)}$$

Let B be $q/(1 - q/s)$, multiplied by a factor $k(k = 0 \text{ to } 1)$; added with $(1 - k)\cdot C_1$ and $(1 - k)\cdot C_2$, respectively, the disutility function becomes a weighted combination of delay and stops with the following coefficients:

$$D1 = (1 - k)\cdot C1 + k\cdot B$$

$$D2 = (1 - k)\cdot C2 + k\cdot B$$

The factor k is specified by the user. If k is set to zero, the disutility objective function becomes minimization of system-wide delay without considering stops. The authors suggest that k be chosen in the range from 0.3 to 0.7. So far, for intersection i, we have two decision variables, $r1_{mi}$ and $r2_{mi}$, with coefficients $D1_i$ and $D2_i$ respectively in the objective function, that is (notation is only for outbound direction):

Minimize

$$\textstyle\sum_i (D1_i r1_{mi} + D2_i r2_{mi})$$

where $r1_m$ and $r2_m$ are subjected to the constraints $r_{\min} \le r1_m \le r_{0.5}$ and $0 \le r2_m \le (r_{\max} - r_{0.5})$, respectively. Moreover, we adjust the arterial delay function value by the progression factor (PF) as used in the *Highway Capacity Manual* (*9*) to account for the effect of platoon progression on delay, assuming there exists a moderately favorable platoon condition under COMBAND-generated timing plans.

Throughput Criteria

In the COMBAND model, an attempt is made to maximize the arterial throughput and minimize the system total delay/stops to increase the system throughput. For intersection i, the arterial throughput is essentially equivalent to the sum of the proportions of vehicles (including arterial through traffic q_1 and sidestreet turning traffic q_2) from upstream intersection passing through the downstream intersection inside and outside the bandwidth during green; for example:

$$q_1 b_i + q_2(g_i - b_i)$$

$$q_1 b_i + q_2(1 - r1_{mi} - r2_{mi} - b_i)$$

$$(q_1 - q_2)b_i - q_2(r1_{mi} + r2_{mi}) + q_2$$

where q_2 can be removed from objective function. Finally, combining delay/stops and throughput objectives, the objective function becomes (notation is only for outbound direction)

Maximize

$$\textstyle\sum_i [(q_1 - q_2)b_i - q_2(r1_m + r2_m) - (D1_i r1_{mi} + D2_i r2_{mi})]$$

Constraints

In addition to the basic MAXBAND formulation, a number of the following constraints are added to enhance the capability of the underlying model. First, the technique of releasing green splits as variables is used to make the optimization of green split possible (*10*). A set of constraints and binary integer variables are included. As shown in Figure 3, total delay increases considerably as traffic flow approaches capacity (i.e., the volume-capacity ratio (q/c) is approaching 1.0). To avoid the escalating delay produced by an overflow queue and maintain a minimum level of service, we restrict volume/capacity ratio below 0.95 and include associated red time upper-bound constraints as used in the MITROP program; for example:

$$r \le r_{\max} = 1 - \frac{q}{0.95s}$$

On the other hand, it is desired to set the minimum effective green time on each approach according to the following considerations:

1. The feasible minimum amount of effective green time pedestrians need to traverse the side streets safely and the time would satisfy the driver's expectancy.
2. Nominal green splits calculated by Webster's method under the control of the local minimum-delay cycle.

An important issue to be considered pertains to the allocation of the slack green time, defined as the excess time of the system optimal cycle beyond the local minimal-delay cycle. Let z and z_l stand for the reciprocals of the system optimal cycle and the local minimal-delay cycle, and zr represents the ratio of these two variables. For each intersection

$$zr = \min\{1, z/z_l\}$$

This equation is equivalent to the following constraints

$$zr \leq 1$$

$$zr \leq z/z_l$$

$$1 - zr \leq M \cdot \delta$$

$$z/z_l - zr \leq M \cdot (1 - \delta)$$

where M is a big number and δ is the binary integer variable. Because z is in unit of cycle time, $M = 1$ is big enough. Therefore, after some rearrangement, the above constraints become

$$zr \leq 1$$

$$zr \leq z/z_l$$

$$zr + \delta \geq 1$$

$$zr - z/z_l - \delta \geq -1$$

As with the intersections carrying the slack time, the green splits of main and side streets will be lowered in proportion to zr, and the slack green time is further reallocated by the underlying model instead of Webster principle.

To make the model closer to real situations, it is assumed that the journey time increases with increasing traffic flow. The BPR travel time prediction function sometimes used in transportation planning models, $t_i = t_0[1.0 + 0.15(q/c)^4]$, is used to characterize the link volume-delay relationship, where t_i is the predicted travel time on link i given a specific traffic flow, t_0 is the free-flow travel time, and c is link capacity. Here in BPR function, the term (q/c) is determined by $2Y/(1 + Y)$, (Y being the sum of flow ratios of the critical movements) under the assumption that the degree of saturation is the same for all critical phases of the intersection for optimum division of the cycle (*11*).

As with the range of system cycle concerned, the upper and lower bounds are selected according to the following considerations. The concept is illustrated in Figure 4.

1. Define the intersection with largest local minimal-delay cycle C_0, as the critical intersection. The calculation of minimal-delay cycle is based on Webster's method, for example, $C_0 = (5 + 1.5L)/(1 - Y)$, where L is total lost time per cycle (*11*).

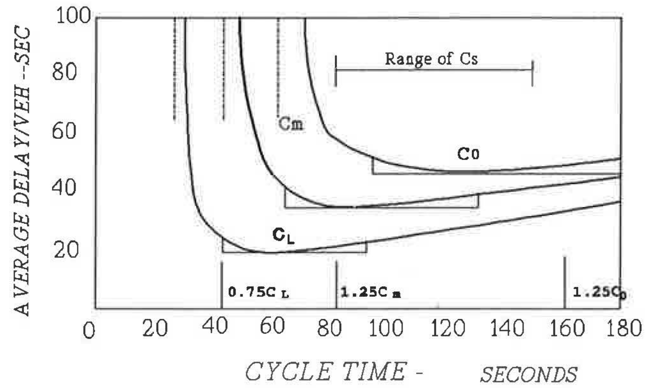

FIGURE 4 Effect on delay of variation of cycle length.

2. According to Webster's method, the minimum cycle is just long enough to allow all the traffic arriving during a cycle to optimally clear the intersection. For deterministic flows, the cycle is given by $C_m = L/(1 - Y)$ (the vertical asymptote to the delay-cycle curve), at which the degree of saturation is close to 1. To ensure the critical intersection is operated under capacity where the level of flow varies appreciably, the system was set at minimum cycle, C_s, at least equal to $1.25L/(1 - Y)$. The lower bound of system cycle is also confined by $0.75*$ (smallest local minimal-delay cycle, C_L).
3. As indicated (*11*), the delay for cycle within the range 0.75 to 1.5 of the optimal value is never 10 to 20 percent more than minimum delay. We further restrict the upper limit as $1.25C_0$ to avoid too much waste of green time.
4. From a practical standpoint, the cycle should be within the range of 40 to 150 sec.

In summary, we suggest the range of system cycle as follows:

$$\max\{40, 0.75C_L, 1.25C_m\} \leq C_s \leq \min\{1.25C_0, 150\}$$

In the basic MAXBAND model, queue clearance time must be supplied by users. However, under the assumption of uniform arrivals, queue length and queue clearance time can be estimated approximately. Assuming that the primary flow can fully utilize the bandwidth without being stopped under a favorable platoon condition, the queue would be produced mainly by the secondary flow consisting of the turning movements of upstream intersection from side streets during arterial red time and through movement from upstream intersection during slack time as shown in Figure 5. Let q_L, q_R, and q_T be the sidestreet left-turn, right-turn, and arterial through movement traffic coming from upstream intersection h, then the number of vehicles traveling on the link between intersection h and i during a cycle are produced by

- Left-turn movement from inbound side street:

$$q_L \cdot \bar{l}_{ch}$$

- Right-turn movement from outbound side street:

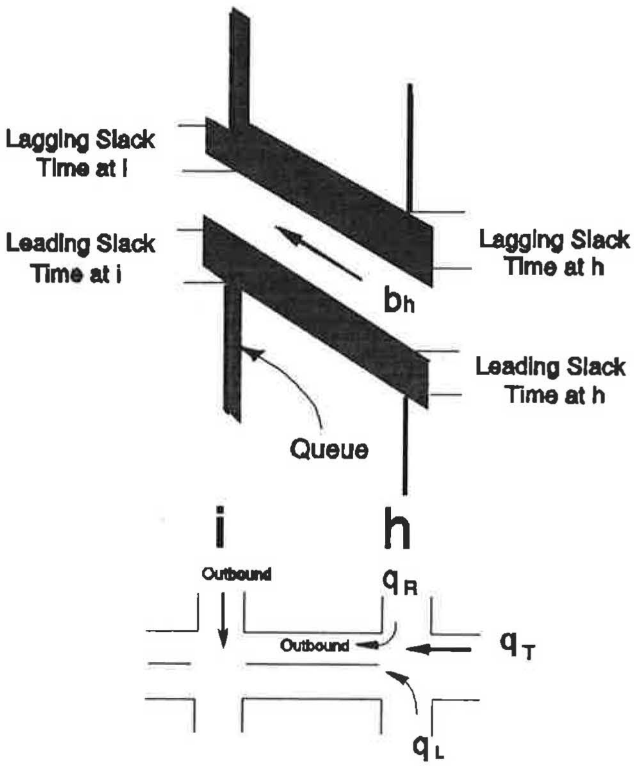

FIGURE 5 Queue produced by secondary flow.

Assuming that every intersection is on the right turn on red (RTOR) operation for right-turn movement, outbound sidestreet right-turn movement is allowed only during outbound sidestreet through phase and inbound arterial left-turn phase. Therefore, the right-turn traffic is equal to

$$q_R \cdot (1 - r1_{ch} - r2_{ch} + \bar{l}_{ch})$$

- Through movement on arterial during leading slack time (w_h) to the progression band:

$$q_T \cdot w_h$$

- Through movement on arterial during lagging slack time $(1 - r1_{mh} - r2_{mh} - w_h - b_h)$ of previous cycle:

$$\max[q_T \cdot (r1_{mi} + r2_{mi} - r1_{mh} - r2_{mh} + w_i - w_h), 0]$$

The flow generated by these four categories is denoted as Q_i and the average discharge headway on the link i as h_i. The required queue clearance time is equal to the start-up loss time plus $h_i \cdot Q_i$, where the start-up time is assumed to be 2 sec. If the available slack time w_i is greater than the required queue clearance time, then there will be no queueing. Thus the queue clearance time is

$$\tau_i = \max[que, 0]$$

where $que = 2.0 \cdot z + h_i \cdot Q_i - w_i$. The equation is equivalent to the following constraints:

$$\tau_i \geq 0$$

$$\tau_i - que \geq 0$$

$$\tau_i - \delta_i \leq 0$$

$$\tau_i - que + \delta_i \leq 1$$

Here, constraints $\tau_i \geq 0$ are implicitly processed in normal LP methodology and need not be specified. The queue forming during lagging slack time can be solved in a similar fashion.

MODEL TESTING AND RESULTS

The LINDO optimization package was used to calculate signal timing plans for the MAXBAND, COMBAND models, and bandwidth-target solution (*12*). Also included were TRANSYT optimization solutions with and without bandwidth constraints and PASSERII-90 solutions into analysis. To measure the effectiveness of performance on the common basis, the TRAFNETSIM network simulation package was employed to evaluate these timing solutions. The exogenous data including traffic volumes, lane configuration, and such were collected from three arterial networks: Skillman Avenue (with four intersections), 12th Street (a.m. peak), and 12th Street (p.m. peak; with seven intersections).

For each timing solution, at least 5 simulation runs with different random number seeds were performed, and each run took 15-min. We choose 15-min periods because the traffic arrival pattern starts to become unstable in a longer time frame so that the initial time solution loses the capability to accommodate the forthcoming traffic conditions. Besides, the TRAF-NETSIM simulation model reflects some degree of variation in the simulation results. Several replications are needed to reduce the variation and to suggest reliability of mean value. Here, the required sample size is five replications, providing a limit on a 95 percent probability that the sample mean will be within a range of acceptable error.

To make consistent comparisons among different methods, cycle length was held constant for each case. The other three timing variables (offsets, green splits, and phasing sequences) were optimized by the MAXBAND and COMBAND models. Moreover, the nonuniform bandwidth concept is used in the COMBAND model since the link-specific bandwidth is weighted by traffic volume (7). The total delay function described previously implies that the red split variables are also weighted by traffic volume. Therefore, the decision variables, bandwidth and red splits, are weighted with respect to their contributions to the overall objective function on the same scale.

The results documented in Tables 1, 2, and 3 are briefly described in the next paragraphs. Figures 6, 7, and 8 show the relationship between delay and bandwidth for each case. These figures also confirm the concept depicted in Figure 1.

1. In some cases, the COMBAND model produces timing solutions with less delay and less bandwidth than the conventional MAXBAND solutions and sometimes the outcome is reversed. However, it is shown, that there is a trend of

TABLE 1 Solutions Based on Bandwidth

Case Name	Model Name	Efficiency (%)	Through-put(vph)	Attain-ability(%)	Selected Cycle Length
SKILLMAN AVENUE	TRANSYT[1]	0.1580	18820	36.6	95 Seconds
	BC-TRANSYT[2]	0.4291	18824	100	
	MAXBAND	0.4291	18804	100	
	COMBAND	0.4340	18944	100	
	PASSER II-90	0.3947	18252	100	
12TH STREET (AM PEAK)	TRANSYT	0.1500	17996	65.9	90 Seconds
	BC-TRANSYT	0.2450	17500	100	
	MAXBAND	0.2450	16440	100	
	COMBAND	0.2650	15560	100	
	PASSER II-90	0.1889	15596	76.0	
12TH STREET (PM PEAK)	TRANSYT	0.1164	21020	38.6	116 Seconds
	BC-TRANSYT	0.2677	21224	88.6	
	MAXBAND	0.2677	21020	88.6	
	COMBAND	0.2578	20539	95.7	
	PASSER II-90	0.2328	20988	80.0	

Note: 1. The bandwidth values in the calculation of efficiency and attainability are directly read from time-space diagram produced by TRANSYT.
2. Bandwidth-Constrained TRANSYT solutions which take MAXBAND solutions as starting solutions and perform the TRANSYT optimization.

TABLE 2 Solutions Based on Throughput

Case Name	Model Name	Throughput (vph) Total	Main Street	Side Street	Selected Cycle Length
SKILLMAN AVENUE	TRANSYT	18820	9524	9296	95 Seconds
	BC-TRANSYT	18824	9696	9128	
	MAXBAND	18804	9688	9116	
	COMBAND	18944	9760	9184	
	PASSER II-90	18252	8630	9392	
12TH STREET (AM PEAK)	TRANSYT	17996	6140	11892	90 Seconds
	BC-TRANSYT	17500	6024	11476	
	MAXBAND	16440	5792	10648	
	COMBAND	15560	4812	10748	
	PASSER II-90	15596	5860	11511	
12TH STREET (PM PEAK)	TRANSYT	21020	5740	15280	116 Seconds
	BC-TRANSYT	21224	5764	15460	
	MAXBAND	21020	5940	14074	
	COMBAND	20539	5897	14642	
	PASSER II-90	20988	5756	15232	

delay being reduced as progression bandwidth decreases. This indicates that a trade-off between delay and bandwidth is made usually under well-timed traffic signal systems.

2. The delay on side streets will decrease significantly if some reallocation of green time from arterial to side streets is made, based on the trade-off between the loss of through green bands and the gain of delay saving, especially when the traffic demand on the side streets increases.

3. The optimization feature in TRANSYT usually produces sound system performance in terms of delay even though the bandwidth layout is wiggly or solutions do not have obvious bandwidth. However, the system performance of TRANSYT-

TABLE 3 Disutility Functions

Case Name	Model Name	Total Delay (veh-hr /hr)	Travel Time (min/ mile)	Average Delay (min/ veh)	No. of Stops per trip	Average Speed (mph)	Fuel consum-ption (mpg)	Emiss-ions (kg/ mile-hr)
SKILLMAN AVENUE	TRANSYT	362.38	4.36	3.33	1.42	13.78	10.74	3.613
	BC-TRANSYT	360.46	4.29	3.29	1.30	13.98	10.90	3.549
	MAXBAND	357.71	3.27	4.25	1.34	14.12	10.94	3.561
	COMBAND	363.86	4.31	3.32	1.22	13.92	10.98	3.507
	PASSER II-90	358.54	3.29	4.43	1.28	13.62	10.70	3.464
12TH STREET (AM PEAK)	TRANSYT	298.53	5.31	2.70	1.28	11.30	9.72	1.983
	BC-TRANSYT	313.44	5.88	2.92	1.20	10.28	9.34	1.923
	MAXBAND	367.04	7.05	3.56	1.20	8.54	8.68	1.953
	COMBAND	402.46	8.60	3.90	1.20	6.98	8.22	1.932
	PASSER II-90	355.39	7.05	3.60	1.26	8.54	8.72	1.884
12TH STREET (PM PEAK)	TRANSYT	289.34	4.98	2.15	1.10	12.06	9.38	2.451
	BC-TRANSYT	293.31	5.01	2.16	1.10	11.98	9.36	2.472
	MAXBAND	328.16	5.68	2.44	1.10	10.58	8.94	2.524
	COMBAND	322.33	5.60	2.44	1.10	10.70	9.00	2.523
	PASSER II-90	279.67	4.73	2.06	1.00	12.72	9.72	2.390

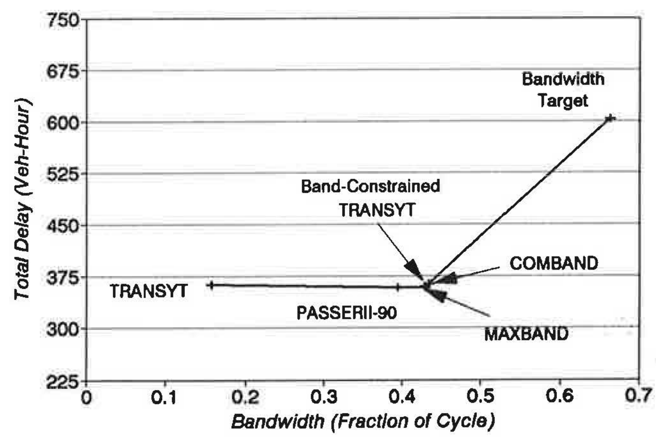

FIGURE 6 Total delay versus bandwidth (Skillman Avenue).

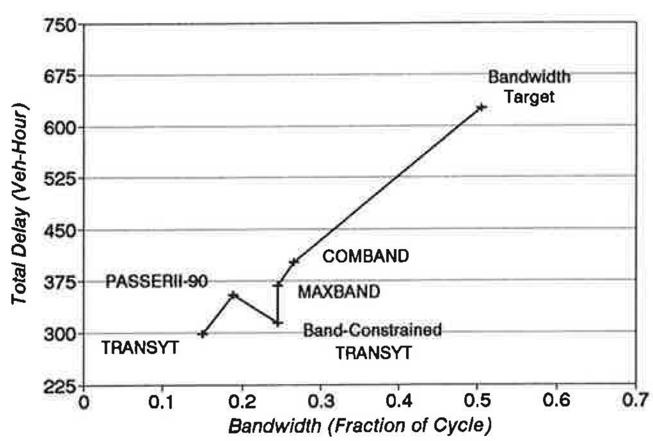

FIGURE 7 Total delay versus bandwidth (12th Street, a.m. peak).

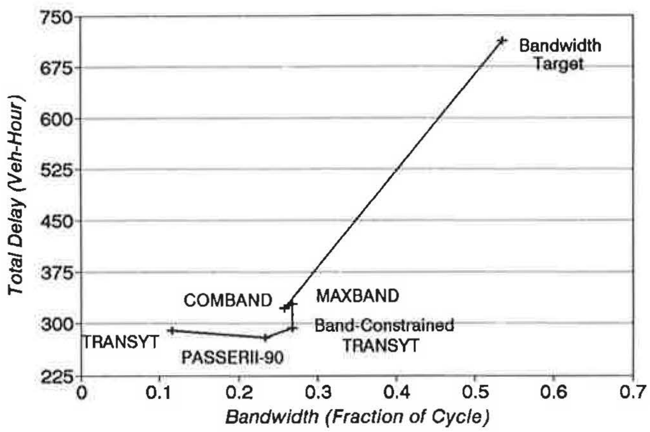

FIGURE 8 Total delay versus bandwidth (12th Street, p.m. peak).

generated solutions is worse than the other models in terms of average number of stops because good progression bandwidth is not provided.

4. Bandwidth-constrained TRANSYT procedures can reduce further delay by subsequently optimizing the MAXBAND initial solutions with the preserved progression bandwidth.

5. PASSER-II 90 shows a tendency to provide a slightly lower bandwidth since it yields more green time for side streets. In general, it produces less delay than the other models.

It is noted that it is not objective to compare these timing solutions by only one criterion. Readers may gain a whole picture of this issue by examining these self-explanatory figures. (Figures 6–8).

CONCLUSIONS

Under a decision-making process, evaluating timing solutions depends on how the decision maker trades off one criterion against the other. In the previous research, two approaches combining delay-minimization and bandwidth-maximization considerations were used to solve the signal coordination problem. One approach is to adjust or fine-tune bandwidth-based timing solutions to further minimize delay by applying a delay-based optimization program. The other approach maximizes bandwidth by modifying the delay-based solutions. These two approaches optimize both timing criteria by either "marrying" two types of programs or adding subprocedures internally or externally. Following concepts similar to those originally proposed in the MITROP model, this paper proposes an alternative approach that provides a viable methodology for simultaneously optimizing two operational criteria delay and progression that are normally conflicting when working with well-timed traffic signal systems.

It is shown that the compromised approach of combining delay/stops and progression bands simultaneously in developing arterial signal timing plans during unsaturated conditions exhibits several advantages over some existing approaches. This approach optimizes the arterial signal timing by performing the trade-off analysis between delay/stop and progression bands criteria simultaneously through MILP method without separately performing the "subprocedures" (such as adjusting the offsets or green splits), although some preprocessing is required. Compared with the basic MAXBAND model, this approach may produce better signal timing plans in terms of delay. The quality of progression bandwidth is still maintained even though some degree of bandwidth may be lost. This approach explicitly optimizes cycle length, system offsets, green splits, and phasing sequences at the same time to achieve a global optimal timing solution. From this methodology, we can either maximize the progression bandwidth at a given user-defined level of service for the side streets or minimize the delay value at some degree of loss in bandwidth.

FUTURE WORK

1. The assumptions and simplifications applied in the underlying model—such as time-stationary flow rate, unsaturated conditions, and no platoon dispersion—can be relaxed further to accommodate real-world situations.

2. The capability of the proposed model can be extended to handle network cases.

3. Multiobjective programming techniques can be applied to deal with signal timing problems having multiple operational criteria.

4. LP-type models suffer from a tremendous computational burden in dealings with a large-scale network problem. Decomposition techniques or other heuristic methods can be introduced to alleviate such a suffering.

ACKNOWLEDGMENTS

This paper is based on research funded by the Texas Department of Transportation (TxDOT) under its Highway Planning and Research Program, whose support is gratefully acknowledged. The authors are solely responsible for all findings, conclusions, and recommendations presented herein.

REFERENCES

1. E. C-P. Chang and C. J. Messer. Minimum Delay Optimization of a Maximum Bandwidth Solution to Arterial Signal Timing. In *Transportation Research Record 1005*, TRB, National Research Council, Washington, D.C., 1985, pp. 80–89.
2. S. L. Cohen and C. C. Liu. The Bandwidth-Constrained TRANSYT Signal-Optimization Program. In *Transportation Research Record 1057*, TRB, National Research Council, Washington, D.C., 1986, pp. 1–9.
3. J. D. C. Little, M. D. Kelson, and N. H. Gartner. MAXBAND: A Program for Setting Signals on Arteries and Triangular Networks. In *Transportation Research Record 795*, TRB, National Research Council, Washington, D.C., 1981, pp. 40–46.
4. N. H. Gartner, J. D. C. Little, and H. Gabbay. Simultaneous Optimization of Offsets, Splits, and Cycle Time. In *Transportation Research Record 596*, 1976, pp. 6–15.
5. N. H. Gartner, J. D. C. Little, and H. Gabbay. Optimization of Traffic Signal Settings by Mixed-Integer Linear Programming—Parts I and II. *Transportation Science*, Vol. 9, No. 4, Nov. 1975, pp. 321–363.

6. E. C-P. Chang, C. J. Messer, and S. L. Cohen. Directional Weighing for Maximal Bandwidth Arterial Signal Optimization Programs. In *Transportation Research Record 1057*, TRB, National Research Council, Washington, D.C., 1986, pp. 10–19.
7. N. H. Gartner, S. F. Assmann, F. Lasaga, and D. L. Hou. A Multi-Band Approach to Arterial Traffic Signal Optimization. *Transportation Research*, Vol. 25B, No. 1, 1991, pp. 55–74.
8. S. Yagar. Minimizing Delay at a Signalized Intersection for Time-Invariant Demand Rates. *Transportation Research*, Vol. 9, 1975, pp. 129–141.
9. *Special Report 209: Highway Capacity Manual*, TRB, National Research Council, Washington, D.C., 1985, Chapter 9.
10. N. A. Chaudhary and T. H. Wu. *Simultaneous Optimization of All Signal Timing Variables*. Working paper. Texas Transportation Institute, Texas A&M University System, College Station, 1991.
11. F. V. Webster. *Traffic Signal Settings*. Road Research Technical Paper 39, 1958.
12. L. Schrage. *User's Manual for Linear, Integer and Quadratic Programming with LINDO*, 3rd ed. The Scientific Press, Redwood City, Calif., 1986.

APPENDIX

NOTATION

$q_{1i}(\bar{q}_{1i}), q_{2i}(\bar{q}_{2i})$ = outbound (inbound) arterial through and sidestreet turning traffic flows of link i (vps);

$b_i(\bar{b}_i)$ = outbound (inbound) bandwidth of link i (cycle);

$r1_{mi}(\bar{r1}_{mi}), r1_{ci}(\bar{r1}_{ci})$ = first component outbound (inbound) main-street red time and sidestreet red time of intersection i (cycle);

$r1_{mi}(\bar{r1}_{mi}), r1_{ci}(\bar{r1}_{ci})$ = second component outbound (inbound) main-street red time and sidestreet red time of intersection i (cycle);

$D1_i(\overline{D1}_i), D2_i(\overline{D2}_i)$ = disutility coefficients associated with each red time variable;

$l_{mi}(\bar{l}_{mi}), l_{ci}(\bar{l}_{ci})$ = outbound (inbound) main-street left-turn green and sidestreet left-turn green time of intersection i;

$w_i(\bar{w}_i)$ = outbound (inbound) leading slack green time outside the bandwidth (cycle);

$\tau_i(\bar{\tau}_i)$ = queue clearance time (cycle);

$t_i(\bar{t}_i)$ = journey time from intersection i to $i + 1$ (i to $1 + i$) (cycle);

$\phi_i(\bar{\phi}_i)$ = internode offsets = time from the beginning of green time at intersection i to beginning of green time at intersection $i + 1$ ($i + 1$ to i) (cycle);

z, C_1, C_2 = reciprocal of common cycle length, lower and upper limits on cycle time;

$e_i, f_i(\bar{e}_i, \bar{f}_i)$ = lower and upper limits on link travel speed from intersection i to $i + 1$ ($i + 1$ to i) (ft/sec);

$g_i, h_i(\bar{g}_i, \bar{h}_i)$ = lower and upper limits on change in link travel speed from link i to $i + 1$ ($i + 1$ to i) (ft/sec); and

n = number of intersections.

FORMULATION

Given

$$q_{1i}(\bar{q}_{1i}),\ q_{2i}(\bar{q}_{2i}),\ D1_i(\overline{D1}_i),\ D2_i(\overline{D2}_i),\ e_i(\bar{e}_i),$$

$$f_i(\bar{f}_i),\ g_i(\bar{g}_i),\ h_i(\bar{h}_i),\ C_1,\ C_2$$

Find

$$b_i,\ \bar{b}_i,\ r1_{mi},\ (\bar{r1}_{mi}),\ r2_{mi},\ (\bar{r2}_{mi}),\ r1_{ci},\ (\bar{r1}_{ci}),$$

$$r2_{ci},\ (\bar{r2}_{ci}),\ l_i,\ \bar{l}_i,\ w_i,\ \bar{w}_i,\ \tau_i,\ \bar{\tau}_i,\ z,\ \delta_i,\ \bar{\delta}_i,\ \lambda_i,\ \bar{\lambda}_i,\ m_i$$

to maximize

$$\sum_i [(q_{1i} - q_{2i})b_i + (\bar{q}_{1i} - \bar{q}_{2i})\bar{b}_i - q_{2i}(r1_{mi} + r2_{mi}) - \bar{q}_{2i}(\bar{r1}_{mi} + \bar{r2}_{mi}) - (D1_i r1_{mi} + \overline{D1}_i \bar{r1}_{mi} + D2_i r2_{mi} + \overline{D2}_i \bar{r2}_{mi})]$$

subject to

$$w_i + b_i \le 1 - r1_{mi} - r2_{mi} \qquad i = 1, \ldots, n - 1$$

$$w_h + b_i \le 1 - r1_{mh} - r2_{mh} \qquad h = 2, \ldots, n$$

$$\bar{w}_i + \bar{b}_i \le 1 - \bar{r1}_{mi} - \bar{r2}_{mi} \qquad i = 1, \ldots, n - 1$$

$$\bar{w}_h + \bar{b}_i \le 1 - \bar{r1}_{mh} - \bar{r2}_{mh} \qquad h = 2, \ldots, n$$

$$(w_h - \bar{w}_h) + (\bar{w}_i - w_i) + t_i + \bar{t}_i + (r1_{mh} + r2_{mh} - \bar{r1}_{mh} - \bar{r2}_{mh}) + (\bar{r1}_{mi} + \bar{r2}_{mi} - r1_{mi} - r2_{mi}) + \delta_h l_h - \delta_i l_i - \bar{\delta}_h \bar{l}_h + \bar{\delta}_i \bar{l}_i - (\tau_h + \tau_i) = m_i \qquad i = 1, \ldots, n - 1, \quad h = 2, \ldots, n$$

(The loop equations have been modified for releasing green split as variables)

$$zr_i = \min\{1, z/z_{1i}\} \qquad i = 1, \ldots, n$$

$$r_{\min,i} \le r1_{mi}(\bar{r1}_{mi}),\ r1_{ci}(\bar{r1}_{ci}) \le r_{0.5,i} \qquad i = 1, \ldots, n$$

$$0 \le r2_{mi}(\bar{r2}_{mi}),\ r2_{ci}(\bar{r2}_{ci}) \le (r_{\max,i} - r_{0.5,i}) \qquad i = 1, \ldots, n$$

$$l_i \ge l_{\min,i} \qquad i = 1, \ldots, n$$

$$\tau_i = \max[que_i, 0] \quad \text{(queue clearance time constraints)} \qquad i = 1, \ldots, n - 1$$

$$r1_{mi} + r2_{mi} + (\text{Webster split on main}) \cdot zr_i - \bar{l}_{mi} \le 1 \qquad i = 1, \ldots, n$$

$$\bar{r1}_{mi} + \bar{r2}_{mi} + (\text{Webster split on main}) \cdot zr_i - l_{mi} \le 1 \qquad i = 1, \ldots, n$$

$$r1_{ci} + r2_{ci} + (\text{Webster split on side}) \cdot zr_i - \bar{l}_{ci}$$

$$\leq 1 \quad i = 1, \ldots, n$$

$$\overline{r1}_{ci} + \overline{r2}_{ci} + (\text{Webster split on side}) \cdot zr_i - l_{ci}$$

$$\leq 1 \quad i = 1, \ldots, n$$

$$l_{mi} - \overline{r1}_{mi} - \overline{r2}_{mi} = \bar{l}_{mi} - r1_{mi} - r2_{mi} \qquad i = 1, \ldots, n$$

$$l_{ci} - \overline{r1}_{ci} - \overline{r2}_{ci} = \bar{l}_{ci} - r1_{ci} - r2_{ci} \qquad i = 1, \ldots, n$$

$$-l_{mi} + \overline{r1}_{mi} + \overline{r2}_{mi} - \bar{l}_{ci} + r1_{ci} + r2_{ci} = 1$$

$$i = 1, \ldots, n$$

$$(d_i/f_i)z \leq t_i \leq (d_i/e_i)z \quad i = 1, \ldots, n - 1$$

$$(\bar{d}_i/\bar{f}_i)z \leq \bar{t}_i \leq (\bar{d}_i/\bar{e}_i)z \quad i = 1, \ldots, n - 1$$

$$(d_i/h_i)z \leq (d_i/d_{i+1})t_{i+1} - t_i \leq (d_i/g_i)z \quad i = 1, \ldots, n - 2$$

$$(\bar{d}_i/\bar{h}_i)z \leq (\bar{d}_i/\bar{d}_{i+1})\bar{t}_{i+1} - \bar{t}_i \leq (\bar{d}_i/\bar{g}_i)z \quad i = 1, \ldots, n - 2$$

(The lower and upper limits on journey time have been modified in COMBAND considering the degree of saturation of each link.)

$$1/C_2 \leq z \leq 1/C_1$$

m_i = integer variables

$\delta_i, \bar{\delta}_i = 0 - 1$ binary variables

This paper does not represent the official views of policies of TxDOT or FHWA. This document does not constitute a standard, specification, or regulation.

Publication of this paper sponsored by Committee on Traffic Signal Systems.

Regional Arterial (Super Street) Concepts

RANDY B. MACHEMEHL AND WILLIAM V. WARD

Freeways in urban areas routinely experience significant traffic congestion during several peak demand hours. Construction of additional freeway lanes could solve this problem, but because of economic, political, and social constraints, this is not always feasible. Diversion of short trips from freeways to arterial streets could help solve the problem without construction of new freeway lanes. However, such diversion will occur only if the traffic speed and capacity of selected urban arterial streets can be improved significantly. The regional arterial, or super street, is proposed as a class of facility that would have the continuity, speed, and capacity characteristics to attract short and medium-length trips. Through several research efforts, design and operational guidelines for regional arterial (super) streets have been developed. These are conceptually described and evaluated in terms of significance to the success of the regional arterial concept. A case study of potential impacts of several alternative regional arterial networks for the Houston/Harris County, Texas, area is described. Within the limitations of simulation, using a large network modeling package, the alternative case study network effects are compared.

During the past three decades, urban areas have experienced extensive growth in population, land area, and traffic congestion. Past approaches to maintaining urban mobility relied on constructing new lane miles of freeways, highways, and streets. The costs of such construction in terms of social, environmental, and political impacts, as well as monetary requirements, have prevented construction activities from keeping pace with urban growth. Today there is little probability that large quantities of lane miles of new facilities will be constructed in congested urban areas. Therefore, there is a great need for new concepts that will enhance the functionality and increase the capacities of existing facilities. The regional arterial or super street is one such concept.

BACKGROUND

A regional arterial as envisioned here would consist of an upgraded arterial street with certain distinct design and operating characteristics. It would have design speeds of 40 to 50 mph, grade separations at railroads and some cross streets, partial access control, and favored treatment for arterial traffic at nongrade separated intersections. Additionally, it would include median barrier separation, very few or no left turns, and an auxiliary or collector-distributor lane to the driver's right functioning as a speed change lane for entering and exiting traffic. However, the key element that is necessary to produce successful regional arterial implementation is a highly disciplined operating policy that guarantees that these features are maintained.

The regional arterial concept is not a freeway, although it has many similar characteristics. The primary differences are lower design speeds, partial access control, and infrequent nongrade separated intersections.

In the worst case, estimated costs of regional arterials are less than half the cost of new freeways, but they would provide half to two-thirds the traffic productivity of a freeway having the same number of lanes. Most regional arterials would be upgraded versions of existing arterials and might require little new construction. Another extremely significant advantage for the regional arterial concept is that the right-of-way requirements for new alignments would be less than those of a freeway with the same number of lanes.

Regional arterial streets can be viewed from two different but complementary contexts. First, they may serve as surrogates for new freeway lane miles that will never be built because of economic, social, environmental, and political constraints. Second, they may serve as a new functional class of urban street that can provide much more appropriate and efficient use of other facility classes.

An ideal representation of urban trip length versus facility class is shown in Figure 1. This figure indicates that the shortest trips would ideally use local streets, and successively longer trips would use successively higher functional facility classes. Freeways, the highest functional class, would serve only for the longest trips. Due to a lack of route continuity, inadequate potential travel speeds, and capacities, arterial streets are typically unable to attract appropriate numbers of intermediate-length trips. Thus, large numbers of short trips are frequently diverted to freeways, the next-higher facility class.

One result of the lack of continuity and the inherent unacceptable travel speeds on nonfreeway routes is the diversion of much more than a proportional share of all travel to the urban freeway system. The ratio of daily vehicle miles of travel (VMT) and facility centerline mileage in the Houston/Harris County area is shown for three facility classes in Figure 2.

The two nonfreeway facility classes noted in the figure are those that provide continuous routes of more than 4 mi and those with less than 4 mi of route continuity. The figure presents the expected conclusion that freeways carry the vast majority of the daily VMT (DVMT) in this automobile-oriented area. Specifically, freeways carry almost 80,000 DVMT/centerline-mi, whereas the facilities with less than 4 mi of continuous routes carry only about 1,600 DVMT/centerline-mi. However, Figure 3 shows the percentage of DVMT for each facility class and the percentage of total highway centerline miles represented by each class. The figure shows that although freeways represent about 2 percent of the total cen-

R. B. Machemehl, Department of Civil Engineering, University of Texas, Austin, Tex. 78712. W. V. Ward, Center for Transportation Research, University of Texas, Austin, Tex. 78712.

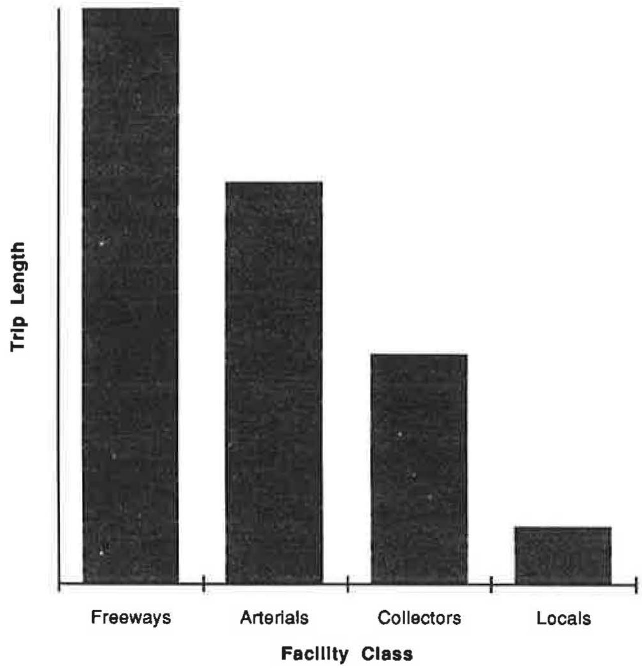

FIGURE 1 Ideal trip length versus facility class.

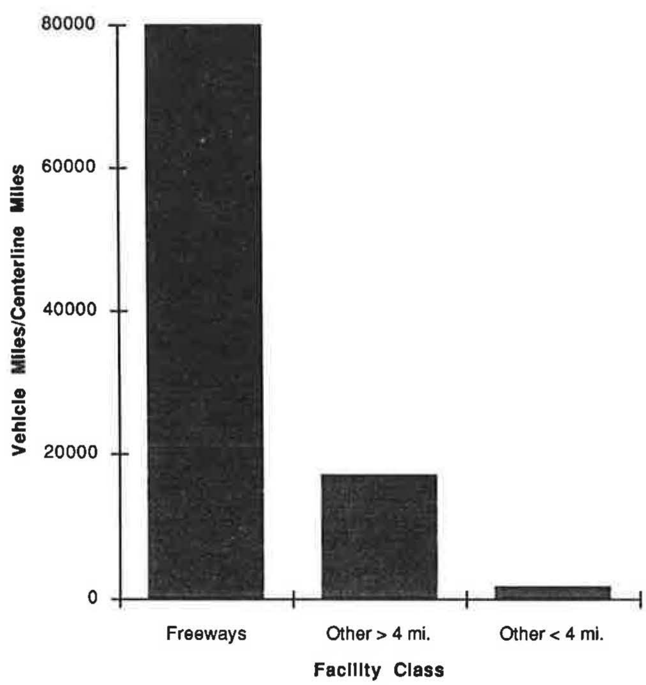

FIGURE 2 Daily vehicle miles per centerline mile by facility class.

terline miles of highway facilities, they carry 40 percent of the DMVT.

Diversion of many short trips to freeways degrades freeway traffic flow capabilities by abnormally increasing entrance and exit operations. Freeway flows are limited almost always by weaving section capacities near entrance and exit ramps. If short trips could be attracted to a lower functional facility

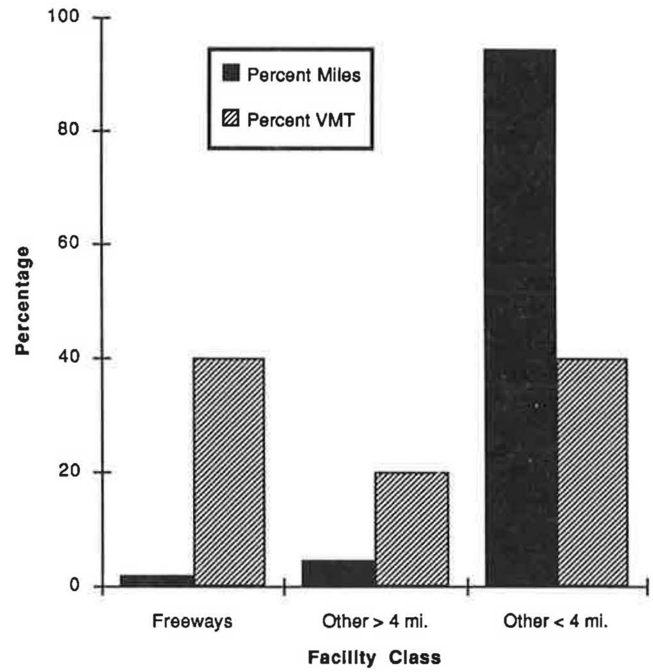

FIGURE 3 Percentage centerline miles and VMT by facility type.

class, away from freeways, the traffic movement capability for the long trips for which freeways were really designed could be enhanced. The regional arterial is proposed as a new facility class that would provide the continuity, travel speed, and capacity that could attract short trips away from freeways.

DESIRABLE DESIGN AND OPERATING CHARACTERISTICS

Two research studies (*1*,*2*) and several years of gathering and analyzing information about regional arterial, strategic arterial, or super streets have produced a compendium of desirable design and operating characteristics.

Continuity

Lack of continuity is one of the most significant problems plaguing arterial streets. Texas cities are notorious for this lack of continuity. In Austin, Texas, for example, it is not possible to travel from east to west between the corporate limits of the city along one arterial street. On the other hand, Houston does have long arterials, but geometrics and traffic control features are so variable along most routes that traffic flow continuity is not maintained.

Regional arterial streets and networks composed of these types of facility must provide sufficient continuity to attract travelers making trips of a length at least equivalent to the average areawide trip length. This will obviously vary among cities depending on many factors; however, it is typically from 5 to 12 mi. A 1984 regional telephone survey of Houston travelers (*1*) indicated the average vehicular trip length was 7 mi and the average commute trip was 11 mi. It is apparent that if a regional arterial is to be effective, it must provide continuity for at least the average trip length.

The concept of route continuity envisioned here is more than a continuous alignment with a common or appropriate geometric cross section. The concept also includes traffic control features necessary to provide continuous high-speed travel.

Signal Timing Policy

The capacity of any approach to a signalized intersection primarily depends on the percentage of the available green time provided to that approach or the effective green time to cycle length (g/C) ratio. If a regional arterial street is to have the potential to carry large volumes of traffic over long continuous street sections, the g/C ratios for all intersection approaches must be appropriately established by operating policy. This policy must clearly establish traffic operations on the regional arterial as preferred movements. A policy of providing 70 percent of the available intersection green time to movements on the regional arterial is considered the desirable minimum.

Because of the relative ease with which signal timing can be changed, the difficulty of maintaining this recommended policy is certainly recognized. Cross streets that do not have regional arterial designation will obviously be penalized; this will clearly create user complaints. However, somewhat like ramp metering on freeways, this policy can produce huge overall user benefits despite increased delay for some users. User complaints and increased delay can be ameliorated through designation of a network of regional arterials and the self-leveling effect of users seeking the fastest, least time-intensive travel paths.

Geometric Design Standards

Geometric design standards for regional arterial streets must be appropriate for high-volume, relatively high speed facilities. Minimum design speeds of 40 to 50 mph should be adopted and strictly maintained. Grade separation structures should be provided for all existing at-grade railroad crossings. Geometric features should be provided to prohibit left-turns from all cross streets and abutting property. In addition, where left-turns are allowed from the arterial, appropriate storage such as turning bays or lanes should be provided. A median barrier for the arterial is recommended as an effective left-turn control measure that enhances overall safety but does not consume extensive, scarce right-of-way areas.

Mass Transit Operations

The high operating speeds and continuity of regional arterials make them ideal for public transportation routes. However, buses stopping in the travel lanes for passenger loading and unloading would be highly disruptive to the advocated kind of traffic service. Therefore, all public transportation passenger boarding and deboarding operations should be performed while buses are in appropriately designed bus turnouts.

Access Control Standards

Positive control of turn movements is a primary element of the regional arterial philosophy of favoring long through trips. As noted previously, median barriers are an excellent means of controlling left turns from the arterial and allowing them only at carefully selected locations. Additionally, right turns from the arterial should be separated from through traffic. One highly desirable means of accomplishing this is the use of collector-distributor lanes from which all turns must be made. Geometrics of the paths to be followed by vehicles making right-turn movements into and off the arterial should resemble freeway entrance-exit ramps—that is, these facilities should follow gradual tapers permitting exits at arterial speeds and appropriate acceleration distances for entering vehicles.

Right-of-way For Undeveloped Areas

Most lane miles of regional arterial streets that are developed in the near future will be in developed areas where right-of-way (ROW) is probably scarce and new acquisitions are very expensive economically, socially, and environmentally. However, where regional arterials are designed for extensions into undeveloped areas, ROW specifications should be developed around desirable, not minimum, standards: when given the chance to do it right, designers should take advantage of the opportunity.

Intersection Grade Separations

Existing and certainly future traffic demands will justify construction of grade separation structures at some locations. Two observations about grade separations are extremely important. First, the regional arterial concepts described here do not assume construction of grade separations at all cross-street intersections. They should be considered only where the crossing traffic volumes are large enough that a critical user delay problem exists and cannot be solved by any less expensive measure. Such locations will be few when the typical regional arterial is first designated and constructed. In other words, the inability of constructing grade separations should not preclude design and implementation of regional arterial streets. Second, whereas AASHTO policy demands very long vertical curves for grade separations, minor changes to these specifications can produce much shorter curves that can be designed to fit between existing intersections. Thus, grade separations can be constructed so that they do not entirely preclude desirable access to abutting property.

CASE STUDIES

The potential impact of a network of regional arterials on a typical large, automobile-oriented urban area was examined through a simulation study of the Houston regional urban area. The simulation was performed using the Texas Large Network Computer Package running on the Texas Department of Transportation mainframe computer system. Assigned traffic consisted of the forecast year 2010 daily highway volumes and the planned 2010 regional highway system.

The simulation consisted of only a very general set of modifications to the planned 2010 regional highway system. These included

1. Addition of links composing 15 and 62 new centerline-mi for the 350- and 600-mi strategic arterial systems, respectively.
2. Speed increases of 5 or 10 mph (two different cases) on the links of the strategic arterial systems. These increases were applied to the normal arterial speeds.
3. Appropriate capacity specifications for the links of the strategic arterials (this generally involved increasing capacity specifications simulating the addition of lanes or modified signal timing).
4. Decreases in link capacities for links crossing the strategic arterials. This represented an effort to simulate effects of signal timing strategies favoring arterial traffic.

Several regional arterial improvement concepts were not simulated. These included control or prohibition of left turns, and grade separations. Additionally, simulation of the 70 percent arterial green recommendation was not attempted.

The simulation case study represented an approximate potential effects worst case. The modeling system used was based on daily traffic, not hourly demands. Therefore, simulation of certain details of regional arterial concepts was somewhat difficult.

Summary statistics for the two alternative regional arterial systems that were examined are presented in Table 1. As indicated in the Table, the approximate 350-mi strategic arterial system was composed of 337 mi of upgraded existing street alignment with only 15 new centerline mi. The more extensive, roughly 600-mi system included 519 mi of upgraded street and only 62 new centerline mi. Thus, both alternatives represented very modest new alignment additions.

Effects of the 350-mi strategic arterial system on other facility classes are presented in Table 2 and Figure 4.

Table 2 indicates that 5 and 10-mph levels of increase compared to existing conditions were simulated for the strategic arterial links. As expected, the 10-mph increase for regional arterials, which produced weighted average speeds of 48 mph, caused more diversion from freeways and other facilities to the regional arterial facility class. Figure 4 compares percentages of centerline miles for each facility class to DVMT before (base) and after (strategic arterial system, or SAS) the strategic arterial addition. The chart indicates the 350-mi system would reduce freeway DVMT by 3 to 4 percent, have minimal impact on other facility classes, and double DVMT on the facilities redesignated as regional arterials. Almost 20 percent of the 1,042 freeway links included in the simulation had daily traffic volumes reduced by more than 10,000, which may be far more significant than the DVMT reductions.

A similar presentation of effects due to the alternative 600-mi strategic arterial system is included as Table 3 and Figure 5. For this system, which featured only 62 lane-mi of new alignment, freeway DVMT was reduced roughly 6 percent, and nearly 30 percent of the 1,042 freeway links had average daily traffic volume reductions of 10,000 or more.

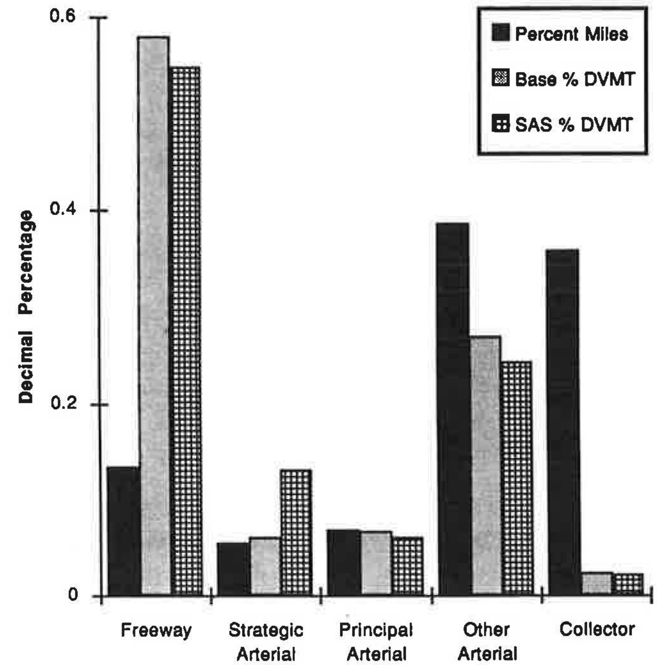

FIGURE 4 Percentage centerline miles and daily vehicle miles for base and 350-mi strategic arterial system networks—Houston/Harris County, Texas, region.

TABLE 1 Alternative Regional Arterial Network for Houston Region

Facility Class	350 Mile Strategic Arterial System Total Centerline Miles	600 Mile Strategic Arterial System Total Centerline Miles
Freeway	875	875
Facility to be upgraded to Strategic Arterial	352[a]	581[b]
Principal Arterial	447	440
Other Arterial	2546	2428
Collector	2352	2294

[a] Includes 15 miles of extended and 337 miles of upgraded facilities.
[b] Includes 62 miles of extended and 519 miles of upgraded facilities.

TABLE 2 Effects of 350-mi Strategic Arterial System

	Assigned Daily Vehicle Miles of Travel (000's)		
		Strategic Arterial Mileage = 350	
	Base	Speed Increase (mph)	
Facility Class	Network	+5	+10
Freeway	55,426	52,655	51,860
Strategic Arterials	5,831[a]	10,827	12,347
Principal Arterial	6,347	5,737	5,753
Other Arterial	25,690	23,528	22,950
Collector	2,132	2,019	1,984

[a] Facilities to be expanded to Strategic Arterial status

TABLE 3 Effects of 600-mi Strategic Arterial System

	Assigned Daily Vehicle Miles of Travel (000's)		
		Strategic Arterial Mileage = 600	
	Base	Speed Increase (mph)	
Facility Class	Network	+5	+10
Freeway	55,426	50,798	49,583
Strategic Arterials	7,939[a]	15,729	17,899
Principal Arterial	6,255	5,598	5,568
Other Arterial	23,872	20,715	20,126
Collector	1,934	1,610	1,577

[a] Facilities to be expanded to Strategic Arterial status

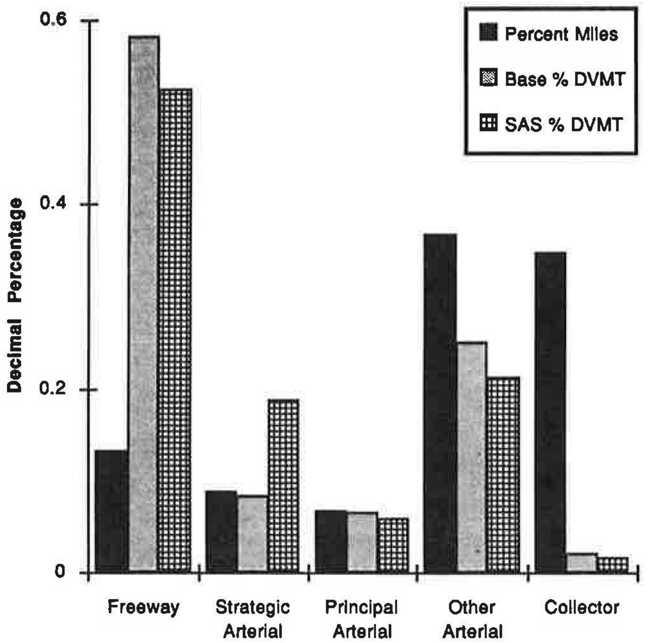

FIGURE 5 Percentage centerline miles and daily vehicle miles for base and 600-mi strategic arterial system networks—Houston/Harris County, Texas, region.

SUMMARY

This paper has summarized regional arterial street concepts that are practically implementable and offer great potential highway user benefits. The concept consists of

- An upgraded, or extended, arterial street with distinct design and operating characteristics that must be maintained through a disciplined policy;
- Design speeds of 40 to 50 mph;
- Grade separations at railroads and some cross streets;
- Partial access control;
- Favored treatment for arterial traffic at nongrade separated intersections;
- Median barrier separation;
- Very few or no left turns; and
- An auxiliary or collector-distributor lane to the driver's right functioning as a speed change lane for entering and exiting traffic.

The key element, however, is a highly disciplined operating policy that guarantees the maintenance of these features.

Estimated costs of regional arterials are, in the worst case, less than half the cost of new freeways, but the arterials would provide half to two-thirds the traffic productivity of a freeway having the same number of lanes. Most regional arterials would be upgraded versions of existing arterials and might require little new construction. Another extremely significant advantage for the regional arterial concept is that right-of-way requirements, for new alignments, would be less than that of a freeway with the same number of lanes.

The regional arterial concept is not a freeway, although it has many similar characteristics. The primary differences are lower design speeds, partial access control, and infrequent nongrade separated intersections.

ACKNOWLEDGMENT

Support for this study was provided by a grant from the U.S. Department of Transportation to the Southwest Region University Transportation Center.

REFERENCES

1. W. V. Ward. *Conceptual Strategic Arterial Street System for Harris County*. Research Report 3-10-88/0-428. Center for Transportation Research, University of Texas, Austin, March 1990.
2. D. L. Christiansen and W. V. Ward. *An Enhanced Role for the Arterial Street System in Texas Cities*. Research Report 1107-1F. Texas Transportation Institute, Texas A&M University, College Station, Nov. 1988.

Publication of this paper sponsored by Committee on Transportation System Management.

International High-Occupancy Vehicle Facilities

KATHERINE F. TURNBULL

The use of high-occupancy vehicle (HOV) facilities in North America, especially those on freeways and separate rights-of-way, has been examined extensively over the past 20 years. However, little is known about the extent of and experience with similar facilities in other parts of the world. This shortcoming is addressed by providing an overview and description of HOV lanes in operation in non-North American countries. The results of initial review indicate that HOV facilities are being used extensively in many parts of the world. Exclusive HOV lanes, on either separate rights-of-way or freeways and arterial streets, are in operation in 16 metropolitan areas around the world. The largest number of international HOV projects fall into the general category of non-exclusive HOV lanes on arterial streets. These types of projects have been identified in at least 75 cities. A general description of these facilities is provided. The similarities and differences between HOV projects in North America and other parts of the world are highlighted. Finally, mechanisms for improving the future exchange of information on international HOV projects are presented.

High-occupancy vehicle (HOV) facilities are one technique being used in many metropolitan areas to address traffic congestion and air quality concerns. Many areas are pursuing a wide spectrum of possible solutions in response to the continued increase in traffic congestion, the projected growth in travel demand, and limited financial resources and right-of-way availability. The use of HOV facilities, which focus on increasing the person-movement efficiency of a roadway or travel corridor, is a viable alternative being considered and implemented in many areas.

Since the 1969 opening of the exclusive bus lanes on Shirley Highway in Washington, D.C., many metropolitan areas in North America have developed priority facilities for high-occupancy vehicles. A variety of HOV treatments are currently in operation, including busways on separate rights-of-way, HOV lanes on freeways, HOV bypass lanes at freeway ramp meters, arterial street HOV lanes, and transit malls. As of April 1990, a total of 40 HOV projects in 20 metropolitan areas, representing 332 mi of HOV lanes, were in operation on freeways or on separate rights-of-way in North America (*1*). Moreover, arterial street HOV facilities and other priority measures are being used in many metropolitan areas.

Although the use of HOV facilities in North America, primarily those on freeways and on separate rights-of-way, has been examined extensively over the past 20 years (*1–5*), little is known about the extent of and experience with similar applications in other parts of the world. An overview and description of HOV lanes in operation throughout the world are provided, and projects in the planning and design stages are identified. In addition, the similarities and differences between HOV applications in North America and other parts of the world are discussed.

The results of this preliminary examination indicate many similarities in the issues, problems, and operating experiences among all projects. However, differences also exist in the approaches and techniques used in different areas. Examination of these similarities and differences can enhance the overall understanding of the role HOV facilities can play as well as the advantages and disadvantages of different approaches. To help facilitate the further sharing of information, which would benefit all groups, areas for additional research and analysis are identified, along with suggested methods to encourage and promote future information exchanges.

METHODOLOGY

Two techniques were used to obtain the information on international HOV applications contained in this report. First, a literature review was conducted to identify existing sources of information. This review identified a few journal articles and listings of projects. However, the general lack of published information supports the previously noted conclusion that few comprehensive data are available on international HOV facilities.

To obtain more detailed information, individual letters were sent to representatives from agencies, transit operators, consulting firms, and university research groups throughout the world. The names of possible contacts were identified by members of TRB's HOV Systems Committee and other individuals, as well as references obtained through the literature review. Approximately 30 letters were initially sent requesting information. More letters were sent on the basis of suggestions received in response to the first mailing.

Response to the initial request was good. Data, including reports and memorandums, were provided on a variety of HOV projects in many parts of the world. However, as could be expected, not everyone contacted has responded. Thus, the information in this paper does not fully address all the HOV projects that are thought to be in operation. In addition, not provided in many instances were detailed operating characteristics such as the number of vehicles and passengers using the facilities. Complete data on some projects are missing, but the results of the literature search and individual correspondence provide a good overview of the types of HOV facilities in operation throughout the world. As such, this paper enhances the level of understanding of international HOV

Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843.

facilities and establishes a base for additional research and analysis.

DESCRIPTION OF INTERNATIONAL HOV FACILITIES

Overview

This section provides a summary of the HOV projects in operation in non–North American countries. The classification system used to describe the different types of HOV projects represents a slight departure from those used in recent documents on HOV facilities in North America (*1,6*). This modification was made to reflect more accurately the types of HOV projects in operation throughout the world. Three general categories are used to characterized international HOV projects. These are exclusive HOV facilities on separate rights-of-way, exclusive HOV facilities on freeways or arterial streets, and nonexclusive HOV lanes on arterial streets.

This research identified six exclusive HOV projects on separate rights-of-way. These projects include two guided busway systems and four busways. Like the busways in Ottawa, Ontario, and Pittsburgh, Pennsylvania, these facilities, with one exception, are open only to buses. Ten exclusive HOV projects were identified in operation on freeways, expressways, and major arterial streets. In all cases, these lanes are restricted to buses only. The largest number of international HOV projects fall into the general category of nonexclusive HOV lanes on arterial streets. Most of these applications are concurrent-flow HOV lanes in downtown areas. These types of applications were identified in 75 cities around the world. This figure probably understates the use of this type of treatment, almost exclusively restricted to buses, because many large and medium-sized cities in Great Britain and Europe use some type of priority lane for buses in congested areas.

The HOV facilities described in this section are oriented much more toward bus-only applications than those in North America, but in many other respects they represent a greater diversity of applications. The types of facility range from guided busways to short queue jumps on local streets. There are also examples of HOV lanes implemented as major components of the transportation infrastructure. These include the development of a new town in Great Britain around a busway and the use of HOV lanes in developing countries to provide a basic level of transportation to large numbers of people. Furthermore, examples exist of the use of both sophisticated advanced technologies to provide additional priorities to transit vehicles and very basic approaches using buses with trailer units and maxi-taxis to increase the capacity of the lanes.

Exclusive HOV Facilities, Separate Rights-of-Way

The six exclusive HOV facilities located on separate rights-of-way are briefly described in this section. Figure 1 shows the location of these facilities, and Table 1 provides a summary of the major characteristics of each. Maps of individual facilities are provided for a few projects.

Adelaide, Australia

A 7.5-mi guided bus system is in operation in Adelaide, Australia. Officially called the Northeast Busway, the facility is commonly referred to as the Adelaide O-Bahn. The system was developed by Mercedes-Benz and was opened in phases between 1986 and 1989. Figure 2 shows the location and general service design of the O-Bahn system. Buses using the facility are equipped with special lateral guidewheels. These allow the vehicles to operate in normal service on local streets and to access the O-Bahn system by connecting the protruding

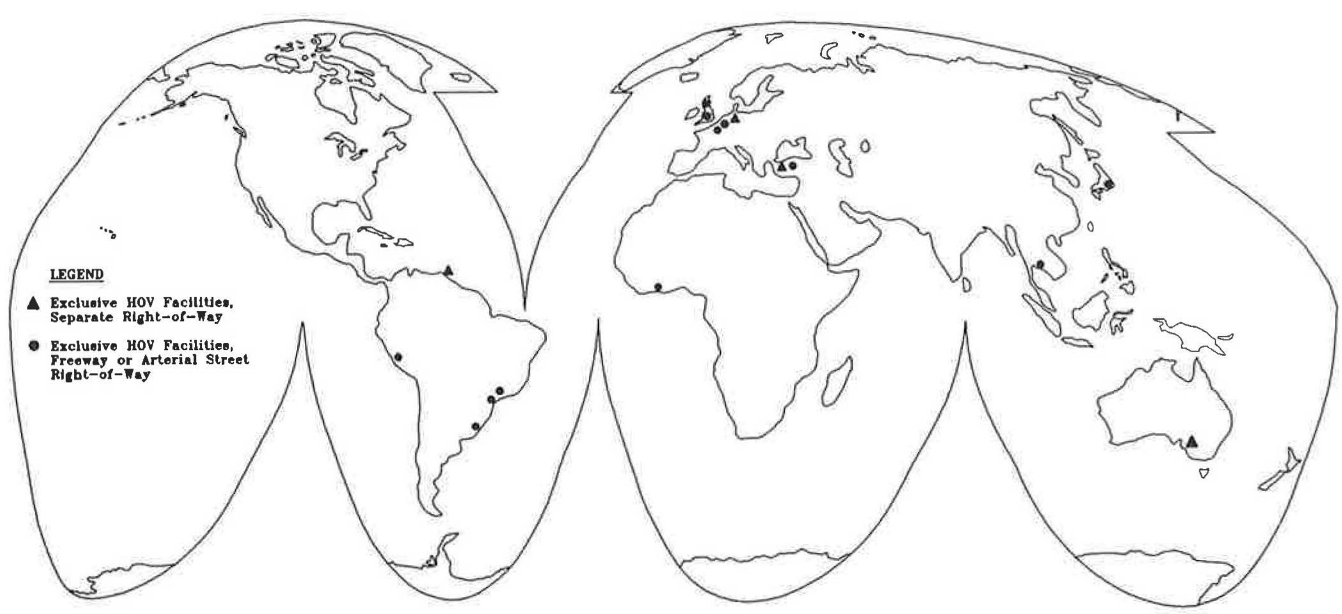

FIGURE 1 Non–North American exclusive HOV lanes.

TABLE 1 Summary of International HOV Projects—Exclusive Facilities on Separate Rights-of-Way (*7–9,11,15*)

Location	Description
Exclusive Facilities Separate Right-of-Way	
Adelaide, Australia	Northeast busway or O-Bahn system, as it is commonly referred to, is a 7.5 mile guided bus system in operation in Northeast Adelaide. Opened between 1986 and 1989, the system provides high speed bus operation on an exclusive guideway. Buses are equipped with special lateral guide wheels allowing operation on both local streets and the O-Bahn. Support services are provided.
Essen, Germany	A 4.5 mile guided busway is currently in operation. Opened in stages between 1980 and 1988 the system allows buses to operate on local streets and access the busway through the use of lateral guide wheels. Further testing of the use of dual-mode trolley bus/diesel buses is planned.
Istanbul, Turkey	A 4 mile two lane two-way bus-only street is in operation. Used by both public and private operators, the facility is located on the alignment of a proposed future rail line.
Redditch, Great Britain	Eight miles of exclusive busways are in operation in the city center and in areas connecting the different villages. The busway system was developed as one element of the overall transportation system designed in the Master Plan for the Redditch New Town.
Runcorn, Great Britain	New Town designed around a 13-mile 2-lane, 2-way busway. The system was developed to provide priority to public transit vehicles, control the use of automobiles and coordinate transportation and land use. The busway is 22 feet wide in most sections, with elevated sections in the city center. Signal preemption and supporting facilities and services are provided.
Port of Spain, Trinidad	Ten miles of a planned 15-mile busway are currently in operation. The busway, which is located in an railroad right-of-way, will link Port of Spain to Arima. Buses and privately owned and operated maxi-taxis are able to use the facility.
Exclusive HOV Facilities, Freeway or Arterial Street Right-of- Way	
Abidjan, Ivory Coast	Two busways are in operation parallel to arterial streets. Both are 2-lane, two-way busways, carrying approximately 150 buses in each direction during the peak-hours. One facility is 5.5 miles long and the other is slightly under 1 mile.
Ankara, Turkey	A 3.3 mile exclusive bus lane is in operation located adjacent to arterial street in major cross-city corridor.
Bangkok, Thailand	A total of 125 miles of bus lanes are in operation. These include both exclusive lanes next to arterial streets and concurrent flow lanes on arterial streets.
Curitiba, Brazil	35 miles of exclusive busways and bus lanes are in operation in four corridors. Daily ridership in all four corridors is approximately 340,000 passengers.
Liege, Belgium	A 6-mile 2-lane, two-way bus facility is located in the median of the Boulevard de la Sauveniere. In addition, 4 miles of non-separated bus lanes are in operation.
Lima, Peru	Exclusive 5-mile 2-lane, two-way busway located in median Paseo de la Republic Expressway. Opened in 1974, ridership levels average 5,000 passengers an hour in both directions.

(continued on next page)

TABLE 1 *(continued)*

Location	Description
Paris, France	Extensive 190 mile system of bus lanes, with some physically separated.
Porto Alegre, Brazil	A total of 18 miles of bus lanes are located in the median of carriageways in seven radial corridors. Using a system of platooning vehicles, volumes of up to 20,000 passengers per hour are carried.
Sao Paulo, Brazil	Physically separated and concurrent flow bus lanes are in operation in a number of corridors. Many of these are 2-lane, two-way facilities located in the median of the roadway. Platooning of buses is used to increase the capacity of the lanes.
Yokohama, Japan	Approximately 15 miles of exclusive busways, 20 miles of priority lanes, and a short segment of a contraflow lane are in operation. Other priority measures are also in use.

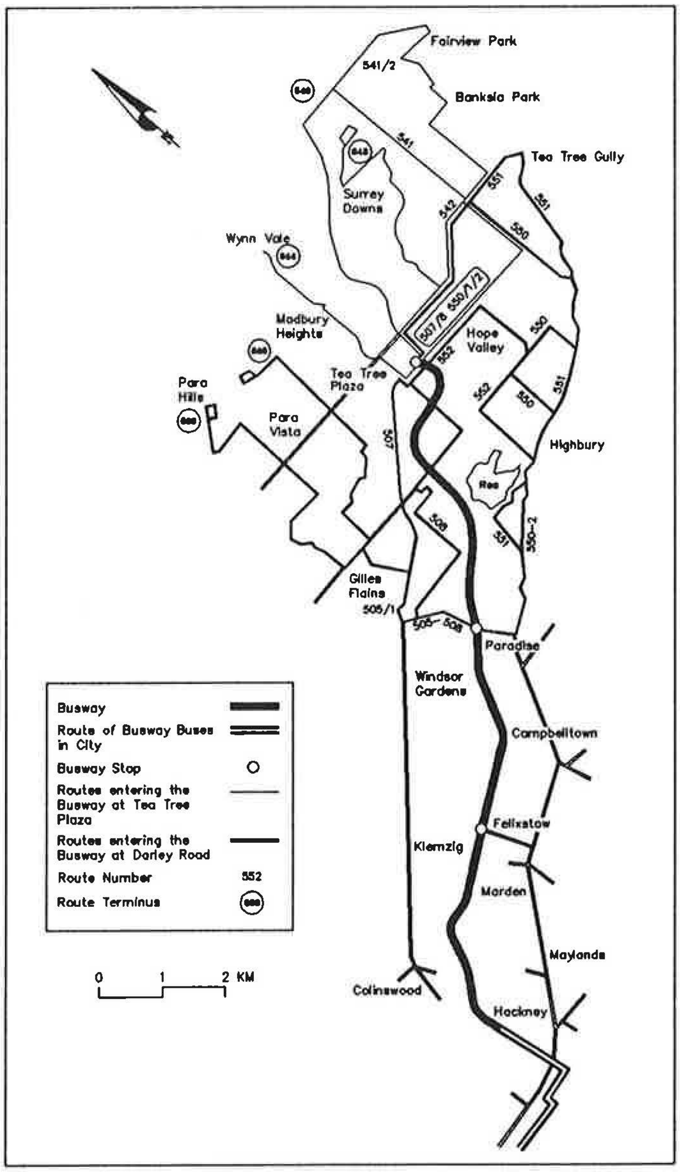

FIGURE 2 Adelaide O-Bahn.

guidewheels to the concrete track. The O-Bahn system allows buses to operate at speeds of up to 50 to 60 mph. The system is supported by park-and-ride lots, feeder buses, express buses operating only on the guideway, and routes that circulate in neighborhoods and then access the guideway (*7,8*).

Essen, Germany

A guided bus system is also in operation in Essen, Germany. Using funding from the Federal Ministry for Research and Technology, the local transit agency, and Daimler-Benz and MAN, a 4.5-mi guided bus system was implemented between 1980 and 1988. Like the system in Adelaide, buses can operate in regular service on local streets and then access the busway by engaging lateral guidewheels into the curbing on the concrete guideway. The next planned step in the project is to test the use of dual-mode trolleybus–diesel buses on the facility (*8*).

Istanbul, Turkey

A 4-mi two-lane, two-way bus-only road is in operation in Istanbul. The busway was developed on an alignment that may be converted to rail at some time. Both public and private buses are allowed to use the facility. Although no ridership levels are available, the facility is noted as being extensively used (*8*).

Port of Spain, Trinidad

Currently, 10 mi of a planned 15-mi busway are in operation in Port of Spain, Trinidad. As shown in Figure 3, when completed the busway will link Port of Spain to Arima. The two-way, two-lane busway was built on an abandoned railroad right-of-way. The facility is used by buses and privately owned and operated maxi-taxis. Ridership figures for 1987 indicated that some 55,000 passengers were being carried in buses; the maxi-taxis accounted for four times as many passengers. Maxi-taxis were prohibited from using the facility for a number of years because of concerns that they were taking potential passengers away from the public buses (*8*).

Redditch, England

Eight miles of exclusive busways were developed as part of the roadway system designed for the new town of Redditch in Great Britain. Located southeast of Birmingham, Redditch was designated as a new town in 1964 to help relieve overcrowding in the West Midlands conurbation. The Redditch master plan included both a roadway and a busway system. The busway system, which was completed in the late 1970s, includes exclusive rights-of-way in the city center and in areas connecting the different villages. Buses operate in the mixed-traffic lanes in other areas (*9–11*).

Runcorn, England

Runcorn, south of Liverpool, is another of the British new towns designated in 1964. In Runcorn, the busway concept was taken a step farther than the design used in Redditch. The master plan, and subsequent development of the new town, was designed around a 13-mi two-lane, two-way busway in the shape of a figure eight. As shown in Figure 4, the city center and shopping district is located at the center of the "eight," with residential, industrial, and smaller commercial

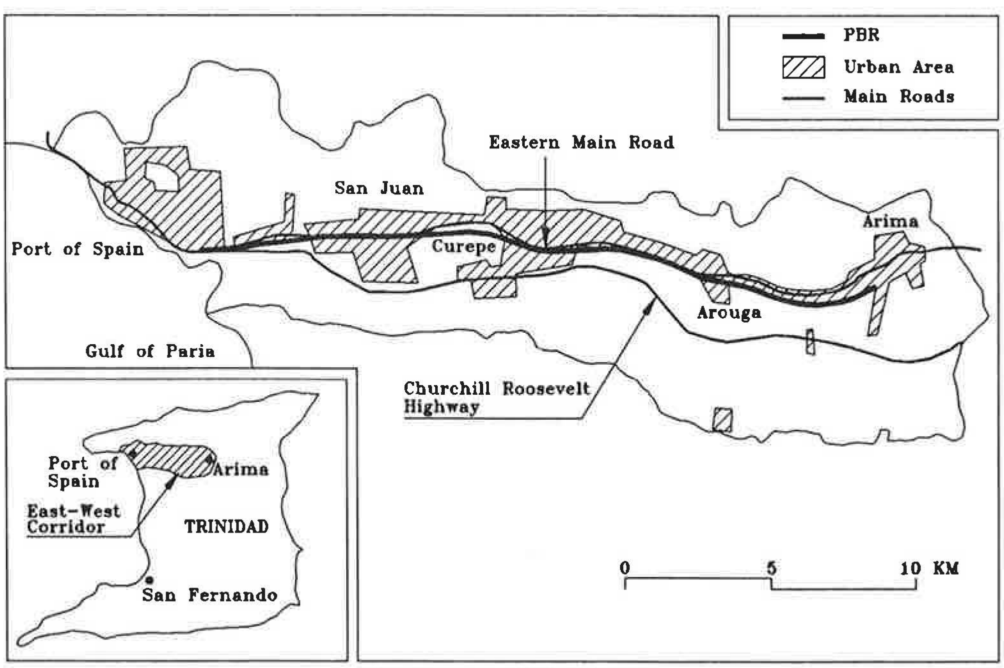

FIGURE 3 Port of Spain busway.

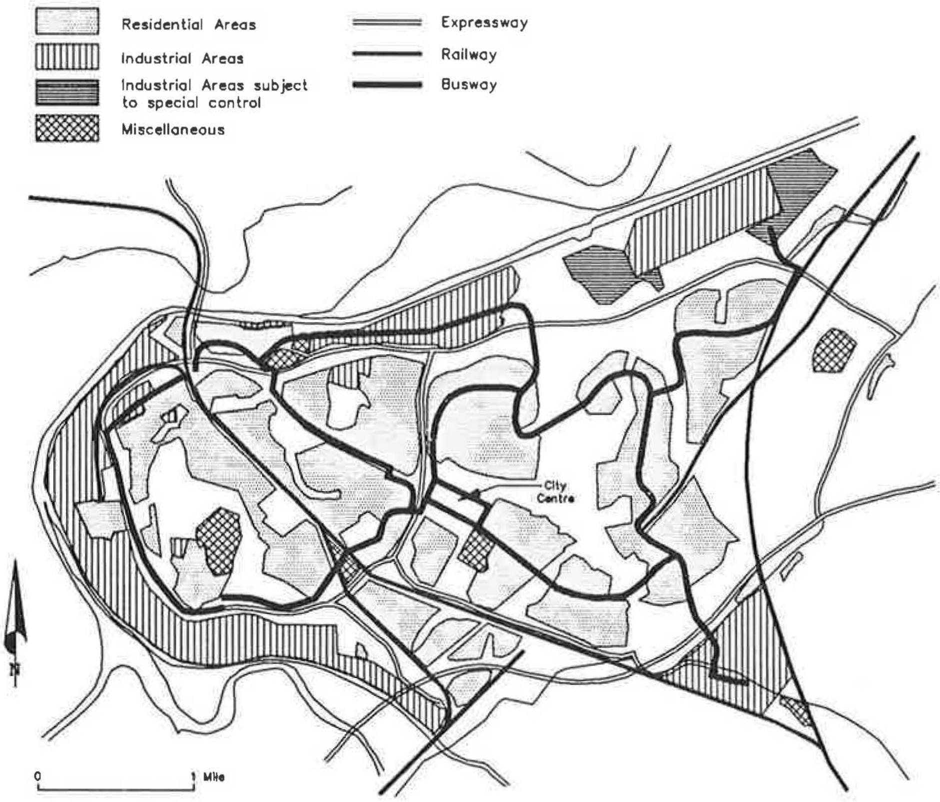

FIGURE 4 Runcorn busway.

areas located along the route. The busway concept was designed specifically to provide public transportation with an advantage, to control the use of the automobile for work trips, and to coordinate transportation and land use. The busway is 22 ft wide in most parts, with elevated sections in the city center. Traffic signal preemption is provided at most intersections to give buses priority. Bus headways range from 5 to 15 min depending on the time of day. Other support services and facilities are provided, and a variety of employer-coordinated programs have been used over the years (*9,12–15*).

Exclusive HOV Facilities, Freeway and Arterial Street Rights-of-Way

The 10 exclusive HOV facilities identified in operation on freeways, expressways, and major arterial streets are described in this section. Figure 1 illustrates the location of these projects, and Table 1 provides a summary of their major characteristics.

Abidjan, Ivory Coast

Two busways are in operation parallel to arterial streets in Abidjan. Both are two-lane, two-way facilities. One is 5.5 mi long and the other is slightly less than 1 mi. Both provide service in the heaviest travel corridors, averaging about 150 buses in each direction in the peak hours. Development of the busway and other transit facilities in Abidjan was financed by the World Bank (*8*).

Ankara, Turkey

A 3.3-mi exclusive bus lane is in operation adjacent to an arterial street in a major cross-city corridor in Ankara. Although no specific statistics are available, bus-rider volumes have been described as being very heavy during the peak periods (*8*).

Bangkok, Thailand

About 125 mi of bus lanes are in operation in Bangkok. These include both exclusive lanes next to arterial streets and concurrent-flow lanes on arterial streets. Both public transit vehicles and privately owned and operated minibuses are allowed to use the lanes (*8*).

Curitiba, Brazil

Approximately 35 mi of exclusive busways and bus lanes are in operation in Curitiba. This system was initiated in the 1970s as part of an integrated land use and transportation policy favoring public transit. Busways and bus lanes of 14, 7, 6, and 8 mi operate in four corridors. Total ridership in all four corridors averages 340,000 passengers a day. Development of the system was financed partially by the World Bank (*8*).

Liège, Belgium

A total of 10 mi of bus lanes are in operation in Liège, including 6 mi of physically separated lanes. The Boulevard de la Sauveniere bus lane is located in the median of an arterial street. It is separated from the general-purpose traffic lanes by a planted curb barrier (*8*).

Lima, Peru

An exclusive bus lane, approximately 5 mi in length, operates in the median of the Paseo de la Republic Expressway in Lima. The two-lane, two-way facility was opened in 1974 on the alignment of a planned rail line. Ridership levels average 5,000 passengers an hour in each direction (*8*).

Paris, France

An extensive system of bus lanes, comprising approximately 190 mi, is in operation throughout Paris. The Ligne Pilote routes use an extensive system of bus lanes. The curb lane is used for much of this system, although only a small amount is actually physically separated from the mixed-traffic lanes. In addition, other preferential treatments, such as traffic signal preemption, are provided along these routes (*8,9*).

Porto Alegre, Brazil

Bus lanes are currently in operation in seven radial corridors in Port Alegre. Located in the center of the roadways, a total of 18 mi are in operation. A convoy system is used to maximize the movement of buses on the lanes. Buses are held back at entry points and travel in platoons, halting at stops in unison. This allows volumes of up to 20,000 passengers an hour at speeds of 12 mph. Buses with trailer units are used on some routes to increase further the capacity of the system (*8*).

São Paulo, Brazil

Physically separated and concurrent flow bus lanes are in operation in a number of corridors in São Paulo. The common design is to locate two-lane, two-way busways in the center of arterials or expressways. The platoon or convoy system noted previously is used in many of these facilities to increase their capacity. This has allowed up to 420 buses an hour to travel in some sections (*8*).

Yokohama, Japan

Approximately 15 mi of exclusive bus lanes, 20 mi of priority lanes, and a short segment of a contraflow bus lane are in operation in Yokohama. In addition, priority traffic signals and computer-controlled bus location devices are being used (*8*).

Nonexclusive HOV Facilities on Arterial Streets

A variety of nonexclusive HOV facilities are being used in cities throughout the world. With a few limited exceptions, these facilities are all oriented toward providing priority treatments for buses. Furthermore, many of these applications are in downtown areas, although projects are also found in other congested travel corridors. Nonphysically separated HOV lanes were identified in 75 cities. In many cases, multiple projects are in operation within a metropolitan area. Figure 5 shows the cities in which these types of projects are in operation, including 10 in England. As noted, this figure probably underrepresents the use of this type of treatment, which appears to be commonly used in many British and European communities.

Table 2 lists the cities identified with operating arterial-street HOV facilities, the total length of all projects in miles,

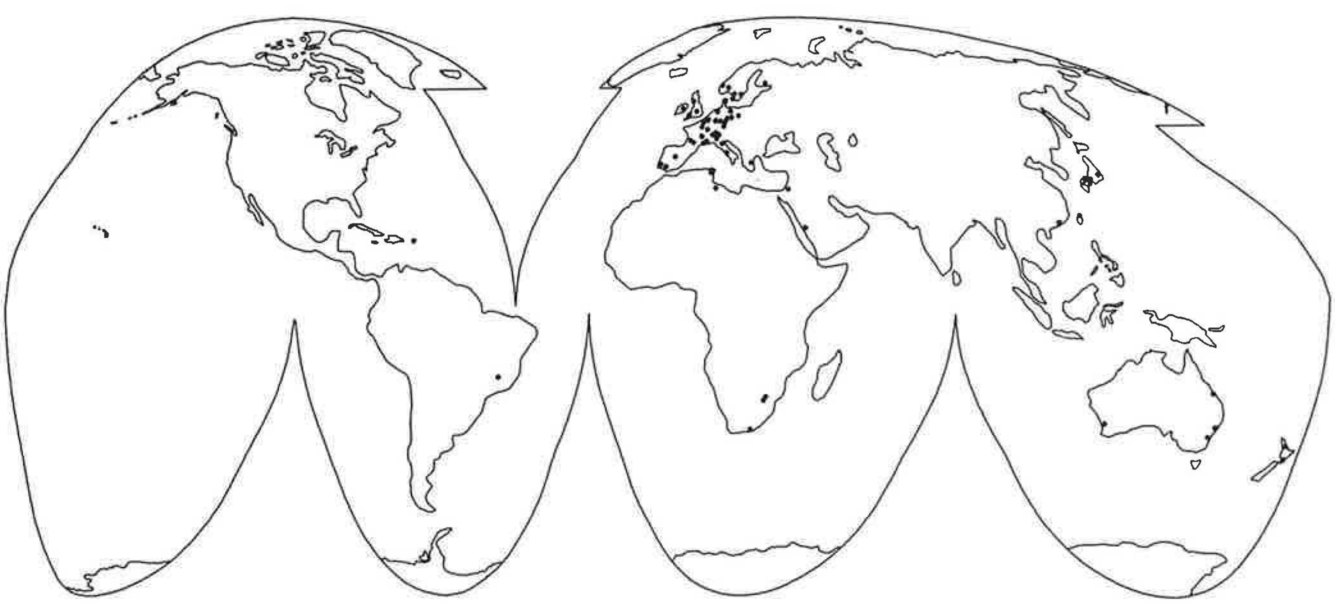

FIGURE 5 Non–North American arterial-street HOV lanes.

TABLE 2 Summary of International HOV Projects—Arterial Street Applications (*8,9*)

Location	Total Length of all Projects (miles)	Other Priority Measures
Australia		
Brisbane	2	X
Canberra	6.5	X
Perth	8.2	X
Sydney	N/A	X
Belgium		
Antwerp	2.0	
Brussels	3.5	
Brazil		
Belo Horizonte	N/A	
Denmark		
Copenhagen	3.2	
England		
Hull	N/A	X
Leeds-Bradford	9	
Leicester	2.2	
London	N/A	
Manchester	N/A	X
Nottingham	N/A	X
Oxford	21	
Reading	.6	
Southampton	1.2	X
Sheffield	2	
Finland		
Helsinki	20	
France		
Bordeaux	6.3	X
Lille	4.4	X
Lyon	31.3	
Marseille	1.5	
Nice	5.6	X
Paris	190	X
Strasbourg	3.5	
Toulouse	4.7	
Germany		
Aachen	3.5	X
Augsburg	1.2	
Berlin	6	
Bochum/Gelsenkirchen	7.5	
Hamburg	N/A	X
Hannover	N/A	
Weisbaden	1	
Wuppertal	3	
Greece		
Athens	1	
Hong Kong		
Kowloon	N/A	X
Ireland		
Dublin	9.1	X
Israel		
Jerusalem	2.4	
Italy		
Bologna	1.4	
Milan	.8	
Rome	1.2	

(*continued on next page*)

TABLE 2 *(continued)*

Location	Total Length of all Projects (miles)	Other Priority Measures
Japan		
Hiroshima	80	X
Kitakyushu	N/A	
Nagoya	6.4	
Okayama	4.8	
Osaka	56	X
Tokyo	145	X
Libya		
Tripoli	11.2	
Luxembourg		
Bouillon	.7	
Netherlands		
Amsterdam	21.5	X
New Zealand		
Auckland	N/A	
Wellington	N/A	X
Norway		
Bergen	5.6	X
Oslo	1.3	
Portugal		
Lisbon	13.4	
Puerto Rico		
San Juan	11.1	X
Saudi Arabia		
Jedda	1.9	
Scotland		
Edinburgh	6.2	
Singapore		
Singapore	43	X
South Africa		
Johannesburg	N/A	X
Port Elizabeth	N/A	X
Pretoria	4.7	
Spain		
Madrid	N/A	
Seville	2.3	
Sweden		
Göteborg	N/A	X
Stockholm	N/A	X
Switzerland		
Bern	0.8	
Lausanne	4.4	
Zurich	7.8	X
Taiwan		
Taipei	N/A	
Tunisia		
Tunis	6.2	
Wales		
Cardiff	1.2	

and an indication if other priority measures are used. The use of traffic signal preemption, giving priority to buses at controlled intersections, is the most common of these measures, but other techniques are also used. Since most of these applications focus on the use of concurrent-flow lanes for buses, the individual projects will not discussed. However, the one exception to this—the Onewa priority scheme in Auckland, New Zealand—is briefly described.

The Onewa priority scheme is the first use of an HOV facility open to buses and carpools in New Zealand. A 0.7-mi HOV lane currently operates in the a.m. peak period on Onewa Road in Auckland. The roadway, which provides access to the Harbour Bridge and Motorway, is severely congested in the morning peak period. Parking was restricted on a short segment of the roadway, and the previous one-lane facility was restriped for two lanes. The curb lane is restricted to buses and carpools, allowing them to bypass the major congestion point. The lane has been in operation since 1982, when it was initiated as a 6-month demonstration. Figures from 1989 indicate that approximately 1,680 vehicles and 4,070 passengers use the lane from 7:00 to 8:30 a.m. Enforcement of the lane occupancy requirements has been an issue, and regular enforcement is provided (W. R. Dragger, personal communication; *16*).

PROPOSED HOV PROJECTS

Two projects in the planning stage are worthy of note, because they both represent the use of HOV lanes on major freeways that would be open to carpools, vanpools, and buses. These projects are similar to many of the applications in operation in North America. One of the planned projects is in the Netherlands; the other is in Taiwan.

Amsterdam, the Netherlands

A tidal-flow carpool lane is in the planning stage on a section of A1 between Amsterdam and 't Gooi/Hengelo in the Netherlands. This facility would be located in the median of the freeway. Reserved right-of-way exists in the median to construct a one-lane facility with shoulders. The occupancy requirement, or demanded seat occupancy per vehicle, is still being determined. Construction on the facility is estimated to begin in 1991, with completion scheduled for 1993 (*17*).

Taipei, Taiwan

A 25-mi HOV lane is currently proposed on the Sun Yat-Sen Freeway in Taipei. This facility would connect Taipei with the CKS International Airport in Taoyuan. The proposal includes concurrent-flow lanes in each direction that would be reserved for carpools and buses. A feasibility study is currently in progress and is expected to be completed by spring 1992 (B. K. Kang, personal communication).

COMPARISONS WITH NORTH AMERICAN HOV FACILITIES

Similarities and differences exist between the international HOV projects examined in this paper and those in operation in North America. The projects identified in this paper were planned and designed to address the same types of problems as those in North America. Furthermore, many projects experience the same types of operating issues and concerns. However, differences also exist. Non–North American applications are more oriented toward bus-only facilities, whereas the use of facilities open to carpools and vanpools, as well as buses, is more prevalent in North America. Moreover, non–North American projects are noted for the use of other priority measures, such as traffic signal preemption, better integration of land use and public transportation, and more extensive use of arterial street HOV applications.

HOV projects in all parts of the world were developed to address common problems related to traffic congestion, environmental concerns, and improving the efficiency of public transportation. In most cases, the non–North American projects focus on providing priority treatment for buses, and those in North America also include carpools and vanpools. This difference may be partially due to the lower levels of automobile ownership in other parts of the world (*18*) For example, many of the HOV applications, especially those in developing countries, provide a basic level of mobility to individuals without other options. However, it is interesting that automobile ownership levels, gasoline consumption, and traffic congestion levels are increasing in many parts of the world (*18–20*). Information on the proposed projects in the Netherlands and Taiwan, as well as some existing projects, specifically note the need to control the increases in traffic congestion in an environmentally sound manner (*16,17*).

The need to coordinate the development and operation of HOV facilities with many agencies and groups was noted in a number of sources (*9,16*). This supports the general experience in North America, where the importance of multiagency coordination and cooperation on successful HOV projects has been noted (*21*). Enforcement of the lane operating requirements was not identified as a major problem with the bus-only facilities, but it was noted as a concern on the Auckland facility. Requests for use by other groups, such as taxis and commercial delivery vehicles, was also noted in Auckland (*16*). Both of these experiences are common with HOV projects in North America.

The introduction of new, expanded, and restructured bus service was noted as often accompanying the introduction of HOV lanes in non–North American cities (*8,9,15,16*). The introduction of services and other service changes often have accompanied many of the HOV projects in the United States and Canada.

The HOV projects in non–North American countries use more extensive priority measures and techniques to support the effectiveness of the facilities than the HOV projects in the United States. The use of traffic signal preemption devices and priority traffic signal timing are the two most common methods. However, the use of a variety of advanced technologies to provide real-time transit information as well as other priority treatments was also noted. The most widespread use of advanced technologies was noted in Europe, Great Britain, and Japan (*8,9,15*). It appears that projects in the United States could benefit from the experience with these types of applications.

The use of HOV facilities to assist in better integrating and coordinating land use and transportation also appears to be more successful outside the United States. The development of the busways as part of Redditch and Runcorn are perhaps the best example of this, but even in other areas the integration of land use and public transportation appears to be more successful than in the United States. However, it is interesting that even in the best examples, many of the same problems currently encountered in the United States are evident. For example, although Runcorn was designed and developed around the busway, parking is free or relatively inexpensive at major employment locations and in the city center. It has been noted that this policy does not encourage use of the busway (*15*).

The widespread use of arterial street HOV lanes is another area in which North American cities could benefit from the experience in other parts of the world. The potential use of arterial street HOV lanes, both as stand-alone projects and as links in a regional HOV system, has generated a great deal of interest recently among transportation professionals. It appears that much could be learned about the design and operation of these types of facilities from the projects in other parts of the world.

CONCLUSION

An overview and description of HOV facilities throughout the world has been provided. A summary of exclusive HOV lanes operating in separate rights-of-way and on freeways, expressways, and major arterials has been presented. In addition, a listing of cities using arterial street HOV lanes was provided, along with a discussion of projects in the planning stage. Finally, a review of the similarities and differences with HOV project in North America was presented.

As noted previously, this paper represents the starting point for an enriched understanding of the application of HOV facilities throughout the world. Obviously, additional information is needed to understand fully the breadth and depth of HOV application outside North America as well as the issues, problems, and experiences associated with their use. It would be beneficial to continue to research the design aspects, operational characteristics, vehicle volumes and passenger levels, land use and transit integration, and potential future projects. It would be appropriate to request additional information through the World Bank, *Jane's Urban Transport Systems*, TRB, and other international transportation organizations.

In addition, transportation professionals and groups in all parts of the world would benefit from greater information sharing and the exchange of experiences and ideas. A variety of methods are appropriate to promote and encourage greater communication on HOV projects. First, the 1992 HOV Facilities Conference, which will be held in Ottawa, provides an excellent opportunity to encourage greater participation from non–North American representatives. The TRB Conference Planning Subcommittee should consider this possibility in developing the program for the 1992 conference. In addition, a session at the TRB Annual Meeting, which is attended by representative from countries throughout the world, may be appropriate. Such a session could be considered for the 1993 Annual Meeting.

The information presented in this paper has provided an initial indication of the scope of HOV applications throughout the world. It has further identified many areas in which additional research and information sharing would benefit transportation professional in all areas. The opportunity exists to build on this effort and to establish an ongoing coordinated approach to facilitate the collection of more detailed information and to share ideas among transportation professionals throughout the world.

ACKNOWLEDGMENTS

To obtain information on HOV facilities in operation in non–North American countries, individual letters were sent to representatives from government agencies, transit operators, consulting firms, and university research groups. Especially helpful in identifying possible contacts were Alan Gonseth, Champagne Associates and ITE International Vice President; Morris Rothenberg, JHK & Associates; Charles Fuhs, Parsons Brinckerhoff Quade & Douglas, Inc.; John Bonsall, Ottawa-Carlton Regional Transit Commission; and Pete Fielding, University of California, Irvine. In addition, Patrick Beck, June Housman, and Robert Viera, TTI, provided assistance with the graphics, tables, and literature review, respectively. The assistance of these individuals is both acknowledged and greatly appreciated.

REFERENCES

1. K. F. Turnbull and J. W. Hanks, Jr. *A Description of High-Occupancy Vehicle Facilities in North America.* DOT-T-92–01. U.S. Department of Transportation, 1990.
2. J. H. Dupree and R. H. Pratt. *A Study of Low-Cost Alternatives to Increase the Effectiveness of Existing Transportation Facilities—Results of Case Studies and Analysis of Busway Applications in the United States.* U.S. Department of Transportation, 1973.
3. R. J. Fisher and H. J. Simkowitz. *Priority Treatment for High-Occupancy Vehicles in the United States: A Review of Recent and Forthcoming Projects.* UMTA-MA-06-0049-78-11. FTA, U.S. Department of Transportation, 1978.
4. *The Effectiveness of High-Occupancy Vehicle Facilities.* ITE, Washington, D.C., 1988.
5. M. J. Rothenberg and D. R. Samdahl. *High-Occupancy Vehicle Facility Development, Operation, and Enforcement.* U.S. Department of Transportation, 1982.
6. C. A. Fuhs. *High-Occupancy Vehicle Facilities: Current Planning, Operation, and Design Practice.* Parsons Brinkerhoff, Inc., New York, N.Y., 1990.
7. C. Bushell. *Jane's Urban Transport Systems 1988.* Jane's Information Group Limited, Surrey, England, 1988.
8. C. Bushell. *Jane's Urban Transport Systems 1990.* Jane's Information Group Limited, Surrey, England, 1990.
9. *Bus Lane and Busway Systems.* Organization for Economic Co-Operation and Development, Paris, France, 1977.
10. *Notes for Students.* Redditch Development Corporation, Redditch, England, 1979.
11. *Progress Report, April 1980.* Redditch Development Corporation, Redditch, England, 1980.
12. M. Harris. The Runcorn Busway. *Modern Transport*, Summer 1978, pp. 186–190.
13. F. J. Osborn and A. Whittick. *New Towns: Their Origins, Achievements and Progress.* Routledge & Kegan Paul, Boston, Mass., 1977.
14. *Runcorn New Town.* Runcorn Development Corporation, Runcorn, England, 1976.
15. *The Busway.* Runcorn Development Corporation, Runcorn, England, 1975.
16. Auckland Regional Council. *Onewa Priority Scheme.* Auckland Regional Commission, New Zealand, 1988.

17. J. Boender. *Measures for Target Groups*. Ministry of Transport and Public Works, the Netherlands, 1991.
18. S. C. Davis and P. S. Hu. *Transportation Energy Data Book*, 11th ed. Oak Ridge National Laboratory, Tenn., 1991.
19. P. W. G. Newman and J. R. Kenworthy. Gasoline Consumption and Cities: A Comparison of U.S.Cities with a Global Survey. *Journal of the American Planning Association*, Winter 1989, pp. 24–37.
20. M. D. Lowe. Rethinking Urban Transport. In *State of the World 1990* (L. R. Brown, ed.). W. W. Norton & Company, New York, N.Y., 1990, pp. 56–73.
21. K. F. Turnbull. *High-Occupancy Vehicle Project Case Studies: History and Institutional Arrangements*. Texas Transportation Institute, College Station, Tex., 1990.

Publication of this paper sponsored by Committee on High-Occupancy Vehicle Systems.

History and Institutional Arrangements of Selected High-Occupancy Vehicle Projects

KATHERINE F. TURNBULL AND DENNIS L. CHRISTIANSEN

High-occupancy vehicle (HOV) facilities are one technique being used in many metropolitan areas to address mobility, traffic congestion, and air quality concerns. The 1990 Clean Air Act amendments, the Intermodal Surface Transportation Efficiency Act, and initiatives at the state and local levels all indicate that HOV facilities will play an even more important role in the future. Providing a better understanding of the characteristics and institutional arrangements that have resulted in the development and operation of successful HOV lane projects can benefit areas initiating the planning and development process, as well as those with projects already in operation. The history and institutional arrangements associated with the development of HOV projects in six case study sites were examined as part of the Assessment of Freeway High-Occupancy Vehicle Lane Projects funded by the Federal Transit Administration through the Texas Department of Transportation. Many similarities were identified in the decision-making and implementation processes; they appear to have assisted in the development of successful HOV projects. Common elements in the decision-making process included significant levels of congestion and projected growth in travel demand, the lack of an adopted fixed-guideway transit plan, planned or scheduled highway improvements, a project champion or champions in positions of authority, and legislative direction and agency policy support. Elements common to the implementation process in the six case study locations included a primary lead agency, interagency cooperation, joint funding, flexibility, adaptability, and support from federal agencies.

High-occupancy vehicle (HOV) facilities are one approach being used in many metropolitan areas to address growing mobility, traffic congestion, and air quality problems. Many areas face significant increases in traffic congestion and projected travel demands beyond what can reasonably be served at current vehicle occupancy rates. Attempting to address these congestion and mobility problems in a time of limited financial resources and right-of-way availability has led many areas to consider pursuing a wide spectrum of potential solutions. The use of HOV facilities, which focus on increasing the person-movement efficiency of the roadway facility, is a practicable alternative in many areas. The recent air quality legislation, initiatives at the state and local level, and the reauthorization of the federal transportation program all point to the continued interest in and development of a variety of HOV projects.

As interest in HOV facilities increases, various issues continue to be addressed related to the most effective use of these types of facilities and to the elements necessary to ensure their successful development, implementation, and operation. To assist in identifying these elements, the Texas Transportation Institute, a part of the Texas A&M University System, conducted an examination of the history and institutional arrangements associated with HOV projects in six case study sites. This analysis was part of a larger study funded by the Federal Transit Administration (FTA) and administered by the Texas Department of Transportation (TxDOT) (*1*). The case study analysis examined HOV facilities in Houston, Texas; Minneapolis, Minnesota; Orange County, California; Pittsburgh, Pennsylvania; Seattle, Washington; and Washington, D.C./Northern Virgina.

The common elements and the unique characteristics leading to the implementation and operation of the HOV facilities in the six case study cities are summarized in this study. They include an examination of the reasons behind the development of the projects, the background and history of the facilities, relevant issues associated with the projects, and the roles and responsibilities of the different agencies and organizations involved in the process. The results of this analysis should be of benefit to areas initiating the planning and development process for HOV facilities, as well as those areas with operating HOV projects. This research advances the current understanding of the characteristics that appear to be significant in the development of HOV projects as well as those characteristics that assist in ensuring the successful implementation and operation of these facilities.

OVERVIEW OF COMMON ELEMENTS

The assessment of the history and institutional arrangements associated with HOV projects in the six case study sites identified a number of common elements. Although these were not present in all case studies to the same degree, the elements occurred often enough to represent common features that appear to be significant in the development of HOV projects. Major similarities among the projects are outlined in the following section. The first elements identify common characteristics that resulted in the decision to implement the HOV facilities; the latter elements relate to similarities during the development of the actual projects. Table 1 presents a summary of the major characteristics common to multiple HOV case study projects.

Common Characteristics of Decision Making

Corridor and Areawide Characteristics

All the case study sites are in major metropolitan areas in the United States. In terms of population, all fall within the coun-

Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843.

TABLE 1 Important Factors in the Development of the Case Study HOV Projects

Features Common to Multiple Projects	Case Study Sites					
	Houston	Minneapolis	Orange County	Pittsburgh	Seattle	Washington, D.C.
Decision Making Process						
Intense Congestion in Corridor	X	X	X	X	X	X
Lack of Agreed upon Fixed-Guideway Transit Plan	X	X	X		X	X[1]
Planned or Scheduled Highway Improvement	X	X	X	X	X	X
Project Champion Within Implementing Agency	X	X	X			
Legislative or Policy Direction		X	X			X
Implementation Process						
Lead Agency in Implementation	X[2]	X	X	X	X	X
Interagency Cooperation	X	X	X	X	X	X
Joint Funding	X	X	X	X	X	X
Support of Federal Agencies, Incl. Funding	X	X	X	X	X	X
Flexibility and Adaptability	X	X	X	X	X	X

[1]In the I-66 corridor, the Washington Metropolitan Transit Authority adopted a plan in 1968 which included a Metro line in the median of I-66 for a portion of the corridor.
[2]The development of the Houston transitways can best be described as multi-agency projects requiring multi-agency decisions.

try's 20 most-populated metropolitan areas. Furthermore, the HOV projects in each site are all in major travel corridors. In all cases, the metropolitan areas and the specific corridors were experiencing significant growth in travel demand at the time the HOV projects began to be considered. In addition, travel demands were projected to increase in all corridors.

The need for major improvements had been identified in all the corridors, and, in many cases, the examination of alternatives and the development of detailed plans had been initiated. HOV facilities became one of the alternatives examined to address the anticipated travel demand and ultimately emerged as a major element of the final recommendation. Thus, an awareness of the need to address increasing traffic congestion problems in a major travel corridor had developed in all of the case studies.

Lack of Fixed-Guideway Transit Plan for Corridor

Another similarity among the case sites was the lack of an agreed-upon or approved long-range fixed-guideway transit plan for the corridor. An approved fixed-guideway transit plan did not exist for most of the case study corridors at the time consideration of an HOV alternative was initiated. In many instances there was disagreement among different agencies over the role transit should play in the corridor and the technology that should be used. In some cases there had been an ongoing debate over this issue.

In addition, in some instances, such as in Seattle, Houston, and Minneapolis, the lack of consensus over the role of transit and the technology to be used applied not just to the corridor, but to the metropolitan area as a whole (*2–4*). In these cases, the debate (which continues today) relates to the implementation of a rail transit component as one element of the overall public transportation system. Thus, in most of the case study sites, no decision had been made on the development of a fixed-guideway transit system in the corridor in which the HOV facility was ultimately developed.

Planned or Scheduled Highway Improvements

Some type of highway improvements was either planned or scheduled in most of the corridors in which the HOV projects were eventually built. These ranged from major new freeways, such as I-394 in Minneapolis, I-66 in Northern Virginia, and I-90 in Seattle, to pavement rehabilitation projects such as Katy (I-10) in Houston and Route 55 in Orange County (*2–6*). Thus, consideration of the HOV project was often initiated as one approach to increasing the person-movement efficiency of the roadway facility.

Once the decision had been made to include the HOV element, steps were initiated to coordinate the planning, design, and construction of both the freeway and HOV elements with maximized available resources and minimized disruptions to the traveling public. Thus, HOV projects in many of the case study sites were considered and implemented as part of larger highway improvements. These ranged from new freeway facilities to pavement rehabilitation projects. This coordination helped maximize available resources and minimize the impacts of implementation on the traveling public.

Project Champions

In most of the case studies one individual, or a small group of individuals, was identified as being instrumental in the development, promotion, and support of the HOV project. These were people, usually within the state transportation department or local transit agency, who had the authority and position to influence the outcome of the process. Their support was often noted as a major reason for the development of the HOV projects in many of the case study areas. These individuals showed a willingness to try new and innovative approaches to dealing with growing traffic congestion problems and to move the projects forward. Many of the projects represented the first use of the different types of HOV facilities in the country, so some risk was associated with their implementation. Thus, individuals in positions of authority in highway and transit agencies supported the HOV project concept and promoted it through the project development and implementation process.

Legislative Direction and Policy Support

In many of the case study sites the consideration of HOV facilities was supported by legislative or policy directives. This took the form of policy directives from the federal level on the I-66 facility in Northern Virginia and the state level on I-394 in Minneapolis (*4,5*). These legislative and policy directives assisted in ensuring that HOV facilities were one of the alternatives considered in the planning process, the directives supported the implementation of the ultimate recommendation. The involvement of Congress and federal agencies in the many aspects of planning, designing, and operating the HOV facilities in the Washington, D.C., area represents a unique feature not found in the other case study sites (*5,7*). Thus, legislative or agency policies and directives played an important role in the decision-making process in some of the HOV case study projects.

Common Characteristics of Implementation

Lead Agency

In general, the agency responsible for making the decision to proceed with the development of the HOV project also had the overall responsibility for implementing the project. In all cases, the state department of transportation or the state highway department was responsible for construction of the actual facility. However, transit agencies were also actively involved in different aspects of many of the case study HOV projects. Thus, while the state department of transportation usually took the lead role, other agencies were actively involved in the process.

The Houston transitways can best be described as multiagency projects requiring multiagency decisions. The Houston Office of Transit, the predecessor agency to the Metropolitan Transit Authority of Harris County (METRO), was the lead agency in the initial contraflow demonstration project. However, on this and subsequent HOV projects, extensive agreements between METRO and TxDOT were used to identify the roles, responsibilities, and financial participation of the two agencies.

Most of the case study projects used some type of project management team or coordinating group. In many cases other agencies also participated in funding some elements of the projects. Thus, one agency, usually the state department of transportation, had overall responsibility for implementing the HOV project. However, transit and other agencies were often involved in some aspects of planning, designing, and, in a limited number of cases, financing the projects.

Interagency Cooperation

All of the HOV projects in the case study sites involved some degree of interagency cooperation. The exact nature and level of this involvement varied substantially between projects. Some type of interagency coordination structure, such as a project management team, was used with many of the HOV projects. These coordinating groups were identified as an important component in ensuring that all groups were adequately involved in the implementation process.

This coordination was noted as especially important because of the unique nature of the HOV projects and the need to involve highway, transit, enforcement, and other groups in the process. In most cases, these committees were active in many aspects of planning, designing, implementing, and operating the facilities. These groups usually involved all the relevant agencies and groups associated with the projects. In several case study sites, the metropolitan planning organization (MPO) was actively involved in the process and openly supportive of the HOV project. Thus, interagency cooperation, including the use of multiagency project management groups, played an important part in the coordinated implementation of most of the case study HOV projects.

Joint Funding

A variety of funding sources were used for many of the HOV projects in the case study sites. Different combinations of funds were obtained from FHWA, FTA, and state and local highway and transit agencies. In addition, many areas, such as Houston and Minneapolis, used a variety of funding approaches and institutional arrangements to develop the HOV projects (*3,4*). Thus, multiple funding sources and innovative financing approaches were used with some of the case study HOV projects.

Support of Federal Agencies

FHWA and FTA were supportive of the HOV projects in the case study sites. This involvement included providing funding for initial demonstration programs, construction of the HOV lanes and supporting elements, and research and evaluation programs; participating in project management teams; providing technical assistance; and providing policy guidance. Thus, support from FHWA and FTA was evident, although in different degrees, in the development of some of the case study HOV facilities.

Flexibility and Adaptability

All the case studies seem to indicate that flexibility and the ability to adapt to change were important elements in both the development and ongoing operation of the HOV facilities. Almost every project has experienced some change in the operating requirements of the HOV facility. These changes have been the result of experience and policy directives. In either case, the need to maintain flexibility to respond to changing travel demands and policies appears to be an important element of the HOV projects in the case study sites.

OVERVIEW OF UNIQUE ELEMENTS

The history and institutional arrangements associated with HOV projects in each of the case study sites identified a few features unique to individual projects.

Congressional Involvement

A unique feature of the two HOV facilities included in the Washington, D.C./Northern Virginia case study, I-66 and the Shirley Highway, is the role that the U.S. Congress and the Secretary of Transportation played in their development and continue to play in their ongoing operation (*5,7*). Other HOV projects, such as I-394 in Minneapolis and I-90 in Seattle, have been influenced by state legislative action, but the I-66 and Shirley Highway HOV facilities are the only projects to draw the specific attention of the federal government in such a significant way (*2,4*). Thus, congressional involvement has been a unique and major feature in the planning, implementation, and operation of the I-66 and Shirley Highway HOV facilities.

Dealing with Past Issues

The development of the Route 55 project in Orange County appears to have been partially influenced by the issues associated with the Santa Monica Diamond Lanes. The approach taken by the California Department of Transportation and other agencies on the Route 55 project appears to have reflected the experience gained from the Santa Monica project. The HOV facility on the San Bernardino Freeway has generally been considered successful since the early 1970s, but the termination of the HOV project on the Santa Monica Freeway in 1976 may have influenced a lack of consideration of HOV projects on other facilities in the Los Angeles–Orange County area (*6*). Thus, the development of HOV projects in Los Angeles and Orange County appear to have been influenced by the termination of the Santa Monica Diamond Lane project.

CONCLUSION

The common elements associated with the development and implementation of HOV facilities in six case study sites have been summarized. The following 10 characteristics, common to all or most of the case studies, appear to be significant in the decision-making and implementation processes.

Decision Making

- Corridor and areawide traffic congestion and growth in travel demand,
- Lack of fixed-guideway transit plan for the corridor,
- Planned or scheduled highway improvements,
- Project champion or champions in positions of authority, and
- Legislative direction and agency policy support.

Implementation

- Lead agency,
- Interagency cooperation,
- Joint funding,
- Flexibility and adaptability, and
- Support of federal agencies.

The analysis of these characteristics presented in this paper should be of benefit to areas considering the development and implementation of HOV projects. Consideration can be given to the extent to which these characteristics are present. Although they do not ensure the success of HOV facilities, these characteristics appear to serve as good benchmarks to assist in measuring the potential for the successful development of HOV facilities. As such, they provide valuable guidelines for the consideration of the potential for successful HOV facilities.

ACKNOWLEDGMENTS

The work described in this paper was undertaken by the Texas Transportation Institute (TTI), a part of the Texas A&M University System. The examination of the history and institutional arrangements associated with HOV projects in six case study cities was one element conducted as part of the larger Assessment of Freeway High-Occupancy Vehicle Lane Projects. This assessment is being funded by FTA through TxDOT.

To ensure accuracy in the documentation of the history and institutional arrangements for the six HOV case study sites, TTI contracted with individuals and organizations knowledgeable with local situations. The following individuals and organizations contributed to this effort: Richard J. Kabat, TTI, Houston, Texas; Charles Fuhs, Orange County, California; Kilareski and Mason, P.C., Pittsburgh, Pennsylvania; G. Scott Rutherford, Seattle, Washington; JHK & Associates, Washington, D.C./Northern Virginia.

The contributions of these individuals and organizations are both acknowledged and appreciated. In addition, the assistance of June Housman, TTI, in the preparation of this paper is acknowledged and greatly appreciated.

REFERENCES

1. K. F. Turnbull. *Draft—High-Occupancy Vehicle Project Case Studies: History and Institutional Arrangements*. Texas Transportation Institute, College Station, Dec., 1990.
2. G. S. Rutherford. History and Institutional Arrangements: HOV Facilities in the Seattle Area. In *Draft—High-Occupancy Vehicle Project Case Studies: History and Institutional Arrangements* (K. F. Turnbull, ed.). Texas Transportation Institute, College Station, Dec., 1990, pp. 225–264.
3. R. J. Kabat. History and Institutional Arrangements: Houston Transitways. In *Draft—High-Occupancy Vehicle Project Case Studies: History and Institutional Arrangements* (K. F. Turnbull, ed.). Texas Transportation Institute, College Station, Dec., 1990, pp. 14–45.
4. K. F. Turnbull. History and Institutional Arrangements: I-394, Minneapolis, Minnesota. In *Draft—High-Occupancy Vehicle Project Case Studies: History and Institutional Arrangements* (K. F. Turnbull, ed.). Texas Transportation Institute, College Station, Dec., 1990, pp. 48–99.

5. JHK & Associates, Inc. History and Institutional Arrangements: I-66. In *Draft—High-Occupancy Vehicle Project Case Studies: History and Institutional Arrangements* (K. F. Turnbull ed.). Texas Transportation Institute, College Station, Dec., 1990, pp. 293–326.
6. C. Fuhs. History and Institutional Arrangements: Orange County, California. In *Draft—High-Occupancy Vehicle Project Case Studies: History and Institutional Arrangements* (K. F. Turnbull, ed.). Texas Transportation Institute, College Station, Dec., 1990, pp. 101–179.
7. JHK & Associates, Inc. History and Institutional Arrangements: Shirley Highway HOV Lanes. In *Draft—High-Occupancy Vehicle Project Case Studies: History and Institutional Arrangements* (K. F. Turnbull ed.). Texas Transportation Institute, College Station, Dec., 1990, pp. 265–291.

The opinions expressed in this paper are those of the authors, not FTA, TxDOT, or the individual case study authors.

Publication of this paper sponsored by Committee on High-Occupancy Vehicle Systems.

Role of High-Occupancy Vehicle Facilities in a Multimodal Transportation System

MELISSA M. LAUBE

High-occupancy vehicle (HOV) treatments are distinguished by their potential to serve demand management objectives while simultaneously increasing the productivity of existing highway facilities in terms of peak-period person-movement. In cities such as Boston, where rapid transit and commuter rail radial lines are used extensively, HOV facilities still may play a significant role in attracting markets that are not served by the rail system. A renewed interest in HOV options has emerged in the Boston area as an outgrowth of planning for the Central Artery–Third Harbor Tunnel. Environmentalists look to expansion of the project's HOV facilities at the regional level as a means of demand management, in the sense of reducing the potential growth in vehicular traffic attendant with a major increase in highway capacity. However, analysis of potential HOV lanes on the radial highways serving Boston indicates that frequently there are trade-offs with respect to mobility and demand management objectives. Moreover, forecasts show that person-movement capacity could be maximized in some cases with non-HOV transportation systems management measures. Right-of-way considerations also are critical in densely developed radial corridors, due to the immediate proximity of homes, commercial development, and, in some cases, environmentally sensitive areas such as wetlands. Because a variety of demand management measures can work for travel to and from the downtown area in Boston, HOV concepts should be evaluated as a component of an integrated regional strategy that optimizes the balance between mobility and demand management objectives.

The Central Artery–Third Harbor Tunnel project in Boston has been designed to incorporate a system of special facilities and preferential treatments for high-occupancy vehicles (HOVs). It is forecast that within the limited confines of the Central Artery/Tunnel (CA/T) project, this system will provide substantial advantages for HOV travel, with direct connections to the downtown street system, Logan Airport, and the rapid transit distribution system serving the downtown area. Because of the short distances covered by the HOV lanes, none of which exceeds 1 mi, travel time savings are forecast to range from only 1.5 to 6.5 min for the various movements provided. However, the system has been designed to connect at the southerly project limit with a much longer HOV line haul facility on the Southeast Expressway, which is being evaluated as a separate project by the Massachusetts Highway Department (MHD). If MHD decides to implement the Southeast Expressway HOV lane, which would be about 8 mi long, travel time savings would be far greater and the connections provided within the project limits would become increasingly important.

CA/T planners recently have analyzed alternative HOV concepts on each of the principal radial highways connecting to the project limits (see Figure 1), for the purpose of evaluating the project's HOV facilities within the broader context of a potential regional system. This analysis showed that exclusive regional HOV lanes would result in substantial travel time savings and increases in HOV use on each of the radial highways analyzed. As a result of implementing a regional system, HOV traffic volume within the project limits would increase to the point at which carpool eligibility requirements would need to be changed to preserve the travel time advantage in the HOV lanes.

The analysis also served to quantify the effects of alternative radial HOV lane configurations in terms of total vehicular volume and person-movement, as well as operating conditions in general-purpose lanes. The results showed that increased

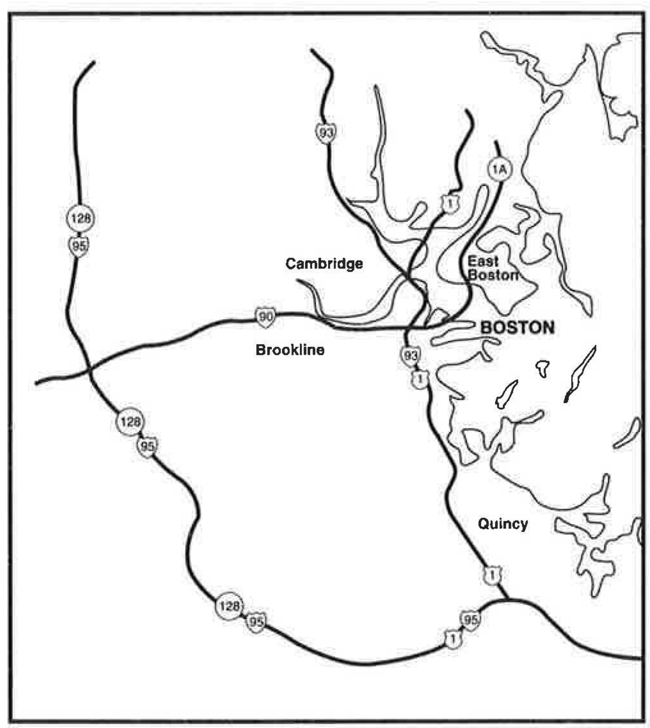

FIGURE 1 Boston metropolitan area highways, 2010.

Parsons Brinckerhoff Quade & Douglas, Inc., One South Station, Boston, Mass. 02110.

HOV use does not necessarily translate into commensurate reductions in the use of low-occupancy vehicles, and that frequently there are trade-offs between improvement in person-movement and demand management. In the Boston area, where radial HOV lanes are being advocated by environmental organizations as a mitigation measure to temper traffic volume growth, these trade-offs are a potential source of controversy with regard to HOV facility planning and design. This paper examines the relationship between the dual objectives of increasing person-movement and reducing vehicular traffic volumes, as it relates to HOV development in an area with practicable multimodal transportation options.

BACKGROUND

The CA/T project will address the need to improve both the capacity and safety of the existing Central Artery and, with a substantial expansion of highway capacity, provide better access to Logan Airport. The existing elevated Central Artery, which is the segment of I-93 that traverses downtown Boston, will be widened and placed in an underground alignment. A new four-lane cross-harbor tunnel on I-90 also will be provided as part of the project, thus doubling existing capacity.

The primary purpose of the HOV system incorporated in the CA/T project is to provide head-of-queue privileges at potential bottlenecks on I-93 and I-90 within the project limits, in a manner that facilitates further expansion of the system outside the boundaries of the project. Although the expansion of general-purpose capacity provided by the project will alleviate the severe traffic congestion experienced today on the Central Artery, there will remain several points of constrained capacity where delays will continue to be experienced in peak hours of travel, specifically at the entrances to the new tunnel on I-90 and at the southerly project limit on the Southeast Expressway (I-93). The CA/T HOV system will allow buses, carpools with two or more occupants, airport limousines, and taxis to bypass these operational bottlenecks, thus providing shorter and more reliable travel times. The system also will provide superior access to downtown Boston, the airport, and South Station, which is a major downtown rail and bus terminal that will serve in the future as a base for airport bus services.

The project's HOV elements will consist of dedicated HOV lanes, priority metering, and managed flow through mixed traffic. Specific facilities include exclusive northbound and southbound lanes on I-93, between the southerly project limit and downtown Boston; an eastbound collector-distributor roadway on I-90, most of which will be for the exclusive use of HOVs; priority metering at the eastbound and westbound I-90 tunnel portals and at the southbound merge with general-purpose traffic at the southerly project limit on I-93; and dedicated ramps between South Station and the HOV lanes (Figure 2). In addition, managed traffic flow downstream of the metering points will ensure uncongested traffic operations through the tunnel and on the roadways connecting to and from Logan Airport.

Each of the projects' HOV elements is forecast to produce measurable travel time savings for HOVs. During the p.m. peak hours in the design year 2010, travel times will be 6.5 min shorter for HOVs than general-purpose traffic between downtown Boston and the new tunnel. Peak-direction traffic volume is forecast to be about 1,000 vehicles for each of the HOV lanes during peak hours; the lanes will each carry between 4,000 and 5,000 peak-hour, peak-direction person trips, which is 2.5 to 4 times the number of people that will be carried in each of the general-purpose lanes. Overall, the system will be relatively small in scale but efficient: increasing person-movement, reducing HOV travel times, and providing the critical connections for HOV distribution in the downtown area.

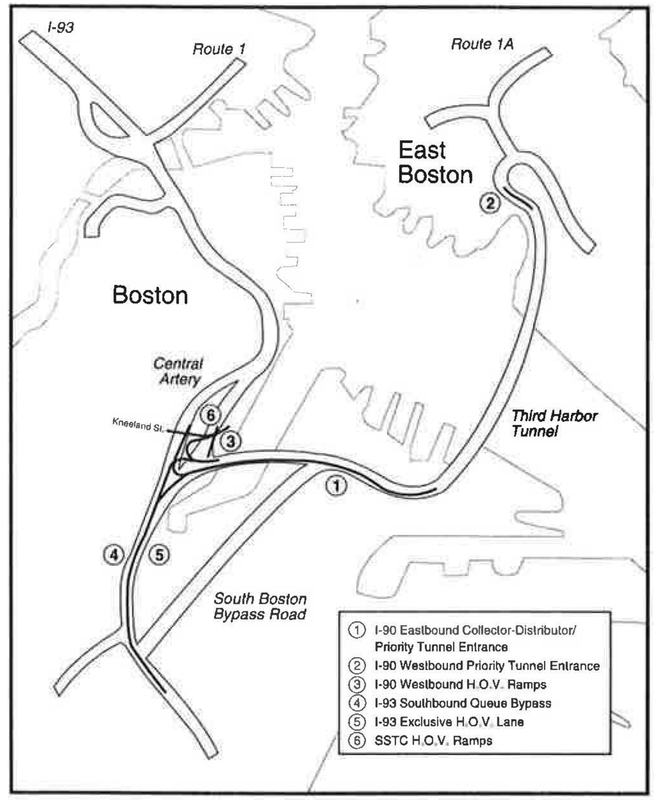

FIGURE 2 Proposed I-90–I-93 HOV system concept.

REGIONAL HOV SYSTEM PLANNING

CA/T Project in Relation to Radial HOV Facilities

From the outset, project planners envisaged a potential connection between the CA/T facilities and a radial HOV lane on the Southeast Expressway, which is the segment of I-93 to the south of the Central Artery. This radial line haul facility would extend about 8 mi, between the southerly CA/T project limit and the circumferential highway, Route 128. The expressway is congested because of high levels of traffic demand relative to available capacity; the right-of-way cannot be expanded appreciably because the highway borders developments and wetlands over much of its length. Because an HOV facility on the expressway would be much longer than the CA/T lanes, and congestion in the general-purpose lanes would be worse on the expressway than within the CA/T project limits (after project completion), potential HOV travel time savings and use of HOVs could be expected to increase substantially with the addition of an expressway HOV lane.

For a long time, the expressway has been regarded as a candidate for HOV facilities. HOV lanes were implemented in two separate trials during the 1970s, with unsuccessful results. A concurrent-flow lane created by borrowing a peak-direction general-purpose lane met with immediate and vociferous public opposition and was summarily terminated. A contraflow lane experiment was cancelled after an MHD employee working on the lane was struck and killed by a moving vehicle. The MHD plans feasibility studies to determine whether design concepts are possible that would avoid the pitfalls and hazards of these previous experiences.

The potential benefits of extending the project's HOV lanes are most apparent in the case of the Southeast Expressway, but planners have also considered the relationship of the CA/T HOV system to the other major radial highways serving downtown Boston: the Massachusetts Turnpike (I-90), which serves the corridor to the west of Boston; I-93 to the north of the Central Artery; and Route 1, which connects to I-93 immediately to the north of the Central Artery. This system-wide analysis approach was requested by FHWA to provide a sound basis for evaluating the HOV facilities planned as part of the CA/T project. The importance of this perspective has become increasingly evident, inasmuch as organizations concerned about the environmental impacts of the project have pressed for the development of extended radial HOV facilities. These facilities have been proposed to mitigate potential increases in vehicular traffic volumes that environmentalists fear will result from the expansion of highway capacity made possible by the CA/T project.

The environmental argument advanced in support of radial HOV lanes in the Boston metropolitan area has focused exclusively on their potential role in demand management. Customarily it is considered good planning and engineering practice to evaluate HOV concepts according to a range of criteria related to both HOV and general-purpose traffic performance (*1*), but the priority for environmentally minded HOV advocates in Boston has been to shift single-occupant automobile drivers to buses and carpools and to give much less consideration to the effects on general-purpose traffic operations. A memorandum of understanding between several of the state transportation agencies and one of Boston's leading environmental advocacy groups exemplifies this orientation. Among other demand management measures, this agreement mandates the implementation of several radial HOV facilities, based solely on their potential benefits for HOVs. In fact, it is stipulated that the new HOV lanes be created on the Massachusetts Turnpike and I-93 north of the city through means other than lane addition. Taking away peak-direction general-purpose lanes implicitly is acceptable within the terms of this agreement, even though this practice may be detrimental to overall traffic operations and proved to be a failure when previously attempted in Boston.

Evaluation of Radial HOV Options

A range of potential HOV concepts has been evaluated to address regional-level issues that could affect the HOV system within the project limits, as well as broader demand management objectives associated with environmental mitigation. In general terms, the concepts tested consisted of the following: concurrent-flow (i.e., same direction) lanes, created by taking away a lane from peak-direction general-purpose traffic; barrier-separated contraflow lanes, created by taking away a lane from off-peak or reverse-direction traffic, assumed to be safe for carpools and vanpools as well as buses; and reversible HOV lanes, created by adding one or two HOV lanes to the existing general-purpose lanes. The analysis was based on year 2010 forecasts of HOV traffic volumes or demand as well as the impacts of alternative concepts on operating speeds in both HOV and general-purpose travel lanes.

These forecasts were generated by the Central Transportation Planning Staff (CTPS) which serves the Boston Metropolitan Planning Organization. The method used to develop the forecasts involved adjustments of existing model networks to incorporate HOV alternatives, assignment of traffic using the Urban Transportation Planning System (UTPS) and pivot-point estimation techniques to relate computer-forecasted travel time differences among the alternatives to vehicle occupancy rates. The results are summarized in Tables 1 and 2. The pivot-point analysis provided estimates of vehicle occupancy by traffic zone, based on empirical relationships between variations in travel time savings and average vehicle occupancy. This methodology was based on the earlier Bolling-Anacostia forecasting model and a model developed by the Metropolitan Washington Council of Governments. A noteworthy characteristic of the forecasting procedure is that the total volume of network vehicle trips is held constant across the alternatives under study. Variations in vehicle occupancy rates correspond to changes in the total number of person trips.

Among the four radial highways studied, potential HOV benefits would be greatest on the Southeast Expressway—which is to be expected, given the magnitude of congestion forecast to occur on that highway. Compared to a year 2010 baseline condition in which no HOV lanes are provided, HOV travel time savings would be about 16 min for all of the expressway concepts analyzed. In comparison, HOV travel time savings on the other radial highways would range from 5 to 9 min. Substantial increases in HOV use would result on all the highways, most particularly on the Southeast Expressway, on which approximately 1,500 HOVs and 5,200 HOV person-trips would be added during the 3-hr a.m. peak period, with HOV eligibility defined as 3 or more occupants per vehicle. Analysis of the impacts on HOV usage within the project limits shows that the eligibility requirement would need to be raised to 3-or-more occupancy, if a radial HOV lane were provided on the Southeast Expressway, because HOV volumes on the CA/T HOV lanes would rise dramatically. With a 3-or-more occupancy eligibility requirement, the peak-hour volume on I-93 north within the project limits would slightly exceed the standard maximum threshold of 1,500 vehicles.

The forecast data provide some revealing comparisons of the effects of different HOV concepts. As might be expected, the concurrent-flow take-away lane options would result in the degradation of traffic operations in the general-purpose lanes, with HOV eligibility defined as 3 or more occupants per vehicle. General-purpose speeds on the Southeast Expressway would decline by an average of 3 mph during the 3-hr a.m. peak period, resulting in an average increase in travel time of 5 min, or 23 percent, over the length of the

TABLE 1 Impacts of HOV Options on 3-hr A.M. Period Traffic Operations, Year 2010 (3+ Occupancy HOV Eligibility)

ROADWAY	INBOUND				OUTBOUND	
	SPEED (MPH)		TRAVEL TIME (MIN)		GENERAL PURPOSE	
	HOV	MIXED FLOW	HOV	MIXED FLOW	SPEED (MPH)	TRAVEL TIME (MIN)
SOUTHEAST EXPRESSWAY						
Baseline	19	19	25	25	52	9
Concurrent	54	16	9	30	53	8
Contraflow	56	21	8	22	45	10
Added Reversible HOV Lane	55	21	9	22	55	8
Reversible Mixed Flow	30	30	16	16	46	10
MASSACHUSETTS TURNPIKE						
Baseline	27	27	23	23	54	12
Concurrent	49	22	13	29	54	12
Contraflow	44	28	14	22	39	16
Added HOV Lane	44	28	14	22	54	12
I–93						
Baseline	31	31	16	16	51	10
Concurrent	57	28	8	18	50	10
Contraflow	57	33	9	15	34	15
Added HOV Lane	57	33	9	15	51	10
ROUTE 1						
Baseline	29	29	19	19	49	11
Concurrent	39	26	15	22	49	11
Contraflow	49	31	12	18	49	12
Added HOV Lane	35	30	16	19	49	11

TABLE 2 Impacts of HOV Options on 3-hr A.M. Peak Period, Peak Direction Traffic Volumes, Year 2010 (3+ Occupant HOV Eligibility)

ROADWAY	VEHICLE TRIPS		PERSON TRIPS	VOR (1)
	HOV	TOTAL		
SOUTHEAST EXPRESSWAY				
Baseline	1360	25340	31870	1.26
Concurrent Flow	2910	22350	32180	1.44
Contraflow	2760	27910	38110	1.37
Added Reversible HOV Lane	2750	26920	36960	1.37
Reversible Mixed Flow Lanes	1920	32370	41150	1.27
MASSACHUSETTS TURNPIKE				
Baseline	1150	18520	23670	1.28
Concurrent Flow	1790	17060	23540	1.38
Contraflow	1520	18910	24990	1.32
Added HOV Lane	1520	18900	24980	1.32
I–93				
Baseline	1490	19920	26060	1.31
Concurrent Flow	2600	14690	22790	1.55
Contraflow	2090	19920	27490	1.38
Added HOV Lane	2080	19890	27430	1.38
ROUTE 1 (2)				
Baseline	1920	11900	14400	1.21
Concurrent Flow	2640	12140	15580	1.28
Contraflow	2960	12540	16400	1.31
Added HOV Lane	2290	11910	14890	1.32

(1) VOR= VEHICLE OCCUPANCY RATE
(2) 2+ OCCUPANT HOV ELIGIBILITY ON ROUTE 1

expressway. In contrast, the contraflow and reversible-added concepts would result in a 2-mph improvement in general-purpose traffic speeds, due to the reduction in per-lane volumes associated with the addition of a peak-direction lane, which would translate into a 3-min decrease in travel time.

The impact of the take-away lane versus the contraflow and added-lane concepts is even more striking in terms of person-movement, which is a critical—if not overriding—measure of overall performance. On the Southeast Expressway, the concurrent-flow take-away lane is forecast to produce a negligible increase in the total number of person trips carried over the entire 3-hr a.m. peak period. In contrast, the contraflow and added-reversible lanes would result in a net peak-period increase of 5,000 to 6,000 person trips. On I-93 and the Massachusetts Turnpike, the take-away concurrent-flow lane concepts are forecast to produce losses in peak-period person-movement, whereas the contraflow and reversible-added concepts would add 1,300 to 1,400 person trips. On the basis of the differences between the options in terms of this indicator, there is little justification for concurrent-flow take-away lane HOV concepts.

Because person-movement is so critical, it is noteworthy that the number of peak-period person trips could be maximized on the Southeast Expressway with a non-HOV transportation systems management (TSM) alternative. The TSM concept consists of the conversion of one northbound and one southbound general-purpose lane into two reversible lanes that would be open to all traffic. The net effect would be to add one mixed-flow lane to the peak direction in both the morning and evening peak periods, yielding a total of five peak-direction lanes. The impact of this concept, in terms of person-movement, was forecast to be an increase of 9,000 peak-period person trips.

Another important benefit of the non-HOV reversible-lane concept would be its effect on operating speeds. All of the concepts that involve expansion of peak-direction capacity (i.e., contraflow HOV, reversible-added HOV, and reversible-added mixed-traffic) would result in some improvement in operating speeds not only for HOVs but for general-purpose traffic as well. All the HOV concepts would provide for HOV travel at about the legal limit of 55 mph. The HOV concepts created by adding peak-direction lanes would also result in a small improvement in general-purpose operating speeds, as described previously. In contrast, the non-HOV reversible-lane concept would increase speeds in all peak-direction lanes to 30 mph, compared with the forecast baseline speed of 19 mph. However, an HOV lane may be superior to a mixed-traffic reversible concept if mixed-flow volumes increase substantially in relation to available capacity. In that case, an HOV lane would provide for greater potential person-movement.

Of course, person-movement and speed are not the sole criteria by which transportation measures are judged. HOV priority is distinguished by its potential to serve the dual objectives of increasing person-movement on highway facilities and reducing the use of low-occupancy vehicles. In terms of demand management, the relative benefits of HOV treatments are illustrated by the forecasts of vehicle occupancy rates (VORs) and vehicular traffic volumes for the various alternatives analyzed. On the Southeast Expressway, the 2010 baseline VOR is forecast to be 1.25 occupants per vehicle. The 3-or-more occupant concurrent-flow take-away lane HOV concept would result in the greatest increase in VOR, to 1.44. As described previously, this increase in vehicle occupancy would be accomplished at substantial cost: degradation of traffic conditions in the already-congested general-purpose lanes and negligible improvement in person-movement. The substantial improvement in VOR is attributable in part to a decrease in the total number of vehicles that would be carried on the highway, reflecting a reduction of about 4,500 low-occupancy vehicle trips during the a.m. peak period without a commensurate increase in HOV trips.

The other HOV concepts on the expressway would result in VOR of approximately 1.37, which represents an increase of about 10 percent over the baseline condition. In contrast, the VOR for the non-HOV reversible-lane concept would be only 1.27, which is essentially equal to the baseline VOR. Overall, the forecast data indicate that the contraflow and reversible-added HOV lanes would strike a balance between the objectives of reducing the volume of vehicular traffic and increasing the number of person trips accommodated during peak periods.

Nevertheless, the analysis results suggest some further questions about the net impact of the various alternatives with respect to areawide, as opposed to facility-specific, transportation objectives. The contraflow and added-lane HOV concepts are forecast to balance mobility and demand management objectives on an individual highway, but an optimal balance for the area served by the highway could be achieved through alternative strategies that involve different modes, such as rail transit.

HOVs and Demand Management

In particular, the forecasts of impacts on the distribution of vehicular traffic volumes indicate the need for a broader analysis of HOV facilities in terms of demand management objectives. It is reasonable to expect that an HOV lane and a parallel rail line might compete for users and that some new HOV users would be diverted from rail transit rather than from single-occupancy vehicles. The forecast data illustrate the potential magnitude of such impacts. The contraflow and added-HOV lane concepts would produce a substantial net increase in the number of low-occupancy vehicles carried on the Southeast Expressway during the a.m. peak period. Although much smaller in magnitude, effects of the same type are forecast for the turnpike and Route 1.

Because the forecasting models used in this analysis do not indicate the modes from which new HOV person trips would be drawn, the results do not provide conclusive evidence of the extent to which HOV users are diverted from rail transit. Even when an increase in vehicular traffic volumes is not forecast in association with HOV facilities, as on I-93, the likelihood is that in the Boston metropolitan area at least some of the increase in HOV use would be at the expense of rail transit ridership. However, such cases probably would mean not a net increase in vehicular travel on an areawide or regional basis, but instead a neutral shift from one high-occupancy mode to another. The increase in total vehicle trips on the Southeast Expressway, in particular, and to a lesser extent on the turnpike and Route 1, raises the prospect that a net increase in the use of automobiles could occur on an areawide basis. Not only total vehicular traffic volume but also the number of low-occupancy vehicle trips are forecast to increase on the expressway if contraflow or reversible-added HOV lanes are implemented.

The implications of the rail diversion factor can be illustrated in terms of VOR. The 1.37 VOR forecast to result from the contraflow HOV lane would decline to 1.29 if just one-third of the forecast increase in peak-period person trips on the expressway is discounted as diversion from rail transit. This magnitude of diversion is plausible and possibly conservative, considering that a rapid transit line parallels the highway and that a radial commuter rail line is scheduled to be in operation by the middle of this decade in the same corridor to serve longer trips. The lower VOR is an indicator of a substantial reduction in benefit with respect to demand management. If half of the increase in person trips results from diversion away from the rail system, and therefore is discounted, the VOR would decline to the baseline level of 1.25. In this case, there may be no net benefit in terms of demand management.

Even if demand management benefits turn out to be less than initially supposed, however, the contraflow and added-lane HOV concepts would still result in a substantial improvement in person-movement. Improved productivity, coupled with some reduction in vehicular travel and improvement in traffic operations, still makes a strong argument in favor of HOV facilities, assuming that they are feasible to design and operate. In the densely developed Boston metropolitan area, however, HOV options tend to be constrained by lack of right-of-way. This is particularly true of the Southeast Expressway. A reversible-added lane may require widening of the roadway by about 20 ft, which probably is not feasible because of the proximity of residential and commercial development in some locations and the likelihood of impacts to wetland areas elsewhere. Construction of the HOV lane in an elevated viaduct is a possible alternative that would entail visual and noise impacts as well as substantial cost. Contraflow is yet another possibility, but highway widening would be required and operating costs could be excessive. Moreover, an analysis conducted by CTPS indicated that reverse-direction general-purpose traffic flow would be impaired by the loss of a travel lane, taking into account projections of future traffic growth by the year 2010.

The physical constraints associated with retrofitting an urban highway with HOV lanes can be significant. For this reason, options that can be accommodated within the existing right-of-way deserve serious consideration. It may be possible to implement the non-HOV reversible lane concept without substantial right-of-way expansion on the Southeast Expressway. This option clearly would not serve demand management objectives—it might actually encourage increased use of automobiles—but its benefits in terms of improved highway operations and person-movement could outweigh its liabilities. To the extent that demand management is motivated by constraints on new highway construction, increasing the productivity of existing highway facilities may effectively serve the same purpose, even if traffic volumes increase.

The goal of environmental protection definitely would not be addressed by increasing the use of automobiles. Nevertheless, because vehicular exhaust emissions are reduced when traffic congestion is alleviated, some increase in traffic volume may be tolerable if traffic operating conditions are improved substantially. They apparently would be improved if the Southeast Expressway, and possibly other highways in Boston and other metropolitan areas, were retrofitted with reversible mixed-flow lanes. This concept appears to warrant consideration in conjunction with HOV planning in some areas.

The areas in which this concept might be most appropriate include Boston and other cities where the objective of reducing vehicular traffic may be addressed through various means, including increased use of bus and rail public transportation. One of the principal incentives to use public transportation downtown is the scarcity and high price of parking, which is reinforced in Boston through public policy, as manifested in controls on the off-street parking supply. Further improvements in the existing public transportation system, such as reducing headways and increasing the supply of commuter parking, could temper growth in vehicular traffic to the core of the metropolitan area. Congestion-pricing and increased gasoline and vehicle taxes are highway-oriented traffic management measures that could be disincentives to driving. All of these measures could be implemented as part of a regional strategy that may or may not include HOV facilities. In many urban areas that cannot be served adequately by rail transit, however, HOV lanes are likely to remain among the more powerful demand management tools.

CONCLUSION

The development of HOV forecasting models, although still in a preliminary stage, has provided the ability to predict a broad range of HOV facility operating characteristics, including demand, impacts on speeds for both HOVs and general-purpose traffic, and changes in person trips as well as vehicular traffic volumes. In the Boston metropolitan area, this forecasting ability has been used to evaluate potential HOV facilities, including those planned from a regionwide systems perspective as part of the CA/T project.

The results of this analysis show that future HOV demand is strong on the radial highways serving Boston and that the function of the project's HOV connections would become increasingly important in the context of a regionwide system.

The analysis also serves to underscore the importance of a multimodal planning framework in cities such as Boston where rail service is frequently an alternative to HOV facilities. The forecasts indicate that total peak-period vehicle trips would increase on some highways as a result of implementing HOV facilities through lane addition. This raises the issue of how many of these trips would be drawn from alternative highway routes versus parallel rail lines. Another result that appears to merit further investigation is the superior performance of a non-HOV reversible-lane concept, in terms of its impacts on operating speeds in all lanes as well as total person-movement.

Some of the HOV concepts studied did appear to combine successfully the benefits of increased automobile-occupancy and person-movement and therefore may prove to be optimal

Examination of Indicators of Congestion Level

SHAWN M. TURNER

This research report examines the relationships between many possible indicators of congestion and estimated congestion level in an attempt to identify and validate indicators for areawide congestion measurement purposes. The study estimates the congestion level for 50 large and medium-sized U.S. urban areas with 3 congestion measures. The estimated congestion level for each urban area is then graphically and statistically compared to indicators composed of travel, facility supply, and demographic characteristics to determine close relationships. Two of the congestion measures used in this study, the roadway congestion index and the congestion severity index, produced results with a high correlation to each other. An index-incorporating congestion duration on freeway segments (lane-mile duration index) suffered from unobtainable data for several urban areas and was consequently less comparable between the other two measures. It was determined that two indicators related to travel intensity—daily vehicle-miles of travel per lane-mile and daily vehicle miles of travel per square mile—had the closest correlation to estimated congestion levels. Average annual daily freeway traffic per hourly capacity, previously used as a facility measure, was identified also as having a close relationship to areawide congestion level.

Over the past decade, traffic congestion in urban and suburban areas has grown from a mere annoyance to a severe problem. Although traffic congestion is not a new problem for residents of the central city, it has spread and intensified to envelop the urban fringe and outlying suburban areas. This rapid increase in traffic congestion has become a major concern of transportation professionals nationwide. Current predictions about congestion offer no relief, either. By the year 2005, freeway delay has been projected to increase from between 300 to 500 percent over the 1985 levels (*1*).

There are several factors that have contributed to the rapid growth of traffic congestion in the United States in the past decade. The number of registered vehicles has increased disproportionately to population and household growth; in turn, vehicle travel had spiraled to more than 2 trillion vehicle-mi by the late 1980s. To compound the increase in travel, construction of new highway facilities has slowed considerably since the near completion of the Interstate system in the early 1970s. Because of increased access to the automobile and the suburban migration of both business and residential properties, a higher percentage of commuters now drive instead of using public transit or walking. This change in commuting patterns in combination with the preceding factors has clogged local street networks and highway facilities.

Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843.

CONGESTION MEASURES

The type of measure used to quantify the level of congestion on a transportation system should deliver comparable results for various systems with similar congestion levels. These measures should accurately reflect the quality of service for any type of system, whether it be a single facility or an entire urban area. A congestion measure should also be simple, well-defined, and easily understood and interpreted among various users and audiences.

Existing congestion measures use assorted variables in equation formats to describe the extent, severity, and duration of congestion. One type of measure uses indicators—or variables closely related to the level of congestion—to quantify congestion. Examples of possible indicators include travel (e.g., vehicle miles of travel), roadway supply (e.g., lane miles of roadway), and population density. Indicators are generally related to the probable causes of congestion. Another type of measure uses variables that are descriptors of the effects of congestion. Vehicle delay, congestion duration, and average travel speed are all examples of variables that characterize the effects of congestion.

This report examines the relationships between many possible indicators and congestion levels in an attempt to identify and validate indicators for congestion measurement purposes. The indicators used in this study include travel characteristics, facility supply characteristics, demographic characteristics, and all combinations thereof.

BACKGROUND

There have been several efforts in recent years to improve the analysis of traffic congestion data. Many of the efforts have concentrated on developing an accurate areawide measure of congestion. However, there has been no clear consensus on which indicator, if any among those currently used, most directly reflects congestion level in urban areas. Additionally, many possible indicators have not been fully examined with regard to their relationship to congestion. The following paragraphs discuss previous research on indicators and measures of congestion.

The level-of-service (LOS) concept as adopted by the 1985 *Highway Capacity Manual* represents a range of operating conditions (*2*). The LOS, or quality of service, of a facility is determined by traffic characteristics such as vehicle density and volume-to-capacity (*v/c*) ratio, depending on the facility type. Most congested traffic conditions fall into the LOS F range, a range "used to define forced or breakdown flow."

solutions. However, the relative merits of HOV priority versus other means of demand management remains an important issue for cities such as Boston, in which a range of public transportation services, including rail transit, have traditionally served as alternatives to driving, particularly for travel to and from downtown. The results of the analysis further suggest that it may be possible to achieve substantial improvements in the operations of at least some existing radial highways, either through implementation of HOV priority or other traffic management measures. The resolution of the issues raised in this paper and the development of effective demand-management strategies can best be accomplished with a systemwide approach that incorporates both highways and other practicable transportation modes.

ACKNOWLEDGMENTS

The author gratefully acknowledges the work of the Central Transportation Planning Staff, and in particular Karl Quackenbush, Florence Ngai, and Ramana Chinnakotla, who developed the forecasts on which this paper is based.

REFERENCE

1. C. A. Fuhs. *High-Occupancy Vehicle Facilities: A Planning, Design, and Operation Manual.* Parsons Brinckerhoff Quade & Douglas, Boston, Mass., Dec. 1990.

Publication of this paper sponsored by Committee on High-Occupancy Vehicle Systems.

It is in this LOS range that the demand of traffic exceeds the capacity of the roadway. Because the *v/c* ratio theoretically cannot exceed 1.0 (volume cannot exceed capacity), past a certain level of demand this ratio and the LOS concept is of little use in distinguishing between levels of congested flow in LOS F. The vehicular density during forced flow conditions is greater than the LOS F threshold density; consequently, the *v/c* ratio decreases as the vehicular density increases in LOS F conditions. Although dense travel corridors now experience many hours of the daily rush hour, the *v/c* ratio has traditionally been used to describe a single peak-hour condition.

In an attempt to better describe the duration dimension of congestion, the California Department of Transportation (Caltrans) uses the number of hours of LOF F service (*3*). For example, LOS F2 represents 2 hr of LOS F service. This combination of the *v/c* ratio and the duration of congested operation enhances the LOS concept and accounts for the "peak spreading" common in many urban areas, which extends the peak period over several hours. This improved measure is relatively easy to calculate, interpret, and communicate, but has been used only in planning analyses by Caltrans.

An analysis technique developed by Lindley used Highway Performance Monitoring System (HPMS) data, traffic distribution patterns, and *Highway Capacity Manual* calculations to determine freeway travel delay (*1*). A congestion severity index was defined as the total freeway delay (vehicle hours) per million vehicle miles of travel (VMT). Urban area freeway systems then were ranked according to the congestion severity index. A methodology was developed to include the delay caused by incidents using an accident data base of breakdown types and rates. Delay on principal arterial streets was not included in this analysis. In Lindley's calculations, the congestion threshold was defined at a *v/c* ratio of 0.77 or higher (LOS D or worse) during 1 or more hr/day. This definition of the beginning of congestion is consistent with values reported to Congress on the status of national highways by the Department of Transportation and values recommended for urban freeway design standards by AASHTO (*4,5*).

Early research by Lomax and Christiansen investigated the use of several variables as indicators of areawide congestion level (*6*). Among those presented as possible indicators were traffic volume per lane, percentage of congested (ADT greater than 15,000) freeway lane miles, *K*-factor, and peak-hour travel distance. Trends in these possible indicators were calculated for 1975 to 1980 for five urban areas in Texas. The study concluded that VMT per lane was perhaps the most reliable indicator and developed a congestion standard that combined weighted values for freeway and principal arterial street VMT per lane.

Subsequent research by Lomax et al. resulted in the development of a roadway congestion index (RCI) (*7–11*). The indicator of daily VMT (DVMT) per lane mile for both freeways and principal arterial streets is weighted and normalized in the index's equation. Major U.S. urban areas then are ranked according to the RCI value. The threshold of congestion was chosen at a *v/c* ratio of 0.77 or higher (LOS D or worse) and was correlated to ADT per lane values for freeways and principal arterial streets through basic assumptions about traffic characteristics.

In a recent report by Cottrell, a lane-mile duration index ($LMDI_F$) is presented as a measure of recurring freeway congestion in urbanized areas (*12*). The $LMDI_F$ represents a summation for an urban area of the congested freeway lane miles multiplied by the respective duration of LOS F service. The use of the duration of LOS F service in intermediate calculations of $LMDI_F$ is similar to Caltrans' reporting of LOS F1, F2, and so on. Traffic distribution patterns in the HPMS Technical Manual were used to relate the value of average annual daily traffic volume per hourly capacity (AADT/C) to congestion duration (*13*). Previous research by Lisco determined that peak-period delay occurred when the AADT/C value reached 8 to 10 (*14*). In his analysis, Cottrell chose LOS F as the congestion threshold, which he correlated to an AADT/C value of 9. The analysis excluded arterial streets and did not consider the effects of incident delay.

As evidenced by the foregoing, no clear consensus exists on where congestion begins or which indicator or measure most accurately reflects congestion level on an areawide basis.

METHODOLOGY

This study attempted to identify indicators that can be used for congestion measurement purposes. For the purposes of this report, an indicator is considered to be a variable directly related to the level of congestion. Three previously developed congestion measures were used to estimate the congestion level in 50 large and medium-sized U.S. urban areas. Possible indicators of congestion were chosen, and the relationships between the indicators and congestion level were examined for 1989. The study methodology is described in detail in the following sections.

Use of Existing TTI Data Base

Researchers at the Texas Transportation Institute (TTI) have compiled congestion-related statistics in an extensive data base. The data base currently contains annual summary statistics from 1982 through 1989 for 50 large and medium-sized U.S. urban areas with populations generally greater than 500,000. These urban areas may be found in Figure 1.

The data base statistics of interest to this study were those relating to travel, facility supply, and demographic charac-

Albuquerque NM	Atlanta GA	Austin TX	Baltimore MD
Boston MA	Charlotte NC	Chicago IL	Cincinnati OH
Cleveland OH	Columbus OH	Corpus Christi TX	Dallas TX
Denver CO	Detroit MI	El Paso TX	Fort Worth TX
Ft. Lauderdale FL	Hartford CT	Honolulu HI	Houston TX
Indianapolis IN	Jacksonville FL	Kansas City MO	Los Angeles CA
Louisville KY	Memphis TN	Miami FL	Milwaukee WI
Minn-St. Paul MN	Nashville TN	New Orleans LA	New York NY
Norfolk VA	Oklahoma City OK	Orlando FL	Philadelphia PA
Phoenix AZ	Pittsburgh PA	Portland OR	Sacramento CA
Salt Lake City UT	San Antonio CA	San Bern-Riv CA	San Diego CA
San Fran-Oak CA	San Jose CA	Seattle-Everett WA	St. Louis MO
Tampa FL	Washington DC		

FIGURE 1 Study cities—50 large and medium-sized U.S. urban areas.

teristics. The possible indicators examined were composed of one, or a combination of two, of these three basic characteristics related to congestion. Values for the indicators and most other data needed to calculate congestion measures were extracted from the existing TTI data base. It did become necessary, however, to extract data necessary for calculation of $LMDI_F$ values from the HPMS data base (*15*).

Estimation of Level of Congestion

To examine the relationships between the various possible indicators and congestion levels, the relative congestion levels for each urban area had to be calculated. This was done by choosing several congestion measures currently in use. The congestion levels calculated from these measures were then compared to ensure similar results among the measures. The choice of several comparable measures also prevented bias toward any particular indicator. The congestion measures were chosen with consideration given to previous results, data availability, and ease of interpretation. Each of the three measures chosen for this study are described in the following paragraphs.

Roadway Congestion Index

The RCI was initially developed by Lomax and others at TTI to study mobility trends in major Texas cities. The RCI analysis was gradually expanded to include 50 urban areas throughout the United States. Urban areas in this analysis are consistent with the boundaries as defined by the Bureau of the Census. The major source of data for the calculation of the RCI comes from the HPMS data base. This data base is supplemented with information collected from local metropolitan planning organizations (MPOs), state departments of transportation (DOTs), cities, counties, and other local or regional agencies for each area.

In calculation of the RCI it is assumed that delay, and consequently congestion, begins to occur at LOS D, corresponding to a *v/c* ratio of 0.77 (*2*). This was determined to be equivalent to approximately 15,000 vehicles per lane per day on freeways, and 5,750 vehicles per lane per day on principal arterial streets. On an areawide basis where averages over many facilities may be misleading, it was determined that lower values were more appropriate. The values of 13,000 vehicles per lane per day for freeways and 5,000 vehicles per lane per day for principal arterial streets were used then on an area basis for the congestion threshold. In the RCI equation, DVMT per lane mile for freeways and principal arterial streets is weighted by the respective amount of DVMT for each urban area. The congestion levels are then normalized (using 13,000 for freeways and 5,000 for principal arterial streets) with an RCI greater than 1.0 representing undesirable areawide congestion. The RCI value for each urban area is calculated with the following equation:

$$\mathrm{RCI} = \frac{\begin{array}{c}[(\text{freeway DVMT/lane-mi}) \\ \times \text{ freeway DVMT}] \\ + [(\text{prin. art. DVMT/lane-mi}) \\ \times \text{ prin. art. DVMT}]\end{array}}{\begin{array}{c}(13{,}000 \times \text{freeway DVMT}) \\ + (5{,}000 \times \text{prin. art. DVMT})\end{array}} \qquad (1)$$

Congestion Severity Index

The congestion severity index (CSI) was originally developed by Lindley as a measure of freeway delay per million vehicle miles of travel (*1*). The measure was modified for this study to include principal arterial street delay, because it was thought that this functional class makes substantial contributions to areawide congestion levels. Delay for both freeways and principal arterial streets was calculated using procedures developed by Hanks and Lomax in the "Roadway Congestion" series (*9–11*). In combining the delay for the two different functional classes, it was believed that delay on freeways and principal arterial streets was roughly equivalent; consequently, the delay values were not weighted with respect to functional class. The CSI value for each urban area is calculated with the following equation:

$$\mathrm{CSI} = \frac{\text{total freeway delay (veh-hr)}}{\text{freeway VMT (million)}} + \frac{\text{total prin. art. delay (veh-hr)}}{\text{prin. art. VMT (million)}} \qquad (2)$$

Lane-Mile Duration Index

The lane-mile duration index, $LMDI_F$, was recently developed by Cottrell as a measure of recurring freeway congestion in urban areas. The analysis technique used the HPMS data base to calculate an AADT/C for urban area freeway segment. This AADT/C value was then related to a congested percentage of ADT by using traffic distribution patterns in the HPMS Technical Manual (*13*). The congestion duration is the product of the AADT/C value and the congested percentage of ADT. The $LMDI_F$ for each urban area, then, is the summation of the product of congested lane miles and congestion duration for all area freeway segments. Cottrell's methodology was used to calculate $LMDI_F$ values, and the following equation applied for each urban area (*12*):

$$\mathrm{LMDI}_F = \sum_{i=1}^{m} [\text{congested lane-mi}_i \times \text{congestion duration}_i \text{ (hr)}] \qquad (3)$$

where i is an individual freeway segment, and m is the total number of freeway segments in an urban area.

Choosing Possible Congestion Indicators

It has been generalized that congestion is related to three basic types of variables: travel, supply, and demographic characteristics (*16*). The indicators chosen for this study were composed of one, or a combination of two, of these three basic types of variables. The indicators were chosen with consideration given to data availability, intuitive relation to the causes of congestion, and logical results. With the exception of one, all indicators were extracted or calculated with data from the existing TTI congestion data base. The indicators that were examined in this study are as follows:

Travel characteristics: DVMT, transit trips, and passenger miles of travel.

Supply characteristics: lane miles and transit revenue miles.

Demographic characteristics: population size, population density, registered vehicles, registered vehicles per square mile, and registered vehicles per capita.

Travel-supply characteristics: DVMT per lane mile and AADT/C.

Travel-demographic characteristics: DVMT per square mile, DVMT per registered vehicle, DVMT per capita, transit trips per capita, and, passenger miles per capita.

Supply-demographic characteristics: lane miles per capita, lane miles per square mile, registered vehicles per lane mile, and revenue miles per capita.

The transit indicators were totals for bus and heavy, light, and commuter rail. The other indicators, with the exception of AADT/C, were calculated for both freeways and principal arterial streets.

Examination of Congestion Relationships

There were two basic steps in the examination of the congestion relationships. The first was a graphical comparison of all possible indicators to the three congestion measures. Each indicator for freeways, principal arterial streets, and the total of both was graphed against each congestion measure, with each urban area representing a datum point on a scatterplot. Each graph was then inspected for variability of data points and ease of constructing a best-fit line, whether it be linear or exponential. This graphical comparison gave a sense of the relationship between the indicator and the estimated congestion level.

The second step in examination of the congestion relationships was a limited regression analysis. This analysis determined the coefficient of determination, r^2, a statistical measure that represents the proportion of variability that is accounted for in a relationship. In general, an r^2-value of .5 or greater was interpreted as a close relationship. Although it may be argued that an r^2 value is not a statistically complete treatment, it was thought that the use of a graphical comparison combined with the r^2-value provided enough information to infer whether some type of relation existed.

RESULTS

Estimation of Congestion Level

The congestion level was estimated using the three measures whose equations were presented in the methodology section of this paper. The measures were calculated for the 50 urban areas for 1989, the most recent year for which data were available. The HPMS data base composition prevented determination of $LMDI_F$ for 19 urban areas in California, Connecticut, Florida, Hawaii, Michigan, Ohio (Cleveland only), Oregon, and Washington. In the past, HPMS reporting procedures did not require states to report traffic data for each urban area individually; consequently, these states chose to submit traffic data with several urban areas grouped into one data set. The use of several transportation agencies' data by TTI in the development of their congestion data base prevented similar deficiencies in calculation of the roadway congestion index and the congestion severity index. A summary of the measures and ranking for the 50 urban areas in this study for 1989 may be found in Table 1. It should be noted that the $LMDI_F$ values in Table 1 do not correspond exactly to the $LMDI_F$ values as reported by Cottrell (*12*). Although a significant effort was made, this study was not able to replicate Cottrell's results.

An important criterion for the measures used to estimate congestion level was comparability of results. On comparing the measures, it was discovered that the roadway congestion index and the congestion severity index were closely related, with an r^2-value of .72. Both measures include freeways and principal arterial streets, but the roadway congestion index uses a travel-to-supply ratio and the congestion severity index uses a delay-to-travel ratio. In a graphical comparison (Figure 2), an RCI value of 1.0 was related to a CSI value of 24,000 (vehicle hours per million vehicle miles of travel) by means of a calculated regression line. The relationship between these two measures and the lane-mile duration index was less distinguished but nonetheless comparable (r^2-values of .60 and .45 for the CSI and the RCI, respectively).

Examination of Congestion Relationships

As described earlier, there were two steps in examination of the congestion relationships: a graphical comparison and determination of r^2. The results of these two steps will be presented for the relationships with the highest correlation.

The indicator of DVMT per lane mile, a travel-to-supply ratio, was found to have the highest correlation among all indicators. Because the roadway congestion index uses DVMT per lane mile in a weighted, normalized equation, it was excluded from this comparison. Also, because of the difficulty of combining this indicator for freeways and principal arterial streets without replicating the roadway congestion index, it was analyzed separately for the two different functional classes. A plot of freeway DVMT per lane mile versus congestion level is presented in Figure 3. The r^2-values for this relationship were .68 for the CSI and .45 for the LMDI. The relationship between arterial DVMT per lane mile and the three congestion measures was less pronounced, with r^2-values between .35 and .45.

The indicator with the next highest correlation to congestion level was DVMT per square mile of urban area. Travel (DVMT) was combined for both freeways and principal arterial streets in this examination. It was found that the roadway congestion index and the congestion severity index were most closely related to this indicator (Figure 4). The r^2-values for this relationship were .48 for the RCI but only .27 for the CSI. It was noted throughout the examination that r^2-values were consistently higher for RCI-indicator relationships than for CSI- or LMDI-indicator relationships.

The third indicator that had a significant relationship to congestion level was AADT/C. Because AADT/C is a facility measure, each HPMS freeway segment in an urban area was classified into a congestion range corresponding to the AADT/C value. In previous research Cottrell assumed that conges-

TABLE 1 Summary of Congestion Measures for 50 Urban Areas, 1989 (*11*)

Urban Area	Roadway Congestion Index Value	Roadway Congestion Index Rank	Congestion Severity Index Value	Congestion Severity Index Rank	Lane-Mile Duration Index Value	Lane-Mile Duration Index Rank [a]
Los Angeles CA	1.54	1	39,938	8	- [b]	-
San Fran-Oak CA	1.36	2	46,817	3	-	-
Washington DC	1.36	2	52,882	1	2773	3
Miami FL	1.25	4	42,930	5	-	-
Chicago IL	1.21	5	34,015	13	1820	4
Seattle-Everett WA	1.21	5	38,215	9	-	-
San Diego CA	1.18	7	16,472	35	-	-
Atlanta GA	1.14	8	30,044	17	1813	5
Houston TX	1.13	9	37,612	10	3392	2
New Orleans LA	1.13	9	33,616	14	95	21
New York NY	1.12	11	48,924	2	4381	1
San Jose CA	1.12	11	43,494	4	-	-
Boston MA	1.09	13	40,551	6	999	6
Honolulu HI	1.09	13	34,885	12	-	-
San Bernardino-Riv CA	1.09	13	40,197	7	-	-
Detroit MI	1.08	16	35,923	11	-	-
Norfolk VA	1.08	16	32,788	15	469	9
Portland OR	1.07	18	29,406	18	-	-
Philadelphia PA	1.05	19	28,011	22	347	13
Phoenix AZ	1.03	20	28,482	19	446	10
Tampa FL	1.03	20	22,472	27	-	-
Charlotte NC	1.02	22	28,151	21	-	-
Dallas TX	1.02	22	28,234	20	887	7 [c]
Denver CO	1.01	24	23,546	26	212	15
Sacramento CA	1.01	24	22,107	28	-	-
Baltimore MD	0.99	26	20,442	31	545	8
Orlando FL	0.98	27	32,731	16	-	-
Jacksonville FL	0.97	28	19,538	33	-	-
Milwaukee WI	0.97	28	14,694	38	125	19
Austin TX	0.96	30	23,632	25	435	11
St. Louis MO	0.96	30	20,298	32	2	28
Cleveland OH	0.95	32	10,642	43	-	-
Nashville TN	0.95	32	16,669	34	31	23
Cincinnati OH	0.94	34	11,242	41	2	28
Albuquerque NM	0.91	35	12,565	40	135	18
Memphis TN	0.91	35	8,714	46	2	28
Minn-St. Paul MN	0.90	37	20,710	30	393	12
Ft. Lauderdale FL	0.89	38	23,679	24	-	-
Hartford CT	0.89	38	14,600	39	-	-
San Antonio TX	0.89	38	16,181	37	259	14
Fort Worth TX	0.87	41	22,037	29	-	-
Louisville KY	0.86	42	10,768	42	86	22
Indianapolis IN	0.85	43	5,267	48	29	24
Columbus OH	0.82	44	16,237	36	96	20
Pittsburgh PA	0.82	44	25,235	23	142	17
Salt Lake City UT	0.81	46	9,624	45	29	24
Oklahoma City OK	0.78	47	9,921	44	15	26
El Paso TX	0.74	48	5,241	49	12	27
Kansas City MO	0.72	49	7,494	47	159	16
Corpus Christi TX	0.70	50	3,497	50	0	31

[a] $LMDI_F$ rank is provided to compare relative positions of urban areas within the $LMDI_F$ analysis. $LMDI_F$ rank should not be compared to RCI or CSI rank.
[b] Missing values in this column are due to "grouped" data in HPMS data base.
[c] In the $LMDI_F$ analysis, Dallas and Forth Worth are combined as one metropolitan area.

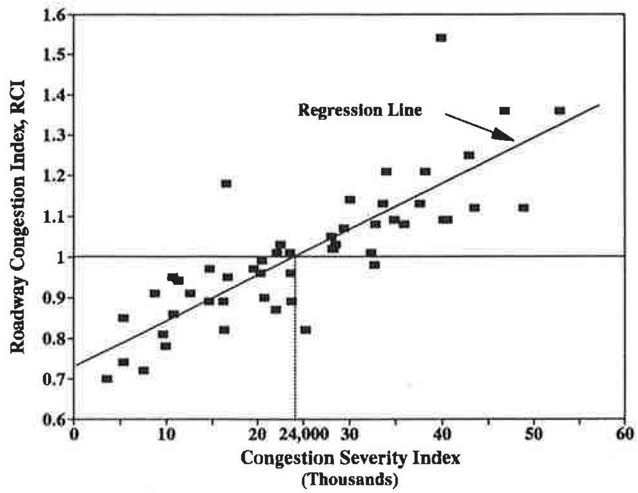

FIGURE 2 Correlation of roadway congestion index and congestion severity index.

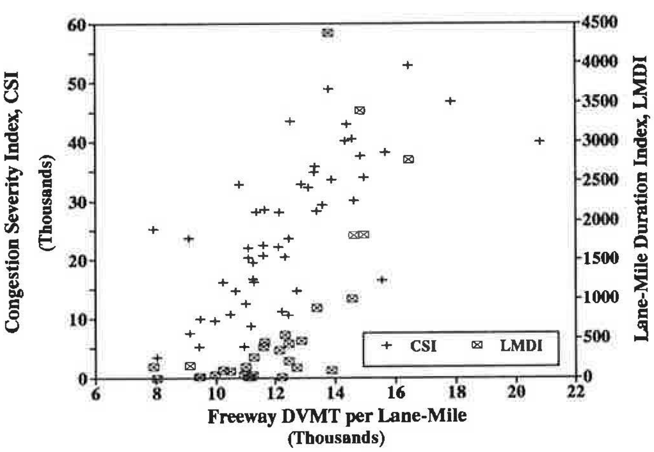

FIGURE 3 DVMT per lane mile versus congestion level.

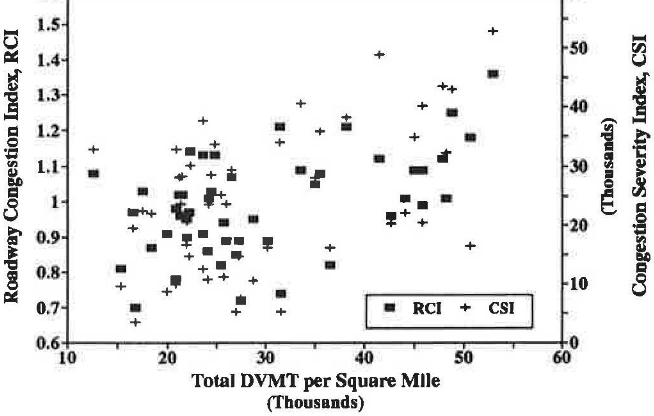

FIGURE 4 DVMT per square mile versus congestion level.

tion occurred on a facility when the AADT/C value was greater than 9 (*12*). It was beyond the scope of this study to determine the accuracy of this value; consequently, for the purposes of this study, an AADT/C value greater than 9 represented congested conditions. The lane miles for all HPMS freeway segments were totaled in the following AADT/C ranges: less than 9, 9 to 11, 11 to 13, 13 to 15, 15 to 17, and greater than 17. Very few urban areas had freeway segments with an AADT/C value higher than 17.

To make a comparison of this indicator to a congestion measure, 31 of the 50 urban areas were grouped according to the RCI value; areas with "grouped" data were excluded. The RCI congestion ranges were 0.70 to 0.85, 0.85 to 0.95, 0.95 to 1.05, 1.05 to 1.20, and greater than 1.20. The comparison is illustrated in Figure 5. It can be seen that, as the percentage of freeway lane miles in higher AADT/C ranges increases, the RCI congestion range also increases. For example, the percentage of lane miles with AADT/C less than 9 (no congestion) is much less in the RCI range of greater than 1.20 (heavy areawide congestion) than in the 0.70 to 0.85 range (none to low areawide congestion). The implication of Figure 5 is that, as the distribution of freeway lane miles shifts toward a higher AADT/C value, the congestion level increases. Because of the nature of this comparison, an r^2-value was not available. The congestion severity index and the lane-mile duration index were not included in this comparison because of the lack of a definition of congestion ranges.

The average AADT/C values for each RCI congestion range are displayed at the bottom of each bar for the respective range. This value was determined by weighting the percentage of lane miles for each AADT/C range by the corresponding average AADT/C value for that AADT/C range. The average AADT/C value at the beginning of areawide congestion (RCI range of 0.95 to 1.05) is 7.1. The initial premise of this particular examination was that congestion on a facility begins at an AADT/C value of 9. The discrepancy in these two numbers—7.1 and 9—may be partly attributed to the translation of AADT/C from a facility measure to an areawide average value.

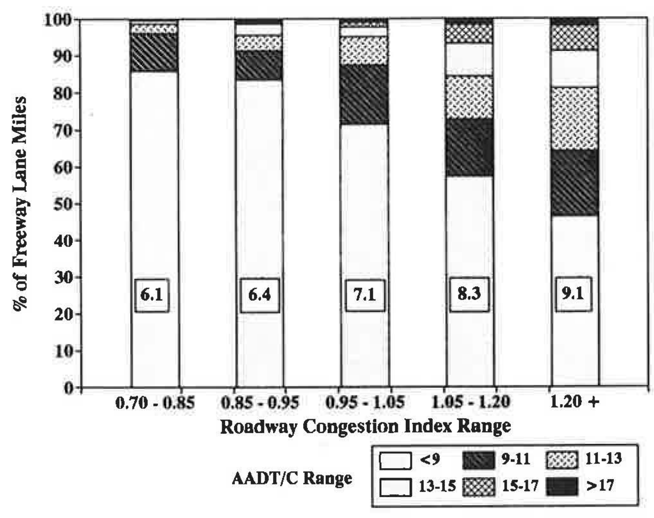

FIGURE 5 Share of freeway lane miles by AADT/C and RCI range.

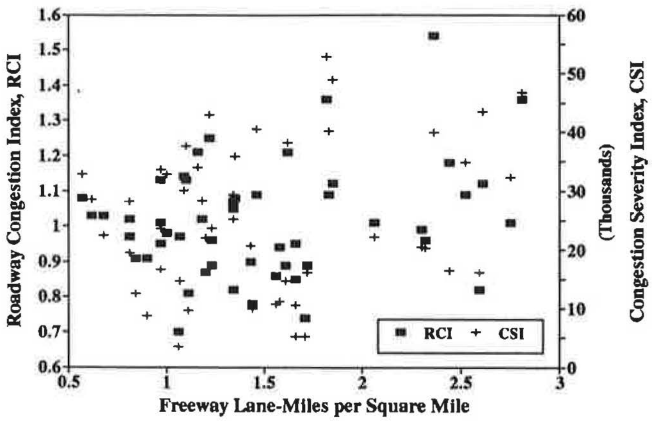

FIGURE 6 Freeway density versus congestion level.

Several indicators had a moderate correlation (r^2-values between .35 and .45) to congestion level. Those indicators are registered vehicles, DVMT, and population density. Most of the transit indicators fared poorly, having r^2-values below .1. Surprisingly, many supply-related indicators had low correlations to the level of congestion. For instance, freeway lane miles per square mile (freeway density) is shown in Figure 6. Freeway density is an indicator used often by automobile clubs and other groups lobbying for construction of new freeway facilities because of congestion. Figure 6 shows that freeway density has a low correlation to congestion level, indicating that there may be other variables that more strongly affect congestion level.

CONCLUSIONS

This report examined the relationships between possible indicators and congestion level as estimated by three congestion measures. The examination included 50 large and medium-sized U.S. urban areas for 1989. The study gathered data from the TTI congestion data base and the HPMS data base.

Indicators of Congestion

This study identified three indicators with a close correlation to congestion level: DVMT per lane mile, DVMT per square mile, and AADT/C for freeways. The indicator of DVMT per lane mile showed the strongest correlation with r^2-values of .68 and .45 for the congestion severity index and the lane-mile duration index, respectively. The indicator of DVMT per square mile had the next highest correlation, with an r^2-value of .48 for the roadway congestion index. It was shown that freeway AADT/C had a clear relationship to the roadway congestion index, although an r^2-value was unobtainable for the type of comparison made.

Several indicators related to roadway supply had low correlations (r^2-values less than .2) to estimated congestion levels. In particular, Figure 6 shows that freeway lane miles per square mile has a low correlation to congestion. The other supply indicators that have a similarly low correlation to congestion include lane miles and lane miles per square mile. It can be concluded safely that, since supply has a low correlation to congestion level, an indicator relying solely on supply characteristics would serve as a poor indicator for areawide congestion levels.

It should be noted that the three indicators with the strongest correlation to congestion level are gauges of the travel intensity for a particular urban area. It is concluded that travel intensity is most directly related to congestion level and would be the most useful type of indicator for areawide congestion measurement purposes. This is not to deny, however, the importance of the effects of roadway supply or demographic factors within an urban area on congestion level.

Congestion Measures

Three congestion measures were used to estimate the congestion level for the urban areas in this study. Two of the measures, the roadway congestion index (using the indicator of DVMT per lane mile) and the congestion severity index (using vehicle delay per million vehicle miles of travel), were found to be very comparable in the results they produced. A regression analysis was used to calculate a best-fit line (Figure 2) through the linearly related data (r^2 = 0.72). Because the roadway congestion index is normalized, an RCI value greater than 1.0 represents the threshold for undesirable areawide congestion. The analysis indicated that this congestion threshold was reasonable; consequently, an RCI value of 1.0 was related to an approximate CSI value of 24,000 using the calculated regression line. It is suggested, then, that a CSI value greater than 24,000 represents undesirable area congestion.

Ideally, congestion measures should provide an accurate representation of congestion levels for a transportation system. Freeways and principal arterial streets are major providers of mobility in urban areas and were included in two of the three measures in this study. Limited data in the HPMS data base and lack of a sound analytical procedure prevented the inclusion of principal arterial streets in the lane-mile duration index (*12*). An illustration of the importance of principal arterial streets is presented in Table 2, in which delay is compared between these two different functional classes.

It can be seen that, for several urban areas, delay (undesirable congestion level) occurs primarily on the freeway sys-

TABLE 2 Delay by Functional Classification for 10 Selected Urban Areas, 1989 (*11*)

	Delay (1000 vehicle-hours)	
Urban Area	Freeways	Principal Arterial Streets
Atlanta GA	133,113 (71 %)	53,928 (29 %)
Chicago IL	246,637 (63 %)	147,008 (37 %)
Detroit MI	165,885 (57 %)	125,612 (43 %)
Houston TX	276,488 (88 %)	38,741 (12 %)
Los Angeles CA	1,119,387 (77 %)	325,949 (23 %)
New York NY	884,915 (72 %)	351,790 (28 %)
Oklahoma City OK	6,116 (38 %)	9,771 (62 %)
Philadelphia PA	66,317 (32 %)	139,422 (68 %)
Phoenix AZ	32,843 (26 %)	95,501 (74 %)
Tampa FL	10,998 (34 %)	20,882 (66 %)

tem. For several areas, however, equal or greater delay occurs on principal arterial streets. This illustrates the importance of principal arterial streets in determination of areawide congestion level.

Further Research

The congestion measures examined in this study attempted to quantify congestion on an areawide basis. For a particular urban area, however, these measures may be somewhat inadequate to quantify the congestion along a particular corridor or route. The increasing use of high-occupancy vehicle (HOV) facilities along highway corridors, in addition to the rapidly developing technologies of intelligent vehicle-highway systems, necessitates the development of congestion measures that accurately reflect travel (car, HOV, bus, and rail) conditions for a particular corridor or urban area.

ACKNOWLEDGMENTS

This research was performed at TTI, Texas A&M University, as a part of the Undergraduate Transportation Engineering Fellows Program. The research was sponsored through the research project entitled *Measuring and Monitoring Urban Mobility in Texas* by the Texas Department of Transportation in cooperation with FHWA, U.S. Department of Transportation.

REFERENCES

1. J. A. Lindley. *Quantification of Urban Freeway Congestion and Analysis of Remedial measures.* Report FHWA/RD-87/052. FHWA, U.S. Department of Transportation, 1986.
2. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
3. *Exhibit E: Highway System Current Level of Service.* California Department of Transportation, District 7 (undated).
4. *The Status of the Nation's Highways: Conditions and Performance.* Report of the Secretary of Transportation to the U.S. Congress. U.S. Government Printing Office, June 1985.
5. *A Policy on Geometric Design of Highways and Streets.* AASHTO, Washington, D.C., 1984.
6. T. J. Lomax and D. L. Christiansen. *Estimates of Relative Mobility in Major Texas Cities.* Research Report 323-1F. Texas Transportation Institute, Texas State Department of Highways and Public Transportation, 1982.
7. T. J. Lomax. *Relative Mobility in Texas Cities, 1975 to 1984.* Research Report 339-8. Texas Transportation Institute, Texas State Department of Highways and Public Transportation, 1986.
8. T. J. Lomax, D. L. Bullard, and J. W. Hanks, Jr. *The Impact of Declining Mobility in Major Texas and Other U.S. Cities.* Research Report 431-1F. Texas Transportation Institute, Texas State Department of Highways and Public Transportation, 1988.
9. J. W. Hanks, Jr., and T. J. Lomax. *Roadway Congestion in Major Urban Areas, 1982 to 1987.* Research Report 1131–2. Texas Transportation Institute, Texas State Department of Highways and Public Transportation, 1989.
10. J. W. Hanks, Jr., and T. J. Lomax. *Roadway Congestion in Major Urbanized Areas, 1982 to 1988.* Research Report 1131–3. Texas Transportation Institute, Texas State Department of Highways and Public Transportation, 1990.
11. J. W. Hanks, Jr., and T. J. Lomax. *1989 Roadway Congestion Estimates and Trends.* Draft Research Report 1131–4. Texas Transportation Institute, Texas State Department of Highways and Public Transportation, 1991.
12. W. D. Cottrell. Measurement of the Extent and Duration of Freeway Congestion in Urbanized Areas. *Proc., ITE 61st Annual Meeting*, Washington, D.C., 1991.
13. *Highway Performance Monitoring System Analytical Process: Volume II—Version 2.1, Technical Manual.* FHWA, U.S. Department of Transportation, 1987.
14. T. E. Lisco. *A Procedure for Predicting Queues and Delays on Expressways in Urban Core Areas.* CTPS Technical Report 36. Feb. 1983.
15. *Highway Performance Monitoring System Field Manual for the Continuing Analytical and Statistical Data Base.* Office of Highway Planning, FHWA, U.S. Department of Transportation, 1984.
16. C. R. Fleet and P. DeCorla-Souza. *Issue 2—Supply and Use of the Nation's Urban Highways.* Highway Planning Technical Report FHWA-ED-89-051. FHWA, U.S. Department of Transportation, Sept. 1989.

Publication of this paper sponsored by Task Force on Transportation Demand Management.